the 10 000 8 19

# UNIVERSITE DES SCIENCES ET TECHNOLOGIES DE LILLE الله على U.F.R. DE BIOLOGIE

Laboratoire d'Ecologie Numérique - UPRES-A 8013 ELICO

Année 1999

n° d'ordre :

#### THESE

pour obtenir le grade de

## DOCTEUR DE L'UNIVERSITE DES SCIENCES ET TECHNOLOGIES DE LILLE

Spécialité : Environnement et Ecosystèmes Marins et Continentaux

par

Patrice CARPENTIER

Caractérisation de structures spatiales et temporelles en écologie. Application aux séries physico-chimiques de qualité des eaux du Réseau National de Bassin en Artois-Picardie et aux peuplements benthiques du site de Gravelines et de la Manche Orientale.

présentée et soutenue publiquement le 8 janvier 1999 devant le jury composé de :

Prof. J.C. Dauvin Dr. P.M.J. Herman Dr. F. Ibañez Prof. S. Frontier Dr. J. Prygiel Dr. A. Leprêtre Université de Lille I Neth. Inst. of Ecology, Yerseke (Pays-Bas) Université Paris VI Université de Lille I Agence de l'Eau Artois-Picardie Université de Lille I Président Rapporteur Rapporteur



50376.

1999 -

489

## SOMMAIRE

I - Séries chronologiques de qualité des eaux douces

1

## Introduction générale

I.1 - Cadre général de l'étude – Introduction.	5
I.2 - Développement et choix d'une méthodologie statistique.	8
I.2.1 Analyse des séries et prévision à court terme	10
I.2.1.a. Régularisation d'une série brute	10
I.2.1.b. Détection des ruptures de tendance	13
I.2.1.c. Tests de tendance	17
I.2.1.d. Prévision	21
I.2.2 Description et typologie des formes des séries temporelles.	22
I.2.2.a. Extraction de la tendance à long terme – Lissage des séries.	23
I.2.2.b. Calcul d'une distance entre séries temporelles	26
I.2.2.c. Lien entre l'indice de co-évolution et le Tau de Kendall.	28
I.2.2.d. Analyse des tableaux de distances par MDS et classification	28
I.2.3 Relation concentration/débit	29
I.3 - Exemples d'applications	32
I.3.1 Evolution des concentrations	32
I.3.2 Flux vers le milieu marin	42
I.3.3 Analyse de co-évolution des profils temporels	44
I.3.3.1 Exemple 1 : co-évolution des séries phosphates de 64 stations du RNB.	44
I.3.3.2 Exemple 2 : co-évolution des différents paramètres dans une station (la Somme à Abbeville)	46
I.3.4 Quelques problèmes rencontrés dans l'évaluation des tendances	48
I.3.4.1 Relation concentration/débit	48
I.3.4.2 Hétérogénéité des tendances saisonnières	53
I.4 – Conclusion	56
Publication jointe :	

Leprêtre, A. Carpentier, P. 1997. Une méthode simple de prévision des tendances appliquée aux séries temporelles de qualité des eaux courantes. *C.R. Acad. Sci. Paris, Sciences de la vie*, 320 : 407-411. 59

## II - Série benthique à long terme de Gravelines (Mer du Nord)

II.1 – Introduction	64
II.1.1 Zone et site d'étude de Gravelines	64
II.1.2 Conditions hydrodynamiques	66
II.1.3 Application d'un modèle hydrodynamique	67
II.1.4 Evolution des conditions climatiques	70
II.2 - Détection des changements de la communauté à Abra Alba	73
II.2.1 Problème posé	73
II.2.2 Analyse de la série benthique	75
II.2.2.1 Méthode	75

II.2.2.2 Application à l'analyse de la série	77
II.2.2.3 Généralisation de la méthode à d'autres métriques	81
II.2.3 - Typologie des profils entre espèces	84
II.3 – Diversité	87
II.3.1 Evolution de la diversité taxonomique sur la série originale	87
II.3.2 Evolution de la diversité à différentes échelles de perception	93
II.3.2.1 Principe et méthode	93
II.3.2.2 Résultats obtenus	94
II.4 - Quelques réflexions sur l'échantillonnage	99
II.4.1 Problèmes d'estimation de la composition quantitative du peuplement	100
II.4.2 Effet de l'effort d'échantillonnage sur l'estimation de la diversité	102
Publication jointe :	
Carpentier, P. Dewarumez, J.M. Leprêtre, A. 1997. Long-term variability of the Abra alba	
community in the southern bight of the North Sea. Oceanol. Acta, 20: 283-290.	105
III – Interpolation spatiale des données benthiques du GDR Manche	
<u>In - Interpolation spatiale des données bentindues du Obit Manene</u>	
III.1 - Introduction - Cadre de l'étude	113
III.2 - Développement d'une méthode d'interpolation des "séries spatiales"	117
III.2.1 Méthode d'interpolation de distributions granulométriques	. 119
III.2.2 Interpolation des données de faune macrobenthique	123
III.2.2.1 Méthode	123

III.2.2.2 Cartographies des espèces et typologies	125
III.2.2.3 Diversité taxonomique	130
III.2.2.4 Répartition spatiale et diversité de groupes écologiques	133
III.3 - Influence du degré de résolution sur la perception de la diversité des	
peuplements benthiques	136
III.4 – Conclusion	142

## IV - Expression numérique de la diversité taxonomique

IV.1- Introduction	144
IV.2 - Données utilisés	145
IV.2.1 Modèles de distribution utilisés	145
IV.2.2 Paramètres choisis pour les modèles	148
IV.3 - Estimation de la Richesse spécifique	150
IV.3.1 Problématique et méthodes d'estimations	150
IV.3.2 Les estimateurs non-paramétriques	154
IV.3.3 Comparaison des performances des estimateurs	157
IV.3.3.1 Simulations réalisées	157
IV.3.3.2 Résultats des simulations	157

IV.3.4 Illustration sur des données réelles	167
IV.3.5 Conclusion sur l'estimation de la richesse	172
IV.4 - Estimation de l'indice H' de Shannon	175
IV.4.1 Quelques rappels	175
IV.4.2 Proposition d'une nouvelle méthode d'estimation de H'vrai : H' CRIC	176
IV.4.2.1 Principe de l'estimation	176
IV.4.2.2 Calcul de l'estimateur H' CRIC	177
IV.4.3 Application à des données entomologiques existantes	180
IV.4.4 Robustesse et efficacité de l'estimateur CRIC	182
IV.4.5 Conclusion sur l'estimation de H'	185
IV.5 - Estimation de la diversité $\beta$	187
IV.5.1 Problème posé	187
IV.5.2 Simulations et résultats	188
IV.6 - Synthèse – Conclusion	192

## Conclusion générale

## **Bibliographie**

#### **ANNEXES**

- Annexe I : rapport de recherche DYSCOP Région Nord-Pas-de-Calais. "Répartition spatiale et évolution temporelle des distributions granulométriques des sédiments dans un secteur côtier. Développement d'une méthode d'amélioration des estimations statistiques". A. Leprêtre, P. Carpentier et G. Chapalain.
- Annexe II : résultats des simulations sur les estimateurs de la richesse spécifique.

Annexe III : résultats des simulations sur l'indice de Shannon et l'estimateur CRIC.

201

194

Introduction générale

## Introduction générale

Développement durable, biodiversité, conservation du patrimoine naturel ou encore changement globaux sont autant de concepts qui se sont développés dans la dernière décennie, traduisant une prise de conscience de la nécessité de comprendre et surveiller l'évolution des écosystèmes. Mais pour répondre à cette demande sociale, manifestée entre autres par l'élaboration de "grands principes" (à Rio en 1992 par exemple) et par l'établissement de normes et de chartes diverses destinées à préserver l'environnement, il devient crucial de disposer d'outils de mesure ou d'observation et d'outils de synthèse et de valorisation des données en vue d'assurer un "monitoring" rigoureux des milieux continentaux ou marins. Des séries de données considérables commencent à exister et à s'étoffer d'année en année grâce à la mise en place d'observatoires ou de réseaux de surveillance. L'exploitation de ces ensembles de données et la capacité d'en réaliser des synthèses deviennent par conséquent des obligations pour des organismes chargés d'opérer les suivis. Ces données sont aussi d'un intérêt majeur pour la recherche en écologie fondamentale, qui dispose au travers de ces réseaux, de mesures et d'observations denses et permanentes que les seuls organismes de recherche n'auraient pas pu financer. Dans cette perspective de valorisation des grands jeux de données, l'écologie numérique doit avoir un rôle important, en s'intercalant entre la recherche fondamentale et la recherche appliquée.

L'écologiste numérique ne doit cependant pas se cantonner à être le prestataire de service, apportant au gestionnaire des réponses synthétiques et au fondamentaliste des méthodes d'analyse des données "clé en main", au risque de n'être plus qu'un développeur, privé de thématique de recherche.

En effet, même si l'outil - en l'occurrence l'outil numérique – a souvent été considéré comme un moyen d'exprimer ou de mettre en œuvre les idées ou réflexions théoriques, on doit admettre maintenant que l'outil puissant que constitue l'ensemble "ordinateur + méthodes numériques, statistiques ou mathématiques" est générateur d'idées nouvelles ou tout au moins permet de mettre à l'épreuve, grâce à sa puissance de calcul, des idées qui ne peuvent être vérifiées par les seules méthodes formelles.

Le travail que nous avons réalisé s'insère dans le cadre unitaire de l'écologie numérique, même si les matériaux ou données sont diversifiés et correspondent à des milieux variés (continentaux et marins) et des objets variés (mesures physico-chimiques, observations d'espèces animales marines ou terrestres, etc...). Le fait d'avoir travaillé sur des données de nature très différentes tient aussi à l'évolution de thématiques au sein du Laboratoire d'Ecologie Numérique au cours de des dernières années.

Dans un premier temps, ce travail a été démarré suite à un DEA, portant sur le suivi à long terme des peuplement benthiques de la région de Gravelines, et réalisé dans le cadre du Programme National d'Océanographie Côtière et plus précisément dans l'opération "Séries à long terme" (PNOC-SLT). Il était alors envisagé de s'intéresser à l'incidence d'événements ponctuels sur l'évolution des communautés benthiques en couplant des observations bionomiques avec des mesures et/ou de la modélisation hydrodynamique à haute fréquence.

Les moyens n'ayant finalement pas pu être réunis pour mener à bien ce projet, une opportunité de travailler sur d'autres séries chronologiques s'est présentée avec la demande formulée par l'Agence de l'Eau Artois-Picardie (AEAP) de valoriser les jeux de données constitués depuis une vingtaine d'années au sein du Réseau National de Bassin (RNB). Cette étude sur les séries temporelles de qualité physico-chimique des eaux douces du bassin Artois-Picardie présentait l'intérêt de rester cohérente avec l'opération PNOC-SLT du point de vue des méthodes d'analyse des données à appliquer ou à développer. Un autre intérêt de ces séries, en rapport avec les thématiques de l'URA 1363, devenue depuis UPRES-A 8013 "ELICO", était de pouvoir évaluer à partir de ces séries les apports en azote et phosphore du bassin versant vers le milieu marin.

Nous avons donc, à côté de cette activité sur les séries de qualité des eaux, continué à nous intéresser à la série benthiques de Gravelines et à chercher à mettre en évidence les tendances d'évolution de cette série multispécifique. Parallèlement, nous avons également abordé, dans la continuité du travail de thèse de Laura Sanvicente-Añorve (1995), le traitement des données benthiques du GDR Manche. Les problématiques et les aspects méthodologiques liés à ces "séries spatiales" présentent une forte analogie avec les séries temporelles de Gravelines.

Enfin, ces deux jeux de données ont été utilisés pour une étude portant sur la diversité et les distributions de fréquences d'espèces, problématique impulsée par Serge Frontier, et qui a toujours dominé au Laboratoire d'Ecologie Numérique depuis sa création en 1981. Nous nous sommes donc intéressé, au vu des questions soulevées par les résultats obtenus à partir de différents jeux de données, aux performances et aux propriétés de différents estimateurs numérique de la diversité.

La première partie de ce travail porte sur l'étude des séries chronologiques de qualité des eaux du bassin Artois-Picardie. Ces séries constituent un large jeu de données (près de 900 séries, d'une vingtaine d'années, avec une fréquence d'observation mensuelle), qui a d'abord nécessité le choix et la mise en place d'une méthodologie statistique "simple" et robuste pouvant s'appliquer à l'ensemble de ces séries. L'objectif pragmatique de cette étude est d'abord de décrire les évolutions pour aboutir ensuite à une prédiction à court et moyen termes. Actuellement, une des préoccupations essentielles de l'Agence de l'Eau Artois-Picardie étant le problème de l'eutrophisation des cours d'eau, nous nous sommes focalisé sur les différentes formes de l'azote et du phosphore, auxquelles se sont ajoutés les paramètres relatifs à l'oxygène (concentration, DBO5, DCO) et les matières en suspension. En complément de ce suivi des concentrations de différents paramètres, nous avons également envisagé ces séries en terme de flux, afin d'estimer les apports du bassin Artois-Picardie vers le milieu marin, soit directement sur le littoral français de la Manche et de la Mer du Nord, soit indirectement, par la Belgique et plus particulièrement par le bassin de l'Escaut. Enfin, nous évoquerons de manière plus détaillée certains problèmes d'ordre méthodologique qui peuvent fausser l'interprétation des tendances et la prévision (effet des débits, hétérogénéité des tendances).

Les deuxième et troisième chapitres de ce mémoire concernent l'étude des peuplements benthiques de la Manche orientale et de la Mer du Nord.

Le deuxième chapitre est consacré au suivi à long terme, de 1978 à 1992, du peuplement des sables fins envasés à *Abra alba* sur le site Gravelines (sud de la Mer du Nord). L'objectif est de mettre en évidence les modifications de la composition quantitative de ce peuplement, et de les mettre en relation avec les variations de facteurs environnementaux, plus particulièrement climatiques. Nous proposons une méthode, qui se base sur les résultats d'une analyse factorielle, et qui permet de visualiser, sous la forme d'une série chronologique univariable, les changements de la structure multispécifique du peuplement. Nous avons ensuite étudié l'évolution temporelle de la diversité de ce peuplement (problématique émergeante au sein du PNOC-SLT), au travers de différents estimateurs. L'objectif principal a été de mettre en évidence l'influence de l'échelle temporelle d'observation sur la perception des tendances de cette diversité. Enfin, nous aborderons, sur la base de quelques relevés de la série, les problèmes liés à la micro-hétérogénéité spatiale de la répartition des individus, ce qui permet en particulier de souligner la difficulté de l'échantillonnage de ce peuplement.

Le chapitre III porte sur les données benthiques du GDR Manche et plus particulièrement sur la partie orientale de la Manche et le sud de la Mer du Nord. Cette contribution fait suite à la thèse de Sanvicente-Añorve réalisée dans le laboratoire (1995). Nous nous sommes penché sur l'exploitation des données spatiales, en nous appuyant sur la base de données qui avait été développée au sein du laboratoire, en collaboration avec Louis Cabioch (Station Biologique de Roscoff). Dans un premier temps, nous avons d'abord appliqué aux données granulométriques une méthode d'interpolation développée dans le cadre du programme DYSCOP sur le site atelier de Wissant. En nous basant sur le principe de cette méthode, nous proposons dans un deuxième temps une méthode d'interpolation des données de la macrofaune benthique qui prend en compte la notion de liaison entre espèces. L'interpolation des données sur un maillage régulier permet la réalisation de représentations cartographiques ainsi que celle de typologies à grande échelle des communautés benthiques. Nous avons aussi illustré les possibilités qu'offre la base de données constituée sur le GDR Manche, par une approche "fonctionnelle" de ces données, basée sur la distribution des régimes alimentaires. Comme pour la série Gravelines, nous nous sommes intéressé à la diversité et à l'influence de l'échelle spatiale d'observation sur la perception des structures.

Dans les applications portant sur les deux jeux de données précédents, et au vu d'autres données détenues par le laboratoire (entomologiques, phytoplanctoniques, etc...), il est apparu nécessaire de mieux connaître les performances et propriétés des estimateurs numériques de la diversité. Le chapitre IV porte donc sur l'étude de deux des estimateurs les plus utilisés : la richesse spécifique et l'indice de Shannon H'. Nous avons réalisé des simulations d'échantillons dans différents peuplements répondant à des distributions de fréquences théoriques, en faisant varier l'effort d'échantillonnage. Ces simulations nous ont d'abord permis d'évaluer le biais introduit par l'échantillonnage sur l'estimation de la richesse spécifique, ainsi que l'efficacité des estimateurs non-paramétriques (Jackknife et Bootstrap) de cette richesse. La même démarche est appliqué à l'indice de Shannon. Nous proposons également un nouvel estimateur de l'indice de Shannon, par l'intermédiaire des Courbes Rangs-Informations Cumulées. Nous envisagerons enfin, la répercussion du biais d'échantillonnage de H' sur l'estimation de la diversité  $\beta$ .

# Chapitre I

Séries chronologiques de qualité des eaux douces

### **I.1 - Cadre général de l'étude – Introduction.**

Dans l'optique d'un suivi de l'évolution de la qualité des eaux, des réseaux de stations d'observation ont été mis en place en France, sous la responsabilité des 6 Agences de l'Eau et sous la tutelle du Ministère de l'Environnement. Dans ces stations, choisies en fonction des problèmes spécifiques de chacun des cours d'eau, des prélèvements sont réalisés périodiquement en vue d'analyses physico-chimiques, et éventuellement d'analyses biologiques et bactériologiques. En ce qui concerne les eaux continentales superficielles, ce suivi correspond à une obligation légale prévue par l'article 3 de la Loi sur l'Eau du 16 décembre 1964 relative au "régime et à la répartition des eaux et à la lutte contre la pollution". Cette loi précisait que les décrets fixeraient d'une part, les spécifications techniques et les critères physiques, chimiques, biologiques et bactériologiques auxquels devraient répondre les cours d'eau, et d'autre part le délai dans lequel la qualité de chaque milieu récepteur devrait satisfaire les intérêts définis à l'article I : "alimentation en eau potable des populations et à la santé publique,...", "l'agriculture, l'industrie, les transports et toutes autres activités humaines d'intérêt général", "la vie biologique du milieu récepteur et spécialement de la faune piscicole ainsi que les loisirs, les sports nautiques et la protection des sites".

En 1971, un groupe interministériel a mis au point un système d'évaluation de la qualité de l'eau, donnant lieu à une grille de qualité comportant 5 classes (1A, 1B, 2, 3 et 4) correspondant à des seuils pour différents types d'utilisation de l'eau. Cette grille, qualifiée de "grille multi-usages", fut adaptée aux conditions locales. Ainsi, dans la grille Artois-Picardie, la qualité 1 (1A et 1B) s'adresse aux eaux permettant l'abreuvage des animaux et la fabrication d'eau potable avec traitement simple alors qu'à l'autre extrême, la classe de qualité 4 ne permet guère plus que la navigation. Le tableau ci-après (tableau 1.1) présente la grille de qualité des eaux actuellement en application en ce qui concerne l'azote et la phosphore. De 1971 à 1987, cette surveillance s'est effectuée dans le cadre de la procédure d'Inventaire National de Pollution (I.N.P), qui fixait les modalités d'une campagne d'inventaire d'une durée de douze mois. Cet inventaire a donc été réalisé tous les 5 ans (en 1971, 1976 et 1981) avec une surveillance allégée entre les campagnes. La surveillance a été accentuée (un prélèvement par mois) à partir de 1981 et, en 1987, cette procédure a été remplacée par le Réseau National de Bassin (R.N.B.) qui comprend plus de 900 stations de mesures réparties sur l'ensemble du territoire national et géré par les Agences de l'Eau.

Cette politique de surveillance a permis la constitution de longues séries chronologiques de données physico-chimiques qui permettent maintenant de retracer et analyser l'évolution de la qualité des eaux dans les différentes stations du RNB, et peuvent être utilisées également dans un but prévisionnel. La prévision des évolutions permet d'une part de fixer des objectifs de qualité pour les cours d'eau et de fixer les redevances des industries et collectivités locales, d'autre part de décider des priorités d'action en matière de restauration de la qualité (choix d'installation de station d'épuration en particulier).

Formes de l'azote et du phosphore	Qualité 1A 'Situation normale'	Qualité 1B 'Pollution modérée'	Qualité 2 'Pollution nette'	Qualité 3 'Pollution importante'	Qualité 4 'Pollution excessive'
<b>NH</b> <sub>4</sub> <sup>+</sup> (mg/l)	< 0.1	0.1 à 0.5	0.5 à 2	2 à 8	> 8
<b>NO</b> <sub>2</sub> <sup>-</sup> (mg/l)	< 0.1	0.1 à 0.3	0.3 à 1	1 à 2	> 2
<b>NO</b> <sub>3</sub> <sup>-</sup> (mg/l)	< 5	5 à 25	25 à 50	50 à 80	> 80
N kjeldahl	< 1	1 à 2	2 à 3	3 à 10	> 10
<b>PO</b> <sub>4</sub> <sup>-</sup> (mg/l)	< 0.2	0.2 à 0.5	0.5 à 1	1 à 2	> 2
P total (mg/l)	< 0.1	0.1 à 0.3	0.3 à 0.6	0.6 à 1	> 1

Tableau 1.1 : Classes de qualité Azote et Phosphore en France

<u>Commentaire du tableau</u> : pour cartographier les différents niveaux de pollution dus aux formes de l'azote et du phosphore, cette grille a été adoptée le 26 juin 1990 en concertation avec le Ministère de l'Environnement, les 6 Agences de l'Eau et les Délégués de Bassin. C'est une harmonisation nationale des différents grilles en usage dans les bassins. Les bornes ne correspondent pas à des moyennes mais à des valeurs atteintes 90% du temps.

Le domaine d'intervention de l'Agence de l'Eau Artois-Picardie couvre le Nord, le Pas-de-Calais et la partie des départements de l'Aisne et de la Somme correspondant au bassin versant du fleuve Somme (fig. 1.1). Jusqu'à une date récente, la politique de l'Agence de l'Eau Artois-Picardie était plus particulièrement axée sur la lutte contre la pollution organique, pollution qui masquait très souvent l'eutrophisation sous-jacente. Le travail est maintenant focalisé sur cette eutrophisation (en particulier sur celle liée au phytoplancton) et la lutte contre ces nuisances principalement liées à l'azote et au phosphore. Le rôle de l'azote et surtout du phosphore comme élément limitant de l'eutrophisation a été décrit depuis longtemps (Hutchinson, 1957). Plus récemment, il a été montré (OCDE, 1982) l'existence d'une forte corrélation entre la chlorophylle et le phosphore, particulièrement sous sa forme orthophosphates. En l'état actuel, aucune station d'épuration ne traitant le phosphore, il est urgent de fixer des priorités d'implantations de stations effectuant ce type de traitement. Dans un premier temps, c'est donc sur les différentes formes de l'azote  $(NH_4^+, NO_2^-, NO_3^-)$ et N total Kjeldahl) et sur le phosphore, plus spécifiquement les orthophosphates  $(PO_4^{-3})$ , qu'a porté l'étude entreprise en collaboration par l'Agence de l'Eau Artois-Picardie et le Laboratoire d'Ecologie Numérique de l'Université des Sciences et Technologies de Lille. Sont venus s'ajouter dans une deuxième phase, les paramètres concernant l'oxygène (concentration et pourcentage de saturation, DBO<sub>5</sub> et DCO) ainsi que les matières en suspension.



Figure 1.1 : Carte du Bassin Artois-Picardie et localisation des stations du Réseau National de Bassin (RNB).

L'étude réalisée (Leprêtre et Carpentier, 1997a), dont les principaux résultats sont présentés ici, porte sur les 67 stations du Réseau National de Bassin du bassin Artois-Picardie (Fig. 1.1). En plus des dix paramètres directement mesurés évoqués ci-dessus, trois autres paramètres sont calculés : l'azote total, le phosphore total et le rapport N/P, indicateur d'une potentielle eutrophisation. On a donc au total 871 (67x13) séries chronologiques portant sur une période de 1972 à 1995. Peu de séries sont réellement complètes sur toute cette période, mais la presque totalité présente des observations de fréquence mensuelle depuis 1980.

Les objectifs de cette étude sont d'abord de représenter et d'évaluer les évolutions des concentrations en azote et phosphore à l'échelle du bassin Artois-Picardie. Une deuxième étape est de mettre en évidence des changements et des ruptures dans ces évolutions, afin de les mettre en relation avec des événements et des aménagements susceptibles de modifier la qualité des eaux. Une suite logique de ce travail est de réaliser ensuite des prévisions à court ou moyen termes de ces évolutions.

Enfin un dernier objectif, directement en rapport avec les thématique de l'unité de recherches "ELICO" (UPRES-A 8013), est d'évaluer les flux et les exportations vers le domaine littoral. Le bassin Artois-Picardie est en effet une source majeure d'apport au milieu marin de la Manche et de la Mer du Nord. Les eaux douces de la région ont du coté ouest, une contribution importante à la constitution du fleuve côtier (Brylinski *et al.* 1991). Côté est, via la Belgique, les apports du Bassin Minier et de l'agglomération Lilloise vont rejoindre le bassin de l'Escaut, élément fondamental dans le fonctionnement des écosystèmes de la baie sud de la Mer du Nord.

## I.2 - Développement et choix d'une méthodologie statistique.

Le traitement des séries chronologiques de qualité des eaux se heurte à de nombreuses difficultés qui rendent l'application de la plupart des méthodes statistiques classiques impossible ou, au mieux, sujette à caution. Les principaux problèmes rencontrés sont les suivants :

1) les dates de prélèvement sont irrégulièrement espacées et l'échantillonnage est d'inégale densité au long de la période considérée.

2) les séries sont souvent courtes : quelques années avec, au mieux, une fréquence d'observation mensuelle.

3) il y a une forte variabilité naturelle des paramètres<sup>1</sup> mesurés (donc un faible rapport signal/bruit).

4) la distribution statistique des paramètres mesurés s'écarte souvent de la normalité et est principalement caractérisée par une forte asymétrie, une variabilité non homogène et par le mélange de différents types de distributions.

5) il existe une forte dépendance (ou autocorrélation) entre les observations.

<sup>&</sup>lt;sup>1</sup> Le terme de "paramètre" est utilisé ici pour désigner les variables physico-chimiques mesurées pour rendre compte de la qualité de l'eau et non pour désigner les estimateurs statistiques.

6) les paramètres étudiés présentent souvent des variations saisonnières marquées.

7) certaines covariables, comme par exemple le débit et la température, ont un effet important sur les variations des paramètres considérés.

8) les valeurs extrêmes ou aberrantes sont nombreuses.

9) il existe souvent une forte proportion de données "censurées" (traduction littérale du terme anglo-saxon "censored data"), c'est-à-dire des valeurs affectées par des limites de détection analytiques et reportées comme étant inférieures à ces limites.

10) à l'approche de ces seuils de détection, les valeurs sont souvent fortement discrétisées.

11) il est fréquent que des changements dans les méthodes de mesures (modification de protocoles ou d'appareils, ...) se produisent au cours de la période considérée.

12) certaines données ne sont pas fiables à cause de problèmes de terrain, de laboratoire ou de transcription.

Même si, fort heureusement, tous ces problèmes ne sont pas systématiquement présents dans chaque série chronologique de qualité des eaux, certains d'entre eux sont suffisamment fréquents pour qu'on ne puisse en faire abstraction. Il a donc été nécessaire de choisir une démarche ainsi qu'un ensemble de méthodes statistiques adapté aux particularités des données. La méthodologie présentée ici permet de prendre en compte l'existence de seuils de détection des méthodes analytiques qui, pour certains paramètres physico-chimiques et dans de nombreuses séries, entraînent une discrétisation des valeurs faibles. Elle permet également la régularisation des séries chronologiques irrégulières, la détection de la dernière rupture de tendance évolutive et la prévision d'une tendance à court ou moyen terme.

#### I.2.1 Analyse des séries et prévision à court terme

#### I.2.1.a. Régularisation d'une série brute

Pourquoi régulariser ?

Les méthodes classiques d'analyse des séries chronologiques et les méthodes statistiques de prévision nécessitent pour la plupart de disposer d'observations réalisées à des intervalles de temps réguliers.

Dans l'ensemble des séries de données disponibles, la densité de l'échantillonnage est variable dans le temps. Il en résulte que certaines périodes sont plus fortement représentées que d'autres dans les séries. Cette irrégularité du suivi a trois origines principales :

1) le changement de l'effort d'échantillonnage, plus important dans les dernières années qu'au début de la surveillance ;

2) la réalisation d'études spécifiques a conduit à un échantillonnage plus important sur certains sites mais sur une courte période ;

3) la non réalisation fortuite de certaines observations due à des problèmes de personnel, de matériel ou à des problèmes analytiques.

La comparaison entre stations peut également poser problème. Certaines stations sont suivies de manières plus intensives que d'autres :

- soit pour des raisons logistiques (techniques, personnel)

- soit de façon délibérée car ces stations sont considérées comme plus sensibles à d'éventuelles perturbations ou susceptibles de présenter une variabilité plus importante.

Enfin, il est impossible de réaliser simultanément les prélèvements sur la totalité du bassin ; les dates de prélèvement ne sont donc pas les mêmes pour toutes les stations.

En conclusion, la régularisation est une étape nécessaire :

- pour assurer l'application de méthodes statistiques d'analyse et de prévision qui nécessitent un pas constant ;

- pour équilibrer l'échantillonnage dans le temps, en vue de l'application des méthodes qui traitent de l'ensemble des données (régressions et ordinations) et éviter ainsi l'introduction d'un biais dans les résultats par une représentation non équitable des différentes périodes de l'étude ;

10

- pour permettre la comparaison entre paramètres et entre stations en réalisant des estimations sur des échelles de temps équivalentes.

#### Comment régulariser ?

La méthode proposée (Leprêtre et Carpentier, 1997b, *cf.* annexe) est dérivée de celle proposée par Ibanez (1991)<sup>2</sup>. Une fonction de pondération a été ajoutée de façon à mieux prendre en compte les observations les plus proches et éviter ainsi un trop fort effet de lissage.

Pour un paramètre physico-chimique donné, la série brute est composée de n valeurs  $x_i$ mesurées à des temps irréguliers  $t_i$ . Une valeur interpolée  $X_j$  est calculée pour des intervalles réguliers aux temps  $T_j$  (avec j = 1 à N). On fait glisser une fenêtre de largeur L sur toute la période qui contient la série brute (fig. 2). Si à un temps  $T_j$  donné correspond exactement un temps  $t_i$  de la série brute, alors la valeur  $x_i$  correspondante est retenue ( $X_j = x_i$ ) sans recourir à l'interpolation. Sinon, le calcul d'une valeur interpolée  $X_j$  (figure 1.2) est donné par:

$$X_{j} = \frac{1}{L} \cdot \frac{1}{\sum_{i=1}^{n} w_{i}} \cdot \sum_{i=1}^{n} x_{i} \cdot F_{i} \cdot w_{i}$$

avec L =largeur de la fenêtre utilisée, centrée sur T<sub>i</sub>

 $Fi = \text{largeur de l'intervalle} \left[ \frac{t_i + t_{i+1}}{2}, \frac{t_{i-1} + t_i}{2} \right] \text{ contenu dans la fenêtre,}$ tel que  $\sum_{Tj-L_2}^{Tj+L_2} F_i = L$ et  $w_i = \frac{1}{d_i} = \frac{1}{|t_i - T_j|}$  à un temps Tj donné.

<sup>&</sup>lt;sup>2</sup> Ibanez fait référence pour la méthode qu'il propose aux travaux de Fox et Brown (1965). Néanmoins, aucune méthode de ce type n'est évoquée dans la publication en question qui ne présente que des moyennes mobiles pondérées classiques à 15 ou 21 termes.

Fox, W.T., Brown, J.A. 1965. The use of time-trend analysis for the environmental interpretation of limestones. *Journal of Geology*, 73: 510-518.



Figure 1.2 : Régularisation d'une série brute de données  $x_i$  en une série régulière de valeurs  $X_j$  au moyen d'une fenêtre de largeur L centrée sur  $X_j$ .

Le coefficient  $F_i$  permet d'affecter un poids à la valeur  $x_i$  en fonction de l'intervalle de temps qui la sépare des valeurs  $x_{i-1}$  et  $x_{i+1}$ . Ce coefficient s'annule en dehors de la fenêtre. Une valeur  $x_i$ , située en dehors de la fenêtre peut contribuer partiellement à l'interpolation si le centre de l'intervalle  $[t_{i-1}, t_i]$  est compris dans la fenêtre. De même, une valeur  $x_i$  comprise dans la fenêtre peut ne contribuer que partiellement si une partie de l'intervalle  $F_i$ correspondant se trouve en dehors de la fenêtre d'interpolation.

Le coefficient  $w_i$  est un coefficient de pondération de  $x_i$  suivant l'inverse de la distance  $d_i$  qui sépare le temps  $t_i$  du centre de la fenêtre  $T_j$ .

Dans les applications réalisées sur le Réseau National de Bassin, une fenêtre de L = 6mois et un intervalle de temps de 1 mois entre les valeurs interpolées  $X_j$  ont été choisis. Un exemple de cette procédure de régularisation, réalisé sur les phosphates mesurés sur la Sambre à Jeumont (station RNB n°4000), est présenté à la figure 1.3.



#### I.2.1.b. Détection des ruptures de tendance

#### Pourquoi détecter les ruptures ?

Dans la réalisation de prévisions à court et moyen termes sur les séries du RNB, la forte variabilité naturelle des données et la faible longueur des séries ne permettent pas l'utilisation des méthodes couramment employée dans d'autres domaines, comme par exemple les méthodes de type Box-Jenkins (Box et Jenkins, 1976). Il n'est pas non plus raisonnable d'envisager la prévision en utilisant des fonctions non-linéaires (polynomiales par exemple). En effet, même si de telles fonctions s'ajustent bien aux données existantes, il est difficile de justifier l'existence et, encore plus, l'extrapolation de relations aussi complexes dans le comportement des paramètres étudiés.

L'observation des séries montre que les évolutions peuvent être assimilées à des tendances linéaires ou peuvent s'y ramener moyennant des transformations simples, rendant donc possible l'application de diverses méthodes de régression linéaire pour la prévision. Cependant, les séries étudiées ne présentent que rarement une tendance strictement monotone et il est fréquent d'observer des modifications et des inversions des tendances. Ces changements sont liés à "l'histoire" propre à chaque série, dépendant en particulier des actions

et aménagements opérés sur le bassin versant : mise en service ou modification de stations d'épuration, installation ou arrêt d'une activité industrielle, modification des rejets urbains, etc. Les profils des séries apparaissent donc le plus souvent comme le reflet d'une succession de différentes périodes présentant chacune une tendance linéaire (ou une absence de tendance) différente. Cette succession ne se fait pas de façon régulière. La tendance observée au cours d'une nouvelle période exprime un changement dans le fonctionnement du système. Détecter les changements de tendance apporte ainsi une information sur l'histoire du cours d'eau, sur les événements qui s'y sont produits.

Une nouvelle tendance caractérise donc une nouvelle situation qui ne correspond plus à celle de la période qui précède. La prévision ne peut pas se faire en prenant en compte une situation passée, qui n'a plus rien à voir avec les conditions actuelles d'évolution des paramètres. Il convient donc d'identifier la dernière partie de la série qui présente une tendance homogène monotone, et de réaliser la prévision uniquement en se basant sur cette dernière partie.

#### Comment détecter les ruptures ?

Dans ce but, on utilise une méthode basée sur les sommes cumulées pour identifier les ruptures de tendance dans les séries. La fonction Cusum est une méthode graphique développée initialement pour le contrôle de la qualité en milieu industriel (Page, 1961). Elle a été employée en hydrologie (Lettenmaier, 1976 ; Cluis, 1983) et plus récemment en océanographie (Ibanez *et al.*, 1993).

A chaque temps *Tj* de la série régularisée, on associe la fonction suivante:

$$Cusum(T_j) = \sum_{j=1}^{T_j} X_j - \frac{T_j}{k} \sum_{j=1}^k X_j \qquad \text{avec } 2 \le k \le N$$

En prenant k = N, on obtient  $Cusum(T_j) = \sum_{j=1}^{T_j} X_j - T_j \cdot \overline{X}$ 

La forme de la courbe obtenue permet de déterminer si une tendance existe et si cette tendance est continue ou discontinue. Le graphique s'interprète de la manière suivante:

- si la série ne montre pas de tendance particulière, alors Cusum(T<sub>j</sub>) reste voisin de zéro et la courbe reste proche de l'axe des abscisses (fig. 1.4 (a) et (b) ).
- si la série subit des changements, la courbe Cusum s'éloigne de l'axe des abscisses.
   Plusieurs cas de figures peuvent être alors envisagés :

si cette série présente des paliers, c'est-à-dire une succession de périodes stationnaires centrées sur des moyennes différentes, la courbe *Cusum* est composée de segments de droites dont les pentes dépendent de l'écart entre la moyenne du palier et la moyenne globale (fig. 1.4 (c) et (d)). Les changements de pentes correspondent aux changements de paliers ;

- si cette série présente une tendance monotone linéaire de pente a, la forme de la courbe *Cusum* est une parabole (*c.f.* démonstration en encadré) (fig. 1.4 (e) et (f)). L'équation de cette parabole est de la forme :

$$Cusum (T) = \frac{a}{2} \cdot T^2 + \frac{a \cdot N}{2} \cdot T$$

En effet, pour la fonction	$Cusum (T) = \sum_{i=1}^{T} X_i - T \cdot \overline{X}$
	si on a $X_i = a \cdot t_i + b$ avec $t_i = 1 a N$
	on a alors :
	Cusum (T) = $\sum_{i=1}^{T} (a \cdot t_i + b) - \frac{T}{N} \cdot \sum_{i=1}^{N} (a \cdot t_i + b)$
	$= a \sum_{i=1}^{T} t_i + T \cdot b - \frac{a \cdot T}{N} \cdot \sum_{i=1}^{N} t_i - T \cdot b$
i	$= a \cdot \frac{T(T+1)}{2} - \frac{a \cdot T}{N} \cdot \frac{N(N+1)}{2}$
	$=\frac{a}{2}\cdot T^{2}+\frac{a\cdot T}{2}-\frac{a\cdot T\cdot (N+1)}{2}$
	$=\frac{a}{2}\cdot T^{2}+T\cdot\left(\frac{a}{2}-\frac{a\cdot N+a}{2}\right)$
donc, finalement ,	Cusum (T) = $\frac{a}{2} \cdot T^2 + \frac{a \cdot N}{2} \cdot T$

- si cette série est composée d'une succession de périodes de tendances différentes, la courbe *Cusum* est formée d'une suite de "paraboles" (fig 1.4 (g) et (h) ). Les changements de pentes correspondent alors aux points d'inflexion de la courbe Cusum.

Notons cependant que pour des séries particulièrement bruitées, il peut être difficile de distinguer ces différents cas schématiques évoqués pour les formes des courbes *Cusum*.



Figure 1.4 : Exemples de séries (à gauche) et courbes Cusum associées (à droite).

#### I.2.1.c. Tests de tendance

#### • Significativité des tendances

Quand il est utilisé sur une série chronologique dans le but de tester l'existence d'une tendance monotone, le coefficient de corrélation de Kendall (Kendall, 1938) est alors qualifié de test de Mann-Kendall (Mann, 1945 ; Kendall, 1975). Il existe de nombreuses références à son utilisation en hydrologie (par exemple : Hirsch *et al.*, 1982 ; Hugues et Millard, 1988 ; Zetterqvist, 1988 ; El-Shaarawi et Niculescu, 1992).

Le calcul du coefficient de corrélation non-paramétrique de Kendall se fonde sur les signes des produits des différences entre toutes les observations de deux séries :

Soit une fonction  $signe(\theta)$  qui prend les valeurs suivantes en fonction d'une expression  $\theta$ :

signe( $\theta$ ) = -1 si  $\theta$  < 0 signe( $\theta$ ) = 0 si  $\theta$  = 0 signe( $\theta$ ) = 1 si  $\theta$  > 0

Soit deux variables :

$$X: x_i; i = 1 a n$$
$$Y: y_i; i = 1 a n$$

on a alors  $\mathcal{T}$ , coefficient de corrélation entre X et Y:

$$\mathcal{T}(X,Y) = \frac{2 \cdot \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \text{signe}[(x_i - x_j)(y_i - y_j)]}{n \cdot (n-1)}$$

Dans le cas d'une série chronologique, la variable Y représente le temps et est donc ordonnée croissante. Le coefficient peut alors être exprimé de la façon suivante :

$$\mathcal{T} = \frac{2 \cdot S}{n \cdot (n-1)}$$

avec

$$S = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} signe(x_{j} - x_{i})$$

Mann (1945) et Kendall (1975) ont montré que même pour une série courte (avec n=10), l'approximation normale pour la distribution de S était excellente. En tenant compte de la présence d'ex-aequo, on a alors :

E(S) = 0 , en l'absence de tendance  
et Var(S) = 
$$\frac{n(n-1)(2n+5) - \sum_{e=1}^{E} t_e(t_e-1)(2t_e+5)}{18}$$

avec : E le nombre de groupes d'ex-aequo avec e = 1 à E $t_e$  la taille du groupe e donné

On peut donc définir une variable normale centrée réduite :

$$z = S / Var(S)^{1/2}$$

Dans le cas d'un test bilatéral, l'hypothèse nulle " $H_0$ :  $\mathcal{T} = 0$  (pas de tendance)", sera acceptée si  $|z| \leq Z_{\alpha/2}$ . Une valeur positive de S indique une tendance croissante et une valeur négative une tendance à la baisse.

Pour prendre en compte la saisonnalité, Hirsch et al. (1982) ont proposé une version modifiée de ce test. Cette forme correspond à une somme pondérée de différents  $\tau$  de Kendall utilisée pour tester l'association en présence d'un "facteur bloquant" qui peut être la saison ou le mois (Korn, 1984 ; Taylor, 1987).

On peut illustrer la procédure sur l'exemple de données recueillies mensuellement. Une valeur du test statistique est calculée pour chaque mois k.

$$S_{k} = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \operatorname{sign} e(x_{jk} - x_{ik})$$
  
avec  $k = 1, 2, ..., 12$   
et  $n = \operatorname{nombre} d'\operatorname{années}.$ 

Pour chaque mois, il est alors possible de tester l'existence d'une tendance significative en utilisant les estimations de l'espérance et de la variance de  $S_k$  telles qu'on les a définies précédemment.

L'existence d'une tendance générale peut être testée en faisant la somme des  $S_k$ .

Soit 
$$S' = \sum_{k=1}^{12} S_k$$

On a :

$$E(S') = \sum_{k=1}^{12} E(S_k) = 0$$
  
Var(S') =  $\sum_{k=1}^{12} Var(S_k) + \sum_{i=1}^{12} \sum_{j=1}^{12} cov(S_i, S_j)$  avec  $i \neq j$ 

Si on suppose les observations indépendantes, alors les  $S_k$ , fonctions de variables indépendantes, sont également indépendantes. On a alors  $cov(S_i, S_i) = 0$  et donc :

$$Var(S') = \sum_{k=1}^{12} Var(S_k)$$

Taylor (1987) propose différents types de pondération. Le premier consiste simplement à donner le même poids à chaque  $S_k$ , ce qui correspond au cas précédent. Un autre système de pondération est préférable lorsque les  $S_k$  ne sont pas calculés sur le même nombre d'observations, ce qui est le cas lorsqu'il y a des données manquantes dans la série :

Soit  $n_k$ , le nombre d'observations réalisés pour le mois k.

La pondération sera égale à  $W_k = n_k \cdot (n_k - 1) / 2$ 

Le test revient alors à comparer la valeur Z à une variable normale centrée réduite, avec :

$$Z = \frac{S'}{\sqrt{\sum_{k=1}^{12} W_k^2 . Var(S_k)}}$$

Hirsh et Slack (1984), en se basant sur les résultats de Dietz et Killeen (1981), ont étendu ce test au cas où les observations ne sont pas indépendantes. En présence de corrélation sérielle, par exemple pour des processus ARMA(1,1), la modification proposée se révèle plus exacte. Néanmoins, cette modification s'avère moins performante sur des petits échantillons et quand les données sont réellement indépendantes. Hirsh et Slack concluent que ce test dérivé de celui de Dietz et Killeen n'est certainement applicable que pour des jeux de données d'au moins 40 ans, pour une fréquence d'observation mensuelle.

#### Homogénéité des tendances

Van Belle et Hugues (1984) proposent de tester l'homogénéité de tendance entre périodes (mois) considérées.

Soit 
$$Z_k = S_k / Var(S_k)^{1/2}$$

Sous l'hypothèse nulle  $H_0: \tau_k=0$ , chaque  $Z_k^2$  suit approximativement une distribution du  $\chi^2$ 

à 1 ddl. On peut donc définir un  $\chi^2_{total} = \sum_{k=1}^{12} Z_k^2$  qui aura approximativement une distribution

du  $\chi^2$  à 12 ddl sous l'hypothèse nulle  $H_0$ :  $\tau_1 = \tau_2 = ... = \tau_{12} = 0$ .

Ce  $\chi^2_{total}$  peut se décomposer de la façon suivante :

 $\chi^2_{total} = \chi^2_{homogénéité} + \chi^2_{tendance}$ 

où 
$$\chi^2_{\text{tendance}} = 12 \cdot \overline{Z}^2$$
 avec  $\overline{Z} = \frac{\sum_{k=1}^{12} Z_k}{12}$ 

 $\chi^2_{homogénéité}$  se déduit par soustraction.  $\chi^2_{homogénéité}$  et  $\chi^2_{tendance}$  ont respectivement des distributions du  $\chi^2$  à 11 et 1 ddl sous les hypothèses  $H_0: \tau_1 = \tau_2 = ... = \tau_{12} = 0$  et  $H_0: \tau = 0$ . L'homogénéité des tendances entre les différents mois peut donc être testée en comparant  $\chi^2_{homogénéité}$  aux tables du  $\chi^2$  à 11 ddl. Si  $\chi^2_{homogénéité}$  est significatif, alors un test sur la tendance générale est inapproprié. Tester indépendamment les tendances de chaque mois reste cependant possible.

#### I.2.1.d. Prévision

Compte tenu des ruptures de tendances, la prévision d'évolution est réalisée à l'aide d'une technique de régression, sur la dernière partie homogène de la série, au sens de la fonction *Cusum*. Dans les premiers traitements que nous avons réalisés (Carpentier *et al*, 1996a), une régression linéaire simple était calculée sur les moyennes mobiles. Constatant que la distribution statistique des valeurs observées s'écarte souvent et quelquefois fortement de la normalité, le choix de la méthode s'est donc ensuite porté sur une régression non paramétrique, ce qui a permis d'améliorer les prévisions en s'affranchissant des hypothèses de distribution et en limitant l'influence des valeurs extrêmes. La méthode choisie permet également de prendre en compte les cycles saisonniers qui caractérisent l'évolution de certains paramètres.

La méthode utilisée est fondée sur l'algorithme de calcul du coefficient de corrélation  $\tau$  de Kendall. Décrite initialement par Sen (1968), elle a été adaptée par Hirsch *et al* (1982, 1991) afin d'éliminer l'effet de la saisonnalité.

Pour une série régularisée, on remplacera la notation  $X_j$  par la notation  $X_{a,m}$ , a étant l'indice de l'année et m l'indice du mois d'observation. Pour des données mensuelles, on aura donc m = 1, 2, ..., 12 et  $1 \le a \le n_m, n_m$  étant le nombre d'observations réalisées pour le mois m dans l'ensemble de la série. Il n'est donc pas nécessaire d'avoir une valeur observée pour chaque couple (a,m) et les valeurs  $n_m$  ne sont pas forcément égales pour tous les mois.

On calcule pour tous les couples  $(X_{a,m}, X_{a',m})$  tels que m = 1, 2, ..., 12 et  $1 \le a \le a' \le n_m$ , la pente :

$$P_{a,a',m} = \frac{\left(X_{a',m} - X_{a,m}\right)}{(a'-a)}$$

L'estimateur de la pente b de la tendance est alors la médiane des pentes  $P_{a,a',m}$ individuelles. L'ordonnée à l'origine de la droite de régression est obtenue en prenant la médiane des valeurs  $(X_{a,m} - \frac{b}{12} \cdot T_{a,m})$ ,  $T_{a,m}$  étant l'abscisse temporelle de la valeur  $X_{a,m}$ . Hirsch *et al* (1991) montrent que cet estimateur de pente est quasiment aussi performant que la régression classique quand les données sont normales, mais qu'il est meilleur dès que l'écart à la normalité devient important (fig. 1.5).



Figure 1.5 : Efficacité relative des estimateurs de pente. L'efficacité est exprimée par le rapport entre les erreurs types (RMSE : Root Mean Square Error) de l'estimation par les régressions non paramétrique et classique sur 1000 simulations par Monte-Carlo, en fonction du pourcentage de mélange entre deux populations différentes (écart à la normalité) et pour deux longueurs de séries différentes (N=6 et N=36). D'après Hirsch *et al.*, 1991.

#### I.2.2 Description et typologie des formes des séries temporelles.

Dès les premières études réalisées sur les séries chronologiques de qualité des eaux à l'échelle du bassin Artois-Picardie, il nous est apparu que beaucoup de séries présentaient des formes d'évolution similaires, c'est-à-dire les mêmes tendances sur les même périodes et des changements simultanés de ces tendances. Cependant si les profils de ces évolutions temporelles étaient souvent similaires, les niveaux atteints par les paramètres mesurés n'étaient pas du tout du même ordre de grandeur d'une variable à l'autre, ou d'une station à l'autre. Notre but a donc été de réaliser une typologie de ces formes qui permette de mettre en évidence une cohérence géographique des ces types de profils : cohérence qui peut s'exprimer par rapport à un bassin ou encore par rapport à un type de situation, un type

"d'environnement" du cours d'eau (occupation des sols, situation à dominante urbaine ou agricole, présence d'activité industrielle, etc...).

Il s'agissait donc de proposer une méthode simple dont le but est de mettre en évidence les stations ou les paramètres qui ont les mêmes formes d'évolution chronologique, abstraction faite des niveaux quantitatifs atteints. On considère pour cela que deux séries chronologiques se ressemblent si elles présentent de façon simultanée les mêmes phases d'augmentation et/ou de diminution ainsi que les mêmes points de changement de tendance. Il ne s'agit pas de faire une différence sur le niveau atteint par les valeurs ou sur les intensités des évolutions, mais uniquement sur le sens (hausse ou baisse) de ces évolutions.

#### I.2.2.a. Extraction de la tendance à long terme – Lissage des séries.

Dans une premier temps, afin de s'affranchir des fluctuations saisonnières et des variations à court terme, la tendance générale des séries est extraite par lissage. Deux types de méthodes ont plus particulièrement été retenues :.

- le lissage par régression locale de type LOWESS ou LOESS ;

- le lissage par la méthode Kernel.

#### • Lissage par Régression Locale (techniques LOWESS et LOESS)

La régression locale est une méthode de lissage ancienne, développée à l'origine pour l'analyse des données de mortalité et le lissage de séries chronologiques. Elle se fonde sur l'ajustement de polynômes locaux à des séries univariables et régulières. On citera les méthodes de moyennes mobiles pondérées, devenues très populaires, comme celles de Spencer (1904) et Henderson (1916). Ce principe avait d'ailleurs été appliqué plus tôt au problème des séries irrégulières (De Forest, 1873).

Plus récemment, ce type d'approche a connu un renouveau, notamment avec les travaux de Cleveland (1979), qui introduit la méthode LOWESS (Robust Locally Weighted Regression). L'introduction d'algorithmes de calcul plus performants pour l'ajustement local, a encore permis d'améliorer la méthode, rebaptisée alors LOESS, et de l'étendre au cas de prédicteurs multiples (Cleveland et Devlin, 1988 ; Cleveland et Grosse, 1991 ; Cleveland *et al*, 1991).

Cette méthode exploratoire a été développée initialement pour mettre en évidence les relations non linéaires dans des diagrammes de dispersion. Le principe de cette méthode la rend applicable à des données irrégulièrement espacées, indépendante d'une quelconque hypothèse de distribution et robuste vis-à-vis des valeurs aberrantes. Elle peut également s'appliquer à l'extraction de tendance dans les séries chronologiques (Bodo, 1989). Il faut cependant remarquer que si à l'origine cette méthode a été définie par Cleveland pour identifier une relation de causalité entre deux variables, une telle causalité n'existe pas dans les séries temporelles : le temps n'est pas la cause de variation d'un paramètre de qualité de *l'eau*.

#### L'algorithme de la méthode LOWESS

Pour *n* couples  $(x_i, y_i)$  d'observations de deux variables avec i = 1, ..., n on réalise une estimation en  $x_i$  de la façon suivante :

1) on définit le voisinage de  $x_i$  en fonction d'un nombre fixé k de voisins les plus proches.

2) on affecte un poids à chacune des observations du voisinage selon une fonction de pondération tri-cube :

$$W_{j}(X_{i}) = W\left(\frac{|X_{i} - X_{j}|}{|X_{i} - X_{k}|}\right)$$

avec  $W(u) = (1-u^3)^3$  pour  $0 \le u < 1$ 

et j = 1,...,k

où les  $x_j$  sont les observations du voisinage ordonnées telles que  $x_k$  est l'observation la plus éloignée de  $x_i$ .

Pour tout point en dehors du voisinage, on a W(u) = 0.

3) on calcule à partir du voisinage une estimation  $\hat{\mathcal{Y}}_i$  par les moindres carrés pondérés.

4) après avoir calculé une estimation en tout point  $x_i$ , on définit des poids de "robustesse" selon une fonction bi-carrée :

$$\mathbf{r}_i = \mathbf{B}(\mathbf{e}_i / 6 \cdot \mathbf{MAD})$$

avec  $B(u) = (1-u^2)^2$  pour  $0 \le u < 1$  B(u) = 0 sinon  $e_i = |y_i - \hat{y}_i|$ et MAD = médiane ( $e_i$ )

5) on recalcule les estimations  $\hat{y}_i$  par les moindres carrés pondérés en ajustant par les poids de robustesse.

6) Les étapes 4 et 5 sont réitérées plusieurs fois pour minimiser l'influence de valeurs aberrantes.

Cette méthode est applicable à des données irrégulièrement espacées. Elle est indépendante d'une quelconque hypothèse de distribution des valeurs et robuste vis à vis des valeurs extrêmes ou aberrantes. Elle peut donc s'appliquer directement aux séries brutes (non régularisées). Elle est capable de "modéliser" des relations complexes, non linéaires et non monotones, notamment pour estimer les relations entre concentrations et débits. Elle permet même de rendre compte des ruptures et changements brutaux.

Le paramètre k, à savoir le nombre de voisins utilisés pour l'estimation en un point donné, est le paramètre le plus important. Il définit le degré de lissage de la méthode : plus k est grand, plus l'effet de lissage est fort. Concrètement, on utilise plutôt un paramètre f, défini comme la proportion de points que représente k par rapport au nombre total de points observés : f = k / n.

Pour les séries chronologiques, le choix de k (ou de f), est important car il permet de choisir si on veut plutôt mettre en évidence des tendances à long terme ou si, au contraire, on veut se focaliser sur les variations locales (à court terme).

Le lissage se fait donc en définissant un nombre de points voisins intervenant dans le calcul d'une estimation et non pas en définissant une largeur de fenêtre. La distribution des ces voisins n'est pas nécessairement symétrique autour du point d'estimation et la largeur de fenêtre considérée n'est donc pas constante.

25

#### • Lissage par la méthode Kernel

Le lissage par Kernel (Härdle, 1990) s'apparente à un lissage par moyenne mobile. Il s'agit en effet de faire glisser une fenêtre sur l'ensemble de la série. Ce type de lissage est de la forme suivante :

$$\hat{y}_{i} = \sum_{j=1}^{n} y_{j} \cdot K\left(\frac{x_{i} - x_{j}}{b}\right) / \sum_{j=1}^{n} K\left(\frac{x_{i} - x_{j}}{b}\right)$$

où b est la largeur de bande et K() la fonction Kernel de pondération, telles que le poids des valeurs situées en dehors d'une fenêtre de largeur b soit nul.

Si on fait exception des fonctions uniforme ou périodique (sinusoïdale), l'ensemble des fonctions de pondération proposées conduisent à des résultats à peu près similaires. Le paramètre critique est principalement la largeur de bande qui détermine le degré de lissage réalisé.

Pour les applications réalisées sur les séries du RNB, il a été choisi d'utiliser un lissage par Kernel avec une largeur de bande de 36 mois et une fonction de pondération triangulaire (décroissance linéaire des poids avec la distance). Sur une série régulière, cette méthode est équivalente à une moyenne mobile pondérée. Par rapport à une moyenne mobile classique, une particularité de cette méthode est de ne pas tronquer les extrémités de la série. Néanmoins, les estimations réalisées aux extrémités peuvent être plus ou moins affectées par des effets de bord et ne sont pas de même qualité que les estimations centrales.

#### I.2.2.b. Calcul d'une distance entre séries temporelles

On calcule d'abord à partir des séries lissées, les séries des différences premières  $\Delta X_i$ suivant la procédure suivante :

Soit une série d'observations  $X_t$  à des temps t = 1, ..., n.

On a  $\Delta X_t = X_t - X_{t-1}$  pour t = 2, ..., n.

On ne s'intéresse ici qu'au signe de ces différences  $\Delta X_t$ : positif si la série montre une augmentation du paramètre et négatif s'il y a diminution. Les séries des valeurs lissées (tendance générale de la série) sont donc transformées en séries de signes correspondants aux sens de variation au long de ces séries. La figure suivante (fig. 1.6) donne un exemple d'une telle transformation d'une série lissée en série de signes.



Figure 1.6 : transformation d'une série lissée (à gauche) en série de signes (à droite).

Pour deux séries X et Y et pour chaque temps t de 2 à n, on distingue les cas suivants en fonction des signes de  $\Delta X_t$  et  $\Delta Y_t$  (augmentation ou diminution) :

	$\Delta Xt < 0$	$\Delta Xt > 0$
$\Delta Yt < 0$	а	b
$\Delta Yt > 0$	с	d

Les valeurs a et d correspondent respectivement au nombre de fois où les deux séries diminuent ou augmentent conjointement. Les valeurs b et c correspondent aux non concordances d'évolution des deux séries. La distance entre deux séries X et Y est alors obtenue en appliquant le coefficient de concordance simple (Sokal et Michener, 1958) :

$$d(X,Y) = (b+c) / (a+b+c+d) = (b+c) / n-1$$

Cette distance d(X,Y) est donc un indice de co-évolution entre deux séries : une valeur proche de 0 indiquera que les deux séries évoluent simultanément (elles baissent ou augmentent en même temps), alors qu'une valeur proche de 1 indiquera au contraire que les tendances des deux séries s'opposent (donc fortes valeurs de *b* et *c* par rapport à *a* et *d*).

#### I.2.2.c. Lien entre l'"indice de co-évolution" et le Tau de Kendall.

Pour le calcul du coefficient de corrélation non-paramétrique de Kendall, on s'intéresse aux signes des produits des différences entre toutes les observations de deux séries, dont on fait la moyenne :

$$\mathcal{T}(x, y) = \frac{2 \cdot \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} signe \left[ (x_i - x_j)(y_i - y_j) \right]}{n \cdot (n-1)}$$

La distance entre profils décrite précédemment correspond à un  $\tau$  de Kendall si on considère non pas tous les couples d'observations possibles, mais uniquement les couples d'observations successives.

$$\mathcal{T}_{1}(x,y) = \frac{\sum_{i=1}^{n-1} signe\left[(x_{i} - x_{i+1})(y_{i} - y_{i+1})\right]}{n-1} = \frac{\sum_{i=2}^{n} signe\left[(\Delta x_{i})(\Delta y_{i})\right]}{n-1}$$

On peut passer de la notion de corrélation, qui s'exprime entre -1 et 1, à celle de distance s'exprimant entre 0 et 1 par la relation suivante :

d (x,y) = (1 - 
$$\tau_{1(x,y)}$$
) / 2

#### I.2.2.d. Analyse des tableaux de distances par MDS et classification

La méthode décrite ci-dessus, appliquée à tous les couples de séries, aboutit à une matrice de distances/dissimilarités à partir de laquelle il est alors possible de réaliser une typologie des séries qui sont comparées.

Dans ce but, il peut être fait appel :

- soit à une technique de représentation en espace réduit (souvent deux dimensions) de sorte que la configuration dans le plan représente le mieux la matrice de distances obtenue. Il s'agit de Multi-Dimensional Scaling ou Analyse en COordonnées Principales (Torgerson ,1958 ; Kruskal, 1964 ; Gower, 1966 ; Escoufier, 1975);

- soit une technique de Classification Hiérarchique (e.g. Kaufman et Rousseuw, 1990).

Une illustration de l'application ces outils sera donné au § I.3.3.

#### I.2.3 Relation concentration/débit

Il est bien connu que dans de nombreux cas les concentrations des éléments dissous et particulaires mesurés dans les eaux courantes sont avant tout fonction du débit des cours d'eau. La forme de cette relation concentration/débit reflète l'origine et le comportement de ces éléments (Meybeck, 1985; Pourriot et Meybeck, 1995).



Figure 1.7 : Types de relations concentration-débit. (1) dilution exponentielle (apports ponctuels). (2) et (3) lessivages des sols. (4) et (5) MES. (6) apports atmosphériques. (7) cycle de dilution. D'après Pourriot et Meybeck (1995).

Quand le débit augmente, différents cas de figure peuvent se présenter :

- on peut observer une baisse (en général une décroissance exponentielle) de la concentration (fig. 1.7- type 1). Il s'agit donc d'un effet de dilution. On rencontre cette relation pour la plupart des ions mais elle s'observe également à l'aval de rejets ponctuels (naturels ou anthropiques). La relation entre la concentration (C) et le débit (Q) est de la forme :

 $C = a Q^b$  avec -1 < b < 0;

la concentration augmente avec le débit, ce qui caractérise surtout le lessivage des nutriments dans les sols agricoles ou un transport en suspension de l'élément considéré (fig. 1.7 - type 2, 3 et 4); - une relative constance des concentrations est plutôt typique des éléments apportés par les précipitations et des systèmes karstiques où les eaux souterraines dominent, quels que soient les débits (fig. 1.7 - type 6);

Ces trois grands types de relations sont rencontrés sur le bassin Artois-Picardie. Notons cependant qu'il existe d'autres types de relations concentration-débit correspondant, soit à des phénomènes plus complexes, soit à la conjonction de deux ou plusieurs des phénomènes évoqués ci-dessus.

Si une forte corrélation existe entre un paramètre et le débit, et si l'évolution des concentrations pour ce paramètre présente une tendance significative, il est possible que cette tendance perçue soit le résultat d'une histoire particulière des débits, liée à des variations climatiques, et non pas la conséquence d'un changement dans les processus d'entrée et de transport de l'élément considéré.

On devra donc regarder si :

i) la tendance observée sur les concentrations n'est qu'un artefact du à une évolution du débit, ou si

ii) cette tendance indique réellement un changement intervenu dans le bassin, tel que la relation concentration/débit a changé dans le temps.

A l'opposé, on devra également vérifier, dans le cas où aucune évolution n'est détectée, que ce soit bien parce qu'aucun changement dans la relation n'est intervenu et non pas parce que l'évolution des débits a masqué l'effet d'un tel changement.

Il apparaît donc important de pouvoir vérifier l'existence de tendances et de quantifier ces tendances en corrigeant les observations de concentration des variations liées au débit. Une conséquence de l'élimination de l'effet du débit est également d'éliminer une part importante de la variabilité dans les données. L'efficacité des méthodes statistiques employées s'en trouve donc accrue.

Une première solution consiste à utiliser le débit (ou une transformation adéquate du débit) comme covariable dans un modèle de régression linéaire multiple (Alley, 1988 ; Smith et Rose, 1991). On peut reprocher à ce type d'approche qu'elle n'est pas satisfaisante pour un certain nombre de problèmes déjà évoqués (distribution des valeurs, données manquantes,
seuils de détection ...) et il est préférable de trouver une estimation plus robuste de la relation concentration/débit.

La deuxième solution, consistant en une approche non-paramétrique, a donc été adoptée (Hirsch *et al.*, 1982, 1991). Elle se décompose en différentes étapes :

1) Trouver un modèle pour la relation concentration/débit, C = f(Q):

- soit par l'intermédiaire de techniques de régression classiques, en ajustant un modèle linéaire ou polynomial.

- soit, pour une approche totalement non-paramétrique et pour modéliser des relations complexes, par l'usage d'une méthode de régression locale de type LOWESS (Cleveland, 1979).

2) Si une relation est bien mise en évidence, calculer une série corrigée des variations du débit en prenant les résidus par rapport aux estimations données par le modèle.

 $e_t = x_t - \hat{x}_t$ 

3) Appliquer les tests de tendances et estimateurs de pente à cette nouvelle série.

On s'aperçoit que cette procédure a pour effet de réduire les effets saisonniers ainsi que l'asymétrie des distributions.

#### **I.3 - Exemples d'applications**

#### I.3.1 Evolution des concentrations

La méthodologie décrite dans le chapitre précédent a été appliquée aux séries chronologiques du Réseau National de Bassin en Artois-Picardie. La réalisation des prévisions d'évolution des concentrations de divers éléments a été faite. Les prévisions portent sur le court ou moyen terme : prévisions pour l'an 2000 à partir des séries disponibles allant jusque 1995, au moment de notre étude. Le choix de l'an 2000 ne résulte pas uniquement de la valeur symbolique de cette date, mais du caractère raisonnable d'une prévision à 5 ans ; une prévision à plus long terme aurait d'ailleurs été impossible dans la majeure partie des cas, puisqu'elle se serait appuyée sur une trop faible série de données ayant une tendance monotone. De plus, ces prévisions avaient pour but d'aider le gestionnaire (en l'occurrence l'Agence de l'Eau) à faire des choix quant aux lieux à privilégier dans la politique d'amélioration de la qualité de l'eau.

Les prévisions ont été réalisées pour tous les paramètres et sur toutes les stations du RNB du bassin, à l'exception d'une seule : la station 92500, la Liane à Boulogne sur Mer, qui n'a commencé à être observée qu'en 1986 et sur laquelle la série a été jugée trop courte pour faire l'objet de prévisions.

Nous ne présenterons ici qu'une synthèse des principaux résultats à l'échelle du bassin Artois-Picardie. Le détail par station et par variable figure dans les rapports réalisés pour l'Agence de l'Eau (Leprêtre et Carpentier, 1997a).

Le tableau ci-dessous (tableau 1.2) indique de manière résumée les changements de classe de qualité prévus entre les moyennes calculées pour la période 1990-1994 et les moyennes données par les prévisions pour l'an 2000. La répartition dans les classes de qualité est faite en tenant compte des valeurs moyennes et non de la valeur atteinte 90% du temps comme dans le système mis en place dans les Agences de l'Eau. En effet, si cette dernière valeur peut être calculée sur les séries observées, il n'était pas possible de l'obtenir pour les prévisions qui se fondent sur l'évolution d'une moyenne annuelle. Les classements de qualité présentés ici pour les années 1990-94 sont donc légèrement plus "optimistes" que ceux réalisés par l'Agence de l'Eau.

On constate que les prévisions annoncent globalement une nette amélioration de la qualité des eaux pour tous les paramètres considérés. Un nombre important de gains de classe de qualité est prévu sur la DBO5, la DCO, le NH4, le NTK, et surtout sur les phosphates pour lesquels ce sont plus de la moitié des stations qui devraient gagner une classe de qualité (au sens des grilles multi-usages). L'amélioration est moins marquée pour les nitrites et l'oxygène. Les exceptions concernent les matières en suspension qui ne semblent pas évoluer et les nitrates pour lesquels il est prévu 11 pertes de classe de qualité.

	NH4	NO2	NO3_	PO4	NTK
Gain de classe	27	18	0	36	27
Pas de changement	39	47	55	30	39
Perte de classe	0	1	11	0	0.
	DBO5	DCO	02	%Sat.	MES
Gain de classe	28	31	9	20	0
Pas de changement	37	34	56	45	64
Perte de classe	0	0	0	0	1

Tableau 1.2: Prévisions de changements de classe entre 1990-94 et 2000.

On peut détailler l'évolution depuis 1980 du classement des stations du RNB par période de 5 ans, ainsi que les prévisions, pour les concentrations de quelques paramètres (fig. 1.8). On constate que les nitrates, après une nette amélioration au cours des années 80 qui s'est poursuivie au début des années 90, connaissent sur les dernières années du suivi une hausse marquée, qui se ressent très fortement dans les prévisions. En revanche, l'évolution des phosphates, qui était plutôt légèrement en perte de qualité depuis 1980, subit un changement brutal d'évolution des concentrations à partir de 1991-92, ce qui aboutit à des prévisions de forte amélioration. L'attention particulière portée précédemment par l'Agence de l'Eau sur la pollution organique et les mesures prises pour l'enrayer se sont traduits par une nette amélioration de la situation. De même, la DBO5 montre une nette amélioration depuis la deuxième moitié des années 80. Cette amélioration concerne aussi l'azote kjeldahl (NTK) et NH4, amélioration qui s'est accentuée dans les années 90, et qui se traduit au niveau des prévisions par un grand nombre de gains de classes de qualité. L'oxygénation des cours d'eau, représentée ici par la concentration en  $O_2$  en pourcentage de saturation, présente une amélioration qui avait démarrée dès le début des années 80. Là encore les prévisions suggèrent que cette amélioration devrait continuer.



Figure 1.8 : évolution de la répartition en classes de qualité des stations du Réseau National de Bassin pour quelques paramètres : nitrates (NO3), phosphates (PO4), azote Kjeldahl (NTK), ammonium (NH4), Demande biologique en O2 (DBO5) et oxygène en pourcentage de saturation (%Sat.).

Si l'on s'intéresse à la localisation géographique de ces principaux changements, on constate qu'ils sont essentiellement situés dans les zones fortement urbanisées, généralement doublées d'une forte industrialisation (cartes 1 à 8).

Les pertes de classes de qualité pour les concentrations en nitrates sont situées sur l'Escaut en amont de Valenciennes, la Scarpe à Raches, le canal de la Deule à sa sortie du bassin minier et avant Lille, sur la Clarence (également en sortie du bassin minier), sur l'Aa en aval de Saint-Omer, ainsi qu'en un point situé sur l'Ancre entre Albert et Amiens (cartes 1 et 2).

Les gains de classes de qualité pour les phosphates (cartes 3 et 4), pour les autres formes de l'azote (cartes 5 et 6) ou pour les paramètres relatifs à l'oxygène (cartes 7 et 8), se trouvent dans les mêmes secteurs géographiques : le bassin de l'Escaut, la Scarpe, la Deule et l'Aa, auxquels viennent s'ajouter la Lys, la Sambre et quelques stations situées sur la Somme (notamment autour de Saint-Quentin et Ham). Les zones portuaires de Calais et Dunkerque montrent aussi une amélioration pour les paramètres relatifs à la pollution organique.

Les cartes 9 à 12 montrent les stations dans lesquelles un changement important dans les concentrations est prévu. En effet, certaines stations connaissent une forte évolution sans pour autant qu'il y ait un changement dans leur classement. C'est le cas des stations proches et en sortie de l'agglomération Lilloise pour les phosphates (carte 9). Les concentrations ont fortement baissé, mais les niveaux atteints restent cependant toujours élevés, ce qui les maintient pour la plupart en classe 4. La même constatation peut être faite pour les nitrates dans les mêmes secteurs géographiques : on constate une augmentation notable des concentrations sans qu'il n'y ait encore perte de qualité (carte 10).

La baisse des teneurs en nitrates s'accompagne le plus souvent d'une baisse des autres formes de l'azote. Il en résulte une baisse de l'azote total surtout sensible, là encore, dans les zones urbaines et industrielles (bassin de l'Escaut, la Scarpe, la Deule, la Lys et l'Aa) (carte 11). On observe aussi sur ces zones une nette amélioration de l'oxygénation de l'eau qui se traduit par une augmentation des concentrations en oxygène exprimées en pourcentage de saturation (carte 12).

#### Nitrates 90-94



Carte 1 : classes de qualité des concentration en nitrates pour la période 90-94 sur les stations du RNB. En vert les stations de qualité 1, en jaune les stations de qualité 2.

Nitrates prévisions



Carte 2 : classes de qualité des prévisions 2000 des concentration en nitrates sur les stations du RNB. En vert les stations de qualité 1, en jaune les stations de qualité 2.

#### Phosphates 90-94



Carte 3 : classes de qualité des concentration en phosphates pour la période 90-94 sur les stations du RNB.

Phosphates prévisions



Carte 4 : classes de qualité des prévisions 2000 des concentration en phosphates sur les stations du RNB.

#### NTK 90-94



Carte 5 : classes de qualité des concentration en azote Kjeldahl pour la période 90-94 sur les stations du RNB.

#### NTK prévisions



Carte 6 : classes de qualité des prévisions 2000 des concentration en azote Kjeldahl sur les stations du RNB.

#### O2 % saturation - 90-94



Carte 7 : classes de qualité des concentration en oxygène pour la période 90-94 sur les stations du RNB.

O2 % saturation - prévisions



Carte 8 : classes de qualité des prévisions 2000 des concentration en oxygène sur les stations du RNB.

#### Phosphates - Changements prévus



Carte 9 : changements prévus des concentration en phosphates sur les stations du RNB. En vert, les baisses > 0.2 mg/l - En jaune, les changements < 0.2 mg/l.

Nitrates - Changements prévus



Carte 10 : changements prévus des concentration en nitrates sur les stations du RNB. En vert, les baisses > 5 mg/l – En jaune, les changements < 5 mg/l – En rouge, les hausses > 5 mg/l.





Carte 11 : changements prévus des concentration en azote total sur les stations du RNB. En vert, les baisses > 1 mg/l – En jaune, les changements < 1 mg/l – En rouge, les hausses > 1 mg/l.

O2 % saturation - Changements prévus



Carte 12 : changements prévus des concentration en oxygène (% de saturation) sur les stations du RNB. En vert, les hausses > 10% – En jaune, les changements < 10% – En rouge, les baisses > 10%.

#### I.3.2 Flux vers le milieu marin

Une estimation des flux d'Azote et de Phosphore exportés vers le milieu côtier (Manche et Mer du Nord) et des flux transfrontaliers a été réalisée (communication au colloque *Interfaces*, Nantes, Carpentier *et al.*, 1996). Les flux moyens en t/an pour les périodes 1985-89 et 1990-94 sont donnés dans le tableau ci-après (tableau 1.3 et fig. 1.9). Ces estimations ont été réalisées soit à partir des valeurs de concentration et des débits mensuels quand les données étaient disponibles (littoral de la Manche), ou à partir d'estimations de débits moyens annuels (surtout pour les cours d'eau transfrontaliers).

On constate, dans les différents secteurs géographiques étudiés, une baisse sensible des flux entre les deux périodes considérées. Pour les cours d'eau débouchant dans la Manche (façade ouest du littoral), les apports sont dominés par ceux de la Somme (presque la moitié des apports), et secondairement par ceux de la Canche et de l'Authie. Les flux d'azote et de phosphore apportés par ces trois cours d'eau sont en baisse importante au cours des dernières années. On constate également une forte baisse sur la Slack. Les exportations d'azote et de phosphore par la Liane et d'azote par le Wimereux sont au contraire en augmentation. Il faut néanmoins remarquer que la baisse globale des flux vers la Manche s'accompagne d'une baisse générale des débits entre les deux périodes, et que ces baisses sont d'une ampleur comparable (à peu près 17%).

L'évolution des flux dans les cours d'eau transfrontaliers (à l'est) et dans ceux donnant directement sur la Mer du Nord est à peu près similaire. On constate partout une baisse assez importante des flux d'azote et de phosphore. En revanche, les débits n'ont par contre pratiquement pas changé. Or ce sont ces cours d'eau transfrontaliers qui réalisent les exportations les plus importantes. La Lys, à elle seule, exporte deux fois plus d'azote que la Somme et trois fois plus de phosphore (soit presque autant d'azote et plus de phosphore que tous les cours d'eau de la façade Manche), alors que son débit est entre une fois et demi et deux fois moins important. A l'extrême, le canal de l'Espierre, avec un débit moyen de 1,5 m3/s, soit environ vingt fois moins que la Somme, exporte autant de phosphore.

Les prévisions à la baisse des concentrations, quasi générales pour les phosphates et concernant de nombreux secteurs pour l'azote, vont de paire avec les tendances observées sur les flux, et devraient conduire à une amélioration très significative de la qualité des cours d'eau, en particulier pour les cours d'eau transfrontaliers pour qui la diminution est la plus marquée.



Figure 1.9 : flux exportés d'azote et de phosphore à partir du bassin Artois-Picardie pour la période 1990-94.

			1985-89		1990-94		
Station	N° RNB	N (t/an)	P (t/an)	Débits (m3/s)	N (t/an)	P (t/an)	Débits (m3/s)
La Somme à Abbeville	130000	6289	421	39.8	4541	301	30.3
L'Authie à Quend	100900	1987	105	12.5	1680	85	11
La Canche à Beutin	095000	2919	179	18	2573	145	15.5
La Liane à Boulogne sur Mer	092500	869	113	4.6	1223	162	4.9
Le Wimereux à Wimille	091000	156	15	1	172	13	1.1
La Siack à Ambleteuse	090000	429	34	2.3	350	20	2
Total exporté vers La Manche		12649	866	78.2	10539	727	64.8
Canal de Marck à Calais	114600						
Canal de Calais à Coulogne	114000	1850	163	3.7	1314	120	3.7
Canal des Pierrettes à Calais	114300				<u> </u>	<u> </u>	
L'Aa à St-Folquin	104000	2985	208	10.1	2203	179	10.6
Exutoire des wateringues à Dunkerque	111500	1237	259	5	1219	195	5
Total exporté vers la Mer du Nord		6072	631	18.8	4735	494	19.3
La Lys à Werwicq	059000	11398	1347	20	9538	1099	20.1
Canal de l'Espierre à Wattrelos	051000	2123	342	1.5	1659	302	1.5
L'Escaut à Mortagne du Nord	019000	8995	497	16	7171	450	16.5
La Sambre à Jeumont	004000	963	156	6	927	146	6
Total exporté par la Belgique		23478	2342	43.5	19296	1996	44.1

Tableau 1.3 : flux exportés vers la Manche et la Mer du Nord.

#### I.3.3 Analyse de co-évolution des profils temporels

L'objectif de la méthode proposée (cf. § I.2.2) est de réaliser une typologie des stations ou des paramètres à l'aide d'un indice quantifiant le degré de similitude des évolutions chronologiques des séries (indice de co-évolution). Comme il s'agit surtout de mettre en évidence les types d'évolution sur le long terme, un lissage préalable est effectué sur les séries afin de s'affranchir des fluctuations à haute fréquence (variations saisonnières et à court terme). On aboutit à des profils de baisse/augmentation qui vont être comparés entre eux, soit par ordination (MDS), soit par classification hiérarchique.

La méthode peut servir à comparer :

- soit l'évolution de différentes stations pour un paramètre donné. On obtient ainsi une typologie des stations, regroupant celles qui présentent une "histoire" similaire (Exemple 1).

- soit l'évolution de différents paramètres sur une station afin d'identifier ceux qui évoluent conjointement (Exemple 2).

#### I.3.3.1 Exemple 1 : co-évolution des séries phosphates de 64 stations du RNB.

Cet exemple concerne les séries phosphates couvrant la période 1980-95. Les stations 1148, 6000 et 92500 sont exclues de l'analyse, car trop fragmentaires sur cette période.

La projection en MDS (fig. 1.10) oppose sur le premier axe les profils montrant une diminution des phosphates sur la fin de la période aux profils plus stationnaires. Le deuxième axe sépare les profils montrant une augmentation de ceux montrant une diminution en début de période.

L'analyse conjointe de la projection et de la classification (fig. 1.11) permet d'identifier un groupe important constitué des stations de la Deûle, l'Escaut, la Lys canalisée, la Sambre, la Ternoise à Saint-Pol et de quelques stations situées sur la Somme. Ces stations présentent toutes une tendance à la hausse pendant les années 80 (un grand nombre de ces stations présentent aussi une baisse pendant l'année 1987) et à la baisse à partir du début des années 90. Deux autres groupes sont mis en évidence:

- un groupe caractérisé par une baisse dans les années 90, mais non consécutive comme précédemment à une hausse continue dans les années 80. On y trouve les stations de l'Aa, de la Lys rivière, de la Scarpe et des autres stations de la Somme.

- un groupe de profils presque plats tels ceux de la Canche, l'Authie, la Liane, le Wimereux et la Slack, à savoir un ensemble de petits fleuves orientés vers le littoral ouest.

Les stations des zones portuaires (Calais et Dunkerque) montrent une forte hétérogénéité.



Figure 1.10 : projection des séries phosphates pour 64 stations du RNB et profils caractéristiques.

La typologie utilisée ici permet donc d'identifier un groupe principal de stations qui présentent des profils similaires d'évolution des concentrations en phosphates. Le point commun entre ces stations est leur situation dans les zones urbaines et/ou dans les zones industrialisées de l'agglomération lilloise, du bassin minier et de la Somme.



Figure 1.11 : classification hiérarchique des séries phosphates pour 64 stations du RNB.

## I.3.3.2 Exemple 2 : co-évolution des différents paramètres dans une station (la Somme à Abbeville)

La méthode permet aussi la comparaison des évolutions de différents paramètres au sein d'une station. L'illustration porte sur la station 130000 (la Somme à Abbeville). En dehors du lien évident entre les phosphates et le phosphore total et entre l'oxygène et le pourcentage de saturation, la projection en MDS (fig. 1.12) et la classification (fig. 1.13) à partir de la matrice de distances (tableau 1.4) confirment le lien fort existant entre l'évolution

des nitrates, celle des matières en suspension et celle des débits. Les nitrates étant ici la forme d'azote prédominante, les concentrations en azote total sont fortement liées aux nitrates. On peut également noter que l'augmentation des concentrations en oxygène dissous semble liée à la baisse de  $NH_4$ ,  $NO_2$  et  $PO_4$  et à la hausse du rapport N/P. La baisse de la DBO5 et de la DCO semble elle être associée à la baisse de  $NH_4$  et NTK.

	NH4	NO2	NO3	PO4	NTK	N total	P total	N/P	DBO5	DCO	O2	% Sat.	MES
NO2	0.387												
NO3	0.672	0.618											
PO4	0.425	0.124	0.624										
NTK	0.215	0.452	0.790	0.457									
N total	0.473	0.677	0.210	0.737	0.581								
P total	0.425	0.124	0.624	0.000	0.457	0.737							
N/P	0.516	0.817	0.306	0.930	0.516	0.194	0.930						
DBO5	0.435	0.543	0.839	0.462	0.274	0.672	0.462	0.608					
DCO	0.489	0.435	0.720	0.452	0.360	0.586	0.452	0.618	0.183				
O2	0.505	0.785	0.468	0.672	0.495	0.473	0.672	0.333	0.457	0.618			
% Sat.	0.559	0.828	0.435	0.704	0.495	0.441	0.704	0.312	0.468	0.629	0.054		
MES	0.785	0.656	0.253	0.694	0.742	0.419	0.694	0.333	0.672	0.565	0.484	0.441	
Débit	0.774	0.677	0.253	0.683	0.720	0.355	0.683	0.344	0.651	0.511	0.538	0.484	0.075

Tableau 1.4 : matrice de distances entre les séries des différents paramètres pour la station 130000.



Figure 1.12 : projection des séries d'évolution des différents paramètres pour la station 130000.



Figure 1.13 : Classification hiérarchique des séries d'évolution des différents paramètres pour la station 130000.

#### I.3.4 Quelques problèmes rencontrés dans l'évaluation des tendances

#### I.3.4.1 Relation concentration/débit

Les concentrations des différents éléments présentent toujours une relation avec les débits, relation qui reflète l'origine et le comportement des ces éléments (Meybeck, 1985). Deux exemples de ces relations sont présentés ici, concernant l'azote total et les phosphates. Les relations concentration/débit sur la Somme à Abbeville (station RNB 130000) sont comparées aux mêmes relations observées sur l'Authie à Quend (station RNB 100900). Pour dégager la tendance principale des ces relations, nous avons utilisé, comme Hirsch & al. (1991), un lissage par la méthode LOWESS (Locally Weighted Regression : Cleveland, 1979).

Pour la Somme, les teneurs azote et celles en phosphates présentent des relations avec le débit qui sont complètement opposées. La concentration en azote tend à augmenter avec le débit, augmentation caractéristique du lessivage des nutriments des sols agricoles, principalement les nitrates qui constituent ici la forme la plus importante d'apport azoté (fig. 1.14). Au contraire, les phosphates présentent une dilution croissante avec le débit (fig. 1.15). Ce type de relation est typique de l'aval de rejets ponctuels relativement constants, naturels ou anthropiques. Ici, il s'agit des rejets de l'agglomération d'Abbeville. La station 130000 est en effet située à la sortie de la ville.

Pour l'Authie, une augmentation des concentrations en azote se produit de la même façon que pour la Somme pour des débits moyens (fig. 1.16). Mais pour des débits plus importants, la concentration stagne et décroît légèrement, type de relation caractéristique des nitrates en région agricole drainée.

Les concentrations en phosphates restent stables quel que soit le débit, ce qui suggère ici une source diffuse (fig 1.17). Les apports en phosphates se limitent essentiellement à un fond géochimique résiduel, sans réels apports anthropiques (urbains ou agricoles).

On voit donc que l'existence de relations entre le débit et les concentrations peut masquer l'évolution réelle d'un paramètre de la qualité de l'eau : une évolution des débits peut induire une évolution des concentrations sans pour autant qu'il y ait un changement des apports à la source. Il paraît donc nécessaire d'examiner systématiquement les séries en termes de flux et de concentration – d'où, d'ailleurs, l'importance de la réalisation des mesures du débit.

Pour les deux exemples évoqués ci-dessus - la Somme à Abbeville de 1980 à 1995, et l'Authie à Quend de 1988 à 1995 – nous avons testé l'existence de tendances pour les différents paramètres concernant les différentes formes de l'azote et du phosphore. Ces tests (cf. § I.2.1.c) ont été réalisés sur les séries régularisées de concentrations et sur les mêmes séries corrigées des variations liées au débit selon la méthode évoquée au chapitre 2.3. Les résultats sont reportés dans les tableaux suivants (tableau 1.5 et 1.6).



Figure 1.14 : relation entre les concentrations en azote et les débits sur la Somme à Abbeville.



Figure 1.15 : relation entre les concentrations en phosphates et les débits sur la Somme à Abbeville.



Figure 1.16 : relation entre les concentrations en azote et les débits sur l'Authie à Quend



Figure 1.17 : relation entre les concentrations en phosphates et les débits sur l'Authie à Quend

On voit que les tests réalisés sur les séries brutes de concentrations pour la Somme indiquent une baisse significative pour le NH4, les nitrates, l'azote Kjeldhal, l'azote total et pour le rapport N/P. Les nitrites et les phosphates ne montrent pas de tendances particulières.

Quand on élimine l'effet des fluctuations du débit, les tendances observées sur les nitrates et le rapport N/P ne sont plus significatives. En revanche, on voit apparaître une baisse significative pour les concentrations en phosphates. Les tendances à la baisse du NH4, de l'azote Kjeldahl et de l'azote total sont confirmées. Les concentrations en nitrites montrent une tendance à la baisse assez marquée bien que non significative au risque de 5%. On peut également noter que le signe du Tau obtenu s'inverse pour les nitrites et le rapport N/P.

	Cor	Concentrations brutes			Ajustées avec débit			
	S	Tau	pr(Tau=0)	S	Tau	pr(Tau=0)		
NH4	-406	-0.297	4.10-8	-401	-0.294	6.10 <sup>-8</sup>		
NO2	36	0.026	0.627	-137	-0.100	0.065		
NO3	-241	-0.177	0.001	-71	-0.052	0.338		
NTK	-408	-0.299	4.10 <sup>-8</sup>	-481	-0.352	8.10-11		
PO4/P	-93	-0.068	0.210	-217	-0.159	0.003		
N	-465	-0.341	4.10 <sup>-10</sup>	-391	-0.286	1.10 <sup>-7</sup>		
N/P	-159	-0.116	0.032	69	0.051	0.352		

Tableau 1.5 : Tests de tendance pour La Somme à Abbeville de 1980 à 1995

Des observations similaires peuvent être faites sur les concentrations des différents paramètres mesurés sur l'Authie. Aucune tendance significative n'est perceptible pour les phosphates et pour NH4, que ce soit sur les concentrations brutes ou sur les concentrations ajustées avec le débit. Une augmentation des concentrations en nitrates est confirmée dans les deux cas. On voit par contre que la tendance à la hausse significative des concentrations en azote total et du rapport N/P observée sur les séries originales ne l'est plus, quand on élimine l'effet du débit. A l'opposé, une augmentation des teneurs en nitrites apparaît. Les concentrations en azote Kjeldahl sont faibles sur cette station, et la série chronologique est fortement affectée par les problèmes de limite de détection analytique. Les résultats pour ce paramètre ne sont donc pas surs et ne sont reportés ici qu'à titre indicatif.

	Concentrations brutes			Ajustées avec débit			
	S	Tau	pr(Tau=0)	S	Tau	pr(Tau=0)	
NH4	-6	-0.020	0.818	-13	-0.043	0.618	
NO2	25	0.083	0.337	61	0.203	0.019	
NO3	152	0.505	5.10 <sup>-9</sup>	93	0.309	0.0004	
NTK	-17	-0.056	n.s.	-123	-0.409	2.10 <sup>6</sup>	
PO4/P	-30	-0.099	0.249	-19	-0.063	0.466	
Ν	117	0.389	7.10 <sup>-6</sup>	25	0.083	0.337	
N/P	59	0.196	0.024	35	0.116	0.179	

Tableau 1.6 : Tests de tendance pour l'Authie à Quend de 1988 à 1995

A l'aide de ces deux exemples, on constate que l'évolution des débits, qui reflète essentiellement l'évolution de la pluviométrie, est un facteur qui peut fortement influencer l'évolution des concentrations des différents paramètres de la qualité de l'eau. Un changement dans le régime des débits peut masquer ou au contraire faire apparaître une évolution des concentrations d'un élément, sans que pour autant il n'y ait eu de modification dans les mécanismes et les sources d'apport de cet élément au niveau du bassin versant. Il est donc important pour la gestion de la qualité de l'eau de pouvoir, à l'aide de méthodes telles que celle présentée ici, de faire la part de l'influence de variations climatiques naturelles de celles des changements réels au niveau des apports.

#### I.3.4.2 Hétérogénéité des tendances saisonnières

Il est possible, sur une série chronologique, que les tendances exprimées par les différentes saisons ne soient pas homogènes. Une tendance peut apparaître à une certaine période de l'année alors qu'elle ne se manifestera pas le reste du temps. Dans ce cas, un test ou une évaluation de la tendance générale de la série est inadéquat et se révélera inefficace.

Les deux exemples suivants illustrent cet aspect. A cet égard, la version saisonnière du test de Mann-Kendall proposée par Hirsch et al. (1982) est utilisée ici, ainsi que le test d'homogénéité des tendances proposé par Van Belle et Hugues (1984) (cf. § 2.1.c).

Le premier exemple (tableau 1.7 et fig. 1.18) porte sur l'évolution du rapport N/P, indicateur d'une eutrophisation potentielle, enregistré à la station 89000 du RNB sur l'Yser de 1980 à 1995. Le test de tendance réalisé sur la série complète ne permet pas de conclure à l'existence d'une quelconque tendance. Cependant, le test d'homogénéité entre les 12 mois de

l'année donne un  $\chi^2$  significatif qui indique des différences entre les tendances qui s'expriment dans les différents mois.

Les valeurs calculées pour le Tau de Kendall au mois par mois, pour la période d'Octobre à Mars, révèlent une tendance à la hausse du rapport N/P qui est significative pour les mois de Décembre et Janvier. A l'inverse, on observe une baisse assez marquée, bien que non significative au risque 5%, de Mai à Août. On voit donc ici une opposition assez nette entre une période hivernale et un période estivale pour lesquelles les tendance d'évolution s'inversent.

Le deuxième exemple (tableau 1.7 et fig. 1.19) porte sur l'évolution des nitrites à la station 32000 du RNB, l'Hogneau à Thivencelles (en amont de sa confluence avec le canal de Mons). Une augmentation très significative est détectée par le test général, mais là aussi le chi2 rejette l'hypothèse d'homogénéité des tendances. Les tests réalisés mois par mois montrent une augmentation significative des nitrites pour une période de Mars à Juillet, alors qu'aucune évolution particulière n'est détectée pour le reste de l'année., On voit donc, là encore, une opposition entre les périodes "automne-hiver" et "printemps-été". La période hivernale, pendant laquelle les concentrations sont faibles, montre une très grande stabilité, alors que la période estivale, caractérisée par une forte variabilité, présente une tendance à la hausse.

	N/P – I'Yser -RNB 89000					
	S	Tau		tau=0)		
Test général	78	0.060		0.273		
	Chi	Chi-2 p				
Test d'homogénéité	24.93 0.009					
	S(j)	Tau(j)	p(tau=0	)) n		
Janvier	45	0.429	0.026	15		
Février	27	0.257	0.181	15		
Mars	19	0.181	0.347	15		
Avril	-9	-0.086	0.656	15		
Mai	-25	-0.238	0.216	15		
Juin	-34	-0.283	0.126	16		
Juillet	-40	-0.333	0.071	16		
Août	-15	-0.143	0.458	15		
Septembre	-1	-0.009	0.960	15		
Octobre	25	0.238	0.216	15		
Novembre	35	0.333	0.083	15		
Décembre	51	0.486	0.012	15		

Tableau 1.7 : tests de tendance, général et pour chaque mois, et test d'homogénéité des tendances pour le rapportN/P sur l'Yser (RNB 89000) et pour les nitrites sur l'Hogneau (RNB 32000)

NO2 – I'Hogneau - RNB 32000							
S	Tau p(tau=0)						
364	0.26	7	9.1	0-7			
Chi-	2		Р				
20.4	20.45 0.039						
S(j)	Tau(j)	p(t	au=0)	n			
30	0.250	0	16				
37	0.308 0		.095	16			
50	0.417		.024	16			
58	0.483	0	0.009	16			
58	0.483	0.009 1					
68	0.567	0.002 10					
56	0.467	0.012 16					
31	0.295	0.125 15					
10	0.095	5 0.620 15					
-7	-0.067		).729	15			
-21	-0.200	0.299 15					
-6	-0.057	-0.057 0.766 15					



Figure 1.18 : évolution mois par mois, de Janvier à Décembre, du rapport N/P sur l'Yser (station RNB 89000).



Figure 1.19 : évolution mois par mois, de Janvier à Décembre, des concentrations en nitrites (en mg/l) sur l'Hogneau (station RNB 32000).

#### **I.4 - CONCLUSION**

A la suite de la première étude réalisée au laboratoire pour l'Agence de l'Eau Artois-Picardie (Leprêtre et Carpentier, 1994), les résultats basés sur les séries de données allant jusque début 1992 avaient abouti à une prévision à la hausse pour les phosphates, ammonium et nitrites et une prévision à la baisse des nitrates. Il avait néanmoins été noté qu'un début d'infléchissement apparaissait sur certaines des séries phosphates, mais ce changement était trop récent et le manque de recul n'avait pas permis qu'il soit pris en compte dans les prévisions. Un premier essai de validation de ces prévisions en 1994 a montré que l'infléchissement remarqué en 1992 s'était poursuivi et que les prévisions faites alors ne correspondaient plus à la nouvelle tendance apparue dans le début des années 90.

Ces résultats sont confirmés par dernières nouvelles analyses et prévisions réalisées. On peut d'abord noter que la baisse des apports organiques se traduit par une plus grande quantité d'oxygène disponible dans les cours d'eau. On voit une nette progression des concentrations en oxygène et des paramètres DBO5 et DCO. Les dernières observations suggèrent que cette tendance devrait persister dans les années à venir.

Alors que les concentrations en nitrates étaient globalement en baisse dans les années 80, elles montrent une augmentation au cours des dernières années sur lesquelles porte l'étude. Cette augmentation des teneurs en nitrates va de paire avec la diminution des autres formes de l'azote (ammonium, nitrites, azote Kjeldahl). On observe également, en de nombreux endroits, une diminution de l'azote total. La hausse des concentrations en nitrates n'est pas donc pas nécessairement due à une hausse des apports, mais elle serait plutôt la conséquence d'un transfert de l'azote des formes de l'azote des formes réduites vers les formes oxydées. L'amélioration globale de la qualité de l'eau et la meilleure oxygénation facilite la reminéralisation et on trouve donc moins d'azote organique et de nitrites qui se transforment en nitrates.

La baisse enregistrée sur les phosphates semble coïncider avec l'apparition sur le marché des lessives compactes et sans phosphates et surtout, avec la mise en place des accords signés entre l'Association des Industries des Savons et Détergents (A.I.S.D.) et le Ministère de l'Environnement sur la baisse des phosphates dans les détergents lessiviels qui s'est opérée dans le courant de l'année 1991. Cette coïncidence est d'autant plus remarquable que la baisse des concentrations en phosphates est notable surtout dans les zones urbaines et

industrielles, où cette source d'apport est la plus importante, et qu'elle est absente ou beaucoup moins nette dans les zones rurales.

On peut donc conclure, à l'échelle du bassin Artois-Picardie, à une amélioration de la qualité générale des cours d'eau, même si les risques d'eutrophisation persistent. Les changements les plus importants, observés et à venir, concernent principalement les zones fortement urbanisées de l'agglomération lilloise et du bassin minier dans son ensemble.

Un lien très fort est aussi mis en évidence entre les concentrations de certains éléments et les débits (principalement les nitrates et les matières en suspension). Dans de telles conditions, la prévision à partir des seules séries de concentration est risquée, car "polluée" par les fluctuations des conditions climatiques, la tendance d'évolution du paramètre étant masquée par sa relation avec le débit. Il convient donc d'apporter une nuance sur les prévisions à la hausse des nitrates. L'augmentation observée sur les dernières années est peutêtre due en partie à un effet du débit, car les années 93 et surtout 94-95 ont été pluvieuses et ont certainement accentué le lessivage des nitrates.

Il est donc intéressant d'envisager de prendre en compte l'influence des débits dans l'analyse de l'évolution des concentrations. De même, une bonne estimation de ces débits est indispensable pour parvenir à une évaluation plus précise des flux.

L'intérêt d'une étude à l'échelle du bassin Artois-Picardie, sur les stations du Réseau National de Bassin, est de permettre une gestion générale de la qualité de l'eau. Les séries chronologiques obtenues permettent de décrire et analyser l'évolution de la qualité de l'eau lors d'événement passés, tels que par exemple la fermeture ou l'ouverture de sites industriels, ou encore l'amélioration de réseaux d'assainissement et de stations d'épuration (Noppe *et al.*, 1997). Il est donc possible ainsi d'apporter l'information nécessaire pour le conseil en vue de nouvelles implantations. L'étude des tendances et la prévision statistique à court et moyen termes permettent de définir des priorités en matière de gestion de l'épuration pour le respect des objectifs de qualité. Ces prévisions de niveaux de qualité permettent également les prévisions budgétaires.

Un des objectifs de cette étude était également d'examiner la possibilité d'une automatisation complète des procédures utilisées dans l'analyse et la prévision des séries chronologiques. On voit qu'une telle automatisation n'est pas envisageable. A différentes étapes du processus, une décision "humaine" est toujours nécessaire. C'est le cas, par exemple, pour la localisation des ruptures de tendances mises en évidence par les *Cusum*. L'application d'une démarche "normalisée" peut convenir à la grande majorité des cas étudiés, mais pour un certain nombre de séries présentant des caractéristiques particulières (par exemple celles qui sont affectées par des seuils de détection), on se trouve dans des conditions qui rendent inapplicables une telle démarche "automatique".

La solution optimale résiderait dans une automatisation relative. La décision de poursuivre la démarche statistique jusqu'à la prévision doit se faire sous contrôle. Il est nécessaire de :

i) visualiser la série originale afin de détecter la présence de valeurs aberrantes ou le voisinage d'un seuil;

ii) visualiser la courbe Cusum pour décider de la date à partir de laquelle doit s'appuyer la prévision.

iii) s'assurer, là encore par une visualisation, de la pertinence de la prévision réalisée.

En revanche, il serait possible d'automatiser les liaisons entre les résultats statistiques (analyses, prévisions ou typologies) et un Système d'Information Géographique. Cela permettrait d'abord de valoriser ces résultats par l'intermédiaire de représentations cartographiques, de faciliter leur interprétation à l'échelle du bassin, et de permettre une meilleure transmission de l'information vers les différents acteurs de la gestion de la qualité de l'eau.

Enfin, outre la parenté méthodologique entre ce travail sur les séries du RNB et celui sur la série benthique de Gravelines, cette étude est aussi en relation avec les problématiques qui concernent le milieu marin. Dans des mers "côtières", comme la Manche ou la Mer du Nord, une connaissance précise des apports continentaux est indispensable à la compréhension du fonctionnement (ou parfois du dysfonctionnement) des écosystèmes marins. La maîtrise des séries en terme de flux est également un préalable à leur injection dans une modélisation des processus au niveau littoral et côtier. Leprêtre, A. Carpentier, P. 1997. Une méthode simple de prévision des tendances appliquée aux séries temporelles de qualité des eaux courantes. *C.R.Acad.Sci. Paris, Sciences de la vie*, 320 : 407-411.

# Une méthode simple de prévision des tendances appliquée aux séries temporelles de qualité des eaux courantes

A simple forecasting method applied to running waters quality time-series

#### ALAIN LEPRÊTRE\*, PATRICE CARPENTIER

Laborasoire d'écologie numérique, Ura CNRS 1363, universisé des sciences et technologies de Lille, 59655 Villeneuve-d'Ascq cedex, France

#### Résumé

Les données manquantes, l'échantillonnage irrégulier, les limites de détection et la non-normalité sont des problèmes fréquemment rencontrés dans l'analyse des séries chronologiques en écologie. Depuis plus de vingt ans, des données de qualités des eaux de surface sont recueillies dans le cadre du Réseau national de bassin. Une méthode est proposée permettant d'abord la régularisation des séries, la détection des changements de tendance et enfin la prévision des évolutions à court terme, à partir de la dernière tendance observée, à l'aide d'une régression non-paramétrique. L'exemple porte sur les mesures de concentration en phosphates et azote réalisées sur la Sambre (Nord).

Mots clés : séries temporelles, qualité des eaux courantes, tendances, prévision

#### ABSTRACT

Missing values, non-regular sampling, censored data and departure from normality are common problems in the analysis of ecological or environmental time series. For over 20 years, in the scope of the National Basin Network in France, numerous water quality time series have been collected. A method is proposed for the regularization of the series, the detection of trend changes and short-term forecasting, on the basis of the last detected trend, using a non-parametric regression. The method is applied to phosphates and nitrogen concentrations measured on the Sambre river (northern France).

Keywords: time-series, running water quality, trends, forecasting

#### Abridged version (see p. 410)

#### Introduction

Pour le suivi de la qualité des eaux courantes, opéré par les agences de l'eau dans le cadre du Réseau national de bassin (loi sur l'Eau de 1964), de nombreuses séries de données physicochimiques et biologiques recueillies sur l'ensemble des cours d'eau, mais à des intervalles irrégu-

Note présentée par Henri Décamps Note remise le 17 juin 1996, acceptée après révision le 7 avril 1997

C. R. Acad. Sci. Paris, Sciences de la vie / Life Sciences 1997, 320, 407-411 liers, ont été constituées en France depuis plus de vingt ans. Un premier objectif est de suivre les évolutions de différents paramètres de la qualité de l'eau (nitrates, phosphates, chlorophylle, DBO, DCO, matières en suspension,...) et de détecter des changements progressifs ou des ruptures brutales de tendance liées par exemple à des interventions humaines (installation de station d'épura-

<sup>\*</sup>Correspondance et tirés à part

tion, législation sur les phosphates des lessives, etc.). Un deuxième objectif est d'effectuer des prévisions d'évolution en vue d'établir les priorités d'intervention en matière de maintien ou de restauration de la qualité de l'eau.

Une méthode est présentée, permettant la régularisation des séries chronologiques irrégulières, la détection de la dernière rupture de tendance évolutive et la prévision d'une tendance à court ou moyen terme. La méthode permet la prise en compte de l'existence de seuils de détection des méthodes analytiques qui, pour certains paramètres physicochimiques et dans de nombreuses séries, entraînent une discrétisation des valeurs faibles.

#### Matériel et méthodes

#### Régularisation d'une série brute

La méthode proposée est dérivée de celle présentée par lbanez [1] suivant les travaux de Fox et Brown [2]. Une pondération a été ajoutée de façon à mieux prendre en compte les observations les plus proches et éviter ainsi un trop fort effet de lissage. Une série brute, concernant un paramètre physicochimique donné, est composée de *n* valeurs  $x_i$  mesurées à des temps irréguliers  $t_i$ . Une valeur interpolée  $X_j$  est calculée pour des intervalles réguliers aux temps  $T_j$  (avec j = 1 à N). On fait glisser une fenêtre de largeur *L* sur toute la période qui contient la série brute (figure 1). Si à un temps  $T_j$  donné correspond exactement un temps  $t_i$  de la série brute, alors la valeur  $x_i$  correspondante est retenue ( $X_j = x_i$ ) sans recourir à l'interpolation. Sinon, le calcul d'une valeur interpolée  $X_j$  (figure 1) est donné par :

$$X_{j} = \frac{1}{L} \cdot \frac{1}{n} \cdot \frac{1}{\sum_{i=1}^{n} w_{i}} \cdot \frac{x_{i}}{\sum_{i=1}^{n} x_{i}} \cdot F_{i} \cdot w_{i}$$

avec L = largeur de la fenêtre utilisée, centrée sur  $T_{\mu}$  Fi =

largeur de l'intervalle  $\left[\frac{t_i + t_{i+1}}{2}, \frac{t_{i-1} + t_i}{2}\right]$  contenu dans



temps  $T_j$  donné.

Le coefficient F, permet d'affecter un poids à la valeur  $x_i$ en fonction de l'intervalle de temps qui la sépare des valeurs  $x_{i-1}$  et  $x_{i+1}$ . Ce coefficient s'annule en dehors de la fenêtre. Une valeur  $x_i$ , située en dehors de la fenêtre peut contribuer partiellement à l'interpolation si le centre de l'intervalle  $[t_{i-1}, t_i]$  est compris dans la fenêtre. De même, une valeur  $x_i$  comprise dans la fenêtre peut ne contribuer que partiellement si une partie de l'intervalle  $F_i$  correspondant se trouve en dehors de la fenêtre d'interpolation.

Le coefficient  $w_i$  est un coefficient de pondération de  $x_i$ suivant l'inverse de la distance  $d_i$  qui sépare le temps  $t_i$  du centre de la fenêtre  $T_i$ .

Dans les exemples présentés, une fenêtre de L = 6 mois et un intervalle de temps de 1 mois entre les valeurs interpolées X, ont été choisis.

#### Détection des ruptures de tendance

La fonction Cusum, méthode graphique développée initialement pour le contrôle de la qualité en milieu industriel [3], a été employée en hydrologie [4] ou en océanographie [5].

À chaque temps  $T_j$  de la série régularisée, on associe la fonction suivante :

$$Cusum(T_j, k) = \sum_{j=1}^{T_j} X_j - \frac{T_j}{k} \cdot \sum_{j=1}^k X_j \quad \text{avec } 2 \le k \le N$$

La forme de la courbe obtenue permet de déterminer si une tendance existe et si cette tendance est continue ou discontinue. Le graphique s'interprète de la manière suivante :

En prenant 
$$k = N$$
, on obtient  $Cusum(T_j) = \sum_{i=1}^{T_j} X_j - T_j \cdot \vec{X}$ 



Figure 1. Régularisation d'une série brute de données x<sub>i</sub> en une série régulière de valeurs X<sub>i</sub> au moyen d'une fenêtre de largeur L centrée sur les X<sub>i</sub>.

C. R. Acad. Sci. Paris, Sciences de la vie / Life Sciences 1997. 320, 407-411

408

- si la série ne montre pas de tendance particulière,  $Cusum(T_i)$  reste voisin de zéro et la courbe reste proche de l'axe des abscisses ;

- s'il existe une tendance, la courbe Cusum s'éloigne de l'axe des abscisses ; la forme de la courbe est parabolique s'il s'agit d'une tendance monotone linéaire ;

- si la courbe présente des ruptures de pentes brusques, c'est que la série subit des modifications de tendance.

#### Prévision

La prévision d'évolution est réalisée sur la partie de la série qui suit la dernière rupture de tendance détectée par la fonction Cusum. Cette dernière partie de la série devra être suffisamment longue pour permettre l'application d'une technique de régression. Dans de nombreux cas, la distribution statistique des résidus obtenus après regression linéaire s'écartant fortement de la normalité, il est préférable d'appliquer une méthode de régression non paramétrique.

La méthode de régression non paramétrique utilisée est fondée sur l'algorithme de calcul du coefficient de corrélation  $\tau$  de Kendall,qui prend en compte les signes des différences entre tous les couples d'observations. Décrite initialement par Sen [6], elle a été adaptée par Hirsch et al. [7, 8] afin d'éliminer l'effet de la saisonnalité.

Pour une série régularisée mensuelle, on remplacera la notation  $X_j$  par la notation  $X_{a,m}$ , a étant l'indice correspondant à l'année et m l'indice du mois d'observation. On aura donc m = 1, 2, ..., n.

Pour un mois donné *m*, on calcule pour tous les couples  $(X_{a,m}, X_{a;m})$  tels que  $1 \le a < a' \le n$ , la pente :  $P_{a,a',m} = \frac{(X_{a',m} - X_{a,m})}{(a' - a)}$ 

L'estimateur de la pente b de la tendance est alors la médiane de toutes les pentes  $P_{a,a,m}$  obtenues, quel que soit le mois. L'ordonnée à l'origine de la droite de régression est obtenue en prenant la médiane des valeurs

 $(X_{a,m} - b/(12 + T_{a,m}))$ ,  $T_{a,m}$  étant l'abscisse temporelle de la valeur  $X_{a,m}$ .

La méthode reste applicable en présence de données manquantes.

#### Résultats

L'illustration de la démarche est donnée sur la série des valeurs obtenues de 1977 à 1995 pour la concentration en phosphates (notée PO<sub>4</sub>) et la concentration en azote total Kjehldal, (notée NTK) dans la station 4000 du Réseau national de bassin, station située sur la Sambre (Nord) à proximité de la frontière belge.

Après régularisation de la série  $PO_4$  (figure 2a), la fonction Cusum est appliquée (figure 2b). Elle met en évidence des ruptures de tendances. La dernière rupture, située début 1991, indique un infléchissement des teneurs en phosphates. C'est donc sur cette dernière partie de la série qu'est réalisée l'extraction de la tendance. Les droites de régression linéaire et non paramétrique sont reportées sur la figure 3a. Les pentes obtenues sont voisines, la régression non paramétrique montrant cependant une pente plus forte et apportant donc une prévision plus « optimiste » sur la baisse des teneurs en phosphates.

La même démarche est appliquée aux teneurs en NTK. La série (figure 3b) est ici affectée par les seuils de détection analytique, ce qui se traduit par un écrêtement de la partie inférieure de la courbe et une discrétisation des résultats de mesures. La distribution statistique des valeurs s'écartant fortement de la normalité, seule la régression non paramétrique est applicable.

#### **Discussion.** Conclusion

L'extraction de tendances évolutives dans les séries de données écologiques ou d'environnement se heurte le plus souvent au problème des valeurs manquantes et à l'irrégularité des séries obtenues. La méthode de régulari-





Figure 2. Régularisation et détection des ruptures de tendance sur une série « phosphates » enregistrée sur la Scarpe. a. Série brute (pointillés) et série régularisée (trait continu) ; b. graphique Cusum (les ilèches indiquent les ruptures de tendance).

C. R. Acad. Sci. Paris, Sciences de la vie / Life Sciences 1997, 320, 407-411



Figure 3. Extraction de la tendance sur la dernière partie homogène d'une série.

a. Régressions paramètrique (trait continu) et non-paramétrique (pointillés) sur la série « phosphates » ; b. régression non-paramétrique sur la série « NTK ».

sation présentée ici a l'intérêt de réaliser les interpolations en affectant un poids plus important aux données mesurées les plus proches, réduisant ainsi l'effet de lissage. La fonction Cusum appliquée aux données régularisées permet d'isoler le dernier tronçon homogène de la série en vue d'établir une prévision fondée sur cette dernière tendance évolutive. La régularisation n'est pas nécessaire à l'application de la fonction Cusum ; elle l'est cependant dans la prévision pour éviter une sous-représentation de certains mois ou certaines périodes et équilibrer la répartition des poids des observations dans la régression.

Le deuxième problème de traitement des données concerne la distribution statistique des valeurs de la série. Cette distribution est très rarement normale ce qui n'autorise pas l'emploi d'une technique de régression linéaire classique. Du fait de l'existence de seuils de détection analytique pour les paramètres physicochimiques étudiés, les valeurs basses sont discrétisées ce qui produit un écrêtement de la partie inférieure des courbes d'évolution et corrélativement une distribution très asymétrique de la distribution des valeurs mesurées. La méthode non paramétrique employée, inspirée du t de Kendall, permet de s'affranchir des conditions de normalité des techniques classiques. L'intérêt de cette méthodologie réside également dans sa facilité de mise en œuvre pour des opérations de traitement très répétitives. Ce travail s'inscrit en effet dans l'exploitation de séries concernant 68 stations du Réseau national de bassin et dix paramètres physicochimiques (NTK, phosphates, mais aussi nitrites, nitrates, matières en suspension, etc.). L'objectif est d'aboutir, au fur et à mesure de l'acquisition de nouvelles données, à la prévision des teneurs dans les années qui suivent, en vue de programmer les interventions sur le terrain : installation de stations d'épuration, amélioration du réseau de collecte des eaux urbaines ou industrielles, etc.

Enfin, sur les données déjà traitées, la technique a permis également de déceler l'influence qu'ont pu avoir certaines mesures agricoles ou industrielles. Sur la majorité des stations implantées en milieu rural, on a pu observé un infléchissement (faible, mais significatif) des teneurs en nitrates faisant suite à la mise en application des bilans azotés. De même, plus récemment, les séries réalisées sur les phosphates dans les zones urbaines ou périurbaines, montrent à partir de 1991–92, comme sur l'exemple présenté ici, une tendance à la baisse consécutive à la mise sur le marché des lessives dites « sans phosphates » [9].

#### ABRIDGED VERSION

#### Introduction

For over 20 years, French Water Agencies have collected water quality data (phosphates, nitrates, BOD, COD, etc.) from all the stations of the National Basin Network (RNB). The statistical analysis of these numerous time series presents several problems: non-regular sampling, departure from normality, censored data, etc. The purpose of this paper is to propose a simple and repetitive method for the regularization

410

of time series, detection of changes in the trends, and shortterm forecasting by a non-parametric technique.

#### Material and methods

#### Regularization

A moving window of length L is used to obtain interpolated values  $X_j$  for regular times  $T_j$  (j = 1-N) from the primary time series of  $x_i$  collected at non-regular times  $t_i$ :

C. R. Acad. Sci. Paris, Sciences de la vie / Life Sciences 1997, 320, 407-411

$$X_j = \frac{1}{L} \cdot \frac{1}{\frac{n}{m}} \cdot \sum_{\substack{i=1\\i=1}}^{n} x_i \cdot F_i \cdot w_i$$

with  $F_i$  = length of the interval  $\begin{bmatrix} t_i + t_{i+1} & t_{i-1} + t_i \\ 2 & 2 \end{bmatrix}$  contained within the window, L = length of the window, centered on  $T_j$  and  $w_i = \frac{1}{d_i} = \frac{1}{|t_i - T_j|}$  for a given  $T_j$ .

#### **Cusum function**

The Cusum function allows us to detect if there are any changes in trends and when these changes occur. For each time  $T_i$ , the associated Cusum function is :

$$Cusum(T_j) = \sum_{j=1}^{T_j} X_j - \frac{T_j}{k} \cdot \sum_{\substack{j=1 \\ j \neq 1}}^{k} X_j \quad \text{with } 2 \le k \le N$$
  
For  $k = N$ ,  $Cusum(T_j) = \sum_{\substack{j=1 \\ j \neq 1}}^{T_j} X_j - T_j \cdot \overline{X}$ 

If there is a monotonic trend, the function is quadratic. If there are changes in the trend, the curve shows sudden breaks of slope.

#### Forecasting

A linear regression method is used to forecast the trend on the last homogeneous part of the series detected by the Cusum function. Distribution of observed values often departs from normality, A non-parametric regression

#### Références

1. Ibanez F. 1991. Treatment of data deriving from the Cost 147 project on coastal benthic ecology: the within-site analysis. In : Space and Time Series Data Analysis in Coastal Benthic Ecology (B. Keegan, ed.), 5-43

2. Fox W.T., Brown J.A. 1965. The use of time-trend analysis for environmental interpretation of limestones. *J Geol* 73, 510-518

3. Page E.S. 1961. Cumulative sum charts. Technometrics 3, 1-9

4. Cluis D.A. 1983. Visual techniques for the detection of water quality : double-mass curves and Cusum functions. *Environ. Monitoring Assessment* 3, 173-184

5. Ibanez F., Fromentin J.M., Castel J. 1993. Application de la méthode des sommes cumulées à l'analyse des séries chronologimethod based on Kendall's Tau is used. To take into account seasonal effects, individual slopes for all couples of observations taken in the same month, in different years, are calculated. The overall slope estimator is the median of all these individual slopes.

#### Results

The method is applied to a phosphates data set collected from 1977 to 1995 on the Sambre river. The Cusum function shows a last breaking off in early 1991. Parametric and non-parametric regression lines are fitted to this last period of the time series. The method is applied to the Kjehldal nitrogen data set collected at the same station. In this case, data are censored by analytical detection limits. Thus, only non-parametric regression can be applied.

#### Discussion and conclusion

As part of the National Basin Network survey, a large number of water quality records has to be analysed. The proposed method can easily be used for such repetitive data treatments. It is robust enough to deal with the particular nature of the statistical distribution of these data.

This method has been able to detect changes in the trend of phosphate concentrations in urban zones from the early 1990s, which may be related to a new policy concerning the chemistry of detergents.

ques océanographiques. C. R. Acad. Sci. Paris, Sciences de la vie 316, 745-748

6. Sen P.K. 1968. Estimates of the regression coefficient based on Kendall's tau. J. Am. Statistical Assoc. 63, 1379-1389

7. Hirsch R.M., Slack J.R., Smith R.A. 1982. Techniques of trend analysis for monthly water quality data. *Water Resources Research* 18, 107-121

8. Hirsch R.M., Alexander R.B., Smith R.A. 1991. Selection of methods for the detection and estimation of trends in water quality. *Water Resources Research* 27, 803-813

9. Carbiener R. 1990. Rapport au secrétaire d'État auprès du Premier ministre chargé de l'Environnement sur les compositions lessivielles avec ou sans phosphates et protection des milieux aquatiques. 133 p.

C. R. Acad. Sci. Paris, Sciences de la vie / Life Sciences 1997, 320, 407-411

### Chapitre II

# Série benthique à long terme de Gravelines (Mer du Nord)

#### II.1 - Introduction.

#### II.1.1 Zone et site d'étude de Gravelines

La zone d'étude de Gravelines se situe dans la partie sud de la Mer du Nord, sur la façade nord du littoral de la région Nord-Pas de Calais, plus précisément entre Calais et Dunkerque. Dans cette zone, soumise à un régime macrotidal, l'intensité des courants de marée est le principal facteur d'organisation et de distribution des sédiments et des peuplements. L'atténuation des courants, depuis le détroit du Pas de Calais, vers l'est, a généré l'installation d'un gradient dans les faciès sédimentaires et dans la répartition des peuplement benthiques (Souplet et al., 1980 ; Prygiel et al., 1988). Cette zone reçoit des apports continentaux assez importants, principalement par l'Aa, mais aussi par les canaux et exutoires des wateringues au niveau des ports de Calais et Dunkerque. Elle peut également, sous certaines conditions, se trouver sous l'influence des eaux en provenance de l'estuaire de l'Escaut (Dewarumez et al., 1993). L'établissement du complexe industriel et portuaire de Dunkerque, sur une longueur de côte de près de 20 km depuis la rive orientale de l'Aa, a profondément remanié le littoral dans ce secteur.

La station de suivi à long terme des espèces macrobenthiques est située, près de la côte, à proximité de l'embouchure de l'Aa, par 8 m de fond. Elle se trouve dans le peuplement des sables fins envasés à *Abra alba*, peuplement qui forme un continuum le long des côtes de toute cette partie sud de la Mer du Nord (Dewarumez et al., 1986). Le sédiment est composé de sables fins (75 à 80%) et de vase (1 à 12%).

Cette station a été créée à l'origine dans le cadre de l'opération de surveillance de l'impact écologique du site de la centrale nucléaire de Gravelines et des installations de l'avant-port ouest de Dunkerque. La série temporelle de données benthiques constituée dans ce contexte comprend 81 relevés pour la période 1978-1992 sur laquelle porte notre travail. Au total, 151 espèces ou taxons ont été identifiés. Chaque relevé est constitué de la réunion de 10 prélèvements élémentaires réalisés avec une benne de Van Veen (Holme et Mc Intyre, 1984), opérant sur une surface de  $0,1 \text{ m}^2$ . Les abondances de chaque espèce sont donc directement exprimées en densité (nombre d'individus /  $m^2$ ).
Ce site et cette série d'observations sont aujourd'hui intégrés dans différents programmes de recherche nationaux et internationaux :

- le programme REcrutement en mer du NORd dans le peuplement à Abra alba (RENORA).

- le Programme National sur le Déterminisme du Recrutement (PNDR).

- le Programme National d'Océanographie Côtière, opération "séries à long terme" (PNOC-SLT).

C'est dans le cadre de ce dernier programme que s'inscrivent les travaux présentés ici. Les principaux objectifs du PNOC-SLT sont (Dauvin et Ibanez, 1995):

1) identifier les changements temporels des écosystèmes côtiers, à différentes échelles de temps;

2) identifier dans le déterminisme des séries les composantes climatiques, anthropiques et biologiques ;

3) identifier, à l'aide d'études intra et inter-sites, les composantes globales et locales;

4) comprendre les accidents ou variations dans la dynamique d'espèces marines ;

5) prédire les changements temporels des écosystèmes.

La série benthique de Gravelines a déjà fait par le passé l'objet de plusieurs études (*e.g.* Dewarumez et al., 1986, 1992). Dans le cadre du PNOC-slt et dans une étude comparant l'évolution de cette série avec trois autres séries prises dans des peuplements similaires de la Manche, Fromentin (1995) montre l'influence des phénomènes climatiques à méso-échelle sur l'évolution de la composition spécifique de ces peuplements. Un cycle d'à peu près 7 ans a pu être mis en évidence. Ce cycle coïncide avec un cycle de la température (Fromentin et Ibanez, 1994). Le site de Gravelines, qui est le plus septentrional des quatre sites étudiés, se révèle être celui qui présente la plus forte sensibilité aux variations climatiques, qui y sont d'ailleurs plus marquées. Fromentin précise néanmoins qu'une grande part de la variabilité du peuplement doit être associée à des variations des conditions locales.

Luczak (1996), grâce à une approche à plusieurs échelles spatiales de l'évolution de ce peuplement sur deux ans, montre l'influence de processus biotiques et abiotiques sur la distributions des organismes benthiques. Il constate néanmoins qu'une grande part de la variabilité observée reste inexpliquée. Il conclut également que les variations du lieu d'échantillonnage dues au positionnement et à la dérive du bateau n'ont pas de conséquences importantes, puisque les événements mis en évidence à l'échelle du suivi à long terme sont comparables à ceux perçus à une échelle spatiale de l'ordre du km<sup>2</sup>. Les résultats du suivi ne peuvent cependant pas être étendus à l'ensemble du peuplement (échelle 40 km<sup>2</sup>). La forte hétérogénéité spatiale de la répartition des organismes, même à très petite échelle, est cependant susceptible de faire apparaître dans la série des relevés "aberrants", qui peuvent fausser l'interprétation de la série temporelle.

## II.1.2 Conditions hydrodynamiques

L'interprétation des tendances, cyclicités ou variations épisodiques des peuplements, passe par une connaissance des conditions hydrodynamiques régnant dans le secteur d'étude. La zone d'étude est caractérisée par de faibles hauteurs d'eau : la profondeur n'y excède pas trente cinq mètres. La caractéristique principale de la morphologie des fonds sous-marins de ce secteur est la présence de nombreux bancs de sables, culminants de 10 à 25 m au-dessus des fonds environnants, et dont les sommets sont recouverts par moins de 5 m d'eau et peuvent être exondés lors des basses mers de vives-eaux (figure 2.1). Ces bancs, orientés parallèlement à la côte, forment des chenaux interrompus.

Le sud de la Mer du Nord est caractérisé par un fort hydrodynamisme de marée dû à la convergence au niveau du détroit du Pas de Calais. Les courants de marée alternatifs sont parallèles à la côte et suivent les chenaux existant entre les bancs. La vitesse de ces courants à l'Ouest de Gravelines atteint 1,8 à 2,3 nœuds (SHOM, 1968).

Le secteur est également soumis à l'action des houles. Les observations recueillies par Météo-France dans cette partie de la Mer du Nord montrent que les houles les plus énergétiques ont lieu de Décembre à Mai. Leur direction dominante est NNE. Le reste du temps, il existe également des houles de Nord qui deviennent prépondérantes en Juillet, Août et Novembre. La moitié de ces houles ont une hauteur comprise entre 1 m et 1,5 m. Les houles d'une hauteur supérieure à 5 m sont relativement rares (7%).

Les observations faites à proximité de la côte par le Laboratoire National d'Hydraulique (LNH, 1971), entre 1960 et 1968, en vue de la construction de l'avant-port Ouest de Dunkerque, montrent une prédominance de la houle de Nord (42%) et une fréquence élevée des houles de NW (18%), de NNW (14%) et de WNW (12%). La rareté des houles de NNE et l'absence de celles de NE au niveau du littoral sont expliquées par un affaiblissement et une rotation par la réfraction due aux bancs de sable. Leur période est généralement comprise entre 5 et 12 secondes, la période moyenne étant de 6 à 6,5 secondes. Ces houles atteignent rarement plus de 4 m.

L'intensité des phénomènes hydrodynamiques et les rythmes qui les caractérisent influencent les conditions de vie des organismes ainsi que les phénomènes de recrutement (Belgrano et al., 1990 ; Dewarumez et al., 1991 ; Lagadeuc, 1992). Les observations réalisées sur le terrain montrent que la station de Gravelines présente une forte sensibilité aux conditions hydrodynamiques et particulièrement à des événements "accidentels" - comme des tempêtes – caractérisés par une variabilité à petite échelle spatio-temporelle.

# II.1.3 Application d'un modèle hydrodynamique

La houle est le type même de phénomènes physiques qui peuvent agir à de telles échelles, d'autant que la faible profondeur à laquelle se situe le peuplement le rend plus particulièrement sensible à ce facteur. Afin de mettre en évidence l'hétérogénéité spatiale des conditions de houle et de révéler des zones de plus ou moins haute énergie, un code d'agitation de houle a été mis en œuvre au sein du Laboratoire.

Le modèle HISWA (HIndcast Shallow water WAves), développé au Department of Civil Engineering de la Delft University of Technology, est utilisé. Il s'agit d'un modèle numérique qui calcule les paramètres de la houle se propageant dans les zones côtières en présence de courant. Le modèle prend en compte la réfraction, la dissipation due au frottement sur le fond et au déferlement ainsi que la dissipation par le courant. La génération de houle par le vent peut être prédite par le modèle. Elle est cependant négligée dans la présente étude.

L'application du modèle couvre une zone qui s'étend sur 25 860 m de longueur sur le littoral et 17 240 m de la côte vers le large (fig. 2.1). Elle est discrétisée selon un maillage régulier de 60x40 mailles. Les calculs concernent un cycle de marée de vive-eau moyenne (coefficient 95) et des conditions de houle incidente de 1,2 m de hauteur significative et de période 6,5 s , provenant du NNE. La discrétisation temporelle consiste à effectuer une simulation à chaque heure marée en réactualisant le champ de courants de marée et les hauteurs d'eau sur l'ensemble de la zone (d'après les données du SHOM, 1968). Treize simulations ont donc été réalisées, de 6h avant à 6h après la pleine mer à Calais.

67

Le modèle calcule la hauteur significative (HS)<sup>1</sup>. La vitesse orbitale près du fond est reliée à cette quantité, ainsi qu'à la profondeur locale. Sans entrer dans le détail des équations, il s'avère que la variabilité spatiale de la hauteur significative se répercute donc sur la vitesse près du fond. Les différences mises en évidence par les hauteurs significatives sont accentuées au niveau du cisaillement de fond, qui est lui, fonction du carré de la vitesse.

La simulation à basse mer (PM -6h) montre l'existence d'un "couloir", situé devant Gravelines, dans lequel la hauteur significative de la houle reste plus forte qu'en d'autres points du littoral (Figure 2.2). Au droit de l'embouchure de l'Aa, où est située la station du suivi à long terme, la hauteur significative atteint des valeurs de 1,1 à 1,2 m, équivalentes à celles rencontrées plus au large. A l'approche de la pleine mer, le couloir de plus forte agitation s'estompe, pour se reformer après la pleine mer.

La simulation numérique révèle, surtout à basse mer, une hétérogénéité du champ de houle, liée à l'hétérogénéité bathymétrique de la zone étudiée, caractérisée par la présence des bancs et chenaux interrompus. Cette hétérogénéité hydrodynamique à petite échelle peut expliquer en partie la forte variabilité, spatiale et temporelle, enregistrée sur la composition spécifique et la structure des peuplements benthiques de cette zone.

<sup>&</sup>lt;sup>1</sup> La hauteur significative est la moyenne du tiers supérieur des vagues.



Figure 2.1 : Bathymétrie de la zone d'étude et situation de la station de suivi à long terme.



Figure 2.2 : Hauteurs significatives de la houle à basse mer (PM -6h) obtenus par simulation avec le modèle HISWA. L'isoligne représentée correspond à une H.S. de 1m.

#### II.1.4 Evolution des conditions climatiques

Une des principales caractéristiques de l'évolution des conditions climatiques pendant la période de 1978 à 1992 est le cycle de la température de 6-7 ans (Fromentin et Ibanez, 1994). Ces fluctuations de température, qui sont ressenties à l'échelle de la Manche, ont une influence sur la dynamique des abondances d'*Abra alba* et du peuplement dans son ensemble (Dauvin *et al.*, 1993 ; Fromentin *et al.*, 1997).

Le cycle de température, qui apparaît déjà sur la série des moyennes mensuelles (fig. 2.3), est clairement mis en évidence quand on réalise des moyennes mobiles annuelles sur cette série (fig. 2.4). On voit l'alternance de périodes froides en 78-79 et 85-86; avec des périodes douces en 82 et 88-89. Ces différences inter-annuelles sont surtout sensibles sur les mois d'hiver (fig. 2.5). Une décomposition saisonnière de la série par la méthode STL (*Seasonal Trend decomposition based on Loess*, Cleveland *et al.*, 1990) révèle, comme les moyennes mobiles, l'existence de ce cycle pluri-annuel (fig. 2.6). Cette méthode permet également de constater que les principales anomalies sont effectivement trouvées en période hivernale (hiver 78-79, fin 81 et hivers 84-85 à 86-87).



Figure 2.3 : moyennes mensuelles des températures de l'air à Dunkerque.



Figure 2.4 : moyennes mobiles sur les moyennes mensuelles de température.



Figure 2.5 : température moyenne mensuelle la plus froide de l'hiver (Janvier ou Février).



Figure 2.6 : série des températures avec extraction de la composante saisonnière par la méthode STL et lissage par la méthode Lowess.

L'évolution de la pression atmosphérique montre une tendance à la hausse constante sur toute la période étudiée (fig. 2.7). Un test de tendance de Mann-Kendall (*cf.* I.2.1.c) montre que cette tendance est très significative ( $\tau = 0.157$ ; p = 0.002).

L'évolution globale de ces paramètres climatiques indique donc une amélioration des conditions sur le site de Gravelines pendant la période considérée pour le suivi à long terme du peuplement à *Abra alba*.



Figure 2.7 : moyennes mensuelles de pression atmosphérique à Dunkerque et lissage par la méthode Lowess.

# II.2 - Détection des changements de la communauté à Abra Alba

## II.2.1 Problème posé

L'étude des changements temporels de la composition et de la structure quantitative des communautés est un objectif important du PNOC-SLT. Il s'agit à la fois d'analyser et de comprendre la dynamique de ces communautés, mais également de réaliser un suivi et une surveillance à long terme afin d'évaluer l'impact que peuvent avoir des perturbations, naturelles ou anthropiques, du milieu environnant. Les techniques d'analyse multivariable, comme l'Analyse des Correspondances (Benzécri 1973 ; Hill, 1973) ou différents types d'analyses de tableaux de distances (Analyse en Coordonnées Principales ou Multidimensionnal Scaling), sont des méthodes heuristiques qui font partie des outils couramment utilisés pour évaluer les variations temporelles des structures multispécifiques. Sans livrer une bibliographie longue et fastidieuse sur ces méthodes, on peut citer les synthèses de Legendre et Legendre (1984) et Jongman et al. (1995), et pour des applications au domaine benthique, les travaux de Warwick & Clarke (1991) et Ibanez et al. (1993).

Ces techniques d'ordination qui s'appliquent à des tableaux de contingence, croisant échantillons (dates d'observation) et espèces, montrent la trajectoire des points échantillons dans un espace réduit. Le plus souvent, la représentation est limitée aux deux premières dimensions. Si la puissance de ces méthodes à faire ressortir l'information de structures complexes est parfaitement établie, un certain nombre de réserves peuvent être émises quant à l'utilisation et l'interprétation des plans factoriels résultants.

Tout d'abord, une forte variabilité naturelle dans les données peut ne pas faire ressortir clairement une trajectoire. Deux dimensions sont souvent insuffisantes pour exprimer une part significative de la variance. Ceci est tout particulièrement vrai quand la taille des jeux de données traités est grande. L'interprétation doit alors tenir compte des axes subséquents, qui peuvent porter une information non négligeable. Il devient alors très difficile de suivre une trajectoire qui se déroule dans plusieurs dimensions et d'interpréter conjointement plusieurs plans factoriels. Un deuxième problème est que la trajectoire peut résulter à la fois d'une tendance à long terme et de composantes à plus haute fréquence comme, par exemple, des variations saisonnières. La superposition de composantes de fréquences différentes peut rendre la trajectoire extrêmement complexe et difficilement interprétable. Il est également possible qu'une des composantes masque les variations d'une autre. Ainsi, une forte variabilité saisonnière peut masquer une tendance globale.

De nombreuses méthodes de pré-traitement des données sont utilisées pour palier ces différents problèmes. Une première voie consiste à réduire la taille des jeux de données, en diminuant au départ le nombre d'espèces prises en compte dans l'analyse (Gray & Pearson, 1982). Les espèces les moins fréquentes et les moins abondantes sont alors considérées comme peu informatives, ou comme génératrices d'un bruit de fond qui nuit à l'analyse. De plus, la difficulté d'échantillonnage des espèces "rares" rend peu fiable l'information apportée par ces espèces. Il est donc couramment proposé d'éliminer les espèces qui ne sont pas présentes régulièrement dans les relevés, ainsi que celles qui n'ont que de faibles effectifs.

Une autre manière d'éliminer une part de variabilité est de réduire le nombre de relevés. Il est possible de s'affranchir des variations saisonnières en regroupant en un seul échantillon les différentes observations faites au cours d'une année. On constitue de cette façon un échantillon moyen annuel. Il est évident que pour permettre une comparaison entre les années, il est nécessaire que l'échantillonnage intra-annuel soit équivalent, c'est-à-dire une même densité d'observation au cours de l'année avec des échantillons pris à des périodes similaires, comparables d'une année sur l'autre. Une autre alternative est de ne s'intéresser qu'à une période précise de l'année.

Enfin, il est possible de réduire la variabilité par différentes méthodes de lissage ou de filtrage des données (Ibanez, 1991 ; Fromentin *et al.* 1997). Ces méthodes permettent d'extraire la tendance principale de l'évolution chronologique de l'abondance de chaque espèces en éliminant des variations supposées aléatoires ou se produisant à petite échelle de temps. Un inconvénient de ces lissages est qu'ils ne se font pas sur la structure multispécifique du peuplement, mais sur les espèces prise séparément les unes des autres. Le risque est donc de perdre ou tout au moins de modifier localement la perception des associations entre espèces.

Si elles peuvent permettre d'obtenir des résultats moins confus et plus facilement interprétables dans les analyses de l'évolution des communautés, ces méthodes présentent quand même un risque de perte d'informations. L'application de critères arbitraires de sélection peut faire disparaître des éléments essentiels. Si l'application de telles solutions se justifie dans la cadre d'étude à grande échelle et de comparaisons inter-sites, elle ne peut convenir à une étude locale, portant sur une seule série, comme celle abordée dans ce travail. Nous avons donc proposé une méthode qui permette de faire ressortir de façon claire l'évolution à long terme d'un peuplement, sans éliminer et sans filtrer *a priori* les données brutes récoltées sur le terrain.

#### II.2.2 Analyse de la série benthique

### II.2.2.1 Méthode

• Pondération des relevés

Les relevés ayant été effectués de manière irrégulière et suivant une fréquence variant d'une année à l'autre, il est nécessaire de procéder à un "rééquilibrage" des poids relatifs des différents relevés grâce à un système de pondération, de sorte que certaines périodes ou certaines années ne soient pas privilégiées a priori dans l'analyse des données.

Les données forment un tableau de contingence croisant échantillons (temps) et espèces. Soit une série observée à des temps irréguliers  $t_j$  avec j = 1 à n et avec le mois comme unité de temps. On dispose d'un tableau X où  $x_{i,j}$  est l'abondance de l'espèce i (avec i = 1 à s) au temps  $t_j$ .

L'échantillon collecté au temps j forme un vecteur X<sub>j</sub>. On considère que chaque valeur de X<sub>j</sub> est une estimation de l'abondance des espèces pour un intervalle de temps d'une longueur F<sub>j</sub> telle que :

$$F_{j} = \left(\frac{t_{j} + t_{j+1}}{2}\right) - \left(\frac{t_{j-1} + t_{j}}{2}\right) = \left(\frac{t_{j+1} - t_{j-1}}{2}\right)$$

Chaque valeur  $x_{ij}$  (ou chaque vecteur  $X_j$ ) est pondéré par le coefficient  $F_j$  correspondant. L'A.F.C. est appliquée à une matrice Y, de mêmes dimensions que X (n dates, s espèces), et avec des valeurs élémentaires  $y_{ij} = F_j \bullet x_{ij}$  (ou  $Y_j = F_j \bullet X_j$ ).

De manière à visualiser la trajectoire moyenne d'un cycle annuel dans l'A.F.C., un échantillon moyen est estimé pour chaque mois m (avec m = 1 à 12). Pour une année donnée y, la valeur attribuée au mois m est la valeur correspondante  $X_j$  si elle existe, alors  $X_{m,y} = X_j$ . Si il n'y a pas de  $X_j$  correspondant dans le jeu de données, alors la valeur du plus proche  $X_j$  est attribuée à  $X_{m,y}$ . Quand la donnée manquante est située exactement entre deux échantillons existants, alors  $X_{m,y}$  est la moyenne de ces deux plus proches voisins.

Soit  $\overline{x_m}\,$  , la moyenne des  $x_{{}_{m,y}}$  pour toute la période considérée, on a :

$$\overline{\mathbf{X}_{m}} = \sum_{\mathbf{y}} \mathbf{X}_{m, \mathbf{y}} \quad \text{and} \quad \sum_{\mathbf{y}} \sum_{m}^{12} \mathbf{X}_{m, \mathbf{y}} = \sum_{j}^{n} \mathbf{F}_{j} \mathbf{X}_{j}$$

Ces valeurs moyennes  $\overline{X_m}$  sont utilisées comme colonnes supplémentaires dans l'A.F.C.

#### Mise en évidence de la tendance à long terme

On utilise les coordonnées de l'échantillon  $X_j$  et des mois moyens  $\overline{X_m}$  sur les k premiers axes (  $\alpha = 1$  à k ) de l'A.F.C. Les variations sur les autres axes sont alors considérées comme des fluctuations résiduelles ou du bruit.

Pour chaque échantillon, la distance  $D_{j,m}$  entre l'échantillon et la moyenne du mois qui lui correspond est calculée :

$$\mathbf{D}_{j} = \sqrt{\sum_{\alpha=1}^{k} \left( \operatorname{Coord}_{\alpha}(\mathbf{X}_{j}) - \operatorname{Coord}_{\alpha}(\overline{\mathbf{X}}_{m}) \right)^{2}},$$

avec Coord<sub>a</sub>( $X_j$ ) coordonnée de l'échantillon j (vecteur  $X_j$ ) sur l'axe  $\alpha$  ( $\alpha = 1$  à k) et Coord<sub>a</sub>( $\overline{X}_m$ ) coordonnée du mois moyen  $\overline{X}_m$  sur cet axe.

On obtient ainsi une expression au cours du temps de la différence entre les échantillons collectés et une situation moyenne, non affectée par les variations saisonnières.

# Détection des variations à court terme de la composition de la communauté

La composition d'une communauté est aussi affectée par des événements soudains : des événements naturels, comme des tempêtes, ou des événement d'origine anthropique comme des pollutions accidentelles. L'objet de cette seconde méthode de visualisation d'une série multivariable est de détecter des modifications inhabituelles de la communauté entre deux périodes d'échantillonnage successives.

On calcule d'abord la distance  $D_{j,j+1}$  entre deux échantillons successifs j and j+1 dans un sous-espace à k dimensions :

$$D_{j,j+1} = \sqrt{\sum_{\alpha=1}^{k} \left( Coord_{\alpha}(X_{j}) - Coord_{\alpha}(X_{j+1}) \right)^{2}}$$

On calcule également la distance  $D_{m,m'}$  entre les moyennes des mois m and m' correspondants à ces deux échantillons :

$$D_{m,m'} = \sqrt{\sum_{\alpha=1}^{k} \left( Coord_{\alpha} \left( \overline{X_{m}} \right) - Coord_{\alpha} \left( \overline{X_{m'}} \right) \right)^{2}}$$

On exprime alors le rapport entre ces deux distances en fonction du temps. L'objectif de cette méthode est de détecter, dans la succession des échantillons, des moments où la différence entre deux échantillons est plus importante (ou plus petite) que la différence moyenne qui aurait due être observée. Ces ruptures dans les profils devraient être mises en relation avec des événements inhabituels, agissant à une petite échelle temporelle, sur la structure de la communauté.

#### II.2.2.2 Application à l'analyse de la série

Les méthodes précédemment décrites ont été appliquées à la série Graveleines. L'interprétation des données est plus détaillée dans la publication jointe (Carpentier *et al.*, 1997). Nous ne ferons ici qu'un résumé rapide des principaux résultats.

Les deux premiers axes de l'A.F.C. (fig. 2.8) représentent 35.3% de l'inertie totale (respectivement 20,9% et 14,4%). L'ordination sur ce plan 1x2 montre un groupe principal

d'échantillons caractérisés par les lamellibranches Abra alba et Tellina fabula et quelques autres espèces associées (*Pectinaria koreni, Nephtys hombergii, Phyllodoce mucosa...*). Opposés sur le premier axe, on trouve :

- les échantillons collectés au printemps 1979 et de 1981 à 1983, caractérisés par la forte abondance de *Lanice conchilega*,

- les échantillons de 1991 et 1992 associés à la présence d'Ensis directus.

Le second axe isole les échantillons de 1986 à 1988, associés principalement à *Spiophanes bombyx* et *Macoma balthica*. La plupart des échantillons de ce groupe ont été collectés au printemps et en été (d'Avril à Septembre).



Figure 2.8 : Projection dans le plan 1x2 de l'AFC des 81 relevés et des 151 espèces. Seules le espèces caractéristiques sont indiquées. Les cercles représentent les relevés originaux et les points noirs représentent les mois moyens.

Les mois moyens  $\overline{X_m}$ , projetés comme colonnes supplémentaires, révèlent un cycle annuel qui s'exprime sur les deux axes. La dispersion des échantillons sur le plan 1x2 (ainsi que sur les autres axes) résulte donc à la fois des variations inter-annuelles et des variations saisonnières. Viennent également s'ajouter les variations résiduelles acycliques. La trajectoire suivie par les relevés successifs sur le plan factoriel est compliquée et difficile à interpréter. La présente méthode permet de visualiser cette trajectoire multidimensionnelle de la communauté en fonction du temps, en éliminant les variations saisonnières et une partie des fluctuations résiduelles.

En se limitant à k = 3 axes (soit 46,3% de l'inertie totale), les distances D<sub>j</sub> ont été calculées (*cf.* § 2.2.1). Le profil chronologique de D<sub>j</sub> montre une variation cyclique à long terme de la communauté (fig. 2.9) dont la période est d'environ 6-7 ans. Les écarts maximums par rapport à la situation moyenne sont trouvés pour les périodes de 1979 à 1981 et de 1986 à 1987. Ces variations sont corrélées aux changements à long terme (inter-annuels) des conditions climatiques. Une moyenne mobile annuelle (12 termes) a été réalisée sur les températures mensuelles mesurées à Dunkerque de façon à éliminer la saisonnalité et à mettre en évidence la tendance globale.



Figure 2.9 : tendance cyclique de la structure de la communauté révélée par la distance entre les relevés et les mois moyens correspondant dans les 3 premières dimensions de l'AFC (en gris). Les moyennes mobiles des températures sont représentées en échelle inversée (en noir).

Le calcul des corrélations entre les distances  $D_j$  (transformées en logarithme) et les moyennes mobiles des températures a été fait en introduisant différents décalages. Ces corrélations avec retard montrent des valeurs négatives importantes (de -0.50 à -0.60; p<1‰ pour le coefficient de corrélation de rang de Spearman) pour des décalages de 6 à 12 mois entre les deux séries. La corrélation (négative) la plus forte ( $r_s = -0.60$ ; p < 1‰) est trouvée

pour un délai de 9 mois. Ce délai pourrait correspondre au temps de réponse de la communauté : les mauvaises conditions climatiques qui peuvent apparaître dès l'automne ne se répercuteraient sur la composition du peuplement qu'au moment du recrutement de l'année suivante. L'influence négative d'une baisse des température sur la reproduction de nombreuse espèces laisse des ressources disponibles et peut favoriser les recrutements importants d'espèces plus opportunistes.

En conclusion, l'intérêt de la méthode est de mettre nettement en évidence des tendances et cycles à long terme sans nécessité de lissage ou d'élimination préalable d'espèces. La méthode des distances D<sub>j</sub> permet d'abord une visualisation claire de ce cycle. En effet, l'interprétation est facilitée, car on obtient une série chronologique univariable et non une trajectoire projetée dans un plan, comme dans la majorité des méthodes de visualisation couramment utilisées. Il est également plus aisé d'établir un lien entre les modifications de la structure multispécifique et des variables externes, comme l'a vu ici avec la température.

La série benthique de Gravelines est donc affectée par un cycle lié à des changements de la température. Il y a une modification de la composition du peuplement d'un état "normal", correspondant à des conditions de température douce, vers une situation perturbée, caractérisée par la dominance forte de quelques espèces, lorsque la température (hivernale surtout) est froide. Cependant, le cycle de la communauté ne se confond pas avec le cycle des espèces prises séparément. Ceci n'est vrai que pour certaines espèces, comme *Abra alba* par exemple, dont les abondances sont aussi liées à l'évolution des température (Dauvin et al., 1993). On constate aussi, que pendant les deux phases de perturbation observées (périodes froides), les espèces qui dominent ne sont pas les mêmes. Il y a donc un déterminisme global de l'évolution de la communauté gouverné par des fluctuations climatiques à méso-échelle. Mais il y a un indéterminisme local dans sa composition précise. Ce sont certainement des phénomènes qui agissent à des échelles de temps plus courtes qui déterminent quelles espèces vont dominer pendant les périodes perturbées.

La deuxième méthode présentée, basée sur les distances entre échantillons successifs, a pour but est de détecter des modifications brutales de la composition du peuplement. Elle met en évidence un pic important correspondant essentiellement à l'apparition d'*Ensis directus* en 1991 (Luczak *et al.*, 1993), puis deux fortes déviations qui correspondent au début des périodes froides (fig. 2.10). Il y a donc, à ce moment là, un changement marqué. Le reste du graphique ne montre pas d'autres irrégularités notables. Il y aurait donc, après le début d'une période froide, maintient de la nouvelle situation et ensuite, retour progressif vers une situation moins perturbée : il n'y a pas de deuxième pic à la fin de la période froide. Les autres modifications qui apparaissent sont difficiles à expliquer. Elles sont certainement à relier à des phénomènes s'exprimant sur de petites échelles de temps et d'espace. Nous ne disposons cependant pas d'informations suffisantes sur de tels phénomènes, et il est impossible d'interpréter les nombreuses variations mises en évidence.



## II.2.2.3 Généralisation de la méthode à d'autres métriques

La démarche qui consiste à calculer une distance à une situation moyenne à partir d'une A.F.C., peut être opérée en s'appuyant sur d'autres types d'analyses multivariables, donc d'autres choix de métriques. Ainsi, nous avons repris cette démarche sur la base d'une ordination des relevés par MDS (positionnement multidimensionnel) forme non-métrique (Kruskal, 1964) appliquée à une matrice de dissimilarités de Bray-Curtis (Bray et Curtis, 1957). Cette métrique est en effet souvent préconisée pour les données benthiques (*e.g.* Warwick et Clarke, 1991). L'analyse porte sur le tableau complet (81 relevés et 151 espèces), auquel on ajoute les 12 mois moyens. Il n'est pas procédé ici à une pondération pour un rééquilibrage dans le temps comme cela a été fait avec l'A.F.C. Le calcul de la matrice de dissimilarités et l'ordination sont réalisés sur l'ensemble des relevés, y compris les 12 mois moyens. Ces derniers sont pris en compte directement dans les calculs, et non en tant qu'éléments supplémentaires, comme dans l'AFC.

La projection sur le premier plan des relevés et des 12 mois moyens montre, comme dans l'AFC, la forte différence entre la période d'Avril à Juillet, marquée par les pics d'abondance de certaines espèces, et le reste de l'année (fig. 2.11).

Le calcul de la distance au mois moyen est réalisé suivant le principe décrit pour l'AFC (§ II.2.2.1). Il se fait à partir des coordonnées des échantillons et des mois moyens obtenues par la projection en 3 dimensions de la MDS.

On retrouve, avec la représentation graphique de la distance, le cycle de 6-7 ans avec un maximum de distance en 79-80 et 86. L'évolution de la distance correspond toujours au cycle de la température (fig. 2.12).



Figure 2.11 : NMMDS sur dissimilarité de Bray-Curtis avec les dates (points) et les mois moyens



Figure 2.12 : distance entre les relevés et les mois moyens correspondant dans les 3 premières dimensions de la NMMDS sur dissimilarité de Bray-Curtis (en gris). Les moyennes mobiles des températures sont représentées en échelle inversée (en noir).

La corrélation avec retard avec la température reste significative pour des délais jusqu'à 12-13 mois. Sa valeur maximale est trouvée pour un délai de 8-9 mois et pour un délai de 3 mois (tableau 2.2).

Délai	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
r <sub>s</sub>	-0.31	-0.32	-0.36	-0.38	-0.37	-0.35	-0.34	-0.35	-0.38	-0.38	-0.33	-0.33	-0.32	-0.27	-0.21	-0.17	-0.19	-0.18	-0.15
p. Ho	<1%	<1%	<1%	<1%	<1%	<1%	<1%	<1%	<1%	<1%	<1%	<1%	<1%	<5%	n.s.	n.s.	n.s.	n.s.	n.s.

 Tableau 2.2 : Coefficient de corrélation de Spearman entre la distance (Bray-Curtis)

 et la température (Moyennes Mobiles) avec différents délais (en mois).

Si on compare aux résultats obtenus à partir de l'AFC, on aboutit globalement à la même interprétation : il y a une modification "cyclique" du peuplement, liée à des fluctuations de la température. La corrélation entre les distances obtenues par l'AFC et la MDS est très significative (Corrélation Spearman  $r_s=0.737$ ;  $p \approx 0$ ) (fig. 2.13). La meilleure corrélation avec la température est obtenue dans les deux cas avec un délai de 8-9 mois.



On peut cependant noter un certain nombre de différences. Les corrélations avec la température sont moins bonnes, et le délai d'action est moins net en MDS qu'avec l'AFC. L'AFC semble également (sur cet exemple au moins) plus à même de faire ressortir la variabilité inter-annuelle par rapport à la variabilité saisonnière intra-annuelle. Les distances à partir de la MDS, même si elles font apparaître le cycle, présentent des variations intra-annuelles qui sont d'un ordre de grandeur quasi équivalent aux variations globales. Il y a donc un effet du choix de la métrique sur la perception des différentes échelles temporelles.

#### II.2.3 - Typologie des profils entre espèces

Nous avons appliqué la méthode, décrite au chapitre I.2.2 et utilisée sur les séries de qualité physico-chimique des eaux douces, pour comparer les profils d'évolutions des principales espèces de la série benthique de Gravelines. La méthode se base sur la définition que deux profils se ressemblent s'ils présentent des phases de montées et descentes simultanées. Pour disposer de séries qui présentent vraiment un profil, c'est à dire pas trop de zéros (absence de l'espèce), l'étude ne porte que sur les espèces qui sont présentent dans au moins la moitié des relevés. Les séries chronologiques d'abondances de 29 espèces sont donc conservées pour l'analyse (tableau 2.3).

Nous avons choisi de réaliser la typologie sur les profils des fonctions Cusum appliquées aux séries d'abondances (cf. § I.2.1.b), car comme le souligne Ibanez et al. (1993), les courbes des sommes cumulées permettent d'évaluer plus efficacement les synchronismes et les changements de tendance dans les séries. Un léger lissage est appliqué aux courbes Cusum afin d'éliminer le bruit et la variabilité locale.

CODE	ESPECE	CODE	ESPECE
ABA	Abra alba	NEO	Eunereis longissima
CAC	Capitella capitata	NOL	Notomastus latericeus
CCB	Cerebratulus sp	OPT	Ophiura texturata
CHS	Chaetozone setosa	OWF	Owenia fusiformis
CRC	Crangon crangon	PEK	Pectinaria koreni
DIB	Diastylis bradyi	PHM	Pholoe minuta
ENA	Ensis arcuatus	PHU	Phyllodoce mucosa
ETN	Eteone longa	SAG	Sagartia troglodytes
EUS	Eumida sanguinea	SPM	Spio martinensis
HAU	Harmothoe lunulata	SPB	Spiophanes bombyx
LAC	Lanice conchilega	SPS	Spisula subtruncata
MAB	Macoma balthica	STB	Sthenelais boa
MAM	Magelona mirabilis	TEF	Tellina fabula
MYB	Mysella bidentata	URP	Urothoe poseidonis
NEH	Nephtys hombergii	ļ	

Tableau 2.3 : liste des 29 espèces sélectionnées et code utilisés dans les figures.

L'analyse de la projection en MDS non-métrique et de la classification (fig. 2.14 et 2.15) montre qu'on peut d'abord distinguer un premier groupe d'espèces qui présentent un (ou plusieurs) pic d'abondance au début de la période du suivi. On retrouve dans ce groupe *Lanice conchilega*, *Sagartia troglodytes*, *Ensis arcuatus* et *Pectinaria koreni*. Au centre de la projection un groupe, composé de *Notomastus latericeus*, *Ophiura texturata*, *Macoma balthica*, *Spisula subtruncata*, *Owenia fusiformis* et *Chaetozone setosa*, présente des pics en fin du suivi. Sur la gauche du plan on trouve les espèces qui, comme *Abra alba*, *Nephtys hombergii*, *Sthenelais boa* ou *Phyllodoce mucosa*, suivent le cycle de la température, avec deux périodes de fortes abondances (la seconde souvent plus marquée) pendant les périodes douces. A l'opposé (en bas à droite) on trouve les espèces avec une dynamique inverse au groupe précédent, avec *Tellina fabula*, *Eteone longa* et *Diastylis bradyi*. Dans une position intermédiaire entre ces groupes, on trouve des espèces présentant les plus fortes abondances plutôt en milieu de période (86-88), comme *Magelona mirabilis*, *Spiophane bombyx* et *Capitella capitata*.

La méthode présentée permet donc d'identifier les groupes d'espèces présentant des évolution conjointes, sans prendre en compte les valeurs des abondances, mais uniquement les sens des évolutions et leur concordance. Elle s'avère d'autant plus efficace qu'elle est couplée ici avec les sommes cumulées qui permettent de mieux faire ressortir les changements et ruptures dans les séries. Il serait également possible de comparer par cette méthode les profils d'évolution des abondances d'espèces avec les profils temporels de différents facteurs environnementaux.



Figure 2.14 : NMMDS sur les indices de co-évolution entre profils Cusum et quelques profils représentatifs.



Figure 2.15 : classification hiérarchique sur les distances entre profils Cusum

# II.3 - Diversité

## II.3.1 Evolution de la diversité taxonomique sur la série originale

Nous nous sommes intéressé à l'évolution de la diversité spécifique sur la série benthique de Gravelines. L'évolution de cette diversité est abordée ici par l'application des descripteurs classiques de la diversité, tels que la richesse spécifique et l'indice de Shannon. Nous y appliquons également la technique d'estimation de H' par la méthode des Courbes Rangs-Information Cumulées (CRIC), proposée au chapitre IV.4, et la méthode des pentes P1-P2 des Diagrammes Rangs-Fréquences (DRF) (Mouillot, 1994; méthode explicitée ciaprès).

L'évolution chronologique de la richesse spécifique (fig. 2.16) montre, sur l'ensemble de la période, une tendance à la hausse qui peut être mise en relation avec celle des différentes variables climatiques (température, pression atmosphérique et vents). On retrouve l'influence des variations cycliques de température : la richesse observée diminue pendant la période froide de 1986. La chute de la richesse en 1991 est à associer à l'"accident" évoqué dans Carpentier *et al.* (1997), dont on a vu qu'il avait provoqué une forte défaunation dans la zone autour du site du suivi à long terme.



Figure 2.16 : évolution chronologique de la richesse spécifique.



Figure 2.17 : évolution chronologique de la densité en nombre d'individus/m<sup>2</sup>.

On constate cependant que la densité des organismes (fig. 2.17) augmente de façon similaire. L'estimation de la richesse étant fortement dépendante du nombre d'individus collectés, on peut se demander si les variations de la richesse qui sont observées ici révèlent une réelle évolution de la diversité ou si, au contraire, elles ne sont qu'un artefact lié à une augmentation de la densité ? Ce problème est évoqué au chapitre IV.3.4.

L'évolution de l'indice de Shannon H' présente une forte variabilité, mais il n'apparaît pas de tendance ou de cycle nettement perceptibles (fig. 2.18). Tout au plus peut-on constater une baisse très légère des valeurs de H' pendant les périodes froides. On observe surtout des chutes ponctuelles de diversité (les valeurs de H' descendent nettement en dessous de 1) qui correspondent aux relevés caractérisés par des abondances exceptionnelles d'une espèce (par exemple *Lanice conchilega* au début du suivi).

Les séries chronologiques obtenues pour la régularité (l'evenness de Pielou, 1975) et pour l'indice de diversité de Simpson montrent les mêmes fluctuations que H' (fig. 2.19 et 2.20). Aucune tendance particulière ne peut être mise en évidence. Ces résultats vont dans le même sens que les observations de Fromentin *et al.* (1997), qui compare les évolutions de la structure de la communauté aux modèles stochastiques de distributions de fréquence de Tokeshi (1990, 1993). La structure de la communauté présente une forte variabilité qui ne peut être reliée, ni aux phénomènes climatiques à méso-échelle, ni aux variations de sa composition spécifique.



Figure 2.18 : évolution chronologique de l'indice de diversité de Shannon H'.



Figure 2.19 : évolution chronologique de l'indice de diversité de Simpson.



L'estimation de H' par la méthode CRIC donne des résultats presque identiques à ceux obtenus avec la formulation classique de l'indice de Shannon (fig. 2.21). Si on se réfère aux propriétés et aux performances que présentent la méthode CRIC (*cf.* § IV.4.2), on peut conclure que l'effort d'échantillonnage réalisé sur la série benthique de Gravelines est suffisant pour permettre des estimations fiables de l'indice de Shannon H'.



Figure 2.21 : évolution chronologique de l'estimation de H' par la méthode CRIC.

La méthode des pentes, P1 et P2, dont le principe et les caractéristiques sont décrits par Mouillot (1994), est basée sur l'observation que les diagrammes rangs-fréquences (DRF), dans leur représentation bilogarithmique, présentent souvent deux parties quasi-linéaires concernant respectivement les espèces de forts et celles de faibles effectifs. Le DRF peut ainsi être résumé à l'aide de deux droites de régression. La méthode utilisée ici est une droite de régression orthogonale ou droite des moindres rectangles (Teissier, 1948). La première droite, donnant la pente P1, est calculée sur les 4 premiers rangs (les 4 espèces les plus abondantes). Des valeurs élevées de cette pente traduisent une situation de nette dominance, où une ou deux espèces présentent des effectifs très importants. Une pente faible caractérisera au contraire une distribution plus équitable des individus entre les premières espèces. La deuxième pente P2 est calculée sur le reste des espèces (à partir du cinquième rang), en excluant les espèces dont les effectifs ne dépassent pas 5 individus dans l'échantillon qui constituent la "queue de distribution" du DRF. Ces espèces sont souvent sous-échantillonnées et peuvent introduire un biais dans l'estimation de P2. Une pente faible caractérisera

L'application de cette méthode à la série Gravelines montre d'abord que l'évolution de la pente P1 est comparable à celle des indices de diversité évoqués précédemment (fig. 2.22). Elle ne présente ni tendance, ni cycle particulièrement marqués. On retrouve les pics de dominance, représentés ici par des pentes fortes et qui sont associés à des valeurs faibles des indices de diversité.



Figure 2.22 : évolution chronologique de la pente P1.

La pente P2 ne présente pas non plus de schéma d'évolution à long terme particulier (fig. 2.23). Elle est par contre, parmi toutes les méthodes que nous avons utilisées pour mettre en évidence l'évolution de la communauté benthique de Gravelines, tant dans sa structure que dans sa composition, la seule qui donne visuellement l'impression d'un cycle saisonnier. On voit que, à quelques exceptions près, les fortes pentes caractérisent les relevés de fin de printemps et d'été, alors que les pentes faibles correspondent surtout aux périodes hivernales. La structure de la communauté, révélée ici par les espèces les moins abondantes, évolue donc selon un cycle saisonnier, avec une régularité plus forte en hiver, et une situation de dominance l'été.

Cette observation rejoint l'hypothèse émise dans la publication ci-jointe (Carpentier et al., 1997), dans laquelle nous faisions remarquer qu'une communauté présente différents niveaux d'organisations, et que les espèces moins abondantes peuvent révéler des évolutions différentes de celles qui apparaissent avec les espèces dominantes. Les espèces les plus abondantes, par leur dynamique souvent "chaotique", peuvent masquer dans les analyses certaines des caractéristiques fondamentales de l'évolution de la communauté.



Figure 2.23 : évolution chronologique de la pente P2.

Le tableau 2.4 présente les corrélations entre les différents indices de diversité que nous avons utilisés. Les indices de Shannon, Simpson, l'estimateur CRIC, la régularité et la pente P1 sont tous très fortement corrélés entre eux. Par contre, la pente P2 et la richesse ne présentent aucune corrélation significative avec les autres indices. La pente P1 a une corrélation un peu plus forte avec l'indice de Simpson. Il caractérise donc bien des situations de dominance.

	<u></u>							
	P1	P2	Shannon	Régularité	Simpson	Richesse		
P2	0.079							
Shannon	-0.878	-0.124						
Régularité	-0.881	-0.156	0.967					
Simpson	-0.934	-0.097	0.962	0.949				
Richesse	-0.044	0.094	0.130	-0.085	0.072			
H'cric	-0.876	-0.123	0.995	0.976	0.958	0.089		

Tableau 2.4 : coefficients de corrélation de Spearman entre les différents indices de diversité. En grisé, les coefficients significatifs au risque  $\alpha = 5\%$ .

# II.3.2 Evolution de la diversité à différentes échelles de perception

#### II.3.2.1 Principe et méthode

L'idée de base est de vérifier si l'évolution de la diversité de la communauté est perçue de manière identique ou suit la même tendance quand on fait varier l'échelle de perception, autrement dit quand le "grain" temporel d'observation varie. Pour ce faire, il est nécessaire, dans un premier temps, de disposer d'une série temporelle régulière à grain fin (à petite échelle de temps) qu'il sera possible ensuite de transformer, par sommation selon un fenêtre mobile de largeur croissante, en série à grain de plus en plus grossier.

La procédure de régularisation de la série est inspirée de celle qui a été appliquée aux données spatiales du GDR Manche (*cf.* chap. III). Cette procédure, qui s'appuie sur les résultats d'une factorielle, a l'intérêt de prendre en compte les interactions entre espèces. La régularisation de la série pour une espèce donnée se fait ainsi en tenant compte de ses relations avec l'ensemble des autres espèces.

Le principe est donc le suivant :

- on soumet le tableau de données (la série chronologique multivariable) à une AFC.
- on applique aux coordonnées des échantillons (dates) sur les différents axes, la méthode de régularisation présentée au chapitre I.2.1.a, qui permet d'aboutir à des

"pseudo-coordonnées" pour des dates régulières ; seuls sont utilisés les axes factoriels principaux jusqu'à concurrence de 95% de l'inertie totale du nuage de points, ce qui constitue donc un filtrage.

on applique la formule de reconstitution (cf. chapitre III.2.2) à partir de ces coordonnées, on arrive ainsi à la constitution d'un "pseudo-tableau de données", donnant les effectifs des différentes espèces à chacune des dates d'une série régulière.

Comme dans l'application réalisée sur les données du GDR, la principale difficulté rencontrée est de passer de fréquences relatives à des fréquences absolues – donc d'estimer l'effectif total d'individus (toutes espèces confondues) à une date donnée.

Nous avons choisi de régulariser sur un pas de temps de 2 mois, d'Avril 1978 à Décembre 1992, soit 6 échantillons par an, ce qui est à peine supérieur à la fréquence moyenne d'échantillonnage de la série originale (81 relevés sur 15 ans). A partir de cette échelle de base, nous avons regardé comment se modifiait la perception de l'évolution temporelle de la diversité, lorsqu'on change la résolution (lorsqu'on augmente le "grain" de l'échantillonnage). Pour passer de l'échelle de 2 mois à des échelles supérieures, on cumule les échantillons, en faisant "glisser" sur la série une fenêtre dont on fait varier la longueur. L'échelle "4 mois" s'obtient donc en cumulant deux échantillons successifs de l'échelle "2 mois", l'échelle "6 mois" en cumulant trois échantillons, etc...

#### II.3.2.2 Résultats obtenus

Les séries chronologiques multivariables obtenues permettent de constituer, à différents degré de résolution, des séries temporelles de l'indice de Shannon H' et des paramètres "P1-P2" (cf. § II.3.1).

En ce qui concerne l'indice de Shannon (fig. 2.24), on constate que l'évolution de H', perçue à l'échelle de 2 mois ou aux échelles immédiatement supérieures, montre une forte variabilité sans laisser apparaître de tendances ou de cycles particuliers. Ce n'est qu'à partir du degré de résolution de 12 mois, donc l'échelle annuelle, qu'une tendance à l'augmentation est détectable. Cette tendance croissant est de plus en plus marquée quand on travaille à des degrés de résolution de plus en plus grossiers. Parallèlement, ces grandes échelles, on voit apparaître une "cyclicité" sous forme de deux vagues de diversité, qui rappelle le cycle d'évolution de la température sans pour autant lui correspondre exactement.

Les pentes P1 et P2 des DRF ne montrent également de tendance particulièrement marquée que pour les échelles de perception supérieure à l'année. La pente P1, calculée sur les quatres espèces les plus abondantes, montre de plus en plus nettement, jusqu'à l'echelle 36 mois, une tendance à la baisse (fig. 2.25). Cette diminution de la pente correspond à une plus grande régularité de la distribution des effectifs entre les espèces dominantes. On constate que le début la période est effectivement marquée par les très fortes abondances *de Lanice conchilega* rencontrées en 79 et de 81 à 83. Sur la fin de la période, cette forte dominance d'une seule espèce et beaucoup moins nette et on observe des changements beaucoup plus fréquents dans les espèces dominantes : *Lanice conchilega* est encore présente, mais on note aussi les fortes abondances de *Spiophanes bombyx* et *Abra alba*, entre autres. Le fait qu'un plus grand nombre d'espèces différentes dominent au cours de la dernière partie du suivi, explique donc aussi bien l'augmentation de l'indice de Shannon, que la diminution de la pente P1, aux plus grandes échelles de perception.

L'évolution de la pente P2 ne présente pas de tendance qui soit en rapport avec celles de H' ou de P1 (fig. 2.26). Elle montre par contre, à grande échelle, une variation cyclique qui est similaire à celle de la température. L'augmentation de P2 avec la température suggère donc une plus grande régularité dans les effectifs des espèces moins abondantes lors des périodes froides. Une structure moins équilibrée s'installe dans les périodes douces. Les changements à méso-échelle de la température, qui ne semblent pas affecter de manière significative la structure et la diversité du peuplement lorsqu'on prend en compte les espèces les plus abondantes, semblent donc se répercuter sur la structure du peuplement au niveau des espèces moins abondantes.



Figure 2.24 : évolution chronologique de l'indice de Shannon H' pour différentes échelles de perceptions.



Figure 2.25 : évolution chronologique de la pente P1 des DRF pour différentes échelles de perceptions.



Figure 2.26 : évolution chronologique de la pente P2 des DRF pour différentes échelles de perceptions.

# II.4 - Quelques réflexions sur l'échantillonnage

A l'image de la majorité des fonds benthiques littoraux, le site de Gravelines est le siège d'une forte hétérogénéité de la micro-répartition spatiale des organismes, qui se traduit à petite échelle spatiale par l'existence de "micro-faciès" de communautés. L'échantillonnage des peuplements benthiques conduit donc souvent à des relevés unitaires dont la composition et/ou la structure quantitative s'écartent nettement de celles de l'ensemble de la communauté. Ces relevés, ressentis alors comme "aberrants", peuvent conduire, comme le fait remarquer Luczak (1996), à une interprétation erronée de l'évolution temporelle du peuplement. Par conséquent, eu égard à la forte variabilité que présente la série benthique de Gravelines et à la difficulté à mettre en évidence ses évolutions, nous nous sommes interrogé sur la valeur de l'échantillonnage pratiqué : celui-ci était-il suffisant pour permettre des analyses quantitatives fiables ?

La justification du choix de l'effort d'échantillonnage remonte aux premières études réalisées sur le site en 1976. L'examen des courbes aire-espèces obtenues sur quelques relevés montrait que les courbes se stabilisaient pour le cumul de 5 réplicats ("coups de benne") et Souplet (1978) en concluait que 5 réplicats étaient suffisants pour obtenir un échantillon correct de la faune benthique dans le peuplement à *Abra alba*. Le choix d'un échantillon constitué de 10 réplicats pour le suivi ultérieur paraissait donc tout à fait adéquat. Notons cependant que la justification de la taille du prélèvement reposait sur un critère qualitatif (le nombre d'espèces) et non pas quantitatif (densité de chaque espèce ou indice de diversité).

Pour quatre relevés de la série – à savoir ceux de Novembre 1983, Mars, Avril et Mai 1984 - nous disposions du détail de la composition des 10 réplicats. A l'aide de techniques de rééchantillonnage numérique, nous avons, dans un premier temps, regardé l'effet de l'hétérogénéité qui s'exprime entre les réplicats sur la qualité de l'estimation de la composition quantitative du peuplement. Nous avons ensuite également observé quel était l'effet de l'augmentation de l'effort d'échantillonnage sur l'évaluation de la richesse spécifique et de l'indice de Shannon.



# II.4.1 Problèmes d'estimation de la composition quantitative du peuplement

On peut tout d'abord remarquer que les quatre relevés utilisés présentent des densités très différentes. En Novembre, la densité est forte et atteint 5436 individus/m<sup>2</sup>. En revanche, elle est faible en Mars (440 ind./m<sup>2</sup>) et augmente à nouveau en Avril (802 ind./m<sup>2</sup>) et en Mai (1831 ind./m<sup>2</sup>). Néanmoins, au regard de la série complète (*cf.* II.3.1), on ne peut pas dire que ces variations soient représentatives d'un cycle saisonnier.

Nous avons d'abord réalisé des intervalles de confiance *Bootstrap* de l'abondance des différentes espèces récoltées. Dans ce but, on tire au hasard et avec remise dix réplicats parmi les dix réplicats observés, et on calcule ensuit l'abondance totale pour chaque espèce. Les intervalles de confiance à 95% sont obtenus à partir de la distribution observée d'après 10 000 échantillons ainsi simulés.

Les résultats montrent qu'il existe effectivement une forte hétérogénéité de la microrépartition des individus. Les intervalles de confiance de la densité donnés par les simulations sont larges et montrent que la densité mesurée pourrait varier du simple au double dans les échantillons prélevés (tableau 2.5).

	Novembre	Mars	Avril	Mai
Densité observée	5436	440	802	1831
IC à 95%	3442 - 7939	316 - 575	671 - 942	1117 - 2797

Tableau 2.5 : densités mesurées et leurs intervalles de confiance Bootstrap.

Cette forte variabilité est liée à l'agrégation que présente la répartition spatiale de certaines espèces. On peut voir que les effectifs récoltés d'*Abra alba* par exemple, sont très variables d'un réplicat à l'autre (fig. 2.27). En novembre 83, deux réplicats (soit 1/5 de l'échantillon) contiennent les 2/3 des individus collectés. Cela conduit à une forte incertitude sur l'estimation de la densité de cette espèce, comme le révèlent les intervalles de confiance à 95% (tableau 2.6). Par contre, ce n'est pas le cas des effectifs de *Tellina fabula* qui montrent une certaine homogénéité entre les réplicats (tableau 2.6 et fig. 2.28, à gauche). La répartition spatiale des individus de cette espèce est beaucoup plus régulière.

Le caractère agrégatif de la répartition des individus est aussi observé sur d'autres espèces, mais parfois seulement de façon temporaire (tableau 2.7). C'est le cas de *Phyllodoce mucosa* (fig. 2.28, à droite) et de *Lanice conchilega* en novembre 83 ou de *Notomastus latericeus* en mai 84.
Tableau 2.6 : effectifs récoltés d'Abra alba et de Tellina fabula et leurs intervalles de confiance Bootstrap.

		Novembre	Mars	Avril	Mai
Abus all s	Densité observée	1305	87	500	1245
	IC à 95%	486 - 2383	54 - 134	378 - 653	679 - 2027
Talling fabula	Densité observée	328	41	33	59
Tenina Jaouta	IC à 95%	268 - 391	29 - 53	16 - 57	40 - 80

Tableau 2.7 : effectifs récoltés de quelques espèces et leurs intervalles de confiance Bootstrap.

	Phyllodoce mucosa (11/83)	Lanice conchilega (11/83)	Notomastus latericeus (05/84)
Densité observée	726	1219	47
IC à 95%	180 - 1598	877 – 1669	9 - 107



Figure 2.27 : répartition des effectifs d'*Abra alba* dans les 10 réplicats des relevés de Nov. 83 (à gauche) et Mai 84 (à droite). Les barres représentent les effectifs et la courbe les pourcentages cumulés. Les valeurs reportées expriment les effectifs en pourcentages et pourcentages cumulés (entre parenthèses).



Figure 2.28 : répartition des effectifs de *Tellina fabula* (à gauche) et *Phyllodoce mucosa* (à droite) dans les 10 réplicats du relevés de Nov. 83. Les barres représentent les effectifs et la courbe les pourcentages cumulés. Les valeurs reportées expriment les effectifs en pourcentages et pourcentages cumulés (entre parenthèses).

# II.4.2 Effet de l'effort d'échantillonnage sur l'estimation de la diversité

Pour mettre en évidence l'effet de l'effort d'échantillonnage sur l'estimation de la richesse spécifique, nous avons appliqué une procédure différente du *Bootstrap* précédemment utilisé. Nous avons calculé la richesse moyenne obtenue en prenant toutes les permutations possibles des réplicats et en prenant des échantillons constitués de 5 à 9 réplicats. Pour 10 réplicats, la seule combinaison possible est en effet l'échantillon observé.

La figure 2.29 présente, pour les quatre relevés, les courbes d'évolution de la richesse moyenne ainsi que les enveloppes, c'est à dire les valeurs maximum et minimum obtenues sur l'ensemble des permutations. On constate que dans aucune des quatre situations, les courbes de la richesse moyenne n'atteignent un palier. Il est évident que 10 réplicats (soit 1 m<sup>2</sup>) ne sont pas suffisants pour obtenir une bonne estimation de la composition spécifique du peuplement. Cependant, on peut voir qu'en Novembre, il est possible avec seulement 5 réplicats, d'atteindre la richesse observée sur la totalité des 10. Le même constat peut être fait pour les trois autres dates avec 7 ou 8 réplicats.

Les différents estimateurs de la richesse spécifique évoqués au chapitre IV.3.2 ont été appliqués aux 4 relevés (tableau 2.8). Au vu des performances réalisées par ces estimateurs au cours des simulations (cf. § IV.3.3), il apparaît également que la richesse observée sur ces échantillons sous-estime fortement la richesse vraie du peuplement, quel que soit le type d'estimateur utilisé, y compris les estimateurs Jackknife, considérés souvent comme les plus performants (Palmer, 1990, 1991 ; Baltanas, 1992). On notera que des écarts importants entre les résultats fournis par les différents estimateurs correspondent, dans les simulations, aux situations dans lesquelles le biais sur l'estimation de la richesse vraie est lui aussi important.

Tableau 2.8 : estimations par différentes méthodes de la richesse spécifique sur 4 relevés de la série Gravelines.

	Novembre 83	Mars 84	Avril 84	Mai 84
S observée	38	36	30	37
Bootstrap	41.0	41.2	34.7	42.7
Jackknife 1	44.3	46.8	40.8	50.5
Jackknife 2	47.0	50.0	47.5	60.4
S∞	38.4	39.6	34.0	41.0



Figure 2.29 : valeurs moyennes, maximales et minimales de la richesse spécifique obtenues pour toutes les permutations des réplicats sur 4 relevés de la série Gravelines en fonction de l'effort d'échantillonnage.

Les mêmes problème sont rencontrés dans l'estimation de l'indice de diversité H' de Shannon. La moyenne et l'intervalle de confiance à 95% de H' ont été calculés à partir de 10 000 simulations *Bootstrap* (tirage avec remise dans les 10 réplicats) avec des efforts d'échantillonnage variant de 5 à 10 réplicats. Les résultats obtenus vont dans le sens les conclusions du chapitre IV, confirmant que des échantillons de relativement petites tailles sont suffisants pour obtenir une estimation "stabilisée" de H'. Cela est d'autant plus vrai que les échantillons de la série présentent souvent une structure de dominance assez marquée. Le fait de passer de 5 à 10 réplicats ne change quasiment pas la valeur du H' d'échantillon (fig. 2.30). Augmenter la taille de l'échantillon n'améliore cependant que très peu la précision de l'estimation, caractérisée par une faible diminution des intervalles de confiance.

La faible dispersion des valeurs de H' observée sur l'échantillon de Novembre 83 peut s'expliquer par le fait que les abondances des quatre principales espèces de ce relevé (*Abra alba, Eumida sanguinea, Lanice conchilega* et *Phyllodoce mucosa*) sont fortement liées : les plus fortes abondances des 4 espèces sont enregistrées au sein des mêmes réplicats. Les proportions relatives de ces espèces restent donc très stables dans les échantillons simulés et les valeurs de H' présentent donc peu de variation. Ce n'est pas les cas pour les relevé suivants.



Figure 2.30 : valeurs moyennes et intervalles de confiance obtenus pour l'indice de Shannon H' par les simulations Bootstrap sur 4 relevés de la série Gravelines en fonction de l'effort d'échantillonnage.

L'étude détaillée de ces quatre relevés montre que l'effort d'échantillonnage qui a été appliqué dans le suivi du site est insuffisant pour permettre une estimation précise des densités des espèces du peuplement. Certaines espèces présentent en effet un caractère agrégatif fortement marqué. L'incertitude sur l'estimation de l'abondance de ces espèces étant importante, les aléas de l'échantillonnage peuvent donc conduire à des distribution de fréquences observées assez variables d'un prélèvement à l'autre. Carpentier, P. Dewarumez, J.M. Leprêtre, A. 1997. Long-term variability of the *Abra alba* community in the southern bight of the North Sea. *Oceanologica Acta*, 20 : 283-290.



Long-term trends Time series Multivariate methods Benthos

> Tendances à long terme Séries chronologiques Méthodes multivariées Benthos

	Patrice CARPENTIER *, Jean-Marie DEWARUMEZ <sup>b</sup> and Alain LEPRÊTRE *						
	<sup>a</sup> Laboratoire d'Écologie Numérique, URA-CNRS 1363, Université des Sciences et Technologies de Lille, Bât. SN3, 59655 Villeneuve-d'Ascq Cedex, France.						
	<sup>b</sup> Station Marine de Wimereux, URA-CNRS 1363, Université des Sciences et Technologies de Lille, BP 80, 62930 Wimereux, France.						
	Received in revised form 28/06/96, accepted 05/07/96.						
ABSTRACT	The Abra alba community from the Gravelines area has been studied since 1976. During this period, sudden and unpredictable variations of the species composition have been observed. These important changes affect a limited number of species without changing the bionomical structure of the community. The species in question are found to be characteristic of this community, and are the more abundant ones. Multivariate statistical methods were used to study quantitative data on the benthic population collected from 1978 to 1992. This analysis was performed in association with air temperature data. Results showed the existence of a six- or seven-year cycle, and this is confirmed by the observation of two complete periods. This cycle, which is found to be related to variations in climatic conditions, is generated by characteristic species of the community. Nevertheless, the species mainly responsible for the two "maxima" phases of the cycle are not the same. A different situation emerges during the years 1991-1992, with a new incoming species, <i>Ensis directus</i> , and modification of the nature of the sediment. The study of the community performed without the representative species, taking into account less abundant and less frequent species, shows a different kind of						
RÉSUMÉ	Variabilité à long terme du peuplement à Abra alba dans la baie sud de la Mer du Nord.						
	Le peuplement à Abra alba de la région de Gravelines fait l'objet d'un suivi à long terme depuis 1976. Pendant cette période, des fluctuations brutales et imprévisibles de sa composition spécifique ont pu être mises en évidence. Ces profondes altérations intéressent un nombre restreint d'espèces dites « caractéristiques du peuplement », en fait les plus abondantes, sans remettre en cause la structure bionomique du peuplement. Le travail porte sur le traitement statistique multivariable des données quantitatives recueillies de 1978 à 1992 sur le peuplement benthique d'une station. Ce traitement est réalisé en relation avec des données climatologiques. Les résultats obtenus ont permis de mettre en évidence l'existence d'un cycle de six ou sept années, confirmée par l'observation de deux périodes complètes						

Ce cycle, qui a pu être mis en relation avec des variations des conditions climatiques, est généré par les espèces caractéristiques du peuplement. Les espèces principalement responsables des deux phases « maxima » du cycle ne sont cependant pas les mêmes. Les années 1991-1992 présentent une situation différente, marquée par l'apparition d'*Ensis directus* et par la modification des caractéristiques granulomètriques du site.

Une étude du peuplement réalisée sans les espèces caractéristiques, en prenant en compte les espèces moins abondantes ou plus rares, montre une évolution différente.

Oceanologica Acta, 1997, 20, 1, 283-290.

#### INTRODUCTION

Benthic communities along the French coast, in the southern bight of the North Sea, have been studied for nearly twenty years. This area is affected by severe anthropogenic perturbations: dense maritime traffic; industrial complexes; harbour installations. It is exposed to particular natural conditions: the proximity of the Dover Strait induces powerful hydrodynamic processes. Gradients of reduced current strength create several macrobenthic assemblages, which are distributed according to the nature of the sediment (Souplet et al., 1980; Davoult et al., 1988; Prygiel et al., 1988; Dewarumez et al., 1991, 1992). The present study examines temporal variations of the Abra alba community. This muddy fine sand community exists along the entire coast, at a depth no greater than 10 m, from Cape Blanc Nez to the Scheldt estuary and beyond (Dewarumez et al., 1978; Dewarumez, 1979; Belgrano et al., 1990).

The purpose of this paper is to detect the existence of cyclical trends or long term cycles in the changes of the benthic community in question, and to relate these changes to variations of global effects, such as climatic conditions. Several studies have pointed to the importance of long-term changes in meteorological features, and especially temperature, on the changes of benthic communities (Glémarec, 1979; Gray and Christie, 1983; Dauvin *et al.*, 1993; Ibanez *et al.*, 1993; Fromentin and Ibanez, 1994). Numerical methods to display these developments and their inter-relationships are proposed.

In this first part of the study, the complete data set was used. As in most quantitative studies of benthic time series, the analysis takes mainly into account the more abundant species, and accords less importance to less abundant and rare species. But the organization of a community shows different levels, and less abundant species can reveal a different pattern of development than the more abundant ones (Frontier and Viale, 1993). The second part of the study focuses on these less abundant but not necessarily rare species, and analyses the evolution of the community without its more representative species.

#### MATERIAL AND METHODS

#### Study area and sampling

The study area is located in the southern part of the North Sea, off the French coast, between Calais and Dunkirk (Fig. 1). Samples were taken in the *Abra alba* community, at a station near Gravelines, situated 1.5 sea miles distant from the mouth of the Aa river, at a depth of 8 metres. Data are quantitative; samples were collected with a Smith McIntyre or a Van Veen grab (10 replicates of 0.1 m). Values are expressed in density: number of individuals/m<sup>2</sup>.

The data relate to the macrobenthic population sampled between 1978 and 1992. During this period, 81 samples were taken, in which a total of 151 species or taxa was described. There were important changes in the frequency of sampling from year to year (from 10 to as few as 4 samples), and sampling was irregularly spaced in time.

#### Data analysis

Correspondence Analysis (or Reciprocal Averaging) on the contingency table crossing samples and species (or taxa) was performed to study the temporal evolution of the community (Benzecri *et al.*, 1973; Hill, 1973; Legendre and Legendre, 1984; Leprêtre, 1988).

The aim of the study is to determine the long-term evolution, *i.e.* inter-annual variations. As already stated sampling was irregular in time, and there was an important heterogeneity in the number of samples collected per year. The first difficulty lay in equilibrating years, in such a way that they were all represented in the the same fashion in the correspondence analysis.

In order to do this, we assumed each sample to be representative of a period of time, and weighted data of each of the samples with the length of the period. We used the month as time unit, and the origin  $t_0$  was set in January 1978. For a series of samples observed at times  $t_1, t_2, \ldots, t_n$ , the weight  $w_i$  associated with the sample taken at time  $t_i$  was:

$$w_i = \frac{t_i + t_{i+1}}{2} - \frac{t_{i-1} + t_i}{2}$$

Values in the sample taken at  $t_i$  were multiplied by  $w_i$ .

Correspondence analysis was performed on the weighted data. We added the annual and monthly means of samples as supplementary variables in the analysis (*i.e.* not active in the analysis). To calculate these mean values, missing samples were replaced by the nearest collected sample in time.

<u>104</u>



Figure 1

Study area.

Interpreting multivariate analysis with a large collection of data is difficult, especially in the case of time series. The presentation of results on a factorial plane is often confused, and a trajectory does not permit a clear indication of long-term trends.

We propose two different methods of presenting the results obtained from the correspondence analysis in order to demonstrate the evolution of the community in a chronological manner. Co-ordinates on a limited number of axes from the correspondence analysis are used to calculate distances between samples.

In the first method, we used co-ordinates on the most significant axes of samples and mean values for each month (added as supplementary variables in the correspondence analysis). For each sample, the distance between the sample and the mean of the month corresponding to the sample is calculated. Then we have an expression in time of the difference between collected samples and a mean situation, unaffected by seasonality.

For the second method, we first calculate the distance between two successive samples and the distance between the means of the months corresponding to the two samples. Then we express the ratio between these two distances as a function of time. The purpose here is to show when, in the succession of samples, the difference between two samples is higher (or lower) than the mean difference that should have been observed. This can reveal when unusual, short-time scale events occur rather than showing up a long-term trend.

#### RESULTS

#### Analysis of the complete data set

First, the correspondence analysis performed on the total set of data shows that species strongly contributing to the first inertia axes are characteristic species of the community. Ordination in the plane of axes I (20.85% of total inertia) and II (14.40%) shows a main group of samples characterized by Abra alba, Tellina fabula, Pectinaria koreni, Nephtys hombergii, Phyllodoce mucosa, Eumida sanguinea and Notomastus latericeus (Fig. 2a). Two subgroups may be distinguished, opposing Abra alba and Pectinaria koreni to Tellina fabula and Nephtys hombergii. The first axis separates from this main group samples collected in 1979 and from 1981 to 1983, associated with Lanice conchilega. The second axis isolates samples from 1986 to 1988, associated with Spiophanes bombyx and Macoma balthica. Most of the samples of the two groups were taken in spring and summer, during the period from April to September. Years 1991 and 1992 are characterized by the appearence of Ensis directus, a North-American species, observed for the first time in this area in June 1991 (Luczak et al., 1993). The third axis (11.0%) opposes 1980 to the main group, associated with Pectinaria koreni, Mya truncata and Cerastoderma edule.

Mean years and mean months are projected as supplementary variables on the axes of this first analysis. Projection on plane I-II of the mean years confirms the inter-annual trend (Fig. 2b). The community departs from its original position in 1978. It stays away during the period 1979-1983 (1980



(b)

### Figure 2

Ordination of samples and species in the plane of axes I and II from the C.A. performed on the complete data set. (2a).
Projection in this plane of mean years and mean months as supplementary variables (2b).

000





- Distance to mean month calculated from the first 3 axes of the C.A. (3a).

- Distance to mean month calculated from the first 5 axes of the C.A. (3b).

- Distance to mean month from first 3 axes of the C.A. (solid line) and 12-term moving average performed on monthly temperatures (dotted line) (3c).

marks a departure from the original situation on the third axis) and then returns in 1984 and 1985. It moves away a

LONG-TERM CHANGES IN AN ABRA ALBA COMMUNITY

second time from 1986 to 1988 and returns to the original situation in 1989-1990. It then shows the special nature of the years 1991 and 1992. Projection of the mean months (as supplementary variables) distinguishes the April to July period from the remainder of the year, and shows that this late spring and early summer period, when recruitment occurs, is determinant for the evolution of the site.

The distance to the mean month calculated from the first three axes of the correspondence analysis shows a cyclical variation of the community (Fig. 3a). An important seasonal variation can also be observed. Performing the calculation on the first five axes shows, in addition, the discontinuity of years 1991-1992 (Fig. 3b). Maximum deviations from the mean situation, found from the end of 1979 to 1981 and from 1986 to 1987, correspond to changes in climatic conditions. A twelve-term moving average is performed on the monthly temperature measured at Dunkirk in order to remove seasonality and to assess the inter-annual trend. The distance calculated on the first three axes follows with a delay of almost one year the temperature variation (Fig. 3c): high distances correspond to low temperatures. A correlation analysis was performed between distance and smoothed temperature data with different time lags. The optimal negative value for Spearman's rank-order correlation coefficient is obtained for a 9-month delay. This negative correlation keeps large values for lags from 4 to 12 months (Table 1).

The ratio of distance between successive samples and distance between mean months shows numerous peaks, especially in the first years of the study (Fig. 4). Two major discontinuities appear in 1980 and at the beginning of 1987.

#### Analysis without characteristic species

The cyclical evolution revealed by the complete data set analysis is due to a few species which are the more abundant ones. A group of 19 species accounts for more than 95% of the contributions to inertia of the five first axes of the correspondence analysis. To determine whether or not less abundant species share the same evolution, we applied the same methodology to a reduced data set. Species with contributions to the inertia of the first five axes of the correspondence analysis higher than the mean of all contributions were removed. To avoid the effects of excessively sporadic species, species appearing less than ten times in the 81 samples were also removed. Forty species were finally retained.

The first axis of the correspondence analysis performed on these data was generated by only one species, *Venerupis* 

Table 1

Spearman's rank-order correlation of distance vs. moving average smoothed temperature with different time lags.

Lag (months)	I	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Spearman's r <sub>s</sub>	-0.29	-0.36	-0.39	0.45	-0.50	-0.53	-0.52	-0.54	-0.60	-0.57	-0.51	-0.42	-0.30	-0.22	-0.17	-0.08
Proba ( $\rho_* = 0$ )	< 5º/o	< 1º/o	< 1º/₀	< 1º/00	< 1º/	< 1º/	< 1º/₀	< 5°/,	n.s	n.s	n.s					

287



Figure 4

Ratio of distance between two successive samples and distance between the corresponding mean months calculated from the first 3 axes of the C.A.

*pullastra*, very abundant in 1982 and 1983. This species was then removed and a second analysis was performed. Ordination in plane I-II with the projection of mean years shows a first group from 1978 to 1983. Years 1984 and 1985 become isolated from this group, associated with *Mysella bidentata* (Fig. 5). Year 1986 returns to the first

group. From 1987 to 1992, years separate from the main group, without any homogeneity between them.

Distance to the mean month is calculated from coordinates on the first five axes of the correspondence analysis (Fig. 6). Results do not show clearly a cyclical variation as does the analysis with all species. In the first part, until 1985, the variation of the distance appears to show a slight opposition to the variation due to main species: a "low" phase in 1981 and a "high" one in 1984-1985. But from 1987, distance seems to increase, following a continuous, regular drift. The ratio of distance between successive samples and between mean months only shows an important discontinuity at the beginning of 1985 (Fig. 7).

#### CONCLUSION-DISCUSSION

The methods of calculating distance to a reference with results from a correspondence analysis proposed here permit the demonstration of changes in a community as a function of time. This method may be easier to read than a factorial plane, and facilitates the establishment



#### Figure 5

Ordination of samples and species in the plane of axes I and II from the C.A. performed without characteristic species. Mean years are added as supplementary variables.



Figure 6

Distance to mean months calculated from the first 5 axes of the C.A.



Ratio of distance between two successive samples and distance between the corresponding mean months calculated from the first 5 axes of the C.A.

of a relationship with environmental data. Distance to the mean month confirms the existence of a periodic modification of this Abra alba community, strongly related to temperature. Following cold years, one or a very few species have high densities, and dominate the community. However, the cyclical variation of the community does not correspond to a cyclical variation of the species, considered separately each from the other. The pattern of change of the community is not the same as the patterns of change of the representative species of this community. During the fifteen years of the study, two cold periods induce a deviation of the community from is basic state, but the main species involved in the different periods of this variation are not the same: Pectinaria koreni, Mya truncata (1980) and Lanice conchilega (1981) for the first period; and Spiophanes bombyx and Macoma balthica for the second period (1986-1987).

This cyclical variation is generated by a limited number of species, representative of the community, which show great abundances and often have chaotic dynamics. Most

#### REFERENCES

Belgrano A., M. Vincx, J.M. Dewarumez, A. Richard, J.A. Craeymeersch, C. Heip (1990). Recruitment of mero-planktonic larvae in the Southern Bight of the North Sea. Oceanis 16, 225-243.

of them are opportunistic, and will react by an important increase in their densities to small, favourable changes in environmental conditions (Dauvin et al., 1993; Dewarumez et al., 1978, 1991, 1992). Species less abundant but with a more balanced population dynamic will certainly be able to reveal more profound changes of the ecosystem. This hypothesis appears to be confirmed by results of the second analysis, performed with strongly contributing species in the analysis of complete data removed. This trend, although it appears less obvious, can be opposed, in a first time, to the cycle due to representative species. A second part of the data, from 1987 to 1992, shows a regular drift of the community from its original state, but not in the direction of any stable situation. This evolution may be related to changes in the sediment composition. In the first years, equally blended middle and fine sand (about 52% of middle sands and 42% of fine sands) characterize the granulometric structure. During the final years, up to 1991, a dominance of fine sands (20% of middle sands and 71% of fine sands) is observed, and the silt content of the sediment appears to show a slight increase (from 4% up to 7.5%).

The particular nature of years 1991 and 1992 is due to the coincidence of two factors: the first recorded presence on the French coast of Ensis directus, a North American species colonizing the North Sea from the German Bight since the end of the 1970s; and a change in the nature of the sediment. Probably after a storm in summer 1991, the granulometric structure of the sediment around the sampling area changed a from muddy fine sands to coarsemiddle sands (from 1.5% of coarse sands before summer 1991 up to 55% in late 1991). This was followed by the disappearance of most of the macrofauna. Only 15 species and a total of 91 ind./m<sup>2</sup> were found in September 1991, as compared with a mean density of 5080 ind./m<sup>2</sup> and a mean species richness of 32 species. The community appears to have rapidly recovered a normal state in 1992, due to an increase of silt content and despite the high content of coarse sand.

Important discontinuities in the variation of this community have been observed over the period of fifteen years. The influence of short spatio-temporal scale events is obvious and is confirmed by observations (or non-observations) of events on neighbouring sites. Numerous hydroclimatic processes are capable of inducing high variability in the community. But the retrievial of information about such local conditions from historical data is not an easy matter, and a local interpretation of the drastic changes of the community remains to be performed. In addition, periodic changes in the community are associated with different species, and the factors determining which of the latter will take over during the cold periods will have to be related to a finer scale.

Benzécri J.P. et al. (1973). L'analyse des données. Tome 1: la taxinomie. - Tome 2: l'analyse des correspondances. Dunod, Paris. 615 p et 619 p.

Dauvin J.C., J.M. Dewarumez, B. Elkaim, D. Bernado, J.M. Fromentin, F. Ibanez (1993). Cinétique d'*Abra alba* (Mollusque-Bivalve) de 1977 à 1991 en Manche-Mer du Nord et relation avec les facteurs climatiques. Analyse descriptive par la méthode des sommes cumulées. *Oceanologica Acta* 16, 413-422.

**Davoult D., J.M. Dewarumez, J. Prygiel, A. Richard** (1988). Carte des peuplements benthiques de la partie française de la Mer du Nord. IFREMER, Région Nord/Pas de Calais, 1 map + 30 p.

**Dewarumez J.M.** (1979). Etude biologique d'*Abra alba*, Wood (mollusque lamellibranche) du littoral de la Mer du Nord. Thèse de troisième cycle. Université des Sciences et Techniques de Lille, 139 p.

Dewarumez J.M., F. Smigielski, A. Richard (1978). Abra alba (mollusque lamellibranche), sa localisation en zone littorale de la Mer du Nord. *Haliotis* 7, 13-19.

Dewarumez J.M., D. Davoult, S. Frontier (1991). Examples of responses of benthic communities to environmental stress (Strait of Dover, France). *Oceanologica Acta* 11, 191-196.

Dewarumez J.M., D. Davoult, L.E. Sanvicente Anorve, S. Frontier (1992). Is the "muddy heterogeneous sediment assemblage" an ecotone between the peebles community and the *Abra alba* community in the southern bight of the North Sea?. *Neth. J. Sea Res.* **30**, 229-238.

Fromentin J.M., F. Ibanez (1994). Year-to-year changes in meteorological features of the French coast during the last halfcentury. Examples of two biological responses. *Oceanologica Acta* 17, 285-296.

Frontier S., D. Pichod-Viale (1993). Ecosystèmes: structure, fonctionnement, évolution. 2nd edition, Masson, Paris, 447 p.

**Glémarec M.** (1979). Les fluctuations temporelles des peuplements benthiques liées aux fluctuations climatiques. *Oceanologica Acta* **2**, 365-371.

Gray J.S., H. Christie (1983). Predicting long term changes in marine benthic communities. *Mar. ecol. prog. ser.* 13, 87-94.

Hill M.O. (1973). Reciprocal averaging: an eigenvector method of ordination. J. Ecol. 61, 237-249.

Ibanez F., J.C. Dauvin, M. Etienne (1993). Comparaison des évolutions à long terme (1977-1990) de deux peuplements macrobenthiques de la Baie de Morlaix (Manche Occidentale): relations avec les facteurs hydroclimatiques. J. Exp. Mar. Ecol. 164, 181-214.

Legendre L., P. Legendre (1984). Écologie numérique. Tome 1: Le traitement multiple des données écologiques. Tome 2: La structure des données écologiques. Masson, Paris et les Presses de l'Université du Québec, 260 p. et 360 p.

Leprêtre A. (1988). Analyse multivariable des peuplements entomologiques. Établissement d'une méthodologie. Application à une situation d'interface écologique. Thèse de doctorat, Université des Sciences et Techniques de Lille, 255 + 155 p.

Luczak C., J.M. Dewarumez, K. Essink (1993). First record of the American jack knife clam Ensis directus on the French coast of the North Sea. J. mar. biol. Ass. UK. 73, 233-235.

Prygiel J., D. Davoult, J.M. Dewarumez, R. Glaçon, A. Richard (1988). Description et richesse des peuplements benthiques de la partie française de la Mer du Nord. *C.R. Acad. Sci. Paris*, sér. III, 306, 5-10.

Souplet A., R. Glaçon, J.M. Dewarumez, F. Smigielski (1980). Distribution des peuplements benthiques littoraux en Mer du Nord du Cap Blanz-Nez à la frontière de Belgique. C.R. Acad. Sci. Paris, 290, 627-630.

# **Chapitre III**

Interpolation spatiale des données benthiques du GDR Manche

# III.1 - Introduction - Cadre de l'étude.

Le travail présenté dans ce chapitre fait suite à celui réalisé par Laura Sanvicente-Añorve (1995) dans le cadre de sa thèse et auquel nous avions contribué pour la mise en place de la base de données du "GDR Manche". Des interrogations subsistaient quant à la validité d'interpolations réalisées sur des variables telles que les indices de diversité, dont on sait qu'il s'agit de variables qui ne sont ni extensives, ni intensives. Par ailleurs, la base de données GDR contient également des données sédimentologiques, en particulier granulométriques, et une collaboration s'était établie à l'époque avec G. Chapalain (URA 719) dans le but de fournir des estimation réalistes des paramètres granulométriques en vue de leur intégration dans un modèle hydrodynamique de transport sédimentaire.

Nous rappellerons donc succinctement dans un premier temps les caractéristique des données du GDR Manche et des explorations statistiques dont elle ont fait l'objet, pour préciser ensuite notre contribution à la valorisation des ces *séries spatiales*. Les méthodes mises en jeu ont en effet de nombreux points communs avec celles utilisées ou développées dans les deux chapitres précédents.

De 1971 à 1976, le RCP "Benthos de la Manche", devenu par la suite "GRECO Manche" puis "GDR Manche" s'était donné comme objectif de réaliser une description bionomique précise des peuplements benthiques de cette mer. Sous la direction de Louis Cabioch (Station Zoologique de Roscoff), et avec plus particulièrement Christian Retière (MNHN, Dinard) et René Glaçon (USTL, Station Marine de Wimereux), une série considérable de prélèvements (environ 6000 récoltes au total) ont été réalisés sur une zone géographique couvrant la Manche entière, la Mer d'Iroise et l'extrême Sud de la Mer du Nord (jusqu'à la frontière Belge). Ces prélèvements ont été effectués avec une drague de type Rallier-du-Baty. Aux relevés biologiques, ont également été associés des relevés de granulométrie du sédiment, des mesures de quelques paramètres physico-chimiques et des estimations d'autres paramètres réalisées à partir des données historiques du SHOM.

L'échantillonnage a été organisé en tenant compte d'une connaissance préalable des peuplements benthiques et de la couverture sédimentaire. Il n'est ni régulier, ni aléatoire, mais peut être qualifié de "raisonné". Il est destiné à s'adapter aux différentes échelles spatiales de variabilité grâce à un effort plus intense dans les zones connues pour leur hétérogénéité

113

spatiale (zones côtières, détroit) ou de manière à prendre en compte les directions privilégiées de certaines structures (bancs de sable).

Dans le cadre de sa participation à l'opération "écologie numérique" du programme GDR Manche, initiée en 1991, le Laboratoire d'Ecologie Numérique de l'ex URA 1363 (devenue UPRES-A 8013 "ELICO") a été chargé d'entreprendre l'exploration de ces données.

Une première exploration des données a été réalisée dans le cadre de la thèse de Laura Sanvicente-Añorve (1995). L'exploitation s'est d'abord limitée à la Manche orientale (à partir du Cotentin) et au sud de la Mer du Nord. Sur cette zone, 1485 récoltes ont été réalisées, totalisant 623 espèces réparties dans 9 phyla et 199 familles.

Dans un premier temps, étant donné la quantité de données à gérer, il a été nécessaire d'organiser cette information sous la forme d'une base de données. La structure de cette base de données a été conçue de façon à faciliter son exploitation. Elle permet d'abord de ne pas stocker les zéros (qu'il est inutile de transporter sur disquette !). Elle permet surtout de disposer de multiples combinaisons de différents types de requêtes et d'extractions à partir de la base générale, ceci selon plusieurs critères (géographiques, taxonomiques, écologiques, etc.).

La deuxième étape a consisté en la réalisation de typologies spatiales des peuplements benthiques à l'aide de méthodes multivariables. En raison de la stratégie d'échantillonnage employée lors des campagnes du GDR (échantillonnage "raisonné"), certains secteurs géographiques ont fait l'objet d'une densité plus importante de prélèvements. Ce déséquilibre de l'effort d'échantillonnage fait que les typologies réalisées sur les données brutes se trouvent contraintes par l'existence de ces zones plus densément échantillonnées. Il a donc été nécessaire de procéder, avant les analyses multivariables, à un rééquilibrage spatial des données, concrétisé par des interpolations aux nœuds d'un maillage régulier.

Différentes méthodes d'interpolations ont été explorées, ainsi que différentes stratégies d'analyse des données, toutes dans le but d'obtenir des typologies spatiales des peuplements macrobenthiques de la Manche orientale. L'un des principaux enseignements de ces travaux et la confirmation de l'interaction existant entre l'amplitude de la zone étudiée et la nature des structures décelées. Les typologies réalisées sur de grandes échelles ne coïncident pas avec celles obtenues à des échelles d'investigation plus restreintes. Seules les analyses se limitant à ces "petites" échelles permettent de retrouver les cartes de communautés déjà obtenues sur ces secteurs (Cabioch et Gentil, 1975 ; Cabioch et Glaçon, 1975, 1977 ; Cabioch *et al.*, 1977).

114



Figure 3.1 : localisation des points de prélèvement, du maillage régulier utilisé pour l'interpolation (à gauche) et du maillage carré utilisé pour évaluer l'effet du changement d'échelle (à droite).

Dans ce travail, nous nous sommes focalisé plus précisément sur la zone la plus orientale de la Manche et au sud de la Mer du Nord, soit une zone de près de 9000 km<sup>2</sup>, s'étendant de Dieppe à la frontière franco-belge, sur laquelle 707 prélèvements ont été effectués (fig. 3.1). Cette zone correspond à celle étudiée par Chapalain et Smaoui dans le cadre de l'application des modèles de transport hydrosédimentaire. Pour rester en cohérence avec ces travaux, nous avons donc réalisé les interpolations sur le maillage utilisé par Smaoui (1996). Ce maillage comporte 1465 nœuds, régulièrement espacés de 2.266 km.

La planche GDR 1 présente quelques caractéristiques morphologiques, physicochimiques et hydrodynamiques de la zone étudiée. La bathymétrie est caractérisée par une faible profondeur, atteignant son maximum au niveau du détroit, et par la présence de nombreux bancs de sable. Les cartes de température et de salinité présentées ici sont réalisées à partir des données stockées dans la base GDR. Les valeurs ont été calculées à partir des moyennes mensuelles sur 60 ans de mesures du SHOM. La température montre un gradient, selon l'axe de la manche, aussi bien nord-sud que est-ouest. La salinité (exprimée en ‰) montre la séparation entre les eaux d'origine atlantique, rencontrées au large, et les eaux constituant le "fleuve côtier" (Brylinski *et al.*,1991), marquées par les apports d'eau douce continentale entre la Baie de Seine et le détroit du Pas-de-Calais. Les valeurs du cisaillement de fond maximum sur un cycle de marée, fournie par la modélisation numérique (Smaoui, 1996), montre l'intensification de l'hydrodynamisme au niveau de la convergence du détroit, ainsi que les zones propices à la sédimentation à proximité des côtes.





Planche GDR 1 : caractéristiques morphologiques, physico-chimiques et hydrodynamiques de la zone d'étude (Manche orientale et sud de la Mer du Nord).

# III.2 - Développement d'une méthode d'interpolation des "séries spatiales"

Compte tenu de la complexité de l'agencement des structures spatiales des fonds benthiques et de leurs communautés, les essais d'interpolations réalisés par Sanvicente-Añorve (1995), par des méthodes de krigeage, spline ou d'interpolation bilinéaire, s'étaient montrés généralement peu satisfaisants. Plusieurs raisons peuvent être invoquées à cette inéquation des méthodes au traitement de données spatiales du GDR :

- La première se rapporte à la nature même des données : les répartitions spatiales des abondances d'espèces sont souvent très éloignées de répartitions aléatoires, souvent caractérisées par une agrégation à petite échelle et présentant, à grande échelle, des structures complexes marquées par des discontinuités brutales, liées à des discontinuités du milieu.
- La stratégie d'échantillonnage utilisée nous l'avons déjà souligné est déséquilibrée du point de vue spatial et donne un poids plus important à certaines zones qu'il était par ailleurs logique de privilégier (les zones littorales en particulier).
- La zone étudiée est particulièrement complexe, soumise à des gradients côte-large avec une côte changeant d'orientation (tantôt est-ouest, tantôt sud-nord), à un transit SW-NE des masses d'eau par le détroit du Pas-de-Calais, à des apports continentaux, etc. Il en résulte une forte anisotropie des structures benthiques (anisotropie de direction variable d'un secteur géographique à un autre) qui rend difficile l'application du krigeage. La solution pourrait résider dans un krigeage par blocs, mais la procédure est complexe, et poserait ensuite le problème d'une remise en cohérence des blocs après des krigeages séparés.

En définitive, l'interpolation par la méthode des voisins les plus proches (NN, *Nearest Neighbour*) est la solution la plus satisfaisante dans une première approche, dans la mesure où elle respecte le choix qui a été fait de densifier l'échantillonnage dans les secteurs *a priori* plus complexe. En effet, l'interpolation NN, en fixant le nombre de points intervenant dans l'interpolation, opère ses estimations à partir de points très rapprochés dans les secteurs complexes qui ont fait l'objet d'un échantillonnage intense, évitant ainsi les lissages excessifs qui sont générés par l'utilisation d'interpolation à partir de distances constantes sur toute la

zone d'étude. Dans les zones dites homogènes, où moins de prélèvements ont été réalisés, le rayon de recherche de la méthode NN est plus grand, donc le lissage est plus intense, ce qui n'est pas gênant si ces zones sont effectivement "homogènes" ou simplement soumises à des gradients, sans discontinuités marquées.

Le choix du nombre de points les plus proches à prendre en compte a fait l'objet d'une étude comparative dans la thèse de Sanvicente-Añorve, étude dans laquelle elle s'est aussi intéressée à la mise en œuvre de différents systèmes de pondération (Sanvicente-Añorve, 1995 ; Sanvicente-Añorve et Leprêtre, 1995). Au vu de ces résultats et d'autres tests complémentaires, nous avons donc retenu une interpolation par les trois points les plus proches, utilisant une pondération par l'inverse de la distance permettant de modérer l'intensité du lissage et de respecter les variations très locales.

Le point le plus délicat et principale critique à apporter aux interpolations réalisées dans le travail de Sanvicente-Añorve, concerne l'absence de prise en compte - lors de l'interpolation spatiale pour une espèce donnée – des relations quantitatives existant avec les autres espèces. Sanvicente-Añorve a en effet procédé à des interpolations indépendantes pour chacune des espèces. Il nous a semblé nécessaire de développer une méthode qui aboutisse à l'interpolation spatiale des abondances sur chacune des espèces, tout en tenant compte de l'existence sous-jacente d'associations ou de communautés. Le principe général de cette procédure, qui a d'abord été développée sur des données granulométriques dans le cadre du programme Etat-Région DYSCOP (cf. rapport en annexe), est donné ci-après (§ 2.1 et 2.2). Elle s'appuie sur le couplage d'une méthode d'analyse factorielle à une méthode géostatistique. La méthode d'analyse factorielle permet, d'une part, de travailler simultanément sur l'ensemble des espèces et, d'autre part, d'assurer un filtrage des données, voire dans une certaine mesure, une "amélioration". La méthode proposée doit en effet permettre de générer, en un endroit précis, des données plus "réalistes" que celles fournies par un échantillon local unique. Par exemple, la méthode doit être susceptible de remettre des espèces, là où elle n'ont pas été observées dans le prélèvement, si le cortège existant suggère que leur présence est fortement probable. Le biais dans l'estimation des espèces rares ou très peu abondantes pourrait ainsi être partiellement éliminé et la diversité mieux estimée.

# III.2.1 Méthode d'interpolation de distributions granulométriques.

Dans le cadre de l'opération "Modélisation statistique et déterministe des processus hydrosédimentaires dans un secteur côtier - la baie de Wissant. Evolution à court et moyen terme" proposé par Chapalain et Leprêtre dans le programme DYSCOP (Etat - Région Nord – Pas-de-Calais), une méthode d'interpolation spatiale a été développée. Elle consiste à réaliser l'estimation, en chaque point d'un maillage régulier, de la distribution de fréquence des différente classes granulométriques à partir de données observées, qui sont inégalement réparties dans le domaine d'étude. *In fine*, ces distributions granulométriques, générées au niveau des nœuds du maillage, et les paramètres statistiques qui les caractérisent (moments statistiques, percentiles, coefficient de classement, etc...; *e.g.* Chamley, 1987) servent de paramètres d'entrée dans la modélisation hydrosédimentaire.

Le problème posé alors était que les interpolations réalisées directement sur les paramètres statistiques (tels que le D50 ou médiane, par exemple) ont un effet de lissage excessif, et ne respectent pas suffisamment les structures marquées, telles que les lisières entre biotopes, donc les passages brusques d'un type de granulométrie à un autre type. Pour pallier cet inconvénient, une méthode a été développée, fondée sur le couplage d'une analyse factorielle et d'une méthode géostatistique (interpolation), ce qui a permis d'aboutir à une interpolation portant sur les distributions elles-mêmes, et non sur les paramètres synthétiques. La méthode doit aussi répondre à une condition fondamentale : la somme des différentes classes granulométriques est forcément égale à 100% (ou à 1). Les variables traitées ne sont dons pas indépendantes, mais contraintes. Le principe de cette méthode a été détaillé dans le rapport DYSCOP joint en annexe I (Leprêtre *et al*, 1997). La méthode a été appliquée dans un premier temps aux données existant sur le secteur de Wissant, en vue de leur injection dans la modélisation. Dans un deuxième temps, nous avons appliqué cette méthode à la zone beaucoup plus vaste et complexe que représente la zone d'étude du GDR précédemment délimitée.

Schématiquement, la méthode réalise les opérations suivantes, qui doivent aboutir à la création de "pseudo-distributions" (fig. 3.2). Le tableau de données granulométriques est d'abord soumis à une Analyse Factorielle Sphérique (ou AFS, variante de l'ACP utilisant la distance de Hellinger ; Domenges et Volle, 1979), ce qui permet d'obtenir les coordonnées factorielles des prélèvements dans un système d'axes factoriels orthogonaux, donc

indépendants linéairement. On a donc remplacé les variables "classes granulométriques" de départ, par des multivariables.

La coordonnée d'un prélèvement sur un axe est équivalente à sa valeur pour cette multivariable. Chaque multivariable peut être cartographiée et interpolée indépendamment des autres. Contrairement au cas de la baie de Wissant, où le krigeage avait été utilisé, l'interpolation des données issus du GDR se base sur la moyenne pondérée par l'inverse de la distance des trois points les plus proches pour les raisons évoquées précédemment. On peut alors obtenir en chaque nœud d'un maillage régulier, ou "pseudo-prélèvement", une estimation de la multivariable. On dispose donc, pour chaque "pseudo-prélèvement", d'un ensemble de valeurs correspondant aux différents axes factoriels.



Figure 3.2 : Etapes de la constitution des distributions granulométriques de pseudo-prélèvements (tableau final) à partir d'un tableau original de distributions granulométriques de prélèvements observés, à l'aide d'une procédure couplant analyse multivariable (AFS) et interpolation.

La formule de reconstitution de l'analyse factorielle (*e.g.* Lebart *et al.*, 1982) permet de reconstituer un tableau de départ "objets × variables" à partir des coordonnées de ces objets ou de ces variables dans l'analyse factorielle du tableau. Appliquée ici aux coordonnées des pseudo-prélèvements, elle va donc pouvoir "reconstituer" le tableau de pseudo-prélèvements.

Cette reconstitution peut n'être réalisée que sur un nombre limité d'axes factoriels, ce qui permet au passage de filtrer le "bruit" que traduisent les axes factoriels portant le moins d'inertie (les derniers axes, en termes d'importance des valeurs propres). On obtient en définitive (moyennant quelques corrections liées au filtrage), une distribution granulométrique pour chaque pseudo-prélèvement.

Les paramètres granulométriques synthétiques, tels que médiane ou autre percentile, coefficient de "sorting" ou autre paramètre de dispersion, ne sont donc pas interpolés mais calculés directement aux nœuds du maillage à partir des distributions estimées en ces nœuds.

La méthode est évidemment plus contraignante que la simple interpolation des paramètres, ou que l'interpolation séparée des classes granulométriques (procédure qu'on devrait s'interdire, vu la non indépendance des proportions). Mais elle permet d'obtenir des estimations beaucoup plus fiables, car elle respecte les patrons de répartition (gradients et coupures). Le rapport joint en annexe présente une comparaison des deux procédures.

Les résultats obtenus par la méthode décrite ici sont très satisfaisants. Les cartographies des paramètres granulométriques classiques (planche GDR 2) restituent fidèlement les structures de cette partie orientale de la Manche et sont en accord avec les travaux de Vaslet et al. (1978) et Larsonneur et al. (1982).

L'hétérogénéité des sédiments est très marquée au niveau de la zone concernée par l'étude. Elle se manifeste sous plusieurs formes: une mosaïque de taches de graviers ou de vases, des structures linéaires (bancs de sable), des gradients (dans les sables littoraux), des coupures franches entre sédiments de natures différentes,...

Il convenait donc de mettre en oeuvre des techniques ou combinaisons de techniques aptes à conserver les patrons existants. L'utilisation "irréfléchie" du krigeage, technique à la mode, sans prise en compte de l'anisotropie, ou sans examen de détail du modèle de variogramme le plus adéquat, ou encore de la portée à donner à ce variogramme, conduit à une diffusion des patrons, et à un lissage des structures essentielles telles que les limites franches entre éléments de la mosaïque.



Planche GDR 2 : paramètres calculés à partir de la distribution granulométrique estimée en chaque point du maillage (les résultats sont exprimés en échelle  $\Phi$ ).

# III.2.2 Interpolation des données de faune macrobenthique.

## III.2.2.1 Méthode

Le principe est similaire à celui qui a été décrit précédemment pour les données granulométriques. La différence réside dans l'utilisation de l'AFC (au lieu de l'ACP), plus adaptée aux données d'abondances d'espèces. Les étapes de la procédure sont les suivantes :

 1) on procède d'abord à une Analyse Factorielle des Correspondances (Benzécri, 1973 ; Hill, 1973) sur le tableau de données initial composé de 707 points de prélèvement et 367 taxons.

2) on réalise ensuite une interpolation sur un maillage régulier des cordonnées des 707 points projetés sur les principaux axes factoriels. La valeur interpolée à chaque nœud du maillage est obtenue en réalisant une moyenne des valeurs des 3 points les plus proches, avec une pondération par l'inverse de la distance

3) parallèlement, on fait une estimation des nouvelles fréquences marginales (lignes et colonnes) qui seront appliquées au nouveau tableau correspondant au maillage régulier. Pour cela, on réalise une interpolation du nombre d'individus collectés en chaque point. L'interpolation s'appuie sur les 3 points les plus proches et utilise une moyenne <u>géométrique</u> pondérée par l'inverse de la distance.

4) on procède ensuite à la reconstitution à partir des coordonnées interpolées, en ne conservant qu'une partie des axes (filtrage).

La formule de reconstitution appliquée est la suivante :

$$\phi_{ij}(p) = f_i \cdot f_j \cdot \left(1 + \sum_{\alpha=1}^p \frac{1}{\sqrt{\lambda_\alpha}} \cdot F_\alpha(i) \cdot G_\alpha(j)\right)$$

 $\phi_{ij}(p)$  est le tableau reconstitué à partir de l'image fournie par la projection sur les p premiers axes factoriels.

 $f_i$  et  $f_j$  sont les fréquences marginales, respectivement des lignes et des colonnes.

 $\lambda_{\alpha}$  est la valeur propre associée à l'axe  $\alpha$ 

 $F_{\alpha}(i)$  et  $G_{\alpha}(j)$  sont les coordonnées de la projection sur l'axe factoriel  $\alpha$  respectivement des lignes et des colonnes du tableau.

5) les quelques valeurs faiblement négatives qui apparaissent après la reconstitution sont éliminées.

6) enfin, pour préserver l'effet de variations biogéographiques (à grande échelle) qui pourraient ne pas être respectées par la méthode employée, on procède à l'élimination, dans les pseudo-relevés reconstitués, des espèces absentes des prélèvements "originaux" situés dans un voisinage de 20 km autour du point reconstitué.

La procédure s'est avérée très satisfaisante, mais on peut d'ores et déjà évoquer ses limites. La reconstitution est performante pour redonner des distributions de fréquences (des fréquences relatives), ce qui ne pose pas de problème pour le calcul de l'indice de Shannon. Le problème est de passer à des abondances. Il faut disposer d'un effectif total, qui sera ensuite distribué selon les fréquences marginales, c'est-à-dire qu'il faut estimer le nombre d'individus qu'on va affecter en chaque point du maillage pour le répartir entre les différentes espèces. Ce nombre total d'individus est obtenus par interpolation à partir des effectifs totaux réellement observés dans les prélèvements voisins.

Avec la reconstitution, on introduit plus d'espèces qu'on en avait à l'origine, ce qui est l'un des objectifs recherché puisque cela permet de donner une meilleure estimation pour les espèces mal échantillonnées. La difficulté réside alors dans le réalisme de la valeur de l'effectif estimé, qui ne peut évidemment pas tenir compte de ces espèces. On peut également reprocher à la méthode qu'elle produit parfois un "lissage" excessif des distributions de fréquences. Pour les distributions qui étaient très "perturbée" par la dominance d'une espèce, on rétablit une distribution un peu plus régulière. Le nombre d'individus remis après reconstitution, qui est estimé d'après l'effectif récolté et donc influencé par l'espèce dominante, va se répartir sur cette nouvelle distribution et provoquer une surestimation de l'abondance des autres espèces (et par conséquent une surestimation de la richesse spécifique).

124

# III.2.2.2 Cartographies des espèces et typologies

L'application la plus évidente de la réalisation d'interpolation des abondances d'espèces sur un maillage régulier est de produire des <u>cartes de répartition</u> de ces espèces. La planche GDR 3 présente, à titre d'exemples, de telles cartes de répartition obtenues pour quatre espèces représentatives des principaux milieux et peuplements identifiés sur la zone étudiée. Nous reprendrons, pour qualifier ces différentes structures, la dénomination utilisée par Sanvicente-Añorve (1995).

La répartition du céphalocordé *Amphioxus lanceolatus* caractérise le groupe "graviers sableux", situé dans la partie centrale au large des côtes picardes et normandes. Celle d'*Ophiothrix fragilis* (ophiuroide) correspond à la zone de cailloutis située dans le détroit du Pas-de-Calais, soumise aux courants les plus forts. La présence de *Gastrosaccus spinifer* (mysidacé) caractérise les zones de sables propres, principalement représentées par les bancs de sable aussi bien en Manche qu'en Mer du Nord. Enfin, la répartition d'*Abra alba* (bivalve) est caractéristique des sédiments fins côtiers. Cette répartition est plus étendue près des côtes anglaises, alors qu'elle ne forme qu'une bande le long du littoral français.

Une des critiques formulées à propos des typologies des communautés réalisées par Sanvicente-Añorve était l'utilisation exclusive de l'Analyse Factorielle des Correspondances (AFC) - et par conséquent d'une métrique  $\chi^2$ . Divers benthologues (*e.g.* Warwick et Clarke, 1991) utilisent plutôt des techniques de positionnement multidimensionnel (Multi Dimensional Scaling, MDS ; Torgerson, 58 ; Kruskal, 1964 ; Escoufier, 1975) appliquée à une matrice de dissimilarités de Bray-Curtis (Bray et Curtis, 1957). Nous avons donc réalisé, en vue de les comparer, les typologies des communautés benthiques obtenues avec ces deux méthodes, appliquée aux pseudo-prélèvements estimés aux nœuds du maillage.

Les cartographies présentées par Sanvicente-Añorve (1995) conservaient une part de "subjectivité" : en effet, la délimitation des différentes communautés, même si elle se basait sur les interprétations des classifications et plans factoriels obtenus, restait une opération essentiellement "manuelle". Au contraire, dans les cartes présentées ici, les différents groupes sont uniquement identifiés d'après la troncature d'une classification hiérarchique. La typologie est donc issue des résultats "bruts" de l'analyse numérique, sans que les contours aient été redéfinis par une "interprétation" de ces résultats.

125





Planche GDR 3 : répartition des abondances de quatre espèces caractéristiques de différents milieux de la Manche orientale et du sud de la Mer du Nord.



Figure 3.3 : répartition des principales structures benthiques identifiées par une classification hiérarchique sur une matrice de dissimilarités de Bray-Curtis. OP = cailloutis à *Ophiothrix fragilis*; GS = graviers sableux; A = sédiments fins envasés; S = sables ; H = hétérogène.

On peut d'abord remarquer que, pour les deux typologies réalisées, il existe une très forte cohérence spatiale des groupes identifiés, alors qu'aucune contrainte de contiguïté n'a été utilisée.

La typologies obtenue à partir d'une classification hiérarchique appliquée à une matrice de dissimilarités de Bray-Curtis montre l'existence de plusieurs grands groupes, spatialement bien délimités (fig.3.3). On distingue, dans la partie centrale le groupe des "graviers sableux" (GS), caractérisé par la présence d'*Amphioxus lanceolatus* (dont les effectifs représentent 17.7% des individus de ce groupe), associé à *Echinocyamus pusillus* (8.1%), *Psammechinus miliaris* (5.7%) et *Ophiura albida* (5.2%). La zone des "cailloutis" (OP) dans le détroit du Pas-de-Calais est marquée par la forte dominance d'*Ophiothrix fragilis* (50.7% des individus) et secondairement de *Pisidia longicornis* (18.4%). Deux groupes occupent les zones côtières : i) le groupe des sédiments fins envasés (A), caractérisée par la dominance d'*Abra alba* (55.8%) et la présence de *Nucula turgida* (7%), *Lanice conchilega* (2.6%) et *Nephtys hombergii* (2.6%), ii) et le groupe des "sables" (S), couvrant les bancs et une grande partie du littoral, très diversifié, composé de *Nephtys cirrossa* (13%), *Echinocardium cordatum* (8%) et *Gastrossacus spinifer, Bathyporeia elegans, Bathyporeia guilliamsoniana, Ophelia borealis, Spiophanes bombyx, Ensis arcuatus* et *Nephtys hombergii* (toutes entre 4 et 5%). Dans la zone des inter-bancs et entre les groupes sables, graviers

sableux et cailloutis à *Ophiothrix fragilis*, on trouve un groupe que nous qualifierons d'"hétérogène" (H), qui présente des caractéristiques de ces trois groupes, avec des espèces ayant des affinités pour des substrats différents (à la fois sables et cailloutis) : *Echinocyamus pusillus* (18.2%), *Ophiura albida* (14.9%), *Ophiothrix fragilis* (6.3%), *Psammechinus miliaris* (5.7%) et *Pisidia longicornis* (2.8%) d'une part, et *Ophelia borealis* (3.5%), *Nephtys cirrosa* (3%) et *Gastrossacus spinifer* (2.2%) d'autre part.

Si on se focalise sur la zone des bancs de sables entre la Baie de Somme et Boulogne sur Mer, secteur dans lequel la distribution des peuplements benthiques a été décrite par Cabioch et Glaçon (1975), on constate une très bonne correspondance entre la carte présentée par ces auteurs et la typologie numérique obtenue (fig. 3.4). Le peuplement des sédiments grossiers à *Amphioxus lanceolatus – Spatangus purpureus* correspond au groupe "graviers sableux" (GS), le peuplement des sables propres à *Ophelia borealis* au groupe "sables" (S), le peuplement des cailloutis et graviers à épibiose sessile et son faciès à *Ophiothrix fragilis*, respectivement aux groupes "hétérogène" (H) et "cailloutis à *Ophiothrix fragilis*" (OP). On ne retrouve pas dans la typologie le peuplement des sables fins légèrement envasés, d'abord parce que peu de prélèvements du GDR ont été faits si près de la côte, mais aussi à cause de la discrétisation du maillage.



Figure 3.4 : distribution des peuplements benthiques de la Baie de Somme au Pas-de Calais d'après Cabioch et Glaçon (1975), à gauche, et d'après la typologie numérique (métrique de Bray-Curtis), à droite.



Figure 3.5 : répartition des principales structures benthiques identifiées par une classification hiérarchique sur une distance du  $\chi^2$ . OP+H = cailloutis à *Ophiothrix fragilis* et zone hétérogène; GS = graviers sableux; A = sédiments fins envasés.

La typologie obtenue à partir d'une classification sur une distance du  $\chi^2$  (fig. 3.5) montre un grand nombre de points communs avec celle obtenue à partir d'une métrique Bray-Curtis : on retrouve les groupes de graviers sableux (GS), dominés par Amphioxus lanceolatus (24.4%), et de sédiments fins envasés (A), fortement dominés par Abra alba (59.4%). Les groupes "cailloutis" à Ophiothrix fragilis et "hétérogène" ne se distingue pas dans cette analyse. Par contre, cette méthode dégage très rapidement (c'est à dire à un niveau élevé dans le découpage de l'arbre) la présence de différents faciès dans le groupe des sables et les zones côtières. On distingue, au sud de la Baie de Somme, deux faciès caractérisés par la présence d'Echinocardium cordatum. Dans l'un (EC1), Echinocardium cordatum est fortement dominant (56.4%), et dans l'autre (EC2), il est associé à Nephtys cirrosa et Ensis arcuatus (respectivement 15.5%, 14.2% et 12.9%). Dans les zones de sédiments fins proches des côtes de la Mer du Nord et entre la Somme et la Canche, on observe également deux faciès dans lesquels on retrouve les mêmes espèces principales : le premier (SF1) occupe une plus grande surface et est caractérisé par Abra alba (12.6%), Nephtys cirrosa (12.4%) et Nephtys hombergii (9.2%). Le second (SF2) est plus localisé, et montre un ordre différent dans la dominance de ces espèces (Nephtys hombergii (18.9%), Abra alba (7.3%) et Nephtys cirrosa (6.6%)). Les bancs de sable, de la Manche à la Mer du Nord, ainsi que les zones côtières les plus proches du détroit, présentent trois aspects différents répartis de manière complexe : l'un

dominé par Spiophanes bombyx (18%) avec Gastrosaccus spinifer (8.3%) et Phyllodoce mucosa (6.7%), un autre caractérisé par Ophélia borealis (16.5%), Gastrossaccus spinifer (16.4%) et Nephtys cirrosa (15.2%), et un dernier caractérisé par Gastrossacus spinifer (12.8%), Nephtys cirrosa (10.2%), Bathyporeia guilliamsoniana (7.4%) et Bathyporeia elegans (5.9%).

Les deux métriques employées ( $\chi^2$  et Bray-Curtis) montrent assez peu de différences dans la détermination des structures à très grande échelle : les grandes unités bionomiques mise en évidence sont très proches dans les deux cas. De même, si on pousse très loin l'analyse des classifications (en procédant à un découpage très fin des arbres), on obtient également une forte similitude des structures fines ainsi isolées. Les différences entre les deux métriques utilisées ici se situent plus dans l'ordre (la hiérarchie) dans lequel elles mettent en évidence les découpages des grandes unités : si la métrique Bray-Curtis distingue facilement les différents aspects des zones de cailloutis et graviers plus ou moins sableux, la métrique  $\chi^2$ a plus de facilité à reconnaître les différents faciès dans les zones de sable. Les deux métriques présentent donc des aptitudes à distinguer les structures fines différentes selon les "milieux". Il n'y a donc pas à affirmer que l'une est meilleure ou plus adaptée que l'autre. De plus, ces conclusions sur les aptitudes des deux méthodes sont uniquement valables pour la zone et les données étudiées ici : on ne peut pas généraliser et conclure, par exemple, que la métrique  $\chi^2$  sera systématiquement plus adéquate dans les zones de sables.

# III.2.2.3 Diversité taxonomique

Une première approche de la structure spatiale de la diversité avait été réalisée, sur ce secteur de la Manche orientale et du sud de la Mer du Nord, par Sanvicente-Añorve *et al.* (1996). Nous avons repris, à partir des nouvelles interpolations réalisées, l'analyse des *patrons* de diversité à l'échelle de cette zone.

La carte de la richesse spécifique obtenue sur le secteur étudiée n'a qu'une valeur "exploratoire" (fig. 3.6 (b)). En effet, les prélèvements effectués, tout comme les interpolations qui en en découlent, sous-estiment fortement la richesse "vraie" des peuplements benthiques. On constate d'abord que les structures mises en évidence ne recouvrent pas les *unités* bionomique identifiées par les typologies réalisées ci-dessus. On observe une plus grande richesse spécifique :

- dans une zone assez étendue, qui occupe la partie est du détroit du Pas-de-Calais;
- dans les zones côtières de sédiments fins et envasés.

Comme nous l'avons fait remarquer dans le cas de la série benthique de Gravelines, on observe sur les données du GDR une forte corrélation entre la richesse observée et l'effectif total de l'échantillons (fig. 3.7 ; coefficient de corrélation de Spearman = 0.88,  $p \approx 0$ ). La carte représentant l'interpolation des effectifs collectés (fig. 3.6 (a)) montre qu'en effet les secteurs de fortes richesses sont les mêmes que ceux où beaucoup d'individus ont été récoltés. Comme nous l'avions évoqué pour la série Gravelines (§ II.3), et comme nous le démontrerons au chapitre IV.3, il est très probable que les différences observées de richesse spécifique sur la zone étudiée ne soient en fait qu'un "artefact" lié à des différences de densité des organismes benthiques.



Figure 3.6 : interpolation du nombre d'individus collectés (a) et richesse spécifique sur l'interpolation des espèces (b)



Figure 3.7 : richesse spécifique en fonction de l'effectif (nbre d'individus) récoltés sur les 707 points originaux.



Figure 3.8 : indice de Shannon H' sur les données reconstituées.

La carte représentant les valeurs de l'indice de Shannon calculés sur les "pseudoprélèvements" ne montre aucune grande structure nettement perceptible (fig. 3.8). On peut cependant identifier :

- une zone de plus faible diversité dans le détroit (très faible diversité au niveau du Cap Gris-Nez), là où la richesse spécifique est pourtant élevée ;
- la zone des bancs de sable, avec une diversité plus faible dans les inter-bancs ;
- une plus grande diversité dans la zone centrale, située plus au large.

Paradoxalement, les zones de faible diversité coïncident avec celles de forte richesse observée. L'explication en est que ces zones riches en espèces sont également celles où l'on constate une très forte dominance de quelques espèces.

A l'échelle des prélèvements effectués lors des campagnes du GDR, la diversité (au sens de l'indice de Shannon) montre une très forte variabilité spatiale. Cette variabilité ne peut être associée à la répartition des structures sédimentaires et bionomiques observée à grande échelle. A l'intérieur d'un même peuplement, au sens de la bionomie benthique comme au sens des typologies numériques, il existe de fortes variations des valeurs de H'. Les communautés, même si elles présentent une certaine homogénéité à grande échelle dans leur composition qualitative, présentent, en interne, une forte hétérogénéité de leur structure quantitative.

# III.2.2.4 Répartition spatiale et diversité de groupes écologiques

La base de données constituée au début du travail de Sanvicente-Añorve (1995) n'a pas seulement pour objectif de "stocker" les données du GDR Manche. Sa structure permet possible à l'utilisateur d'extraire l'information à partir de requêtes portant sur de nombreux critères différents. Un élément essentiel de cette base de données est le "Dictionnaire d'espèces". Ce dictionnaire ne se contente pas de référencer l'ensemble des 624 espèces trouvées en Manche orientale, mais contient également, pour chacune de ces espèces un certain nombre d'informations d'ordre biologique ou écologique :

- trois niveaux taxonomiques, de l'embranchement à la famille.
- le régime alimentaire : déposivore, suspensivore, carnivore, etc...
- l'habitat : épifaune, endofaune, commensal, etc...

Ces différents éléments peuvent être recoupés avec les résultats issus des interpolations et des analyses multivariables. Une telle application est illustrée par les cartes de la planche GDR 4. Nous avons mis en relation le fichier des pseudo-prélèvements avec la base de données pour réaliser des cartographies "fonctionnelles", à partir des régimes alimentaires.

La carte "a" montre la répartition spatiale des proportions d'individus suspensivores estimées aux nœuds du maillage. On voit que les suspensivores dominent dans les secteurs de graviers sableux de la zone centrale et surtout dans la zone de cailloutis du détroit du Pas-de-Calais. cette dominance est quantitative (en nombre d'individus) dans ces deux secteurs, mais la composition qualitative présente par contre des différences. La carte "b" montre que la majorité des espèces de la zone "graviers sableux" sont effectivement des suspensivores. En revanche, la zone à cailloutis ne présente pas une majorité d'espèces suspensivores, mme si quantitativement, c'est une espèce – à savoir *Ophiothrix fragilis* – qui domine très fortement le peuplement.

La carte "c", figurant les proportions d'individus déposivores, apparaît (sans surprise) comme le "négatif" de celle des suspensivores. Les organismes déposivores dominent dans les zones de sédiments homogènes, bancs de sable et zone cotières à sables fins plus ou moins envasés.

Nous avons appliqué la formule de l'indice de Shannon à ces données de régimes alimentaires. Cet indice devient donc un indice de <u>diversité fonctionnelle</u> et non plus taxonomique. Les valeurs obtenues présentent une homogénéité assez marquée. On distingue cependant des zones de plus faible diversité au niveaux des bancs de sable ou dans certaines zones du détroit, comme la pointe du cap Griz-nez par exemple. Une forte diversité est observée face à Douvres : cette zone est caractérisée par la présence de tous les régimes alimentaires, sans que l'un d'eux ne domine les autres.


Planche GDR 4 : répartition spatiale et diversité de différents groupes écologiques

# III.3 - Influence du degré de résolution sur la perception de la diversité des peuplements benthiques

Afin d'étudier l'effet du degré de résolution sur la perception de la diversité taxonomique, nous nous sommes focalisé sur une zone située en face des estuaires de la Canche et de l'Authie. La distribution des peuplements benthiques de cette zone a été décrite par Cabioch et Glaçon (1975). La figure 3.1 (à droite) matérialise l'emplacement des pseudo-prélèvements qui ont servi de support à cette étude, soit un carré de 20x20 mailles. L'application a été réalisée ici sur l'indice de diversité H' de Shannon.

L'indice est d'abord calculé sur chaque pseudo-prélèvement pris séparément, comme c'est le cas pour la carte présentée précédemment à la figure 3.8. L'évaluation de H' aux différentes échelles est réalisée sur un maillage intérieur à celui utilisé pour les pseudoprélèvements, soit 19x19 points. Cette évaluation est faite <u>sans procéder à des moyennes</u> <u>d'indices</u> qui sont critiquables. Les indices de diversité ne sont en effet, ni des variables extensives, ni des variables intensives (Frontier, 1995 ; Leprêtre, 1997) : les valeurs ne sont ni additives, ni proportionnelles à une surface ou un volume considéré.



Figure 3.9 : schéma de construction des différentes échelles de perception. Points recrutés successivement pour les échelles 1 (en bleu), 2 (en rouge), 3 (en vert), etc...

Pour augmenter le degré de résolution de l'observation, on procède à la sommation des prélèvements situés dans des cercles de diamètres croissants. Cette sommation se fait selon le schéma présenté figure 3.9, en agrandissant la zone de cumul de manière centrifuge à partir du

point central (en noir). Une nouvelle échelle est déterminée à chaque fois qu'une série de nouveaux points est prise en considération. L'échelle 1 correspond donc à la sommation de quatre pseudo-prélèvements (en bleu), l'échelle 2, quand on ajoute les six suivants (en rouge), etc... Dix échelles successives sont ainsi déterminées, pour des rayons de 2 à 25 km.

La sommation des effectifs se justifie, car les fréquences d'espèces sont en effet cumulables sans ambiguïtés, contrairement aux indices. A chaque échelle, l'indice de Shannon doit être recalculé sur les effectifs cumulés à l'intérieur du cercle correspondant.

Les cartographies obtenues aux différentes "échelles" sont représentées sur la planche GDR 5. On constate qu'aux échelles 1 et 2, apparaissent des structures obliques qui caractérisent le lien entre la diversité et les structures linéaires des bancs et des inter-bancs au large de la côte (orientation SW-NE). En revanche, à mesure que le degré de résolution décroît (c'est-à-dire que le cercle s'agrandit), ces structures côte-large s'estompent et c'est un gradient de diversité sud-nord qui apparaît. Les images perçues sont donc très différentes suivant le degré de résolution employé.

A petite échelle, l'organisation spatiale de la diversité est donc gouvernée par les structures bathymétriques et sédimentaires qui caractérisent cette zone. La structure des bancs détermine des zones de sables propres (sommet des bancs) et des zones envasées plus hétérogènes (inter-bancs) pour lesquelles la diversité est différente. Près de la côte, la diversité tend à diminuer quand la profondeur augmente (fig.3.10). Vers le large, et surtout au niveau des bancs, la situation est souvent plus complexe : on observe des différences de H' entre le bas, le sommet et les flancs des bancs ainsi que des différences entre les cotés des bancs tournés vers le large ou qui le sont vers la côte (fig. 3.11).



Planche GDR 5 : cartographies de H' à différents degrés de résolution (de 2 à 20 km de rayon).



Figure 3.10 : Indice de Shannon en fonction de la profondeur (en mètres) sur la bande côtière (largeur  $\approx 12 \text{ km}$ )Corrélation Spearman :  $r_s = -0.458$ p.c.  $\approx 0$ 



Figure 3.11 : évolution H' à l'échelle 1 (trait plein) en fonction de la profondeur (pointillé) le long d'un transect Ouest-Est (du large vers la côte), en face de la baie de Canche.

A grande échelle, il semble que ce soit le facteur hydrodynamique qui joue. A l'approche du détroit, la vitesse des courants augmente et le "stress" induit une baisse de la diversité. En effet, si à petite échelle, aucune relation ne peut être mise en évidence entre les valeurs de H' et celle du cisaillement de fond ( $\tau_b$ ) (fig. 3.12), en revanche on décèle, à grande échelle, un lien évident entre H' et le cisaillement (fig. 3.13).



Figure 3.12 : relation entre l'indice de Shannon H' et le cisaillement de fond ( $\tau_b$ ) à l'échelle 1 (grain fin).



Figure 3.13 : relation entre l'indice de Shannon H' et le cisaillement de fond  $(\tau_b)$  à l'échelle 10 (grain grossier).





La figure 3.14 représente l'évolution de l'indice H' en fonction du rayon de prospection. La figure 3.14 (a) donne le H' moyen obtenu sur l'ensemble des points du carré Canche-Authie en fonction de ce rayon (en log). Cette augmentation est linéaire, avec un changement de pente à une échelle donnée (ici, la rupture de pente se fait quand le rayon de recherche dépasse 12 km): voir les deux segments de droite ajustés (avec p = pente, et k = ordonnée à l'origine). La figure 3.14 (b) correspond au H' moyen obtenu sur les 15 points du coin inférieur gauche du carré (points situés dans une zone "homogène" à cailloutis. L'augmentation du rayon (en log) se traduit par une augmentation linéaire de H'. On a donc une relation du type H' = p.log(R) + k, ou  $e^{H'} = e^k R^p$ , liant le nombre de diversité de Hill,  $N_I = e^{H'}$  (Hill, 1973), à l'échelle d'observation. Cette "power law", dans laquelle  $R^p$  représente un facteur de transfert d'échelle, suggère une relation fractale entre la diversité et l'espace.

La relation diversité-degré de résolution se présente, dans le premier cas observé (correspondant à la prise en compte de tout le carré), sous forme de deux droites successives. Il y a en fait deux relations, l'une qui caractérise l'échelle locale (celle qui est dominée par les processus côte-large), et l'autre qui caractérise l'échelle globale (dominée par les gradients sud-nord.

La même procédure, appliquée aux 15 points les plus proches de la baie de Canche, montrent une augmentation de la diversité en fonction du rayon, puis une chute brutale (fig. 3.14 (c)). Cette chute correspond au "recrutement" progressif dans le peuplement à *Ophiothrix fragilis*, que cette espèce domine très fortement, faisant chuter l'indice de diversité.

### III.4 - Conclusion

Un certain nombre de réserves ont souvent été émises vis à vis de la "qualité" des données du GDR Manche. Les prélèvements effectués à l'aide d'une drague ne sont peut-être pas les mieux adaptés pour l'obtention de véritables données quantitatives (Holme et Mc Intyre, 1984). On peut en effet s'interroger sur les performances relatives de cet outil de prélèvement, utilisé ici sur des substrats très différents (de sédiments fins envasés au bedrock) : une part de la variabilité des structures mises en évidence n'est-elle pas liée à des différences de rendement de la drague ? D'autre part, les prélèvements obtenus ne sont pas ponctuels, et dans des secteurs très hétérogènes et organisés en mosaïque, tels que ceux existant près des

côtes de la Manche et de la Mer du Nord, le "trait de drague" peut traverser différents types de substrats, donc affecter des peuplements ou des faciès benthiques différents et mélanger des espèces ayant des affinités différentes.

Néanmoins, les données s'avèrent tout à fait satisfaisantes pour des approches et des comparaisons à grandes échelles. On constate en effet que les unités bionomiques identifiées par les typologies réalisées sur des critères quantitatifs ne présentent pas de discordance majeure avec les connaissances de "terrain" (*e.g.* Cabioch et Glaçon, 1975, 1977). D'autre part, les analyses portant sur l'indice de diversité de Shannon mettent en évidence des structures tout à fait cohérentes, ce qui montre que l'imprécision à petite échelle n'empêche pas un bon rendu des structures aux échelles supérieures. En revanche, l'estimation de la richesse spécifique est (comme nous le verrons au chapitre IV) fortement dépendante de l'effort d'échantillonnage. La forte corrélation qui existe entre la richesse observée et le nombre d'individus collectés affecte l'estimation du nombre d'espèces. Paradoxalement, les résultats obtenus sur les donnée du GDR Manche sont donc plus satisfaisants sur les critères quantitatifs (structure du peuplement, H', ...) que sur le critère plus qualitatif qu'est la richesse.

La base de données du GDR Manche constitue un jeu de données d'une ampleur exceptionnelle, tant par son extension géographique que par la densité de l'échantillonnage. Si pour le moment, les travaux réalisés ont été essentiellement descriptifs et n'ont apporté que peu d'éléments nouveaux à la connaissance des milieux benthiques, ces données peuvent constituer une base pour des développement méthodologiques et pour l'analyse, à grande échelle, de la répartition spatiale des organismes benthiques en fonction de facteurs environnementaux. Il conviendrait de pousser plus loin les investigations dans la mise en relation des données biologiques avec les résultats issus des mesures ou de la modélisation des processus physiques, ce que nous n'avons qu'effleuré ici. Nous avons pu voir également que les informations contenues dans la base de données du GDR permettent d'envisager une approche à grande échelle des écosystèmes benthiques de la Manche sous un aspect fonctionnel, et non plus seulement par leur composition taxonomique.

# **Chapitre IV**

# Expression numérique de la diversité taxonomique

# **IV.1-Introduction**

La diversité est un concept majeur de l'écologie. Ce concept a connu un regain d'intérêt avec les préoccupations grandissantes concernant les changements globaux et le développement durable. L'émergence du terme biodiversité (Wilson et Peter, 1988) a conduit à un ensemble de réflexion (Hawksworth, 1995; Gaston, 1996) portant sur les divers points de vue qu'elle peut recouvrir : taxonomique, phylogénétique, génétique, moléculaire, etc. Les travaux réalisés ici se cantonnent à la diversité taxonomique, dont la déclinaison la plus usuelle est la diversité spécifique.

Cette diversité avait été qualifiée de "non-concept" par Hurlbert (1971) dans un article très largement cité par la suite. Cette attitude se justifiait dans la mesure où les auteurs avaient généralement tendance à confondre le concept lui-même et les outils d'expression numérique de cette diversité.

Les enjeux actuels en matière d'évaluation et de suivi de la diversité des milieux – qu'ils soient naturels ou perturbés – incitent à une nouvelle attitude de la part des scientifiques désireux de transférer des outils fiables. Dans ce chapitre, nous nous sommes consacré à la mise en évidence des propriétés d'indices de diversité comptant parmi les plus utilisés (Pielou, 1975; Magurran, 1988) : l'indice de richesse spécifique et l'indice de diversité de Shannon. Ce travail s'appuie sur la réalisation de simulations, permettant d'apprécier la robustesse de ces indices en fonction de l'intensité de l'échantillonnage et de la structure quantitative des communautés.

La difficulté majeure viens en effet du fait qu'en échantillonnant une communauté, on collecte forcément moins d'espèces (au mieux, toutes) que n'en recèle la communauté. Il s'ensuit, du point de vue de la richesse spécifique, une estimation nécessairement biaisée, une sous-estimation systématique. En ce qui concerne l'indice de Shannon, dont la valeur dépend à la fois du nombre d'espèces et de l'équilibre quantitatif entre les effectifs de chacune d'elles, le problème de l'estimation est différent : le biais est moins marqué, mais la précision de l'estimation est fortement dépendante des modalités de l'échantillonnage (en termes de nombre de prélèvements, de leur répartition, et des effectifs capturés) et de la structure quantitative de la communauté (d'une part, les proportions représentées par les différentes espèces, et d'autre part, la microrépartition spatiale de chacune des espèces).

# IV.2 - Données utilisés

Pour tester les propriétés et performances des différents estimateurs de la diversité, on a fait appel dans ce qui suit, à des données réelles (données de la série benthiques de Gravelines, données entomologiques entre autres), mais surtout à des simulations d'échantillons tirés dans des communautés correspondant à des modèles théoriques de distributions, tels que les modèles de Motomura ou de Preston. Ces simulations sur ordinateur ont permis de décrire et de quantifier les biais et précisions sur les indices, dans la mesure où ils s'appuient sur des compositions et des distributions quantitatives connues, puisque fixées par l'expérimentateur.

## IV.2.1 Modèles de distribution utilisés

Différents modèles de distributions de fréquences d'espèces ont été en vogue dans les années 70 et 80. On citera les synthèses de May (1975), Daget (1976), Sugihara (1980), Magurran (1988) ou Frontier et Viale (1995).

En tant que modèles, ces distributions avaient pour but de donner une interprétation aux différentes modalités de dominances rencontrées généralement dans les communautés. Ainsi une distribution de type Motomura mime souvent bien des situations de communautés pionnières, tandis que le modèle de Preston, par exemple, correspond plutôt aux distributions observées pour des peuplements plus équilibrés, plus mature.

En dehors des modèles "classique" de Mac-Arthur (1957) ou modèle du bâton brisé, de Motomura (1932), de Preston (1948) ou modèle log-normal, et de Zipf-Mandelbrot (Mandelbrot, 1953 ; Frontier, 1977), des modèles fondés sur différents types de partage séquentiel de niche ont été proposés par Tokeshi (1990, 1993) et utilisés par exemple par Fromentin *et al.* (1997).

Dans la mesure où nous n'utilisons ces modèles que pour générer sur ordinateur des jeux de données en vue de tester les qualités des estimateurs de diversité, nous ne nous appesantirons pas sur leur description et n'en ferons qu'un rappel succinct dans les encadrés cidessous.

#### Le modèle de Preston

Le modèle log-normal de Preston (1948, 1962), ou modèle log-normal canonique, est sans doute le plus célèbre, et de nombreux auteurs le trouvent correspondre à la distribution des fréquences de nombreuses communautés. Il a été explicité au départ à partir de la distribution du nombre d'espèces présentes dans des intervalles successifs d'abondance - ces intervalles étant établis suivant une progression logarithmique. Dans la représentation de type Diagramme Rang Fréquence (DRF), la distribution par ordre décroissant des logarithmes des fréquences prend une allure sigmoïde si les rangs sont espacés régulièrement en abscisse. Une droite est obtenue après transformation probit sur les rangs (e.g. Frontier et Viale, 1995, p.303).

Les interprétations données à cette log-normalité sont variées. Au départ, elle était interprétée par Preston en terme de partage/partition de niche. May (1975) se fonde sur les processus de spéciation. Stenseth (1979) en se basant sur l'hypothèse de la "Reine Rouge" (Van Valen 1973, 1976 et 1977) propose un modèle mathématique qui prédit que la distribution d'abondance des espèces est de type log-normal en l'absence de perturbation et log-linéaire dans des situations perturbées. Sugihara (1980) voit dans cette log-normalité le résultat d'une division de la ressource entre espèces au hasard et de manière séquentielle. Une explication alternative, fondée sur l'hétérogénéité spatiale des ressources est développée par Huston (1994).

Gray (1981a) montre que le modèle log-normal s'ajuste bien à la plupart des données de benthos marin, en particulier à des échantillons larges et à des communautés hétérogènes. Il résulterait de la combinaison entre de nombreux facteurs environnementaux agissant de façon aléatoire. L'action combinée de ces facteurs serait, multiplicative et non additive. Pour lui, ce modèle représente une communauté à l'équilibre, avec une balance entre immigration et émigration, et où les espèces se sont divisées la ressource. La log-normalité pourrait également résulter du caractère "patchy" des populations benthiques (Ugland et Gray, 1982), à savoir leur organisation spatiale en taches. Les écarts par rapport au modèle log-normal peuvent ainsi être utilisés, par exemple, pour mettre en évidence les effets de pollutions (Gray et Mirza 1979 et Gray 1981b). Il a été montré également que les tempêtes et les recrutements peuvent provoquer un écart au modèle log-normal, mais que de manière générale le retour à la lognormalité se fait rapidement.

#### Le modèle de Motomura

Le modèle log-linéaire de Motomura (1932), ou modèle géométrique ("log series" ou "geometric series", *in* Whittaker, 1972), a été utilisé initialement pour des données de benthos lacustre. Les logarithmes des effectifs des différentes espèces portés en ordonnée par rapport aux différentes espèces rangées par ordre décroissant d'abondance en abscisse, suivent une droite. On a donc, pour l'espèce occupant un *x*-ième rang, un effectif  $n_x$ , tel que log  $n_x = b - k.x$ , où *b* et *k* sont des constantes. Les valeurs de *k* sont fonction du degré de dominance relative des espèces. L'interprétation écologique qui en est généralement donnée est celle d'un partage séquentiel de la ressource par les différentes espèces : chaque espèce s'appropriant la même proportion des ressources disponibles restantes, après que les espèces plus dominantes se soient déjà "servies". Ce modèle est souvent considéré comme caractéristique de communautés pionnières, donc d'écosystèmes "jeunes" ou "juvéniles", ou évoluant dans des environnements rudes (ou constamment "rajeunis").

#### Autres modèles

Il existe d'autres modèles de distribution de fréquences comme par exemple le modèle de Mac-Arthur (1957), ou modèle du bâton brisé, ou encore le modèle de Zipf-Mandelbrot (Mandelbrot, 1953 ; Frontier, 1977, 1985, 1987).

Tokeshi (1990, 1993) propose un certain nombre de modèles basés sur différents types de partage séquentiel de la niche. Il précise bien cependant que ces modèles stochastiques n'ont de sens que si on les applique à des espèces "écologiquement très proches" : il les utilise sur une "communauté" composée de 6 espèces de larves de chironomides épiphytes récoltées sur la partie apicale d'une espèce précise de myriophylle sur un site d'une rivière anglaise. Bersier et Sugihara (1997) appliquent ces modèles à 4 espèces d'oiseaux. L'aspect stochastique des liaisons entre espèces est un point fondamental de ces modèles. Il n'est donc pas justifié de les appliquer sur des collections hétérogènes, composées d'espèces taxonomiquement différentes, avec des comportements et des régimes alimentaires divers, et n'ayant pas forcément d'interactions directes, comme celles obtenues par exemple dans les relevés macro-benthiques.

#### IV.2.2 Paramètres choisis pour les modèles

L'objectif était d'obtenir des distributions de fréquences présentant des diagrammes rangs-fréquences assez différents, de manière à simuler des communautés également assez différentes, nous avons donc utilisé les modèles suivants :

- deux modèles Motomura (log-linéaire) :
  - modèle M1 : S = 20 espèces et de pente 0.5 (H' = 2.000 ; J' = 0.463)
  - modèle M2 : S = 30 espèces de pente 0.7 (H' = 2.937 ; J' = 0.599)
- deux modèles Preston (log-normal) :
  - modèle P1 : S = 50 espèces (H' = 3.510 ; J' = 0.622)
  - modèle P2 : S = 200 espèces (H' = 4.279 ; J' = 0.560)

- un modèle Mixte avec S = 50 espèces (H' = 2.783 - J' = 0.493), avec les individus des 5 premières espèces distribués selon un modèle Motomura et les 45 autres selon un modèle Preston.

où H' représente l'indice de Shannon et J' l'indice de régularité de Pielou.

Le modèle "mixte" permet de travailler sur une distribution telle qu'on en rencontre souvent sur le terrain (dans les données de la série Gravelines par exemple) : une distribution d'espèces équilibrée (type Preston) sur laquelle se surimposent quelques espèces surabondantes, et donnant lieu à des diagrammes rangs-fréquences en "S", tels que ceux décrits en tant que DRF de type "intermédiaire" (Frontier, 1977).

Pour chacun de ces modèles, les simulations ont été réalisées avec différents efforts d'échantillonnage, en faisant varier le nombre d'individus récoltés dans chaque échantillon et le nombre de réplicats réalisés. La figure ci-après (fig. 4.1) montre l'allure des distributions de fréquences correspondant aux modèle M1, M2, P1 et Mixte, sous forme de DRF en échelle semi-logarithmique (fig a), sous forme de DRF en métrique bi-logarithmique (fig b) suivant Frontier (1977), ou encore sous forme de diagrammes de K dominance (fig c) tels que ceux utilisés par Lambshead *et al.* (1983).



Figure 4.1 : représentations graphiques des modèles de distribution utilisés.

- (a) représentation des fréquences en échelle logarithmique de base 10. (b) représentation bilogarithmique. (c) courbes de K dominance.
- 1 modèle P1 : Preston à 50 espèces
- 2 modèle Mixte
- 3 modèle M2 : Motomura à 30 espèces
- 4 modèle M1 : Motomura à 20 espèces

## IV.3 - Estimation de la Richesse spécifique

#### IV.3.1 Problématique et méthodes d'estimations

La Richesse Spécifique, qui correspond au nombre d'espèces présentes dans la communauté, constitue a priori la plus simple des expression numérique de la diversité. Elle est généralement estimée à partir du nombre d'espèces observées dans l'échantillon réalisé, puisqu'on ne peut qu'exceptionnellement procéder à un véritable recensement conduisant à l'obtention de la richesse "vraie". Quel que soit l'échantillon, la richesse observée sera forcément inférieure à la richesse vraie, au mieux, elle lui sera égale. La richesse observée est donc une sous-estimation systématique.

Par ailleurs, l'ampleur de la sous-estimation (le biais) dépend fortement de l'effort d'échantillonnage :

- du nombre de réplicats réalisés
- de l'aire couverte par chacun de ces réplicats (ou de la durée de l'échantillonnage)
- du nombre d'individus collectés.

La dépendance du nombre d'espèces observées par rapport à l'effort d'échantillonnage a abouti à l'étude de la relation aire-espèces (Arrhenius, 1921 ; Gleason, 1922). Il s'agissait alors seulement de l'aire prospectée, donc du domaine écologique concerné, dans la mesure où ces auteurs réalisaient à l'intérieur de la dite aire, un recensement d'espèces (en l'occurrence d'espèces végétales).

En ce qui concerne les communautés d'invertébrés, on a presque obligatoirement recours, non à un recensement, mais à un échantillonnage dans le domaine que l'on a délimité. On a donc une relation effort d'échantillonnage - nombre d'espèces qui s'ajoute à la relation classique aire-espèces.

Diverses méthodes ont été proposées pour corriger le biais existant entre richesse observée et richesse vraie. Une de ces méthodes consiste à ramener les échantillons à une taille commune. C'est ce que réalise la méthode dite de raréfaction (Sanders, 1968; Hurlbert, 1971; Simberloff, 1972, 1978, 1979). Le principe est le suivant : si l'échantillon avait consisté en n individus (avec n inférieur au nombre d'individus de l'échantillon collecté), quel est le nombre d'espèces qui aurait dû être trouvé ? Cependant, si cette méthode permet de pouvoir comparer des échantillons entre eux, elle ne permet pas l'extrapolation en vue d'estimer la vraie richesse spécifique d'une communauté.

D'autres méthodes s'appuient sur le principe d'une extrapolation, telle que celle utilisée pour la relation aire-espèces. On peut l'appeler relation "effort-espèces". Elle consiste à représenter l'augmentation du nombre d'espèces observées au fur et à mesure qu'on cumule les échantillons élémentaires (réplicats). Le nombre d'espèces collectées en fonction de l'effort d'échantillonnage tend sur ces courbes à s'approcher de manière asymptotique de la richesse vraie (fig. 4.2). Cette méthode, qui ne se restreint donc pas uniquement à une question d'aire prospectée, est qualifiée de "Courbe d'Accumulation d'Espèces" par Colwell et Coddington (1995).



Figure 4.2 : Courbe d'Accumulation d'Espèces

L'estimation de la richesse vraie suppose une extrapolation de la courbe d'accumulation, s'appuyant sur son caractère asymptotique. Plusieurs méthodes, fondées sur la régression linéaire, sont utilisées :

- régression sur le nuage de points nombre d'espèces vs log de l'aire (Gleason, 1922) ;

- régression sur le nuage de points log du nombre d'espèces vs log de l'aire (Arrhenius, 1923);

- régression sur la réciproque de l'aire et de la richesse (Lauga et Joachim, 1987), considérant que la courbe aire-espèces soit une fonction de Monod,.

Un des principaux inconvénients de ces méthodes d'extrapolation est que la forme de la courbe dépend fortement de l'ordre dans lequel on cumule les réplicats. Les différentes permutations de l'ordre de ces réplicats peuvent amener à des conclusions très différentes. L'exemple en est donné fig.4.3 avec le faisceau de courbes formé par les différentes permutations des dix coups de bennes d'un échantillon de la série Gravelines (Nov. 83 – Richesse observée = 38 espèces). On voit que dans certains cas on atteint déjà un palier pour 5-6 réplicats (0,5-0,6 m<sup>2</sup>) alors que dans d'autres, la courbe est toujours croissante et ne converge pas encore vers une asymptote pour les 10 réplicats (1 m<sup>2</sup>). Dans un cas, l'échantillon parait très suffisant et les 38 espèces observées sont une très bonne estimation de la richesse, alors que dans l'autre cas, la valeur obtenue est fortement sous-estimée.



Figure 4.3 : Courbes aire-espèces obtenues avec les différentes permutations des 10 réplicats de l'échantillon de Nov. 83 de la série benthique de Gravelines.

Une solution à ce problème est proposée par plusieurs auteurs (voir par exemple Pielou, 1975, ou encore Stout et Vandermeer, 1975, et Karakassis, 1995). Elle consiste à réaliser l'extrapolation, non pas sur une réalisation de la courbe aire-espèces, mais sur une courbe moyenne obtenue soit à partir de toutes les permutations, soit au moins à partir d'un nombre suffisant de celles-ci.

Plus récemment, le développement des méthodes de ré-échantillonnage (Efron, 1982 ; Diaconis et Efron, 1983) a ouvert de nouvelles perspectives. Ces nouvelles techniques statistiques, fondés sur le calcul intensif sur ordinateur, permettent de s'affranchir des conditions restrictives de la statistique classique ; telles que normalité des distributions ou homoscédasticité. Parmi les estimateurs non-paramétriques les plus utilisés, on citera surtout ceux issus des méthodes dite du "Jackknife" (Quenouille, 1956; Tukey, 1958) et du "Bootstrap" (Efron, 1979).

Ces techniques demandent néanmoins de disposer de réplicats et ne sont guère applicables quand l'échantillon est constitué d'un seul et unique prélèvement.

Dans ce dernier cas de figure, la solution repose sur l'utilisation des distributions des individus entre les espèces (Fisher *et al.*, 1943 ; Efron et Thisted, 1976 ; Slocomb et Dickson, 1978). Une méthode a été plus particulièrement utilisée dans ce cas. Elle se base sur l'intégration de la distribution log-normale (Preston, 1948, 1962) ou de la distribution log-normale tronquée (Cohen, 1959, 1961). L'inconvénient est ici qu'on suppose que la distribution d'abondance est une distribution log-normale ; la méthode est donc inadéquate pour tout autre type de modèle de distribution.

Au vu des études comparatives des performances des divers estimateurs (Palmer, 1990, 1991 ; Baltanas, 1992 ; Colwell et Coddington, 1995) il ressort que tous les estimateurs sont biaisés, mais que les méthodes fondées sur le Jackknife et le Bootstrap sont les plus satisfaisantes. Palmer (1990) montre en particulier que les estimations obtenues par Jackknife et Bootstrap, à défaut d'être non biaisées, montrent une forte corrélation avec la richesse vraie, et sont donc intéressants, au moins pour effectuer des comparaisons entre différents sites.

Ces estimations montrent cependant des limites. Palmer (1991) montre en particulier que les estimations par Jackknife et Bootstrap ne peuvent pas donner une valeur supérieure à deux fois la richesse observée.

Nous nous sommes donc intéressé aux méthodes Jackknife et Bootstrap, ainsi qu'à une nouvelle méthode proposée plus récemment (Karakassis, 1995) également basée en partie sur des opérations de rééchantillonnage sur ordinateur.

Une de nos interrogations a porté sur la validité d'une extrapolation de la richesse à partir d'un échantillon qui ne contiendrait même pas la moitié des espèces réellement présentes dans le milieu étudié. Peut-on vraiment espérer obtenir une estimation fiable, et ce quelle que soit la méthode employée, quand l'échantillon est aussi restreint ? L'inadéquation n'est elle pas ici celle de l'échantillonnage et des moyens mis en œuvre par rapport à l'objectif plutôt que celle de l'estimateur ?

#### IV.3.2 Les estimateurs non-paramétriques

Le Jackknife et le Bootstrap sont des méthodes non-paramétriques de rééchantillonnage pour obtenir l'estimation d'un paramètre, le biais et la variance. Elles sont désormais couramment utilisées dans différents domaines (Efron, 1982 ; Diaconis et Efron, 1983 ; Efron et Tibshirani, 1993 ; Davison et Hinkley, 1997). Le Jackknife a notamment été utilisé pour l'estimation d'indices de diversité (Zahl, 1977).

Dans ce type d'approche, il est nécessaire de disposer de réplicats (quadrats) de tailles équivalentes. Les réplicats sont alors considérés comme des échantillons indépendants tirés de la même distribution.

Heltshe et Forrester (1983a et b) ont proposé comme estimateur de la richesse spécifique le Jackknife d'ordre 1 :

 $JACK1 = SO + [r_1(n-1)/n]$ 

avec n = nombre de réplicats

 $r_1$  = nombre d'espèces présentes dans seulement 1 réplicat

SO = richesse spécifique observée sur les n réplicats

Smith et Van Belle (1984) ont étendu ces résultats et présenté la formule générale pour un estimateur Jackknife d'ordre k:

$$J_{k} = SO + \left\{ \sum_{j=1}^{k} r_{j} \cdot \sum_{i=j}^{k} (-1)^{i+1} \binom{k}{i} (n-i)^{k} \binom{n-j}{i-j} / \binom{n}{i} \right\} / k!$$

avec n = nombre de réplicats

 $r_j$  = nombre d'espèces présentes dans exactement j réplicats

SO = richesse spécifique observée sur les n réplicats

Pour l'estimateur d'ordre k = 2 on a alors :

$$JACK2 = SO + [r_1(2n-3)/n] - [r_2(n-2)^2/(n(n-1))]$$

avec n = nombre de réplicats

 $r_1$  = nombre d'espèces présentes dans seulement 1 réplicat

 $r_2$  = nombre d'espèces présentes dans seulement et seulement 2 réplicats

SO = richesse spécifique observée sur les n réplicats

Smith et Van Belle (1984) ont également proposé un estimateur Bootstrap :

$$B O O T = S O + \sum_{i=1}^{SO} (1 - p_i)^n$$

avec n = nombre de réplicats

p<sub>i</sub> = proportion de réplicats qui contiennent l'espèces i

SO = richesse spécifique observée sur les n réplicats

Dans les revues de Palmer (1990, 1991) et Baltanas (1992), ces estimateurs nonparamétriques se montrent plus performants (moins biaisés et plus précis) que les autres méthodes (extrapolation de la courbe aire-espèces, intégration de la distribution log-normale). Le Jackknife d'ordre 2 semble meilleur que le Jackknife d'ordre 1 et que le Bootstrap.

Dans leur comparaison, Colwell et Coddington (1995) montrent que l'estimateur "Chao2" (Chao, 1984) a des performances au moins aussi bonnes que le Jackknife 2. Il se définit de la façon suivante :

 $CHAO2 = SO + (r_1^2 / 2.r_2)$ 

avec  $r_1$  = nombre d'espèces présentes dans seulement 1 réplicat

 $r_2$  = nombre d'espèces présentes dans seulement et seulement 2 réplicats

SO = richesse spécifique observée sur l'ensemble des réplicats

Les résultats obtenus par les simulations sur les modèles de Preston et de Motomura (M2) ont effectivement montré que cet estimateur présente des performances proches de celles du Jackknife2, en restant toujours quand même légèrement moins bon (Annexe II). Il pose cependant un problème : il n'est pas défini (il y a division par 0) si il n'y a pas d'espèces présentes seulement dans 2 réplicats. Ce cas de figure s'est présenté dans un grand nombre de simulations ; pour les plus grandes tailles d'échantillons, c'est dans plus de la moitié des distributions simulées qu'il est impossible de l'évaluer.

Plus récemment, Karakassis (1995), s'appuyant sur des données benthiques, propose une méthode qui utilise les deux aspects, ré-échantillonnage et extrapolation de la courbe aireespèces. La méthode peut se résumer de la façon suivante :

1) on réalise d'abord p permutations aléatoires des n réplicats qui donnent autant de courbes aire-espèces différentes (Karakassis propose d'utiliser p = 150 permutations).

2) à partir de ces p permutations, on calcule alors une courbe aire-espèces moyenne.

Soit j = 1 à n, on a une courbe aire-espèces moyenne composée des

$$\overline{S}j = \frac{\sum_{i=1}^{p} Sij}{p}$$

avec  $\overline{S}j$  = nombre moyen d'espèces observées dans j réplicats

 $S_{ii}$  = nombre d'espèces observées dans j réplicats pour la i-ème permutation

n = nombre de réplicats

p = nombre de permutation

3) en se basant sur l'observation que la courbe aire-espèces s'approche de la richesse vraie de manière asymptotique, d'une façon similaire à un modèle de croissance exponentiel négatif de Von Bertalanffy, Karakassis propose d'estimer la valeur de l'asymptote en utilisant le principe du diagramme de Ford-Walford (*e.g.* Frontier et Viale, 1995). Le principe est le suivant : on représente graphiquement  $S_{i+1}$  en fonction de  $S_i$ . La relation entre  $S_{i+1}$  et  $S_i$  est une droite. L'intersection entre cette droite et la diagonale d'équation y = x donne la valeur de S $\infty$ . Cette valeur estimée de l'asymptote de la courbe est donc une estimation de la richesse spécifique vraie de la communauté observée.

Cette méthode est similaire à celle proposée par Stout et Vandermeer (1975), et citée par Baltanas (1992), qui se base également sur une courbe moyenne obtenue après permutations, mais qui estime l'asymptote directement par ajustement d'un modèle de croissance exponentiel négatif par les moindres carrés non-linéaires.

#### IV.3.3 Comparaison des performances des estimateurs

#### IV.3.3.1 Simulations réalisées

Les différents estimateurs (Jackknife1, Jacknife2, Bootstrap et S $\infty$ ) ainsi que la Richesse observée sont calculés pour 1000 échantillons simulés, tirés au hasard dans les modèles choisis (*cf* 2.2).

Pour chaque modèle, on fait varier deux paramètres dans les simulations :

- le nombre de réplicats qui constituent un échantillon : 10, 20 et 40 réplicats.

- l'effectif (nombre d'individus) de chacun de ces réplicats : 50, 100, 500, 1000, 5000 et 10000 individus.

On dispose donc au total de 18 séries de 1000 simulations pour chacun des modèles. Les tailles globales des échantillons simulés varient donc de 10\*50 = 500 individus à  $40*10000 = 400\ 000$  individus.

Ces différentes simulations couvrent une large gamme d'efforts d'échantillonnage, du sous-échantillonnage qui va largement sous estimer la richesse, jusqu'à de très larges échantillons dans lesquels on va retrouver presque toutes les espèces et obtenir des estimations très proches de la richesse vraie.

Afin d'évaluer les performances des différents estimateurs, on calcule pour chaque série de 1000 simulations la moyenne, la médiane et un intervalle de confiance à 95% basé sur les quantiles. L'ensemble des résultats est reporté dans les tableaux en annexe II.

## IV.3.3.2 Résultats des simulations

La figure 4.4 montre l'évolution, en fonction du nombre d'individus collectés, des moyennes obtenues sur 1000 simulations dans chacun des cas testés pour les modèles Preston à 50 espèces, Motomura à 20 et 30 espèces (M1 et M2). On visualise ainsi le biais par rapport à la richesse vraie (en pointillé) dans les différentes situations d'échantillonnage envisagées.

Une première remarque concerne les performances de l'estimateur S $\infty$ . Dans le meilleur des cas, à savoir sur les petits échantillons (10 réplicats), il aboutit en moyenne à une estimation de richesse à peine supérieure à la simple richesse observée sur les échantillons eux-mêmes. Dès qu'on augmente le nombre de réplicats à 20 et 40, la moyenne des estimations obtenues par cette méthode passe même en dessous de la richesse moyenne observée. Cet estimateur est donc encore plus biaisé que la richesse observée pour évaluer la richesse vraie d'un peuplement. Cette observation sur les simulations est confirmée par les résultats obtenus sur des échantillons réels : l'estimation obtenue de la richesse vraie est le plus souvent à peine supérieure (d'une ou deux espèces) à la richesse observée, ce qui est peu réaliste, et dans certains cas, cette méthode a effectivement conduit à une estimation de la richesse inférieure à la richesse observée dans l'échantillon analysé.

Il apparaît donc que l'utilisation de cette méthode est à proscrire. Même si le fait d'utiliser les permutations des réplicats pour construire la courbe d'accumulation d'espèces peut sembler une amélioration par rapport aux autres méthodes d'extrapolations de ces courbes, le modèle exponentiel négatif de type Von Bertalanffy et surtout l'estimation de l'asymptote par le principe du diagramme de Ford-Walford semblent inadéquats. Quelques essais ont été réalisés en utilisant le diagramme de Lockwood. Les estimations sont alors légèrement supérieures mais n'arrivent cependant pas au niveau des estimateurs Jackknife et Bootstrap. On peut noter que Baltanas (1992) en testant la méthode de Stout et Vandermeer (1975), basée sur un principe voisin, concluait déjà que cette méthode était la plus mauvaise de celles qu'il avait testées.

En ce qui concerne les autres méthodes, on constate que quel que soit le modèle de distribution ou l'effort d'échantillonnage, l'ordre des performances des différents estimateurs reste globalement toujours le même : l'estimateur Jackknife 2 donne les plus fortes valeurs, devant respectivement le Jackknife 1 et le Bootstrap. La différence est surtout marquée sur les petits tailles d'échantillons, où le Jackknife 2 est nettement moins biaisé que les autres méthodes. Il a cependant tendance à surestimer légèrement la richesse pour des échantillons de taille intermédiaire. Cette surestimation s'observe aussi de manière moins marquée et pour des échantillons plus importants pour le Jackknife 1. Elle est presque inexistante pour le Bootstrap.

Pour des échantillons relativement grands (de plusieurs milliers ou de plus de 100 000 individus), les estimateurs Jackknife 1 et Bootstrap deviennent presque aussi performants et même quelquefois meilleurs que le Jackknife 2 (Tableau 4.1).

L'illustration en est évidente sur le modèle Preston, où l'on voit que le Jackknife 2 est le moins biaisé sur les petits échantillons, que le Jackknife 1 est meilleur pour les échantillons moyens de 20000 individus (20x1000 et 40x500), et que le Bootstrap vient ensuite sur les échantillons plus importants.

Une séquence tout à fait similaire s'observe pour le modèle Motomura 2, décalée vers des tailles d'échantillons plus importantes. On n'atteint visiblement pas ce stade pour le modèle Motomura 1 avec la gamme d'efforts d'échantillonnage utilisée : le Jackknife 2 reste dans ce cas toujours le moins biaisé des estimateurs.

Tableau 4.1 : estimateur le moins biaisé dans les différents cas envisagés.

Modele Preston									
	50	100	500	1000	5000	10000			
10									
20									
40									

Modele Motomura2								
	50	100	500	1000	5000	10000		
10								
20								
40								

	50	100	500	1000	5000	10000
10				a start and		and the second
20						
40						

Légende : Jackknife1 Jackknife2 Bootstrap

Il faut cependant relativiser ces résultats pour les plus gros échantillons. Les différences de biais observées entre les trois méthodes sont en effet peu importantes. Dès que les valeurs des trois estimateurs se rapprochent fortement de la valeur vraie de la richesse, les écarts entre les moyennes du meilleur et du moins bons des trois sont souvent de l'ordre de quelques 1/10<sup>es</sup> voire quelques 1/100<sup>es</sup> d'espèces.

On voit par exemple que pour les très gros échantillons (40\*5000, 20\*10000 et 40\*10000) sur le modèle Preston, le Jackknife 2 est l'estimateur le moins biaisé. La différence maximale entre les estimateurs n'est cependant que de 0.04 espèce. On ne peut pas considérer dans ce cas qu'un des estimateurs est réellement meilleur que les autres : d'une part parce qu'avec une aussi faible différence, on ne peut exclure qu'une autre série de 1000 échantillons simulés aurait donné un résultat différent, et d'autre part parce qu'avec ces tailles d'échantillons, la richesse observée ne présente quasiment aucune différence avec la richesse vraie.

Le biais - à savoir l'écart entre la moyenne des valeurs observées et la valeur vraie n'est pas le seul critère à prendre en compte lors du choix d'un estimateur. La précision, c'està-dire la variabilité des estimations autour de la moyenne réalisée, est également un facteur important. Un estimateur peut s'avérer juste, sans biais, mais être en revanche peu précis, montrant une trop grande dispersion.

Les résultats montrent que, pour ce critère, l'ordre des estimateurs est inversé : le Bootstrap, bien que globalement plus biaisé en moyenne, est celui qui présente la plus faible dispersion. L'intervalle de confiance à 95% (obtenu à partir des quantiles sur la distribution observée des 1000 valeurs simulées) est à peu près deux fois moins large pour le Bootstrap que pour le Jackknife 2. Le Jackknife 1 présente une situation intermédiaire. Il est à noter que la dispersion des valeurs fournies par les estimateurs est toujours plus importante que celle que présente la richesse observée.





Figure 4.4 : Performance des estimateurs de richesse pour les modèles Preston (a), Motomura 1 (b) et Motomura 2 (c).

Figure 4.5 : Intervalles de confiance à 95% et moyennes des estimateurs Bootstrap, Jackknife 1 et 2 pour le modèle Preston à 50 espèces



10 réplicats

La figure 4.5 présente les moyennes et les intervalles de confiance à 95% des valeurs obtenues par les trois estimateurs pour le modèle Preston à 50 espèces.

La forte variabilité que montre le Jackknife 2 fait que cet estimateur, bien qu'étant celui qui se rapproche le plus rapidement de la valeur juste, ne constitue pas forcément le meilleur choix. Pour des échantillons "moyens", pour lesquels le biais des estimateurs Jackknife 1 et Bootstrap devient moins important, la variabilité plus réduite de ces estimateurs les rend plus aptes à donner de manière quasi systématique une valeur plus proche de la richesse vraie. Le Jackknife 2, bien que centré sur la valeur juste du fait de sa plus grande variabilité, aura de plus fortes chances de produire des valeurs plus éloignées.

Par exemple, on voit dans les tableaux suivants (tableau 4.2 et 4.3), réalisés à partir des résultats pour le modèle Preston à 50 espèces, que pour un échantillon de taille 10x1000 individus, alors que le Bootstrap présente le biais moyen le plus important (1.82 contre 0.47 pour le Jackknife 1 et 0.02 pour le Jackknife 2), il a une plus grande probabilité de donner une estimation contenue dans un intervalle de  $\pm$  10% autour de la richesse vraie : 97.9% des valeurs données par la méthode Bootstrap se situe entre 45 et 55 espèces, contre 96.8% pour le Jackknife 1 et seulement 85.3% pour le Jackknife 2. De la même manière, pour des échantillons de 40x500 individus, 89.9% des estimations Bootstrap sont dans un intervalle de  $\pm$  5% autour de la richesse vraie, contre 83.1% pour le Jackknife 1 et 56.1% pour le Jackknife 2, alors que là encore le Bootstrap présente le biais le plus important.

10	0	0.8	80.0	97.9	100	100
20	1.2	18.4	97.3	100	100	100
40	17.6	58.5	100	100	100	100
			Jackknife1		,	
	50	100	500	1000	5000	10000
10	4.2	24.7	87.0	96.8	100	100
20	23.5	51.9	96.5	99.8	100	100
40	50.1	75.7	99.8	100	100	100
		<u> </u>	Jackknife2			
	50	100	500	1000	5000	10000
10	25.5	42.4	74.1	85.3	97.8	99.9
20	43.3	54.6	83.4	92.2	100	100

92.2

Tableau 4.2 : Pourcentage d'estimateurs dans un intervalle de  $\pm 10\%$  autour de la richesse vraie. Bootstrap

500

50

54.4

40

100

65.3

1000

96.0

5000

100

10000

163

Tableau 4.3 : Pourcentage d'estimateurs dans un intervalle de  $\pm 5\%$  autour de la richesse vraie.

20010444								
The label	50	100	500	1000	5000	10000		
10	0	0	30.5	67.0	99.7	100		
20	0	1.8	67.9	90.7	100	100		
40	1.3	15.4	89.9	98.6	100	100		

Bootstrap
Doomah

Jackknife1								
	50	100	500	1000	5000	10000		
10	1.0	7.4	61.4	74.9	98.0	100		
20	9.2	28.3	72.2	84.6	100	100		
40	26.6	46.8	83.1	96.3	100	100		

Jackknife2								
u.	50	100	500	1000	5000	10000		
10	11.0	21.2	42.9	53.7	81.8	97.6		
20	20.5	30.6	46.5	59.6	97.3	99.9		
40	30.4	37.4	56.1	63.0	100	100		

Il ne semble pas que, a effectif constant, le fait d'augmenter le nombre de réplicats améliore les estimations que ce soit en terme de biais ou de variabilité. Les performances des trois estimateurs ne semblent dépendre que du nombre d'individus collectés. Cela est peut-être lié à la forte homogénéité qui existe entre les réplicats utilisés dans les simulations.

Grâce à ces simulations, on peut voir également l'importance du type de distribution de fréquence. Bien qu'ils présentent une richesse vraie plus faible, il est plus difficile (il faut des échantillons plus grands) d'obtenir une bonne estimation de la richesse avec les modèles Motomura qu'avec le modèle Preston. Une distribution des individus entre espèces présentant une plus grande régularité (présentant aussi une plus grande diversité au sens de l'indice de Shannon) facilite l'obtention d'une bonne estimation de la richesse. Pour une taille d'échantillon équivalente (en nombre d'individus), les espèces peu fréquentes ont une plus grande probabilité de se trouver dans les échantillons pris dans un modèle de Preston, que dans des modèles présentant une forte dominance de quelques espèces. Le biais de la richesse observée sera moins important pour des distributions plus régulières, et cette amélioration se répercute sur les autres estimateurs.

Cependant, au sein de ces estimateurs, il n'en est pas un qui serait plus particulièrement adapté à un type de distribution de fréquence donné. Aussi bien en termes de biais que de précision, les performances relatives des estimateurs restent similaires, quel que soit le modèle de distribution utilisé.

Les simulations sur le modèle Preston à 200 espèces confirment les observations faites sur les autres modèles : l'estimateur Jackknife 2 reste le moins biaisé mais montre une dispersion plus importante (fig. 4.6 et 4.7). On constate surtout avec ce modèle qu'il est très difficile d'obtenir une estimation juste de la richesse pour des communautés comportant un nombre important d'espèces. Bien que le modèle Preston soit, semble-t'il, un des plus propice à une bonne estimation de la richesse, on voit ici que seul les échantillons les plus importants (40x5000 et surtout 40x10000 individus) arrivent à une estimation correcte par les méthodes Jackknife.



Figure 4.6 : comparaison des différent estimateurs sur le modèle Preston à 200 espèces.

V.3.4 Illustration sor de

10 réplicats



Figure 4.7 : Intervalles de confiance à 95% et moyennes des estimateurs Bootstrap, Jackknife 1 et 2 pour le modèle Preston à 200 espèces

### IV.3.4 Illustration sur des données réelles

On a vu que le nombre d'individus collectés agissait sur l'estimation de la richesse. Le plus souvent dans les études de terrain, ce n'est pas ce nombre d'individus qui est fixé, mais la surface du prélèvement, son volume, ou encore la durée. On dispose donc de différents prélèvements dont les effectifs différent. Les problèmes peuvent être illustrés à partir des relevés de la série benthique de Gravelines.

On observe globalement sur toute la période de 1978 à 1992 une augmentation de la richesse spécifique (fig : 4.8). Un test de Mann-Kendall (*cf.* chap. I) confirme que cette tendance est significative.



Figure 4.8 : évolution chronologique de la richesse observée

On constate cependant que dans le même temps, il y a augmentation de la densité c'est-à-dire du nombre d'individus récoltés par m<sup>2</sup> (fig. 4.9). Là encore, un test révèle que cette augmentation est significative au risque 5%.



Figure 4.9 : évolution chronologique de la densité en nombre d'individus/m<sup>2</sup>

On obtient donc une corrélation significative (coefficient de corrélation de Spearman :  $r_s = 0.514$ ;  $p(r_s=0) \approx 0$ ) entre le nombre d'individus récoltés dans chaque échantillon et le nombre d'espèces ainsi observées (fig. 4.10).



Figure 4.10 : Richesse observée en fonction de la densité d'individus collectés.

Sachant que l'estimation de la richesse est fortement dépendante de l'effort d'échantillonnage, on peut donc se demander si l'augmentation de la richesse observée dans le cadre de cette série correspond réellement à une augmentation de la richesse du peuplement. En d'autre termes, peut-on conclure qu'il y a vraiment plus d'espèces sur le site, ou bien la hausse observée n'est elle qu'un "*artefact*" lié au fait qu'il y a de plus en plus d'individus dans les échantillons collectés et donc plus de chances d'observer de nouvelles espèces ?

La question posée est donc : retrouve-t-on toujours la tendance quand on ramène tout les relevés à un effort d'échantillonnage homogène en nombre d'individus ? Pour y répondre nous avons appliqué deux méthodes.

La première méthode utilisée est la méthode de Raréfaction (Sanders, 1968 ; Hurlbert, 1971 ; Simberloff, 1972, 1978, 1979). Elle permet d'obtenir la richesse qu'on peut espérer observer sur un échantillon d'un effectif n, inférieur à l'effectif N de l'échantillon collecté. L'estimation de cette richesse espérée est obtenue par la formule suivante :

$$E(S_n) = \sum_{i=1}^{S} \left[ 1 - \frac{\binom{N-N_i}{n}}{\binom{N}{n}} \right]$$

avec  $E(S_n) = richesse espérée pour un échantillon aléatoire de n individus$ S = richesse observée $<math>N_i = nombre d'individus de l'espèce i$  $N = nombre total d'individus collectés = \Sigma N_i$ n = effectif de l'échantillon choisi (nombre d'individus)

La deuxième méthode s'appuie sur un rééchantillonnage : on tire au hasard, dans les distributions de fréquences observées, des échantillons de tailles égales en nombre d'individus. Pour chaque relevé (date), le tirage est répété 200 fois et la moyenne de ces 200 répétitions est prise comme estimation de la richesse espérée.

Ces deux méthodes ont été appliquées à l'estimation de la richesse espérée pour des effectifs d'échantillons simulés de 100, 200 et 500 individus. Trop d'observations dans la série n'ont pas assez d'individus pour tester l'existence d'une tendance avec des effectifs plus importants. On ne peut obtenir d'estimation par la Raréfaction (l'effectif n de l'échantillon simulé est plus grand que l'effectif collecté N) pour 1 date d'observation avec n=100 individus, 2 avec n=200, 8 avec n=500, 23 avec 1000 et 59 avec n=5000. Afin de ne pas trop diminuer la
longueur de la série pour réaliser les tests, seuls les trois premières tailles d'échantillons évoquées ci-dessus ont été utilisées.

Les résultats obtenues avec les deux méthodes sont tout à fait similaires pour les trois efforts d'échantillonnage proposés. La méthode de raréfaction donne toujours une estimation de la richesse légèrement supérieure au rééchantillonnage, mais la valeur du coefficient de corrélation entre les deux reste autour de 0,99. Afin d'évaluer l'existence ou non d'une tendance, les deux méthodes sont donc équivalentes. De même, les différentes tailles d'échantillons, 100, 200 ou 500 ne semblent pas montrer de grandes différences en ce qui concerne l'évolution chronologique.

La tendance nette présentée par la série originale disparaît quand les échantillons sont ramenés à une taille équivalente en nombre d'individus (fig 4.11). Quel que soit l'effort ou la méthode appliqués, les tests de tendances (tableau 4.4) confirment qu'il n'existe en fait aucune tendance significative.

En conclusion, l'augmentation de la richesse spécifique observée sur la série chronologique est donc un artefact lié à l'augmentation des effectifs en place. Cette augmentation de la densité des peuplements, donc des échantillons, donne l'"illusion" d'une augmentation de la richesse, due au fait qu'un effectif d'échantillon croissante augmente la "probabilité" de trouver des espèces rare.

Dans le cadre d'un programme de suivi de la diversité, ou pour des comparaisons entre différents sites, il s'avère donc nécessaire de travailler à effectif constant, ou tout au moins, de ramener le calcul de la richesse à un effectif donné de manière à s'affranchir des changements de densité qui affectent la qualité de l'estimation. Cette approche est d'ailleurs appliquée, par exemple, pour les analyses quantitatives du phytoplancton en eaux douces (Uthermöl, 1958), qui consiste à ne compter dans le prélèvement qu'un nombre fixé d'individus.



500 individus

Figure 4.11 : évolution de la richesse espérée pour des effectifs d'échantillond de 100, 200 et 500 individus par rééchantillonnage (à gauche) et par la méthode de raréfaction (à droite).

### IV.3.5 Conclusion sur l'estimation de la richesse

Les résultats obtenus dans cette étude montrent qu'aucune des méthodes que nous avons envisagées n'offre de réponse parfaite au problème de l'estimation de la richesse spécifique. Elles permettent cependant d'obtenir, dans certaines conditions, une estimation de la richesse qui, si elle reste fortement biaisée dans de nombreux cas, s'approche de la richesse vraie.

L'estimateur Jackknife 2, qui présente généralement le biais le moins important, surtout pour les petits échantillons, n'est pas forcément le meilleur choix. En effet, sa forte variabilité le rend peu fiable. En effet, quand les échantillons deviennent assez importants, le Bootstrap est l'estimateur qui va offrir la plus grande précision. Les performances du Jackknife 1, aussi bien pour le biais que pour la précision, se situent entre les deux : son espérance présente un biais qui n'est que légèrement inférieur à celle du Jackknife 2 et la dispersion des valeurs produites n'est pas beaucoup plus forte que celle du Bootstrap. Il apparaît donc comme le meilleur compromis dans une grande majorité des cas.

Comme critique des méthodes Jackknife, différents auteurs ont évoqué des problèmes de surestimation liés à ces estimateurs (Heltshe et Forrester, 1983b et Karakassis, 1995 par exemple). On a vu ici que ces problèmes ne se produisent que rarement et que l'ampleur de la surestimation est alors en moyenne très faible : elle ne représente, au pire, que 2% de la richesse vraie. La variabilité de ces estimateurs, en particulier celle du Jackknife 2 qui est importante, fait qu'effectivement ils peuvent produire des valeurs qui surestiment assez nettement la richesse vraie. Néanmoins, il n'y a pas beaucoup plus de chances qu'ils ne surestiment cette valeur qu'ils ne la sous-estiment.

Dans le cadre d'un suivi temporel ou d'une comparaison inter-sites, il est nécessaire d'être très prudent sur l'utilisation des différents estimateurs. Palmer (1991), en observant que tous les estimateurs sont fortement corrélés avec la richesse vraie, conclut que ces estimateurs sont donc valables pour établir des comparaisons dans l'espace ou dans le temps. Cependant, les simulations montrent que le biais des estimateurs de richesse est influencé par un certain nombre de facteur. Baltanas (1992) montre par exemple que l'agrégation a un effet sur la qualité de l'estimation. Dans ce travail, on a vu aussi que le type de distribution de fréquences, en particulier sa régularité, change fortement le biais que présente les estimateurs. En toute rigueur, il conviendrait donc de ne comparer entre elles que des situations où les distributions de fréquences sont voisines. Si on soupçonne un changement de distribution, il est impossible de savoir si une différence de richesse n'est due qu'à ce changement de distribution ou si au contraire elle est bien due à une changement réel de la richesse. Contrairement à l'affirmation de Karakassis (1995) à propos de  $S\infty$ , il n'y a pas d'estimateur indépendant de la distribution de fréquence. Ce n'est pas parce que la distribution des individus entre espèces n'intervient pas dans le calcul de l'estimateur, que celui-ci en est indépendant. Cette distribution influence directement la représentativité et la qualité de l'échantillon, donc aussi la richesse observée sur cet échantillon, ce qui se répercute forcément sur tous les estimateurs.

Quoique le concept de richesse spécifique soit très simple, la mesure de la richesse spécifique est donc particulièrement délicate. Les méthodes présentées ici permettent d'accéder à des estimations assez réalistes, mais à la condition d'un effort d'échantillonnage déjà important. En particulier, les estimateurs Jackknife et Bootstrap fournissent une estimation peu biaisée et précise dés lors qu'ils s'appuient sur des échantillons dont la richesse observée n'est pas trop éloignée de la richesse vraie. A ce sujet, Slocomb et Dickson (1978), repris par Baltanas (1992), avaient montré qu'une estimation fiable de la richesse demandait qu'au moins 80% du nombre total d'espèces soit déjà présents dans les échantillons. En ce qui nous concerne, les simulations (Tableaux 4.5 et 4.6) indiquent plutôt un minimum de 90% !

On ne peut donc espérer procéder à une estimation de la richesse de vastes espaces écologiques ou de milieux complexes ou hétérogènes sur la base d'échantillonnages légers, ainsi qu'on aurait pu le vouloir dans le cadre d'opérations de suivi écologique, entrant par exemple dans l'esprit de la surveillance des changements globaux et de l'évaluation des modifications de la biodiversité.

1	7	4
•		

m 11 4 6		• •	1		• • •			1	1.	• •	•
Lableau 4	\ ' f	nchesse.	onservee	expr	imee	en i	nourcentage	de.	12 1	nchesse	vrale
Laciona ino		10110000	00001100	- APA		<b>W</b> AA .	pourooniu <sub>b</sub> o		100 1	101100000	11410.

	Modèle Preston 50										
	50	100	500	1000	5000	10000					
10	62.58	72.23	89.26	93.78	98.90	99.72					
20	72.32	80.53	93.76	96.67	99.68	99.97					
40	80.52	87.07	96.64	98.64	99.97	99.99					

	Modèle Motomura 1								
	50	100	500	1000	5000	10000			
10	46.55	50.94	62.99	67.85	79.06	84.51			
20	51.07	55.69	67.66	73.05	84.03	89.22			
40	55.45	60.11	72.46	77.95	88.96	93.46			

	Modèle Motomura 2									
	50	100	500	1000	5000	10000				
10	53.76	60.17	74.97	81.20	93.08	97.09				
20	59.76	66.47	81.16	86.58	96.88	99.26				
40	65.88	72.11	86.61	91.80	99.15	99.95				

	Modèle Preston 200										
	50	100	500	1000	5000	10000					
10	28.42	36.08	57.21	65.44	82.66	87.92					
20	36.43	44.82	65.80	73.55	88.08	92.05					
40	45.47	53.81	73.71	80.47	92.31	95.02					

Tableau 4.6 : estimation Jackknife 2 exprimée en pourcentage de la richesse vraie.

	Modèle Preston 50										
	50	100	500	1000	5000	10000					
10	82.26	89.82	98.48	99.95	100.63	100.34					
20	89.94	93.98	100.15	100.78	100.17	99.92					
40	94.58	96.79	100.77	101.33	99.95	100.00					

	50	_100	500	1000	5000	10000
_10	55.58	60.20	71.83	77.14	89.49	94.02
20	59.99	64.88	77.44	83.53	94.03	98.18
40	65.29	69.33	83.87	89.79	98.34	100.44

	Modèle Motomura 2										
	50	100	500	1000	5000	10000					
10	66.10	72.38	86.71	92.08	101.70	101.92					
20	72.54	78.05	92.62	97.74	101.94	100.43					
40	78.79	84.78	97.53	102.10	100.87	99.76					

	50	100	500	1000	5000	10000
10	44.87	53.57	74.81	81.77	93.88	96.51
20	55.34	63.59	82.73	88.33	97.05	98.26
40	65.03	73.02	87.90	92.92	98.50	99.54

# Modèle Motomura 1

## Modèle Preston 200

# IV.4 - Estimation de l'indice H' de Shannon

### IV.4.1 Quelques rappels

Les indices de diversité font l'objet d'une abondante bibliographie ainsi que de très nombreuses revues et études comparatives. On peut se reporter à Leprêtre (1997) pour une liste quasi exhaustive (bien que l'auteur s'en défende !) des principales références sur le sujet.

Ces indices, basés sur l'hétérogénéité de la distribution des individus entre espèces, semblent tout comme la richesse spécifique bénéficier d'une nouvelle vague d'intérêt liée aux problématiques de la conservation des milieux et du patrimoine biologique.

Parmi les nombreux indices proposés, le plus populaire et le plus souvent utilisé est certainement l'indice de Shannon-Wiener H', indice basé sur la théorie de l'information (Shannon et Weaver, 1949) et introduit en écologie par Margalef (1956, 1958).

L'expression en est: 
$$H' = -\sum_{i=1}^{S} p_i \cdot \log_2(p_i)$$
,

où  $p_i$  est la proportion d'individus représentée par l'espèce *i* (parmi les *S* espèces présentes) dans la collection considérée. Il correspond, sous la forme  $e^{H'}$ , au nombre de diversité de Hill d'ordre 1 (Hill, 1973).

Compte tenu de la sous-évaluation systématique du nombre d'espèces par un échantillon, cet indice aura tendance également à sous-estimer la diversité H' réelle ou  $H'_{vraie}$  (Good, 1953; Adams et McCune, 1979; Magnussen et Boyle, 1995). Néanmoins, contrairement à la richesse qui est systématiquement sous-estimée, ceci n'est vrai ici qu'en espérance mathématique. Un prélèvement quelconque dans une communauté peut donner lieu à une valeur H' observée ( $H'_{ech}$ ) supérieure à la valeur  $H'_{vraie}$ . On peut noter que cet indice appliqué le plus souvent à des effectifs, peut être utilisé sur des valeurs de biomasses.

Comme la richesse, l'indice de Shannon H' souffre d'une sous-estimation : en moyenne, on constate que  $H'_{ech}$  est très inférieur à  $H'_{vrai}$ . On consultera Bowman (1971) pour une description des propriétés statistiques de H' (et d'autres indices de diversité). Zahl (1977) et Adams et McCune (1979) proposent une méthode efficace d'estimation de  $H'_{vrai}$  fondée sur le Jackknife. Mais le défaut de la méthode Jackknife est de reposer sur la nécessité d'un certain nombre de réplicats (souvent une dizaine, au moins), qui n'est pas souvent réaliste par rapport aux moyens mis en œuvre sur le terrain

#### IV.4.2 Proposition d'une nouvelle méthode d'estimation de H'vrai : H' CRIC

### IV.4.2.1 Principe de l'estimation

Devant les difficultés posées par les ajustements d'une part, et du fait des critiques portant souvent sur le côté trop synthétique des indices de diversité d'autre part, une première idée a été de décomposer graphiquement l'indice de diversité de Shannon de manière à faire apparaître non seulement sa valeur, mais aussi la façon dont il se forme.

En reprenant le principe des courbes de k-dominance, on représente graphiquement le cumul progressif des  $p_i \cdot log_2(p_i)$  en fonction des rangs des taxons. Cette courbe rangsinformations cumulées (ou CRIC) représente donc la fonction

$$CRIC(k) = -\sum_{i=1}^{k} p_i \cdot \log 2(p_i)$$
 avec k = 1 à S

La courbe obtenue atteint son maximum à la valeur  $H' = -\sum_{i=1}^{s} p_i \log_2 p_i$ , c'est à dire

quand on a cumulé tous les composants de l'information, correspondant aux S espèces ou taxons. Cette courbe coupe nécessairement (ou au mieux se confond avec) le segment de droite correspondant à un échantillon de S espèces ayant des effectifs égaux (régularité J = 1). Ce segment atteint sa valeur maximale à la valeur  $H_{max}^{*} = \log_2 S$ .

Quelle que soit la distribution de fréquences d'origine, la forme des courbes CRIC est toujours bien régulière et lisse ; elle ne semble pas vraiment "marquée" par les problèmes de discrétisation, classiques dans la représentation de type diagrammes rangs-fréquences (fig. 4.12).



Figure 4.12 : représentation en Courbe Rangs-Informations Cumulées (CRIC) des différents modèles utilisés dans les simulations.

1 - modèle de Preston à 50 espèces

2 – modèle Mixte

3 – modèle de Motomura à 30 espèces

4 – modèle de Motomura à 20 espèces

Un autre intérêt de cette méthode de représentation est que les courbes CRIC obtenues, sur les modèles théoriques complets ou sur des échantillons importants, atteignent leur valeur maximum H' de manière asymptotique. Il apparaît donc que l'extrapolation de la courbe de manière à obtenir la valeur de l'asymptote devrait permettre de donner une estimation du H' vrai du peuplement, à partir d'un échantillon. Cette estimation devrait surtout se révéler intéressante quand on a des échantillons de relativement petite taille ou quand les espèces peu abondantes sont sous-représentées dans l'échantillon et que l'allure des courbes apparaît alors comme tronquée, n'atteignant pas de valeur asymptotique.

## IV.4.2.2 Calcul de l'estimateur H' CRIC

On suppose que la courbe CRIC est proche d'un modèle de type croissance exponentielle : modèle de Von Bertalanffy, équation d'Ivlev ou modèle exponentiel utilisé pour la modélisation des variogrammes (fig. 4.13).

Pour une espèce de rang i, on a la valeur de la courbe CRIC :  $C_i = C_{\infty} [1 - e^{-ki}]$ 

avec C<sub>∞</sub> asymptote de la courbe et k coefficient de "croissance instantanée".



Figure 4.13 : courbe CRIC suivant un modèle de Von Bertalanffy.

On utilise le principe du diagramme de Lockwood pour linéariser cette courbe et calculer ainsi la valeur de l'asymptote (cf. Frontier et Viale, 1995).

Soit C<sub>i</sub> la valeur de la courbe CRIC pour l'espèce de rang i.

On représente graphiquement les valeurs suivantes :

$$X_i = (C_i + C_{i+1}) / 2$$
  
 $Y_i = C_{i+1} - C_i$ 

Une régression linéaire appliquée à ces valeurs (X,Y) permet d'obtenir une droite de pente -k, dont l'intersection avec l'axe des abscisses donne la valeur de l'asymptote  $C_{\infty}$  (fig. 4.14).

Cette valeur  $C_{\infty}$  donne l'estimation de H' telle que semble le laisser supposer la forme de la courbe CRIC. Le paramètre k caractérise la régularité de la distribution ou la dominance relative des espèces. Il est plus élevé pour les distributions dominées par très peu d'espèces et prend des valeurs faibles pour les distributions régulière, dans lesquels les individus se partagent équitablement entre un nombre important d'espèces.



Figure 4.14 : diagramme de Lockwood correspondant à la courbe CRIC de la fig.4.13.

On remarque cependant que les courbes CRIC observées (réelles aussi bien que simulées) ne sont souvent linéarisées qu'imparfaitement, ou ne le sont que sur une partie de leur profil. Cet écart à la linéarité affecte généralement les extrémités du diagramme et concerne donc :

- les espèces peu fréquentes pour lesquelles on a une discrétisation des valeurs de la courbe (et donc une forme en paliers).

- les premières espèces, à savoir les plus abondantes, qui se situent souvent au-dessus de la droite d'ajustement.

Les courbes CRIC semblent donc ne répondre à un modèle de type exponentiel que partiellement, d'où notre choix d'appliquer, non pas une régression linéaire classique, mais plutôt une régression robuste qui va se montrer plus résistante à l'effet des points aberrants.

Deux méthodes de régression robuste ont été utilisées :

- la Least Trimmed Squares Regression (LTS) proposée par Rousseuw et Leroy (1987) qui consiste à minimiser la somme des q plus petits carrés des résidus, avec q un peu plus grand que n/2.

- une méthode de régression (RREG), basée sur un M-estimateur (Huber, 1981) développée dans le logiciel S-PLUS (Heiberger et Becker, 1992).

Ces deux méthodes s'avèrent plus performantes que la régression classique pour l'estimation de H'. La méthode RREG apparaît comme celle donnant en moyenne l'estimation la moins biaisée de H'. Néanmoins, la méthode LTS, bien que donnant de moins bons résultats en général, semble plus à même de s'adapter à des distributions très particulières et à donner une estimation correcte, même quand un nombre important de points s'écartent de la linéarité.

Dans tous les cas, un examen visuel du diagramme de Lockwood semble une étape indispensable pour valider la méthode de régression employée et l'estimation de H' qui en résulte.

#### IV.4.3 Application à des données entomologiques existantes

Nous donnons ici une illustration de la méthode d'estimation du H' vrai par la méthode CRIC, en utilisant la méthode de régression robuste. Les données utilisées sont des données d'échantillonnage d'insectes tirées d'une étude de Leprêtre (1986) portant sur l'influence des lignes à haute tension sur les peuplements entomologiques en forêt ardennaise.

Un exemple de CRIC est donné à la figure 4.15, réalisée à partir de 2 distributions de fréquences observées sur des échantillons d'insectes recueillis dans une coupe forestière sous ligne à haute tension. Le point *5cal* correspond à une callunaie, le point *5mol* à une moliniaie, dans laquelle la diversité est moindre, et présentant un DRF plus pentu (régularité moindre). Sur la figure 4.16, les deux segments correspondant aux régularités maximales ont été superposés aux CRIC. On lit donc en ordonnée la valeur de H' des deux échantillons ainsi que la valeur  $H'_{max}$ . L'abscisse des derniers points de chaque courbe donne la richesse ( $S_{mol}$  et  $S_{cal}$ ). Les moindres diversité et régularité du peuplement de *5mol* apparaissent sous trois aspects: moindre valeur de S, moindre valeur de H', et CRIC plus linéaire à pente faible.

La figure 4.16 montre l'allure des diagrammes de Lockwood obtenus pour les points 5mol et 5cal utilisés précédemment. Une régression linéaire est ensuite réalisée, de manière à évaluer la valeur limite (baptisée  $H'_{CRIC}$ ), qui est donnée par l'intersection avec l'axe des abscisses (tableau 4.7).



Figure 4.15 : Distributions de fréquences observées sur 2 échantillons d'insectes recueillis dans une callunaie (5cal) et une moliniaie (5mol). A gauche, représentation sous forme de DRF (échelle bilogarithmique) A droite, représentation sous forme de CRIC (Courbes Rangs-Informations Cumulées)



Figure 4.16 : Diagrammes de Lockwood appliqués aux CRIC des échantillons d'insectes 5cal et 5mol, estimations de H'CRIC et pentes de la régression robuste

	H'	H' cric	k
5cal	3.381	3.421	0.164
5mol	1.782	1.861	0.132

Tableau 4.7 : valeurs de H', H'cric et pente k pour les 2 échantillons 5cal et 5mol.

## IV.4.4 Robustesse et efficacité de l'estimateur CRIC

Dans l'exemple précédent, il impossible de juger de l'efficacité de l'estimation de H' dans la mesure où nous ne connaissons pas la véritable valeur de H', d'où la nécessité de valider la méthode sur des échantillons simulés tirés de distributions connues.

Les simulations sont réalisées à partir de tirages dans les quatre modèles présentés au 2.2 : le modèle Preston à 50 espèces, les deux modèles Motomura M1 et M2 et le modèle mixte. Six efforts d'échantillonnage différents sont employés : 50, 100, 500, 1000, 5000 et 10000 individus.

Dans chacun de ces cas, 1000 échantillons sont simulés et on calcule sur chaque échantillon l'indice de Shannon H' dans sa formulation classique ainsi que l'estimateur de H' à partir des courbes CRIC en utilisant les différents types de régressions.

Les résultats obtenus sont synthétisés dans les figure 4.17 à 4.20 et dans les tableaux en annexe III. On constate d'abord qu'effectivement, l'indice de Shannon H' est sous-estimé par les échantillons surtout quand les effectifs collectés sont faibles. Cependant, on peut voir également que, d'une part, le biais n'est jamais très important (au contraire de ce qu'on a pu voir pour la richesse spécifique), et que d'autre part, ce biais devient très faible et presque nul dès que les tailles des échantillons dépassent quelques milliers d'individus. De même que pour le biais, la variabilité de H' décroît très rapidement avec l'augmentation de l'effort d'échantillonnage. Avec les échantillons de 5000 et 10000 individus, on a, quel que soit le modèle, une estimation très juste et très précise de l'indice de Shannon.

La différence entre les modèles est nettement moins marquée que pour la richesse. La situation apparaît même inversée par rapport au problème de l'estimation de la richesse spécifique : une bonne estimation semble plus facile à obtenir avec les modèles Motomura qu'avec le modèle Preston.

Les résultats sur les estimations à partir des courbes CRIC sont beaucoup plus mitigés. Si on arrive à des résultats satisfaisants avec les modèles Preston et Mixte, ce n'est pas le cas avec les modèles Motomura. Dans toutes les situations testées, les régressions robustes (RREG ou LTS) ont généralement fait mieux que la régression linéaire classique pour le biais. La méthode est très performante dans le cas du modèle de Preston, surtout avec l'estimation de l'asymptote par la régression robuste (fig. 4.17). L'estimation de la valeur réelle de l'indice de Shannon est systématiquement meilleure que par la formulation habituelle de H'. Pour une taille d'échantillon de 100 individus, le biais est déjà quasiment inexistant. La variabilité des estimations par les courbes CRIC est à peine supérieure à celle que présente les valeurs classiques de H'.

Les résultats sont un peu moins satisfaisants avec le modèles mixte. Dans ce cas, c'est la méthode de régression LTS qui donne les meilleurs résultats (fig. 4.18). L'estimateur CRIC converge plus rapidement vers la vraie valeur de H' pour les plus petits échantillons. Pour les échantillons moyens et plus importants, le biais qu'il présente en surestimation est du même ordre de grandeur que la sous-estimation produite par le H' classique. La dispersion des valeurs obtenues est assez importante, ce qui s'observe de manière générale avec la régression LTS.

La méthode d'estimation CRIC n'est par contre pas du tout performante sur les modèles Motomura (fig. 4.19 et 4.20), pour lesquels la surestimation est toujours trop forte, et dépasse nettement le biais du H' classique.



Figure 4.17 : simulations à partir du modèle Preston à 50 espèces. moyennes et IC95% de H' et de l'estimateur CRIC avec régression robuste



Figure 4.18 : simulations à partir du modèle mixte : moyennes et IC95% de H' et de l'estimateur CRIC avec régression LTS



Figure 4.19 : simulations à partir du modèle Motomura 1 : moyennes et IC95% de H' et de l'estimateur CRIC avec régression robuste



Figure 4.20 : simulations à partir du modèle Motomura 2 : moyennes et IC95% de H' et de l'estimateur CRIC avec régression robuste

## IV.4.5 Conclusion sur l'estimation de H'

L'estimation de l'indice de diversité de Shannon par l'extrapolation des courbes CRIC semble particulièrement adéquate dans le cas de communauté d'espèces distribuées selon un modèle de Preston ou tout autre modèle présentant une certaine régularité de distribution des individus entre les différentes espèces. Ce résultat est donc encourageant, montrant que la méthode est opérationnelle dans de nombreuses situations.

Cependant, les mauvaises performances sur les modèles de distribution où existe une forte dominance (tels que Motomura), tiennent au choix de la méthode pour estimer l'asymptote. Nous avons supposé ici que les courbes CRIC avaient une forme de type exponentiel négatif ou Von Bertalanffy. Ceci ne semble être vrai que sur une partie de la courbe – ce qui explique les meilleures performances des régressions robustes face à la régression classique. Ces méthodes robustes diminuent en effet (voire annulent) l'influence des premiers points de la courbe (les espèces les plus abondantes) dans l'estimation de l'asymptote.

Il conviendrait donc, pour que la méthode CRIC puisse donner des estimation fiables de H' dans toutes les situations, de disposer d'un modèle (ou d'un petit nombre de modèles) de courbe asymptotique suffisamment flexible pour s'ajuster aux différentes formes de courbes CRIC qu'on peut obtenir. Dans cette optique, nous avons essayé l'ajustement au modèle de Von Bertalanffy par les moindres carrés non-linéaires, l'ajustement à un modèle de type Michaelis-Menten ou celui à divers modèles utilisés en géostatistique pour les variogrammes, mais aucune de ces solutions n'a encore fournis d'estimations de meilleure qualité.

Nous proposons donc de tracer systématiquement le diagramme rangs-fréquences à partir de l'échantillon, et de choisir le mode d'estimation au vu du profil obtenu. Si la distribution est relativement régulière (type Preston ou Mandelbrot), alors H' cric est efficace. Sinon, lorsque la distribution montre une forte dominance des premières espèces (type Motomura), la mesure H' d'échantillon est une bonne estimation de H' vrai.

## **IV.5 - Estimation de la diversité** β

#### IV.5.1 Problème posé

Whittaker (1972) a décrit la "diversité bêta" comme étant l'augmentation de la diversité quand la surface prospectée (l'effort d'échantillonnage) augmente. Il suggère en effet de nommer "diversité alpha" ( $H'_{\alpha}$ , pour l'indice de Shannon), la diversité en une station d'observation donnée, et "diversité gamma" ( $H'_{\gamma}$ ), cette même diversité calculée sur une plus grande échelle. La "diversité bêta" ( $H'_{\beta}$ ), qu'il vaudrait mieux dénommer "distance bêta" ou "changement bêta", est donnée par la différence entre  $H'_{\gamma}$  et la moyenne des  $H'_{\alpha}$ .

Dans le cas de 2 observations (2 stations ou 2 réplicats) dans un secteur donné, on a 2 mesures de  $H'_{\alpha}$ , et on a donc:

$$H'_{\beta} = H'_{\gamma} - \frac{1}{2} (H'_{\alpha_1} + H'_{\alpha_2})$$
 (Pielou, 1975; Blondel, 1986)

On peut généraliser cette formule pour plus de 2 réplicats :

$$H'_{\beta} = H'_{\gamma} - moy(H'_{\alpha})$$

Les résultats précédents (§ 4) ont montré que le biais sur l'estimation de  $H'_{éch}$  est très sensible à l'effort d'échantillonnage et dépend de la taille de l'échantillon : le biais diminue quand l'effectif augmente, plus l'échantillon est petit et plus la valeur de H' a de chances d'être sous-estimée.

Il y a donc un problème pour l'estimation de la diversité  $\beta$ . Dans la formule ci-dessus, le H'<sub> $\gamma$ </sub> est calculé sur un échantillon global, somme de tous les échantillons élémentaires. Il va donc être moins biaisé que les H'<sub> $\alpha$ </sub> (et que leur moyenne), calculés sur des échantillons plus petits. Donc même si on tire au hasard des échantillons dans une communauté théorique homogène, s'il n'y a pas d'effet "mosaïque" et pas de diversité  $\beta$  (H' $\beta$  = 0), le simple fait que H'<sub> $\gamma$ </sub> soit réalisée sur un échantillon plus grand que les H'<sub> $\alpha$ </sub> va systématiquement donner une valeur de H' $\beta$  supérieure à 0. La diversité  $\beta$  ainsi obtenue n'exprime alors rien d'autre qu'un effet de l'échantillonnage statistique et pas une augmentation de diversité due au recrutement de nouveaux biotopes.

Une première solution pour éviter ce biais pourrait consister à n'utiliser que des estimations de  $H'_{vrai}$ , comme celles obtenues à partir des courbes CRIC, de manière à être garanti d'une plus grand indépendance vis-à-vis de l'effort d'échantillonnage et de ne percevoir et mesurer ainsi que l'effet du changement d'échelle.

Une autre possibilité est de procéder à du rééchantillonnage. En faisant la somme des échantillons élémentaires, on réalise l'échantillon  $\gamma$  et on tire dans cette réunion un nouvel échantillon, de taille équivalente aux échantillons élémentaires. Le H' $\gamma$  est ensuite calculé sur cet échantillon. On élimine donc le problème de dépendance du biais par rapport à l'effort d'échantillonnage, puisque le H' $\gamma$  et les H' $\alpha$  sont estimés à partir d'échantillons de même taille. Deux procédures ont été testées : l'une avec tirage avec remise, l'autre avec tirage sans remise.

Cette dernière méthode peut encore être améliorée en répétant de nombreuses fois l'opération de tirage de l'échantillon  $\gamma$ . La moyenne de ces différentes réalisations fournit ainsi une estimation encore plus fiable de la diversité  $\beta$ .

### IV.5.2 Simulations et résultats

Pour évaluer le biais des différentes méthodes d'estimation proposées on réalise des simulations réalisées à partir de tirages d'échantillons pris dans un modèle de Preston à 50 espèces. Comme tous les échantillons sont tirés dans une même communauté théorique parfaitement homogène, il n'y a pas ici d'effet "mosaïque" et donc pas de diversité  $\beta$ . La valeur réelle de H' $\beta$  est donc égale à 0.

Les échantillons constituant les réplicats (échantillons  $\alpha$ ) sont des tirages de 100 et 500 individus. Dans chacun de ces deux cas, H' $_{\beta}$  est estimé avec 2, 5, 10 et 20 réplicats.

Pour chacune des méthodes envisagées, on obtient 8 combinaisons différentes pour lesquelles 1000 simulations sont réalisées.

Les deux figures 4.21 et 4.22 présentent graphiquement les résultats obtenus, qui sont aussi synthétisés dans le tableau 4.8. On voit clairement dans ces résultats, que l'estimation de H' $\beta$ , en utilisant la formule classique de l'indice de Shannon, est systématiquement biaisée. La valeur réelle de H' $\beta$ , qui est égale à 0, ne se situe jamais dans l'intervalle de confiance à 95% de ces estimations. Les valeurs obtenues sont toujours positives et plus le nombre d'échantillons  $\alpha$  est grand, plus ce biais augmente car l'écart entre le nombre d'individus utilisés pour les H' $\alpha$  et pour le H' $\gamma$  est de plus en plus important.

Les estimations réalisées à partir des courbes CRIC sont beaucoup plus performantes. Elles présentent une variabilité légèrement plus importante que les estimations classiques mais le biais est considérablement réduit. On peut remarquer que l'évolution de ce biais quand on augmente le nombre d'échantillons  $\alpha$  est différente pour les 2 tailles d'échantillons testées. Ce biais est négatif et tend à diminuer pour les échantillons de 100 individus alors qu'il est positif et augmente pour les échantillons de 500 individus. Pour le dernier cas - à savoir 20x500 individus - la valeur 0 n'est pas dans l'intervalle de confiance à 95%.

Les deux méthodes de rééchantillonnage, qui permettent de ramener les échantillons  $\gamma$ à la taille des échantillons  $\alpha$  à l'aide de tirages avec et sans remise, donnent des résultats quelque peu différents. La dispersion des estimations obtenues par ces deux méthodes est très importante. Elle reste stable pour les tirages avec remise alors qu'elle augmente avec le nombre de réplicats pour les tirages sans remise. Les tirages avec remise présentent un biais négatif relativement faible, qui diminue quand on ajoute des échantillons  $\alpha$ . Le biais est par contre presque inexistant pour les tirages sans remise et il reste stable.

La méthode de tirage sans remise dans l'échantillon  $\gamma$  d'un échantillon d'un effectif équivalent aux échantillons  $\alpha$  semble celle qui minimise le plus le biais. On peut résoudre le problème de sa forte variabilité en répétant le processus de tirage pour chaque échantillon. C'est ce que propose la dernière méthode testée, en répétant 200 fois le tirage pour chacun des échantillons simulés. En conservant comme valeur la moyenne de ces 200 répétitions comme estimation, on obtient ainsi une évaluation de H' $\beta$  très peu biaisée et très précise.

La qualité de l'estimation de H' dépend fortement des effectifs collectés dans les prélèvements, et cette dépendance provoque, comme on le voit dans les simulations réalisées, une erreur systématique sur l'estimation de la diversité  $\beta$ . Les méthodes de rééchantillonnage proposées ici permettent de s'affranchir de ce problème et d'obtenir ainsi une très bonne

estimation. Ces méthodes restent quand même relativement lourdes en termes de calcul et de développement informatique. L'utilisation des estimations CRIC peut alors représenter une alternative plus simple et d'assez bonne qualité.

				H'				
		100 indi	ividus			500 indi	vidus	
	2	5	10	20	2	5	10	20
IC inf	0.060	0.144	0.183	0.208	0.019	0.042	0.050	0.055
Moyenne	0.106	0.183	0.213	0.230	0.030	0.050	0.057	0.060
IC sup	0.158	0.227	0.246	0.254	0.043	0.060	0.065	0.066
				CRIC				
		100 in	dividus			500 in	dividus	
	2	5	10	20	2	5	10	20
IC inf	-0.186	-0.157	-0.114	-0.087	-0.028	-0.014	-0.002	0.008
Moyenne	-0.033	-0.030	-0.023	-0.018	0.008	0.017	0.022	0.025
IC sup	0.138	0.097	0.068	0.052	0.047	0.050	0.046	0.044
Tirage avec remise								
	100 individus					500 in	dívidus	
	2	5	10	20	2	5	10	20
IC inf	-0.424	-0.359	-0.364	-0.361	-0.206	-0.201	-0.185	-0.183
Moyenne	-0.074	-0.027	-0.011	-0.003	-0.026	-0.015	-0.008	<b>-0</b> .007
IC sup	0.268	0.316	0.330	0.341	0.149	0.163	0.170	0.163
			Tirage	e sans remise				
	_	100 in	dividus		500 individus			
	2	5	10	20	2	5	10	20
IC inf	-0.263	-0.312	-0.354	-0.374	-0.128	-0.155	-0.181	-0.184
Moyenne	0.002	0.006	0.003	0.001	-0.001	-0.003	-0.003	-0.003
IC sup	0.243	0.334	0.314	0.338	0.113	0.148	0.173	0.170
			Firage sans re	emise avec ré	pétition			
		100 in	dividus			<b>500</b> in	ldividus	
	2	5	10	20	2	5	10	20
IC inf	-0.0405	-0.0367	-0.0346	-0.0320	-0.0128	-0.0148	-0.0127	-0.0123
Moyenne	0.0012	0.0004	0.0007	0.0007	0.0005	0.0003	0.00006	0.0002
IC sup	0.0560	0.0449	0.0350	0.0308	0.0180	0.0152	0.0131	0.0130
			-					

Tableau 4.8 : limites de l'intervalle de confiance à 95% et moyenne des valeurs de H' $_{\beta}$  obtenues par les différentes méthodes et pour les différents efforts d'échantillonnage simulés.











## IV.6 - Synthèse - Conclusion

Pour résumer l'essentiel des résultats de cette étude sur les indices de diversité, une première conclusion est la confirmation que la richesse spécifique, bien qu'étant un des plus simple paramètre en écologie des communautés, est aussi certainement l'un des plus difficiles à estimer correctement. En effet, les simulations ont montré qu'il est nécessaire de collecter un nombre considérable d'individus pour espérer obtenir une estimation très proche de la vérité, même dans le cas de communautés supposées répartis de manière homogène dans l'espace. L'estimation de la richesse serait donc plus accessible dans le cas de communautés organisées en mosaïque – ce qui est la modalité écologique la plus fréquente !

Les estimateurs non-paramétriques, tels que le Jackknife et le Bootstrap, jugés dans d'autres études similaires comme les plus aptes à produire une bonne estimation, n'apportent en aucun cas une solution valable pour des échantillons trop restreints : ils n'aboutissent à une estimation juste que lorsque la richesse observée approche les 90% de la richesse vraie, ce qui implique un échantillonnage conséquent. Ils sont cependant toujours moins biaisés que la richesse observée.

L'estimation de la richesse dépend évidemment fortement des effectifs collectés (effort d'échantillonnage) mais également de la façon dont ces effectifs se répartissent entre les espèces. Il est beaucoup plus simple d'obtenir une bonne estimation de la richesse pour une distribution de fréquences qui présente une forte régularité. La richesse spécifique est donc un paramètre difficile à utiliser pour comparer la diversité de différents sites ou du même site lors d'un suivi temporel. Une forte différence peut apparaître, du simple fait d'un changement de la structure de la communauté ou même d'un changement de la densité des individus, et ce, sans que la richesse n'ait évolué.

L'indice de Shannon paraît beaucoup moins exigeant : une estimation assez précise requiert des échantillons plus petits. Les résultats des simulations montrent que pour des effectifs de 1000 à 10000 individus, le H' d'échantillon donne une estimation déjà très juste du H' de la communauté. Dans le cas d'échantillons comportant des effectifs plus réduits, la proposition d'une estimation basée sur l'extrapolation des courbe CRIC donne des résultats très satisfaisants quand on s'adresse à des distributions de fréquences assez régulières (type

192

Preston). La méthode CRIC devrait être améliorée pour s'appliquer aux cas de distributions présentant un fort degré de dominance (type Motomura).

Dans le cadre de suivis ou de comparaisons de communautés, l'indice de Shannon est donc un indicateur numérique assez fiable de la diversité taxonomique.

Les problèmes d'estimation constatés sur la richesse spécifique se répercutent évidemment sur un grand nombre d'autres indicateurs de diversité qui utilisent la valeur de la richesse observée dans leur formulation (par exemple Margalef, 1958 ou Menhinick, 1964).

Ainsi, par exemple, pour l'indice de régularité, ou *evenness* de Pielou (1975), dont l'expression est  $J' = H' / \log_2(S)$ , on peut s'attendre sur des échantillons de tailles raisonnables à ce qu'il y ait peu de biais sur l'estimation de H', alors la richesse S, risque d'être fortement sous-estimée. Cette différence entre les deux indices va induire une surestimation de la régularité.

On peut voir dans l'étude menée ici que l'utilisation de données simulées peut se révéler très utile pour établir les propriétés de méthodes numériques utilisé en écologie. Palmer (1990) fait remarquer que les simulations peuvent ne pas correspondre à des situations réelles et préfère donc une approche empirique basée sur un échantillonnage intensif d'une communauté réelle. Une telle approche est forcément très coûteuse, voire même impossible à mener à bien dans certains milieux ou pour certains organismes. L'intérêt des simulations est qu'elles peuvent fournir une large gamme de *patrons* supposés typiques de diverses communautés et des jeux de données de grandes tailles, permettant une véritable approche statistique. Rien n'empêche d'ailleurs de simuler des distributions qui s'inspirent de distributions observées (ce que nous avons fait avec le modèle "mixte").

Dans la suite des travaux, il serait intéressant – pour être en accord avec les situations de terrain – de générer des jeux de données qui simulent aussi la répartition hétérogène des espèces sur le terrain. C'est d'ailleurs ce qui est actuellement développer par D. Mouillot (thèse en cours à l'Université de Corte), avec la fabrication de "champs" sur lesquels les différentes espèces sont réparties avec des degrés d'agrégation variés. Il sera ainsi possible de tester non seulement l'effet de l'effort d'échantillonnage (en terme d'effectifs collectés et de nombre de réplicats), mais aussi l'effet du plan d'échantillonnage utilisé en rapport avec l'organisation spatiale des communautés.

193

Conclusion générale

## **Conclusion générale**

La première partie de ce travail, consacrée à l'étude des séries chronologiques de qualité des eaux de surface dans le bassin Artois-Picardie, a permis la mise en place d'une méthodologie opérationnelle de traitement de ces séries en vue de leur utilisation pour réaliser des prévisions à court et moyens termes. Les résultats obtenus sur les 67 points de mesures du Réseau National de Bassin ont montré une amélioration globale de la qualité de l'eau, surtout perceptible dans les secteurs fortement urbanisés de l'agglomération lilloise et du bassin minier. Cette amélioration concerne aussi bien l'azote et le phosphore que les paramètres liés à l'oxygène, révélateurs de la pollution organique. L'augmentation des concentrations en nitrates peut en grande partie s'expliquer par le transfert, vers cette forme, des formes organiques de l'azote. Ce passage de l'azote sous sa forme la plus oxydée est l'indice d'une amélioration globale de la qualité de l'eau (meilleure oxygénation en particulier) et d'une meilleure gestion des apports au niveau du bassin versant. L'exemple de l'effet de la mise en service de la nouvelle station d'épuration de Douai sur la qualité des eaux de la Scarpe est particulièrement représentatif du phénomène (Noppe et al., 1997) : la chute brutale de l'azote d'origine organique (N Kjeldahl et NH4<sup>+</sup>) s'est accompagnée d'une forte hausse des concentrations en nitrates, bien que l'azote total soit en baisse.

L'évolution des flux reflète l'évolution des concentrations : le début des années 90 a donc vu une baisse globale à l'échelle du bassin Artois-Picardie des exportations en azote et phosphore. Cette diminution est surtout sensible à l'est du bassin (secteur frontalier avec la Belgique), et plus particulièrement sur le bassin de l'Escaut et de ses affluents. Il convient d'être plus prudent sur l'évaluation des tendances pour les cours d'eau débouchant à l'ouest sur le littoral de la Manche. On peut mettre en évidence sur ces cours d'eau une relation très forte entre les débits et les concentrations de différents éléments, comme les nitrates ou les matières en suspension. Par conséquent, prévoir de manière fine l'évolution des concentrations supposerait que l'on puisse préalablement prédire l'évolution des conditions climatiques ! De même, la diminution des flux observée sur cette façade de la Manche, pour la période 90-94, n'est peut être due qu'à la baisse des débits sur cette même période, et on ne peut pas conclure à une réelle diminution des apports potentiels du bassin versant.

L'étude de la série benthique de Gravelines montre, comme Fromentin (1995) l'a mis en évidence sur les différentes séries benthiques du PNOC-SLT, une évolution cyclique du peuplement, liée aux facteurs climatiques et surtout à la température. La méthode que nous proposons, consistant à regarder l'évolution temporelle de la distance entre les coordonnées des relevés et les coordonnées d'une situation moyenne issues d'une analyse factorielle, peut être particulièrement intéressante pour révéler le lien entre l'évolution de la structure multispécifique du peuplement et celle des variables du milieu (sédimentologiques, hydrodynamiques, climatiques, etc...).

L'étude de la diversité sur la série brute ne permet pas de mettre en évidence des modifications temporelles significatives. L'augmentation de la richesse observée sur les échantillons n'est qu'un "artefact", dû à la forte corrélation existant entre cette richesse observée et la densité des organismes qui a augmenté depuis la mise en place de la surveillance du site de Gravelines. Il y a donc eu seulement une augmentation de la richesse "relative", à savoir de la richesse par m<sup>2</sup>, mais on ne peut conclure à une augmentation réelle de la richesse du peuplement. La tendance qui s'exprime sur la richesse est donc liée à un changement d'échelle de l'organisation spatiale des individus de ce peuplement. De même, aucune tendance marquée ne peut être mise en évidence sur l'évolution d'autres indicateurs de la diversité, tels que l'indice de Shannon H'. Il est ainsi nécessaire d'utiliser un grain de perception plus grossier (une fenêtre plus large) que le grain d'observation utilisé dans le suivi. Ce n'est qu'avec une perception à l'échelle annuelle (au moins) que l'on arrive à dégager des tendances à long terme sur les descripteurs de la diversité. Cette tendance à l'augmentation des indices de diversité à l'échelle de la série complète est surtout due à des changements dans le dominance relative des espèces les plus abondantes de la communauté. Si on ne s'intéresse qu'aux espèces moins abondantes (en utilisant un indicateur tel que la pente P2), on ne détecte pas de tendance à l'augmentation, mais uniquement des variations cycliques de même période que le cycle de la température. Nous avions déjà évoqué un phénomène similaire en examinant les changements de la composition du peuplement (Carpentier et al., 1997).

Le fait qu'il faille se placer à grande échelle pour percevoir une évolution de la diversité peut s'expliquer par la forte hétérogénéité de la structure spatiale du peuplement dans la zone du suivi. Ainsi, Luczak (1996) montre que les variations temporelles du peuplement observées à petite échelle dans le cadre du suivi à long terme sont comparables à celles qui existent à une échelle de quelques km<sup>2</sup>. A l'image des systèmes emboîtés (Morin, 1977 ; Allen et Starr, 1982), les modifications temporelles du peuplement sont donc plus rapides à petite

195

échelle. Aux deux échelles, celle de la station et celle du km<sup>2</sup>, les compositions spécifiques sont similaires. Il y a donc homogénéité du point de vue qualitatif. Mais il y a hétérogénéité du point de vue quantitatif : le peuplement présente une structure quantitative en mosaïque, se traduisant par l'existence de différents micro-faciés. Ces variations à petite échelle d'espace dans la zone sud de la Mer du Nord sont à rapprocher de la forte dynamique des structures sédimentaires (e.g. Beck et al., 1991, Vicaire, 1991 et le rapport Dyscop en annexe I), qui est liée elle-même à la variabilité des conditions hydrodynamiques à cette petite échelle. On peut aussi penser que cette organisation du peuplement en mosaïque mouvante (Price et al., 1984 ; Clark, 1991 ; Reise, 91) est due à de nombreux facteurs (Menge et Olson, 1990) : non seulement à la mobilité des sédiments, mais aussi à une répartition inégale des ressources alimentaires, qui dépend également des conditions hydrodynamiques. Des facteurs biotiques peuvent également agir : inégalité du recrutement et de la sédentarisation, pressions de prédation, compétition pour la ressource. L'organisation en mosaïque et sa "mouvance" sont des facteurs reconnus de persistance des écosystèmes (Levin et Paine, 1974 ; Whittaker et Levin, 1977; Caswell et Cohen, 1991). Le fort degré d'hétérogénéité de ce peuplement, par rapport à des peuplement similaires en Manche, peut constituer un moyen d'assurer sa stabilité dans le contexte de stress important de ce secteur.

On arrive parfaitement à détecter, sur cette série temporelle, les changements à mésoéchelle, car on dispose alors de suffisamment de "recul" pour s'affranchir des effets de cette mouvance du peuplement. De même, notre investigation à différents degré de résolution temporels ne met peut-être pas seulement en évidence une hiérarchie dans les niveaux d'organisation temporelle du peuplement, mais reflète aussi en partie la hiérarchie de son organisation spatiale. L'organisation en mosaïque peut en effet expliquer la forte variabilité perçue à petite échelle de temps : on observe des changements qui en réalité n'existent pas au niveau du peuplement (niveau d'organisation hiérarchique que le suivi est censé représenter) et qui ne sont dus qu'au fait que le peuplement peut se présenter sous des aspects différents. Par rapport à l'objectif 5 du PNOC-SLT (prédire les changements temporels des écosystèmes), on peut conclure à une prédictabilité de l'évolution à long terme (méso-échelle) du peuplement en fonction des variations climatiques, mais à une imprédictibilité à petite échelle de temps et d'espace, due à l'organisation en mosaïque.

Ces résultats ont aussi des implications importantes sur l'échantillonnage. D'abord, il conviendrait de mieux connaître la micro-répartition spatiale des organismes à petite échelle : il existe trop d'incertitude sur l'estimation des effectifs de certaines espèces à l'échelle du suivi à long terme. Face à ces phénomènes d'agrégation spatiale, l'optimisation de l'échantillonnage

devrait passer, comme le souligne Luczak (1996), par un suivi à plusieurs échelles : de l'échelle spatiale utilisée pour le suivi à long terme (de l'ordre de 50 000 m<sup>2</sup>) à l'échelle du km<sup>2</sup> au moins. En adoptant cette stratégie d'observation à plusieurs échelles, il serait également plus aisé (grâce aux observations à l'échelle du km<sup>2</sup>) de mettre en évidence les variations à court terme (prédictibilité à petite échelle de temps) de l'évolution du peuplement : par exemple des variations saisonnières – ce qui n'a pu être fait à l'échelle du suivi à long terme tel qu'il a été pratiqué jusqu'à maintenant.

Il serait également souhaitable de mettre en évidence les effets d'événements brutaux et apériodiques, tels que les tempêtes, d'abord sur la modification à petite échelle temporelle du peuplement, mais aussi sur son organisation (hétérogénéité) spatiale. Dans ce but, on pourrait coupler des observations faunistiques à haute fréquence avec des mesures physiques et des données issues de la modélisation hydrosédimentaire en particulier.

Dans le traitement des différents jeux de données abordés dans ce travail, la "régularisation" des séries, qu'elles soient temporelles ou spatiales, s'est avérée une étape essentielle avant l'application de méthodes destinées à mettre en évidence les tendances et les grandes structures sous-jacentes. Tout d'abord, un certain nombre de méthodes statistiques. surtout dans le domaine des séries chronologiques, s'appuient sur des observations réalisées à intervalles réguliers. De même, dans le domaine spatial, la réalisation d'estimations sur un maillage régulier facilite une valorisation cartographique. Cependant, les contraintes logistiques ou de terrain, qui empêchent la réalisation de certaines observations, sont une source d'irrégularités dans les séries qu'il semble difficile d'éliminer a priori. De plus. l'utilisation d'un plan de prélèvement "raisonné" (donc à maille inégale), destiné à mieux apprécier l'hétérogénéité de certaines périodes ou de certains secteurs géographiques, conduit à un déséquilibre de la densité de l'échantillonnage. Dans l'application de techniques d'analyses statistiques qui s'appuient sur l'ensemble des données, comme les méthodes d'ordinations et de régressions, le déséquilibre de l'échantillonnage peut induire une surreprésentation des zones ou périodes les plus fortement échantillonnées. Il est donc nécessaire, non seulement de régulariser, mais aussi de rééquilibrer a posteriori l'échantillonnage dans le temps ou l'espace, de manière à ce que toutes les zones (ou périodes) soient également représentées et aient le même poids dans les analyses effectuées.

Nous avons pu également mettre en évidence, dans l'évaluation de la diversité par l'indice de Shannon, l'influence de l'échelle d'observation (plus précisément du degré de résolution ou grain) sur la perception des structures et des tendances. L'exploitation des données spatiales du GDR Manche, a permis de constater que la structure morphologique et granulométrique des bancs de sable, de la côte vers le large (direction est-ouest) gouvernait, à petite échelle, les variations de l'indice de diversité. Quand on se place à grande échelle, cette structure côte-large s'estompe et laisse la place à un gradient sud-nord de diminution de la diversité qui peut être reliée au gradient hydrodynamique à l'approche du détroit du Pas-de-Calais. Il a été possible de dégager de ces observations un opérateur de changement d'échelle de la diversité, qui est différent selon le degré d'hétérogénéité du secteur concerné. La série temporelle de Gravelines, fortement bruitée et sans tendance particulière à petite échelle (échelle des prélèvements), fait apparaître, quand le degré de résolution dépasse l'échelle annuelle, une nette tendance à l'augmentation de la diversité.

Il convient néanmoins de nuancer les résultats avancés en terme de diversité. Les résultats présentés au chapitre IV ont permis de mieux apprécier le problème de l'évaluation de la diversité, et surtout de sa dépendance vis à vis de l'effort d'échantillonnage. L'effort d'échantillonnage ne doit pas être envisagé seulement en fonction de l'aire prospectée, mais aussi en terme d'effectifs collectés. Dans le cas d'un échantillonnage à surface constante, comme celui qui est pratiqué pour la série benthique de Gravelines, une augmentation de la densité des organismes peut conduire à observer une richesse spécifique plus élevée, uniquement parce qu'en collectant plus d'individus, il y a une plus forte probabilité d'observer des espèces rares, et non pas parce qu'il y a une augmentation réelle de la richesse du peuplement. Les simulations réalisées à partir de distributions théoriques ont permis de montrer que la richesse est un paramètre très difficile, voire impossible à estimer dans le cadre d'un échantillonnage contraint par les réalités de terrain. Les estimateurs non-paramétriques, Jackknife et Bootstrap, qui offrent pourtant la meilleure solution pour obtenir une estimation de la richesse vraie, ne sont vraiment performants que lorsque les échantillons sont déjà presque exhaustifs. En revanche, l'évaluation de l'indice de diversité de Shannon est moins exigeante et s'est révélée satisfaisante dans les conditions de l'échantillonnage pratiquées dans les différents cas concrets que nous avons abordé. De plus, la méthode d'extrapolation de H' à partir des Courbes Rang-Information Cumulées (CRIC) peut permettre, dans certains cas, d'améliorer l'estimation de H' lorsque l'échantillon est de petite taille.

En guise de conclusion finale, quelques perspectives de recherche peuvent être dégagées :

- Concernant les travaux sur les données de qualité des eaux du bassin versant, il serait intéressant de lier les séries physico-chimiques et les prévisions réalisées aux données biologiques, en particulier aux observations du phytoplancton mises en œuvre par Karine Noppe (thèse en cours au laboratoire) pour caractériser l'eutrophisation des cours d'eau. L'ensemble de ces données devrait également servir d'entrée dans un modèle de fonctionnement des écosystèmes littoraux ; les apports du bassin versant sont en effet relativement très importants dans un secteur du détroit où on a également mis en évidence (Brylinski *et al.*, 1991) un fleuve côtier, à savoir une masse d'eau littorale clairement individualisée.
- L'exploitation plus fouillée de la série benthique de Gravelines serait utilement complétée par une meilleure connaissance de l'incidence, à plusieurs échelles spatiales, d'événements climatiques brusques. La suggestion qui avait été faite, il y a quelques années, de réaliser une opération d'échantillonnage intensif du benthos et de mesures hydrodynamiques et sédimentaires au moment d'une tempête continue de se justifier. Un tel échantillonnage devrait être réalisé suivant une stratégie à plusieurs échelles d'observation.
- Les premières exploitations par Laura Sanvicente-Añorve (1995) des fichiers constitués dans le cadre du GDR Manche avaient été limitées par les capacités des moyens informatiques de l'époque et n'avaient pu permettre d'explorer la répartition des peuplements à de forts degrés de résolution (maillage fin). De plus, les travaux sont encore limités à la partie est de la zone concernée (Manche est sud Mer du Nord). Compte tenu des moyens informatiques actuellement disponibles et des algorithmes de traitement qui ont été améliorés, on pourrait envisager le traitement du fichier GDR complet (de la mer d'Iroise au sud de la Mer du Nord), ce qui permettrait de travailler finement à des échelles spatiales assez considérables. On pourrait aussi envisager de coupler ces données benthiques avec les données de la modélisation hydrosédimentaire existant à ces échelles. Une approche plus fonctionnelle (fondée par exemple sur les groupes trophiques) est envisageable grâce à la base de données qui a été développée.
- Les capacités des micro-ordinateurs donnent maintenant accès à des simulations particulièrement intéressantes pour étudier les propriétés des indices de diversité. Dans ce travail, nos simulations se sont limitées à des tirages dans des "urnes" où les individus de la communauté avaient des probabilités égales d'être observées. A ce tirage dans une urne, on

pourrait substituer un tirage dans des espaces simulés, dans lesquels les organismes ont été placés en tenant compte de leur tendance à l'agrégation et des interactions existant entre espèces. Cette approche est actuellement initiée par David Mouillot (thèse à l'université de Corte, mais dirigée par le laboratoire). La poursuite de tels travaux serait tout à fait enrichissante pour évaluer l'effet des modalités de l'échantillonnage (effort, type de plan, échelle utilisée) sur la perception des paramètres écologiques (densité, diversité, etc...) et, en retour, apporterait des informations pour l'optimisation de futures stratégies d'échantillonnage. Bibliographie

# **Bibliographie**

Allen T.F.H., Starr T.W., 1982. *Hierarchy : perspectives for ecological complexity*. University of Chicago Press, Chicago.

Alley W.M., 1988. Using exogeneous variables in testing for monotonic trends in hydrologic time series. *Water Resources Research*, 24 : 1955-1961.

Adams J.E., McCune E.D., 1979. Application of the generalized jackknife to Shannon's measure of information used as an index of diversity. In J.F. Grassle, G.P. Patil, W. Smith et C. Taillie (eds.), *Ecological diversity in theory and practice*. International Co-operative Publishing House, Fairland (Maryland). p. 117-131.

Arrhenius O., 1921. Species and area. Journal of Ecology, 9: 95-99.

Arrhenius O., 1923. Statistical investigation in the constitution of plant associations. *Ecology*, 4:68-73.

Baltanas A., 1992. On the use of some methods for the estimation of species richness. *Oikos*, 65 : 484-492.

Beck C., Clabaut P., Dewez S., Vicaire O., Chamley H., Augris C., Hoslin R., Caillot A., 1991. Sand bodies and sand transport paths at the English Channel – North Sea border : morphology, hydrodynamics and radioactive tracing. *Oceanologica Acta*, 11 : 111-122.

Belgrano A., Vincx M., Dewarumez J.M., Richard A., Craeymeersch J., and Heip C., 1990. Recruitment of meroplanktonic larvae in the Southern Bight of the North Sea. *Océanis* 16, 225-243.

Benzecri J.P., 1973. L'analyse des données. Tome 2: l'analyse des correspondances. Dunod, Paris. 632p.

Bersier L.F., Sugihara G., 1997. Species abundance patterns : the problem of testing stochastic models. *Journal of Animal Ecology*, 66 : 769-774.

Blondel J., 1986. Biogéographie évolutive. Masson, Paris. 221 p.

Bodo B.A., 1989. Robust graphical methods for diagnosing trend in irregularly spaced water quality time series. *Environmental Monitoring and Assessment*, 12: 407-428.

Box G.E.P., Jenkins G.M., 1976. Time series analysis : forecasting and control. Holden-Day, San-Francisco.

Bowman K.O., 1971. Comments on the distribution of indices of diversity. In G.P. Patil, E.C. Pielou, W. Waters (eds.), *Statistical Ecology*. . International Co-operative Publishing House, Fairland (Maryland). p. 315-359.

Bray R.J., Curtis J.T., 1957. An ordination of the upland forest communities of southern Wisconsin. *Ecological Monograph* 27, 325-349.

Brylinski J.M., Lagadeuc Y., Gentilhomme V., Dupont J.P., Lafite R., Dupeuble P.A., Huault M.F., Auger Y., Puskaric E., Wartel M., Cabioch L., 1991. Le "fleuve côtier" : un phénomène hydrologique important en Manche orientale. Exemple du Pas-de-Calais. *Oceanologica Acta*, sp. 11 : 197-203/

Cabioch L., Gentil F., 1975. Distribution des peuplements benthiques de la partie orientale de la Baie de Seine. Comptes-Rendus de l'Académie des Sciences de Paris - Sciences de la vie, 280 : 571-574.

Cabioch L., Glaçon R., 1975. Distribution des peuplements benthiques en Manche orientale, de la Baie de Somme au Pas-de-Calais. Comptes-Rendus de l'Académie des Sciences de Paris - Sciences de la vie, 280 : 491-494.

Cabioch L., Glaçon R., 1977. Distribution des peuplements benthiques en Manche orientale, du Cap d'Antifer à la Baie de Somme. Comptes-Rendus de l'Académie des Sciences de Paris - Sciences de la vie, 285 : 209-212.

Cabioch L., Gentil F., Glaçon R., Retière C., 1977. Le macrobenthos des fonds meubles de la Manche : distribution générale et écologie. In : B.F. Keegan, P.O. Ceidigh et P.J. Caston eds. Biology of benthic organisms. Pergamon Press, Oxford. p. 115-128

Carpentier P., Leprêtre A., Prygiel J., 1996a. Evolution et prévision des concentrations en phosphates dans les eaux superficielles en Artois-Picardie. Intérêt des réseaux de surveillance dans la gestion de la qualité des eaux. *In* : *Rencontres hydrologiques Franco-Roumaines*. J.P. Carbonnel, P. Serban, P. Hubert et H. Bendjoudi (Ed.), UNESCO, Paris, p. 317-332.

Carpentier P., Leprêtre A., Prygiel J., 1996b. Apports en azote et phosphore des cours d'eau du bassin Artois-Picardie au littoral Nord Pas-de-Calais et impact sur l'eutrophisation côtière. 3ème Congrès International Limnologie-Océanographie "Interfaces", Nantes, Octobre 1996.

Carpentier P., Dewarumez J.M., Leprêtre A., 1997. Long term variability of the Abra alba community in the southern bight of the North Sea. *Oceanologica Acta*, 20: 283-290.

Caswell H., Cohen J.E., 1991. Communities in patchy environments : a model of disturbance, competition and heterogeneity. In : J. Kolasa, S.T.A. Pickett, eds. *Ecological heterogeneity*. Springer Verlag, New-York. p. 97-122.

Chamley H., 1987. Sédimentologie. Dunod, Paris. 175p.

Chao A., 1984. Non-parametric estimation of the number of classes in a population. *Scandinavian Journal of Statistic*, 11: 265-670.

Clark J.S., 1991. Disturbance and population structure on the shifting mosaic landscape. *Ecology*, 72: 1119-1137.

Cleveland W.S., 1979. Robust locally weighted regression and smoothing scatterplots. Journal of the American Statistical Association, 74: 829-836.

Cleveland W.S., Devlin S.J., 1988. Locally weighted regression : an approach to regression analysis by local fitting. *Journal of the American Statistical Association*, 83 : 596-610.

Cleveland R. B., Cleveland W. S., McRae J. E., Terpening I., 1990. STL : a seasonal-trend decomposition procedure based on loess. *Journal of Official Statistics*, 6 : 3-73.

Cleveland W.S., Grosse E., 1991. Computational methods for local regression. *Statistics and Computing*, 1: 47-62.

Cleveland W.S., Grosse E., Shyu W.M., 1991. Local regression models. In Chambers J.M. and Hastie T., ed., *Statistical Models in S.* Wadsworth, Pacific Grove, California : 309-376.

Cluis D.A., 1983. Visual techniques for the detection of water quality: Double-mass curves and Cusum functions. *Environmental Monitoring and Assessment*, 3: 173-184.

Cohen A.C.J., 1959. Simplified estimators for the normal distribution when samples are singly censored or truncated. *Technometrics*, 1: 217-237.

Cohen A.C.J., 1961. Tables for maximum likelihood estimates : singly truncated or singly censored sample. *Technometrics*, 3 : 535-541.

Colwell R.K., Coddington J.A., 1995. Estimating terrestrial biodiversity through extrapolation. In D.L. Hawksworth (ed.), *Biodiversity : measurement and estimation*. The Royal society, Chapman & Hall, London. p. 101-118.

Daget J., 1976. Les modèles mathématiques en écologie. Masson, Paris. 172 p.

Dauvin J.C., Dewarumez J.M., Elkaim B., Bernardo J.M., Fromentin J.M., Ibanez F., 1993. Cinétique de Abra alba (mollusque-bivalve) de 1977 à 1991 en Manche-Mer du Nord, relation avec les facteurs climatiques. *Oceanologica Acta*, 16 : 413-422.

Dauvin J.C., Ibanez F., 1995. PNOC. Opération séries à long terme. Rapport PNOC, septembre 1995. 25 p.

Davison A.C., Hinkley D.V., 1997. Bootstrap methods and their application. Cambridge University Press, Cambridge.

De Forest E.L., 1873. On some methods of interpolation applicable to the graduation of irregular series. Annual Report of the Board of Regents of the Smithsonian Institution for 1871 : 275-339.

Dewarumez J.M., Quisthoudt C., Richard A., 1986. Suivi pluriannuel du peuplement à Abra alba dans la partie méridionale de la Mer du Nord (région de Dunkerque-France). Hydrobiologia, 142 : 187-197.

Dewarumez J.M., Davoult D., Frontier S., 1991. Examples of responses of benthic communities to environmental stress (Strait of Dover, France). *Oceanologica Acta* 11, 191-196.
Dewarumez J.M., Davoult D., Sanvicente-Anorve L.E., Frontier S., 1992. Is the "muddy heterogeneous sediment assemblage" an ecotone between the peebles community and the *Abra alba* community in the southern bight of the North Sea ? *Netherland Journal of Sea Research*, 30: 229-238.

Dewarumez J.M., Belgrano A., Craeymeersch J., Duquesne S., Heip C., Hilde D., Vincx M., 1993. Influence de la circulation des masses d'eaux dans la dynamique du peuplement à Abra alba de la baie sud de la Mer du Nord. *Journal de la Recherche Océanographique*, 18 : 1-4.

Diaconis P., Efron B., 1983. Computer-intensive methods in statistics. Scientific American, 248: 116-130.

Dietz E.J., Killeen T.J., 1981. A nonparametric multivariate test for monotone trend with pharmaceutical applications. *Journal of the American Statistical Association*, 76: 169-174.

Domenges D., Volle M., 1979. Analyse factorielle sphérique : une exploration. Annales de l'INSEE, 35 : 3-84.

Efron B., Thisted R., 1976. Estimating the number of unseen species : how many words did Shakespeare know ? *Biometrika*, 63 : 35-41.

Efron B., 1979. Bootstrap methods : another look at the Jackknife. Annals of Statistics, 7 : 1-26.

Efron B., 1982. The Jackknife, the Bootstrap, and other resampling plans. Society of Industrial and Applied Mathematics, Philadelphia.

Efron B., Tibshirani R., 1993. An introduction to the Bootstrap. Chapman and Hall, New-York.

El-Shaarawi A.H., Niculescu S.P., 1992. On Kendall's Tau as a test of trend in time series data. *Environmetrics*, 3: 385-411.

Escoufier Y., 1975. Le positionnement multidimensionnel. Revue de Statistique Appliquée, 23 (4) : 5-14.

Fisher R.A., Corbet A.S., Williams C.B., 1943. The relationship between the number of species and the number of individuals in a random sample of animal population. *Journal of Animal Ecology*, 12: 42-58.

Fromentin J.M., Ibanez F., 1994. Year-to-year changes in meteorological features of the french coast during the last half-century. Examples of two biological responses. *Oceanologica Acta* 17, 285-296.

Fromentin J.M., 1995. Evolution à long terme de communautés zooplanctoniques et macrozoobenthiques. Relation avec les changements climatiques. Thèse de doctorat, Université Paris VI. 170 p.

Fromentin J.M., Ibanez F., Dauvin J.C., Dewarumez J.M., Elkaim B., 1997a. Long-term changes of four macrobenthic assemblages from 1978 to 1992. *Journal of Marine Biological*. *Association U.K.* 77, 287-310.

Fromentin J.M., Dauvin J.C., Ibanez F., Dewarumez J.M., Elkaim B., 1997b. Long-term variations of four macrobenthic community structures. *Oceanologica Acta*, 20: 43-53.

Frontier S., 1977. Réflexions pour une théorie des écosystèmes. *Bulletin d'Ecologie*, 8 : 445-464.

Frontier S., 1985. Diversity and structure in aquatic ecosystems. Oceanographic and Marine Biology Annual Review, 23: 253-312.

Frontier S., 1987. Applications of fractal theory to ecology. In P. et L. Legendre eds. Developments in numerical ecology. Springer Verlag, Berlin Heidelberg. p. 335-378.

Frontier S., Pichot-Viale D., 1995. *Ecosystèmes. Structure, fonctionnement, évolution.* Masson, Paris. 2<sup>ème</sup> ed. 447 p.

Gaston K.J., 1996. Biodiversity : a biology of numbers and difference. Blackwell Science, Oxford. 396 p.

Gleason H.A., 1922. On the relation between species and area. *Ecology*, 3 : 156-162.

Good I.J., 1953. The population frequencies of species and the estimation of population parameters. *Biometrika*, 40: 237-264.

Gower J.C., 1966. Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika*, 53: 325-338.

Gray J.S., Mirza F.B., 1979. A possible method for detecting pollution induced disturbance on marine benthic communities. *Marine Pollution Bulletin*, 10: 142-146.

Gray J.S., 1981a. The ecology of marine sediments. Cambridge University Press, Cambridge. 185p.

Gray J.S., 1981b. Detecting pollution induced changes in communities using the log-normal distribution of individuals among species. *Marine Pollution Bulletin*, 12: 173-176.

Gray J.S., Pearson T.H., 1982. Objective selection of sensitive species indicative of pollutioninduced change in benthic communities. I. Comparative methodology. *Marine Ecology Progress Series*. 9, 111-119.

Gray J.S., Christie H., 1983. Predicting long term changes in marine benthic communities .*Marine Ecology Progress Series.* 13, 87-94.

Härdle W., 1990. Smoothing techniques : with implementation in S. Springer-Verlag. 261p.

Heiberger R.M., Becker R.A., 1992. Design of an S function for robust regression using iteratively reweighted least squares. *Journal of Computational and Graphical Statistics*, 1: 181-196.

Heltshe J.F., Forrester N.E., 1983a. Estimating species richness using the Jackknife procedure. *Biometrics*, 39: 1-11.

Heltshe J.F., Forrester N.E., 1983b. Estimating diversity using the quadrat sampling. *Biometrics*, 39: 1073-1076.

Henderson R., 1916. Note on graduation by adjusted average. *Transaction of the Actuarial Society of America*, 17: 43-48.

Hill M.O., 1973. Diversity and evenness : a unifying notion and its consequences. *Ecology*, 54 : 427-432.

Hill M.O., 1973. Reciprocal averaging: an eigenvector method for ordination. Journal of Ecology 61, 237-24.

Hirsch R.M., Slack J.R., Smith R.A., 1982. Techniques of trend analysis for monthly water quality data. *Water Resources Research*, 18: 107-121.

Hirsh R.M., Slack J.R., 1984. A nonparametric trend test for seasonal data with serial dependence. *Water Resources Research*, 20: 727-732.

Hirsh R.M., Alexander R.B., Smith R.A., 1991. Selection of methods for the detection and estimation of trends in water quality. *Water Resources Research*, 27: 803-813.

Holme N.A., Mc Intyre A.D., 1984. Method for the study of marine benthos. 2<sup>ème</sup> éd. Blackwell, Oxford. 387 p.

Huber P.J., 1981. Robust Statistics. John Wiley and Sons, New-York.

Hugues J.P., Millard S.P., 1988. A tau-like test for trend in the presence of multiple censoring points. *Water Resources Bulletin*, 24 : 521-531.

Hurlbert S.H., 1971. The non-concept of species diversity : a critique and alternative parameters. *Ecology*, 52 : 577-586.

Hutchinson G.E., 1957. A treatise on limnology. Volume I: Geography, physiscs and chemistry. John Wiley & Sons, New-York. 1015 p.

Hutchinson G.E., 1967. A treatise on limnology. Volume II: Introduction to lake biology and the limnoplankton. John Wiley & Sons, New-York. 1115 p.

Hawksworth D.L., 1995. *Biodiversity : measurement and estimation*. The Royal society, Chapman & Hall, London. 140 p.

Ibanez F., 1991. Treatment of data deriving from the Coast 147 project on coastal benthic ecology : the within-site analysis. In : Space and time series data analysis in benthic ecology, B. Keegan, ed. : 5-43.

Ibanez F., Fromentin J.M., Castel, J. 1993. Application de la méthode des sommes cumulées à l'analyse des séries chronologiques océanographiques. Comptes-Rendus de l'Académie des Sciences de Paris, Sciences de la vie, 316 : 745-748.

Ibanez F., Dauvin J.C., Etienne M., 1993. Comparaison des évolutions à long terme (1977-1990) de deux peuplements macrobenthiques de la Baie de Morlaix (Manche occidentale): relations avec les facteurs hydrodynamiques. J. Exp. Mar. Biol. Ecol. 164, 181-214.

Jongman R.H.G., Ter Braak C.J.F., Van Tongeren O.F.R., 1995. Data analysis in community and landscape ecology. 2<sup>nd</sup> ed., Cambridge University Press, Cambridge. 299p.

Kaufman L., Rousseeuw P.J., 1990. Finding groups in data : an introduction to cluster analysis. John Wiley & Sons, New-York. 342 p.

Karakassis I., 1995. S $\infty$ : a new method for calculating macrobenthic species richness. *Marine Ecology Progress Series*, 120: 299-303.

Kendall M.G., 1938. A new measure of rank correlation. *Biometrika*, 30:81-93.

Kendall M.G., 1975. Rank correlation methods. Charles Griffin, London, 4th ed.

Korn E.L., 1984. Kendall's tau with a blocking variable. *Biometrics*, 40: 209-214.

Kruskal J.B., 1964. Nonmetric multidimensional scaling : a numerical method. *Psychometrika*, 29 : 115-129.

Laboratoire National d'Hydraulique 1971. Nouvel avant-port de Dunkerque. Mesures en nature. Rapport n°3 HC 042/05.

Lambshead P.J.D., Platt H.M., Shaw K.M., 1983. The detection of differences among assemblages of marine benthic species based on an assessment of dominance and diversity. *Journal of Natural History*, 17: 859-874.

Larsonneur C., Bouysse P., Auffret J.P., 1982. The superficial sediments of the English Channel and its Western approaches. *Sedimentology*, 29: 851-864.

Lauga J., Joachim J., 1987. L'échantillonnage des populations d'oiseaux par la méthode des E.F.P. : intérêt d'une étude mathématique de la courbe de richesse cumulée. Acta Oecologica Oecologia Generalis, 8 : 117-124.

Lebart L., Morineau A., Fénelon J.P., 1982. Traitement des données statistiques. Méthodes et programmes. Dunod, Paris. 510p.

Legendre L., Legendre P., 1984. *Ecologie numérique*. Tome 1 et 2. Masson, Paris et Presses de l'Université Laval, Québec. 260p. et 360p.

Leprêtre A., 1986. Impact des tranchées forestières liées aux lignes H.T. et T.H.T. sur les invertébrés dans les Ardennes. A.M.B.E./E.D.F.-C.E.R.T., Bruay-sur-Escaut (Nord). 133p.

Leprêtre A., Carpentier P., 1994. Etude statistique des séries chronologiques obtenues dans le cadre du RNB en Artois-Picardie pour les paramètres azote et phosphore. Rapport USTL-Ecologie Numérique/Agence de l'Eau Artois-Picardie. 800 p.

Leprêtre A., 1997. Hétérogénéité, variabilité et diversité dans les écosystèmes. Concepts, outils, applications. Habilitation à Diriger des Recherches, Université des Sciences et Technologies de Lille. 261 p.

Leprêtre A., Carpentier P., 1997a. Etude des séries chronologiques obtenues dans le cadre du RNB en Artois-Picardie pour les paramètres azote, phosphore, oxygène et matières en suspension. Rapport USTL-Ecologie Numérique/Agence de l'Eau Artois-Picardie/Adrinord. 950 p.

Leprêtre A., Carpentier P., 1997b. Une méthode simple de prévision des tendances appliquée aux séries temporelles de qualité des eaux courantes. *Comptes-Rendus de l'Académie des Sciences de Paris - Sciences de la vie*, 320 : 407-411

Leprêtre A., Carpentier P., Chapalain G., 1997. Répartition spatiale et évolution temporelle des granulométries des sédiments dans un secteur côtier. Développement d'une méthode d'amélioration des estimations statistiques. Rapport recherche Programme Dyscop Région Nord Pas-de-Calais. 27p & annexes.

Lettenmaier D.P., 1976. Detection of trends in water quality data from records with dependend observations. *Water Resources Research*, 12 (5): 1037-1046.

Levin S.A., Paine R.T., 1974. Disturbance, patch formation and community structure. *Proc. Nat. Acad. Sci. USA*, 71: 2744-2747.

Luczak C., Dewarumez J.M., Essink K., 1993. First record of the american jackknife clam *Ensis directus* on the French coast of the North Sea. *Journal of Marine Biological.* Association U.K. 73, 233-235.

Luczak C., 1996. Dynamique du recrutement au sein du peuplement à Abra alba dans la baie sud de la Mer du Nord : approche systémique. Thèse de doctorat, Université de Lille 1, 144p.

MacArthur R.H., 1957. On the relative abundance of bird species. Proceedings of the National Academy of Sciences of the USA, 43: 293-295.

Magnussen S., Boyle T.J.B., 1995. Estimating sample size for inference about the Shannon-Weaver and Simpson indices of species diversity. *Forest Ecology and Management*, 78 : 71-84.

Magurran A.E., 1988. Ecological diversity and its measurement. Princeton University Press, Princeton (New-Jersey). 179 p.

Mandelbrot B., 1953. Contribution à la théorie mathématique des communications. Thèse Université Paris, Publ. Inst. Stat. Univ. Paris., 2: 121p.

Mann H.B., 1945. Nonparametric tests against trend. Econometrica, 13: 245-259.

Margalef R., 1956. Information y diversidad especifica en las communidades de organismos. *Investigacion pesq.*, 3 : 99-106.

Margalef R., 1958. Information theory in ecology. General Systems, 3: 36-71.

May R.M., 1975. Patterns of species abundance and diversity. In M.L. Cody et J.M. Diamond (eds.), *Ecology and evolution of communities*, The Belknap Press of Harvard University Press, Cambridge. p. 81-120.

McIntosh R.P., 1967. An index of diversity and the relation of certain concepts to diversity. *Ecology*, 48 392-404.

Menge B.A., Olson A.M., 1990. Role of scale and environmental factors in regulation of community structure. *Trends in Ecology & Evolution*, 5: 52-57.

Meybeck M., 1985. Variabilité dans le temps de la composition chimique des rivières et de leurs transports en solution et en suspension. Sciences de l'Eau, 4 : 93-121.

Menhinick E.F., 1964. A comparison of some species-individuals diversity indices applied to samples of field insects. *Ecology*, 45 : 859-861.

Morin E., 1977. La méthode. 1. La nature de la nature. Editions du Seuil, Paris. 399p.

Motomura I., 1932. On the statistical treatment of communitites. Japanese Journal of Zoology, 44: 379-398. (en japonais).

Mouillot D., 1994. Analyse des écosystèmes par la méthode des diagrammes rangsfréquences (DRF) : problèmes de traitements numériques. Diplôme d'Etudes Approfondies, Université de Paris VI et VII. 15p.

Noppe K., Prygiel J., Coste M., Carpentier P., Leprêtre A., 1997. Impact de la station d'épuration de Douai dur le phytoplancton et sur la qualité des eaux de la Scarpe aval. *CILEF* 5, Namur, Juillet 1997.

O.C.D.E. 1982. Eutrophisation des eaux. Méthodes de surveillance, d'évaluation et de lutte. O.C.D.E., Paris. 157p.

Page E.S., 1961. Cumulative sum charts. *Technometrics*, 3 : 1-9.

Palmer M.W., 1990. The estimation of species richness by extrapolation. *Ecology*, 71 : 1195-1198.

Palmer M.W., 1991. Estimating species richness : the second-order Jackknife reconsidered. *Ecology*, 72 : 1512-1513.

Pielou E.C., 1975. Ecological diversity. John Wiley and Sons, New-York.

Pourriot R., Meybeck M., 1995. Limnologie générale. Masson, Paris. 956 p.

Preston F.W., 1948. The commonness and rarity of species. *Ecology*, 29: 254-283.

Preston F.W., 1962. The canonical distribution of commonness and rarity. *Ecology*, 43 : 185-215, 410-432.

Price P.W., Slobodchikoff C.N., Gaud W.S., 1984. A new ecology. Novel approaches to interactive systems. John Wiley & sons, New-York. 515p.

Prygiel J., Davoult D., Dewarumez J.M., Glaçon R., Richard A., 1988. Description et richesse des peuplements benthiques de la partie française de la Mer du Nord. *Comptes-Rendus de l'Académie des Sciences de Paris, série III* 306, 5-10.

Quenouille M.H., 1956. Notes on bias in estimation. *Biometrika*, 43: 353-360.

Reise K., 1991. Mosaic-cycles in the marine benthos. In : H. Remmert, ed. *The mosaic-cycle concept of ecosystems*. Springer Verlag, Heidelberg. p. 61-82.

Rousseeuw P.J., Leroy A.M., 1987. Robust regression and outlier detection. John Wiley and Sons, New-York.

Sanders H.L., 1968. Marine benthic diversity : a comparative study. *American Naturalist*, 102 : 243-282.

Sanvicente-Añorve L.E., 1995. Détermination des structures benthiques spatiales en Manche orientale au moyen de méthodes d'analyse multivariable et de techniques d'interpolation. Thèse de doctorat, Université de Paris VI. 169p. + annexes.

Sanvicente-Añorve L.E., Leprêtre A., 1995. Typologie des stations océanographiques en Manche orientale : comparaison des méthodes d'interpolation spatiale. *Journal de la Recherche Océanographique*, 20 : 27-32.

Sanvicente-Añorve L.E., Leprêtre A., Davoult D., 1996. Large-scale spatial pattern of the macrobenthic diversity in the Eastern English Channel. *Journal of Marine Biological*. *Association U.K.*, 76: 153-160.

Sen P.K., 1968. Estimates of the regression coefficient based on Kendall's tau. Journal of the American Statistical Association, 63: 1379-1389.

Shannon C.E., Weaver W., 1949. The mathematical theory of communication. University of Illinois Press, Urbana. 117p.

Service Hydrographique et Océanographique de la Marine 1968. Courants de marée dans la Manche et sur les côtes françaises de l'Atlantique. S.H.O.M., Paris. 287p.

Simberloff D., 1972. Properties of the rarefaction diversity measurement. American Naturalist, 106:414-418.

Simberloff D., 1978. Use of rarefaction and related methods in ecology. In K.L. Dickson, J. Cairns, Jr., and R.J. Livingston (eds.), *Biological data in water pollution assessment : quantitative and statistical analyses*, American Society for Testing and Materials STP 652, Philadelphia. p. 150-165.

Simberloff D., 1979. Rarefaction as a distribution-free method of expressing and estimating diversity. In J.F. Grassle, G.P. Patil, W. Smith, and C. Taillie (eds.), *Ecological diversity in theory and practice*, p. 159-176. International Cooperative Publishing House, Fairland, Md.

Slocomb J., Dickson K.L., 1978. Estimating the total number of species in a biological community. In K.L. Dickson, J. Cairns, Jr., and R.J. Livingston (eds.), Biological data in water pollution assessment : quantitative and statistical analyses, American Society for Testing and Materials STP 652, Philadelphia. p. 38-52.

Smaoui H., 1996. Modélisation numérique tridimensionnelle de l'hydrodynamique et des transports sédimentaires en Manche orientale et dans le sud de la Mer du Nord. Thèse de doctorat, Université des Sciences et Technologies de Lille. 160p. + annexes.

Smith E.P., Van Belle G., 1984. Nonparametric estimation of species richness. *Biometrics*, 40 : 119-129.

Smith E.P., Rose K.A., 1991. Trend detection in the presence of covariates : stagewise versus multiple regression. *Environmetrics*, 2 : 153-168.

Sokal R.R., Michener C.D., 1958. A statistical method for evaluating systematic relationships. *Univ. Kansas Sci. Bull.*, 38 : 1409-1438.

Souplet A., 1978. Le domaine benthique. In : *Etude écologique de projet du site de Graveline*. CNEXO-IBMRW. 237p.

Souplet A., Glaçon R., Dewarumez J.M., Smigielski F., 1980. Distribution des peuplements benthiques littoraux Mer du Nord du Cap Blanc-Nez à la frontière belge. Comptes-Rendus de l'Académie des Sciences de Paris, série III 20, 627-630.

Spearman C., 1904. The proof and measurement of association between two things. *Journal of Psychology*, 15: 88-97.

Spencer J., 1904. On the graduation of the rates of sickness and mortality. Journal of the Institute of Actuaries, 38: 334-347.

Statistical Sciences 1995. S-PLUS Guide to statistical and mathematical analysis, version 3.3. StatSci, a division of MathSoft, Inc., Seattle.

Stenseth N.C., 1979. Where have all the species gone ? On the nature of extinction and the Red Queen hypothesis. *Oikos*, 33 : 196-227.

Stout J., Vandermeer J., 1975. Comparison of species richness for stream-inhabiting insects in tropical an mid-latitude streams. *American Naturalist*, 109 : 263-280.

Sugihara G., 1980. Minimal community structure : an explanation of species abundance patterns. *American Naturalist*, 116 : 770-787.

Taylor J.M.G., 1987. Kendall's and Spearman's correlation coefficients in the presence of a blocking variable. *Biometrics*, 43 : 409-416.

Teissier G., 1948. La relation d'allométrie : sa signification statistique et biologique. *Biometrics*, 4 : 14-53.

Tokeshi M., 1990. Niche apportionment or random assortment : species abundance patterns revisited. *Journal of Animal Ecology*, 59 : 1129-1146.

Tokeshi M., 1993. Species abundance patterns and community structure. Advances in Ecological Research, 24: 111-186.

Torgerson W.S., 1958. Theory and methods of scaling. John Wiley & Sons. London.

Tukey J., 1958. Bias and confidence in not quite large samples. Annals of Mathematical Statistics, 29:614.

Ugland K.I., Gray J.S., 1982. Lognormal distributions and the concept of community equilibrium. *Oikos*, 39: 171-178.

Uthermöhl H., 1958. Zur Vervollkommnung der quantitativen Phytoplankton-Methodik. *Mitt. Internat. Verein. Limnol.*, 9 : 1-38.

Van Belle G., Hugues J.P., 1984. Nonparametric tests for trend in water quality. *Water Resources Research*, 20: 127-136.

Van Valen L., 1973. A new evolutionary law. Evolution Theory, 1:1-30.

Van Valen L., 1976. The Red Queen lives. Nature, 260: 575.

Van Valen L., 1977. The Red Queen. American Naturalist, 111: 809-810.

Vaslet D., Larsonneur C., Auffret J.P., 1978. Carte des sédiments de la Manche au 1/500 000. BRGM, Orléans.

Vicaire O., 1991. Dynamique hydro-sédimentaire en Mer du Nord méridionale (du Cap Blanc-Nez à la frontière belge). Thèse de doctorat, Université des Sciences et Technologies de Lille. 264p.

Warwick R.M., Clarke K.R., 1991. A comparison of some methods for analysing changes in benthic community structure. J. Mar. Biol. Ass. U.K. 71, 225-244.

Whittaker R.H., 1972. Evolution and measurement of species diversity. Taxon, 21: 213-251.

Whittaker R.H., Levin S.A., 1977. The role of mosaic phenomena in natural communities. *Theoretical Population Biology*, 12: 117-139.

Wilson E.O., Peter F.M., 1988. Biodiversity. National Academy Press, Washington D.C.

.

Zahl J.H., 1977. Jacknifing an index of diversity. *Ecology*, 58: 907-913.

.

Zetterqvist L., 1988. Asymptotic distribution of Mann's test for trend for m-dependent seasonal observations. *Scandinavian Journal of Statistics*, 15: 81-95.

# **ANNEXE I**

# Rapport de recherche DYSCOP

Répartition spatiale et évolution temporelle des distributions granulométriques des sédiments dans un secteur côtier.

Développement d'une méthode d'amélioration

des estimations statistiques.

Alain Leprêtre, Patrice Carpentier et Georges Chapalain.

# UNIVERSITE DES SCIENCES & TECHNOLOGIES DE LILLE

## 1997

# Rapport de recherche

# **DYSCOP** Région Nord Pas-de-Calais

Répartition spatiale et évolution temporelle des distributions granulométriques des sédiments dans un secteur côtier. Développement d'une méthode d'amélioration des estimations statistiques.

Travail réalisé au Laboratoire d'Ecologie Numérique par:

Alain LEPRETRE \*, Patrice CARPENTIER \*& Georges CHAPALAIN\*\*

appliqué à des données recueillies par Bernadette TESSIER\*\*

USTL. Laboratoire d'Ecologie Numérique. URA 1363/ERS 395. Bât.SN3.

\*\* USTL. Laboratoire de Sédimentologie & Géodynamique. URA 719. Bât.SN5

Ce travail a été réalisé dans le cadre du programme DYSCOP Région Nord-Pas-de-Calais par Alain LEPRETRE et Patrice CARPENTIER (Laboratoire d'Ecologie Numérique -URA 1363), en collaboration étroite avec Georges CHAPALAIN (URA 719) en vue d'une articulation ultérieure avec la modélisation hydrosédimentaire développée par Georges CHAPALAIN & Hassan SMAOUI.

Le développement a été effectué sur des données granulométriques de l'atelier « Wissant », transmises par Bernadette TESSIER (URA 719).

Une version étoffée des interprétations données par les sédimentologues et océanographes physiciens pourra être réalisée, mais la présente opération, proposée initialement par Georges CHAPALAIN et Alain LEPRETRE et intégrée à la demande de support du programme "Modélisation hydrosédimentaire à court et à moyen termes", ne concernait que les aspects statistiques développés dans ce rapport.

Une version condensée concernant spécifiquement la méthodologie développée en matière d'estimation et de répartition spatiale des distributions granulométriques est préparée pour soumission à publication dans « Journal of Coastal Research » sous l'intitulé « Estimating and mapping grain-size distributions and related parameters of sea-bottom sediments in a coastal area » par A.Leprêtre, P.Carpentier & G.Chapalain.

#### 1. INTRODUCTION. PROBLEMATIQUE.

Sous les influences conjuguées de la houle, des courants et des apports continentaux, les sédiments des zones côtières présentent du point de vue granulométrique une hétérogénéité spatiale souvent marquée, doublée d'une variabilité temporelle en relation avec les processus de transport sédimentaire. La caractérisation spatiale de la distribution granulométrique des sédiments nécessite la réalisation de campagnes de prélèvements le plus souvent réalisées suivant un plan d'échantillonnage régulier. Les conditions météorologiques et techniques ne permettent pas toujours d'appliquer exactement le plan d'échantillonnage prévu, conduisant soit à un échantillonnage incomplet, soit à une modification de la localisation des prélèvements par rapport aux campagnes précédentes. Il est donc nécessaire de disposer de méthodes d'interpolation spatiale permettant l'estimation des caractéristiques du sédiment en des points non échantillonnés en vue de la cartographie de ces propriétés et de leur comparaison dans le temps.

Il existe de nombreuses techniques d'interpolation (splines, krigeage, ...) permettant de procéder à ces cartographies (Burgess & Webster, 1980a,b; Volz & Webster, 1990; Cressie, 1993; Brus *et al.*, 1996; Theocharopoulos *et al.*, 1997). Elles s'adressent en général à un seule variable: une variable brute (concentration en un élément par exemple) ou un paramètre donné (par exemple granulométrie moyenne ou médiane). Ainsi, la distribution granulométrique d'un sédiment, décrite par les proportions observées pour différentes classes de diamètre de particules, est une variable multiple dont l'interpolation spatiale est réalisée suivant deux schémas: soit par interpolations séparées de chacune des classes granulométriques, soit par interpolation únique d'un paramètre statistique de distribution: un moment statistique tel que la moyenne, ou un percentile tel que la médiane ou autre centile, décile ou quartile.

La méthodologie proposée dans ce travail a pour objectif l'estimation des distributions granulométriques (donc de la multivariable) en tous points d'un maillage, permettant ensuite, d'une part la cartographie de n'importe quel paramètre statistique (moyenne, percentile ou indice de

- 1 -

dispersion), et d'autre part l'obtention en chaque point des proportions de chaque classe granulométrique, valeurs requises par exemple en entrée des modèles physiques de transport multiclasses.

La méthode d'estimation proposée est réalisée sur l'ensemble des classes prises simultanément, prenant en compte la corrélation existant entre les différentes classes granulométriques et évite le biais dû au fait que ces distributions granulométriques soient des distributions fermées (somme égale à 100%), les différentes proportions n'étant donc pas statistiquement indépendantes (Butler, 1979). Cette estimation, faite à l'aide d'une méthode d'analyse multivariable couplée à une méthode de krigeage, permet en outre un filtrage des données et surtout assure la conservation des limites spatiales franches (« sharp boundaries ») entre types sédimentairès, qui sont artificiellement lissées par les méthodes classiques (Brus et al., 1996).

La méthode développée a été appliquée à des données granulométriques recueillies au cours de deux campagnes de prélèvement effectuées dans le site-atelier de la baie de Wissant en juillet 1991 et janvier 1992 (figure 1).

### 2. <u>METHODOLOGIE</u>

#### • 2.1. Principe général de la méthode d'estimation des distributions granulométriques

Dans un espace géographique donné, on dispose de n prélèvements de sédiments. Chacun de ces prélèvements donne une distribution granulométrique, à savoir une série de proportions correspondant aux différentes classes granulométriques obtenues sur une colonne de tamis. Dans le tableau de données constitué, on appellera donc 'objets' les prélèvements (i = 1 à n) et 'variables' les différentes classes granulométriques (j = 1 à k) exprimées en %.



Figure 1: Localisation de la zone-atelier DYSCOP et plans de prélèvement des campagnes de juillet 1991 et de janvier 1992

On a donc, pour un prélèvement *i* donné (*i* = 1 à n), des proportions  $p_{i1}$ ,  $p_{i2}$ ,

.....
$$p_{ij}$$
,.... $p_{ik}$ , avec  $\sum_{j=1}^{k} p_{i,j} = 1$ .

On cherche à estimer ces proportions en chaque noeud d'un maillage de cette zone, en vue d'une cartographie par exemple. On doit donc faire une estimation, pour chaque point i' (i'=1) à N, des proportions  $p_{i'1}$ ,  $p_{i'2}$ , ...., $p_{i'j}$ , ...., $p_{i'k}$ .

Une première méthode consiste à estimer séparément, variable par variable  $(j = 1 \ a \ k)$ , les  $p_{i'j}$  par interpolation à partir des  $p_{ij}$  observés. Cette méthode, communément utilisée, ne tient pas compte de la relation existant entre les différentes catégories granulométriques, qui d'une part sont liées par leur somme égale à 1, et d'autre part sont corrélées. En effet, l'existence d'un type de grains grossiers peut favoriser un type de grains fins qui va se placer dans les interstices qu'il a générés; les deux types de grains s'en trouvent alors corrélés. Par conséquent, de telles interpolations séparées des différentes variables ne sont pas satisfaisantes.

La méthode proposée ici repose sur le traitement simultané de l'ensemble des variables par analyse factorielle. Les différentes catégories granulométriques n'étant pas des variables indépendantes, puisqu'on a des distributions fermées (la somme des proportions est égale à 1), le tableau "relevés X classes granulométriques" ne devrait pas être soumis tel quel, aux analyses en composantes principales ou ACP (ce qui se fait pourtant souvent...). Il est néanmoins nécessaire de rappeler quelques éléments sur l'ACP avant de présenter la méthode d'analyse factorielle retenue.

En ACP, la métrique utilisée est une métrique euclidienne. La distance entre les objets est calculée par la distance euclidienne dans l'espace des k variables décrivant ces objets:

Soit k variables indépendantes,  $x_1, x_2, \dots, x_j, \dots, x_k$ , chaque objet *i* prend les valeurs  $x_{i,1}$ ,  $x_{i2}, \dots, x_{ij}, \dots, x_{ik}$  pour ces différentes variables. L'objet peut donc être représenté dans un espace euclidien à k dimensions (orthogonales), grâce à ses coordonnées  $x_{ij}$ . La distance euclidienne entre

2 objets *i* et *l*, est donnée simplement par : 
$$d_{i,l} = \sqrt{\sum_{j=1}^{k} (x_{i,j} - x_{l,j})^2}$$
.

On procède souvent en ACP, pour donner une même importance *a priori* à chacune des variables, à un centrage-réduction des valeurs de chaque variable, ou quelquefois à un simple centrage si l'objectif se limite à donner une origine (un zéro) commune à tous les vecteurs variables impliqués.

Pour une variable  $x_i$  donnée, le centrage-réduction transforme les valeurs  $x_{i,j}$  en valeurs

 $X_{i,j}$ , avec:  $X_{i,j} = \frac{x_{i,j} - \overline{x_j}}{s_j}$ ,  $\overline{x_j}$  étant la moyenne observée sur l'ensemble des objets pour la

variable  $x_j$ , et  $s_j$  étant l'estimation de l'écart-type.

L'ACP appliquée alors à ces données transformées est dite ACP centrée réduite. Par cette transformation, les vecteurs variables ont tous la même norme 1.

Dans le cas des distributions fermées, avec des variables "proportions" non indépendantes telles que les  $p_j$ , la perte d'un degré de liberté rend l'ACP inapplicable en toute rigueur, car les points-objets ne sont pas véritablement dans un espace à k dimensions, mais sur l'hyperplan à k-1 dimensions coupant chacun des axes à la valeur 1.

L'ACP a cependant des propriétés intéressantes, en particulier la propriété de reconstitution (e.g. Lebart et al., 1982). Il est en effet possible, à partir des coordonnées des pointsobjets sur les différents axes (k axes) de l'ACP et des valeurs propres correspondant à ces axes factoriels, de reconstituer parfaitement le tableau de données de départ. Cette propriété peut être mise à profit pour "filtrer" les données, à savoir négliger dans une reconstitution les k-h derniers axes, c'est-à-dire ceux dont la contribution est la plus faible (par construction). La méthode employée ici, proposée initialement par Gonzalez (1979) pour une analyse comparative de proportions d'espèces animales (benthiques), consiste à profiter des avantages offerts par l'ACP, en évitant l'écueil de la non indépendance des proportions. Il est proposé de travailler avec une métrique de Hellinger (Matusita, 1955; LeCam, 1970), où:

$$d_{i,l} = \sqrt{\sum_{j=1}^{k} \left(\sqrt{p_{ij}} - \sqrt{p_{lj}}\right)^2}$$
, qui revient à réaliser une distance sur les racines-carrées

des proportions. Cette distance peut aller de 0 à  $\sqrt{2}$ 

Gonzalez (1981) montre que l'on peut travailler dans cette métrique en utilisant un programme d'ACP; il suffit de réaliser une ACP centrée non réduite, appliquée aux racines-carrées des proportions. Cette méthode avait été proposée préalablement dans un contexte très différent (analyse de données économiques) par Domenges et Volle (1979) sous le nom d'analyse factorielle sphérique (qu'on notera AFS).

Dans ce cas, les points-objets ne sont plus répartis dans un espace euclidien quelconque,

mais situés à la surface d'une hypersphère de rayon 1, puisque  $\sum_{j=1}^{k} \left( \sqrt{p_{i,j}} \right)^2 = \sum_{j=1}^{k} p_{i,j} = 1$ .

Comme les proportions sont positives par construction, les points sont situés en réalité sur un seul orthant de l'hypersphère. La distance de Hellinger entre deux points-objets correspond à la longueur de la corde joignant ces deux points.

On peut donc appliquer une "ACP Hellinger" ou AFS au tableau de données "prélèvements X classes granulométriques", et obtenir une ordination des points-objets (à savoir ici des prélèvements). Chaque prélèvement a un jeu de coordonnées correspondant aux différents axes de l'analyse factorielle.

Le but recherché ici n'est pas uniquement de filtrer les données mais de "fabriquer" des données aux points du maillage en vue d'une cartographie ou d'une injection dans un modèle hydrosédimentaire. Or, chaque axe factoriel correspond à une combinaison multivariable caractérisée par un vecteur propre, et ces vecteurs propres sont indépendants par construction (principe même des analyses factorielles). Le nuage de points-prélèvements est donc décrit, après analyse factorielle, par k multivariables indépendantes. Il est donc maintenant permis de procéder de manière séparée à des calculs sur chacune de ces variables.

L'objectif étant de trouver des granulométries aux points d'un maillage, il suffit de calculer pour chacune des multivariables, la valeur prise en ces "pseudo-points" d'observation. Cette estimation est faite par une procédure géostatistique, en l'occurrence le krigeage.

Pour un axe factoriel j donné, les coordonnées  $C_{ij}$  des points i (i = 1 a n) sur cet axe servent à établir le ou les variogrammes qui sont à la base de l'interpolation. Sur certains axes, on constate en effet une anisotropie spatiale qui requiert d'utiliser des variogrammes directionnels.

Pour chaque axe j, après krigeage, on obtient donc N valeurs  $C_{i'j}$  correspondant aux "pseudo-points" du maillage i'  $(i'=l \ \alpha \ N)$ , à savoir un tableau "N pseudo-prélèvements X k multivariables", qu'il suffit de "reconstituer" à l'aide des valeurs propres de l'analyse initiale. On limite la reconstitution à p = k-h axes, de manière à filtrer le "bruit de fond" exprimé par les axes factoriels mineurs. Les valeurs reconstituées sont recentrées pour obtenir des  $\sqrt{p_{i'j}}$  qu'il suffit ensuite de mettre au carré pour obtenir les proportions dans chaque classe granulométrique pour les N pseudo-points i'.

Cette reconstitution amène néanmoins à de légers *arte fact* du fait, en particulier du filtrage. On obtient en effet des valeurs légèrement négatives pour certains  $\sqrt{p_{i'j}}$ . Un algorithme a dû être mis au point pour rectifier les profils granulométriques présentant ces valeurs négatives, avant la mise au carré.

Les algorithmes trouvés en bibliographie pour de tels redressements (Bartnicki, 1989, en particulier) sont peu respectueux de la forme des profils et s'intéressent essentiellement à la conservation des masses, car généralement appliqués à des problèmes de diffusion numérique.

- 7 -

Notre objectif étant au contraire de préserver au maximum la forme du profil granulométrique, et non la masse (puisqu'il s'agit de pourcentages, qu'on peut toujours ramener à une somme de 1), nous avons utilisé l'algorithme suivant.

Sur le profil de k valeurs d'un point-objet i présentant des valeurs négatives, on cherche d'abord la plus forte des valeurs négatives, autrement dit la moins négative parmi les négatives. Soit  $x_{iM}$  cette valeur.

On calcule la valeur absolue de la différence  $|x_{ij} - x_{iM}|, \forall j$  puis la somme  $\sum_{j=1}^{k} |x_{ij} - x_{iM}| \text{ de ces valeurs absolues des différences. On ajoute ensuite à toutes les valeurs <math>x_{ij}$  la

valeur  $|x_{iM}| \times \frac{|x_{ij} - x_{iM}|}{\sum_{i=1}^{k} |x_{ij} - x_{iM}|}$ , ce qui revient à répartir en positif sur toutes les classes j = 1 à k, et

proportionnellement au poids de chacune des classes, la partie négative initiale.

Chaque valeur  $x_{ij}$  prend donc une nouvelle valeur  $x'_{ij} = x_{ij} + |x_{iM}| \times \frac{|x_{ij} - x_{iM}|}{\sum_{j=1}^{k} |x_{ij} - x_{iM}|}$ , et

la classe initialement négative est amenée à  $x_{iM} = 0$ .

A la phase suivante, la même procédure est appliquée s'il y a encore des valeurs négatives, et ainsi de suite. On arrête le processus quand tous les  $x_{ij}$  sont  $\geq 0$ .

Peu de points ont présenté cet *arte fact* "valeurs négatives", et dans ces cas, les valeurs ne l'étaient que peu, comme le montre l'exemple donné à la figure 2.

<u>Remarque pratique</u>: les prélèvements effectués sur placages ou affleurements rocheux ne sont pas écartés du calcul, mais codés à une proportion de 100% pour une classe granulométrique "fictive" plus élevée que la plus grande des classes granulométriques véritablement observées. De même, les placages vaseux sont codés 100% pour la classe la plus fine.



Figure 2: Recalage d'un profil présentant des valeurs négatives

#### • 2.2. Mise en évidence de modifications temporelles des distributions et des

#### paramètres

Outre celui de procurer des valeurs à injecter dans un modèle hydrosédimentaire ou de pouvoir réaliser une cartographie, l'intérêt d'obtenir, pour chaque campagne de prélèvement une estimation de la distribution granulométrique en chaque point d'un maillage, est de pouvoir comparer ces distributions d'une période à une autre, ou d'une année à l'autre, en vue de localiser les zones où se sont produits des remaniements du sédiment (apports et exportations). Cette comparaison peut être faite sur les valeurs obtenues pour le même paramètre de distribution  $\theta$  aux temps  $t_i$  et  $t_{i+1}$ , et on pourra alors cartographier par exemple, les valeurs d'une différence  $d_{i+1,i}$ correspondant à  $\theta_{i+1} - \theta_i$ . Une autre manière de procéder consiste à comparer les distributions complètes, et non directement les paramètres statistiques, aux temps  $t_i$  et  $t_{i+1}$ , par le calcul d'une distance entre les profils granulométriques du même point aux temps  $t_i$  et  $t_{i+1}$ , Cependant, le calcul d'une distance entre profils par un  $\chi^2$  classique n'est pas permise dans la mesure où il s'agit de distributions fermées. Nous suggérons donc d'effectuer cette comparaison à l'aide de la distance de Hellinger et de cartographier les distances de Hellinger obtenues.

### 3. <u>SITE D'ETUDE ET DONNEES</u>

Le jeu de données utilisé correspond à deux campagnes de prélèvement réalisés avec une benne Van Veen dans la baie de Wissant (zone-atelier DYSCOP: rectangle indiqué à la figure 1).

Ces campagnes ont été menées en juillet 1991 et janvier 1992. Les prélèvements ont été faits suivant un plan d'échantillonnage "approximativement" régulier: la figure 1 (1b et 1c) donne la répartition des stations et la numérotation employée dans les campagnes. La position géographique 'exacte' des prélèvements est connue (positionnement GPS), et enregistrée en vue du traitement géostatistique. L'effort d'échantillonnage (nombre de prélèvements) n'a pas été identique dans les deux campagnes: en particulier, la zone nord (voir à droite de la figure 1c) n'a pas été échantillonnée lors de la deuxième campagne.

La distribution en 27 classes granulométriques a été obtenue sur tamis conformes à la norme AFNOR (de raison  $10^{1/10} = 1.2589$ ). Les données correspondantes sont jointes en annexe I.

La baie de Wissant, située entre le Cap Gris-Nez et le Cap Blanc-Nez est connue comme présentant une forte instabilité des dépôts littoraux (Clabaut & Chamley, 1986). La partie de la zone d'échantillonnage située près de la côte est caractérisée par des sables, tandis que la partie la plus éloignée, proche du banc à la ligne est beaucoup plus hétérogène spatialement et comporte des "faciès sédimentaires" plus variés (voir figure 3, extraite de Clabaut, 1988)

- 10 -



Figure 3: Hétérogénéité spatiale des structures sédimentaires dans la zone-atelier Dyscop de Wissant (extrait de Clabaut, 1988)

L'objectif n'étant ici que méthodologique - à savoir trouver une méthode capable de faire les meilleures interpolations dans le cadre d'un suivi et sans *a priori*, - les informations existant dans la littérature sur la structure spatiale des sédiments dans cette zone ne sont pas détaillées (*cf.* Clabaut, *op. cit.*, Clabaut & Chamley, *op. cit.*). Il est néanmoins intéressant de savoir au départ que des structures nettement délimitées devraient être révélées au large, tandis qu'au niveau de la côte on doit s'attendre à la présence de gradients dans les sables.

La description des distributions granulométriques n'est pas non plus l'objectif de ce travail. Un choix de distributions observées a été représenté sous forme de polygones de fréquences, et placé en annexe II. Ces distributions sont représentées en échelle  $\phi$ , échelle communément utilisée depuis Krumbein (1934), et telle que  $\phi = -\log_2$  (diamètre en mm). Les classes de granulométrie fine sont donc situées sur la droite des graphiques.

### 4. <u>RESULTATS</u>

Les données obtenues, représentées en annexe II par des polygones de fréquences en échelle  $\phi$ , montrent une grande variété<sup>i</sup> de distributions granulométriques: tantôt symétriques, tantôt asymétriques, souvent bimodales. Les tableaux de données 'prélèvements X proportions', pour chacune des deux campagnes, sont maintenant soumis à l'AFS.

#### • 4.1. Analyse factorielle des jeux de données des 2 campagnes.

Pour l'AFS appliquée à la campagne de juillet 1991, le plan 1X2 des deux premiers axes factoriels (figure 4) oppose globalement les stations situées au large et au nord, qui ont des granulométries plus grossières, aux stations caractérisées par les sables, situées plus près de la côte et protégées par le Cap Gris-Nez.



 Figure 4: Projection des points de prélèvements (en haut: 4a) et des vecteurs-variables (en bas: 4b) dans le plan des 2 premiers axes de l'Analyse Factorielle Sphérique du tableau de données de la campagne de juillet 1991

Cette séparation en deux groupes géographiques est visualisée par la cartographie des coordonnées des stations sur l'axe 1 (figure 5a). En 4b, on voit sur le plan de projection des vecteurs-variables, cette opposition entre les classes à grains fins (à gauche) et les classes à granulométrie grossière (à droite). Le deuxième axe factoriel sépare les stations à granulométrie très grossière (> à 16-20mm) ou situées sur des placages rocheux (en bas sur la figure 4a: stations F3, F9, E11, *etc.*) aux autres stations.



Figure 5: Cartographie des coordonnées factorielles des points estimés aux noeuds du maillage: Axe 1 en juillet 1991 (fig.5a) - Axe 1 en janvier 1992 (fig. 5b) - Axe 2 en janvier 1992 (fig. 5c)



Figure 6 (6a & 6b): Projection des points de prélèvements (en haut: 6a) et des vecteurs-variables (en bas: 6b) dans le plan des 2 premiers axes de l'Analyse Factorielle Sphérique (AFS) du tableau de données de la campagne de janvier 1992



Figure 6 (6c & 6d): Projection des points de prélèvement dans le plan des axes 2 et 3 (en haut: 6c) et histogramme des valeurs propres de l'AFS du tableau de données de la campagne de janvier 1992 Les zones de substrat rocheux apparaissent très nettement sur l'axe 1 de l'AFS portant sur la deuxième campagne en janvier 1992 (figure 6a) formant un groupe compact (à droite sur le plan factoriel). Leur localisation est nette (en noir) sur la figure 5b, qui cartographie les coordonnées des stations sur cet axe 1. Les autres stations s'étalent le long de l'axe 2 et leur typologie est mieux exprimée sur le plan 2X3 (figure 6c). L'agencement de stations sur ce plan 2X3 ressemble à celui qui a été trouvé pour les stations dans le plan 1X2 de la première AFS. La cartographie des coordonnées des points suivant l'axe 2 (figure 5c) donne d'ailleurs une image semblable à celle qui avait été obtenue pour l'axe 1 dans la première AFS: opposition des zones sableuses proches de la côte et du Cap, aux zones à sédiment plus grossiers et hétérogènes situées au large et au nord.

### 4.2. Krigeage et "reconstitution" de pseudo-valeurs aux noeuds du maillage

Après un examen détaillé des plans factoriels successifs, 10 axes sont retenus en vue de la "reconstitution" des granulométrie des pseudo-points. Pour vérifier la validité d'une telle sélection, la reconstitution est appliquée aux n points vrais. On constate sur 2 exemples de janvier 1992 (figure 7); superposant le profil original et le profil reconstitué avec 10 axes, que le filtrage n'induit que des modifications très mineures des profils granulométriques, ce qui conforte dans le choix de ces 10 axes. L'histogramme des valeurs propres (variances) de l'AFS (figure 6d) montre d'ailleurs l'importance très faible des axes suivants. Un filtrage plus sévère, ne conservant que 5 axes (plus de 90% de la variance cumulée - cf. figure 6d), donne d'ailleurs des résultats encore très acceptables.

En ce qui concerne l'interpolation réalisée pour chacun des axes, pour estimer des coordonnées à chacun des pseudo-points du maillage, celle-ci est réalisée par krigeage. Dans un premier temps, les variogrammes sont établis pour chaque jeu de coordonnées d'un axe donné. On vérifie également la présence d'une éventuelle anisotropie spatiale de la variance, en réalisant des variogrammes directionnels. Chaque variogramme observé est ajusté à un modèle.

Janvier 1992











On constate que suivant les axes le type de modèle n'est pas le même: sur l'exemple donné à la figure 8 (variogrammes des 4 premiers axes factoriels de la campagne de juillet 1991), on a un modèle sphérique pour les trois premiers axes et un modèle exponentiel pour le quatrième. Chaque axe est donc porteur d'une structure différente de l'hétérogénéité: gradient ou coupures, aboutissant à des modèles différents , avec des tailles variables des structures élémentaires plus ou moins grandes, aboutissant à des portées différentes des variogrammes.

Sur l'ensemble des analyses géostatistiques réalisées dans ce travail, ce sont quasiment toujours les modèles sphériques et exponentiels qui présentent les meilleurs ajustements, constatation semblable à celle de Webster (1985) pour les sols. Les variogrammes des coordonnées factorielles situés au delà du cinquième axe sont isotropes et s'ajustent très généralement à des modèles sphériques à portée faible, ce qui montre qu'on détecte alors des microstructures.

La figure 8 donne les variogrammes des 4 premiers axes factoriels de la campagne de juillet 1991et les modèles ajustés pour chacun. L'axe 2 présente une anisotropie (orientation des axes de l'ellipse des variances à 70 et 160°); les paramètres du modèle sphérique ajusté ne sont donc pas les mêmes suivant les directions. Les autres axes ne révèlent pas d'anisotropie et donnent des variogrammes "omnidirectionnels".

Pour juger les performances de la méthode employée, nous avons comparé les cartographies obtenues pour différents paramètres granulométriques usuels (médiane ou  $D_{50}$ , médiane de Folk, coefficient de sorting,  $D_{35}$ ,  $D_{90}$ , etc.) obtenus suivant deux procédures:

- la procédure 1 qui vient d'être présentée: ce sont les distributions qui sont estimées grâce à la méthode couplant AFS, krigeage et reconstitution de pseudo-prélèvements; les paramètres "" synthétiques (statistiques) ne sont donc calculés qu'en fin de processus, au point par point (noeuds du maillage).

- la procédure 2, usuellement employée, consistant à calculer d'abord le paramètre de distribution sur les points vrais, pour appliquer ensuite une méthode géostatistique (spline ou krigeage)

- 19 -



Figure 8: Ajustements des variogrammes obtenus à partir des coordonnées des points observés sur les 4 premiers axes factoriels de la campagne de juillet

Les résultats obtenus pour la médiane (ou  $D_{50}$ ) sont présentés simultanément sur trois transects (figure 9).

- Sur le transect "ligne 8 du maillage", qui est parallèle et très proche de la série de points vrais de la ligne C du plan de prélèvement, les résultats sont donnés à la figure 9a. Les valeurs des points vrais sont positionnées. On constate que les deux profils de  $D_{50}$ , obtenus par les 2 méthodes,

. . .

se suivent et s'ajustent correctement aux points vrais. Cependant, la méthode 1 donne, à l'endroit des deux ruptures de profil liées au passage de zones de cailloux, un profil plus tranché, montrant sa plus forte capacité à mettre en évidence des lisières franches.

- Cette constatation est plus nette sur le deuxième exemple, correspondant à un transect confondu avec la ligne E de départ. Dans cet exemple, les deux méthodes donnent des résultats nettement différents. La méthode classique de krigeage direct des paramètres (méthode 2, en pointillés) donne un profil beaucoup moins tranché que la nouvelle méthode proposée. Cette incapacité de la méthode 2 est due à une diffusion (ou dilution) des valeurs du D<sub>30</sub>, L'interpolation sur paramètres a tendance à lisser exagérément les lisières franches en instaurant un passage progressif d'une faible à une forte valeur.

On donne en annexe III les profils granulométriques estimés sur cette ligne du maillage la plus proche de la ligne E de prélèvement, entre les pseudo-points 116 à 125 qui se succèdent parallèlement aux point initiaux E4 à E7. On y constate par exemple entre E6 et E7, qui correspondent sur la ligne de points estimés à 122 et 125, que les points intermédiaires 123 et 124 ne réalisent pas une déformation progressive entre deux profils très différents mais conservent, par leur différence marquée, le caractère tranché de la lisière franche existant entre deux sédiments très différents.

- Le troisième exemple est très caractéristique du lissage excessif créé par la méthode 1: il s'agit d'un transect "ligne 10 du maillage", très proche de la ligne D du plan de prélèvement, mais pas confondue avec elle. La méthode 2 épouse parfaitement les points vrais, tandis que l'autre méthode montre un profil présentant deux décrochements "suspects". On s'aperçoit qu'il s'agit d'une diffusion des valeurs très fortes du D<sub>50</sub> (ou faibles du  $\phi_{50}$ ) de points situés assez loin du transect.

En conclusion, la carte du  $D_{50}$  donnée par notre méthode (figure 10) respecte beaucoup plus les lisières franches, qui sont nombreuses dans la zone d'étude. Ce qui se voit dans une comparaison des deux cartes.



Figure 9: Comparaison des valeurs de D50 obtenues sur une ligne du maillage par les 2 méthodes d'interpolation avec les valeurs réellement mesurées sur un transect original proche Méthode 1 (directe) en pointillés; méthode 2 via couplage AFS-krigeage en trait continu; valeurs réellement observées (cercles)

- Fig. 9a: "ligne 8" du maillage proche du transect C (suivant fig. 1)
- Fig. 9b: "ligne 12", quasi-confondue avec le transect E (id.)
- Fig. 9c: "ligne 10", proche du transect D (id.)

Dans la carte obtenue avec la méthode 2, des taches à granulométrie grossière ( $D_{50} = 16$ mm) situées au centre de la zone, sont entourées de zones à granulométrie beaucoup plus fine ( $D_{50} = 0.50$ mm). Le passage d'une granulométrie à l'autre se réalise très rapidement. Alors qu'au même endroit, avec la méthode 1, le passage d'une granulométrie à l'autre est beaucoup plus graduel. La méthode ancienne a donc tendance à créer des gradients qui n'existent pas. La nouvelle méthode respecte les lisières franches, ce qui ne l'empêche pas de détecter les véritables gradients.

L'exemple a été donné ici pour le paramètre  $D_{50}$ . Des résultats similaires sont obtenus pour les autres paramètres de position (moyenne de Folk,  $D_{50}$ , etc.) ou pour des paramètres de dispersion (sorting de Trask, ...): voir par exemple les figures de l'annexe IV.

#### 4.3: Comparaison des résultats de juillet 1991 et de janvier 1992

Disposant donc d'une méthode d'estimation performante des distributions granulométriques aux noeuds d'un maillage, on peut décrire les changements qui se sont produits de juillet 1991 à janvier 1992. La figure 11 donne la cartographie de la distance de Hellinger calculée entre les profils de juillet et de janvier pris deux à deux. Seule est représentée la zone qui n'a pas demandé d'extrapolation importante. En effet, la deuxième campagne a donné lieu à une investigation de moindre étendue (en particulier vers le nord), et il serait peu prudent pour cette campagne de pousser des extrapolations à plusieurs kilomètres en dehors de la zone observée.

On constate que les changements les plus forts (distances les plus élevées) se sont produites au large et au nord de la zone représentée, alors que les points situés dans la baie de Wissant, plus proches de la côte et protégés par le Cap, montrent des changements moins marqués.

L'interprétation peut être abordée en considérant les sorties du modèle de transport sédimentaire extraites de Chapalain *et al.* (1993), repris en annexe V. Les zones de fort changement constatées à l'aide de la distance de Hellinger entre juillet et janvier, correspondent aux secteurs de transport les plus forts calculés sous l'action combinée des courants de marée et de la houle.




Figure 10: Comparaison des cartes de D50 obtenues avec les 2 méthodes d'estimation .:

Méthode par couplage AFS-krigeage (2, en haut) et méthode directe (1, en bas)



Figure 11: Cartographie de la distance de Hellinger entre les 2 cartes de granulométrie des campagnes de juillet 1991 et janvier 1992.

#### <u>CONCLUSION</u>

La conclusion concernant la répartition spatiale des sédiments et la modification des profils granulométriques entre deux campagnes reviendra au sédimentologue et à l'océanographe physicien. Mais l'objectif recherché ici ne concernait que l'optimisation de l'estimation des distributions granulométriques. Les résultats obtenus par la méthode couplant une analyse factorielle avec le krigeage, et utilisant des procédures de filtrage et de reconstitution, a donné des résultats qui sont plus conformes aux résultats attendus d'après les connaissances préalables de la zone, que les interpolations effectuées suivant une démarche plus classique consistant en une interpolation directe des paramètres.

Dans une phase ultérieure, ces calculs seront reconduits sur un ensemble de quatre campagnes réalisées dans cette même zone (ces deux campagnes et deux autres non présentées). La méthode, garante du respect des patrons spatiaux souvent complexes des réaprtition spatiale des types de sédiments en zone côtière, pourra également être généralisée à d'autres situations: plages par exemple. Une variante de la méthode a également été développée dans un contexte écologique, pour l'interpolation de distributions quantitatives d'espèces animales benthiques sur l'ensemble Manche Est - sud Mer du Nord.

#### <u>Références citées</u>

- Bartnicki, J. (1989). "A simple filtering procedure for removing negative values from numerical solutions of the advection equation." <u>Environmental Software</u> 4(4): 187-201.
- Brus, D. J., J. J. De Gruijter, et al. (1996). "The performance of spatial interpolation methods and chloropleth maps to estimate properties at points: a soil survey case study." <u>Environmetrics</u> 7: 1-16.
- Burgess, T. M. and R. Webster (1980). "Optimal interpolation and isarithmic mapping of soil properties. I. The semi-variogram and punctual kriging." Journal of Soil Science 31. 315-331.
- Burgess, T. M. and R. Webster (1980). "Optimal interpolation and isarithmic mapping of soil properties. II. Block kriging." Journal of Soil Science 31: 333-341.
- Butler, J. C. (1979). "Effects of closure on the measures of similarity between samples." <u>Mathematical Geology</u> 11(4): 431-440.
- Chapalain, G., H. Smaoui, et al. (1993). <u>Modélisation du transport sédimentaire sous l'action combinée des courants de marée et de la houle dans la partie sud du Détroit du Pas-de-Calais</u>. Actes du 4ème Congrès Français de Sédimentologie, Lille, 17-19 nov., AFS ed., Paris: 87-88.
- Clabaut, P. and H. Chamley (1986). "Observations sédimentologiqes sur le littoral de Wissant (Site des caps, Boulonnais)." <u>Ann. Soc. Géol. Nord</u> CV: 169-178.
- Clabaut, P. (1988). Dynamique sédimentaire dans le détroit du Pas-de-Calais. Thèse de Doctorat de l'Université de Lille 1. 251p.
- Cressie, N. A. C. (1993). Statistics for spatial data. New-York, John Wiley & Sons: 900p.
- Domenges, D. and M. Volle (1979). "Analyse factorielle sphérique: une exploration." <u>Annales</u> <u>de l'INSEE</u> 35: 3-84.
- Gonzalez, P. L. (1979). Contribution au choix de modèles de distributions d'abondances. Montpellier, Mémoire de DEA, Univ. Sc. Techn. Languedoc. 67p.
- Krumbein, W. C. (1934). "Size frequency distributions of sediments." Journal of Sedimentary Petrology 4(2): 65-77.
- Lebart, L., A. Morineau, et al. (1982). <u>Traitement des données statistiques</u>. Méthodes et programmes. Paris, Dunod-Bordas: 510p.

- LeCam, L. (1970). "On the assumptions used to prove asymptotic normality of maximum likelihood estimates." <u>The Annals of Mathematical Statistics</u> **41**(3): 802-828.
- Matusita (1955). "Decision rules based on the distance for the problem of fit two samples and estimation." <u>Ann. Inst. Statistic. math. Tokyo</u> 26
- Theocharopoulos, S. P., P. V. Petrakis, et al. (1997). "Multivariate analysis of soil grid data as soil classification and mapping tool: the case study of a homogeneous plain in Vagia, Viotia, Greece." <u>Geoderma</u> 77: 63-79.
- Voltz, M. and R. Webster (1990). "A comparison of kriging, cubic splines and classification for predicting soil properties from sample information." <u>Journal of Soil Science</u> 41: 473-490.
- Webster, R. (1985). "Quantitative spatial analysis of soil in the field." <u>Advances in Soil Science</u> 3: 1-70.

:

#### ANNEXE I

Données granulométriques des campagnes de juillet 1991 et de janvier 1992

ĸ

- . -

Juillet 1991

:

•

;

~

00         00           00         00           00         00           00         00           00         00           00         00           00         00           00         00           00         00           00         00           00         00           00         00           00         00           00         00           00         00
$ \begin{array}{c ccccccccccccccccccccccccccccccccccc$
00         00           00         00           00         00           00         00           00         00           00         00           00         00           00         00           00         00           00         00           00         00           00         00           00         00           00         00
00 00 00 00 00 00 00 00 00 00 00 00 00 00
00 00 00 00 00 00 00 00 00 00 00 00
$ \begin{array}{c ccccccccccccccccccccccccccccccccccc$
$ \begin{array}{c ccccccccccccccccccccccccccccccccccc$
00 00
0.2 40.2
00 00
00 00
00 00
00 00
0.0 0.0
00 00
00 00
00 00
00 00
2.6 0.0
00 19
1.5 74.9
4.1 6.5
36 360
00 00

Julliet 1991

:

• .

.

٠

•

٠

:

:

.

E5	0.0	0.0	0.1	0.7	1.1	1.5	3.0	4.1	3.1	1.6	1.2	1.0	1.2	21	2.4	3.1	3.9	5.8	7.5	9.0	8.8	7.9	10.4	6.6	13.9
E7	0.0	0.0	0.1	0.4	1.0	2.2	1.8	3.5	3.9	1.3	0.6	0.3	0.2	0.1	0.1	0.1	0.1	0.1	0.4	0.4	0.2	0.5	0.5	0.5	81.7
E8	0.0	0.1	0.3	1.6	2.2	4.2	4.1	20.7	30.4	8.1	3.5	1.6	1.1	1.0	0.7	0.8	1.5	1.8	0.9	2.5	4.3	2.0	2.5	4.0	0.0
E9	0.0	0.0	0.0	0.2	0.3	1.2	3.5	32.8	47.1	8.8	2.6	1.0	0.5	0.5	0.2	0.3	0.4	0.3	0.1	0.0	0.0	0.0	0.0	00	0.0
E10	0.1	0.1	0.3	1.4	2.8	6.6	5.8	20.8	19.3	9.2	7.7	5.3	4.4	4.3	2.9	2.1	1.5	1.9	0.8	1.3	0.8	0.6	0.0	0.0	0.0
E11	0.1	0.3	0.7	3.1	4.0	6.2	3.3	7.1	5.0	0.8	0.4	0.2	0.2	0.3	0.3	0.5	0.9	0.8	1.3	2.4	3.4	1.7	3.1	1.4	52.6
F4	0.0	0.0	0.0	0.0	0.1	0.2	0.9	13.6	34.3	16.3	9.8	5.6	4.1	3.7	2.2	1.8	1.9	1.5	1.3	1.8	0.9	0.0	0.0	0.0	0.0
F7	0.0	0.0	0.0	0.3	0.6	2.4	7.7	51.6	24.6	2.8	1.5	1.0	0.9	1.1	0.7	0.7	0.5	0.4	0.7	0.7	1.6	0.0	0.0	00	0.0
F8	0.0	0.0	0.1	0.7	1.6	5.8	31.4	43.5	11.8	2.2	0.9	0.4	0.2	0.1	0.1	0.2	0.1	0.6	0.3	0.0	0.0	0.0	0.0	0.0	0.0
B11	0.1	0.2	0.5	2.9	4.8	4.2	4.0	4.9	2.0	0.6	0.4	0.3	0.3	0.6	0.6	0.9	1.2	1.4	1.6	2.2	3.8	5.2	4.8	6.2	46.2
£5	0.0	0.0	0.1	0.7	1.1	1.5	3.0	4.1	3.1	1.6	1.2	1.0	1.2	2.1	2.4	3.1	3.9	5.8	7.5	9.0	8.8	7.9	10.4	6.6	13.9
£7	0.0	0.0	0.1	0.4	1.0	2.2	1.8	3.5	3.9	1.3	0.6	0.3	0.2	01	0.1	0.1	0.1	0.1	0.4	0.4	0.2	0.5	0.5	0.5	81.7

#### Janvler 1992

.•

•

.

1

Ech.01/92	0.050	0.063	0.080	0.100	0.125	0.160	0.200	0.250	0.315	0.400	0.500	0.630	0.800	1.000	1.250	1.600	2.000	2.500	3.150	4.000	5.000	6.300	8.000	10 000	12.500	16.000
A4	0.0	00	02	1.9	18.7	40.1	24.4	6.8	1.9	05	04	04	06	4.2	_00	_00_	_00_	00	00	0.0	00	_00	00	_00	00	00
A5	00	00	0.1	1.6	20.2	48.2	21.8	50	16	00	03	_02_	0.2	0.1	_00_	00	00	0.0	00	_00_	_00_	_00_	0.0	_00_	00	00
A6	0.0	0.0	02	1.3	16.3	46.8	26.2	5.7	1.7	04	02	_0.2	_02_	_0.1	0.3	_03_	_0.0	_00_	_00	_00_	00	00	_00_	_00_	_00_	_00_
A7	00	0.1	0.1	0.9	9.4	36.8	37.9	10.3	29	_07	0.3	0.2	10	01	_01_	_01_	_00	0.0	00_	00	00	_ <u>00</u> _	_00_	_00	_00_	_00_
84	00	00	0.1	0.2	3.0	148	36.8	28.5	10.7	24	1.1	0.5	_04	03	_04_	09	00	00	_00_	_00	_00_	_00_	_00_	_00_	00	_0.0
A9	0.0	0.0	00	04	4.0	17.9	30.5	23.8	168	42	1.2	04	_02_	02	_02_	0.4	_0.0_	_00_	_00	00	00	_00_	_00_	_00_	_00	_00_
A10	00	00	00	0.5	11.0	47.0	31.5	54	1.4	0.5	0.3	02	0.3	02	_0.3_	12	_00_	_00_	00_	00_	_00_	00	_00_	00	_00_	00
B3	00	00	00	0.3	3.2	14.5	27.3	27.3	<u>11.4</u>	2.4	0.9	05	_0.4_	_02	04	_02_	_0.1	_03_	_0.3	0.3	_03_	_00_	0.1	_00_	-09	<u>-8.7</u>
B4	00	0.0	00	0.7	11.7	45.9	32.6	68	14	0.3	02	_0.2	_01_	01	_00_	_00_	00_	_00_	_00_	_00_	_0.0_	_00_	_00_	_00_	_00	_00_
B5	00	0.1	_0.1_	<u> </u>	13.8	47.2	26.8	_70_	_2.8	0.7	0.3	_01	_0.1_	<u>_00</u>	_00_	00	_00_	_00	_00_	_00_		0.0	00	-00	$-\frac{00}{2}$	-00
86	00	0.0	0.2	1.5	18.5	52.2	21.6	3.5	1.2	04	0.2	0.1	0.1	_01_	0.1	_0.1_	_00_	_00_	_00_	_00_	00	$-\frac{00}{2}$	_00	00	-00-1	
B7	00	02	0.3	2.7	21.1	48.8	18.3	43	2.1	0.9	0.5	03	02	_0.1	0.1	0.1	_00	0.0	<u>_00</u>	00	$-\frac{00}{2}$	_00_	-00-	-0.0	-00-	_00_
88	00	0.2	0.2	1.8	17.8	43.3	20.6	7.4	_5.1	2.4	_08_	_0.1_	_0.1	_0.1_	01	0.1	_00_	-0.0	_00	_00_	-00-	_00_	-00	_00_	-00-	
89	00	1.9	0.2	1.6	17.9	42.8	20.4	58	4.3	_22_	08	_03_	_01_	_0.1	0.1	_1.5_	_00_			00	-00-1	-00	-00	-00-	-00	
B10	00	0.1	0.1	_08_	7.4	26.0	_29.1	12.3	9.4	5.3	30	<u> </u>	<u>    1.2    </u>	_06_	_07	28	_00	_00_	-0.0	00	_00_	-00	-00-		-00	_00_
811	00	00_	0.1	0.5	7.0	23.9	25.3	13.1	11.2	_51	3.6	2.1	1.8	_10_	<u>_]. </u>	_4.3_	00	_0.0_	-00	_00_	_00	-00	_00_			-00-
B12	00	0.2	0.2	1.2	6.5	12.4	7.8	80	24.4	20.4	8.9	2.1	1.2	_0.5_	0.7	5.5	00_	00	_00	_00_	_0.0_	<u>-00</u>	-00-1	-00	-00-	-00
814	4.9	0.3	0.3	1.2	8.5	24.2	242	<u>_11.7_</u>	62	2.3	<u>i.0</u>	04	0.5	_05_	_10_	.06_	0.8_			<u>_1.9</u>				-3/-	-20-	-00
B15	5.4	0.5	0.4	08_	3.1	12.7	17.4	9.5	11.8	55	1.8	0.6	_06	05	09	05	.0.9	0.9-	0.7	_09_	1.2		- 2.1	- 39	-33	-10.2
C0	00	0.1	0.1	08	9.1	25.9	32.3	194	76	_2.0_	<u>_!.!</u> _	06	04		_02_	_02	_0.0	_00		_00	-00-1	-00	_00_			
	00	00	0.1	0.5	6.9	25.7	37.1	18.1	_53_	10	_0.7	04	04	03	04	_3.1	00	00	$-\frac{0.0}{0.0}$	_00	.00	-00-	-00			
C2	00	00	00	0.6	9.3	47.5	35.2	5.8	12	0.2	0.1	_01_	_00_	_0.0_	_00_	-00_	.00.	00		_00		-00-1	-22-1		-00-1	
C3	00	00	0.0	0.8	14.5	50.9	29.6	3.2	_0.7_	_02_	0.1	01	00_	_00_	00	00	.00	. 00	-00	_0.0	-00	-00	-0.0	-00-	-00-1	-00-
C4	00	0.0	0.0	0.6	10.4	45.4	362	5.2	1.3	0.4	0.3	_0.2	0.2	_00_	_00	.00	00	.00	- 50	- 00-	-00-	-00-1	-00-	-00-		
C5	00	0.0	0.1	15	17.3	56.5	21.3	_20_	_06_	0.2	_0.1	02	02	01	- 00	_00	.00	<u>.00</u> .	-00-	-00-	-00-	-00-1			-00-1	-00-
C6	00	0.1	0.2	1.5	12.0	342	34.6	10.7	46	<u> </u>	05	_01_	_01_	<u>. 00</u> _	_00	-00	-00	-00			-00				-00-1	
C7	0.0	00_	00_	0.6	6.4	159	16.4	148	_20.0	<u></u>	7.8	34	2.1					-00	-00.		-00	-00-	-00-	-00-	-00	
C8	0.0	02	0.2	1.1	9.7	22.1	152	80	139	107	8.6	42	-3.2	-1.2		08		-00-		-00		-00	-00			-00-
C9	00	00	0.0	0.2	1.3	1.9	22_	5.4	180	16.9	155	9.4	85	-49	_52	_33		- 1.9				00	· 00-	00	- 00-	00
C10	00	00	01	06	_3.7	_5.2_	4.6_	95_	26.7	18.9	10.9	4.1	$-\frac{32}{2}$	19	25			09	1.1		140	-00	-00	00	00	
CII	0.0	00	0.0	_03_	1.8	_5.3_	. 86	12.7	190	12.2	8./	_3.5_		-21	- 20-		- 1 4.	_ <u>1,4</u>	-00	-00-	- 00	-00-	-00-	$-\frac{0}{00}$	-00-1	-00-
C13	00	00	0.0	0.3	1.8	40	4.0	10.8	_30.2	25.5	15.7	44	<u> </u>	05	_04	4	-:			-16	1.6		-20-	-30-	11	-217-
C14	00	0.9	0.1	06	2.4	32	2.1	_46_	154	135	7.6	-2.5	1.6	08	_1.2	-08						- 00			-00-1	- 00
DO	00	0.0	0.0	_0.2_	2.3	15.5	350	27.4	128	30		04	<u>_0.3</u>	02	-02-	. 10			- 00-		-00	- 00-1	- <u>00</u> - [	-00-	-00-1	
DI	00	00	_00_	0.3	3.7	19.4	439	23.4	_7.1_	14	-05-	-02-	-01-				00	00.		- 00-		-00-1		- 00-	-00-1	
D2	00	00	00	0.1	40	29.3	47.0	16.2	_2.5_	-0.5	$\frac{02}{02}$		- 01-	- 00			1.00		- 00-		-00-		-00			-00-1
03	00	00	_01	03	63	29.0	450	155	28	05	0.2	$\left  -\frac{0.1}{0.0} \right $		-0.1		- 00			- 00		- 00		-00-1	-00-		-00-1
D4	00	00	0.1	0.3	4.6	23.3	45.1	17.6	4.4	1.2	1.1	0.9		-02						-0.0	-00			-00-1	-00-1	-00-1
	0.0	0.0	00	01	07	1.4	1.3	1.7	13.1	238	24.9	12.8	<u> 91</u>	30	30	12	00		0.5		11.7	00	00	_00		

v

•

#### Janvier 1992

:

.

.

e

.

																							_			
Dó	0.0	0.0	0.0	0.1	0.4	0.8	1.2	3.1	21.9	33.4	23.9	7.6	3.8	1.2	0.9	1.7	0.0	0.0	00	0.0	0.0	0.0	0.0	0.0	0.0	00
D7	0.0	0.0	0.0	0.0	0.3	0.4	0.4	0.7	7.8	15.1	16.9	10.8	10.3	6.1	7.0	5.0	3.9	3.1	2.9	2.4	1.6	1.7	0.3	0.0	3.3	00
D8	0.0	0.0	_0.1	0.2	1.0	1.4	1.8	6.1	17.5	14.8	8.4	4.2	4.4	3.2	4.4	3.5	3.8	3.8	4.7	3.5	4.6	4.7	2.0	0.6	1.2	0.0
D9	0.0	0.0	0.0	0.4	1.2	1.6	1.9	5.9	17.4	17.3	12.5	6.4	6.1	3.9	5.1	3.6	3.6	3.4	3.5	2.3	4.0	0.0	0.0	0.0	00	00
D10	0.0	0.0	0.0	0.2	0.8	1.5	4.5	19.2	41.4	19.2	8.0	2.5	1.4	0.5	0.3	0.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	00	0.0	00
D11	0.0	0.0	0.0	0.3	1.4	3.1	5.9	16.7	35.6	20.1	8.4	2.7	1.9	0.9	0.9	2.1	0.0	00	0.0	0.0	0.0	00	0.0	00	0.0	0.0
D13	0.0	0.1	0.1	0.2	1.1	1.8	2.6	5.8	21.3	29.4	19.2	6.9	4.2	1.9	1.8	3.8	00	0.0	00	0.0	0.0	0.0	0.0	0.0	0.0	00
D14	0.0	0.0	0.0	0.1	0.3	0.5	0.8	1.8	12.4	26.9	27.0	12.2	7.8	3.2	2.9	4.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	00
D15	66.8	4.9	3.9	6.7	9.1	5.1	2.0	0.5	0.3	0.3	0.1	0.1	1.0	00	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	00	0.0
E1	0.0	0.0	0.0	0.1	1.5	6.3	9.0	6.7	4.1	3.0	3.6	3.0	- 4.1	36	5.9	5.9	6.2	6.0	6.2	5.4	6.0	3.9	3.4	2.2	2.2	1.5
E2	0.0	0.3	0.1	0.6	2.8	4.1	3.1	2.6	5.9	5.6	4.2	2.4	2.3	1.6	2.4	2.2	2.4	28	3.7	2.7	4.6	4.1	4.4	4.8	1.7	28.8
E3	0.0	1.5	0.1	0.5	2.3	3.3	2.9	4.2	12.6	8.1	4.4	2.1	2.0	1.5	22	1.9	2.2	2.9	2.9	1.9	3.9	6.9	4.8	6.2	2.9	15.8
E4	0.0	0.1	0.1	0.4	1.6	2.6	3.8	4.5	13.1	8.2	6.2	4.7	60	4.7	64	48	4.6	4.8	3.7	4.5	3.8	3.3	3.2	2.0	1.7	1.2
£5	0.0	0.0	0.0	0.0	0.4	1.1	1.6	1.6	3.9	36	4.2	4.5	7.5	7.3	14.0	11.4	10.9	9.6	6.5	4.0	3.8	1.1	0.9	1.0	0.0	0.9
E7	0.0	0.1	0.1	0.2	0.9	1.6	2.2	3.0	7.1	10.1	10.9	8.0	8.5	6.3	8.6	6.7	5.4	5.0	4.3	2.9	3.5	1.1	0.4	0.2	0.7	2.4
E9	0.0	0.1	0.1	0.7	1.8	1.9	2.3	1.9	10.1	11.2	7.4	5.4	6.2	4.2	5.5	4.3	4.1	4.0	3.9	4.1	4.9	4.0	4.2	4.9	1.7	1.3
E10	0.0	1.4	0.2	0.8	2.5	3.6	5.1	4.7	17.0	14.5	5.5	2.2	2.1	1.3	2.0	1.6	2.0	2.6	3.3	2.8	3.7	3.8	2.8	6.4	1.9	61
E11	0.0	0.3	0.2	0.7	2.5	4.2	7.6	5.1	13.3	8.0	2.0	0.7	0.6	04	0.7	0.9	0.9	1.4	2.1	2.3	5.3	4.0	6.2	4.0	6.4	20.1
E12	0.0	0.0	0.1	0.3	1.2	2.4	6.2	9.6	35.7	16.4	5.3	2.2	1.7	0.9	0.9	0.4	0.4	0.5	0.4	0.4	0.4	1.1	1.5	2.4	00	9.8
E13	0.0	1.2	0.2	0.7	2.3	2.7	4.9	9.2	42.1	17.1	4.1	1.3	0.9	0.6	0.8	0.6	0.7	0.9	0.6	1.1	0.7	1.4	1.5	0.8	0.2	3.3
E14	0.0	0.0	0.2	0.4	1.4	1.8	3.6	12.3	42.6	15.5	6.1	2.8	2.8	1.8	2.4	1.4	1.3	0.9	0.7	0.2	1.8	0.0	0.0	0.0	0.0	0.0
F3	0.0	0.5	0.1	0.3	1.1	1.7	3.6	4.1	6.0	4.5	2.8	1.3	1.3	1.0	1.5	1.3	1.4	2.0	2.4	2.3	3.3	4.2	30	6.8	7.2	36.4
F4	0.0	0.0	0.0	0.1	0.3	0.6	1.3	1.7	14.9	17.7	14.2	7.5	6.4	3.7	4.6	3.5	3.1	3.2	2.9	2.7	3.1	2.2	1.1	02	1.0	4.2
F5	0.0	0.0	0.0	0.0	0.2	0.4	1.2	3.5	39.2	34.6	12.2	3.0	1.7	0.7	0.8	0.3	0.4	0.3	0.6	0.0	0.9	0.0	0.0	00	00	00
F6	0.0	0.0	0.0	0.0	0.1	0.4	1.9	9.2	58.3	21.5	4.3	1.3	0.8	0.3	04	1.5	0.0	0.0	0.0	0.0	0.0	00	0.0	00	0.0	00
F7	0.0	0.0	0.1	0.2	0.8	3.0	14.5	42.3	31.2	3.7	1.2	0.5	0.4	0.2	02	1.9	0.0	0.0	0.0	0.0	0.0	00	0.0	00	00	0.0
F12	0.0	0.0	0.0	0.1	0.6	2.1	9.9	29.0	12.4	2.5	1.9	1.4	1.6	1.0	1.5	0.9	0.9	0.9	0.8	0.8	0.8	1.0	1.8	2.4	3.8	21.9
G4	0.0	0.0	0.0	0.0	0.1	0.4	2.0	11.8	39.3	16.8	10.1	5.4	4.7	2.7	2.6	4.2	0.0	0.0	0.0	0.0	0.0	00	00	0.0	00	00

:

#### ANNEXE $\Pi$

Exemples de distributions granulométriques observées au cours des 2 campagnes

## de juillet 1991 et de janvier 1992

×,

;

Juillet 1991



•

diametre

Janvier 1992



•· •

#### ANNEXE III

Comparaison de distributions estimées et observées sur des lignes parallèles à la côte

en janvier 1992. cf. texte §.4.2.

۰

,

Janvier 1992



٠



•

....

Janvier 1992

#### ANNEXE IV

Cartographie des moyennes de Folk estimées en juillet 1991 et janvier 1992, et du coefficient de sorting de Trask en juillet 1991

# moy. Folk 07/91







#### ANNEXE V

## Transports sédimentaires calculés dans la zone-atelier Dyscop (Chapalain *et al.*, 1993 - légende ci-dessous)



Transports sédimentaires calculés (a) sous l'action combinée des courants de marée et de la houle, (b) sous l'action des courants de marée seuls.

# **ANNEXE II**

Résultats des simulations sur les estimateurs

de la richesse spécifique

.,

### Modèle P1 : Preston à 50 espèces

#### Richesse observée

·		Preston – R	ichesse observée	10 réplicats		
	50	100	500	1000	5000	10000
Moyenne	31.290	36.114	44.631	46.890	49.452	49.861
IC inf	27	32	42	45	48	49
Médiane	31	36	45	47	50	50
IC sup	35	40	47	49	50	50
		Preston – R	ichesse observée	20 réplicats		
	50	100	500	1000	5000	10000
Moyenne	36.161	40.264	46.880	48.333	49.840	49.985
IC inf	32	37	45	47	49	50
Médiane	36	40	47	48	50	50
IC sup	40	44	49	50	50	50
	. <u> </u>	Preston – R	ichesse observée	40 réplicats		
	50	100	500	1000	5000	10000
Moyenne	40.259	43.534	48.319	49.322	49.987	49.999
IC inf	37	41	46	48	50	50
Médiane	40	44	48	49	50	50
IC sup	44	46	50	50	50	50

### Jackknife 1

		Preston	– Jackknifel 10	réplicats		
	50	100	500	1000	5000	10000
Moyenne	38.222	42.288	48.377	49.527	50.215	50.214
IC inf	30.6	35.6	43.8	45.8	48.0	49.0
Médiane	38.2	42.3	48.5	49.7	- 50.0	50.0
IC sup	45.9	49.9	53.3	54.4	51.8	51.8
		Preston	– Jackknifel 20	réplicats		
	50	100	500	1000	5000	10000
Moyenne	42.561	45.497	49.369	49.896	50.165	50.046
IC inf	35.85	38.95	45.00	46.95	49.00	50.00
Médiane	42.60	45.70	48.90	49.90	50.00	50.00
IC sup	50.40	52.50	53.75	52.85	50.95	50.95
		Preston	– Jackknife1 40	réplicats		
	50	100	500	1000	5000	10000
Moyenne	45.547	47.583	49.972	50.348	50.067	50.002
IC inf	38.950	41.95	47.000	48.000	50.000	50
Médiane	45.825	47.875	49.95	50.000	50.000	50
IC sup	52.775	53.776	52.925	52.925	50.975	50

		Preston	– Jackknife2 10	réplicats		
	50	100	500	1000	5000	10000
Moyenne	41.130	44.909	49.242	49.975	50.315	50.171
IC inf	29.945	34.541	41.567	43.867	47.289	48.289
Médiane	40.911	44.633	48.956	49.689	50.278	50.0
IC sup	53.267	57.267	57.496	57.067	54.389	51.7
		Preston	– Jackknife2 20	réplicats		
	50	100	500	1000	5000	10000
Moyenne	44.970	46.990	50.073	50.388	50.083	49.960
IC inf	34.138	37.437	43.135	45.288	48.147	49.147
Médiane	44.689	46.837	49.992	50.142	50.00	50.00
IC sup	56.939	57.384	57.397	56.547	51.896	51.85
		<b>D</b>	* ** ** **			
		Preston	– Jackknife2 40	replicats		·····
	50	100	500	1000	5000	10000
Moyenne	47.290	48.396	50.385	50.667	49.973	49.999
IC inf	37.220	39.444	45.149	47.0	49.074	50
Médiane	46.922	48.072	49.999	50.074	50.0	50
IC sup	57.787	58.477	56.774	55.775	51.925	50

## Bootstrap

		Preston	– Bootstrap 10 r	éplicats		
	50	100	500	1000	5000	10000
Moyenne	34.682	39.177	46.530	48.177	49.908	50.054
IC inf	29.151	34.268	42.847	45.315	48.085	49.006
Médian <b>e</b>	34.701	39.103	46.544	48.194	50.199	50.136
IC sup	40.319	44.389	50.016	51.254	51.052	50.698
		Preston	– Bootstrap 20 1	éplicats		
	50	100	500	1000	5000	10000
Moyenne	39.202	42.784	48.231	49.256	50.068	50.042
IC inf	34.591	38.112	44.935	46.640	49.004	50.000
Médiane	39.341	42.666	48.279	49.124	50.148	50.015
IC sup	44.290	47.103	51.240	51.607	50.605	50.359
<i>i</i> .		Preston	– Bootstrap 40 i	réplicats		
	50	100	500	1000	5000	10000
Moyenne	42.903	45.470	49.132	49.931	50.046	50.005
IC inf	38.592	41.635	46.541	48.104	50.000	50.000
Médiane	42.898	45.495	49.009	49.856	50.029	50.000
IC sup	46.949	49.342	51.499	51.164	50.365	50.044

		Prest	on – Sinf 10 rép	licats		
	50	100	500	1000	5000	10000
Moyenne	33.223	36.747	44.831	46.737	49.375	49.827
IC inf	28.254	32.413	41.736	44.610	47.976	48.988
Médiane	33.105	36.893	44.997	46.711	49.720	49.946
IC sup	38.875	41.469	47.820	49.554	50.066	50.032
		Prest	ton – Sinf <u>20 r</u> ép	licats		
	50	100	500	1000	5000	10000
Moyenne	36.065	39.681	46.196	47.656	49.640	49.922
IC inf	31.902	36.189	43.723	45.458	48.862	49.704
Médiane	35.814	39.622	46.237	47.589	49.726	49.955
IC sup	40.899	42.779	48.319	49.396	49.940	49.998
		Pres	ton – Sinf 40 rép	licats		
	50	100	500	1000	5000	10000
Moyenne	38.927	41.844	47.278	48.690	49.858	49.951
IC inf	35.878	39.069	45.539	47.484	49.613	49.837
Médiane	38.737	41.833	47.246	48.902	49.886	49.969
IC sup	41.745	44.376	48.897	49.459	49.956	49.993

## Chao 2

		Chao2	2 - Preston 10 réj	olicats		
	50	100	500	1000	5000	10000
Moyenne	41.37	44.25	48.6	49.64	50.07	50.04
IC inf	28.74	34.12	42.5	45.11	48	49
Médiane	38.4	42.12	47.6	48.67	50	50
IC sup	72.76	69.04	63	59.5	52	50.5
NA	5	5	19	63	373	542
		Chao	2 - Preston 20 ré	plicats		
	50	100	500	1000	5000	10000
Moyenne	44.39	45.76	49.54	49.93	50	50
IC inf	34.14	38	45	47	49	50
Médiane	42.06	44.12	48.5	49.5	50	50
IC sup	68.05	64.5	59.5	57	50.5	50
NA	7	13	82	187	569	838
		Chao	2 - Preston 40 ré	plicats		
	50	100	500	1000	5000	10000
Moyenne	46.4	47.79	49.86	50.07	50.01	50
IC inf	38.08	41	47	48	50	50
Médiane	44.5	46.6	49.25	50	50	50
IC sup	64.5	63	57	52.2	50	50 <sup>·</sup>
NA	18	23	220	311	828	991

## Modèle P2 : Preston à 200 espèces

### Richesse observée

		Preston 200 -	Richesse observ	ée 10 réplicats		
	50	100	500	1000	5000	10000
Moyenne	56.841	72.167	114.423	130.871	165.323	175.837
IC inf	49	65	106	124	159	170
Médiane	57	72	114	131	165	176
IC sup	64	80	122	138	172	181
		Preston 200 –	Richesse observ	ée 20 réplicats		
	50	100	500	1000	5000	10000
Moyenne	72.868	89.644	131.608	147.095	176.167	184.091
IC inf	65	81.975	124	140	171	179
Médiane	73	89	132	147	176	184
IC sup	81	98	139	155	182	189
		Preston 200 –	Richesse observ	ée 40 réplicats		
	50	100	500	1000	5000	10000
Moyenne	90.943	107.616	147.412	160.944	184.628	190.043
IC inf	83	100	140	154	180	186
Médiane	91	108	147	161	185	190
IC sup	99	116	155	167	189	194

#### Jackknife 1

		Preston 20	0 – Jackknifel 1	0 réplicats		
	50	100	500	1000	5000	10000
Moyenne	77.296	95.496	138.615	154.190	181.996	188.798
IC inf	64.197	82.197	124.197	141.197	170.800	179.2
Médiane	77.000	95.500	138.400	154.300	182.100	188.6
IC sup	92.000	109.602	154.302	167.302	193.305	198.8
		Preston 20	0 – Jackknifel 2	0 réplicats	·····	
	50	100	500	1000	5000	10000
Moyenne	97.153	114.607	155.058	168.444	190.171	194.301
IC inf	82.200	100.2	142.100	155.250	179.599	185.6988
Médiane	96.800	114.6	154.800	168.800	190.300	194.400
IC sup	113.251	130.2	169.254	182.501	200.050	203.200
		D . 00				
		Preston 20	0 – Jackknifel 4	0 réplicats		
	50	100	500	1000	5000	10000
Moyenne	117.244	133.952	168.438	179.320	194.926	197.204
IC inf	102.499	119.500	155.573	167.650	185.899	189.925
Médiane	117.325	134.300	168.475	179.475	194.750	196.850
IC sup	133.075	149.101	181.372	191.425	203.626	203.725

### Jackknife 2

•		Preston 20	0 – Jackknife2 1	0 réplicats		
	50	100	500	1000	5000	10000
Moyenne	89.730	107.140	149.619	163.546	187.767	193.025
IC inf	68.086	85.937	128.333	141.959	170.733	177.200
Médiane	89.700	106.539	149.111	163.239	187.978	192.694
IC sup	110.848	130.981	173.659	185.584	205.367	209.458
		Preston 20	0 – Jackknife2 2	0 réplicats		
	50	100	500	1000	5000	10000
Moyenne	110.680	127.187	165.467	176.65+	194.108	196.512
IC inf	88.763	104.366	144.100	155.405	177.969	182.716
Médiane	110.606	126.727	165.387	176.468	193.522	196.401
IC sup	134.374	151.091	188.672	198.830	211.445	211.776
		Preston 20	0 – Jackknife2 4	0 réplicats		
	50	100	500	1000	5000	10000
Moyenne	130.069	146.0408	175.805	185.833	197.004	199.080
IC inf	105.378	122.835	154.358	166.978	182.515	187.221
Médiane	129.868	145.979	175.466	185.542	196.919	198.847
	154,195	169.121	198.352	206.645	211.246	211.636

## Bootstrap

		Preston 20	00 – Bootstrap 10	) réplicats		
	50	100	500	1000	5000	10000
Moyenne	66.149	82.533	125.855	141.833	173.682	182.310
IC inf	56.579	72.463	115.503	131.997	166.429	175.143
Médiane	65.964	82.598	125.868	142.001	173.494	182.160
IC sup	76.830	92.793	136.566	151.490	182.178	189.374
		Preston 2	)0 Bootstrap 2	0 réplicate		
	50	100	50 - DOUISITAD 2		<b>5</b> 000	10000
		100	500	1000	5000	10000
Moyenne	83.750	100.977	142.629	157.148	183.245	189.060
IC inf	72.956	90.671	132.627	147.768	175.842	183.032
Médiane	83.656	101.054	142.433	157.110	183.464	189.044
IC sup	94.841	112.490	152.725	166.107	190.365	195.174
•		Preston 20	00 – Bootstrap 4	0 réplicats		
	50	100	500	1000	5000	10000
Moyenne	102.684	119.463	157.314	169.885	189.851	193.949
IC inf	91.430	109.217	148.262	161.243	183.214	188.710
Médiane	102.760	119.421	157.182	169.875	189.854	194.087
IC sup	114.067	129.997	166.469	178.344	195.766	198.715

## Chao 2

Chao2 - Preston 200 10 réplicats								
	50	100	500	1000	5000	10000		
Moyenne	89.33	103.96	141.21	155.58	181.09	187.72		
IC inf	62.44	78.88	119.12	135.71	166.56	174.89		
Médiane	85.18	101.00	138.13	153.53	179.64	186.23		
IC sup	136.01	143.06	174.66	187.51	202.73	212.34		
NA	0	0	0	0	0	0		

## Modèle M1 : Motomura à 20 espèces

#### Richesse observée

		Motomural -	Richesse observ	ée 10 réplicats		•
	50	100	500	1000	5000	10000
Moyenne	9.310	10.188	12.598	13.571	15.812	16.902
IC inf	8	9	11	12	14	15
Médiane	9	10	13	14	16	17
IC sup	11	12	14	16		19
		Motomura1 –	Richesse observ	ée 20 réplicats		
	50	100	500	1000	5000	10000
Moyenne	10.214	11.138	13.533	14.610	16.806	17.845
IC inf	9	9	12	13	15	16
Médiane	10	11	13.5	15	17	18
IC sup	12	13	15	17	19	20
		Motomura1 –	Richesse observ	ée 40 réplicats		
	50	100	500	1000	5000	10000
Moyenne	11.090	12.023	14.493	15.591	17.792	18.692
IC inf	9	10	13	14	16	17
Médiane	11	12	14	16	18	19
IC sup	13	14	16	17.025	19	20

## Jackknife 1

		Motomura	1 – Jackknifel 1	0 réplicats		
	50	100	500	1000	5000	10000
Moyenne	10.537	11.492	13.851	14.798	17.241	18.231
IC inf	8.000	9.00	11.0	12.0	14.000	15.0
Médiane	10.800	11.00	13.9	14.8	16.900	17.9
IC sup	13.722	14.82	17.6	18.6	20.722	21.7
		Motomura	<u>1 – Jackknifel 2</u>	0 réplicats		
	50	100	500	1000	5000	10000
Moyenne	11.496	12.364	14.834	15.936	18.145	19.074
IC inf	9.00	9.00	12.0	13.0	15.00	16.00
Médiane	10.95	11.95	14.9	15.9	17.95	18.95
IC sup	14.85	15.85	18.8	19.8	21.85	22.80
		Motomura	1 – Jackknifel 4	0 réplicats		
	50	100	500	1000	5000	10000
Moyenne	12.387	13.297	15.952	17.148	19.113	19.747
IC inf	9.000	10.000	13.000	14.000	16.000	17.000
Médiane	11.975	12.975	15.950	16.975	18.975	19.975
IC sup	15.925	16.925	19.925	20.900	21.974	22.925

## Jackknife 2

		Motomura	1 – Jackknife2 1	0 réplicats		
	50	100	500	1000	5000	10000
Moyenne	11.117	12.041	14.365	15.429	17.898	18.805
IC inf	6.867	7.867	10.286	11.289	13.289	14.578
Médiane	10.700	11.700	13.989	14.700	17.700	18.694
IC sup	17.089	17.800	20.800	21.800	23.807	24.800
		Motomura	1 – Jackknife2 2	0 réplicats		
	50	100	500	1000	5000	10000
Moyenne	11.999	12.976	15.488	16.706	18.807	19.636
IC inf	7.986	9.000	11.147	12.147	14.147	15.295
Médiane	11.850	12.145	14.997	15.997	18.145	18.997
IC sup	18.400	18.550	21.547	23.400	25.400	24.550
		Motomura	11 – Jackknife2 4	0 réplicats		
	50	100	500	1000	5000	10000
Moyenne	13.057	13.867	16.773	17.958	19.667	20.089
IC inf	9.000	9.149	12.074	13.074	15.147	16.147
Médiane	12.925	13.074	15.999	17.925	18.999	19.925
IC sup	18.777	18.849	23.700	24.700	24.777	25.775

#### Bootstrap

		Motomura	al – Bootstrap 10	) réplicats		
	50	100	500	1000	5000	10000
Moyenne	9.865	10.771	13.154	14.138	16.508	17.526
IC inf	7.995	9.012	11.034	12.034	14.035	15.034
Médiane	9.569	10.538	13.332	13.805	16.564	17.563
IC sup	12.425	13.423	15.401	16.502	19.172	20.032
		Motomura	al – Bootstrap 2	0 réplicats		
<u> </u>	50	100	500	1000	5000	10000
Moyenne	10.767	11.746	14.105	15.169	17.447	18.430
IC inf	8.122	9.122	12.043	13.042	15.042	16.050
Médiane	10.530	11.580	13.756	14.845	17.530	18.520
IC sup	13.319	14.435	16.319	17.473	20.209	20.473
		Motomur	1 Bootstrop A	0 réplicate		
	<b>E</b> 0	100				10000
	50	100	500	1000	5000	10000
Moyenne	11.673	12.636	15.082	16.369	18.385	19.232
IC inf	9.045	10.044	13.024	14.089	16.045	17.088
Médiane	11.536	12.537	14.731	16.542	18.536	19.455
IC sup	14.263	15.233	17.458	18.581	20.453	21.180

		Motom	ura1 – Sinf 10 re	eplicats		
	50	100	500	1000	5000	10000
Moyenne	9.723	10.329	13.001	13.641	16.165	17.024
IC inf	7.929	8.949	10.998	11.911	14.042	15.005
Médiane	9.180	10.097	13.019	13.348	16.059	17.034
IC sup	12.682	13.233	16.218	17.083	18.765	19.095
		Motor	ura1 – Sinf 20 r	éplicats		
	50	100	500	1000	5000	10000
Moyenne	10.246	11.108	13.427	14.614	16.811	17.907
IC inf	8.799	8.999	11.839	12.857	14.861	15.972
Médiane	9.996	10.915	13.393	14.841	16.848	17.938
IC sup	12.943	13.373	15.328	16.674	19.334	19.832
		Motom	nura1 - Sinf 40 r	éplicats		
-	50	100	500	1000	5000	10000
Moyenne	10.750	11.757	14.347	15.349	17.553	18.438
IC inf	8.958	9.929	12.724	13.709	15.863	16.882
Médiane	10.716	11.684	14.577	15.571	17.666	18.654
IC sup	12.753	13.791	15.854	16.990	18.858	19.842

-----

.

,,

## Modèle M2 : Motomura à 30 espèces

### Richesse observée

		Motomura2 –	Richesse observ	ée 10 réplicats		
÷.	50	100	500	1000	5000	10000
Moyenne	16.129	18.052	22.492	24.359	27.925	29.126
IC inf	14	16	20	22	26	27
Médiane	16	18	23	24	28	29
IC sup	19	21	25	27	30	30
		Motomura2 –	Richesse observ	ée 20 réplicats		
	50	100	500	1000	5000	10000
Moyenne	17.929	19.941	24.348	25.974	29.064	29.779
IC inf	15	17	22	24	27	29
Médiane	18	20	24	26	29	30
IC sup	20	23	27	28	30	30
		Motomura2 –	Richesse observ	ée 40 réplicats		
	50	100	500	1000	5000	10000
Moyenne	19.764	21.632	25.982	27.540	29.744	29.986
IC inf	17	19	24	25	29	30
Médiane	20	22	26	28	30	30
IC sup	22	24	28	30	30	30

### Jackknife 1

_		Motomura	2 – Jackknifel 1	0 réplicats		·
	50	100	500	1000	5000	10000
Moyenne	18.636	20.558	24.974	26.599	29.849	30.435
IC inf	13.997	16.0	20.78	22.9	26.0	27.997
Médiane	18.700	20.7	24.80	26.7	29.8	30.800
IC sup	23.402	25.5	30.40	31.4	33.6	32.700
		Motomura	12 – Jackknifel 2	0 réplicats		
	50	100	500	1000	5000	10000
Moyenne	20.496	22.314	26.709	28.281	30.383	30.351
IC inf	16.0	17.95	22.00	24.00	27.926	29.0
Médiane	19.9	21.90	26.85	27.90	30.000	30.0
IC sup	25.7	27.70	31.70	32.75	32.850	31.9
		Motomura	12 – Jackknifel 4	0 réplicats		
	50	100	500	1000	5000	10000
Moyenne	22.384	24.206	28.286	29.697	30.387	30.059
IC inf	17.975	19.975	24.000	25.975	29.00	29.000
Médiane	21.950	23.950	27.950	29.925	30.00	30.000
IC sup	27.875	29.850	32.876	33.875	31.95	30.975

		Motomura	2 – Jackknife2 1	0 réplicats		
	50	100	500	1000	5000	10000
Moyenne	19.831	21.715	26.013	27.623	30.509	30.575
IC inf	12.866	14.578	19.578	21.567	24.867	26.578
Médiane	19.678	21.678	25.967	27.389	30.400	30.278
IC sup	27.489	29.778	34.200	34.796	36.089	35.100
•		Motomura	12 – Jackknife2 2	0 réplicats		
	50	100	500	1000	5000	10000
Moyenne	21.761	23.415	27.785	29.323	30.582	30.130
IC inf	15.147	16.295	21.000	23.000	26.295	27.442
Médiane	21.697	22.995	27.845	28.992	30.145	30.000
IC sup	29.250	31.395	35.397	36.397	35.550	33.700
		Motomura	12 – Jackknife2 4	0 réplicats		
	50	100	500	1000	5000	10000
Moyenne	23.638	25.435	29.259	30.629	30.261	29.929
IC inf	16.149	18.074	23.074	24.223	27.149	29.000
Médiane	23.849	25.072	28.924	30.850	30.000	30.000
IC sup	31.625	33.699	36.699	36.796	33.850	31.925

### Bootstrap

		Motomura	a2 – Bootstrap 10	) réplicats		
	50	100	500	1000	5000	10000
Moyenne	17.405	19.120	23.729	25.317	28.851	29.793
IC inf	14.469	16.176	20.435	22.328	26.166	27.377
Médiane	17.264	19.130	23.773	25.166	28.913	29.778
IC sup	20.942	22.887	27.072	28.701	31.425	31.268
		Motomur	a2 – Bootstrap 2	0 réplicats		
	50	100	500	1000	- 5000	10000
Moyenne	19.190	21.022	25.386	27.096	29.757	30.146
IC inf	16.208	17.503	22.401	24.184	27.371	29.051
Médiane	19.168	21.042	25.235	27.086	29.768	30.260
IC sup	22.926	24.336	28.797	30.076	31.357	30.842
		Motomur	a2 – Bootstrap 4	0 réplicats		
	50	100	500	1000	5000	10000
Moyenne	20.934	22.882	26.935	28.552	30.153	30.062
IC inf	17.732	19.516	24.163	25.506	29.036	29.044
Médiane	21.034	23.016	26.990	28.685	30.318	30.049
IC sup	24.241	26.128	30.075	31.325	30.870	30.408

•

	50	100	500	1000	5000	10000
Moyenne	16.416	18.599	22.867	24.402	28.186	29.280
IC inf	13.015	15.417	19.920	21.886	26.008	27.844
Médiane	16.402	18.395	22.981	24.296	28.057	29.148
IC sup	19.848	23.126	25.852	27.534	30.821	30.596
	50	100	500	1000	5000	10000
	50	100	500	1000	5000	10000
Moyenne	17.895	20.075	24.227	25.711	28.959	29.681
IC inf	14.908	16.745	21.794	23.191	27.694	28.717
Médiane	17.805	19.891	23.919	25.775	28.811	29.828
	01.171	22 202	37 697	20 000	20.052	20.005

Motomuraz – Sim 40 replicats									
	50	100	500	1000	5000	10000			
Moyenne	19.464	21.116	25.511	27.053	29.372	29.818			
IC inf	17.312	19.371	23.617	25.080	28.491	29.579			
Médiane	19.434	21.276	25.428	27.275	29.575	29.852			
IC sup	21.605	23.610	28.283	28.593	29.885	29.954			

### Chao 2

Chao2 – Motomura2 10 réplicats 50 100 500 5000 10000 1000 Moyenne 19.41 21.42 26.18 27.31 29.79 30.24 26 29.125 IC inf 14 16 20.5 22.25 28 18.25 30 34.2 Médiane 20.125 24.5 26.25 IC sup NA 37.5 31.78 32.5 38.25 37 174 182 197 175 201 220

# **ANNEXE III**

Résultats des simulations sur l'indice de Shannon H' et l'estimateur CRIC

### Modèle P1 : Preston à 50 espèces

H' vrai = 3.510

H' Shannon								
	50	100	500	1000	5000	10000		
Ecart à H'vrai	0.419	0.246	0.068	0.036	0.008	0.005		
IC inf	2.596	2.923	3.258	3.353	3.447	3.466		
Moyenne	3.091	3.264	3.442	3.474	3.502	3.505		
IC sup	3.614	3.596	3.612	3.592	3.557	3.541		

H' CRIC

	50	100	500	1000	5000	10000
Ecart à H'vrai	-0.075	0.043	0.078	0.067	0.050	0.047
IC inf	2.875	3.039	3.231	3.306	3.406	3.422
Moyenne	3.585	3.467	3.432	3.443	3.460	3.464
IC sup	4.459	3.934	3.614	3.569	3.518	3.501

H' CRIC - régression robuste

<u> </u>	50	100	500	1000	5000	10000
Ecart à H'vrai	-0.145	-0.015	0.023	0.015	-0.001	-0.004
IC inf	2.916	3.059	3.269	3.359	3.453	3.472
Moyenne	3.655	3.525	3.487	3.495	3.511	3.514
IC sup	4.593	4.034	3.687	3.627	3.568	3.553

#### Modèle M1 : Motomura à 20 espèces

H' vrai = 2.000

H' Shannon								
	50	100	500	1000	5000	10000		
Ecart à H'vrai	0.112	0.065	0.020	0.011	0.003	0.002		
IC inf	1.479	1.670	1.840	1.902	1.956	1.970		
Moyenne	1.880	1.935	1.979	1.989	1.997	1.998		
IC sup	2.236	2.188	2.101	2.074	2.035	2.023		

H' CRIC								
	50	100	500	1000	5000	10000		
Ecart à H'vrai	-0.233	-0.135	-0.053	-0.042	-0. 030	-0.026		
IC inf	1.662	1.799	1.898	1.942	1.989	1.995		
Moyenne	2.233	2.135	2.053	2.042	2.030	2.026		
IC sup	2.971	2.525	2.202	2.139	2.071	2.055		

H' CRIC – régression robuste								
	50	100	500	1000	5000	10000		
Ecart à H'vrai	-0.233	-0.131	-0.048	-0.036	-0.015	-0.010		
IC inf	1.660	1.797	1.897	1.936	1.969	1.979		
Moyenne	2.233	2.131	2.048	2.036	2.015	2.010		
IC sup	2.996	2.521	2.201	2.137	2.059	2.044		

#### Modèle M2 : Motomura à 30 espèces

H' vrai = 2.937

H' Shannon								
50 100 500 1000 5000 10000								
0.218	0.120	0.034	0.019	0.005	0.003			
2.337	2.566	2.766	2.831	2.892	2.907			
2.719	2.817	2.903	2.919	2.932	2.934			
3.085	3.071	3.034	3.003	2.971	<b>2</b> .960			
•	50 0.218 2.337 2.719 3.085	50         100           0.218         0.120           2.337         2.566           2.719         2.817           3.085         3.071	Fill         Snannon           50         100         500           0.218         0.120         0.034           2.337         2.566         2.766           2.719         2.817         2.903           3.085         3.071         3.034	50         100         500         1000           0.218         0.120         0.034         0.019           2.337         2.566         2.766         2.831           2.719         2.817         2.903         2.919           3.085         3.071         3.034         3.003	50         100         500         1000         5000           0.218         0.120         0.034         0.019         0.005           2.337         2.566         2.766         2.831         2.892           2.719         2.817         2.903         2.919         2.932           3.085         3.071         3.034         3.003         2.971			

H' CRIC								
	50	100	500	1000	5000	10000		
Ecart à H'vrai	-0.311	-0.188	-0.077	-0.061	-0.042	-0.037		
IC inf	2.637	2.761	2.853	2.892	2.936	2.943		
Moyenne	3.249	3. 125	3.015	2.998	2.979	2.974		
IC sup	4.089	3.541	3.167	3.097	3.021	3.004		

### H' CRIC - régression robuste

	50	100	500	1000	5000	10000
Ecart à H'vrai	-0.316	-0.183	-0.062	-0.043	-0.022	-0.016
IC inf	2.640	2.755	2.841	2.876	2.915	2.921
Moyenne	3.253	3.121	2.999	2.981	2.959	2.954
IC sup	4.190	3.551	3.155	3.084	3.004	2.985

### Modèle Mixte à 50 espèces

H' vrai = 2.783

H' Shannon								
	50	100	500	1000	5000	10000		
Ecart à H'vrai	0.234	0.150	0.048	0.029	0.007	0.004		
IC inf	2.164	2.354	2.589	2.652	2.729	2.748		
Moyenne	2.549	2.633	2.735	2.754	2.776	2.779		
IC sup	2.895	2.926	2.879	2.851	2.823	2.807		

H' CRIC								
	50	100	500	1000	5000	10000		
Ecart à H'vrai	-0.218	-0.056	0.048	0.050	0.040	0.036		
IC inf	2.481	2.528	2.599	2.640	2.698	2.718		
Moyenne	3.001	2.839	2.735	2.733	2.743	2.747		
IC sup	3.590	3.171	2.868	2.823	2.788	2.773		

H' CRIC – régression robuste

	50	100	500	1000	5000	10000
Ecart à H'vrai	-0.222	-0.047	0.060	0.058	0.040	0.035
IC inf	2.487	2.529	2.585	2.631	2.697	2.718
Moyenne	3.005	2.830	2.723	2.725	2.743	2.748
IC sup	3.634	3.153	2.860	2.817	2.789	2.777

H' CRIC - régression LTS

	50	100	500	1000	5000	10000
Ecart à H'vrai	-0.223	-0.056	-0.052	-0.032	-0.008	-0.005
IC inf	2.341	2.475	2.590	2.680	2.744	2.756
Moyenne	3.006	2.839	2.835	2.815	2.791	<b>2</b> .788
IC sup	3.671	3.203	3.142	2.961	2.840	2.818