

Société Française de Biométrie

**BIOMETRIE
ET
HALIEUTIQUE**

1998

**Textes présentés lors des journées réalisées
par la Société Française de Biométrie**

**à Rennes
du 26 au 28 mai 1998**

Edité par

Duby Camille, INA-PG Paris

Gouet Jean-Paul, ITCF, Boigneville

Laloë Francis, IRD, Montpellier

N° 15

Table des matières

Méthodes d'étude de l'activité de pêche dans les pêcheries composites D. Pelletier	1
Le GAM, un outil d'estimation des biomasses locales. Application au thon albacore (Thunnus albacares) de l'Atlantique O. Maury, L. Millischer, D. Gascuel, A. Fonteneau	11
Représentation simulée d'un processus de recherche et de capture de bancs de poissons répartis selon un processus ponctuel marqué P. Petitgas, F. Laloe	21
Actualisation d'un modèle du cycle biologique du saumon V. Trenkel, R. Faivre, J. Badia, M. Goulard	27
L'estimation des puissances de pêche par modélisation linéaire des capturabilités L. Millischer, O. Maury, D. Gascuel	37
Modèles matriciels et colonisation de milieu vierge : l'exemple de la truite commune aux îles Kerguelen M. Jarry, P. Davaine, E. Beall	47
Utilisation des méthodes de segmentation pour la construction de cartes illustrant les relations ressources halieutiques/environnement S. Salaun, J. Ferraris	59
Comparaison de stratégies de pêche par analyses factorielles sur tableaux d'échanges Jocelyne Ferraris, Catherine Pardoux	71
Estimation des taux de mortalité M et F pour l'anchois : présentation générale et premiers résultats O. Gavart, M. Artzrouni	85
Structuration spatiale de la ressource halieutique : exemple d'une pêcherie démersale à Vanuatu (Océanie) Espérance Cillauren, Monique Simier, Laurence Blanc	93

Modèle de mortalité exponentiel étendu et estimation semi-paramétrique. Application à l'évaluation de la production des œufs d'anchois dans le Golfe de Gascogne B. Truong-van	107
Estimation de la biomasse des anchois dans le Golfe de Gascogne. Une étude statistique des estimateurs B. Truong-van, J.C. Turlot	113
Méthodes de vraisemblance pour les captures successives. Applications aux saumons dans l'Adour-Nive B. Truong-van, C. Kokonendji, P. Prouzet	123

METHODES D'ETUDE DE L'ACTIVITE DE PECHE DANS LES PECHERIES COMPOSITES

Dominique Pelletier

Laboratoire MAERHA, IFREMER BP 21105, 4311 Nantes Cedex 03

Introduction

Une définition courante de l'halieutique est la science qui étudie les populations marines exploitées par la pêche dans l'optique de mettre en adéquation les moyens de pêche avec les potentialités des ressources. Cette définition transparait clairement dans la terminologie anglophone puisque halieutique se dit « fisheries science », c'est-à-dire science des pêcheries. Or, traditionnellement, la recherche halieutique s'est surtout focalisée sur les populations exploitées. Relativement peu de recherches s'intéressent à la composante humaine, c'est-à-dire le pêcheur, et les facteurs qui sous-tendent son activité. Cette focalisation transparait de manière évidente dans les articles publiés dans les revues halieutiques. Il y a une dizaine d'années, Hilborn (1985) notait déjà le peu d'études relatives à la caractérisation et à la dynamique de l'activité de pêche.

Cependant, que l'objet central de l'étude soit la ressource ou le pêcheur, une bonne connaissance des caractéristiques de l'activité de pêche est indispensable à une représentation correcte de la dynamique des populations exploitées. En effet, la majeure partie de la mortalité subie par les populations exploitées est le fait de la pêche. Il est donc impossible d'ignorer cette source de mortalité dans la dynamique des populations.

S'intéresser à l'activité de pêche, c'est se placer d'emblée - du fait de la prééminence de l'homme dans le système- dans une perspective pluridisciplinaire où écologie marine, géographie, économie, sociologie et anthropologie ont toutes leur mot à dire. Cependant, ce n'est pas l'étude de la complexité liée à cet aspect pluridisciplinaire que je souhaite aborder dans cette communication, il s'agit plutôt de l'étude des conséquences de l'hétérogénéité de l'activité de pêche induite par la position centrale de l'homme dans les pêcheries, et dans une moindre mesure par la diversité des peuplements exploités. Cette hétérogénéité s'exprime par le terme de "pêcherie composite", c'est-à-dire une pêcherie caractérisée par de nombreuses espèces exploitées concurremment par divers engins à partir de différents bateaux, selon des modes opératoires variables au cours de l'année. Je tenterai dans un premier temps de montrer ces composantes de la diversité à partir d'exemples. Puis, je donnerai un aperçu des différentes méthodes adoptées en halieutique pour décrire, analyser et modéliser l'activité de pêche dans de telles pêcheries. Pour ne pas sortir de mon domaine de compétence, je me limiterai à l'articulation entre l'homme et les ressources qu'il exploite, bien que je sois

persuadée de l'importance des autres points de vue. Ainsi, les relations entre l'environnement abiotique et les ressources d'une part, et le pêcheur et le système économique dans lequel il évolue, sortent du cadre de cet exposé.

1. La diversité de l'activité de pêche.

Cette partie sera illustrée par l'exemple de la flotte de pêche commerciale bretonne (Berthou et al. 1997). En 1994, celle-ci comprenait environ 30% des navires de pêche français et 35% des marins inscrits en France. Pour ce qui est des ventes sous criée, la valeur de sa production est d'environ 45% de la production française. Cette flotte est donc économiquement importante et active (77% des navires sont actifs toute l'année). Cependant, elle a subi une évolution rapide au cours des quinze dernières années, avec une réduction de 45% du nombre de navires entre 1983 et 1995. Cette évolution affecte les navires très différemment selon le type de pêche. En évaluer les conséquences pour la flotte et pour les ressources requiert donc une analyse fine des composantes de l'activité de pêche.

La diversité de l'activité de pêche s'exprime à la fois dans les moyens déployés pour pêcher et dans les captures qui en résultent.

1.1. Les moyens mis en œuvre pour capturer les ressources.

Ces moyens, à la fois d'ordre technique et humain - et donc foncièrement intangibles - sont extrêmement divers. Il faut cependant les quantifier d'une manière homogène pour évaluer leur impact sur les ressources.

1.1.1. Moyens techniques.

Il s'agit ici non seulement des caractéristiques des navires, des engins de pêche, mais aussi des outils technologiques d'aide à la pêche. En ce qui concerne les zones d'exploitation, la flotte bretonne se compose essentiellement de navires côtiers, c'est-à-dire qui exploitent exclusivement (72% des navires) ou majoritairement (13% des navires) une bande côtière de 12 milles nautiques. Le reste des navires exploite des zones beaucoup plus lointaines (Ouest de l'Ecosse par exemple). On note une grande spécialisation des ports de ce point de vue, les navires du Nord de la Bretagne étant nettement plus côtiers.

En conséquence, les caractéristiques techniques des navires sont très variables, avec des longueurs s'échelonnant de 4 à 54 m, la majorité des navires mesurant entre 6 et 16 m. Corrélativement, la puissance motrice des navires s'étale de 50 à 600kW, la majorité des navires possédant un moteur de 50 à 350 kW. Pour une même longueur, deux navires peuvent être très différents en fonction de leur année de construction du fait de l'évolution très rapide des technologies marines. La plupart des navires ont été construits entre 1970 et 1990.

1.1.2. Moyens humains.

La variété des pratiques de pêche est grande d'un pêcheur à l'autre, et de plus un même pêcheur est susceptible de recourir à différentes pratiques de pêche durant la saison de pêche (l'année en général) en fonction des variations de la conjoncture économique et de la disponibilité saisonnière des ressources. Dans l'enquête sur la flotte bretonne (Berthou et al. 1997), le métier est défini comme la façon dont le pêcheur décrit son activité à l'échelle de l'opération de pêche. Au cours de cette étude, 130 métiers avaient été recensés initialement ; ils ont ensuite été regroupés en 31 métiers dont les principaux sont le chalut de fond à poissons, le chalut de fond à langoustine, le filet à petites mailles à poissons, le casier à

grands crustacés et la drague à coquille Saint-Jacques. Ces cinq métiers sont pratiqués par 20% des bateaux et représentent 60% des mois d'activité de la flotte. Les navires exercent donc plusieurs métiers, soit simultanément, par ex. poser des filets et des casiers, soit successivement en fonction de la disponibilité saisonnière des ressources. Cette polyvalence est variable d'un port à l'autre, et elle est plus importante chez les navires côtiers.

Ainsi, la multiplicité des caractéristiques des navires, des engins de pêche, des zones de pêche et des habitudes locales génère une grande diversité de l'activité de pêche. Des enquêtes auprès des pêcheurs et l'analyse de calendriers d'activité permettent de caractériser cette activité, notamment en ce qui concerne les métiers (choix d'un engin et d'une ou plusieurs espèces-cibles) et les stratégies de pêche (enchaînement et/ou association des métiers au cours de l'année). Cependant, cette approche se place du point de vue du pêcheur, c'est-à-dire qu'elle ne répond pas à la question de la quantification de l'impact de chaque activité sur les ressources. Celle-ci nécessite des informations sur les captures réalisées par chaque type d'activité.

1.2. Les ressources capturées.

La multiplicité des espèces capturées est une autre cause de diversité de l'activité de pêche. En effet, la capture est généralement constituée d'un ensemble de populations des peuplements de la zone exploitée. Dans le cas de la pêche, ces peuplements sont très fortement modifiés par l'intensité de la pression de pêche. La dynamique du peuplement s'en trouve probablement grandement affectée. Dans l'optique de cet exposé, le fait important est qu'en pêchant à un endroit et un instant donnés, le pêcheur capture non seulement une fraction de la population ciblée (espèce-cible), mais aussi "accidentellement" des individus d'autres populations (espèces accessoires). Cet aspect doit absolument être pris en compte lorsqu'on veut évaluer l'effet de la pêche sur les populations, et inversement lorsqu'on s'intéresse à l'adaptation de l'activité de pêche aux captures réalisées.

Pour illustrer ce point, considérons les chalutiers qui exploitent la Mer Celtique, qui font partie de la flotte bretonne prise comme exemple jusqu'ici. La Mer Celtique est la mer épicontinentale qui s'étend au large de la Bretagne et du Pays de Galles, et limitée au Nord par l'Irlande et à l'Est par la Manche. Cette zone est exploitée majoritairement par des chalutiers issus principalement de ports du Sud de la Bretagne. En 1993, les captures réalisées par ces navires ne comprenaient pas moins de 104 espèces différentes, parmi lesquelles une quinzaine d'espèces dominantes, dont le merlan, le cabillaud, la baudroie, la langoustine, etc.... La capture totale de ces chalutiers n'est cependant pas représentative de la composition spécifique des captures individuelles de chaque navire (voir § 3.4). Ainsi, pour des navires relativement semblables et exploitant une même zone, l'activité de pêche recouvre des résultats très variables du fait du savoir-faire, de la compétence professionnelle et du comportement décisionnel des pêcheurs. Les espèces-cibles et les pratiques de pêches varient sensiblement d'un pêcheur à l'autre.

Même dans le cadre restrictif fixé en préambule, l'activité de pêche peut être considérée comme la résultante d'interactions entre des éléments de natures fondamentalement différentes comme des populations naturelles, des moyens techniques (bateaux, engins de pêche...), des aspects humains (expérience et comportement du pêcheur), chacune de ces composantes recouvrant une réalité multiple. L'évaluation de l'impact de l'activité de pêche sur les

ressources nécessite d'identifier et de quantifier ces composantes de façon à obtenir une mesure standardisée de la mortalité par pêche subie par chaque ressource. Cette analyse nécessite des données à l'échelle de chaque opération de pêche de chaque navire.

2. L'information.

Les données utiles à l'étude de l'activité de pêche sont de deux ordres : des données relatives à des campagnes expérimentales scientifiques, et des données relatives à la pêche commerciale. Les premières fournissent des indices d'abondance indépendants de la pêche, mais leur couverture temporelle est souvent très faible. En effet, on dispose au maximum d'une ou deux campagnes annuelles d'une durée habituelle de 3 à 4 semaines dans une zone donnée. A l'opposé, les données de la pêche commerciale sont en général beaucoup plus nombreuses et elles permettent de décrire l'activité de pêche. Elles comprennent dans le meilleur des cas, des données de capture et d'effort par unité de pêche et par unité de temps, ces unités variant selon les pêcheries. Elles fournissent aussi les mesures de certaines caractéristiques des bateaux, des engins, de l'équipage, etc...

Selon les cas, les données commerciales proviennent de systèmes nationaux de collecte de statistiques ou d'enquêtes de terrain auprès des pêcheurs.

2.1. Les statistiques de pêche.

En France, des obligations réglementaires nationale et européenne imposent aux pêcheurs qui exploitent certaines ressources, en particulier les espèces soumises à une gestion communautaire, des déclarations sous forme de journaux de pêche (log-books) après chaque sortie en mer (marée). Il existe donc un réseau national qui saisit a) les caractéristiques de chaque unité de pêche immatriculée, b) pour chaque marée d'un navire, le temps d'utilisation de chaque engin de pêche dans chaque zone fréquentée, c) pour chaque vente sous criée d'un navire, la capture par espèce. De fait, ce système couvre seulement les navires qui vendent leur capture sous criée, il en résulte donc une couverture médiocre de la pêche côtière et a contrario une bonne couverture de la pêche hauturière. Pour cette dernière, ce système obligatoire garantit une certaine exhaustivité et une continuité dans le temps.

2.2. Les enquêtes de terrain.

Les enquêtes menées auprès de pêcheurs sont réalisées par des équipes de recherche halieutique. Par limitation de moyens, ces études sont souvent ponctuelles, et pour combler les lacunes des systèmes statistiques officiels, elles concernent en général les petites pêches côtières. Les enquêtes collectent deux types d'information, les modalités de la pêche et les captures. En général, les modalités de l'activité de pêche sont recensées, à savoir engin(s) utilisé(s), caractéristiques techniques du navire, taille de l'équipage, zone(s) de pêche, mesure de l'effort de pêche, espèce(s)-cible(s), et éventuellement cause de non activité. Les captures sont souvent échantillonnées suivant des protocoles assez complexes liées à la structuration des débarquements, c'est-à-dire lieux et jours de débarquement, navires, caisses de poissons, etc... Une description précise d'un protocole mis en place pour étudier la pêche artisanale sénégalaise peut être trouvée dans Ferraris (1995).

Une autre caractéristique à prendre en compte dans l'étude de l'activité de pêche est l'étendue géographique de certaines pêcheries, c'est-à-dire de leur zone d'influence à terre, et des zones d'exploitation. Avec le fait que les populations exploitées ne sont pas directement observables,

l'étendue géographique constitue une contrainte dans l'étude de l'activité de pêche, qui a des conséquences importantes sur la collecte des données.

3. Analyse de l'activité de pêche.

3.1. Les objectifs.

Les exemples cités plus haut font apparaître une activité de pêche multiforme, soumise à une grande variabilité individuelle. Cette activité est analysée via une information nombreuse et essentiellement multivariée. En premier lieu, l'analyse de l'activité de pêche vise donc à synthétiser l'information disponible en identifiant à partir d'un grand nombre d'observations (par ex. captures et effort de pêche pour chaque sortie d'un navire) un nombre restreint de types d'activité qui ont des impacts similaires sur les ressources. Ainsi, une typologie de l'activité de pêche regroupe dans une même classe les navires qui pêchent de la même manière (mêmes espèces-cibles, mêmes lieux de pêche, mêmes engins...), et qui donc sont supposés engendrer des mortalités par pêche semblables sur les populations exploitées. Distinguer les types d'activité permet d'améliorer l'évaluation de ces populations, et de mieux comprendre le fonctionnement des pêcheries.

Ensuite, à partir de l'impact de chaque type d'activité sur les ressources, elle fournit des éléments en vue d'orienter des mesures de gestion spécifiquement sur certaines activités. Si de plus, l'analyse appréhende l'évolution temporelle des types d'activité de pêche, elle ouvre alors la voie à l'exploration de la dynamique de la pêcherie sous l'influence de différents scénarii de gestion ou pour différents états des ressources. Simultanément, l'analyse de l'activité de pêche cherche à identifier les facteurs déterminants de l'activité de pêche, c'est-à-dire les variables explicatives qui influencent le plus la structuration des données. Enfin, elle permet d'évaluer les variations individuelles et temporelles des efficacités de pêche des navires. Ceci est important pour arriver à standardiser les efforts de pêche entre navires, en vue d'évaluer l'impact de chacun sur les ressources. Par exemple, on peut se poser la question de l'impact sur le stock de merlu du Golfe de Gascogne, de 100h de chalutage contre 30km de filets calés pendant 12h.

3.2. Les différentes échelles de description de l'activité de pêche.

L'activité de pêche se rapporte à différents processus selon l'échelle à laquelle on se place (Tableau 1).

Activité de pêche	Echelle de temps	Type d'activité associé
Opération de pêche	de la minute à quelques heures ex : coup de chalut, pose d'un filet, d'un casier	métier (ou tactique de pêche)
Sortie de pêche (marée)	de quelques heures à quelques jours limitée par départ/retour au port peut coïncider avec opération	parfois associé à la tactique de pêche ou au métier
Saison de pêche	année ou partie de l'année	stratégie de pêche annuelle enchaînement des tactiques
Durée de vie d'une unité de pêche (bateau et équipage)	plusieurs années	stratégie à long terme (entreprise)

Tableau 1. Les différentes échelles de l'activité de pêche.

3.3. Les méthodes d'analyse de l'activité de pêche.

Compte tenu de la nature de l'information et des objectifs de l'analyse, les analyses descriptives multivariées ont été largement utilisées, et notamment les analyses factorielles et les techniques de classification. Ainsi, des types d'opération de pêche ont pu être identifiés à partir de données de pêche commerciale (Murawski et al. 1983 ; Biseau & Gondeaux 1988 ; Lewy & Vintheer 1994 ; Ferraris 1995 ; He et al. (1997), etc...). Dans certains cas, ces types ont été mis en relation avec des variables illustratives comme le port d'origine ou les caractéristiques techniques des navires. Bertignac (1992) et Rochet & Durand (1995) ont plutôt cherché à décrire l'évolution temporelle de l'activité de pêche, en particulier à partir d'analyses multitableaux pour la deuxième référence. Pech & Laloë (1997) présentent une application un peu différente, en ce sens qu'ils cherchent à identifier les sources de variation des captures par ACP sur variables instrumentales.

Une deuxième classe d'outils utilisés comprend les modèles linéaires généraux ou généralisés. L'estimation de puissances de pêche par navire ou groupe de navires a été très largement abordée grâce à des analyses de variance (Robson 1966 ; Laurec 1977 ; Stocker & Fournier 1984 ; Hilborn & Ledbetter 1985, etc...). Plus récemment, Gaertner et al. (1996) ont modélisé les tactiques de pêche aux thonidés en fonction de variables explicatives en recourant à des modèles log-linéaires et Logit.

3.4. Un exemple d'analyse : les tactiques de pêche des chalutiers de la Mer Celtique.

La tactique de pêche est le type d'activité de pêche défini ici comme le choix lors d'une marée, d'une ou plusieurs espèces-cibles et d'un lieu de pêche, ce choix étant éventuellement associé à une période précise de l'année. La question du choix de l'engin ne se pose pas dans cet exemple puisque tous les navires considérés sont des chalutiers. Les données utilisées pour réaliser cette analyse sont celles du système statistique officiel, à savoir pour chaque marée, l'effort de pêche par rectangle statistique, et pour chaque vente la capture en poids par espèce. Un rectangle statistique est une petite zone d'environ 30 milles nautiques de côté. L'année étudiée est 1993 pour laquelle 265 navires ont effectué 5262 marées en Mer Celtique ou dans les zones immédiatement environnantes. Les résultats, issus d'analyses factorielles et de

classification, (Pelletier & Ferraris 1999) ont mis en évidence 7 types de marées, c'est-à-dire 7 tactiques, et ont permis de les caractériser par certaines variables (Tableau 2).

Une tactique de pêche se présente ici sous la forme d'une conjonction de modalités de variables. Le formalisme des objets symboliques a d'ailleurs été utilisé par Ferraris et al. (1996) pour décrire des tactiques de pêche au Sénégal.

Une fois les tactiques obtenues, il est simple de calculer l'effort de pêche et les captures par tactique. La composition spécifique des captures varie énormément d'une tactique à l'autre, reflétant les choix d'espèce-cible. Ces résultats prouvent la nécessité d'analyser l'activité de pêche pour l'évaluation des ressources.

%	Espèce-cible	Lieux de pêche	Période de l'année	Caractéristiques des navires
31	Langoustine	Nord et Centre M.C. Sud Porcupine Grande Vasière	80% de Mars à Août	65% de Douarnenez, St Guénolé et Loctudy 25 % Concarneau et La Rochelle Unités peu puissantes
21	Baudroie	Sud M.C. Petite Sole Chapelle	saisonnalité non marquée	78% du Guilvinec Artisans, Petites unités
6	Gadidés	Nord-Est M.C. Canal de Bristol	38% de Fév. à Mars 19% d'Août à Sept.	66% de Lorient, Concarneau et La Rochelle Unités grandes, puissantes, assez anciennes
20	Tacauds et Grondins	Nord Bishop Manche-Est	saisonnalité non marquée	51% de Caen et 42% de Cherbourg Artisans, Unités puissantes et assez récentes

Tableau 2. Principales tactiques de pêche mises en évidence pour les chalutiers de Mer Celtique en 1993 (d'après Pelletier & Ferraris 1999). La première colonne indique la proportion de marées classées dans la tactique.

4. Modélisation de la dynamique de pêcheries composites

Une fois les composantes de l'activité de pêche analysées et quantifiées, il convient de pouvoir établir un diagnostic sur l'état des ressources, de prévoir leur évolution, et si possible d'explorer les conséquences de mesures de gestion sur la pêcherie. Des modèles mathématiques peuvent être construits à cet effet. Si ces modèles abondent dans le cas de pêcheries monospécifiques, ils restent peu fréquents dans le cas des pêcheries composites.

Des modèles diagnostics développés par Murawski (1984), Pikitch (1987), Mesnil & Shepherd (1990), évaluent les conséquences à long terme d'une situation initiale où les métiers et l'effort par métier sont fixés. Murawski & Finn (1986) et Marchal & Horwood (1996) ont mis au point des modèles d'optimisation qui permettent de trouver la répartition de l'effort de pêche entre plusieurs métiers, ou le maillage qui garantissent une exploitation durable des populations. Laloë & Samba (1991) et Laurec et al. (1991) ont modélisé la dynamique de l'effort de pêche et le comportement du pêcheur. Ce type de modèle a aussi été développé par Allen & McGlade (1986), mais sans vraiment s'appuyer sur des données réelles.

La nécessité de prendre en compte la complexité des pêcheries pour mieux les comprendre a favorisé ces dernières années le développement de modèles multi-agents issus de l'intelligence artificielle. L'objectif de ces modèles est de faire émerger des macrophénomènes à partir des interactions entre comportements individuels (Bousquet 1994 ; Le Fur 1995). Le problème majeur auquel se trouvent confrontés ces modèles réside dans leur calibration avec des données. Actuellement, cet obstacle empêche ce type de modèle de devenir des outils permettant de quantifier la dynamique de ressources ou de captures.

Conclusion

Dans un contexte où la majorité des pêcheries sont composites, l'analyse de l'activité de pêche est indispensable pour réduire la dimension du problème. Une modélisation qui s'appuie sur ces synthèses peut déboucher sur des outils puissants permettant d'évaluer la soutenabilité de la pêcherie, aussi bien que d'explorer la pertinence de mesures alternatives de gestion.

Bien que la recherche halieutique ait considérablement progressé dans ces domaines ces dernières années, peu de travaux ont conduit à des résultats opérationnels à court terme, c'est-à-dire chiffrés, permettant des diagnostics sur l'évaluation des ressources, ou étayant des propositions en matière de gestion ou d'aménagement des pêches. L'analyse de l'activité de pêche devrait être systématisée, car les outils statistiques sont disponibles et relativement standards. La modélisation de pêcheries composites est un domaine où la recherche reste très ouverte. Un compromis reste à trouver entre des modèles très complexes mais incapables d'aboutir à des diagnostics ou des quantifications, et des modèles simplistes qui laissent de côté des aspects essentiels de la dynamique de telles pêcheries. Les modèles futurs devront prendre en compte les composantes spatiale et saisonnière de la dynamique des ressources et de l'activité de pêche, la succession des tactiques au cours de l'année, et la dynamique interannuelle de l'effort de pêche. Etant donné le caractère composite de la plupart des pêcheries, c'est à ce prix que des progrès pourront être réalisés dans la quantification de l'impact de la pêche sur les ressources et des conséquences de différentes mesures de gestion sur les captures et les populations exploitées.

Bibliographie

- Allen P.M. & J. MacGlade, 1986. Dynamics of discovery and exploitation : the case of the Scotian shelf groundfish fisheries. *Can. J. Fish. Aquat. Sci.* 43, 1187-1200.
- Berthou P., C. Talidec, M. Jezequel, et P. Lespagnol. 1997. La flotte de pêche commerciale bretonne. Description des métiers et des flottilles. Rapport interne IFREMER. 103 p. + annexes.
- Bertignac M., 1992. Les rendements par espèce de la pêche chalutière hauturière française de l'Atlantique Nord-Est entre 1972 et 1989. *Aquat. Living Resour.* 5, 261-275.
- Biseau A., & E. Gondeaux. 1988. Apport des méthodes d'ordination en typologie des flottilles. *J. Cons. Int. Explor. Mer* 44, 286-296.
- Bousquet F., 1994. Des milieux, des poissons, des hommes : étude par simulations multi-agents. Le cas de la pêche artisanale dans le delta central du Niger. Thèse de doctorat de l'Université Claude Bernard - Lyon I. ORSTOM Editions, Bondy. 175 p.
- Ferraris J. 1995. Démarche méthodologique pour l'analyse des comportements tactiques et stratégiques des pêcheurs artisans sénégalais, pp. 263-295, *In* : Laloë F., H. Rey & J.L.

- Durand (éds), Questions sur la dynamique de l'exploitation halieutique, Série Colloques et séminaires, ORSTOM Editions, 542 p.
- Ferraris J., D. Pelletier, & M.J. Rochet, 1996. Modélisation des activités de pêche à l'aide de méthodes numériques-symboliques. Journées du Programme Environnement, Vie et Sociétés, Paris, 15-17 janvier.
- Gaertner D., M. Pagavino & J. Marcano, 1996. Utilisation de modèles linéaires généralisés pour évaluer les stratégies de pêche thonière à la senne en présence d'espèces associées dans l'Atlantique ouest. *Aquat. Living Resour.* 9, 305-323.
- He X., K.A. Bigelow & C.H. Boggs, 1997. Cluster analysis of longline sets and fishing strategies within the Hawaii-based fishery. *Fish. Res.* 31, 147-158.
- Hilborn R., 1985. Fleet dynamics and individual variation : why some people catch more fish than others. *Can. J. Fish. Aquat. Sci.* 42, 2-13.
- Hilborn R. & M. Ledbetter, 1985. Determinants of catching power in the British Columbia salmon purse seine fleet. *Can. J. Fish. Aquat. Sci.* 42, 51-56.
- Laloë F. & A. Samba, 1991. A simulation model of artisanal fisheries of Senegal. *ICES mar. Sci. Symp.*, 193, 281-286.
- Laurec A., 1977. Analyse et estimation des puissances de pêche. *J. Cons. int. Explor. Mer* 37, 173-185.
- Laurec A., A. Biseau, & A. Charreau, 1991. Modelling technical interactions. *ICES mar. Sci. Symp.* 193, 225-236.
- Lewy P. & M. Vinther, 1994. Identification of Danish North Sea trawl fisheries. *ICES J. mar. Sci.* 51, 263-272.
- Le Fur J., 1995. Représentation de la complexité des exploitations halieutiques, pp. 167-185, *In* : Laloë F., H. Rey & J.L. Durand (éds), Questions sur la dynamique de l'exploitation halieutique, Série Colloques et séminaires, ORSTOM Editions, 542 p.
- Murawski S.A., A.M. Lange, M.P. Sissenwine & R.K. Mayo, 1983. Definition and analysis of multispecies otter-trawl fisheries off the northeast coast of the United States. *J. Cons. int. Explor. Mer* 41, 13-27.
- Pech N. & F. Laloë, 1997. Use of principal component analysis with instrumental variables (PCAIV) to analyse fisheries catch data. *ICES J. Mar. Sci.* 54, 32-47.
- Pelletier D. & J. Ferraris, 1999. Defining fishermen exploitation tactics from the analysis of logbooks and commercial sampling data. *Can. J. Fish. Aquat. Sci.* (accepté)
- Robson D.S., 1966. Estimation of the relative fishing power of individual ships. *Int. Comm. Northwest Atl. Fish. Res. Bull.* 3, 5-14.
- Rochet M.J. & J.L. Durand, 1995. Dynamique à moyen terme des flottilles artisanales du Mor-Braz, pp. 331-352, *In* : Laloë F., H. Rey & J.L. Durand (éds), Questions sur la dynamique de l'exploitation halieutique, Série Colloques et séminaires, ORSTOM Editions, 542 p.
- Rogers J.B. & E.K. Pikitch, 1992. Numerical definition of groundfish assemblages caught off the coasts of Oregon and Washington using commercial fishing strategies. *Can. J. Fish. Aquat. Sci.* 49, 2648-2656.
- Stocker M. & D. Fournier, 1984. Estimation of relative fishing power and allocation of effective fishing effort, with catch forecasts, in a multi-species fishery. *Bull. Int. North Pac. Fish. Comm.* 42, 3-9.

LE GAM, UN OUTIL D'ESTIMATION DES BIOMASSES LOCALES. APPLICATION AU THON ALBACORE (*Thunnus albacares*) DE L'ATLANTIQUE.

Olivier MAURY⁽²⁾, Laurent MILLISCHER⁽¹⁾, Didier GASCUEL⁽¹⁾, Alain FONTENEAU⁽²⁾.

(1) ENSAR halieutique, 65 route de S^t Brieuç, F35042 Rennes cedex France.

(2) ORSTOM HEA, 911 av. Agropolis B.P. 5045, F34032 Montpellier cedex 1 France

résumé :

L'usage de données de CPUE (captures par unité d'effort de pêche) issues de la pêche commerciale comme indices d'abondance locaux de la ressource demande d'identifier les variations potentielles de la probabilité de capture des poissons par différents engins (capturabilité). Dans cette optique, on cherche à identifier et à caractériser l'influence de différents facteurs sur la capturabilité des thons albacore (*Thunnus albacares*) de l'Atlantique aux flottilles de senneurs et de palangriers. On utilise des modèles additifs généralisés (GAM) pour étudier les relations non linéaires entre l'efficacité des navires de pêche et différents facteurs liés à la pêche et à l'environnement. On met en évidence des effets de l'effort de pêche et de la profondeur de la thermocline. Les relations obtenues sont discutées et interprétées.

abstract :

The use of commercial fishing CPUE data (catches per unit of effort) as local abundance indices requires the estimation of potential variations of the catchability by gear. In that purpose, we try to identify and to characterize the effects of various factors on atlantic yellowfin tunas (*Thunnus albacares*) catchability by both purse seiners and long liners fleets. We use generalized additive models (GAM) to study non linear relationships between fishing ships efficiency and different factors linked either to fishing activity or to the hydrological environment. We bring to the fore fishing effort and thermocline depth effects. Obtained relationships are discussed and interpreted.

INTRODUCTION

Les thons albacore (*Thunnus albacares*) sont répartis dans l'ensemble de l'Atlantique tropical. L'étude de la distribution spatiale de telles espèces à larges aires de répartition ne peut, aujourd'hui, se faire qu'en utilisant des données de captures commerciales. On utilise alors le rendement des navires (les captures par unité d'effort CPUE) comme une mesure de l'abondance locale des poissons. Le problème est que la capturabilité q (la probabilité pour une unité de biomasse d'être capturée par une unité d'effort de pêche) liant CPUE et biomasses locales ($CPUE=q.B$) est extrêmement variable, en fonction des tactiques utilisées par les pêcheurs pour localiser les poissons, et des conditions hydrologiques locales.

Dans cette étude, on analyse les relations non linéaires liant les CPUE des flottilles de senneurs français et de palangriers japonais aux efforts de pêche et à la profondeur de la thermocline (la thermocline est une zone de variation thermique brutale séparant les eaux de surface chaudes des eaux profondes plus froides). Pour caractériser ces relations sans hypothèse de linéarité, on utilise les techniques de modélisation GAM (Generalized Additive Models) (Hastie et Tibshirani, 1990).

MATERIEL ET METHODES

• Les données environnementales

Météo-France nous a fourni les données hydrologiques. Elles sont issues de simulations du modèle OPA7.1 (Delecluse *et al.*, 1993) développé au LODYC (Laboratoire d'Océanographie DYnamique et de Climatologie, Paris VI) dans le cadre du projet OPERA (Morlière, 1989 ; Morlière *et al.*, 1989). Le modèle est forcé par des champs de vents observés et simulés et il assimile des observations ponctuelles de paramètres océanographiques comme des profils thermiques. Dans ces conditions, un tel modèle numérique peut être considéré comme un très puissant interpolateur spatio-temporel d'observations océanographiques. Les données simulées couvrent l'ensemble de l'atlantique tropical de 20° de latitude nord à 20° de latitude sud. Le modèle n'est pas adapté à la circulation des eaux côtières pour lesquelles il ne fournit pas de données. Les frontières côtières du modèle correspondent donc aux frontières du plateau continental. La résolution spatiale est une grille horizontale régulière de 0,5°x 0,5° sur 16 niveaux bathymétriques. La résolution temporelle est de 5 jours de 1980 à 1995.

Les paramètres utilisés dans la présente étude sont la température de l'eau en degrés Celsius (à 5m de profondeur assimilée à la SST) et la profondeur de l'isotherme 20°C en mètres (Z20) qui est généralement utilisée pour approximer la profondeur de la thermocline dans l'océan Atlantique tropical. L'étude d'autres paramètres (Maury, 1998 ; Maury et Gascuel, soumis) n'est pas présentée ici.

• Les données de pêche

La Commission Internationale pour la Conservation des Thons de l'Atlantique (ICCAT) centralise les données statistiques concernant la plupart des pêcheries thonières dans l'ensemble de l'océan Atlantique. Dans cette étude, nous avons utilisé des données de CPUE en provenance de deux flottilles très différentes : la flottille de senneurs FIS (Franco-Ivoiro-Sénégalais) pendant la période 1980-1991 et la flottille palangrière japonaise pendant la période 1980-1993.

Ces deux flottilles ont été choisies pour leurs avantages respectifs :

- Les senneurs FIS visent principalement l'albacore (*Thunnus albacares*), le patudo (*Thunnus obesus*) et le listao (*Katsuwonus pelamis*) sur une grande étendue géographique, mais seulement dans l'Atlantique tropical Est. Pendant la période 1980-1991, cette flottille a gardé un diagramme d'exploitation relativement stable couvrant presque tout le spectre de taille des albacores de 30cm (âge 0) à 180cm (âge 5+) (Anon., 1997). Pour cette flottille, les données de CPUE ont été calculées par âge (0 à 5+) à l'aide de la méthode de décomposition polymodale des « clefs ajustées » (Gascuel, 1994) étalonnée avec la courbe de croissance estimée par Gascuel *et al.* (1992). La conversion des tailles en âges à été réalisée mensuellement. Les CPUE FIS par âge sont exprimées pour les six groupes d'âge en nombre de poissons par heure de recherche ($CPUE_{\text{âge, zone, période}} = \text{captures en nombre}_{\text{âge, zone, période}} / \text{effort}_{\text{zone, période}}$). Leur résolution spatiale est de 1° de latitude x 1° de longitude et leur résolution temporelle est 15 jours pour chacune des 12 années de 1980 à 1991.

- La flottille palangrière japonaise couvre l'ensemble de l'océan Atlantique, recouvrant totalement et excédant probablement l'aire de répartition des albacores. Les palangriers japonais pêchent en eau profonde (de 100 à 150m de profondeur pour les palangres traditionnelles pendant les années 80, à 300m de profondeur pour les palangres profondes des années 90) et visent principalement des patudos adultes et des thons rouges (*Thunnus tyngnus*). Les albacores adultes constituent des captures accessoires importantes pour cette flottille pendant la période considérée (Anon., 1997). Leur diagramme d'exploitation est resté stable pendant la période 1980-1993, mais comme ils ne capturent que des albacores âgés dont la taille est comprise entre 110cm and 170cm (âge 4 et 5+), nous n'avons pas cherché à éger les captures. Les CPUE palangrières japonaises sont exprimées en tonnes par millier d'hameçons ($CPUE_{\text{zone, période}} = \text{captures en poids}_{\text{zone, période}} / \text{effort}_{\text{zone, période}}$). Leur résolution spatiale est 5°x5° et leur résolution temporelle est 30 jours.

• Choix des méthodes GAM

Différentes méthodes statistiques permettent la modélisation du lien entre une variable à expliquer et plusieurs variables explicatives. Schématiquement, elles peuvent être classées dans deux grandes catégories fondamentales (Causeur, 1998) (Fig. 1) : les méthodes paramétriques pour lesquelles la

relation recherchée est spécifiée de manière paramétrique *a priori* et les méthodes non paramétriques pour lesquelles elle n'est pas spécifiée *a priori*¹. Les méthodes paramétriques peuvent être soit linéaires comme les méthodes de régression linéaire dans le cadre gaussien ou linéaires généralisées à différentes lois de distributions des erreurs (GLM), soit non linéaires comme les méthodes de régression non linéaire par exemple. Parmi les méthodes non paramétriques, les méthodes additives sont fondées sur une interpolation linéaire des données par lissage non paramétrique avec hypothèse de distribution gaussienne des erreurs, les modèles GAM en sont une généralisation telle que la distribution du terme d'erreur puisse être choisie parmi l'ensemble des lois de distribution de la famille exponentielle (Hastie et Tibshirani, 1990). Les méthodes non paramétriques non additives comprennent, entre autre, les réseaux de neurones.

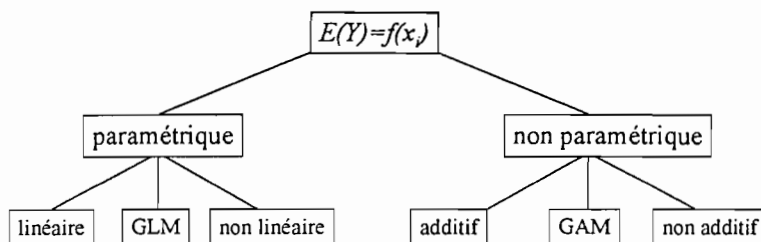


Fig. 1. : Classification schématique des 6 grandes catégories de méthodes de modélisation du lien entre variables explicatives et variable à expliquer (Causeur, 1998).

Nous ne connaissons rien *a priori* de la forme des relations que nous recherchons entre l'abondance des thons et l'environnement. Le cadre non paramétrique est donc adapté. Dans cette catégorie, le caractère paramétrique de l'hypothèse de distribution des erreurs des modèles additifs généralisés (GAM) permet l'usage de tests statistiques classiques. C'est donc les GAM que nous avons choisi d'utiliser pour modéliser les relations albacore-environnement.

Ces méthodes statistiques récentes commencent à être de plus en plus fréquemment utilisées en halieutique (Mendelsohn et Cury, 1987 ; Cury et Roy, 1989 ; Cury *et al.*, 1995 ; Swartzman *et al.*, 1992 ; Swartzman *et al.*, 1994 ; Swartzman *et al.*, 1995 ; Jacobson et MacCall, 1995 ; Welch *et al.*, 1995). Mais, à notre connaissance, elles n'ont jamais été appliquées à l'estimation de l'abondance locale à partir de CPUE commerciales, ni à l'étude des relations thons-environnement.

• Principe des GAM

Les modèles additifs généralisés sont des généralisations non paramétriques des régression linéaires multiples (Hastie et Tibshirani, 1990). Les méthodes GAM s'appuient sur des transformations empiriques des variables explicatives par des techniques de lissage locaux². La formulation générale d'un modèle additif est la suivante :

$$S(Y_j) = \sum_i T_i(X_i(j)) + \varepsilon(j)$$

avec Y_j la variable expliquée pour l'observation j ; X_i les variables explicatives ; S , une fonction paramétrique liée à l'extension du cadre gaussien à la famille exponentielle appelée fonction de lien ; T_i

¹ Le terme non paramétrique concerne ici la nature du lien entre variables explicatives et variable expliquée et pas le terme d'erreur (Hastie et Tibshirani, 1990). Comme l'écrit Gauthiez (1997) à propos des méthodes de lissage non paramétriques, « on peut en même temps formuler des hypothèses paramétriques en ce qui concerne la loi des erreurs. Il ne faut pas confondre avec les modèles semi-paramétriques ou non paramétriques au sens habituel de la statistique, où le caractère non paramétrique ou semi-paramétrique porte sur la modélisation des erreurs. »

² Parmi les différentes méthodes de lissage possibles, les méthodes de moyenne mobile dites « à noyau », la méthode des droites mobiles (loess), les approximations polynomiales type spline, sont les plus utilisées.

sont des fonctions non spécifiées estimées de manière non paramétrique par des techniques de lissages locaux et ε est une variable aléatoire dont la loi de distribution doit être choisie dans la famille exponentielle.

Dans les GAM, la loi de distribution de l'erreur peut être choisie dans toutes les lois de la famille exponentielle (qui comprend la loi normale, la loi de Poisson, la loi binomiale, la loi gamma, ...). Dans notre application, après examen des distributions, nous avons fait l'hypothèse d'une distribution normale de l'erreur. Dans ce cas, la fonction de lien S est simplement la fonction identité (Hastie et Tibshirani, 1990 ; Swartzman *et al.*, 1994). L'espérance de la variable Y conditionnellement aux variables X_i peut donc s'exprimer comme une somme de fonctions de lissage des variables explicatives :

$$E(Y / X_i) = \sum_i T_i(X_i)$$

Une transformation logarithmique a été utilisée d'une part pour symétriser la distribution dissymétrique des fréquences de CPUE et, d'autre part, en raison de la nature *a priori* multiplicative des processus considérés. En effet, pour pouvoir interpréter les CPUE comme des indices d'abondance locaux (des densités D), les effets liés à la capturabilité (q) doivent être identifiés et retirés. Ainsi, conformément à la formule générale $CPUE = q \cdot D$, la forme générale des GAM que nous avons ajustés est la suivante :

$$\ln(CPUE + 1) = \sum_i T_i(X_i) + \sum_j T_j(X_j)$$

avec X_i les variables explicatives liées à la capturabilité et X_j les variables liées à l'abondance locale des albacores.

Pour estimer les fonctions T , nous avons utilisé un lisseur spline cubique implémenté dans le logiciel S-plus (© AT&T Bell Laboratories).

• Mise en oeuvre des GAM

Un grand nombre de modèles a été testé pour mettre en évidence le rôle des variables environnementales sur les densités locales d'albacores. Une approximation du test F permettant de tester la significativité de la contribution de chaque variable (Hastie et Tibshirani, 1990) a été systématiquement calculée. Dans chaque modèle, toutes les variables supposées être liées à la capturabilité ont été systématiquement prises en compte quand elles étaient significatives, associées à la variable environnementale étudiée. Des effets d'interaction avec d'autres variables environnementales ont été également ajoutés quand ils modifiaient la relation considérée. Un pseudo coefficient R^2 , défini comme la fraction de la déviance totale expliquée par le modèle, a été calculé systématiquement pour tous les modèles testés (Swartzman *et al.*, 1992). Bien que n'ayant pas grand sens dans l'absolu, ce coefficient permet de comparer différents modèles ayant la même structure. Une analyse de déviance a, enfin, été systématiquement réalisée pour avoir une idée de la part de l'inertie du modèle expliquée par chaque facteur.

- L'analyse des données de CPUE des senneurs FIS a été menée à la résolution spatio-temporelle de $1^\circ \times 1^\circ \times 15$ jours. Les données environnementales issues du modèle OPA7 ont été agrégées à cette échelle à l'aide d'une moyenne pondérée par la surface ; à l'inverse, les données CZCS ont été désagrégées à cette échelle. Le fichier de données obtenu contient 35725 observations. Le modèle ajusté a la forme générale suivante :

$$\begin{aligned} \ln(CPUE + 1) &= s(\text{année}) + s(\text{quinzaine} * \text{zone ICCAT}) + s(\text{effort}) + s(\text{vent}) + s(CI) + s(Z20) \\ &+ s(gZ20) + s(\text{variable environnementale}) + s(\text{effets d'interaction avec d'autres variables}) \\ &= A + s(\text{variable environnementale}) + s(\text{effets d'interaction avec d'autres variables}) \end{aligned}$$

A est la somme des transformations des variables supposées être liées à la capturabilité des albacores aux senneurs FIS. S est la fonction de lissage spline.

L'effet année est supposé prendre en compte les fluctuations de la biomasse totale et l'évolution tendancielle potentielle de la puissance de pêche qui a déjà été observée pour les pêcheries d'albacores de l'Atlantique (Gascuel *et al.*, 1993). Le terme (quinzaine * zone) mesure un effet spatio-saisonnier moyen sur la période considérée. Il permet de donner un poids égal à chaque zone et période, quelle que soit la répartition de la flottille de pêche.

Pour chaque effet considéré, une analyse générale sur l'ensemble du jeu de données et une analyse par zone ont été réalisées. Une analyse par âge a également été réalisée en considérant les 6 groupes d'âge usuels utilisés habituellement par le comité scientifique de l'ICCAT (âge 0 à 5+).

- L'analyse des CPUE des palangriers japonais a été menée à une résolution spatio-temporelle de 5°x5°x30 jours. Les données environnementales issues du modèle OPA7 ont été agrégées à cette échelle à l'aide d'une moyenne pondérée par la surface et les données CZCS ont été désagrégées à cette échelle. Le fichier de données obtenu contient 5035 observations. Le modèle ajusté a la forme générale suivante :

$$\begin{aligned} \ln(\text{CPUE} + 1) &= s(\text{année}) + s(\text{latit.} * \text{longit.} * \text{mois}) + s(\text{effort}) + s(\text{Z20}) \\ &+ s(\text{variable environnementale}) + s(\text{effets d'interaction avec d'autres variables}) \\ &= B + s(\text{variable environnementale}) + s(\text{effets d'interaction avec d'autres variables}) \end{aligned}$$

avec *B*, la somme des transformations des variables supposées être liées à la capturabilité des albacores aux palangriers japonais et *s* les fonctions de lissage spline.

Seules les relations concernant l'effort de pêche et la profondeur de la thermocline sont analysées ici. Les relations concernant d'autres variables, bien qu'incorporées aux modèles étudiés, ne sont pas détaillées ici.

RESULTATS

• Relation CPUE-effort

Pour les senneurs FIS comme pour les palangriers japonais, l'effort de pêche est systématiquement hautement significatif et explique la plus importante partie de la variabilité des CPUE pour tous les modèles considérés (93% pour les senneurs). La relation générale entre $\ln(\text{CPUE}+1)$ et l'effort de pêche est non linéaire pour les deux flottilles. Pour les senneurs, la courbe croît rapidement pour les valeurs d'effort inférieures à un seuil de l'ordre de 500 heures de pêche. Au delà, soit la courbe continue à croître plus lentement pour les poissons d'âge 0, 1, 2 et 3, soit elle décroît pour les poissons d'âges 4 et 5+ (Fig. 2).

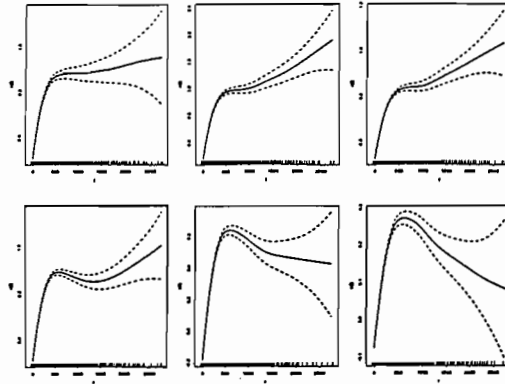


Fig. 2. : Résultats de la régression GAM (trait continu) de $\ln(\text{CPUE}+1)$ des senneurs FIS en fonction de l'effort de pêche (en heures de pêche) pour les 6 groupes d'âge (0 à 5+ de gauche à droite et de haut en bas) avec l'intervalle de confiance ($\alpha=0.05$) en pointillé. Les tirets sur l'axe des x localisent les données.

Pour les palangriers, la courbe est continûment décroissante avec une forme exponentielle négative (Fig. 3).

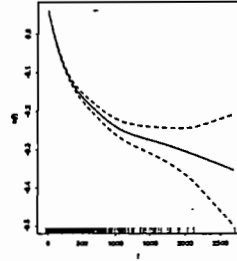


Fig. 3. : Résultats de la régression GAM (trait continu) de $\ln(\text{CPUE}+1)$ des palangriers japonais en fonction de l'effort de pêche (en milliers d'hameçons) avec l'intervalle de confiance ($\alpha=0.05$) en pointillé. Les tirets sur l'axe des x localisent les données.

• Relation CPUE/thermocline :

Pour les senneurs FIS, la profondeur de la thermocline a un effet clairement négatif sur les CPUE pendant la période considérée (Fig. 4). Plus la thermocline est profonde, plus faibles sont les rendements. Les CPUE sont supérieures à la moyenne quand la profondeur de la thermocline est inférieure à 50m et elles sont plus faibles au-delà. Un tel effet est constant avec l'âge mais il explique une part plus importante de la variance chez les vieux poissons (âges 3, 4 and 5+) que chez les jeunes (âges 1 and 2).

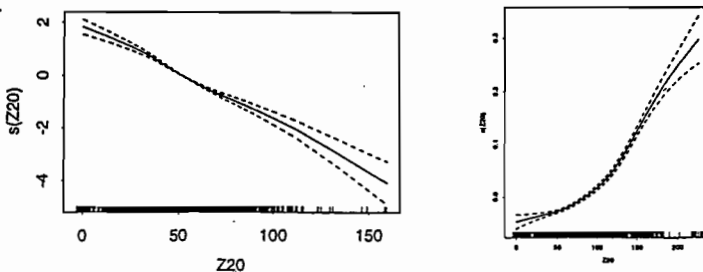


Fig. 4. : Résultats de la régression GAM (trait continu) de $\ln(\text{CPUE}+1)$ des senneurs FIS en fonction de la profondeur de la thermocline (en m) avec l'intervalle de confiance ($\alpha=0.05$) en pointillé. A gauche, les senneurs FIS et à droite, les palangriers japonais. Les tirets sur l'axe des x localisent les données.

Il est important de prendre en compte les effets d'interaction entre la profondeur de la thermocline et la température de surface car ces deux facteurs ne sont pas indépendants : plus la thermocline est proche de la surface, plus la SST est basse et réciproquement. Sans l'effet d'interaction avec la température, la relation entre les CPUE et la profondeur de la thermocline n'est pas clairement discernable.

Pour les palangriers japonais qui pêchent plus profondément que les senneurs, la profondeur de la thermocline a un effet monotone positif (Fig. 4) et explique une part importante de la variance du modèle (17,4%).

DISCUSSION : l'utilisation des CPUE comme un indice d'abondance local

La CPUE est l'indice d'abondance le plus fréquemment utilisé en halieutique. Pourtant, son emploi est délicat et pose des problèmes dus à la non linéarité de la relation CPUE/abondance. Un certain nombre de phénomènes bien connus sont à l'origine de cette non linéarité. Ils sont en général liés à l'hétérogénéité spatiale de la ressource et au comportement de recherche des navires (Hilborn et Walters, 1992 ; Gauthiez, 1997). Dans le cas des pêcheries thonières, l'utilisation des CPUE comme indices d'abondance locaux est encore compliquée par des phénomènes spécifiques responsables d'une non linéarité et d'une forte variabilité de la relation CPUE/effort locale :

- l'accroissement d'efficacité des flottilles de senneurs quand leur effort de pêche augmente ;
- les très forts taux d'exploitation locaux qui entraînent des phénomènes de surexploitation locale (Fonteneau *et al.*, 1997).
- les importantes variations spatio-temporelles de capturabilité dues à l'environnement et aux comportements migratoires et agrégatifs saisonniers.

Nos résultats aident à préciser ces différentes sources de biais des CPUE thonières.

• Interprétation des relations CPUE/effort

Il est vraisemblable que quand l'effort des senneurs augmente (*i.e.* le nombre d'heures de recherche), la puissance locale des navires s'accroît car les bateaux coopèrent (ou s'espionnent) pour repérer les mattes de thon. Elle atteindrait ensuite un palier quand le nombre total d'heures de recherche de l'ensemble des bateaux est suffisant pour repérer presque à tous les coups les plus fortes concentrations de thons de la zone considérée. Ce seuil serait d'environ 500 heures de recherche par carré de 1 degré pendant 15 jours (Fig. 2). Pour les jeunes poissons, très nombreux dans le milieu, les taux d'exploitation locaux demeurent faibles malgré l'augmentation de l'effort et les CPUE continuent à augmenter quand l'effort augmente. Pour les poissons âgés moins nombreux dans le milieu, l'augmentation de puissance locale entraîne un phénomène de surexploitation locale (Fig. 5) et les CPUE des navires tendent alors à décroître.

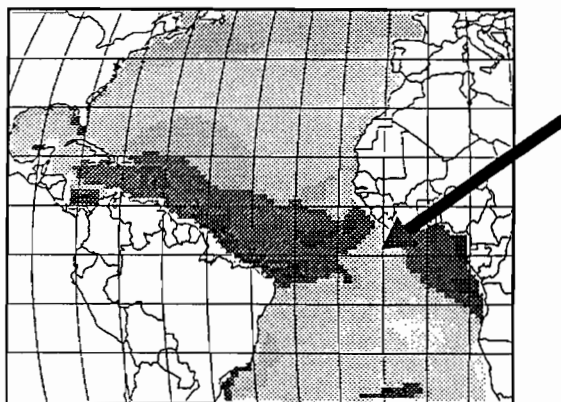


Fig .5. : Simulation par un modèle d'advection-diffusion forcé par l'environnement (Maury, 1998) de la répartition des albacores adultes (âge 5+ en février 1991) prenant en compte les captures réalisées par les senneurs FIS. Les très faibles concentrations observées au large du Golfe de Guinée (où est exercée une forte pression de pêche) sont dues au phénomène de surexploitation locale des albacores âgés.

De la même manière, les palangriers épuisent vraisemblablement localement leur zone de pêche quand ils sont nombreux et leurs rendements chutent quand leur effort local augmente. La diminution locale d'abondance qu'ils provoquent peut être compensée par l'arrivée de poissons par diffusion. Ces poissons proviennent alors d'une fraction « cryptique » de la biomasse (Fonteneau *et al.*, 1997) localisée dans des zones non exploitées (ou non exploitables).

La combinaison des deux facteurs antagonistes, coopération et surexploitation locale, est vraisemblablement responsable des relations non linéaires observées entre les CPUE et l'effort de pêche.

Dans cette optique, Maury (1998) et Maury et Gascuel (soumis) ont étudié à l'aide du simulateur SHADYS³ la non linéarité de la relation CPUE/effort pour des pêcheries thonières (Fig. 6). Ils ont montré qu'une telle relation non linéaire peut être, effectivement, interprétée en terme de stratégie de répartition et de coopération entre navires pour la partie ascendante de la courbe et de surexploitation locale pour la partie décroissante (Fig. 7).

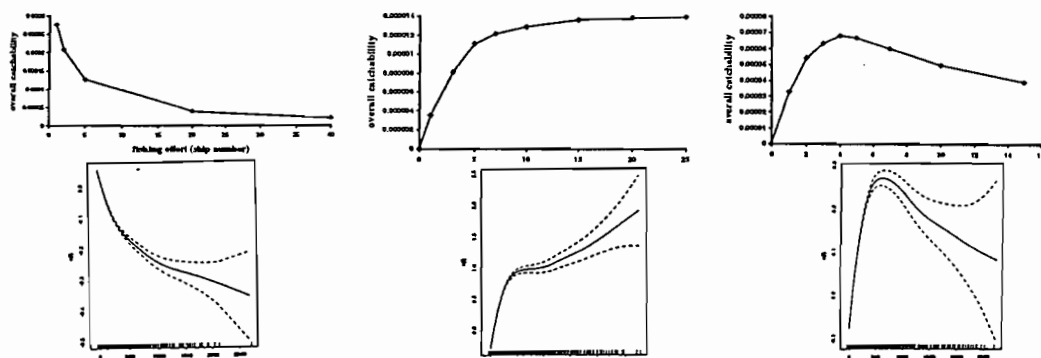


Fig. 6. : Comparaison des relations observées et simulées sous SHADYS (Maury, 1998 ; Maury et Gascuel, soumis). En haut, relations simulées dans SHADYS (voir texte). En bas, relations CPUE/effort obtenues par analyse GAM des CPUE (cf. quatrième partie). A gauche, courbe obtenue pour les palangriers japonais. Au milieu, CPUE des senneurs concernant les jeunes poissons. A droite, CPUE des senneurs concernant les albacores âgés.

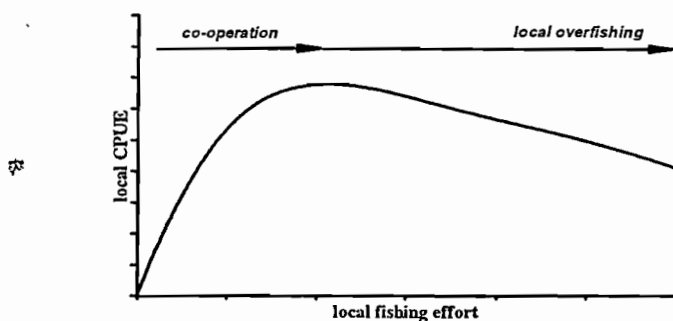


Fig. 7. : Relations théoriques entre CPUE locales et effort de pêche. La non linéarité de la relation est interprétée comme résultant de deux processus antagonistes : d'une part la coopération entre les bateaux (tactique) et la saisonnalité des choix de zones de pêche (stratégie) qui conduisent à une augmentation de l'efficacité de pêche et, d'autre part, la diminution locale de la biomasse qui explique la décroissance des rendements locaux.

³ Simulateur HALieutique de DYnamiques Spatiales fondé sur le couplage d'un modèle de l'environnement, un modèle d'advection-diffusion et un modèle de comportement des pêcheurs.

• Influence de la profondeur de la thermocline sur l'efficacité des flottilles

Pour les senneurs, il est bien établi que les facteurs environnementaux modifient la capturabilité (Fonteneau et Marcille, 1988 ; Marsac, 1992). Les résultats obtenus confirment une telle hypothèse et contribuent à préciser le rôle fonctionnel de la profondeur de la thermocline sur la capturabilité des albacores aux senneurs FIS pendant la période 1980-1991. Plus la thermocline est profonde et son gradient faible, plus les niveaux de capturabilité observés sont faibles. Un tel effet négatif de l'enfoncement de la thermocline sur la capturabilité est observé dans l'océan Pacifique par Inada *et al.* (1997) pour des profondeurs de la thermocline supérieures à 50m. Ce phénomène semble être dû au fait que la thermocline agit comme une barrière pour les albacores, particulièrement pour les jeunes qui ne sont pas encore capables de plonger rapidement dans les eaux froides (Fig. 8). Quand la thermocline est profonde ou faiblement marquée, les thons s'échappent facilement des sennes tournantes alors que quand elle est proche de la surface et très marquée, les thons sont pris au piège et ils ne peuvent s'échapper du filet (Sharp et Dizon, 1978 ; Evans *et al.*, 1981 ; Marsac, 1992 ; Fonteneau, 1996).

Un tel phénomène est probablement responsable, entre autre, des très faibles niveaux de captures réalisés au premier trimestre de 1984. En effet, cette année là, une anomalie environnementale comparable au phénomène *el Niño* a entraîné de faibles rendements des senneurs probablement liés à un enfoncement de la thermocline (Fonteneau, 1991). Le phénomène de baisse de la capturabilité semble être plus marqué quand la profondeur de la thermocline dépasse 60-70m, profondeur qui correspondait à la chute des sennes utilisées par les senneurs FIS pendant la période considérée. Les choses sont probablement différentes aujourd'hui, puisque les senneurs FIS utilisent des sennes beaucoup plus profondes (150 à 200 mètres de chute).

Pour les palangriers, la capturabilité des albacores est, également, manifestement liée à la profondeur de la thermocline, mais l'effet est inverse. En effet, plus la thermocline est profonde, plus les eaux profondes sont chaudes, plus les albacores peuvent descendre en profondeur et être accessibles aux palangres mouillées en profondeur pour capturer des patudos qui chassent en profondeur (Fig. 8).

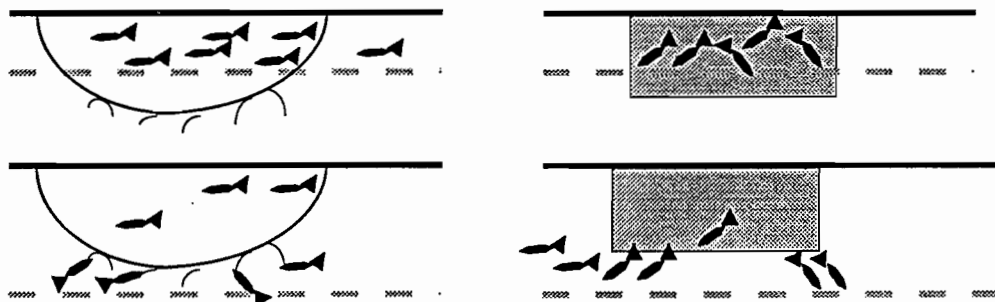


Fig. 8. : Schématisation de l'importance de l'effet « barrière » de la thermocline (en tiret gris) et de ses conséquences sur la capturabilité des albacores selon que la thermocline est proche de la surface (en haut) ou profonde (en bas). A gauche, la pêche à la palangre et à droite, la pêche à la senne tournante.

CONCLUSION

Malgré leurs limitations, les données provenant de pêcheries commerciales sont bien souvent le seul moyen d'estimer la répartition des thons à grande échelle. Pour cela, l'identification des relations entre la capturabilité et les facteurs liés à l'activité de pêche ou liés à l'environnement, est une étape nécessaire à l'utilisation des CPUE comme indices d'abondance locaux. Dans cette optique, les modèles GAM, permettant l'identification de telles relations non linéaires, constituent des outils extrêmement intéressants pour la recherche halieutique.

BIBLIOGRAPHIE

- Anon., 1997. Report of the working group on tropical tunas. *Rec. Doc. Scient. SCRS, ICCAT* Madrid.
- Causeur D., 1998. Modélisation des liens entre variables. Cours de statistique. Spécialisation d'écologie quantitative. ENSAR.
- Cury P. et C. Roy 1989. Optimal environmental window and pelagic fish recruitment success in upwelling areas. *Can. J. Fish. Aquat. Sci.* 46, 670-680.
- Cury P., C. Roy, R. Meridelsohn, A. Bakun, D.M. Usby et R.H. Parrish, 1995. Moderate is better : exploring nonlinear climatic effects on the Californian anchovy (*Engraulis mordax*). *Can. J. Fish. Aquat. Sci.*
- Delecluse P., M. Imbard, C. Lévy, G. Madec, 1993: OPA Ocean General Circulation Model, reference manual, 93/03.
- Evans R.H., D.R. McLain et R.A. Bauer, 1981. Atlantic Skipjack Tuna : Influences of Mean Environmental Conditions on Their Vulnerability to Surface Fishing Gear. *Marine Fisheries Review* 43(6), 1-11.
- Fonteneau A., 1991. La surexploitation du stock d'albacore en 1984 : mythe ou réalité ? *Rec. Doc. Scient. ICCAT. Vol. XXXVI*, pp 348-379.
- Fonteneau A., 1996. Introduction aux problèmes des relations thons-environnement dans l'Atlantique. In : proceedings of the ICCAT Symposium, Les Açores, Juin 1996. Sous presse.
- Fonteneau A. et J. Marcille, 1988. *Ressources, pêche et biologie des thonidés tropicaux de l'Atlantique Centre-Est*. FAO document technique sur les pêches. Fonteneau A. Et Marcille J. Eds. 391p.
- Fonteneau A., D. Gascuel, P. Pallares Soubrier 1997. Vingt cinq ans d'évaluation des ressources thonnières dans l'Atlantique : quelques réflexions méthodologiques. in : Ouvrage du Symposium ICCAT, Les Açores, Juin 1996. A paraître.
- Gascuel D., 1994. Une méthode simple d'ajustement des clefs taille/âge : application aux captures d'albacores (*Thunnus albacares*) de l'Atlantique est. *Can. J. Fish. Aquat. Sci.* 51:723-733.
- Gascuel D., A. Fonteneau et C. Capisano, 1992. Modélisation d'une croissance en deux stances chez l'albacore (*Thunnus albacares*) de l'Atlantique est. *Aquat. Living Resour.*, 5(3), 155-172.
- Gascuel D., A. Fonteneau et E. Foucher 1993. Analyse de l'évolution des puissances de pêche par l'analyse des cohortes : application aux senneurs exploitant l'albacore (*Thunnus albacares*) dans l'Atlantique est. *Aquat. Living Resour.*, 6(1), 15-30.
- Gauthiez F. 1997. Structuration spatiale des populations de poissons marins demersaux. Caractérisation, conséquences biométriques et halieutiques. Thèse de doctorat. ENGREF.
- Hastie T. and R. Tibshirani, 1990. *Generalized Additive Models*. Chapman and Hall, London. 335p.
- Hilborn R., C. J. Walters 1992. Quantitative fisheries stock assessment. Choice, dynamics and uncertainty. Chapman and Hall, 570p.
- Jacobson L.D. et A.D. MacCall, 1995. Stock-recruitment models for Pacific sardine (*Sardinops sagax*). *Can. J. Fish. Aquat. Sci.*, 52, 566-577.
- Inada H., J. Sekine, H. Kim, M. Nemoto, S. Takeuchi, M. Kagoshi, Y. Anzai et K. Yabuki, 1997. The influence of environmental conditions on fish catches during purse seining operations for skipjack *Katsuwonus pelamis* and yellowfin tuna *Thunnus albacares* in the western tropical Pacific fishing grounds. *Tokyo Suisandai Kempo*, 83, 129-138.
- Marsac F. 1992. Etude des relations entre l'hydroclimat et la pêche thonnière haiturrière tropicale dans l'Océan Indien. Thèse de doctorat. Université de Bretagne Occidentale. 353p.
- Maury 1998. Modélisation spatiale en halieutique. Approche par simulateur sous SIG. Application à la modélisation hiérarchique de la population de thons albacore (*Thunnus albacares*) de l'Atlantique. Thèse de doctorat de l'ENSAR. 350p.
- Maury and Gascuel (soumis). The local overfishing concept. Simulations in SHADYS. *Can. J. Fish. Aquat. Sci.* Soumis.
- Mendelsohn R. et P. Cury 1987. Fluctuations of a fortnightly abundance index of the Ivoirian coastal pelagic species and associated environmental conditions. *Can. J. Fish. Aquat. Sci.* 44: 408-428.
- Morliere A., 1989: Climatologie mensuelle d'une simulation de l'océan Atlantique tropical, *Rapp Int. LODYC* 89/07, 76p.
- Morliere A., P. Delecluse, P. Andrich et B. Camusat, 1989: Une évaluation des champs thermiques simulés par un modèle de circulation générale océanique dans l'Atlantique tropical, *Oceanol. Acta.*, 12(1), 9-22.
- Sharp G.D. et A.E. Dizon 1978. The physiological ecology of tunas. Sharp and Dizon (eds.). Academic Press, New York: 485p.
- Swartzman G., E. Huang and S. Kaluzny, 1992. Spatial Analysis of Bering Sea Groundfish Survey Data Using Generalized Additive Models. *Can. J. Fish. Aquat. Sci.* 49: 1366-1378.
- Swartzman G., W. Stuetzle, K. Kulman et M. Powojowski, 1994. Relating the distribution of pollock schools in the Bering Sea to environmental factors. *ICES J. Mar. Sci.*, 51 : 481-492.
- Swartzman G., E. Silverman, N. Williamson, 1995. Relating trends in walleye pollock (*Theragra chalcogramma*) abundance in the Bering Sea to environmental factors. *Can. J. Fish. Aquat. Sci.* 52 : 369-380.
- Welch D.W., A.I. Chigirinsky et Y. Ishida 1995. Upper thermal limits on the oceanic distribution of Pacific salmon (*Oncorhynchus spp.*) in the spring. *Can. J. Fish. Aquat. Sci.*, 52, 489-503.

Représentation Simulée d'un Processus de Recherche et de Capture de Bancs de Poissons
Répartis selon un Processus Ponctuel Marqué

Pierre Petitgas ⁽¹⁾ et Francis Laloë ⁽²⁾

⁽¹⁾ IFREMER, BP 21105, 44311 cedex3, Nantes

⁽²⁾ IRD, BP 5045, 34032 cedex 1. Montpellier

1. Introduction

Les résultats obtenus par une unité de pêche lors d'une sortie peuvent être décrits par de nombreuses variables, telles que la prise par sortie, le nombre de coups de pêche, le temps de recherche, la prise par coups de pêche, la prise par temps de mer etc... On peut, à partir de la distribution de chacune de ces variables tirer des indicateurs porteurs d'information sur l'état de la ressource, la nature de sa distribution spatiale etc... Mais il est évident que ces distributions sont le résultat d'une interaction dépendant à la fois du processus humain d'exploitation (recherche et capture) et du processus biologique de répartition spatiale de la ressource.

Les relations entre la capture et la biomasse dépendent de la capturabilité du poisson lors de l'opération de pêche, de paramètres de la structure spatiale du poisson (agrégation et répartition spatiale) et enfin de la stratégie de pêche mise en oeuvre. A l'échelle globale population.année, la théorie classique des pêches s'appuie sur l'hypothèse d'une capture proportionnelle à la biomasse. Or, cette hypothèse n'est pas vérifiée pour les poissons dont la distribution spatiale est agrégée et particulièrement pour ceux qui forment des bancs (Paloheimo and Dickie 1964) mais aussi dans les cas où la répartition de la biomasse est densité-dépendante (MacCall 1990).

Alors, quel formalisme proposer pour rendre compte de la combinaison entre le processus de répartition de la ressource et celui de recherche et capture ? Paradoxalement, cette question a très peu été explorée en halieutique. Mangel et Beder (1985) ont proposé un modèle à petite échelle d'espace et de temps de distribution au hasard pour les poissons et de recherche au hasard pour les pêcheurs. Gauthiez (1997) a proposé une représentation basée sur la combinaison de lois où le processus de capture est paramétré par un échantillonnage de la loi des densités de poissons. Ces représentations ne tiennent toutefois pas compte soit de l'agrégation des poissons soit de la continuité spatiale dans le déplacement des unités de pêche. Pour résoudre ce problème, nous présentons ici un prototype de simulateur qui s'attache à décrire la distribution spatiale agrégée des poissons, les comportements des pêcheurs et à les combiner. Nous allons effectuer une simulation d'un champ de densité pour le poisson spatialement structuré sur lequel nous allons faire intervenir des acteurs auxquels nous donnerons des règles de comportement pour la recherche et la capture. Nous nous intéressons ici à la pêche à la senne tournante sur une ressource en bancs par des unités de pêche artisanale au Sénégal.

Dans ce cas nous choisissons de représenter la ressource selon un processus ponctuel marqué, où les bancs sont les points et où leur biomasse (et éventuellement d'autres

variables) sont les marques. Le processus de recherche fait intervenir la nature des choix de déplacement des unités de pêche et celle de la prise de décision de donner un coup de filet en fonction de leurs observations. La question des variations de l'intensité du processus ponctuel dans l'espace est intéressante d'un point de vue biologique et s'impose de toute façon dès lors que les pêcheurs considèrent l'existence de telles variations pour prendre des décisions de déplacement. Selon que ces décisions sont faites au vu d'observations locales ou bien à partir d'une connaissance a priori des zones de plus forte abondance de bancs, l'hypothèse de stationarité du processus pourra être justifiée ou non. Les mêmes remarques peuvent bien entendu porter sur la répartition des marques...

L'objet de cette présentation n'est pas de présenter une revue de toutes les situations et de toutes les "solutions" qui pourraient y être adaptées. Il s'agit de décrire un cas d'étude inspiré d'une situation particulière et de tenter d'illustrer, à partir de cet exemple, la richesse du sujet des points de vue biologique, technologique et biométrique. Le simulateur combinant les deux processus est fondée sur la connaissance de la pêcherie sénégalaise à la senne tournante.

2. Processus de répartition de la ressource

Le processus de répartition de la ressource est adapté aux connaissances disponibles sur la répartition en bancs des sardinelles (Marchal et Petitgas, 1993). Nous prenons en compte les deux échelles structurales de la ressource: les caractéristiques des bancs et l'organisation spatiale des bancs en agrégats de bancs.

La zone de pêche est une grille de 40×40 carrés de côté unité. Dans chaque carré x , le nombre de bancs est une réalisation d'une loi de Poisson de paramètre $\lambda(x)$ spatialement structuré. La valeur $\lambda(x)$ est elle-même obtenue par un processus spatial caractérisé par une valeur moyenne M_λ , un coefficient de variation CV_λ et un variogramme de portée a_λ . La loi de λ est lognormale. Dans chaque case, les bancs sont répartis selon une distribution uniforme.

Le nombre total N de bancs et la biomasse totale Q sont fixés et la biomasse de chaque banc est tirée dans une distribution lognormale dont le coefficient de variation est égal à 3.

La simulation géostatistique d'un champ de densité $u(x)$ est effectuée par la méthode des bandes tournantes de Matheron (1973) en utilisant l'algorithme de Lantuéjoul (1994). La loi de u est alors par construction une multigaussienne dont le variogramme a pour portée a_λ . Le champ des valeurs $\lambda(x)$ est déduit de celui des valeurs $u(x)$ par exponentiation. Le paramètre a_λ est donc la portée de la lognormale associée à la gaussienne dont la portée est a_λ .

3. Processus de recherche et de capture

Le processus de recherche et de capture est lui aussi adapté aux connaissances disponibles sur la pratique des unités artisanales qui exploitent les bancs de sardinelle en utilisant la sennes tournante sur la Petite Côte du Sénégal et sur les contraintes liées à ce type d'exploitation.

Les pirogues ont une contenance maximale $CP = 20$ tonnes, les sennes ont une capacité maximale $CS = 20t$, les pêcheurs choisissent de pêcher sur un banc dès lors que celui-ci a une biomasse supérieure à une valeur SB . Un banc ne peut être détecté que s'il se trouve à une distance inférieure à $RU = 0.6$ de la pirogue. Au cours d'un pas de temps élémentaire, la distance parcourue est $DU = 0.6$, la durée d'un coup de senne est $TP = 10$ et celle maximum d'une sortie est $TM = 100$ (pendant une sortie, une pirogue peut avoir un parcours dont la longueur maximum est de l'ordre de celle de la diagonale de la grille).

A chaque pas de temps Δu , la pirogue se déplace selon une direction tirée au hasard. Si un banc de taille suffisante pour mettre en oeuvre sa capture est détecté, il est capturé dans sa totalité ou non selon que sa biomasse est inférieure à CS ou non. Dans ce cas la pirogue reste sur place pendant la durée de l'action de pêche.

La stratégie d'une unité est décrite par trois paramètres St_i ($i = 1,2,3$). Lorsque plusieurs bancs sont détectés, l'unité de pêche s'intéresse au banc le plus proche ($ST1 = 0$) ou au plus gros ($ST1 = 1$). Ce banc est pêché quelque soit sa biomasse ($ST2 = 0$) ou après une décision ($ST2=1$). La capture est décidée seulement si la biomasse du banc est supérieure à un seuil calculé en fonction de SB , des captures déjà réalisées au temps t (Qt) et du nombre de coups pouvant être donnés avant la fin de la marée. Au temps t , ce nombre est égal à $nt = (TM - t)/TP$ et on peut encore pêcher $Rt = CP - Qt$. Le seuil à partir duquel la pêche est décidée est égale au min de SB et Rt/nt . Enfin, lorsque la pêche a eu lieu, l'unité de pêche poursuit sa route initiale ($ST3 = 0$), ou bien elle reste aux environs de sa dernière pêche pendant un temps fixé ($ST3 = 1$) ou fonction de la portée du variogramme ($ST3 = 2$).

4. Résultats: combinaison entre ressource, recherche et capture

Les résultats présentés ici concernent trois situations possibles de répartition de la ressource (Tableau 1).

Cas	a_λ	M_λ	CV_λ	Mb	CVb	N	Q
1	15.05	0.625	1.5	5.0	3	1000	5000
2	2.5	0.312	2.5	2.5	3	500	1250
3	2.5	0.311	2.5	10.0	3	500	5000

Tableau 1 : paramètres de distribution de la ressource selon trois cas de figure. N = nombre de bancs, Q = biomasse, M_λ = moyenne du nombre de bancs par maille sur la grille de simulation, CV_λ = coefficient de variation du nombre de bancs par maille, a_λ = portée du variogramme du Log du nombre de bancs par maille, Mb = moyenne de la biomasse par banc, CVb = coefficient de variation de la biomasse par banc

Le cas 1 correspond à une ressource abondante et répartie spatialement en de nombreux petits bancs formant de grands agrégats. Le cas 3 correspond à une ressource abondante et répartie spatialement en un petit nombre de gros bancs formant de petits agrégats de bancs. Le cas 2 correspond à une ressource peu abondante et répartie spatialement en un petit nombre de petits bancs formant de petits agrégats de bancs.

Dans chaque cas, on considère une flotte de pêche dont toutes les unités ont la même stratégie parmi deux possibles. La première stratégie correspond à des unités « témoins », la

seconde correspond à des unités qui appliquent la règle de décision quant au choix de mettre ou non en pêche en fonction de la taille du banc et qui, après la pêche restent dans la zone pendant le temps qui serait nécessaire pour parcourir la portée du variogramme du processus ponctuel. Dans chaque cas, 20 simulations ont été réalisées avec à chaque fois une flotte de dix pirogues. La Figure 1 montre le résultat d'une simulation.

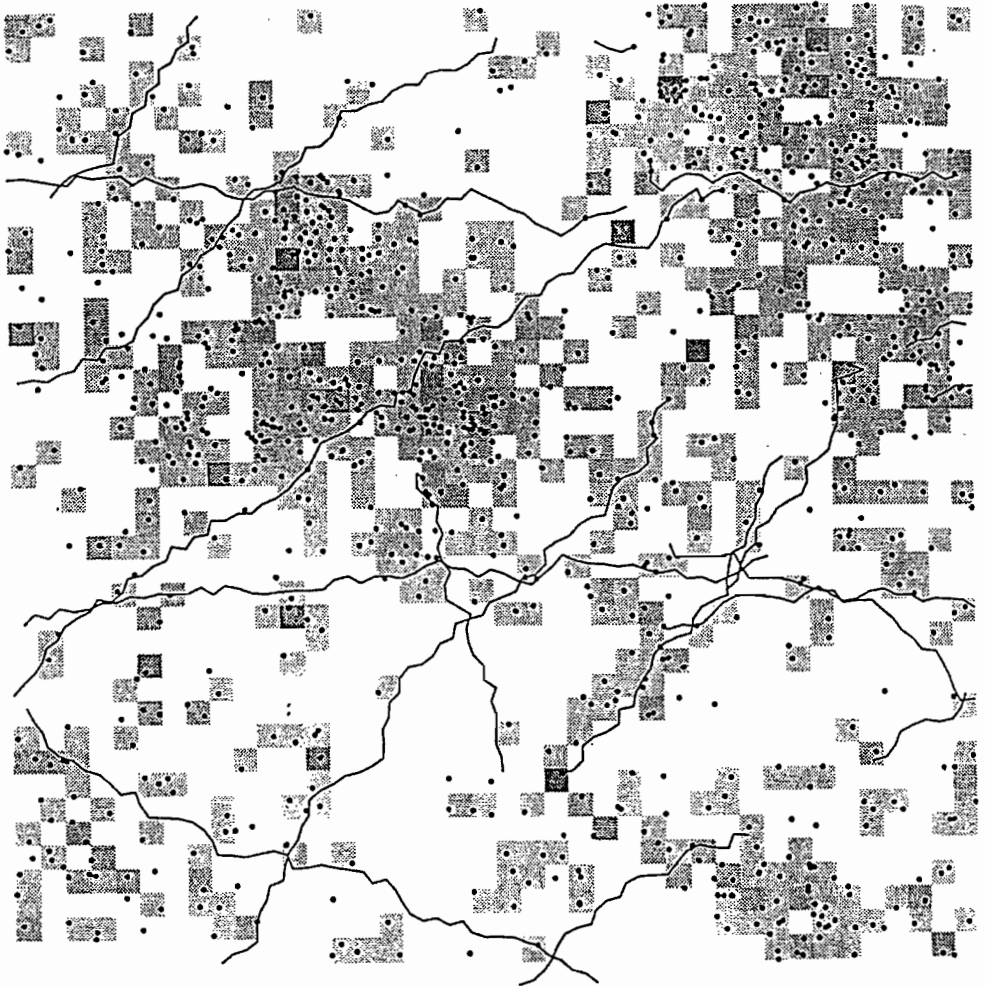


Figure 1 : Une des simulations du cas 1 de distribution du poisson et de sa pêche avec la stratégie (0,1,2). Les grisés représentent la contribution de chaque case à la biomasse totale. Les points représentent des bancs. Les traits représentent les trajets des pirogues.

Les résultats des moyennes obtenues pour plusieurs variables sont donnés ci-dessous pour chacune des six combinaisons (Tableau 2):

Cas	Stratégie	Prise/ temps mer	Prise/ temps recherche	Prise/ coup	Durée sortie	Temps recherche	nb.coup/ temps mer	Taux remplissage pirogue
1	0,0,0	0.23	0.14	3.17	77.1	20.7	0.073	0.9
1	0,1,2	0.24	0.05	10.72	64.4	49.8	0.023	0.7
2	0,0,0	0.13	0.06	1.97	93.1	31.8	0.066	0.6
2	0,1,2	0.13	0.04	3.69	88.9	56.8	0.036	0.6
3	0,0,0	0.32	0.15	4.82	80.8	21.3	0.050	0.9
3	0,1,2	0.22	0.06	6.59	71.1	47.4	0.034	0.8

Tableau 2 : Valeurs moyennes pour différents indicateurs de la biomasse après simulations pour différentes répartitions spatiales de la ressources et différentes stratégies de pêches.

Les résultats montrent une relative efficacité de l'absence de stratégie (0,0,0) en termes de capture par sortie, mais les temps de sortie sont plus longs que dans le cas où les unités de pêche ont la stratégie (0,1,2). Ces dernières ont des prises par coup de senne nettement plus élevées, mais bien évidemment avec un nombre de coups plus réduits. Il apparaît ainsi que la distribution de la taille des bancs capturés dépend beaucoup de la stratégie, ce qui était prévisible, par contre il était moins évident a priori que l'absence de stratégie conduise à une augmentation des temps de mer, et surtout -dans les cas simulés ici- à une augmentation des captures par sortie, dont l'ampleur est liée à la nature de la répartition des bancs.

Les contraintes techniques de l'exploitation (taille des pirogues et des sennes, durée de la sortie) influencent logiquement les valeurs de la taille des bancs capturés mais cette dépendance n'a pas été étudiée.

5. Conclusion

Le simulateur présenté ici est bien entendu améliorable, mais il apparaît que ce type d'outil pourrait s'avérer très utile pour l'inférence de paramètres de la population à partir des résultats de pêche, i.e. pour la recherche d'indicateurs fiables d'abondance et pour celle de la nature des relations entre capture et abondance. Il illustre en outre l'intérêt de recherches combinées sur la distribution spatiale de la ressource à plusieurs échelles, et sur les stratégies d'exploitation des unités de pêche. De telles analyses peuvent en particulier être réalisées à partir de données collectées par des observateurs embarquant sur les navires de pêche (Gaertner et al 1999).

L'exemple traité ici est relatif à un modèle construit et paramétré à partir de connaissances disponibles sur la ressource et sur la pêche. Bien que cet exercice soit assez sommaire, il illustre l'intérêt d'une réflexion sur la collecte et l'utilisation d'informations et d'observations sur le milieu et les populations exploitées et sur les pratiques de pêche.

L'utilisation halieutique de l'information sur la distribution spatiale du poisson dans les campagnes d'évaluation directe engendre le besoin de connaître celle des captures et de connaître comment les décisions des pêcheurs sont prises.

L'exemple présenté ici reste assez sommaire. Mais il montre l'apport décisif des disciplines de modélisation des processus spatiaux et temporels à l'halieutique.

Références

Gaertner D., Pagavino M., and Marcano J. 1998, Influence of fisher's behaviour on the catchability of surface tuna schools in the venezuelian purse seiner fishery in the Carribean Sea, *Canadian Journal of Fisheries and Aquatic Sciences*, 56 : 394-406

Gauthiez F. 1997. Structuration spatiale des populations de poissons marins démersaux. Caractérisation, conséquences biométriques et halieutiques. Thèse de l'Université Claude Bernard Lyon I.

Lantuéjoul C. 1994. Non conditional simulation of stationary isotropic multigaussian random functions, in *Geostatistical simulations*, Ed. Armstrong A. and P. Dowd, Kluwer Academic Publishers, *Quantitative geology and geostatistics*, 7: 147-177.

MacCall A., 1990, *Dynamic geography of marine fish populations*, University of Washington Press, Seattle, 153p.

Mangel M., Beder J., 1985, Search and stock depletion : theory and applications, *Canadian Journal of Fisheries and Aquatic Sciences*, 42 : 150-163.

Matheron G. 1973. The intrinsic random functions and their applications. *Advanced Applied Probabilities*, 5: 439-468.

Marchal E. and P. Petitgas, 1993. Precision of acoustic fish abundance estimates : separating the number of schools from the biomass in the schools. *Aquatic Living Resources*, 6, 211-219.

Paloheimo J., Dickie L., 1964, Abundance and fishing success, *Rapports et Procès Verbaux CIEM*, 155 : 152-163.

ACTUALISATION D'UN MODÈLE DU CYCLE BIOLOGIQUE DU SAUMON

TRENKEL V., FAIVRE R., BADIA J., GOULARD M.

*Unité de Biométrie et d'Intelligence Artificielle
INRA - Centre de Recherche de Toulouse
BP 27 - 31326 Castanet-Tolosan Cedex
FRANCE*

Résumé:

L'utilisation d'un modèle stochastique de dynamique de population pour la gestion pose le problème de la connaissance des effectifs réels de la population. Pour les états non observables, il existe des méthodes qui permettent d'utiliser au mieux les données disponibles sur le système. Nous étudions ici, dans le cadre des salmonidés migrateurs, comment mettre à jour les effectifs simulés par un modèle. Deux méthodes d'actualisation sont présentées: le filtre de Kalman généralisé et le calcul par simulation de distributions *a posteriori* en considérant plusieurs lois d'observation. Les méthodes proposées ont été mises en œuvre en situation simulée à partir du modèle et sur des observations. Le filtrage est facile à mettre en œuvre et ses performances sont acceptables. La méthode bayésienne avec hypothèse binomiale est également satisfaisante mais coûteuse en temps de calcul.

Mots-Clés: actualisation, filtre de Kalman généralisé, approche bayésienne, cycle du saumon, simulation.

Summary:

The need to know the true state of the system occurs when a stochastic population dynamics model is used for management purposes. A number of methods exist for estimating the unobservable states of the system based on some system observations. We investigate for the example of migrating salmon how the unobservable states can be updated. Two updating methods are presented which are the generalized Kalman filter and Bayesian estimation by simulation assuming three different likelihood functions. All methods have been applied to data simulated using the population dynamics model and some real observations. The Kalman filter is easy to implement and seems to perform well. Bayesian estimation with binomial likelihood functions provides satisfying results but is expensive to calculate.

Keywords: updating, generalized Kalman filtering, bayesian approach, salmon life-cycle, simulation.

Introduction

Dans le cadre de l'étude de l'impact des activités humaines sur le devenir des populations de salmonidés migrateurs (Prouzet *et al.* 1996), nous avons développé le modèle CBS du Cycle Biologique du Saumon (Charron, 1994; Faivre *et al.*, 1997). Ce modèle simule l'évolution d'une population en se basant sur les stades "clés" du développement du saumon atlantique (Dumas *et al.*, 1996). Les effectifs de saumons passant d'un stade à l'autre sont modélisés par des fonctions paramétrées prenant en compte les conditions environnementales en rivière, les pressions de pêches, la compétition territoriale, *etc.* (Figure 1). Les paramètres de ces fonctions peuvent être fixés ou aléatoires (variables d'une année à l'autre, leurs valeurs sont tirées dans des distributions fixées *a priori*).

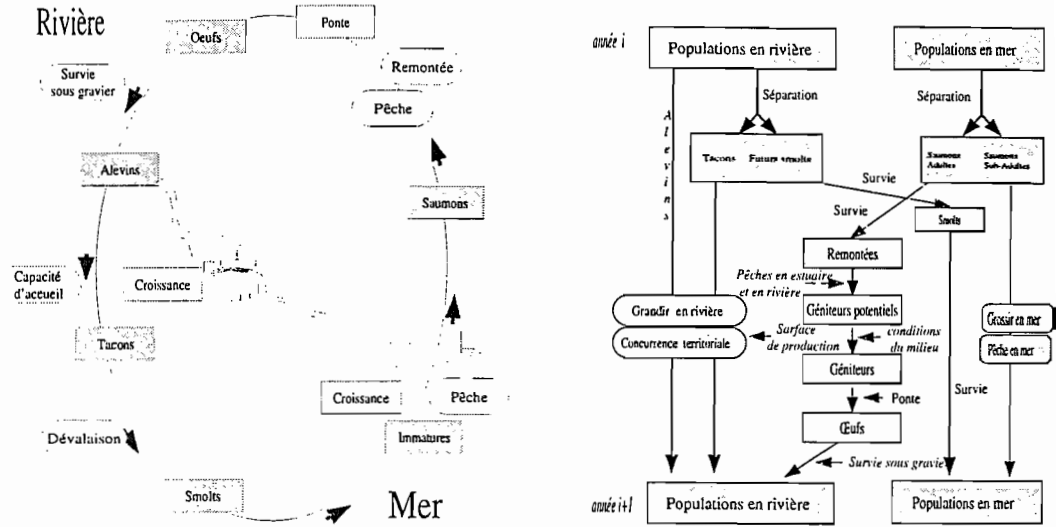


FIG. 1 – Cycle biologique du Saumon et organigramme du modèle CBS.

Le caractère très aléatoire de la dynamique du saumon impose, dans l’optique d’une gestion adaptée, de suivre l’évolution du peuplement en temps réel (dans notre cas, d’année en année). Ce suivi doit permettre de réduire la gamme des simulations du modèle pour la gestion. Dans ce travail, notre objectif est de réaliser l’actualisation du modèle pour prendre en compte dans les simulations des observations partielles de la population : captures par pêche, sondage en rivière, relevés des passes à poissons, etc. Jager *et al.* (1997) ont utilisé une approche d’actualisation basée sur la réestimation des paramètres et sur la reformulation du modèle. Notre procédure d’actualisation concerne la mise à jour du vecteur d’état de la population simulée.

Pour ce type de problème on trouve dans la bibliographie deux approches : le filtre de Kalman et l’approche bayésienne. Le filtre de Kalman (Kalman, 1960) est une méthode de filtrage linéaire utilisée dans la domaine halieutique depuis quelques années pour estimer par exemple la taille du stock et du taux de mortalité par pêche avec des données de capture par pêche (Gudmundsson, 1994 ; Gudmundsson, 1995 ; Sullivan, 1992). Le filtre de Kalman généralisé est la version étendue aux modèles à erreur d’observation dépendant de l’état (Zenwirth, 1988).

Une alternative aux méthodes de filtrage est la recherche de lois *a posteriori* obtenues par simulation (Smith et Gelfand 1992). Cette approche bayésienne a été appliquée à la gestion des stocks de poissons par McAllister *et al.* (1994).

Dans la section 1, on présente le modèle CBS du cycle biologique et la description du processus d’observation puis on introduit dans la section 2 les différentes méthodes d’actualisation de CBS. Elles sont basées sur le filtre de Kalman généralisé ou sur une approche bayésienne en considérant diverses alternatives en terme de loi d’observation et en calculant les distributions *a posteriori*. Les méthodes proposées ont été mises en œuvre dans une situation simulée à partir du modèle CBS et pour des observations réelles sur la Nivelle. Les résultats sont présentés dans la section 3 et sont discutés en section 4.

1 Modèle et Méthodes

1.1 Modèle du cycle biologique du saumon

Soit x_t le vecteur simplifié des états du cycle biologique des saumons. On peut le décomposer en trois groupes de populations : rivière, mer et remontant ;

$$x'_t = \begin{bmatrix} \text{(Rivière)} & & \text{(Mer)} & & \text{(Remontant)} \\ R1_t, FS2_t & | & S1_t, S2_t, & SBA11_t, SBA12_t & | & RMT11_t, RMT12_t, RMT21_t, RMT22_t \\ = & [& x'_{R,t} & | & x'_{M,t} & x'_{S,t} & | & x'_{B,t} \end{bmatrix}$$

où R1 désigne la population des alevins de 1 an, FS2 les futurs smolts de 2 ans, S1 et S2 les smolts de 1 et 2 ans, SBA11 (SBA12) les saumons subadultes de 1 an en rivière et 1 an (2 ans) en mer, RMTjk les saumons remontant après j années passées en rivière et k années passées en mer (Figure 2).

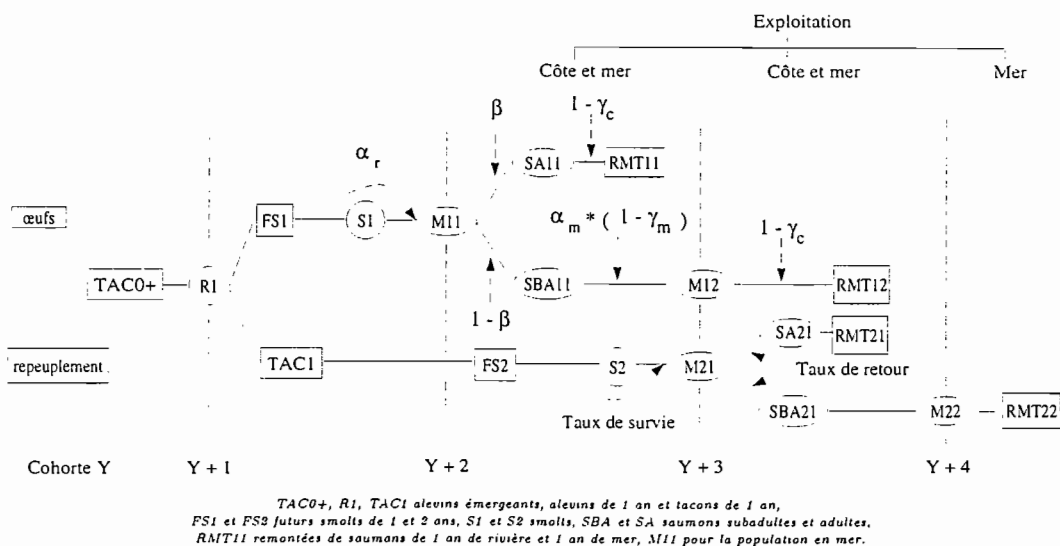


FIG. 2 – Histoire d'une cohorte - Schéma simplifié.

Des équations d'état décrivent le changement des nombres de saumons dans chaque classe d'âge de l'année t à l'année $t+1$ (Charron. 1994 ; Dumas *et al.*. 1996). Parmi celles-ci l'équation liant le nombre de saumons remontant à t et le nombre de saumons en mer à $t-1$ est

$$x_{B,t} = (1 - \Gamma_c) B_t A_m \begin{bmatrix} x_{M,t-1} \\ x_{S,t-1} \end{bmatrix} \quad (1)$$

où A_m est le taux de survie en mer, B_t le taux de retour en rivière au temps t et Γ_c le taux de pêche côtière.

1.2 Modèle d'observation des remontées

Pour chaque classe k de saumons observée, le comptage des saumons remontant en rivière, conditionnellement à $\mathbf{x}_{B,t}$, est modélisé par une loi binomiale $(y_{k,t}|\mathbf{x}_{B,t}) \sim \text{Bin}(H_{k,t} \mathbf{x}_{B,t}, p_{k,t})$.

- $H_{k,t}$ est la k -ième ligne de la matrice \mathbf{H}_t , de dimension $m \times n$, qui décrit le processus d'observation. Si tous les stades sont observés séparément \mathbf{H}_t est la matrice d'identité.
- $p_{k,t}$, supposé connu, est le taux d'observation pour la catégorie d'observation k des saumons de l'année t .

De façon plus synthétique on peut aussi écrire pour le vecteur \mathbf{y}_t de tous les comptages

$$\mathbf{y}_t = \mathbf{P}_t \mathbf{H}_t \mathbf{x}_{B,t} + \mathbf{w}_t$$

- $\mathbf{P}_t = \text{diag}(p_{k,t})$ étant la matrice diagonale de dimension $m \times m$ qui contient les taux d'observation $p_{k,t}$,
- \mathbf{w}_t étant l'erreur de mesure; c'est une variable aléatoire d'espérance nulle et de variance Σ_t^w dépendant de l'état de la population $\mathbf{x}_{B,t}$ échantillonnée pendant l'année t .

L'espérance de \mathbf{y}_t , conditionnellement aux observations $\mathbf{x}_{B,t}$ est $E[\mathbf{y}_t|\mathbf{x}_{B,t}] = \mathbf{P}_t \mathbf{H}_t \mathbf{x}_{B,t}$ et la matrice de variance-covariance, avec le même conditionnement, $V[\mathbf{y}_t|\mathbf{x}_{B,t}] = (\mathbf{I} - \mathbf{P}_t) \mathbf{P}_t \mathbf{H}_t \text{diag}(\mathbf{x}_{B,t})$. L'espérance de \mathbf{y}_t sans conditionnement est $E[\mathbf{y}_t] = \mathbf{P}_t \mathbf{H}_t \mu_{B,t}$ et la variance-covariance $V[\mathbf{y}_t] = (\mathbf{I} - \mathbf{P}_t) \mathbf{P}_t \mathbf{H}_t \text{diag}(\mu_{B,t}) + \mathbf{P}_t \mathbf{H}_t \Sigma_{B,t} \mathbf{H}_t' \mathbf{P}_t'$.

2 Méthodes d'actualisation

Pour l'actualisation des états du système, les deux classes de méthodes auxquelles nous nous intéressons sont le filtre de Kalman généralisé (Zenwirth, 1988) et l'approche bayésienne par simulation (Smith et Gelfand, 1992).

Le filtre de Kalman généralisé est le prédicteur linéaire $\mathbf{x}_{B,t}^*$ du nombre de saumons remontant en rivière défini par

$$\mathbf{x}_{B,t}^* = \mu_{B,t} + \mathbf{K}_t (\mathbf{y}_t - \mathbf{P}_t \mathbf{H}_t \mu_{B,t}) \quad (2)$$

où $\mu_{B,t}$ est l'espérance de l'état, $\mathbf{K}_t = \Sigma_{B,t} (\mathbf{P}_t \mathbf{H}_t)' [\mathbf{P}_t \mathbf{H}_t \Sigma_{B,t} (\mathbf{P}_t \mathbf{H}_t)' + (\mathbf{I} - \mathbf{P}_t) \mathbf{P}_t \mathbf{H}_t \text{diag}(\mu_{B,t})]^{-1}$ le gain du filtre et $V[\mathbf{x}_{B,t}^*] = (\mathbf{I} - \mathbf{K}_t \mathbf{P}_t \mathbf{H}_t) \Sigma_{B,t}$ la variance du prédicteur.

Cet estimateur est le meilleur prédicteur linéaire, au sens du risque quadratique. Si le système est gaussien, il est optimal parmi tous les prédicteurs. Conditionnellement au comptage, la distribution du nombre de saumons remontant en rivière suit une loi binomiale négative $(H_{k,t} \mathbf{x}_{B,t} | y_{k,t}) \sim \text{Neg}(y_{k,t}, p_{k,t})$. La distribution binomiale négative pouvant être approchée par une loi normale quand $y_{k,t}$ est grand, le prédicteur linéaire (eq. 2) est alors proche de l'optimalité.

En pratique il faudra estimer $\mu_{B,t}$ et $\Sigma_{B,t}$ pour appliquer le filtre de Kalman.

L'approche bayésienne par simulation consiste à engendrer un échantillon de la distribution jointe *a posteriori* des nombres de saumons dans les différentes catégories d'observation. On peut décrire la méthode de la façon suivante :

- On utilise CBS pour simuler des valeurs $x_{B,t}^b$, $b = 1, \dots, r$
- On effectue un tirage de s valeurs parmi les $x_{B,t}^b$, $b = 1, \dots, r$, ce tirage utilise comme poids les valeurs de vraisemblance standardisée.

Cette procédure donne un échantillon de taille s de la distribution *a posteriori* du vecteur d'état $(\mathbf{x}_{B,t} | \mathbf{y}_t, \mathbf{y}_{t-1})$. Nous avons considéré trois procédures suivant qu'on utilise pour le calcul de la vraisemblance le modèle de base ou des approximations. Dans le modèle de base, les vraisemblances pour l'observation \mathbf{y}_t sont obtenues en supposant une loi binomiale pour chaque catégorie donnant ainsi des distributions indépendantes *a posteriori*. Nous avons fait le calcul en considérant des approximations des distributions de vraisemblance par des lois normale et multinormale. Dans le cas multinormal, on obtient une distribution jointe *a posteriori* des différentes catégories de saumons, il n'y a plus indépendance.

Pour la loi binomiale, si toutes les prédictions par CBS sont inférieures à l'observation, la vraisemblance est égale à zéro pour toutes les simulations. Dans ce cas on prend l'observation divisée par le taux d'observation comme estimateur et on fixe à zéro la variance de cet estimateur.

Sur la base de la relation (éq. 1), l'actualisation du nombre, non observable, de saumons restés en mer $(\mathbf{x}_{S,t})$ est obtenue par l'équation

$$\mathbf{x}_{S,t} = \Gamma^{-1} \mathbf{B}_t^{-1} (\mathbf{I} - \mathbf{B}_t) \mathbf{x}_{B1,t}$$

avec $\mathbf{x}'_{B1,t} = [RMT11_t, RMT12_t]$ les classes de saumons ne remontant pas en totalité en rivière l'année t , $\Gamma = \text{diag}(1 - \gamma_c^j)$ avec γ_c^{ij} les taux de pêche côtière considérés constants et $\mathbf{B}_t = \text{diag}(\beta_t^{ij})$ les taux de retour des saumons qui ont passés i années en rivière et j années en mer ($i=1,2$ et $j=1,2,3$). Ce taux de retour est supposé aléatoire, $\beta_t^{ij} \sim N(\theta^{ij}, \sigma_\theta^{ij})$ et $0 \leq \beta_t^{ij} \leq 1$. Ainsi $E[\mathbf{B}_t] = \text{diag}(\theta^{ij})$ et $V[\mathbf{B}_t] = \text{diag}(\sigma^{ij})$.

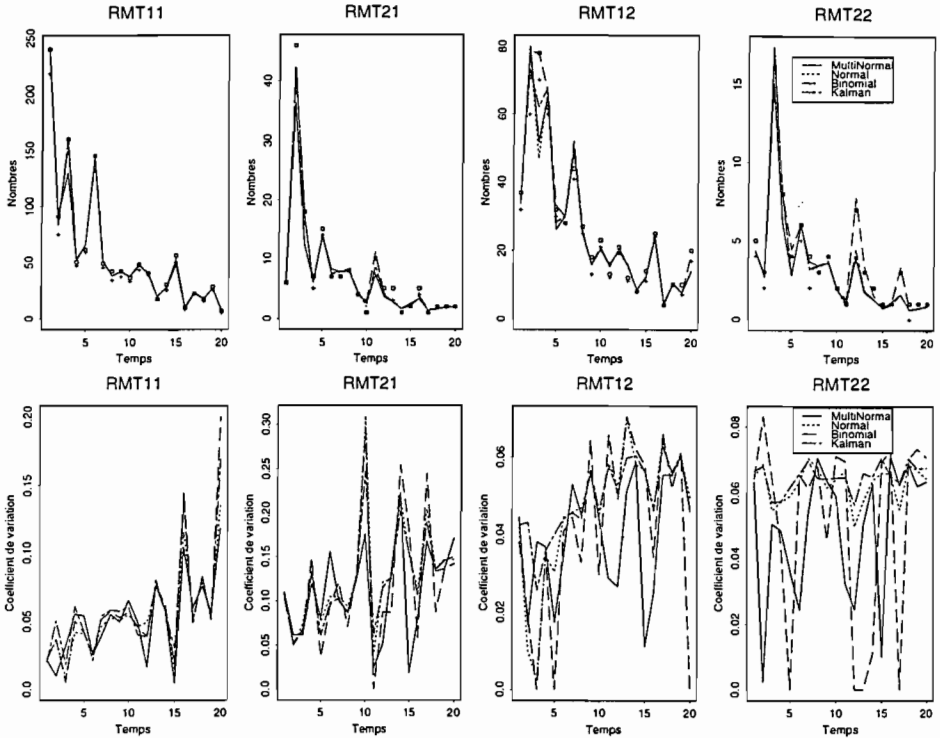
Le prédicteur de $\mathbf{x}_{S,t}$ est $\mathbf{x}_{S,t}^* = \Gamma^{-1} E[\mathbf{C}_t] \mathbf{x}_{B1,t}^*$ où $\mathbf{x}_{B1,t}^*$ est, par définition de $\mathbf{x}_{B1,t}$, obtenu par projection de $\mathbf{x}_{B,t}^*$ et $\mathbf{C}_t = \text{diag}((1 - \beta_t^{ij})/\beta_t^{ij})$. La matrice de variance-covariance de $\mathbf{x}_{S,t}^*$ s'obtient à partir de celle de $\mathbf{x}_{B1,t}$ et des moments des β_t^{ij} .

3 Comparaison des méthodes d'actualisation

Les quatre méthodes d'actualisation (filtre de Kalman généralisé, approches bayésiennes par simulation pour des lois de vraisemblance multinormale, normale et binomiale) ont été comparées sur des données simulées à partir de CBS et sur les données observées sur la Nivelle.

Pour les simulations, le nombre initial de saumons dans chaque classe a été obtenu à partir du modèle CBS après une phase d'initialisation qui simule dix ans. Le nombre de simulations de CBS pour l'approche bayésienne a été fixé à $r = 20000$ pour arriver finalement à un échantillon de la distribution *a posteriori* de $s = 1000$. Pour le filtre de Kalman généralisé, 400 simulations du modèle CBS ont été utilisées pour estimer $\mu_{B,t}$ et $\Sigma_{B,t}$. Les valeurs des paramètres du modèle CBS ou leur loi de distribution sont décrites dans Charron (1994).

Le processus d'actualisation se fait de la façon suivante : le modèle CBS prévoit les distributions des nombres de saumons pour l'année t (distributions *a priori*), les méthodes d'actualisation fournissent les distributions *a posteriori* pour la même année t . Les espérances *a posteriori* sont les points de départ pour les prévisions de l'année $t + 1$ par CBS. Les distributions *a priori* sont donc différentes pour chaque méthode sauf la première année.



RMT11 remontées de saumons de 1 an de rivière et 1 an de mer. Les □ indiquent les vraies valeurs simulé par CBS et les + les valeurs "observés".

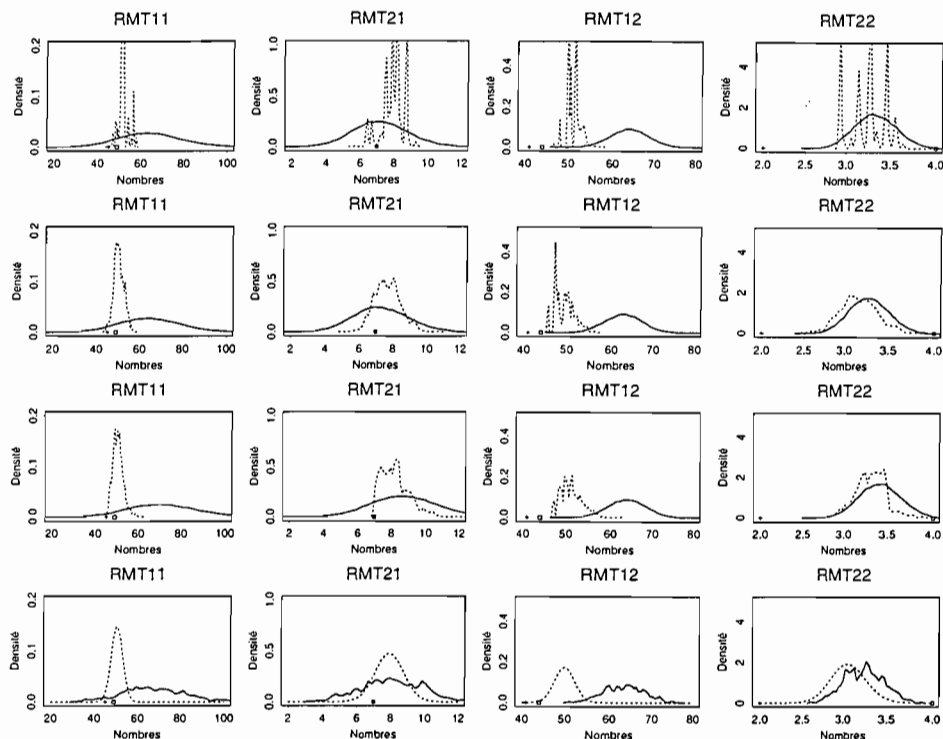
FIG. 3 – Étude de simulation pour le nombre de saumons actualisé par quatre méthodes, en haut les moyennes, en bas les coefficients de variation.

3.1 Performance en situation contrôlée

Pour simuler les observations, 20 années d'une population ont été simulées avec le modèle CBS et l'observation résulte d'un échantillonnage effectué indépendamment pour chaque classe d'âge avec un taux d'observation de 90 %.

Les moyennes des distributions *a posteriori* des nombres de saumons remontant en rivière obtenues avec les quatre méthodes sont semblables (Figure 3). La méthode bayésienne avec la vraisemblance multinormale donne en général les coefficients de variation les plus petits, les autres méthodes donnent des coefficients de variation comparables et faibles. Quand le modèle CBS prévoit des nombres inférieurs aux comptages, par exemple l'année 5 pour RMT12, les coefficients de variation sont fixés à zéro pour la méthode binomiale.

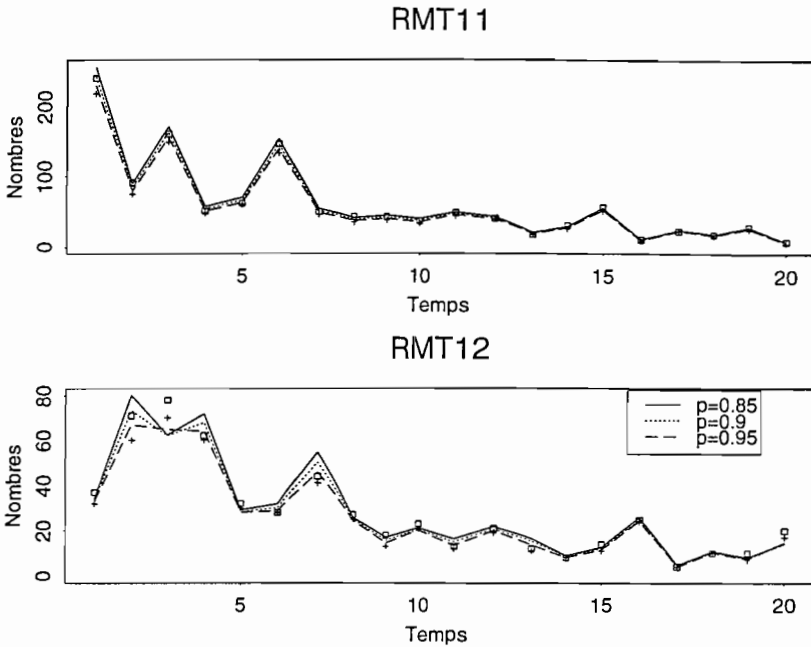
Pour comparer les estimations des différentes méthodes plus en détail, nous présentons les distributions *a posteriori* de l'année 7 (Figure 4). Le filtre de Kalman généralisé est représenté par une distribution gaussienne (eq. 2). On peut noter l'asymétrie des distributions *a posteriori* obtenues avec la vraisemblance binomiale, et dans une moindre mesure pour la normale. La nature très irrégulière des courbes obtenues avec la vraisemblance multinormale suggère que le nombre de simulations n'était peut-être pas assez grand. Les prévisions *a priori* obtenues par CBS sont loin des valeurs "observées" pour les RMT12 (à noter les échelles différentes entre les 1 an et 2 ans de rivière). L'observation pour les 2 ans de mer qui pour l'année considérée est très faible, correspond à un événement rare comme par exemple une mauvaise survie en mer entre la première et la seconde année en mer.



RMT11 remontées de saumons de 1 an de rivière et 1 an de mer.

FIG. 4 – Étude de simulation pour le nombre de saumons actualisé par quatre méthodes. En ligne continue les distributions *a priori* données par le modèle CBS pour l'année 7 et en ligne pointillée les distributions *a posteriori* obtenues avec la méthode multinormale (première ligne), la méthode normale (deuxième ligne), la méthode binomiale (troisième ligne) et le filtre de Kalman généralisé (dernière ligne).

Pour les résultats précédents, les estimations des nombres de saumons remontant en rivière ont été réalisées en utilisant le vrai taux d'observation. En répétant l'estimation avec des taux d'observation de 0.85 et 0.95, nous avons étudié la sensibilité de l'estimation vis-à-vis de ce taux. Quand on utilise le filtre de Kalman généralisé, les moyennes ne diffèrent que très peu (Figure 5). Les résultats pour les trois autres méthodes sont similaires.



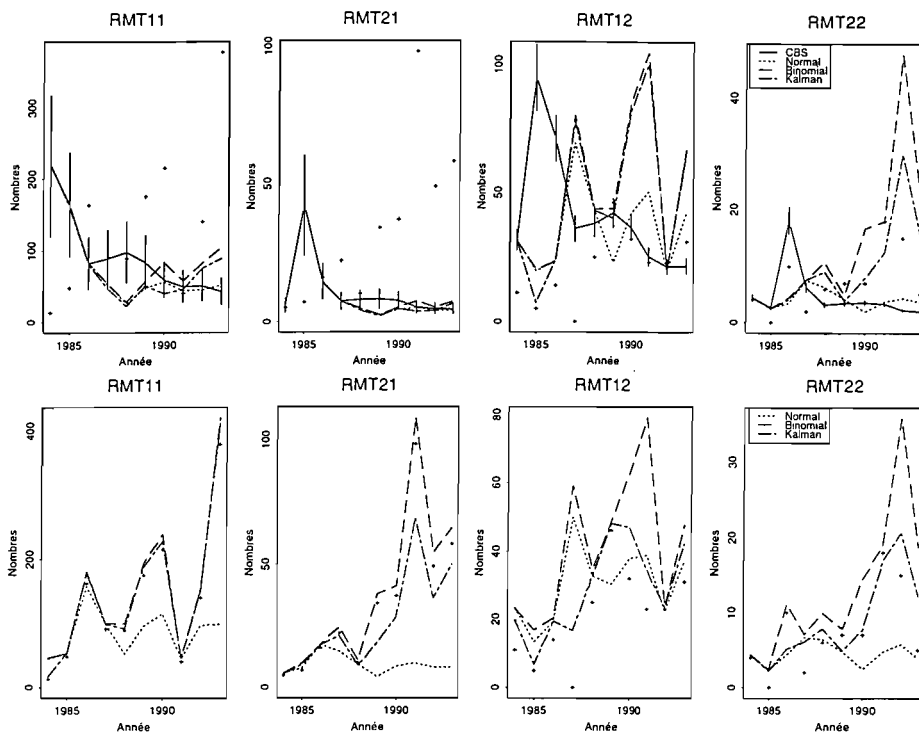
RMT11 remontées de saumons de 1 an de rivière et 1 an de mer. Les \square indiquent les vraies valeurs et les $+$ les valeurs observées.

FIG. 5 - Étude de simulation pour l'espérance du nombre de saumons actualisé par le filtre de Kalman généralisé pour trois taux d'observation.

3.2 Application aux données de la Nivelles

Les comptages des nombres de saumons ayant remonté la Nivelles entre 1984 et 1993 ont été utilisés (Dumas, 1994). Le taux d'observation pour cette population a été fixé à 0.9 pour toutes les années et pour toutes les classes de saumon.

Les trois méthodes utilisées (Kalman généralisé, méthodes bayésiennes avec les lois normale et binomiale) donnent des estimations des remontées *a posteriori* très différentes (Figure 6). Les procédures de mises à jour des effectifs l'année t (Figure 6, 2-ème ligne) sont acceptables sauf pour une loi d'observation normale. Par contre, les prévisions pour l'année $t + 1$ (Figure 6, 1-ère ligne) diffèrent suivant le nombre d'années passées en mer. Pour les 1 an de mer, on n'observe pas de correction du modèle qui pourrait être incorrectement paramétré. Les valeurs des paramètres taux de retour après 1 an de mer ou taux de survie en mer semblent être sous évaluées; les effectifs des remontées sont croissants alors que les effectifs simulés par le modèle et ses corrections ne suivent pas cette tendance. Cependant, pour les 2 ans de mer, effectifs observés et effectifs corrigés évoluent de façon comparable. Le nombre prédit de saumons remontant la Nivelles l'année $t + 1$ est directement lié au nombre restant en mer l'année t . Il semblerait donc que le nombre prédit de saumons restant en mer l'année t soit correctement prédit. Le filtre de Kalman généralisé et la méthode bayésienne avec la vraisemblance binomiale donnent des résultats proches. Ces résultats sont indépendants du taux d'observation fixé ici à 90 %. On arrive à des conclusions identiques avec un taux d'observation fixé à 0.95.



RMT11 remontées de saumons de 1 an de rivière et 1 an de mer. Les comptages sont indiqués par +.

FIG. 6 – Estimation du nombre de saumons dans la Nivelle par trois méthodes d'actualisation - en haut les prévisions par CBS, en bas les effectifs actualisés.

4 Discussion

La comparaison des méthodes pour des données simulées n'a pas montré de grandes différences entre les quatre méthodes d'actualisation. Pour les données de la Nivelle, le filtre de Kalman généralisé et l'approche bayésienne avec une loi d'observation binomiale semblent les plus appropriées : le filtrage étant en général plus stable numériquement. D'une manière générale, les méthodes bayésiennes par simulation sont coûteuses en temps de calcul. Pour la vraisemblance multinormale, il faut un très grand nombre de simulations. La méthode normale (unidimensionnelle) n'est pas efficace dans les cas extrêmes où l'observation est un événement rare de la distribution *a priori*. La méthode binomiale est satisfaisante en général mais la stratégie a nécessité une modification quand l'observation était un événement rare de la distribution *a priori*. Le filtre de Kalman est intéressant pour le calcul, il est robuste dans le cas des observations extrêmes mais la loi *a posteriori* gaussienne pour les faibles valeurs n'est pas réaliste.

Pour l'application aux données de la Nivelle, l'actualisation des nombres de saumons remontant en rivière n'a pas amélioré les prévisions *a priori* des effectifs de saumons remontant après une année en mer seulement. Pour corriger ces effectifs, des observations en rivière après la survie sous-gravier permettrait de mieux apprécier les départs en mer et donc les effectifs potentiels pour la remontée après 1 an en mer. Une très forte variabilité des effectifs résulte des aménagements du milieu et leur interaction avec les événements climatiques difficilement prévisibles (effet de crues par exemple).

On pourrait également envisager une technique d'actualisation basée sur la réestimation des distributions des paramètres utilisés dans le modèle. Nous avons dans cette étude privilégié l'actualisation des états du système. Il serait intéressant de comparer les deux approches.

Vis-à-vis du gestionnaire, il est trop tôt pour conclure sur les capacités des méthodes d'actualisation. La faible efficacité constatée dans le cas des données de la Nivelle nous amène à nous interroger sur le caractère non contrôlable du système. Il peut être dû à une dynamique purement aléatoire du système ou à la méconnaissance des paramètres des lois régissant la dynamique. Une étude plus poussée sur le déterminisme de la survie des populations de salmonidés est donc nécessaire.

Références

- CHARRON, M.H. 1994. Modélisation stochastique du cycle biologique des salmonidés migrateurs: application à la modélisation du cycle du Saumon Atlantique de la Nivelle et de l'Adour. Dipl. Etud. Sup. spéc. Méthodes Informatiques et Modèles Mathématiques, Univ. Paul Sabatier, Toulouse, France, 86 pp.
- DUMAS J. 1994. La population de saumons adultes de la Nivelle en 1993. *Station d'Hydrobiologie*, INRA, St-Pée-sur-Nivelle, 22 p.
- DUMAS J., FAIVRE R., CHARRON M.H., BADIA J., DAVAINÉ P., PROUZET P. 1996. Modélisation stochastique du cycle biologique du saumon atlantique (*Salmo salar* L.): bases biologiques, implémentation informatique et interprétation. In *Méthodes d'étude des systèmes halieutiques et aquacoles*, Ferraris, Pelletier, Rochet, ORSTOM (Ed.), 211-218.
- FAIVRE R., DUMAS J., CHARRON M.H., BADIA J., PROUZET P., 1997. River basin management using a stochastic model of the salmon life cycle. MODSIM'97: International Congress on Modelling and Simulation, Hobart, Tasmania (Australia), 8-11 December 1997, 1536-1541.
- GUDMUNDSSON G. 1994. Time series analysis of catch-at-age observations. *Applied Statistics*, 1, 117-128.
- GUDMUNDSSON G., 1995. Time series analysis of catch-at-length data. *ICES Journal of Marine Sciences*, 52, 781-795.
- JAGER H.I., CARDWELL H.E., SALE M.J., BEVELHIMER M.S., COUTANT C.C., VAN WINKLE W., 1997. Modelling the linkages between flow management and salmon recruitment in rivers, *Ecological Modelling*, 103, 171-191.
- KALMAN, R.E., 1960. A new approach to linear filtering and prediction problems, *Transaction of the American Society of Mechanical Engineers, Series D*, 82, 35-45.
- MCALLISTER, M.K., PIKITCH, E.K., PUNT, A.E., HILBORN, R., 1994. A Bayesian approach to stock assessment and harvest decisions using the sampling/importance resampling algorithm. *Canadian Journal of Fisheries and Aquatic Sciences*, 51, 2673-2687.
- PROUZET, P., BADIA, J., FAIVRE, R., 1996. Action des facteurs naturels et anthropiques sur la dynamique d'une ressource migratrice amphihaline: le saumon atlantique (*Salmo salar* L.). Programme "Environnement, Vie et Sociétés" du CNRS; journées "Tendances nouvelles en modélisation pour l'environnement", Post. 1-2, 149-159.
- SMITH, A.F.M., GELFAND, A.E., 1992. Bayesian statistics without tears: A sampling-resampling perspective, *The American Statistician*, 46, 84-88.
- SULLIVAN, P.J., 1992. A Kalman filter approach to catch-at-length analysis, *Biometrics*, 48, 237-257.
- ZENWIRTH, B. 1988. A generalization of the Kalman filter for models with state-dependent observation variance, *Journal of the American Statistical Association*, 83, 164-167.

L'ESTIMATION DES PUISSANCES DE PECHE PAR MODELISATION LINEAIRE DES CAPTURABILITES.

Laurent MILLISCHER, Olivier MAURY, Didier GASCUEL .

ENSAR Halieutique
65, rue de Saint-Brieuc, 35042 Rennes Cedex
tel : 02-99-28-75-32 / fax : 02-99-28-75-35 / email : millisch@roazhon.inra.fr

Abstract :

We suggest a statistical method to estimate the fishing power of vessels (i.e. their fishing efficiency), based on linear modeling of catchabilities using the SAS-GLM procedure. This method, applied to the industrial fleet of Lorient, permits an identification of strategic choices of fleets and an analysis of their evolution.

Résumé :

On propose une méthode d'estimation des puissances de pêche des navires (c'est à dire de leur efficacité de pêche), reposant sur une modélisation linéaire des capturabilités, à l'aide de la procédure SAS-GLM. Cette méthode, appliquée ici à la flottille industrielle de Lorient, permet d'identifier les choix stratégiques des flottilles et d'analyser leur évolution.

INTRODUCTION : QUELQUES POINTS DE DEFINITION.

Pour l'essentiel, l'évaluation des stocks halieutiques, et la gestion des pêcheries, reposent sur les données de captures et d'effort de pêche, établies à partir d'un suivi statistique des exploitations commerciales. Les observations directes, établies par les navires océanographiques restent en effet rares, souvent partielles, et toujours coûteuses. Ainsi, la relation liant les captures commerciales, l'effort de pêche et l'abondance des stocks exploités constitue l'une des questions centrales de la recherche en dynamique des populations halieutiques. Le pivot théorique de cette relation est le facteur de capturabilité, selon les formulations usuelles suivantes : $U = q \cdot B$; $F = q \cdot f_n$ (avec U , la prise par unité d'effort ; B , la biomasse totale ; F , le coefficient de mortalité par pêche ; f_n , l'effort total, dit aussi effort nominal ; et q , la capturabilité). De façon synthétique, le facteur q fait le lien entre le développement d'une activité (quantité d'effort nominal de pêche, comptabilisé par exemple en heures de pêche) et son impact sur la ressource (mortalité par pêche F). Ainsi définie, la capturabilité peut être considérée comme le produit de deux termes : le premier, fondamentalement dépendant de l'activité du pêcheur, est la puissance de pêche P_g développée sur le stock ; le second, dépendant fondamentalement de la ressource, est la disponibilité d du stock (Gascuel, 1993) :

$$q = \frac{F}{f_n} = \frac{d \cdot f_e}{f_n} = d \cdot P_g \quad (1)$$

La puissance de pêche fait donc le lien entre le développement d'une activité de pêche (quantité d'effort nominal f_n) et la pression réelle exercée sur le stock par cette activité (effort effectif f_e). En un mot, elle correspond à l'efficacité de l'activité de pêche, pour une disponibilité de la ressource donnée. Son estimation s'avère souvent délicate. Dans la plupart des cas, elle s'appuie sur l'étude des prises par unité d'effort U, en s'appuyant sur la relation $U = q.B$ (Robson, 1966 ; Laurec, 1977 ; Laurec et Le Guen, 1981). Mais il est alors souvent difficile de discerner, au sein des variations de prise par unité d'effort, ce qui relève de changements de puissance de pêche, et donc de capturabilité, et ce qui relève de variations d'abondance.

On propose ici une méthode d'estimation des puissances de pêche des navires s'appuyant sur une modélisation G.L.M. (General Linear Model) des capturabilités, quantifiées à partir des coefficients de mortalité par pêche F issues de l'analyse des cohortes (généralement notée V.P.A., d'après la dénomination anglo-saxonne "Virtual Population Analysis"). A titre d'illustration, la méthode est appliquée au cas de la flottille industrielle de Lorient.

METHODE DE QUANTIFICATION DES PUISSANCES DE PECHE.

Rappel : G.L.M. et Halieutique

L'utilisation en halieutique des méthodes de modélisation de type G.L.M. ne constitue pas une nouveauté. Notamment, l'estimation d'indices d'abondance annuelle des ressources halieutiques, ou encore l'estimation des puissances de pêche, à partir des données de Prises par Unité d'Effort (P.U.E.) fait couramment appel à ce type de méthode (Beverton et Holt, 1957 ; Robson, 1966 ; Kimura, 1981 ; Lo et al., 1992).

Cet intérêt tient à deux grands points. Tout d'abord, la grande difficulté à mesurer et à suivre directement l'abondance des ressources halieutiques conduit à chercher et à séparer au sein des indicateurs disponibles, en l'occurrence les P.U.E., des tendances reflétant l'effet des facteurs pris en compte dans ces indicateurs. Cette recherche et cette séparation de tendances trouvent un appui assez naturel dans les modèles d'analyse de variance. D'autre part, la plupart des données fiables servant d'indicateurs proviennent de l'activité de pêche elle-même, et rarement de campagnes océanographiques. De ce fait, ces données ne correspondent pratiquement jamais à un plan d'échantillonnage équilibré. Au regard de ces deux points, le G.L.M. s'avère être un outil précieux d'analyse des données halieutiques.

Ainsi, partant de l'hypothèse classique de proportionnalité entre les P.U.E. et l'abondance, à un facteur de capturabilité près, la modélisation par modèle linéaire de ces P.U.E. permet de calculer des indices d'abondance. Les modèles les plus courants, à deux ou trois facteurs¹, font intervenir un effet A_n , interprété comme un indice annuel de l'abondance sous l'hypothèse que les puissances de pêche des navires considérées ont varié sans tendance inter-annuelle, sur l'ensemble de la période d'étude (Gascuel et Thiam, 1994). Cette hypothèse, bien qu'acceptable dans de nombreux cas, constitue une forte restriction à la validité de ces modèles.

La méthode proposée ici permet précisément de quantifier ces puissances de pêche, et de suivre ainsi leur évolution. Elle se base sur le même principe d'utilisation du modèle linéaire, appliqué cette fois directement sur des estimations de capturabilité, et non plus sur des P.U.E.. Ainsi, posant la relation théorique (1) : $q = d.P_g$, l'hypothèse qui soutient l'interprétation de l'effet des facteurs impliqués dans les modèles porte ici sur le facteur de disponibilité d, dont on suppose la fluctuation sans tendance.

¹ Robson (1966) propose un modèle à trois facteurs, qui s'exprime comme suit :

$U_{an, saison, zone, navire} = IA_{an} * q_{saison, zone} * P_{g, navire}$. Laurec (1977) montre l'intérêt d'un modèle à deux facteurs :

$U_{an, saison, zone, navire} = IA_{an} * P_{g, navire}$.

Application à l'estimation des puissances de pêche

Les mortalités par pêche F sont estimées en routine chaque année, pour les principaux stocks exploités, par les groupes scientifiques internationaux chargés d'établir des recommandations de gestion de la pêche². De ces mortalités globales sont déduites des mortalités par navire, calculées au prorata des captures de chaque navire ; celles-ci sont alors divisées par les efforts de pêche annuels individuels. Les capturabilités de chaque stock exploité sont ainsi estimées pour chaque navire, appartenant à l'une des flottilles considérées. Elles peuvent s'interpréter comme la probabilité (pour une année et un stock donnés) qu'un poisson d'être capturé par une unité d'effort du bateau considéré, sachant que ce bateau se trouve sur l'aire de répartition du stock.

Un modèle linéaire des capturabilités est ensuite ajusté à l'aide de la procédure G.L.M. (General Linear Model) du logiciel SAS (SAS Institute Inc., 1987). A une capturabilité correspond un stock, un navire, une flottille (celle à laquelle appartient le navire), et une année de pêche (sur laquelle est calculée la capturabilité). A partir de ces différents facteurs explicatifs, deux types de modèles sont ajustés. D'une part, les capturabilités sont regroupées dans différents fichiers classés par stock ; pour chaque stock, les effets statistiques An, Flottille, et l'effet d'interaction An*Flottille sont testés. D'autre part, les capturabilités sont regroupées par flottille, et sont alors testés pour chaque flottille, les effets An, Stock, et l'interaction An*Stock. La part de chaque effet dans la variabilité totale et sa significativité sont quantifiées par analyse de variance. Les différentes moyennes utilisées par l'analyse de variance, moyennes cellulaires et marginales, sont estimées par la fonction "LsMeans" du logiciel, suivant le critère des moindres carrés non pondérés (Searle et al, 1980 ; Littell et al, 1987). Enfin, la nature multiplicative des processus en cause et la distribution asymétrique des variables étudiées conduisent à travailler sur le logarithme des capturabilités.

Les différents modèles linéaires appliqués peuvent être formalisés de la manière suivante. Soit : $Y_{i,j,k} = \text{Ln}(q_{i,j,k} \cdot 10^7 + 1)$, avec $q_{i,j,k}$ la capturabilité d'un stock i , l'année j , relative à la flottille k . Le multiplicateur 10^7 est utilisé pour minimiser l'influence de l'unité ajoutée, les valeurs de capturabilité étant toujours très proches de zéro, de l'ordre de 10^{-6} à 10^{-8} .

Le modèle par stock s'écrit, pour un stock i donné³:

$$E(Y_{j,k,n}) = \mu_{j,k} = \mu + \alpha_j + \beta_k + \gamma_{j,k} \quad (2)$$

avec : - $Y_{j,k,n}$, la $n^{\text{ième}}$ observation de Y (pour le $n^{\text{ième}}$ navire) correspondant à la combinaison (j, k) ;

- μ le terme moyen ;

- α_j et β_k les effets simples dues respectivement, à l'année j et la flottille k ;

- $\gamma_{j,k}$ le terme d'interaction.

De la même façon, le modèle par flottille s'écrit, pour une flottille k donnée⁴ :

$$E(Y_{i,j,n}) = \mu_{i,j} = \mu + \alpha_i + \beta_j + \gamma_{i,j} \quad (3)$$

avec : - $Y_{i,j,n}$, la $n^{\text{ième}}$ observation de Y correspondant (pour le $n^{\text{ième}}$ navire) à la combinaison (i, j) ;

- μ le terme moyen ;

² Plus précisément, les groupes scientifiques estiment des mortalités pour chaque groupe d'âge et en déduisent une mortalité moyenne, calculée pour les principaux groupes d'âge exploités.

³ Ce modèle par stock correspond à un modèle d'analyse de variance de type : $Y_{j,k} = E_j + E_k + E_{j,k}$, avec E_j , E_k , $E_{j,k}$ les effets statistiques correspondant respectivement aux facteurs An, Flottille, et à l'interaction An*Flottille.

⁴ Ce modèle par flottille correspond à un modèle d'analyse de variance de type : $Y_{i,j} = E_i + E_j + E_{i,j}$, avec E_i , E_j , $E_{i,j}$ les effets statistiques correspondant respectivement aux facteurs Stock, An, et à l'interaction Stock*An.

- α_i et β_j les effets simples dues respectivement, au stock i et à l'année j ;
- $\gamma_{i,j}$ le terme d'interaction.

Interprétation des effets

Trois analyses sont conduites successivement.

La première découle de la modélisation par stocks et permet d'étudier l'évolution de l'efficacité de l'heure de pêche appliquée à chacun des stocks pris en compte. On mesure ainsi l'évolution tendancielle des capturabilités annuelles moyennes, toutes flottilles confondues (effet statistique An). Théoriquement, cette évolution peut être le fait de variations de la puissance de pêche des navires, de la disponibilité des stocks, ou bien sûr, des deux à la fois. Néanmoins, sous l'hypothèse d'une évolution sans tendance de la disponibilité, l'évolution inter-annuelle des capturabilités est interprétée comme une dérive des puissances de pêche. Pour chaque stock étudié, l'effet statistique An permet ainsi d'analyser l'évolution tendancielle de la puissance de pêche globale, c'est à dire de l'efficacité réelle de l'heure de pêche d'un navire, toutes flottilles confondues.

La seconde analyse est commune aux deux types de modèles, par stock et par flottille. Elle conduit à calculer les "moyennes cellulaires" des modèles, c'est à dire les capturabilités moyennes par stock, par flottille et par an. Sous l'hypothèse ici aussi d'une fluctuation sans tendance de la disponibilité des différentes espèces, l'évolution de ces capturabilités rend compte des changements de puissance de pêche spécifique. On analyse ainsi comment chaque flottille est devenue plus ou moins efficace vis-à-vis de chacune des espèces étudiées.

Enfin, la troisième analyse découle des modèles par flottilles. Elle conduit à étudier cette évolution en termes d'allocations spécifiques des puissances de pêches, au travers de l'effet d'interaction An*Stock. Pour chaque flottille considérée, l'évolution globale de l'efficacité de pêche, toutes espèces confondues, est prise en compte dans l'effet statistique An. L'effet Stock est interprété comme un indice de disponibilité, constant sur la période. L'effet An*Stock, quant à lui, permet de mesurer les différences d'évolution de la capturabilité d'un stock à l'autre. On calcule ainsi un indice d'allocation de la puissance de pêche, noté APg, à partir des estimateurs de moyennes marginales, correspondant aux effets simples, et de moyennes cellulaires, correspondant à l'effet croisé. Ces estimateurs (notés LSM, pour "LsMeans") sont calculés de la manière suivante (Searle et al., 1980).

A partir du modèle (3), $E(Y_{i,j,n}) = \mu_{i,j} = \mu + \alpha_i + \beta_j + \gamma_{i,j}$:

$$\begin{aligned}
 LSM(\alpha_i) &= \sum_{j=1}^J \frac{\hat{\mu}_{i,j}}{J} = \sum_{j=1}^J \frac{\bar{Y}_{i,j.}}{J} \\
 LSM(\beta_j) &= \sum_{i=1}^I \frac{\hat{\mu}_{i,j}}{I} = \sum_{i=1}^I \frac{\bar{Y}_{i,j.}}{I} \\
 LSM(\gamma_{i,j}) &= \hat{\mu}_{i,j} = \bar{Y}_{i,j.}
 \end{aligned}
 \tag{4}$$

Un facteur correctif (Laurent, 1963) est appliqué lors de la transformation inverse, effectuée pour le retour aux valeurs brutes. L'indice APg, correspondant à l'effet d'interaction seul, s'exprime alors comme suit :

$$APg_{i,j,estimée} = \frac{[\exp(LSM(\gamma_{i,j}) + \frac{\sigma_{i,j}^2}{2}) - 1]}{[\exp(LSM(\alpha_i) + \frac{\sigma_i^2}{2}) - 1] \cdot [\exp(LSM(\beta_j) + \frac{\sigma_j^2}{2}) - 1]} \quad (5)$$

avec :

- $LSM(\alpha_i)$, $LSM(\beta_j)$, $LSM(\gamma_{i,j})$, les estimateurs de moyennes marginales, relatifs respectivement aux effets Stock, An et Stock*An ;
- σ_i , σ_j , $\sigma_{i,j}$, les erreurs standards d'estimation associées ;

L'effet d'interaction An*Stock est ainsi interprété comme une mesure de la répartition de la puissance de pêche des navires, entre les différents stocks considérés. Autrement dit, pour chaque flottille, caractérisée par une efficacité globale donnée, on analyse la manière dont l'heure de pêche est plus ou moins efficace, vis-à-vis de chacun des différents stocks exploités. Cette analyse permet d'identifier des stratégies de pêche par flottille, ciblées sur telle ou telle espèce parmi celles prises en compte, et d'étudier l'évolution de ces stratégies d'année en année.

Le modèle de capturabilité sur lequel repose cette dernière analyse peut donc être formalisé de la manière suivante :

$$q_{i,j,k} = ID_{i,k} \cdot IE_{j,k} \cdot APg_{i,j,k} \quad (6)$$

- avec :
- $q_{i,j,k}$: la capturabilité moyenne par stock i , année j , et relative à la flottille k ;
 - $ID_{i,k}$: l'indice de disponibilité du stock i , supposé constant sur la période relativement à la flottille k ⁵ ;
 - $IE_{j,k}$: l'indice d'efficacité globale de la flottille k , pour l'année j ;
 - $APg_{i,j,k}$: l'indice d'allocation spécifique de la puissance de pêche de la flottille k , sur le stock i , par année j .

Ceci revient à considérer un modèle de puissance globale de pêche Pg à deux facteurs :

$$Pg_{i,j,k} = IE_{j,k} \cdot APg_{i,j,k} \quad (7)$$

Cette représentation de l'efficacité de pêche individuelle consiste donc finalement à scinder en deux les causes de ses variations : entre, d'une part, l'appartenance à telle ou telle flottille, plus ou moins efficace de façon globale par rapport aux autres flottilles, et d'autre part, l'identité du stock auquel se rapporte cette puissance de pêche, au sein d'une même flottille.

⁵ Soit : $ID_{i,k} = \exp[LSM(\alpha_i) + \frac{\sigma_i^2}{2}] - 1$

⁶ Soit : $IE_{j,k} = \exp[LSM(\beta_j) + \frac{\sigma_j^2}{2}] - 1$

UN EXEMPLE D'APPLICATION : LA FLOTTILLE INDUSTRIELLE DE LORIENT.

La méthode présentée a été utilisée sur les données des flottilles industrielles de Bretagne Sud (Millischer et al., 1999 ; Millischer, 1996). Nous ne présentons ici que les résultats concernant la flottille industrielle de Lorient, sur la période 1983-1994. Les stocks considérés sont la Morue, le Merlan, l'Eglefin et le Lieu Noir, capturés à l'ouest de l'Ecosse.

Les paramètres d'ajustement des différents modèles sont donnés dans les tableaux 1 et 2. Ceux-ci sont globalement très significatifs, à 99.99% de certitude. Les modèles par stock permettent d'expliquer entre 49 et 70% de la variabilité totale de l'échantillon des capturabilités, selon le stock considéré. Le modèle par flottille, appliqué à la flottille de Lorient, permet d'expliquer quant à lui, 91% de la variabilité totale.

Stocks	Effets testés	ddl	% d'inertie	P level
Morue	Modèle	35	48.7	0.0001
zone VI	Effet Flottille	2	43.1	0.0001
	Effet An	11	42.9	0.0001
n = 706	Effet croisé	22	14.0	0.0001
Merlan	Modèle	35	59.9	0.0001
zone VI	Effet Flottille	2	2.7	0.0001
	Effet An	11	87.4	0.0001
n = 706	Effet croisé	22	9.9	0.0001
Eglefin	Modèle	35	63.1	0.0001
zone VI	Effet Flottille	2	21.7	0.0001
	Effet An	11	73.2	0.0001
n = 706	Effet croisé	22	5.1	0.0001
Lieu Noir	Modèle	47	68.7	0.0001
zone VI	Effet Flottille	3	80.6	0.0001
	Effet An	11	12.9	0.0001
n = 778	Effet croisé	33	6.5	0.0001

Tableau 1. - Paramètres d'ajustement des modèles par stocks (n : nombre total d'observations ; ddl : nombre de degrés de liberté ; zone VI : zone CIEM Nord-Ouest Ecosse). Les pourcentages d'inertie expliquée sont donnés : relativement à la variance totale pour le modèle complet ; relativement à la variance expliquée par le modèle pour chacun des facteurs du modèle (effet Flottille, effet An et effet croisé).

Flottilles	Effets testés	ddl	% d'inertie	P level
IND Lorient	Modèle	59	90.9	0.0001
	Effet Stock	4	90.3	0.0001
n = 1204	Effet An	11	5.1	0.0001
	Effet croisé	44	4.6	0.0001

Tableau 2. - Paramètres d'ajustement des modèles par flottille (n : nombre total d'observation ; ddl : nombre de degrés de liberté ; IND LO : Industriels de Lorient). Les pourcentages d'inertie expliquée sont donnés : relativement à la variance totale pour le modèle complet ; relativement à la variance expliquée par le modèle pour chacun des facteurs du modèle (effet Stock, effet An et effet croisé).

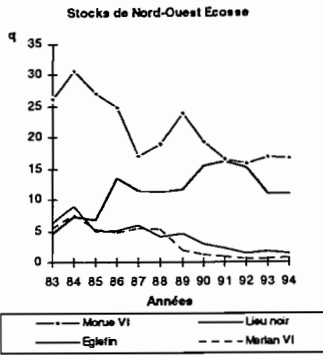


Figure 1 : Evolution des capturabilités spécifiques moyennes relatives aux flottilles industrielles de Bretagne Sud (effet An des modèles par Stocks).

Dans un premier temps l'analyse préliminaire par stocks montre que les stocks d'Eglefin, de Morue et de Merlan de la zone Nord-Ouest Ecosse présentent une forte décroissance de capturabilité sur l'ensemble de la période 1983-1994, traduisant une chute de l'efficacité de l'heure de pêche moyenne d'un navire industriel sur ces stocks (fig. 1). Cette décroissance est de l'ordre de -5.3% par an pour la Morue, de -15.1% par an pour l'Eglefin, et de -23.6% par an pour le Merlan. A l'inverse, le stock de Lieu Noir présente un accroissement très important de sa capturabilité entre 1983 et 1991, de l'ordre de 14.5% par an.

Ces résultats font ainsi apparaître un transfert d'efficacité de l'heure de pêche moyenne, pour l'ensemble des flottilles industrielles exploitant la zone : depuis les stocks de Morue, de Merlan et d'Eglefin, vers le stock de Lieu Noir.

Dans un second temps, l'analyse par flottille appliquée aux navires de Lorient permet de quantifier la part, en termes de puissance de pêche, que cette flottille a eu dans ces variations globales de capturabilités. Ainsi, l'examen des capturabilités par navire moyennes (fig. 2 et tab.3), nous donne l'évolution de l'efficacité de l'heure de pêche d'un navire "moyen" de la flottille lorientaise. On a ainsi une indication des variations de puissance de pêche de cette flottille, et donc de l'impact réel de son activité sur chacun des stocks étudiés.

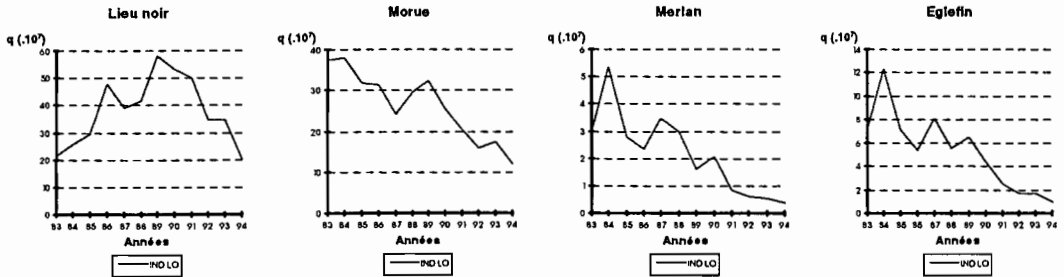


Figure 2 : Evolution de la capturabilité moyenne par navire du Lieu Noir, de la Morue, du Merlan et de l'Eglefin, relatives à la flottille industrielle de Lorient.

Ainsi, la puissance individuelle de pêche moyenne développée sur le Lieu Noir de N-W Ecosse présente une séquence en deux temps : un accroissement de 16.5% par an entre 1983 et 1989, suivi d'une forte dérive négative, de l'ordre de -18% par an entre 1989 et 1994. La chute de l'efficacité de pêche sur les stocks de Morue, de Merlan et d'Eglefin, observée pour l'ensemble des flottilles de Bretagne sud, se retrouve au niveau de la flottille de Lorient, avec des taux moyens de croissance annuelle respectifs de -9%, -19%, -17.5%.

Lieu Noir	Morue	Eglefin	Merlan
83-89: +16.5 %			
89-94: -18.0 %	-8.9 %	-17.5 %	-19 %

Tableau 3. - Taux d'accroissement inter-annuels des capturabilités individuelles spécifiques de la flottille industrielle de Lorient.

Aux vues de telles variations de la puissance de pêche des navires, il semble raisonnable, pour les expliquer, de penser à des phénomènes de changement de stratégie de la part des pêcheurs, se traduisant par des reports de puissance de pêche d'un stock sur un autre. L'analyse des allocations de puissance de pêche (fig.3) permet en effet d'interpréter ces variations d'efficacité moyenne en termes stratégiques, et notamment d'identifier les orientations stratégiques adoptées par la flottille au cours de la période considérée.

Le schéma d'évolution de la flottille lorientaise comporte ainsi deux séquences successives. Entre 1983 et 1989, on observe un accroissement important de la puissance de pêche orientée sur le Lieu Noir ; en sept ans, cette puissance spécifique est multipliée par trois. Une telle dérive se fait, comme le montre la figure 3, au détriment des puissances allouées aux stocks d'Eglefin et de Merlan, qui sont progressivement délaissés par la flottille. L'allocation de puissance relative à la Morue reste quant à elle approximativement constante. A partir de 1989, le Lieu Noir semble à son tour délaissé, comme l'indique la dérive négative de -18% par an de l'efficacité de pêche correspondante. On passe ainsi d'une stratégie "Lieu Noir" à une stratégie de pêche orientée sur d'autres espèces, et notamment sur les espèces profondes, qui ne sont pas prises en compte dans la présente analyse.

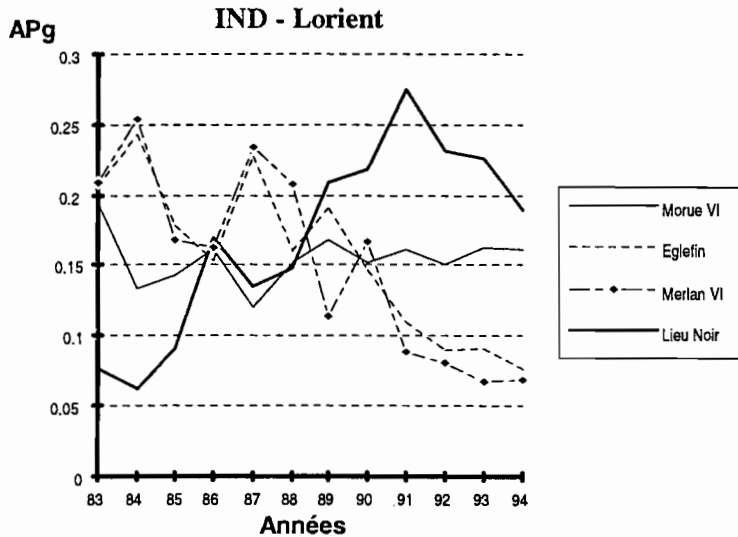


Figure 3 : Evolution des allocations spécifiques de puissances de pêche individuelles (Interaction An*Stock des modèles par flottilles), de la flottille industrielle de Lorient.

On note qu'entre 1989 et 1991, l'allocation spécifique de puissance "Lieu Noir" augmente, alors que la puissance spécifique correspondante décroît. Ceci s'explique par la chute plus importante des puissances de pêche orientées sur la Morue, l'Eglefin et le Merlan. En effet, il faut rappeler que l'indice d'allocation donne une valeur relative de la puissance de pêche en fonction des autres stocks considérés.

Ces fortes variations de puissance de pêche mettent en évidence d'importants processus de régulation au sein de la pêcherie. Ainsi, la période allant de 1983 à 1988-1989 correspond à une chute importante de l'abondance des stocks d'Eglefin, de Merlan et de Morue dans la zone Nord-Ouest Ecosse (Anonyme, 1995), voit les navires effectuer un report de puissance (un transfert d'efficacité de pêche) vers le stock de Lieu Noir. Il est à noter que ce report se fait à effectif de la flottille à peu près constant (autour de 23 navires à Lorient), et n'est donc pas le fait d'un arrêt

d'activité de certains pêcheurs. Au contraire, la chute d'abondance du Lieu noir, qui commence sensiblement en 1986, est suivie par un renforcement de la stratégie "Lieu noir" de la part des Lorientais, confirmant les tendances des premières années de la période.

A partir de 1989, le stock de Lieu Noir ayant été presque divisé par deux en deux ans, l'effectif de la flottille régresse fortement, jusqu'à 13 navires. Cette seconde période semble donc correspondre à une sélection des navires les plus efficaces, le Lieu Noir restant une espèce cible.

Parallèlement se développe à Lorient l'exploitation du Grenadier, qui peut expliquer la baisse de la puissance individuelle de pêche moyenne "Lieu Noir", et de l'allocation APg correspondante, dans les dernières années (Biseau, 1996). En ce sens, Lorient aurait adopté une stratégie de différenciation des captures, en réponse à la chute d'abondance du Lieu Noir, alors que la flottille de Douarnenez par exemple, est restée spécialisée sur le Lieu Noir.

CONCLUSION.

La principale limitation de cette méthode d'estimation des puissances de pêche réside en ce qu'elle nécessite l'existence d'évaluations des coefficients de mortalité par pêche pour chacun des stocks. Ainsi, l'absence d'évaluation sur le stock de Grenadier et sur les stocks de Lingues franches et bleues empêche de quantifier la part effective de ces stocks, en termes de puissances allouées. Pourtant, les Lingues représentent une part importante des captures de Lorient sur toute la période, et le développement de la pêche du Grenadier est une des caractéristiques fortes de la période. Il est certain que l'existence de telles évaluations aurait sensiblement amélioré la qualité de l'analyse des allocations de puissances de pêche.

Néanmoins, la méthode présentée permet de mettre en évidence de fortes variations de puissances de pêche au sein des flottilles étudiées. Ce résultat n'est pas négligeable, connaissant les hypothèses couramment émises concernant le modèle de capturabilité, qui soutiennent les méthodes de calibration de l'analyse des cohortes : modèle constant, modèle linéaire, ou modèle exponentiel. D'une façon générale, la dérive des capturabilités est considérée comme une simple perturbation, pouvant être, sinon négligée, du moins ramenée à un processus très secondaire. Cette conception reste valable dans de nombreux cas. Toutefois, cette étude montre que cette dérive peut également constituer un aspect essentiel de l'histoire d'une pêcherie.

Enfin, l'analyse a permis de traduire ces variations d'efficacité en termes de comportements stratégiques des pêcheurs lorientais : mesurant les effets, nous avons déduits les causes probables. Ainsi, cette étude montre bien la nécessité, qui se pose aujourd'hui à la recherche halieutique, de l'étude de ces comportements de pêche, qui constituent l'un des déterminants majeurs de l'évolution des pêcheries.

REFERENCES BIBLIOGRAPHIQUES.

- Anonyme, 1995. Report of the working Group on the Assesment of the Northern Shelf Demersal Stock. *ICES CM 1996/Assess:1*.
- Beverton R.J.H., Holt S.J., 1957. On the dynamics of exploited fish populations. *Fishery Invest., Lond. Ser. 2, 19:533 pp.*

- Biseau A., 1996. Definition of a directed fishing effort, and its impact on CPUE trends : the case of the French demersal fishery off the west coast of Scotland *ICES CM 1996/G:17*.
- Gascuel D., 1993. Efforts et puissances de pêche : redéfinition des concepts et exemple d'application. in : *Actes du Premier Forum Halieumétrique, ORSTOM Editions coll. colloques et séminaires, 159-181*.
- Gascuel D., Thiam M., 1994. Evolution de l'abondance des ressources démersales sénégalaises : estimation par modélisation linéaire des P.U.E. in : *L'évaluation des ressources exploitées par la pêche artisanale au Sénégal, Symposium, Dakar, 8 au 13 février 1993*.
- Kimura D.K., 1981. Standardized measures of relative abundance based on modelling log(c.p.u.e.), and their application to Pacific ocean perch (*Sebastes alutus*). *J. Cons. int. Explor. Mer*, 39: 211-218.
- Laurent A.G., 1963. Lognormal distribution and the translation method : description and estimation problems. *J. Amer. Statist. Ass.*, 58 :231-5.
- Laurec A., 1977. Analyse et estimations des puissances de pêche. *J. Cons. Int. Explor. Mer*, 37(2): 173-185.
- Laurec A., Le Guen J.C., 1981. Dynamique des populations marines exploitées. Tome 1 : concepts et modèles. *Rap. Scient. et techn. N°45. Publications du CNEOX*.
- Littel R.C., Freund R.J., Spector P.C., 1991. SAS System for Linear Models, Third edition. Cary, NC: SAS Institute Inc., 1991, 329 pp.
- Lo N.C., Jacobson L.D., Squire J.L., 1992. Indices of Relative Abundance from Fish Spotter Data based on Delta-Lognormal Models. *Can. J. Fish. Aquat. Sci.* 49: 2515-2526.
- Millischer L., 1996. Evolution des puissances et des stratégies de pêche des navires hauturiers de Bretagne sud. *Rapport DAA Laboratoire Halieutique, DEERN, ENSA-Rennes*.
- Millischer L., Gascuel D., Biseau A., 1999. Estimation of the overall fishing powers : a study of the dynamics and fishing strategies of Brittany's industrial fleets, *Aquat. Living Resour.* 12 (2) (1999) 89-103.
- Robson D.S., 1966. Estimation of the relative Fishing Power of individual ships. *ICNAF Research Bulletin*, (3), 1966.
- SAS Institute Inc. SAS/STAT Guide for Personal Computers, Version 6 Edition. Cary, NC : SAS Institute Inc., 1987, 1028 pp.
- Searle S.R., Speed F.M., Milliken G.A., 1980. Population marginal means in the linear model : an alternative to least squares means. *The american statistician*, November 1980, 34 (4), 216-221.

MODÈLES MATRICIELS ET COLONISATION DE MILIEU VIERGE : L'EXEMPLE DE LA TRUITE COMMUNE AUX ÎLES KERGUELEN.

Marc JARRY ¹ Patrick DAVAINÉ ² et Edward BEALL ²

¹ Laboratoire d'Écologie Moléculaire, IBEAS-UPPA, Avenue de l'Université, 64000 Pau.

² INRA, Station d'Hydrobiologie, BP 3, 64310 Saint Pée sur Nivelle

Summary

We present a matrix model for the study of the population dynamics of brown trout *Salmo trutta* L., introduced in the '60s in the virgin aquatic ecosystems of the Kerguelen Islands. This species clearly acclimatized very well : a portion of the population became migratory and spent a part of its life cycle in the sea, which allowed the rapid colonization of two rivers close to the stream of origin in the same bay.

These migratory trout can smolt at 2, 3, or 4 years of age. The model takes into account age and smolt age structures and in a first step considers the fish from the bay as belonging to a single population. The transition matrix looks like a 32 x 32 Leslie matrix in which some survival rates are not on the subdiagonal. They represent survival after the first sea migration and are particularly important for the dynamics of the whole population.

The estimate of demographic parameters was obtained from a data base containing information collected in the field since 1970. The model was calibrated on the population size estimate and the stock structure of the migratory trout in 1979. Population size was estimated by tagging-recapture and monitoring of the migratory trout in freshwater when they returned to overwinter or reproduce. Under the hypothesis of a constant survival rate for all ages and categories of fish, it was possible to determine relationships between the annual population growth rate and survival rates at first downstream migration, for which no direct estimate was available.

These constraints on the model induced paradoxical results. For instance, an increase in survival rate reduced migratory trout numbers in the first years of colonization. These data suggested that the average survival rate should be around 0.30 and not 0.50 as surmised in previous studies. However, the model systematically underestimated numbers of migratory trout during the first years of development. Thus, to improve the model it will be necessary to introduce survival rates varying with time. Another possible approach would be to consider the population as three subunits corresponding to the three colonized rivers.

Résumé

Un modèle matriciel est proposé pour l'étude de la dynamique des populations de la truite commune *Salmo trutta* L., introduite dans les écosystèmes aquatiques vierges de Salmonidés des îles Kerguelen dans les années 1960. Cette espèce s'est remarquablement bien acclimatée : une fraction de la population s'est mise à effectuer une partie de son cycle en mer et a pu ainsi rapidement coloniser deux autres rivières proches de la rivière d'origine situées dans une même baie.

Ces truites migratrices peuvent smoltifier à 2, 3 et 4 ans. Le modèle prend en compte la structure en âge et en catégories de smoltification, mais néglige, dans un premier temps, l'existence possible de trois sous-populations correspondant aux trois rivières de la baie étudiée. La matrice de transition se présente comme une matrice de Leslie de dimension 32×32 dont certains taux de survie ne sont pas dans la sous-diagonale : ils représentent des taux de survie à la première dévalaison et sont particulièrement importants pour la dynamique globale de la population.

L'estimation des paramètres démographiques a été réalisée à partir d'une banque de données rassemblant les informations recueillies depuis 1970 sur le terrain. Le modèle a été calé sur l'estimation de l'effectif et de la structure du stock de truites migratrices en 1979, établie par marquage-recapture et contrôle lors du retour des truites en rivière pour hiverner et se reproduire. Sous l'hypothèse d'un taux de survie constant pour tous les âges et toutes les catégories, on peut étudier les relations entre le taux de croissance annuel de la population et les taux de survie à la première dévalaison dont on n'avait aucune estimation directe disponible.

Ces contraintes imposées au modèle induisent des résultats paradoxaux. Par exemple, l'augmentation du taux de survie diminue les effectifs de truites migratrices dans les premières années de la colonisation. L'ensemble des résultats suggère que ce taux de survie moyen doit être de l'ordre de 0.30 et non de 0.50 comme le laissaient supposer des études précédentes. Il reste que le modèle sous-estime systématiquement les effectifs de truites migratrices au cours des premières années de colonisation. On devra donc envisager d'introduire des taux de survie variables avec le temps. Une autre approche est de considérer cette population structurée en sous-populations correspondant aux trois rivières colonisées.

1. Introduction

Dans le domaine de l'halieutique, on dispose assez souvent de données structurées en âge. Le modèle matriciel de Leslie (1945, 1948) est un outil adapté à ce type de données (Caswell *et al.*, 1984 ; Caswell 1989). Il a été en particulier appliqué à des espèces d'intérêt commercial comme le bar *Morone saxatilis* (Cohen *et al.*, 1983) ou le labre *Tautoglabrus adspersus* (Horst, 1977). Dans ce cadre matriciel, diverses façons de traiter les relations de dépendance à la densité ont été proposées : fonction de compensation pour modéliser la compétition entre juvéniles (De Angelis *et al.*, 1980), fonction de Ricker directement intégrée dans les paramètres de recrutement (Levin & Goodyear, 1980). Chez les Salmonidés, on trouvera des exemples d'application à la gestion du saumon atlantique (Browne, 1988) et de la truite commune (Sabaton *et al.*, 1996).

Une telle approche est appliquée à un problème d'introduction d'une espèce nouvelle, la truite commune *Salmo trutta* L., dans les écosystèmes aquatiques vierges de poissons d'eau douce des îles Kerguelen (Davaine & Beall, 1982a et b, 1997). Cette espèce s'est remarquablement bien acclimatée : spontanément, une fraction de la population s'est mise à effectuer une partie de son cycle en mer et a pu ainsi rapidement coloniser des rivières proches de la rivière d'origine. Ces observations ont particulièrement contribué à éclaircir le statut de la « truite de mer » (Guyomard *et al.*, 1984), toujours considérée par certains comme une sous-espèce (*Salmo trutta trutta*) différente de la « truite sédentaire » (*Salmo trutta fario*) comme le rappelle Elliott (1994).

Parmi les nombreuses perspectives qu'offre une telle modélisation, nous nous restreindrons, dans ce présent travail, à l'estimation des paramètres de survie à la migration en mer qui sont extrêmement difficiles à obtenir directement à partir des données de terrain disponibles.

2. Milieu d'étude, données biologiques et démographiques

2.1. Milieu d'étude : la baie Norvégienne des îles Kerguelen

Les îles Kerguelen (6500 km²) sont situées aux limites de la partie indienne de l'océan Antarctique, ce qui leur confère un environnement marin plus froid que la plupart des îles subantarctiques. En hiver, le couple salinité-température en surface (34 pour mille, 1°C) est trop défavorable pour permettre aux migrateurs amphihalins comme la truite de séjourner en mer (Davaine & Beall, 1997).

La baie Norvégienne du Sud-Est de la péninsule Courbet, abritée des grands courants d'Ouest, peu profonde, présente un gradient de salinité entretenu par l'apport d'eau douce de trois cours d'eau importants, rivière du Château, rivière Norvégienne et rivière des Albatros, drainant environ 200 km² de bassins versants (Thomas *et al.*, 1981). Ces conditions environnementales particulières la rendent plus favorable aux truites de mer qui n'en sont pas sorties pendant les vingt premières années de leur acclimatation, comme l'attestent les campagnes de marquage individuel initiées en 1970.

Les rivières, bien que froides et peu productives, conviennent cependant aux exigences des Salmonidés qui, en l'absence de prédateurs et de compétiteurs aquatiques, exploitent efficacement cette niche écologique vacante (Davaine & Beall, 1997).

2.2. Historique des introductions de la truite commune

Les truites communes acclimatées aux îles Kerguelen sont toutes issues de la même souche de pisciculture du Pays Basque (France), aux environs de Bidarray, sur un affluent de la Nive. Elles ont été importées en mars 1958 sous la forme d'œufs proches de l'éclosion (Davaine & Beall, 1982a ; Davaine & Beall, 1997), élevées dans des bassins de la base de Port aux Français, et les truitelles ont été lâchées dans deux étangs voisins en octobre 1960. Entre mai et octobre 1962, 22 truites de 5 ans ont été déversées dans la rivière du Château, et comme certaines ont été contrôlées mûres au déversement, il n'est pas exclu qu'elles aient frayé en juillet ou août 1962 (cohorte 1962).

Un premier contrôle en 1971 a permis de constater la colonisation des deux autres rivières de la Baie Norvégienne, rivière Norvégienne et rivière des Albatros (Davaine & Beall, 1982b). Des inventaires plus précis sur la rivière Norvégienne ont commencé en 1972. L'analyse des différentes cohortes permet de formuler l'hypothèse que les 3 rivières de la baie Norvégienne ont été colonisées en 1968 par des truites migrantes âgées de 5 ans appartenant à la cohorte de 1963 de la rivière du Château. Des campagnes de marquage intensif des truites de mer en 1978 et 1979 ont montré l'importance des échanges entre rivières (Thomas *et al.*, 1981).

Nous considérerons, dans un premier temps, que l'ensemble des truites de la baie Norvégienne forme une seule population. Cette proposition est *a priori* acceptable au moins jusqu'en 1983, première année pour laquelle on enregistre la colonisation de rivières hors de la baie Norvégienne.

2.3. Cycle de vie

La figure 1 résume les phases les plus importantes du cycle de vie de la truite commune, reconstitué d'après les très nombreuses observations de terrain qui couvrent une période de plus de 20 années (1970-1992, Davaine & Beall, 1997). On notera en particulier que les truites migratrices (dites « truites de mer ») peuvent smoltifier à 2, 3 et 4 ans : on conservera dans la modélisation cette structuration en trois catégories de truites migratrice notées M2, M3 et M4. En général, les femelles migratrices ne commencent à se reproduire qu'à 5 ans,

mais une fraction non négligeable des femelles ayant smoltifié à deux ans peuvent se reproduire à 4 ans.

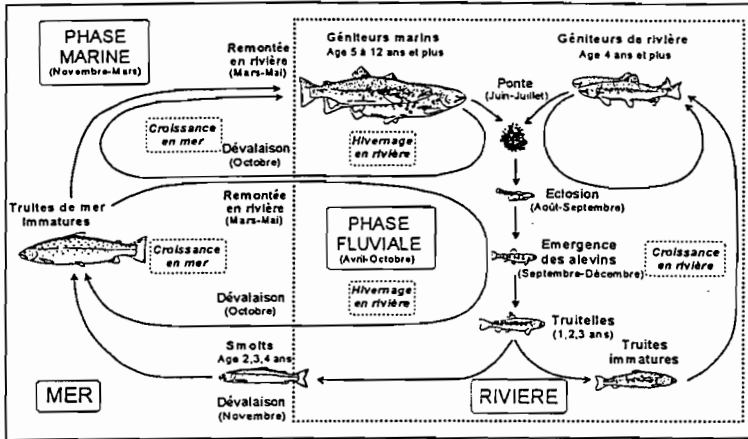


Figure 1. Cycle de vie de la truite commune aux îles Kerguelen.

2.4. Données démographiques

Les données sont présentées en détail dans Jarry *et al.* (1998). On ne rappellera ici que les méthodes d'obtention des paramètres démographiques.

Taux de recrutement

Les taux de recrutement f_k d'une truite d'âge k se décomposent de la façon suivante :

$$f_k = \alpha_k w p_k l \quad (1)$$

où α_k est la fréquence de femelles mûres, w est le nombre moyen d'ovules par gramme de poids corporel, p_k le poids corporel moyen et l la survie de l'ovule à l'alevin (considéré dans le modèle comme le premier groupe d'âge).

La fréquence de femelles mûres α_k a été déterminée par dissection pour les sédentaires et par examen des marques de frai sur les écailles et détermination de l'âge de premier frai chez des femelles migratrices ayant frayé au moins une fois (Beall et Davaine, 1988 ; Beall *et al.*, 1992). Le nombre moyen d'ovules par gramme de poids corporel (w) a été fixé à 3 ovules par gramme d'après les travaux de Thomas *et al.* (1981). La relation entre le poids p_k et l'âge k a été établie par régression linéaire. La survie de l'ovule à l'alevin (l) n'a pu être estimée qu'une seule fois (une femelle de poids connu a frayé dans une zone bien déterminée où on a pu estimer la production en alevins l'année suivante) : elle a été fixée à 0.30.

Structure en catégories et âge des truites migratrices

Pour obtenir une structure d'âge non biaisée par la sélectivité du mode de capture, seules les captures de 1978 et 1979 à la station de contrôle de la rivière Norvégienne ont été considérées.

La structure en catégories (M2, M3, M4) se conserve en 1978 et 1979 ($\chi^2_{ddl} = 0.71$; $P = 0.70$), de même que la structure en âge ($\chi^2_{ddl} = 3.06$; $P = 0.38$). On utilisera donc les valeurs cumulées 1978 + 79 pour quantifier la structure en catégories (Tab. 1) et décrire la structure d'âge en 1979.

Année	Âge de smoltification	Âge de mer				Total
		1	2	3	≥ 4	
1978	M2	27	13	13	6	59
	M3	47	14	2	0	63
	M4	4	0	0	0	4
	Total	78	27	15	6	126
1979	M2	29	4	5	4	42
	M3	36	10	3	1	50
	M4	5	0	0	0	5
	Total	70	14	8	5	97
1978 + 1979	M2	56	17	18	10	101
	M3	86	24	5	1	113
	M4	9	0	0	0	9
	Total	148	41	23	11	223
	%	66.37	18.39	10.31	4.93	100

Tableau 1. Structure en catégories (âge de smoltification) et en âge des échantillons de truites migratrices (mâles et femelles) obtenus en 1978 et 79 par pêche électrique, barrage et piège dans la rivière Norvégienne.

Effectif de truites migratrices en 1979

L'effectif global de truites migratrices de la baie Norvégienne en 1979 a été calculé à partir des recaptures en 1979 des truites marquées en 1978 dans différentes rivières, en tenant compte d'un taux de survie annuel moyen s_M constant quel que soit l'âge ou la catégorie. Il dépend de la survie s_M et de l'efficacité de piégeage E (Jarry *et al.*, 1998).

3. Construction d'un modèle, étude de ses propriétés et algorithme de simulation

3.1. Le modèle matriciel

La construction d'un modèle matriciel impose de choisir un pas de temps et une date de recensement. Le choix d'un pas de temps annuel est cohérent avec le mode de reproduction annuel de l'espèce. La date de recensement est plus délicate à choisir car les inventaires se font à différentes périodes de l'année selon les stades de développement. Nous avons choisi de placer cette date entre les remontées en rivière (mars-mai) et la ponte (juin-juillet). Ce choix nous obligera à donner un âge de 1 an aux alevins traditionnellement notés 0^+ .

Quelques simplifications ont été faites par rapport aux observations de terrain :

- l'âge maximal des femelles sédentaires (R_1 à R_8) a été limité à 8 ans et celui des truites migratrices à 11 ans. Des longévités supérieures sont possibles mais restent assez exceptionnelles ;

- on ne considère que 3 catégories de truites migratrices (smoltification à 2, 3 et 4 ans), notées $M_{i,j}$, d'âge $i + j$, i étant l'âge de smoltification et j l'âge de mer. Les autres catégories ont été considérées comme négligeables (1.4 % du total pour les smolts de 1, 5, 6 et 7 ans).

Les taux d_2 , d_3 et d_4 intègrent le taux de migration et la survie annuelle des individus d'un groupe d'âge qui smoltifient à 2, 3 et 4 ans. Les survies globales aux âges 2, 3 et 4 ans valent donc $s_2 + d_2$, $s_3 + d_3$ et $s_4 + d_4$.

Les taux de recrutement f et les survies s des truites sédentaires R_i sont indexées par un seul indice (i) et celles des truites migratrices $M_{i,j}$ par deux indices i et j .

La matrice de transition A , de dimension 32×32 , s'écrit donc :

$$A = \begin{bmatrix} f_1 & f_2 & f_3 & f_4 & f_i & \cdots & f_{2,j} & \cdots & f_{3,j} & \cdots & f_{4,j} & \cdots \\ s_1 & & & & & & & & & & & \\ & s_2 & & & & & & & & & & \\ & & s_3 & & & & & & & & & \\ & & & s_4 & & & & & & & & \\ & & & & s_i & & & & & & & \\ d_2 & & & & & 0 & & & & & & \\ & & & & & & s_{2,j} & & & & & \\ & & d_3 & & & & & 0 & & & & \\ & & & & & & & & s_{3,j} & & & \\ & & & & d_4 & & & & & 0 & & \\ & & & & & & & & & & s_{4,j} & \\ & & & & & & & & & & & 0 \end{bmatrix}$$

avec $f_1, f_2, f_3, f_{2,j}, f_{3,j} = 0$ et tous les autres éléments non figurés égaux à 0.

Le modèle donnant les effectifs des différentes catégories de truites femelles l'année t par rapport à l'année $t-1$ s'écrit :

$$N(t) = A N(t-1) \quad (2)$$

ou
$$N(t) = A^t N_0 \quad (3)$$

avec $N_0 = 11$, effectif estimé de la population femelle à son introduction dans la rivière du Château en 1962.

3.2 Propriétés du modèle

La matrice positive A est irréductible et primitive. Elle admet donc une valeur propre λ réelle, dominante et strictement positive (Théorème de Perron-Frobenius, *in* Caswell, 1989) qu'on interprétera comme le taux de croissance annuel de la population lorsque la stabilité de la structure en âge des différentes catégories est atteinte. Le vecteur propre associé à λ donnera la distribution des âges des différentes catégories de truites.

Cette propriété nous permet d'établir des relations entre les taux de survie-migration, la structure en âge de smoltification, et le taux de croissance λ . En effet, en régime stable :

$$M_{21} = \frac{d_2}{\lambda} R_2 ; M_{31} = \frac{d_3 s_2}{\lambda^2} R_2 ; M_{41} = \frac{d_4 s_2 s_3}{\lambda^3} R_2. \quad (4)$$

En posant $\frac{M_{31}}{M_{21}} = a$ et $\frac{M_{41}}{M_{21}} = b$ on déduit de (4) :

$$d_3 = a\lambda \frac{d_2}{s_2} ; d_4 = b\lambda \frac{d_2}{s_2 s_3}. \quad (5)$$

On peut montrer que les constantes a et b peuvent être estimées à partir des rapports des totaux des truites des différentes catégories de smoltification, données disponibles dans le tableau 1 (Jarry *et al.* 1998).

3.3. Un algorithme pour la simulation

On supposera, en analyse exploratoire, un taux de survie constant s pour les truites sédentaires et migratrices, indépendant de l'âge et des catégories de smoltification ce qui impose :

$$s_2 + d_2 = s_3 + d_3 = s_4 + d_4 = s \quad (6)$$

Les taux de recrutement f étant calculés par régression linéaire, les simulations numériques se font en deux étapes :

- pour a , b et s donnés, on introduit différentes valeurs de d_2 et une valeur initiale de λ . Le programme calcule les paramètres manquants de la matrice A par les relations (6) et (5), puis la valeur propre dominante de la matrice A ; cette valeur est réintroduite en entrée et le processus est répété jusqu'à ce que les deux valeurs de λ convergent pour une précision donnée. Un programme de simulation utilisant la relation (2) donne alors année par année l'état de la population ;

- on recherche la (les) valeur(s) de d_2 compatible(s) avec la fourchette d'estimation de la population de truites migratrices faite en 1979, calculée pour deux valeurs extrêmes de l'efficacité de piégeage $E_{\min} = 0.65$ et $E_{\max} = 0.90$. Ces différents programmes sont écrits à l'aide du logiciel *Mathematica*®.

4. Résultats et discussion

	Taux de survie		
	$s = 0.30$	$s = 0.40$	$s = 0.50$
d_2	0.02805	0.00646	0.00081
d_3	0.17847	0.03127	0.00355
d_4	0.18089	0.0115	0.00112
% M	0.58	0.17	0.02
Rap. M/R	27.93	0.71	0.06
% BM	35.76	9.22	1.11
% Rec.M	88.94	29.65	3.60

Tableau 2. Estimation des taux de survie-migration et importance des truites migratrices femelles pour différents taux de survie (valeurs estimées par le modèle).

- % M : pourcentage de truites migratrices par rapport à la population totale ;
- Rap. M/R : rapport migratrices/sédentaires, les truites sédentaires étant constituées des catégories R4 à R8, les stades juvéniles R1 à R5 étant indifférenciés ;
- % BM : pourcentage de la biomasse totale dû aux truites migratrices ;
- % Rec.M : pourcentage de recrutement dû aux truites migratrices.

L'absence de différence significative dans la structure en âge des échantillons de truite de mer obtenus en 1978 et 79 (Tab. 1) était un bon argument en faveur de l'existence d'une structure en âge stabilisée. Les résultats des simulations indiquent clairement que cet équilibre n'est pas encore atteint. Il est toutefois possible que les fluctuations par rapport à la structure

en âge asymptotique soient du même ordre de grandeur que les fluctuations d'échantillonnage, ce qui expliquerait l'absence de différence significative entre les échantillons de 1978 et 79.

Les valeurs obtenues pour les taux de survie-migration décroissent lorsque le taux de survie augmente (Fig. 2). Cette décroissance est particulièrement forte pour les taux correspondant aux smoltifications à 3 et 4 ans. Toutes ces valeurs sont, par construction, compatibles avec la structure en âge de smoltification des truites migratrices observée en 1978 et 79, mais la variation de s induit des modifications importantes de la structure globale de la population et de la contribution des truites migratrices au recrutement (Tab. 2).

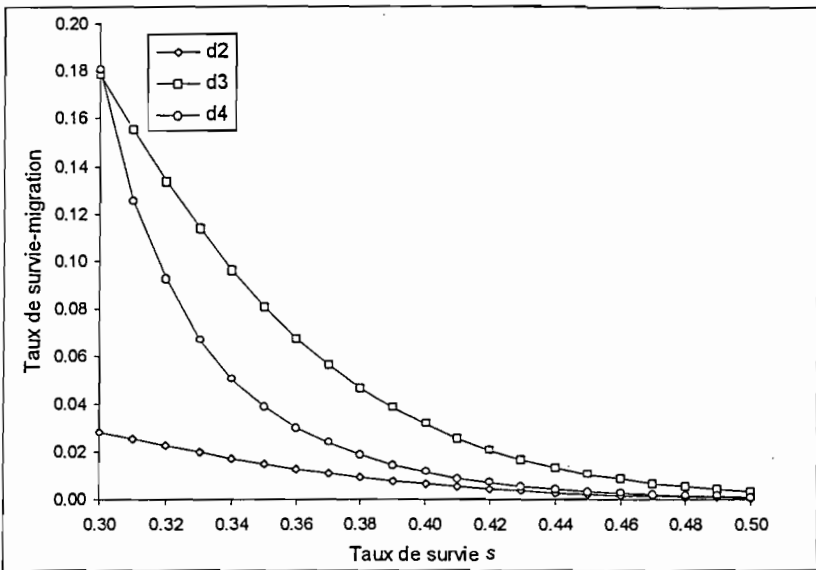


Figure 2. Évolution des taux de survie-migration en fonction du taux de survie s (constant pour toutes les catégories et classes d'âge de truite).

Malgré des taux de survie-migration faibles, les truites de mer peuvent être d'une grande importance dans le recrutement, ce qui est en accord avec les observations de terrain. En 1979, dans la rivière Norvégienne, 62 % des alevins provenaient d'une seule frayère de truite de mer (Thomas *et al.*, 1981). Ces valeurs observées montre qu'il est peu raisonnable de proposer des survies moyennes supérieures à 0.40. Dans l'ensemble, les prédictions du modèle sont satisfaisantes pour des taux de survie moyens proches de 0.30 ; les résultats deviennent de plus en plus mauvais lorsqu'on augmente ce taux de survie.

Le taux de croissance de la population (λ) augmente avec le taux de survie, tout en restant toujours supérieur à 1.5 pour l'intervalle du taux de survie considérée (0.30-0.50). La dynamique globale est donc de type exponentiel, mais, compte tenu des relations entre λ et les taux de survie-migration (d_2 , d_3 et d_4), la dynamique des truites migratrices dépend aussi du taux de survie s : les taux de survie-migration étant relativement plus forts pour des valeurs peu élevées de s (Fig. 2), les effectifs de truites migratrices sont, jusqu'en 1974, d'autant plus grands que s est petit (Fig. 3). En 1978, les trois simulations donnent des résultats très proches ; à partir de cette date, les effectifs de truites migratrices sont d'autant plus forts que le taux de survie moyen est fort.

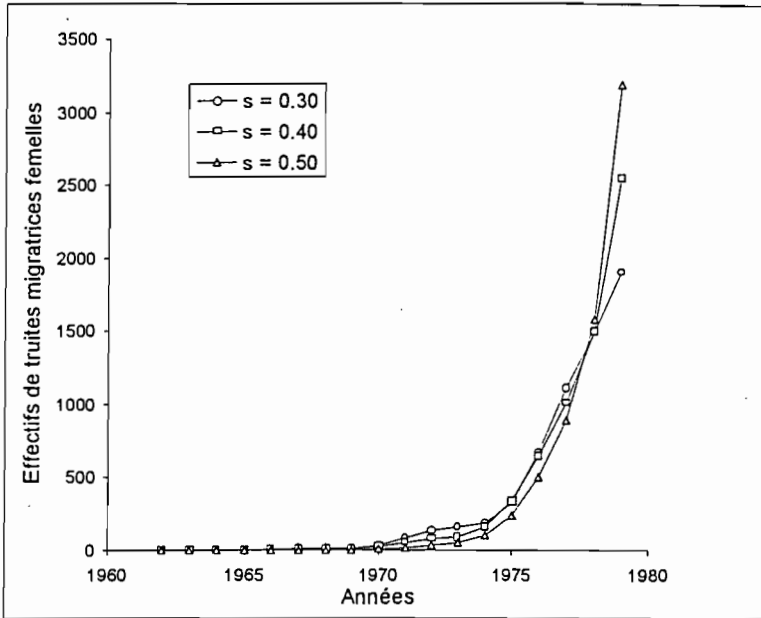


Figure 3. Simulation de l'évolution des effectifs annuels de truites migratrices femelles de la baie Norvégienne de 1962 à 1979 pour différentes valeurs du taux de survie s (0.3, 0.4 et 0.5), constant pour tous âges et toutes catégories de truite.

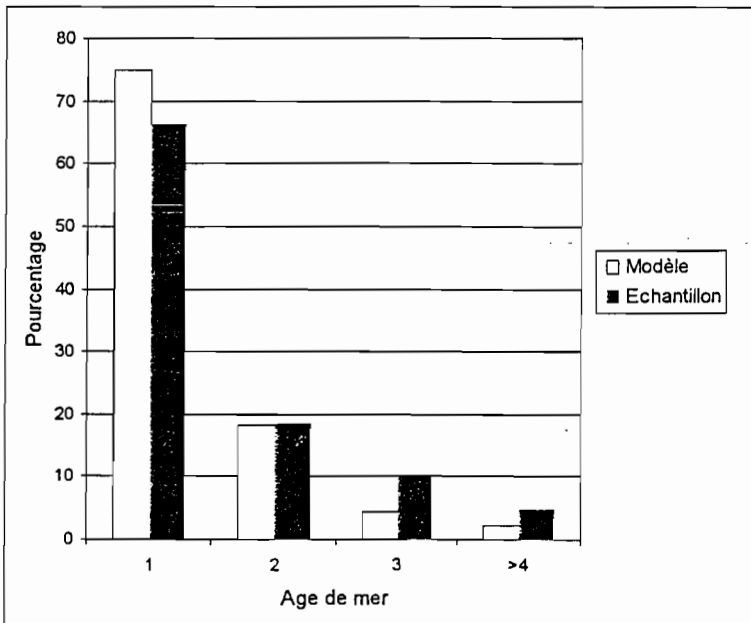


Figure 4. Structure en âge des truites migratrices femelles de la baie Norvégienne en 1979 :
 - □ : valeurs estimées par le modèle ($s = 0.30$) ;
 - ■ : valeurs calculées d'après les deux échantillons de 1978 et 1979.

En 1971, 127 truites migratrices femelles ont été capturées par différentes techniques (pêche à la ligne, pêche électrique...). Le modèle donne comme estimations 86.22, 51.01 et 19.76 pour des survies respectives de 0.30, 0.40 et 0.50. Il y a donc une très forte sous-estimation qui ne peut être compensée (bien au contraire) par une augmentation du taux de survie moyen.

La structure en âge estimée par le modèle est d'allure comparable à celle fournie par les échantillons de 1978 et 79, mais avec une surestimation des truites d'un an et une sous estimation des truites de plus de 3 ans (Fig. 4). Nous n'avons présenté qu'un exemple correspondant à une survie de 0.30 car cette structure est peu sensible à ce paramètre. Là encore, l'augmentation du taux de survie n'améliore pas l'ajustement (le pourcentage de truites migratrices d'âge ≥ 4 ans passe de 2.37 à 1.95 lorsque s augmente de 0.30 à 0.50).

5. Conclusion

Dans l'état actuel d'avancement de ce travail, ce modèle fonctionne surtout comme un outil d'analyse des données. Il s'agit en effet de faire le bilan des données de terrain obtenues au cours des diverses missions et d'en apprécier la fiabilité en terme d'estimations de paramètres démographiques.

L'hypothèse d'une survie égale pour toutes les catégories de truite et tous les âges nous a permis de mettre clairement en évidence les relations entre taux de croissance et taux de survie-migration pour une structure en âge de smoltification des truites migratrices donnée. Ces relations sont la cause d'effets paradoxaux du modèle (l'effectif de truites migratrices peut décroître alors que le taux de survie en mer augmente). Cependant, cette hypothèse est peu réaliste.

Les résultats montrent que la population était encore (en 1978-79) en expansion (voire en très forte expansion) et que le taux apparent de survie est alors un très mauvais estimateur de la survie. Il semble donc que des taux de survie en mer de l'ordre de 0.50, comme cela avait été proposé (Thomas *et al.*, 1981) ne soient pas raisonnables.

Avec un taux de survie constant de 0.30, on obtient des prédictions relativement correctes pour la proportion de truites migratrices et leur importance dans le recrutement. Ce taux donne cependant une très mauvaise estimation de la population de 1971.

Il est donc très probable que ces taux de survie aient dû évoluer au cours de la croissance de cette population. La survie aurait pu être relativement forte dans les premières années (absence de prédateurs, pas de pêche de loisirs) puis diminuer au cours du temps (évolution du comportement des prédateurs... dont l'homme).

On sait également que l'un des paramètres très sensibles de la dynamique des Salmonidés migrateurs est le taux de survie embryo-larvaire, que nous avons étendu ici au stade alevin. L'analyse de la sensibilité du modèle à ce paramètre est à faire et quelques simulations seront tentées, en tenant compte des effets possibles de brusques variations des conditions climatiques. Par exemple, l'hiver 1977 a été particulièrement rude et il a été compté très peu d'alevins en 1978 (Davaine & Beall, 1982b).

Ces premiers résultats sont cependant très encourageants et confirment l'intérêt de l'approche matricielle dans des problèmes de dynamique des Salmonidés. Les auteurs utilisent souvent ce cadre pour la facilité de mise en œuvre de simulations (Browne, 1988, Sabaton *et al.*, 1996). Ce type de modèle est également utilisé pour l'estimation de paramètres difficiles à mesurer sur le terrain, mais généralement sous l'hypothèse restrictive d'une population à l'équilibre (par exemple Sabaton *et al.*, 1996, pour la survie des 0⁺) ; il est alors paradoxal de modifier ensuite le modèle et d'obtenir des taux de croissance différents de 1. Lorsque cela est possible, il est préférable de travailler avec une hypothèse plus large et d'estimer à la fois le paramètre recherché et le taux de croissance (Jarry *et al.*, 1996a et b).

D'autres perspectives s'ouvrent avec la prise en compte de la structuration spatiale (3 sous-populations correspondant aux trois rivières de la baie Norvégienne) en utilisant des matrices multi-sites dont l'analyse mathématique a fait des progrès récents (Lebreton, 1996). On pourra enfin introduire des effets de dépendance à la densité pour étendre cette étude aux années postérieures à 1979, période au cours de laquelle la truite commune a colonisé une grande partie de la péninsule Courbet.

Remerciements

Les campagnes aux Îles Kerguelen ont été financées par la Mission de Recherche des Territoires des Terres Australes et Antarctiques Française (TAAF), puis par l'Institut Français pour la Recherche et la Technologie Polaires (IFRTP). Ce travail a été effectué dans le cadre des programmes du GDR-CNRS n° 1107, Modèles et Théories pour le Contrôle de Ressources Vivantes et la Gestion de Systèmes Écologiques (CoRev) et du Pôle de Recherches DRAES (Dynamique des Ressources Aquatiques en Environnement Sensible).

Références

- Beall, E. et P. Davaine (1988). Analyse scalimétrique de la truite de mer (*salmo trutta* l.) : formation des anneaux et critères d'identification chez les individus sédentaires et migrateurs d'une même population acclimatée aux îles kerguelen (TAAF). *Aquat. Living Resour.* 1 : 3-16.
- Beall, E., P. Davaine et D. Bazin (1992). Étude scalimétrique de la truite commune à Kerguelen, TAAF. Principales difficultés et validation. In : Tissus Durs et Âge Individuel des Vertébrés. Colloque national, Bondy, France (4-6 mars 1991), Baglinière J.L., Castanet J., Conand F., Meunier F.J. (Eds), Colloques et Séminaires ORSTOM-INRA : 199-210.
- Browne, J. (1988). The use of Leslie matrices to assess the Salmon population of the river Corrib. In *Atlantic Salmon : Planning for the Future*, D. Mills and D. Piggins eds, Proc. 3th Int. Atlantic Salmon Symp., Croom Helm, London : 275-300.
- Caswell, H. (1989). *Matrix population models*. Sinauer, Sunderland, Massachusetts, USA, 328 p.
- Caswell, H., R.J. Naiman and R. Morin (1984). Evaluating the consequences of reproduction in complex life cycles. *Aquaculture*, 43 : 123-134.
- Cohen, J.E., S.W. Christensen and C.P. Goodyear (1983). A stochastic age-structured population model of striped bass (*Morone saxatilis*) in the Potomac River. *Can. J. Fish. Aquat. Sci.*, 40 : 2170-2183.
- Davaine, P. et E. Beall (1982a). Introduction de Salmonidés dans les Terres Australes et Antarctiques Françaises. *CNFRA*, 51 : 289-300.
- Davaine P. et E. Beall (1982b). Acclimatation de la truite commune, *Salmo trutta* L., en milieu subantarctique (Îles Kerguelen) : Stratégie adaptative. *CNFRA*, 51 : 399-411.
- Davaine P. et E. Beall (1997). Introduction de Salmonidés en milieu vierge (Iles Kerguelen, Subantarctique) : enjeux, résultats, perspectives. *Bull. Fr. Pêche Pisc.*, 344/345 : 93-110.
- De Angelis, D.L., L.J. Svoboda, S.W. Christensen and D.S. Vaughan (1980). Stability and return times of Leslie matrices with density-dependant survival : applications to fish populations. *Ecological Modelling*, 8 : 149-163.
- Elliott, J.M. (1994). *Quantitative Ecology of the Brown Trout*. Oxford University Press, Oxford, UK, 286 p.
- Guyomard, R., C. Grévisse, F.X. Oury et P. Davaine (1984). Evolution de la variabilité génétique inter et intra-populations de populations de Salmonidés issues de mêmes pools génétiques. *Can. J. Fish. Aquat. Sci.*, 41,7 : 1024-1029.
- Horst, T.J. (1977). Use of the Leslie matrix for assessing environmental impact with an example for a fish population. *Trans. Am. Fish. Soc.*, 106 : 253-257.

- Jarry M., M. Khaladi and J.P. Gouteux (1996a). A matrix model for studying tsetse fly populations. *Ent. Exp. Applic.*, 78 : 51-60.
- Jarry M., J.P. Gouteux and M. Khaladi (1996b). Are tsetse fly populations close to equilibrium ? *Acta Biotheoretica*, 44 : 317-333.
- Jarry M., P. Davaine and E. Beall (1998). A matrix model to study the colonization by brown trout of a virgin ecosystem in the Kerguelen Islands. *Acta Biotheoretica*, sous presse.
- Lebreton, J.D. (1996). Demographic models for subdivided populations : the renewal equation approach. *Theor. Pop. Biol.*, 49(3) : 291-313.
- Leslie, P.H. (1945). On the use of matrices in populations mathematics. *Biometrika*, 33 : 183-212.
- Leslie, P.H. (1948). Some further notes on the use of matrices in population mathematics. *Biometrika*, 35 : 213-245
- Levin, S.A. and C.P. Goodyear (1980). Analysis of an age-structured fishery model. *Journal of Mathematical Biology*, 9 : 245-274.
- Sabaton C., L. Siegler, V. Gouraud, J.L. Baglinière and S. Manné (1996). Presentation and first application of a dynamic population model for brown trout (*Salmo trutta*) – Aid to river management. *Proc. Ecohydraulique 2000, Québec* : 531-542.
- Thomas T., P. Davaine et E. Beall (1981). Dynamique de la migration et reproduction de la truite de mer *Salmo trutta* L., dans la rivière Norvégienne, îles Kerguelen, TAAF. CNFRA, 47 : 5-42.

UTILISATION DES MÉTHODES DE SEGMENTATION POUR LA CONSTRUCTION DE CARTES ILLUSTRANT LES RELATIONS RESSOURCES HALIEUTIQUES / ENVIRONNEMENT

S. SALAUN ⁽¹⁾⁽²⁾ et J. FERRARIS ⁽¹⁾⁽³⁾

(1) Antenne ORSTOM, IFREMER, BP 21105, 44311 NANTES Cedex 03

(2) Adresse actuelle : IFREMER, Laboratoire ECOHAL, BP 21105, 44311 NANTES Cedex 03

(3) Adresse actuelle : ORSTOM, B.P A5, 98848 NOUMEA, Nouvelle-Calédonie

MOTS CLÉS : Segmentation, analyse de données, cartographie, ressources halieutiques, environnement

INTRODUCTION:

Les données géoréférencées collectées lors de campagnes océanographiques en vue d'étudier les relations entre les ressources biologiques et l'environnement peuvent être analysées en tenant compte de la dimension spatiale grâce notamment à l'utilisation d'un Système d'Information Géographique. Un tel logiciel permet de plus de restituer les résultats de ces analyses sous forme de cartes. Mais la construction de cartes implique de faire de nombreux choix quant aux variables à représenter et à leur mode de représentation sous forme de gradient (exprimant un continuum) ou sous forme discontinue par la construction de classes. Ces choix sont souvent plus ou moins arbitraires mais peuvent aussi être orientés par une analyse statistique préalable. Les données biologiques et environnementales étant souvent multidimensionnelles, toutes les variables ne sont pas visualisables telles quelles sur la même carte. Une solution pour réduire la dimension du problème consiste à construire des indicateurs synthétiques par exemple grâce à la réalisation d'analyses factorielles et/ou de méthodes de classification qui fournissent une représentation condensée des données (FERRARIS et PELLETIER, 1997).

Les méthodes de segmentation quant à elles fournissent des résultats beaucoup plus explicites pour l'interprétation de la structure et l'identification des facteurs discriminants ; et ceci particulièrement lorsque l'on souhaite expliquer une variable caractérisant la ressource par des variables environnementales.

L'objectif de cette étude est de montrer l'apport de telles méthodes pour la construction de cartes thématiques mettant en relation un type de données biotiques (biomasse, présence/absence d'une espèce, etc.) et une combinaison de facteurs environnementaux (caractérisant la sédimentologie, l'hydrologie, la bathymétrie).

Après une présentation **générale** des méthodes de segmentation et de cartographie **envisagées dans une telle problématique**, nous nous intéresserons à une application concernant

l'analyse des relations entre les poissons récoltés par campagnes de chalutages et leur environnement sur le plateau continental guinéen.

MÉTHODES :

- La segmentation :

→ Principe général :

Le principe de base des méthodes de segmentation est de partir de la population totale et de procéder à des dichotomies successives de celle-ci, de façon à obtenir au fur et à mesure du processus, des populations qui soient le plus homogènes possible vis à vis des classes d'une partition a priori (BREIMAN et al, 1984 ; PERINEL, 1996). Ces méthodes permettent ainsi de hiérarchiser les variables les plus discriminantes mais aussi de mettre en évidence des effets de seuil.

Ainsi, lorsque l'on est en présence d'un tableau de données contenant une variable privilégiée y à expliquer par les autres variables du tableau x_1, x_2, \dots, x_p , la méthode de segmentation consiste à rechercher d'abord la variable x_j qui explique le mieux la variable y . Cette variable définit la première division de l'échantillon en deux sous ensembles appelés segment. Puis on réitère cette procédure à l'intérieur de chacun des deux segments et ainsi de suite (LEBART et al, 1995). A partir de la population totale on construit ainsi un arbre de décision binaire (*figure 1*). Chaque chemin de l'arbre (de la racine à la feuille) est donc une combinaison de propriétés décrivant une sous population. Dans un arbre binaire complet, chaque segment contient un seul individu.

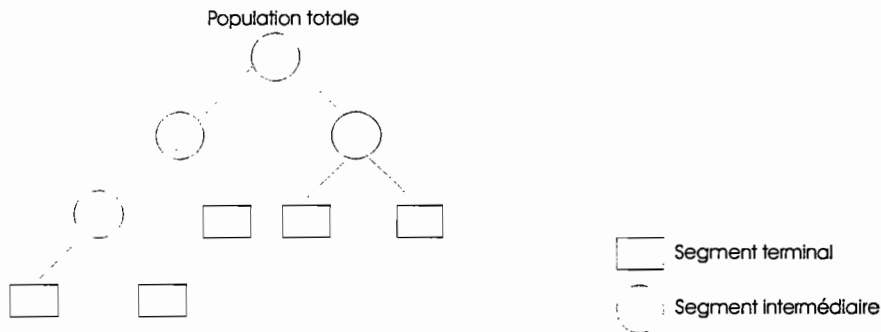


figure 1 : Arbre de décision binaire

Outre cet objectif descriptif, les techniques de segmentation fournissent une règle de décision utile si l'on cherche à prévoir la place d'un nouvel individu i . Ainsi, suivant les valeurs que i présente pour x_1, x_2, \dots, x_p , il va parcourir tel ou tel chemin de l'arbre et tombera dans un des segments terminaux. Ce nouvel individu sera alors affecté à l'un des groupes constitués (dans le cas d'une variable nominale à expliquer) ou sera affecté d'une valeur y_i qui sera la moyenne dans le segment terminal (l'écart-type correspondra aussi à celui du segment).

La segmentation par arbre de décision binaire présente de nombreux avantages tels que la lisibilité des règles d'affectation et la facilité d'interprétation des résultats. De plus, elle

permet d'utiliser en même temps comme variables explicatives des variables continues et nominales et d'analyser soit une variable nominale, soit une variable continue.

Il faut néanmoins prendre garde à cette simplification des résultats et tenir compte, lors de l'interprétation des coupures, de la corrélation qui peut exister entre les variables discriminantes. La variable discriminante qui a entraîné la coupure de l'arbre peut être fortement corrélée avec une autre variable en fonction de laquelle il paraît plus pertinent d'interpréter une coupure. Afin de palier à ces inconvénients, la procédure offre la possibilité de décrire les divisions « equi-réductrices » qui correspondent aux meilleures divisions successives du segment après celle qui a provoqué la partition. De plus, l'existence de divisions « equi-divisantes » permet de classer un nouveau sujet présentant des données manquantes concernant les variables explicatives.

→ Critères de division des nœuds :

Lorsque la variable à expliquer y est continue, le critère de sélection de la meilleure division d'un nœud est tel que la variance résiduelle du nœud, qui correspond à la moyenne pondérée des variances de y à l'intérieur de chacun de ses segments descendants, soit minimale. Ainsi, pour toute division d_j^m d'un nœud t par une variable x_j , on calcule la moyenne pondérée des variances de y à l'intérieur de chacun de ses segments descendants t_g et t_d , c'est à dire la variance résiduelle du nœud t :

$$\text{var}(d_j^m, t) = \left(\frac{n_g}{n_t} s_g^2\right) + \left(\frac{n_d}{n_t} s_d^2\right)$$

où n_g , n_d , n_t sont respectivement les effectifs des segments t_g , t_d , t et s_g^2, s_d^2 , sont les variances de la variable continue y à l'intérieur des segments t_g et t_d . On retient la meilleure division d_j^* réalisée par la variable x_j qui correspond à la variance résiduelle minimale :

$$\text{var}(d_j^*, t) = \min[\text{var}(d_j^m, t)]$$

où m appartient à d_j et où d_j est l'ensemble des divisions de la variable x_j .

Parmi toutes les meilleures divisions d_j^* obtenues à partir des p variables explicatives, la meilleure division (globale) du nœud t est effectuée à l'aide de la variable qui assure :

$$\text{var}(d^*, t) = \min_{j=1, \dots, p} [\text{var}(d_j^*, t)]$$

Lorsque la variable y est nominale et répartit les individus en k classes, il faut que le mélange des classes soit moins important dans les segments descendants que dans le nœud parent (critère de la pureté maximale). En effet, à chaque segment est associée une mesure de l'impureté $i(t)$ définie par :

$$i(t) = \sum_r \sum_s^k P(r|t)P(s|t)$$

avec $r \neq s$ et où $P(r|t)$ et $P(s|t)$ sont les proportions d'individus dans les classes c_r et c_s , dans le segment t . Un segment est pur s'il ne contient que des individus d'une seule classe,

dans un tel cas : $i(t) = 0$. Plus le mélange des classes dans le segment t est important, plus l'impureté $i(t)$ est élevée.

Chaque division d_j^m du nœud t par la variable x_j entraîne une réduction de l'impureté qui s'exprime par :

$$\Delta_j^m = i(t) - p_g i(t_g) - p_d i(t_d)$$

où p_g et p_d sont les proportions d'individus du nœud t dans les segments descendants t_g et t_d . Ainsi, pour chaque variable x_j , la meilleure division d_j^* est telle que la réduction de l'impureté Δ_j^* est maximale :

$$\Delta_j^* = \max[\Delta_j^m]$$

où m appartient à d_j , l'ensemble des divisions de la variable x_j .

Sur l'ensemble des p variables, la division du nœud t est effectuée à l'aide de la variable qui assure :

$$\Delta^* = \max_{j=1, \dots, p} [\Delta_j^*]$$

(LEBART et al, 1995).

→ Construction du meilleur sous-arbre:

À partir de l'arbre binaire complet, une phase d'élagage permet de produire un sous arbre optimal exploitable en se fondant sur l'estimation de l'erreur d'affectation ou de prévision à l'aide, soit d'un échantillon test, soit de la validation croisée.

La méthode des échantillons tests, la plus couramment utilisée, consiste à effectuer la discrimination sur une partie seulement de la population initiale (80% souvent) et de tester les règles d'affectation sur les 20% restant.

Dans le cas de petits échantillons, il est conseillé d'utiliser la validation croisée car elle permet de prendre en compte tous les sujets de l'échantillon à la fois pour construire et pour tester l'arbre (GUEGUEN et NAKACHE, 1988). En effet, dans ce cas, l'échantillon est divisé en m parties égales ; la discrimination se fait sur l'échantillon de $m-1$ parties et le taux d'erreur sur la partie restante, ce qui peut être fait de m façons différentes (GUEGUEN et al, 1996).

Les analyses ont été ici réalisées grâce aux procédures REGAR (dans le cas de y continue) et DISAR (dans le cas de y nominale) du logiciel SPAD.S (GUEGUEN et al, 1996).

- Réalisation des cartes :

Les cartes ont été réalisées par le biais d'un Système d'Information Géographique (logiciel SAVANE, © ORSTOM, version 5.02). Un système d'information géographique (SIG) est un système informatisé d'acquisition, de gestion, d'analyse et de représentation de données à référence spatiale. Il permet de croiser des données de natures diverses telles que les données d'enquêtes, des cartes thématiques, des données topographiques, des images satellites, des modèles numériques de terrain. Un SIG stocke les deux composantes de l'information décrites par une carte : la description des objets spatiaux (coordonnées géographiques en longitude et en latitude) et leurs caractéristiques thématiques (valeur de chaque objet pour telle caractéristique).

Les méthodes de segmentation permettent de sélectionner les variables environnementales les plus discriminantes vis à vis d'une variable biotique. Ce sont ces deux types de variables qu'il s'agit de représenter sur la même carte.

Afin de visualiser les variables environnementales dans l'espace, on utilise ici une interpolation déterministe permettant de connaître en tout point de l'espace considéré les valeurs prises par la variable à cartographier. La méthode d'interpolation utilisée estime la valeur en un point à partir des valeurs mesurées dans un certain nombre de stations voisines. On choisit le nombre de ces stations à retenir en fonction de la densité des stations sur la zone d'étude. Les résultats de ces interpolations sont des modèles numériques de terrain que l'on peut « découper » en un certain nombre de classes afin de les visualiser. Les valeurs de coupure entre ces classes sont celles qui ont provoqué les coupures binaires des arbres de segmentation correspondants.

On peut par la suite superposer plusieurs variables environnementales traitées de la sorte, ceci dans la limite du « visuellement acceptable ». Puis une dernière couche d'information, correspondant à la variable biologique expliquée par ces variables environnementales discriminantes, est ajoutée à la carte.

Il faut aussi noter que les résultats des interpolations des variables environnementales sont aussi utilisés dès la première phase de construction des tableaux de données qui vont être utilisés pour la segmentation. En effet, les données, lorsqu'elles ont été collectées à des moments différents peuvent différer tant au niveau du nombre de stations que de la situation géographique de ces stations. Une solution pour mettre en relation ces données d'origines différentes consiste alors à interpoler la variable environnementale par exemple ; et ceci surtout si les points de sondage auxquels elle correspond sont plus rapprochés que les stations de chalutage correspondant à la variable biologique car on minimise ainsi l'erreur due à l'interpolation. Puis on récupère, au niveau des stations de chalutages biologiques, l'information issue de l'interpolation. Ainsi, les deux types de données deviennent homogènes puisqu'elles concernent le même nombre de stations et la même position géographique.

APPLICATION:

A titre d'illustration, on présentera ici une application de ces méthodes de segmentation à l'analyse des relations entre les ressources démersales du plateau continental guinéen (voir *fig.2*) et leur environnement (SALAUN, 1997).



figure 2 : Position géographique de la Guinée sur le continent africain

Cette démarche a pour but de décrire la répartition des ressources démersales par l'intermédiaire de variables biologiques synthétiques en liaison avec les variables environnementales, et plus particulièrement la sédimentologie. Ce travail, réalisé à l'antenne ORSTOM de Nantes dans le cadre d'un stage de DEA, et fait suite au projet FAO « Systèmes d'information géographique appliqués aux pêcheries de l'Afrique de l'Ouest » (1993-96) qui couvrait le Sénégal, la Mauritanie, le Maroc et la Guinée.

La plupart des données proviennent de campagnes scientifiques de chalutages démersaux sur 150 stations (FONTANA et MORIZE, 1995), seules les données de sédimentologie ont été récoltées lors d'une autre campagne en 800 points de sondage (DOMAIN et BAH, 1993).

Les relations ressources/ environnement peuvent être décrites à différentes échelles de complexité biologique et on traite ici les exemples de présence/absence d'une espèce (variable nominale à deux modalités) puis de taille d'une espèce (variable continue). Les facteurs explicatifs entrant dans l'analyse sont la profondeur, les salinités et températures de fond et de surface, ainsi que les teneurs du sédiment en lutites, sables fins, sables grossiers, graviers et carbonates. Les deux espèces considérées ici : *Pagellus bellottii* (Pageot à taches rouges) et *Sparus caeruleostictus* (Pagre à points bleus). Elles appartiennent à la famille des Sparidae, présente sur l'ensemble de la zone et constituant la plus forte biomasse lors des campagnes.

→ Présence du Pagre à points bleus sur la zone d'étude :

La procédure de discrimination par arbre de décision binaire appliquée à la variable présence-absence du Pagre, au cours de la campagne correspondant à la saison sèche, aboutit à la création de trois segments terminaux intéressants (fig. 3)

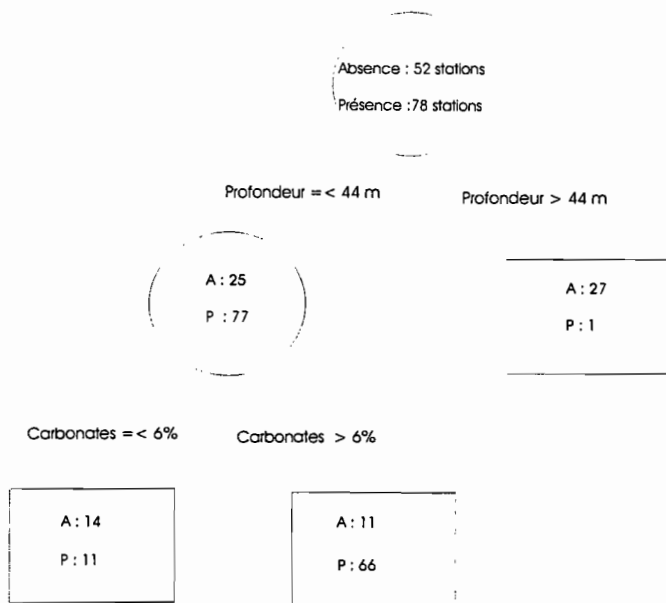


figure 3 : Résultats de la segmentation appliquée à « présence/absence du Pagre »

D'après ces résultats, lorsque le Pageot est présent, il s'agit à 85% de fonds de profondeur inférieure à 44 m et pour lesquels la teneur en carbonates est supérieure à 6%. Afin de représenter ceci sous forme de cartes, les variables environnementales apparaissant les plus discriminantes, c'est à dire la profondeur et la teneur en carbonates, ont été interpolées à l'ensemble de la zone puis scindées en deux classes selon les valeurs de coupure obtenues par la segmentation (44 m pour la profondeur et 6% pour la teneur en carbonates). Enfin, chaque segment terminal de ce sous-arbre correspond à une combinaison de variables environnementales que l'on peut alors représenter sous formes de plages de couleurs lors de la représentation cartographique (*fig. 4*).

Le Pageot apparaît peu présent dans les zones les plus profondes. De plus, au niveau de la sédimentologie, le facteur expliquant le mieux sa répartition est la teneur en carbonates et non un type de granulométrie. Ceci peut être lié au fait que les poissons de la famille des Sparidae ne vivent pas directement sur le fond mais s'en rapprochent pour se nourrir. Cette relation avec les sédiments carbonatés trouverait alors une explication dans le régime alimentaire.

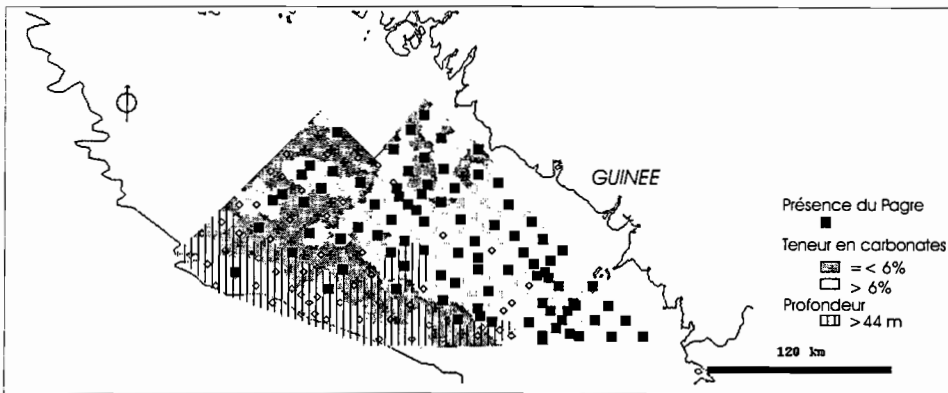


figure 4 : Présence du Pageot sur la zone d'étude en liaison avec l'environnement

→ Taille du Pageot à taches rouges sur la zone d'étude :

À une échelle biologique plus fine, on peut raisonner au niveau des tailles des individus. Le Pageot à taches rouges (*Pagellus bellottii*) est abondamment présent sur le plateau continental guinéen mais sa faible taille reste bien souvent un frein à sa commercialisation. On dispose des données de capture en poids et en nombre de poissons pour chaque espèce et grâce à une clé longueur/poids disponible (FONTANA et MORIZE, 1995), on peut facilement calculer la taille moyenne individuelle du Pageot par station. Cette taille est une variable quantitative que l'on peut expliquer par les variables environnementales grâce à la segmentation.

Il se dégage alors trois segments terminaux intéressants (*fig. 5*). Ces résultats montrent que les plus grandes tailles apparaissent sur des profondeurs inférieures à 58.5 m et où la teneur en sables gros des sédiments est supérieure à 25%.

Les variables explicatives (sables grossiers et profondeur) ont été là aussi interpolées à l'ensemble de la zone d'étude avant d'être partitionnées en fonction des valeurs de coupure données par les résultats de la segmentation.

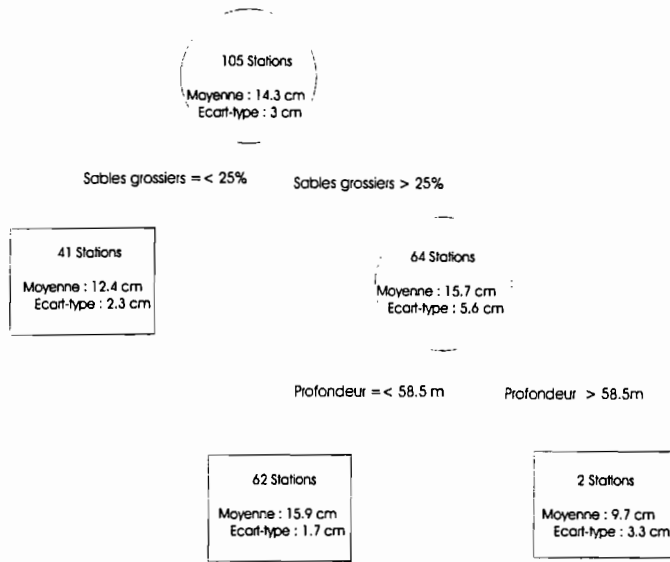


figure 5 : Résultats de la segmentation appliquée à la variable « taille du Pageot »

Par rapport à la question qui nous intéressait ici, c'est à dire identifier les zones où la taille du Pageot est maximale, on a choisit une représentation cartographique des tailles qui doit permettre une interprétation rapide (fig.6). On voit ainsi que les fortes tailles, correspondant comme nous l'avons vu au plus fortes teneurs en sables gros sont présentes au centre du plateau continental. De plus, cette carte, associée aux résultats de la segmentation, met en évidence l'effet de seuil dû à la profondeur : la taille du Pageot augmentant de la côte vers le large jusqu'à une certaine profondeur (autour de 58 m) à partir de laquelle la tendance s'inverse, ce qui est confirmé par la figure 7.

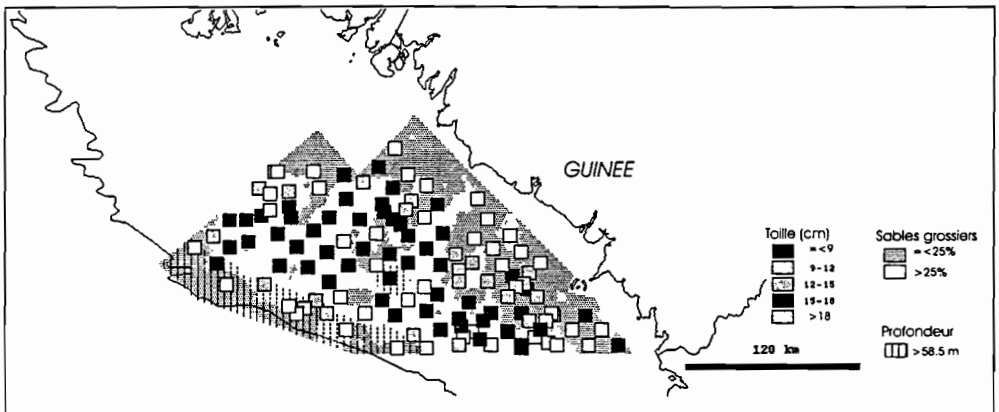


figure 6 : Tailles moyennes individuelles du Pageot en relation avec l'environnement

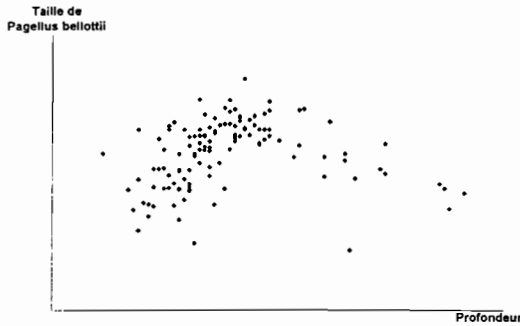


figure 7 : Tailles moyennes individuelles du Pageot en fonction de la profondeur

CONCLUSION:

L'objectif de cette étude était de présenter un exemple de complémentarité entre analyse de données et cartographie, et plus particulièrement de décrire l'apport que peuvent représenter les méthodes de segmentation lorsque l'on souhaite expliquer un facteur biotique par un ensemble de facteurs abiotiques. L'approche spatiale de ces relations ressources-environnement a été facilitée ici au niveau méthodologique par l'utilisation d'un Système d'Information Géographique. Ce logiciel a été utilisé tout au long de cette analyse. Une des fonctions les plus utiles dans cette étude est la possibilité d'effectuer des interpolations. Cette interpolation est utile tant pour la représentation des variables environnementales sur toute la zone d'étude que lors de l'étape préliminaire de construction de tableaux de données où à chaque variable biologique correspond un ensemble de variables discriminantes. Le SIG est bien sûr aussi indispensable pour la visualisation sur une carte des résultats de la segmentation (représentation des segments sous forme de plages de couleurs, superposition des représentations des différentes variables). Il y a donc un va-et-vient permanent entre le logiciel d'analyse de données et le logiciel SIG et on peut donc regretter que l'analyse de données proprement dite ne se fasse au sein même du logiciel SIG, ce qui faciliterait l'analyse. De plus, l'analyse pourrait être optimisée en prenant en compte la dépendance spatiale des données notamment par exemple par l'utilisation de méthodes de géostatistique ce que ne permettait pas ce logiciel SIG.

Quoiqu'il en soit, l'application développée ici montre l'importance de se baser sur une démarche statistique tant pour identifier les variables environnementales les plus intéressantes à représenter que pour construire une partition de ces variables environnementales basé sur une réalité biologique. Ce découpage de l'espace pourrait aussi conditionner le plan d'échantillonnage de campagnes à venir si l'on souhaite, par exemple, décrire un paramètre biologique précis. Ce travail montre aussi l'intérêt de construire des cartes thématiques différentes en fonction du phénomène biologique à interpréter.

De plus, comme le soulignent GUEGUEN et NAKACHE (1988), cette méthode basée sur la construction d'un arbre de décision binaire fournit une règle de décision simple, est robuste vis à vis de données aberrantes et tient compte naturellement des interactions qui peuvent exister entre les données.

Les méthodes de segmentation possèdent donc de nombreux avantages lorsque l'on souhaite construire des cartes en se basant sur des critères statistiques, dont la facilité

d'interprétation directe et de visualisation rapide des résultats mais aussi, comme nous l'avons vu, la possibilité de détermination de seuils.

Ces méthodes de segmentation peuvent aussi être utilisées par exemple à la suite d'une typologie afin de retrouver comment sont combinées entre elles les différentes variables caractéristiques de la classe et ainsi de mieux décrire sa variabilité intrinsèque (TONG et PERINEL, 1996).

Enfin, l'approche décisionnelle de la segmentation, qui permet d'affecter de nouveaux individus dans des segments déjà caractérisés, paraît intéressante en halieutique lors de l'élaboration de règles de gestion (TAQUET et al, 1997).

REFERENCES BIBLIOGRAPHIQUES :

BREIMAN L., FRIEDMAN J. H., OLSHEN R. A., STONE C. J., 1984 : « Classification and Regression Trees », Wadsworth International Group.

DOMAIN F. et BAH M.O., 1993 : « Carte sédimentologique du plateau continental guinéen à 1:200 000 et notice explicative » n° 108, ORSTOM-CNSHB, Paris, 15p.

FERRARIS J., PELLETIER D., 1997 : « Deuxième table ronde sur les SIG et Journée thématique sur le spatial », Lettre de l'Association Française d'Halieumétrie, n°5.

FONTANA A., MORIZE E., 1995 : « Projet protection et surveillance des pêches de la ZEE guinéenne- Volet scientifique- Rapport de fin d'étude », ORSTOM-CNSHB, 137p.

GUEGUEN A., NAKACHE J. P., NICOLEAU-MOLINA J., 1996 : « SPAD.S, version 3, Procédures de segmentation », CISIA.

GUEGUEN -A., NAKACHE J. P., 1988 : « Méthodes de discrimination basée sur la construction d'un arbre de décision binaire », Rev. Stat. Appl. 36, 19-38.

LEBART L., MORINEAU A., PIRON M., 1995 : « Statistique exploratoire multidimensionnelle », Ed. DUNOD, 440p.

PERINEL E., 1996 : « Segmentation et analyse de données symboliques - Application à des données probabilistes imprécises », Thèse, Université Paris Dauphine, 346p.

SALAUN S, 1997 : « Analyse des relations ressources-environnement au sein d'un système d'information géographique - Application aux ressources démersales du plateau continental guinéen », Rapport de DEA, ORSTOM-UBO, 30p.

TAQUET M., GAERTNER J.-C. , BERTRAND J., 1997 : « Typologie de la flottille chalutière du port de Sète par une méthode de segmentation », Aquat. Living Resour. , 10, 137-148

TONG H., PERINEL E., 1996 : « Une approche numérique/symbolique pour l'extraction et la formalisation de connaissances : Application à la description de tactiques de pêche artisanale

au Sénégal ». In : Méthodes d'étude des systèmes halieutiques et aquacoles. J. Ferraris, D. Pelletier, M.J. Rochet édts. Colloques et Séminaires, ORSTOM, 157-164.

COMPARAISON DE STRATÉGIES DE PÊCHE

PAR ANALYSES FACTORIELLES SUR TABLEAUX D'ÉCHANGES

Jocelyne FERRARIS¹ et Catherine PARDOUX²

1 IRD
BP A5
98848 - NOUMEA
Nouvelle-Calédonie
ferraris@noumea.ird.nc

2 LISE-CEREMADE
Université Paris IX - Dauphine
Place du M^l de Lattre de Tassigny
75775 - PARIS Cedex 16
Catherine.Pardoux@dauphine.fr

1. CONTEXTE

La compréhension de la dynamique des pêcheries composites, telles que les petites pêcheries côtières caractérisées par un grand nombre d'engins et d'espèces capturées, nécessite de développer de nouvelles approches pour mieux rendre compte de la variabilité spatio-temporelle des populations exploitées et de leur exploitation. Les systèmes d'enquête mis en place pour l'étude des pêcheries artisanales sont souvent basés sur l'échantillonnage des captures et des efforts afin d'estimer la production et la mortalité par pêche, mais tendent à évoluer en fonction des nouvelles questions qui se posent en halieutique (Morand & Ferraris, 1997). La prise en compte du comportement du pêcheur devient une nécessité pour comprendre ses réactions aux changements, analyser les capacités d'adaptabilité des pêcheries et mieux répondre aux enjeux de la gestion halieutique. La dynamique de la pêcherie maritime artisanale au Sénégal est ainsi décrite en modélisant les tactiques et stratégies des pêcheurs en fonction de deux typologies distinctes relatives d'une part aux unités de pêche et d'autre part aux actions de pêche (Laloë & Samba, 1991, Pech, 1998). Une enquête spécifique est menée en 1992, auprès de 200 unités de pêche suivies pendant un an, afin de préciser les notions de tactiques et stratégies de cette pêcherie et d'étudier les comportements de mobilité et de polyvalence des pêcheurs. La **tactique** décrit l'activité quotidienne du pêcheur par la combinaison des choix effectués au cours de ses opérations de pêche (Gettler-Summa et al., 1994). La **stratégie** est appréhendée par la succession des tactiques au cours d'une période donnée. La dynamique temporelle de chaque stratégie peut être décrite par un processus stochastique (chaîne de Markov d'ordre 1 homogène dans le temps) appliqué à la matrice de transition associée à des séries chronologiques qualitatives. Le graphe valué associé représente les passages d'un état à un autre et permet d'illustrer les transferts d'efforts de pêche et les comportements-types sur une période homogène (Ferraris et al., 1996). De par la variabilité saisonnière et la structure hétérogène de la flottille, la pêcherie doit être partitionnée en segments homogènes du point de vue temporel (saisons) et structurel (groupes d'unités de pêche). La description de la dynamique de la pêcherie consiste alors à analyser les **comportements temporels observés sur chacun** des segments et à comparer les différentes stratégies pour **mettre en évidence leur dynamique propre**.

2. OBJECTIFS

Ce travail vise plusieurs objectifs :

- la mise en évidence des **différentes stratégies d'exploitation observées** au cours du temps et au sein de la pêcherie artisanale sénégalaise,
- la description des stratégies sous une forme synthétique traduisant la dynamique temporelle du phénomène,
- la comparaison des différentes stratégies mises en évidence.

A partir du suivi des activités individuelles de pêche, une approche typologique par **classification automatique** vise à décomposer la dynamique de la pêcherie en périodes et groupes de pêcheurs homogènes de comportements similaires. La comparaison des tableaux de fréquences décrivant les changements observés d'une tactique à une autre pour chacun des groupes mis en évidence, est effectuée par des **analyses factorielles** afin de définir un référentiel commun aux différentes matrices et de mettre en évidence les effets "période" et "groupe de pêcheurs", et la dynamique propre à chacune des stratégies.

3. DONNEES

On s'intéresse ici au cas du port de Kayar (Côte nord du Sénégal) où 55 pêcheurs ont participé sur une base de volontariat au suivi d'units de pêche comportant une enquête sur leur activité quotidienne (tactique) au cours d'un an. La tactique est codée en 9 états (elle est réduite ici à sa plus simple expression, soit le type d'opération décrit par l'engin de pêche utilisé ou la cause de non pêche) :

L	==>	<i>pêche à la ligne</i>
FD	==>	<i>pêche au filet</i>
Lpo	==>	<i>pêche à la turlutte (ligne à poulpe)</i>
LL	==>	<i>pêche utilisant simultanément plusieurs types de lignes (L + LPo)</i>
LF	==>	<i>pêche utilisant simultanément ligne et filet (L + FD)</i>
Rep	==>	<i>repos ou arrêt court (météo, non satisfaction du marché, ...)</i>
Arr	==>	<i>arrêt long (panne, entretien, maladie, voyage, ...)</i>
Cam	==>	<i>campagne de pêche à l'extérieur du port</i>
FF	==>	<i>pêche utilisant simultanément plusieurs types de filets (FD+FD)</i>

Les unités sont suivies pendant 366 jours (du 15/12/91 au 15/12/92) : chaque unité de pêche est caractérisée chaque jour par l'une des neuf tactiques.

Les données peuvent être présentées sous trois formes :

1 - Première matrice : chaque colonne associée à un pêcheur est une série chronologique (de longueur 366) d'une variable qualitative E_j à 9 modalités (correspondant aux 9 tactiques).

<i>unité de pêche</i>	ω_1	...	ω_j	...	ω_N
<i>Jour 1</i>					
<i>Jour t</i>		...	E_j		
<i>Jour T</i>					

2 - Deuxième matrice : tableau de fréquences où chaque individu-pêcheur (ligne) est décrit par 81 variables quantitatives construites à partir des 81 transitions possibles entre les 9 états (à chaque état à la date t , correspond un état à $t + 1$).

3 - Troisième matrice : tableau construit pour un groupe de pêcheurs et une période, décrivant les échanges ou les transitions entre les 9 états :

Soit T = durée totale de la période considérée

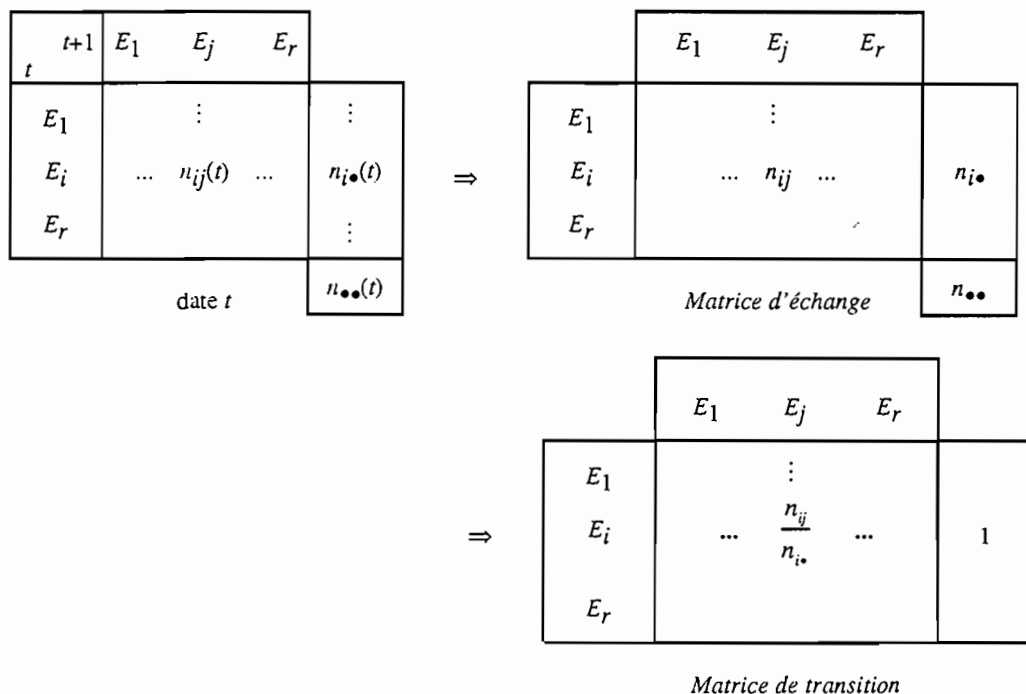
r = nombre d'états

$n_{ij}(t)$ = nombre d'individus en E_i à la date t et E_j à la date $(t + 1)$

$$n_{i\bullet}(t) = \sum_{j=1}^r n_{ij}(t)$$

$$n_{ij} = \frac{1}{T-1} \sum_{t=1}^{T-1} n_{ij}(t)$$

$$n_{i\bullet} = \sum_{j=1}^r n_{ij}$$



4. METHODES

Les méthodes choisies pour atteindre les objectifs fixés dans ce travail font partie de la "Boîte à Outils" offerte par le logiciel SPAD¹, complété du logiciel ADDAD². L'algorithme de classification automatique est basé sur le critère de Ward avec coupure de l'arbre hiérarchique et consolidation de la partition par une technique de type "nuées dynamiques". Ces logiciels offrent également plusieurs méthodes d'analyse factorielle et notamment, des méthodes adaptées au traitement de tableaux ternaires.

L'approche adoptée pour définir les périodes homogènes et les groupes de pêcheurs à comportement similaire est présentée dans un travail précédent sur les stratégies des unités de pêche de Kayar (Ferraris et al., 1996). Elle comporte deux étapes :

- 1- une classification automatique sous contrainte temporelle de la première matrice préalablement projetée sur les axes factoriels d'une analyse des correspondances multiples (analyse adaptée au traitement des variables qualitatives),
- 2- une analyse factorielle des correspondances suivie d'une classification automatique de la deuxième matrice pour partitionner l'échantillon de pêcheurs à partir des fréquences de transition observées au cours de l'année.

Nous retenons les groupes jugés les plus représentatifs. Ayant divisé l'année en périodes, nous avons ainsi à analyser et à comparer plusieurs tableaux de fréquence particuliers que sont les matrices

¹ SPADTM Version 3 pour Windows, CISIA 1, avenue Herbillon 94160 Saint-Mandé (France).

² ADDAD, 22 rue Charcot 75013 Paris (France)

d'échanges et les matrices de transition. Plusieurs approches sont envisagées pour comparer ces différents tableaux, basées sur le principe de l'Analyse Factorielle des Correspondances. On cherche à utiliser d'une part, les propriétés barycentriques de l'AFC pour analyser simultanément les relations entre les états de départ (temps t) et les états d'arrivée (temps $t + 1$) en termes de distance du χ^2 , et d'autre part, ses extensions à l'Analyse des Correspondances Multiples et à l'Analyse Factorielle Multiple pour l'analyse des interactions entre les différentes dimensions des tableaux, notamment les périodes et les stratégies :

- 1- Analyses des Correspondances de suites de tableaux d'échanges : analyse du tableau "Somme", avec projections des lignes et des colonnes des 12 tableaux en éléments supplémentaires [Cazes (1990), Escofier (1983, 1984)] et analyse du tableau global [Escofier & Drouet (1983)] avec projection des modalités supplémentaires des variables qualitatives "Période", "Groupe de pêcheurs" et leurs interactions.
- 2- Analyses des Correspondances des tableaux de Burt associés à chaque stratégie de pêche [Escofier & Pagès (1993)].
- 3- Analyses des tableaux ternaires [Dazy & Le Barzic (1996), Escofier & Pagès (1993)] : Analyse Factorielle Multiple de suites de matrices de transition basée sur l'Analyse des Correspondances Multiples avec des variables codées floues.

5. RESULTATS

1- Recherche de périodes homogènes

Une classification hiérarchique avec contraintes de contiguïté est réalisée sur le tableau dont les colonnes sont les facteurs résultant d'une analyse des correspondances multiples de la première matrice (définie au § 3) de dimension 366×48 (7 unités de pêche sont mises en supplémentaire en raison d'un grand nombre de valeurs manquantes). On retient une partition en quatre périodes contiguës de durée inégale :

- période 1 : 137 jours (15/12/91 - 30/04/92)
- période 2 : 92 jours (01/05/92 - 31/07/92)
- période 3 : 104 jours (01/08/92 - 12/11/92)
- période 4 : 33 jours (13/11/92 - 15/12/92)

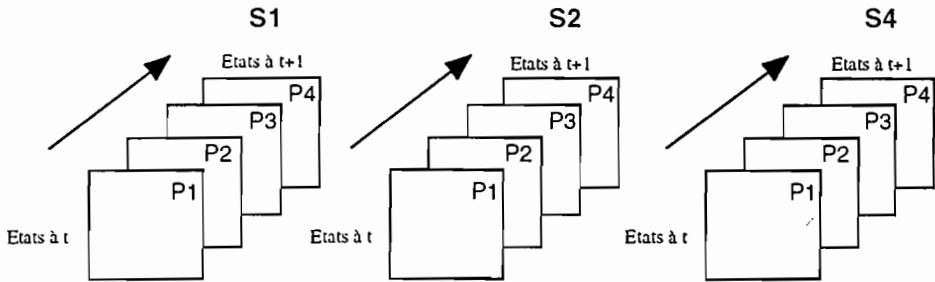
2- Recherche de groupes homogènes

Le tableau de fréquences indiquant le nombre de fois où les unités de pêche, ayant pratiqué une activité de pêche effective au cours des quatre périodes, transitent par les 81 combinaisons possibles entre les 9 états (deuxième matrice définie au § 3 de dimension 33×81), est soumis à une Analyse des Correspondances suivie d'une Classification ; l'arbre hiérarchique met clairement en évidence 6 classes d'unités de pêche, dont trois correspondent à des comportements atypiques :

DECOMPOSITION DE L'INERTIE : QUOTIENT (INERTIE INTER/INERTIE TOTALE) = 0.7941

	INERTIES	EFFECTIF	POIDS	DISTANCE
INERTIE INTER-CLASSES INERTIES INTRA-CLASSE	1.5059			
CLASSE 1 / 6	0.1830	15	5322.00	0.7289
CLASSE 2 / 6	0.0641	3	1086.00	1.5857
CLASSE 3 / 6	0.0000	1	365.00	6.8590
CLASSE 4 / 6	0.1197	10	3646.00	0.8994
CLASSE 5 / 6	0.0237	3	1095.00	2.5035
CLASSE 6 / 6	0.0000	1	365.00	10.3096
INERTIE TOTALE	1.8964			

La stratégie 5, caractérisée par un arrêt prolongé, et les stratégies 3 et 6, décrites par un seul pêcheur, ne sont pas considérées dans la suite de ce travail, car elles sont peu représentatives de l'ensemble de la pêcherie ; ces stratégies peuvent cependant être pertinentes pour l'étude de la dynamique de la pêche (phénomène d'innovation, diversification des activités, etc.). La caractérisation des stratégies 1, 2 et 4 sur chaque période nous amène à construire 12 matrices d'échanges et 12 matrices de transition. La comparaison des stratégies consiste alors à analyser trois suites de quatre tableaux :



Les 4 périodes étant de durée inégale et les 3 classes de pêcheurs ne comportant pas le même nombre d'individus, une étape préliminaire à l'analyse comparative a consisté à "normaliser" les 12 tableaux en les rapportant à un même nombre jours-individus (effectifs totaux des tableaux égalisés).

3- Analyses des Correspondances de suites de tableaux d'échanges

L'Analyse des Correspondances est effectuée sur les tableaux d'échanges (définis au § 3) selon deux approches :

1- Tableau « Somme » formé de 8 lignes et 8 colonnes actives (un des états étant absent pour les trois stratégies conservées) auxquelles ont été juxtaposées en éléments supplémentaires les fréquences de chaque stratégie-période.

2- Tableaux juxtaposés verticalement : 96 lignes \times 8 colonnes actives où chaque ligne est caractérisée par trois variables qualitatives supplémentaires indiquant la stratégie, la période et la stratégie-période.

Les résultats similaires obtenus par ces 2 approches indiquent des tendances communes très fortes :

1. $\lambda_1 \approx 1 \Rightarrow$ un état quasi-absorbant définit le 1^{er} axe (tactique "Campagne").
2. Une décroissance régulière des valeurs propres : 98% de l'inertie pour les 6 premiers axes : un axe est généralement défini par un état.
3. La représentation simultanée montre des positions quasi-confondues des états à t et à $t + 1$,
4. Mise en évidence des liens entre stratégies, périodes et tactiques par les variables supplémentaires.

Le 1^{er} plan factoriel (figure 1) obtenu par la 2^{de} approche (40% de l'inertie) avec projection des variables supplémentaires, résume les principales sources de variabilité du jeu de données : l'état "campagne" (Cam), plus caractéristique de la troisième période et des stratégies S1 et S2, définit le premier axe de par le peu de relations avec les autres états ; le deuxième axe est expliqué par l'activité "filet mixte" (FF), peu fréquente et associée à l'activité "filet" (FD) plus caractéristique de la stratégie S2.

Cette approche met davantage en évidence la variabilité entre les états, au détriment de la variabilité entre les différents tableaux (stratégies ou périodes). La tendance commune très forte entre les différents tableaux s'explique par une diagonale chargée, quelque soit la stratégie-période, due à des passages fréquents d'un état sur lui-même de t à $t + 1$ indiquant que les pêcheurs ont tendance à reproduire la même activité d'un jour à l'autre.

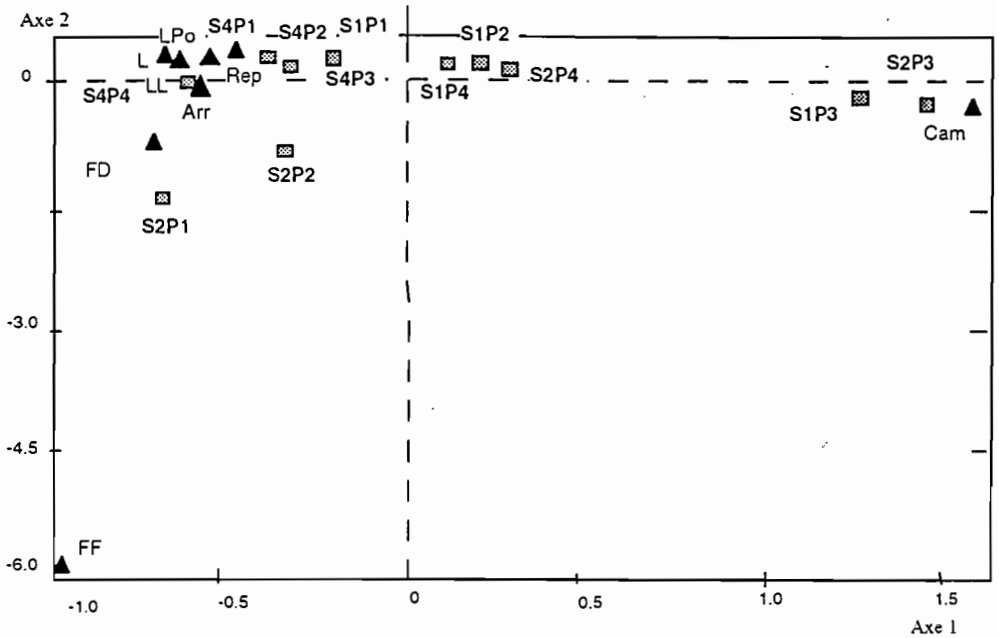


Figure 1

4- Analyses des Correspondances des trois tableaux de Burt

A chaque stratégie est associé un tableau de Burt mettant en relation trois ensembles : états au temps t , états au temps $t + 1$ et périodes. On a un tableau de Burt, de dimension 20×20 , propre à chacune des stratégies, soit : une matrice symétrique, indiquant sur la diagonale le nombre d'occurrences des observations pour chaque état de départ, chaque état d'arrivée ou chaque période, au-dessous et au-dessus de la diagonale, le nombre d'observations correspondant à chaque échange pour toute période confondue, le nombre d'observations pour chaque état de départ par période et pour chaque état d'arrivée par période (ex. : tableau de Burt pour la stratégie S1 en annexe).

Les Analyses des Correspondances appliquées aux trois tableaux de Burt donnent les résultats suivants :

- des paliers dans la décroissance des valeurs propres,
- les trois premiers facteurs résument l'essentiel du comportement de S1 et de S2, alors que les deux premiers facteurs résument l'essentiel du comportement de S4,
- résumé pertinent des stratégies avec une bonne lecture des interactions entre les tactiques et les périodes.

On donne en exemple le premier plan factoriel (figure 2) de la stratégie S1 (41% de l'inertie). De nouveau, on constate une superposition des états au temps t et au temps $t + 1$, reflétant l'inertie du système et la capacité des pêcheurs à reproduire leur activité. Les deux premiers axes isolent les deux états montrant le moins de relations avec les autres activités : d'une part, l'état "Campagne" plus caractéristique de la période 3 et d'autre part, l'état "Arrêt". La première période est caractérisée par une activité de pêche plus marquée, basée sur la pêche à la ligne, présentant des échanges importants avec le "Repos" et avec "Lignes mixtes" (LL), associant la ligne-poisson (L) à la ligne-poulpe (Lp). L'axe III est défini par les lignes mixtes, plus caractéristiques de la période 2. Au cours du temps, de la période 1 à 4, on note une translation des activités de pêche à la ligne vers les départs en campagne de pêche, avec en période 4 retour au port avec des arrêts marqués.

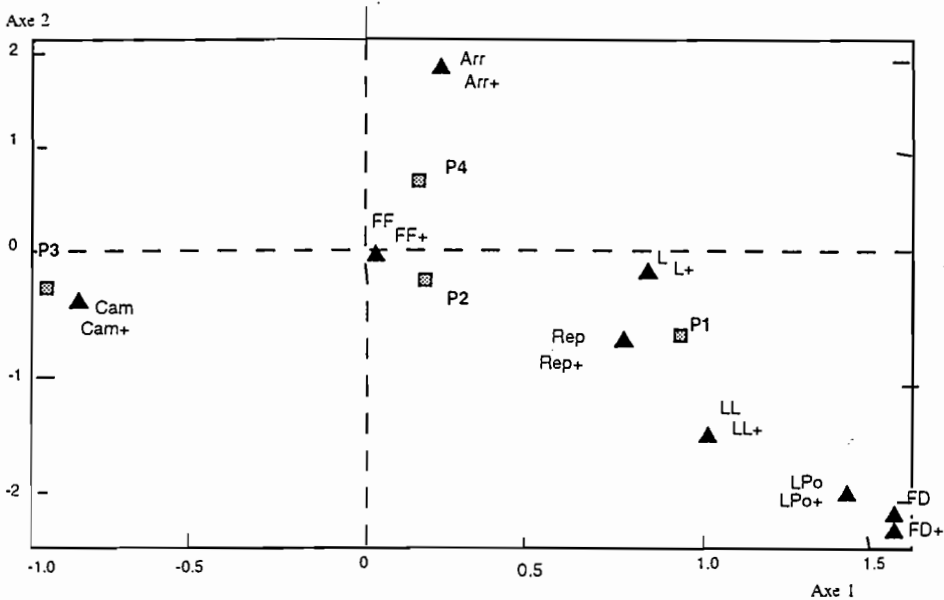


Figure 2

5- Analyses Factorielles Multiples sur les matrices de transition

La comparaison des stratégies est effectuée par Analyses Factorielles Multiples basée sur l'Analyse des Correspondances Multiples appliquée aux matrices de transition. Chaque stratégie-période est décrite par sa matrice de transition (soit les profils en ligne des probabilités de passer d'un état au temps t à un état au temps $t + 1$) : on peut considérer que chaque ligne représente les modalités de la variable "Etat au temps $t + 1$ " avec un codage flou (somme des probabilités = 1). Afin de comparer les groupes (stratégies et / ou périodes), il est proposé d'appliquer l'AFM avec codage flou : les individus du tableau sont les états de départ (temps t), les colonnes du tableau sont soit, 4 variables qualitatives définies par les profils de transitions d'une même stratégie aux 4 périodes, soit, 3 variables qualitatives définies par les profils de transition des 3 stratégies pour une période donnée. La troisième dimension du tableau ternaire est respectivement la stratégie (3 stratégies) ou le temps (4 périodes). L'ACM opérée sur chaque matrice de transition permet de mettre en correspondance les modalités (variables indicatrices) des variables qualitatives décrivant les transitions observées pour chacun des états de départ, puis l'AFM permet de rechercher les dimensions communes aux tableaux comparés. On effectue deux Analyses Factorielles Multiples sur :

1. Trois groupes « Stratégies » comprenant chacun quatre variables « Périodes » codées flou :

	S1				S2				S4			
Etats à t	t + 1											
	P1	P2	P3	P4	P1	P2	P3	P4	P1	P2	P3	P4

2. Quatre groupes « Périodes » comprenant chacun trois variables « Stratégies » codées flou :

	P1			P2			P3			P4		
Etats à t	t + 1											
	S1	S2	S4	S1	S2	S4	S1	S2	S4	S1	S2	S4

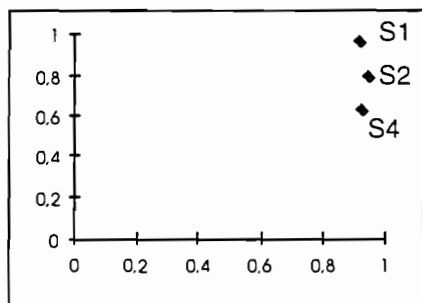
La première valeur propre de chaque groupe étant voisine de 1, les axes globaux des deux analyses sont quasiment les mêmes. Les résultats de la première analyse (comparaison des trois stratégies de pêche) sont donnés en annexe avec le premier plan factoriel de l'analyse globale (position des individus moyens et partiels). Le premier axe correspond à 24% de l'inertie totale du nuage, les 4 premiers axes correspondent à plus de 78% de l'inertie totale avec un saut marqué après le 4^{ème} axe. La signification des axes de l'analyse globale est donnée par l'interprétation des individus moyens ayant les plus fortes contributions et les plus faibles inerties intra-groupes (représentants de la structure commune). Le premier axe, commun à tous les groupes, est défini par l'activité "Arrêt prolongé", activité la moins spécifique à une stratégie d'exploitation, pouvant se produire à n'importe quelle période pour un groupe de pêcheurs. Les deux premiers axes principaux sont définis par les activités de non pêche (Arrêt et Campagne) communes aux différents groupes de par leur faible relation avec les autres activités. L'axe III oppose les activités "Repos" et "Ligne" des activités "Filets mixtes" et "Filet". L'axe IV est caractérisé par les lignes, notamment "Lignes mixtes".

L'indice de liaison et le coefficient RV^3 nous informent sur la dimensionalité des groupes et leurs relations : les stratégies S1 et S2 présentent une dimensionalité plus marquée que S4, S1 présentant une structure commune plus riche avec S4, puis avec S2 et enfin S2 avec S4. Pour les périodes, on note une décroissance de la dimensionalité au cours du temps (donc passage à une structure plus simple). Les structures communes les plus riches sont entre P1 et P2, P3 avec P1-P2 et enfin P4 qui montre le moins de structure commune avec les autres périodes.

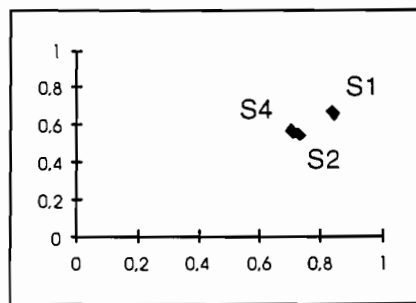
Les corrélations entre les variables canoniques et les facteurs de l'analyse globale indiquent que l'on a bien des structures communes mises en évidence entre les différents groupes pour tous les axes de l'analyse globale. Ces structures communes, qui correspondent à des directions d'inertie plus ou moins importantes pour les groupes permettent une bonne qualité de représentation des groupes.

Représentation des groupes

1. Trois groupes « Stratégies » comprenant chacun quatre variables « Périodes » codées flous



Axes 1 et 2



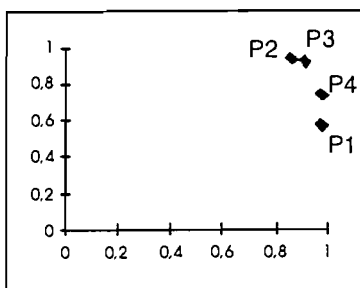
Axes 3 et 4

Qualité de représentation des groupes sur les axes

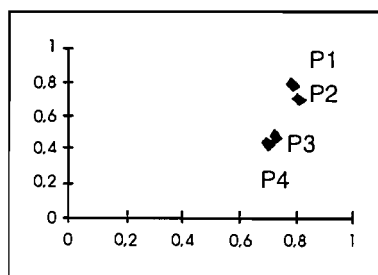
Groupe	Axes 1 et 2	Axes 3 et 4	Axes 1 à 4
S1	0,528	0,431	0,869
S2	0,544	0,293	0,837
S4	0,497	0,326	0,823

³ Coefficient de corrélation vectorielle qui est un coefficient d'association entre deux tableaux de données et qui a été défini par Robert et Escoufier (1976).

2. Quatre groupes « Périodes » comprenant chacun trois variables « Stratégies » codées flous



Axes 1 et 2



Axes 3 et 4

Qualité de représentation des groupes

Groupe	Axes 1 et 2	Axes 3 et 4	Axes 1 à 4
P1	0,360	0,351	0,711
P2	0,496	0,349	0,845
P3	0,606	0,272	0,878
P4	0,626	0,284	0,910

Dans le cas du codage flou, les coordonnées des modalités des variables qualitatives correspondent aux centres de gravité des individus ayant cette modalité, pondérés par les valeurs de la modalité. Les relations entre les variables qualitatives (profils de transition d'une stratégie-période) et les facteurs sont analysées globalement à partir du carré des rapports de corrélation. Ces informations permettent de comparer globalement les différents groupes.

Comparaison des trois stratégies de pêche

Carrés des rapports de corrélation entre les facteurs et les variables (*1000) et rapport d'inertie inter / inertie totale associé à la partition définie par la variable (INR) :

Stratégie S1	F1	F2	F3	F4	F5	INR
P1	795	899	718	689	782	772
P2	848	886	813	629	757	714
P3	902	918	226	635	246	678
P4	940	971	844	595	238	703

Stratégie S2	F1	F2	F3	F4	F5	INR
P1	777	223	751	665	126	562
P2	853	942	753	866	163	743
P3	697	650	384	75	1	399
P4	865	894	552	259	1	537

Stratégie S4	F1	F2	F3	F4	F5	INR
P1	871	285	545	662	620	539
P2	740	898	695	546	610	640
P3	831	931	703	564	505	731
P4	899	200	595	324	168	437

On obtient des valeurs à peu près identiques pour les tableaux des 4 périodes décrites par les 3 stratégies. Ces tableaux permettent de résumer les spécificités de chacun des groupes de la structure commune, en rapport à la signification des axes de l'analyse globale.

6. CONCLUSION

La comparaison de la dynamique de différentes stratégies d'exploitation de la pêche artisanale sénégalaise est abordée par l'analyse de tableaux de fréquences particuliers que sont les tableaux d'échanges, sur le principe de l'Analyse Factorielle des Correspondances. La démarche, basée sur trois approches successives, a fait ressortir les points suivants :

1. Les données, caractérisées par une forte inertie du système et de grands écarts entre la fréquence des différents états, se traduisent par des matrices d'échanges présentant une forte concentration sur la diagonale et des lignes de poids de niveau très différent.
2. L'Analyse Factorielle des Correspondances sur les matrices d'échanges révèle :
 - la forte variabilité entre les tactiques (au détriment de la variabilité entre les groupes),
 - le rôle particulier de la diagonale : reconduction de la même activité au cours du temps,
 - les tactiques spécifiques aux stratégies - périodes (spécialisation des stratégies d'exploitation).
3. L'Analyse Factorielle des Correspondances des tableaux de Burt obtenus en croisant les catégories de variables deux à deux, révèle :
 - les interactions d'ordre deux entre les tactiques et les périodes,
 - un résumé pertinent des stratégies au détriment de la dynamique intra - période.
4. La description complète du système a été faite pour chaque période par un modèle de Markov qui a donné un ajustement de bonne qualité de la dynamique du système (Proceedings of *OSDA '95*, Paris). L'Analyse Factorielle Multiple sur les matrices de transition met finalement en évidence :
 - la relation entre les stratégies, les périodes et les tactiques en ne différenciant pas les tactiques rares des tactiques fréquentes,
 - la caractérisation globale de la dynamique des groupes et la mise en évidence des structures communes et de leurs différences.

Bibliographie

- Cazes P. (1990) Quelques Considérations sur l'Utilisation des Eléments Supplémentaires en Analyse Factorielle. *Revue Modulad* n°5, pp. 29-44.
- Dazy F. et J. F. Le Barzic (1996) L'analyse des données évolutives : méthodes et applications. Technip, Paris. 227 p.
- Escofier B. et D. Drouet (1983) Analyse des différences entre plusieurs tableaux de fréquence. Les cahiers de l'Analyse des Données. Vol. VIII, N°4, pp. 491-499.
- Escofier B. (1984) Analyse factorielle en référence à un modèle application à l'analyse de tableaux d'échanges. *Revue de Statistique Appliquée*, Vol. XXXII, N°4, pp. 25- 36.
- Escofier B. et J. Pages (1993) Analyses factorielles simples et multiples : objectifs, méthodes et interprétation. Dunod, Paris. 267 p.
- Ferraris J., M. Gettler-Sunuma, C. Pardoux and H. Tong (1996) Knowledge extraction using stochastic matrices : application to elaborate fishing strategies. *In "Ordinal and symbolic data analysis"*, E. Diday, Y. Lechevallier & O. Opitz Eds. Springer-Verlag, Series Studies in classification, data analysis and knowledge organization, pp 103-112.
- Gettler-Sunuma M., E. Perinel and J. Ferraris (1994) Automatic aid to symbolic cluster interpretation. *In "New approaches in classification and data analysis"*, E. Diday, Y. Lechevallier, M. Schader, P. Bertrand & B. Burtschy Eds. Springer-Verlag, Series Studies in classification. pp 404-413.
- Laloë F. et A. Samba (1991) A simulation model of artisanal fisheries of Senegal. *ICES Mar. Sci. Symp.* 193 : 281 - 286.
- Pech N. (1998) Approche statistique d'une exploitation halieutique : le cas de la pêche artisanale au Sénégal. Thèse de l'Université Montpellier II, 250 p.
- Robert P. and Escoufier Y. (1976) An unifying tool for linear multivariate methods : the Rv coefficient. *Applied Statistics*, 25, (.3), pp. 257-265.

ANNEXE

Tableau de BURT construit pour la stratégie S1 : (les états de départ, au temps t, sont indiqués avec un +)

	L+	FD+	LPo+	LL+	Rep+	Arr+	Cam+	FF+	L	FD	Lpo	LL	Rep	Arr	Cam	FF	P1	P2	P3	P4
L+	14770	0	0	0	0	0	0	0	13123	0	37	71	1186	297	56	0	5807	3919	779	4265
FD+	0	26	0	0	0	0	0	0	5	21	0	0	0	0	0	0	26	0	0	0
LPo+	0	0	0	625	0	0	0	0	0	37	0	519	5	59	5	0	0	537	88	0
LL+	0	0	0	1813	0	0	0	0	85	5	18	1507	166	20	12	0	756	911	45	101
Rep+	0	0	0	0	2638	0	0	0	1138	0	42	190	1165	50	53	0	1285	678	309	366
Arr+	0	0	0	0	0	4026	0	0	316	0	5	26	32	3626	21	0	422	904	527	2173
Cam+	0	0	0	0	0	0	0	16313	0	24	0	5	0	63	59	16162	0	1220	3555	8390
FF+	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
L	13123	5	37	85	1138	316	24	0	14728	0	0	0	0	0	0	0	5775	3883	805	4265
FD	0	21	0	5	0	0	0	0	0	26	0	0	0	0	0	0	26	0	0	0
Lpo	37	0	519	18	42	5	5	0	0	0	626	0	0	0	0	0	538	88	0	0
LL	71	0	5	1507	190	26	0	0	0	0	0	1799	0	0	0	0	763	904	51	81
Rep	1186	0	59	166	1165	32	63	0	0	0	0	0	2671	0	0	0	1316	641	308	406
Arr	297	0	5	20	50	3626	59	0	0	0	0	0	0	4057	0	0	426	904	534	2193
Cam	56	0	0	12	53	21	16162	0	0	0	0	0	0	0	16304	0	1209	3635	8352	3108
FF	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
P1	5807	26	537	756	1285	422	1220	0	5775	26	538	763	1316	426	1209	0	10053	0	0	0
P2	3919	0	88	911	678	904	3555	0	3883	0	88	904	641	904	3635	0	0	10055	0	0
P3	779	0	0	45	309	527	8390	0	805	0	0	51	308	534	8352	0	0	0	10050	0
P4	4265	0	0	101	366	2173	3148	0	4265	0	0	81	406	2193	3108	0	0	0	0	10053

Comparaison des trois stratégies de pêche par AFM : (Analyse de l'ensemble des groupes)

Inertie totale (= Somme des valeurs propres) = 11.682

Num	Inertie	%	%Cum.	Variation	Diagramme des inerties (= Valeurs propres)
1	2.7720	23.730	23.730	23.730	***** ***** ***** *****
2	2.3977	20.525	44.255	3.205	***** ***** ***** *****
3	2.2478	19.242	63.498	1.283	***** ***** ***** *****
4	1.7930	15.349	78.847	3.894	***** ***** ***** *****
5	1.1382	9.743	88.590	5.606	***** ***** ***** *****
6	.9350	8.004	96.594	1.739	***** ***** ***** *****
7	.3978	3.406	100.000	4.599	***** ***** ***** *****

Comparison of Fishing Strategies with Factorial Analysis on Exchange Arrays

Ferraris Jocelyne, IRD, BP A5, 98848 - NOUMEA, & Pardoux Catherine, LISE-CEREMADE, Université Paris-IX, Place M^{al} de Lattre de Tassigny, 75775 - Paris Cedex 16.

Dynamics of the senegalese fishery is analysed from multivariate categorical time series and frequency arrays describing the exchanges between fishing tactics. The comparison of temporal strategies, built by cluster analysis, is made with Correspondence Analysis applied to a sequence of arrays and to Burt's arrays and with Multiple Factorial Analysis on fuzzy variables.

**ESTIMATION DES TAUX DE MORTALITE M ET F
POUR L'ANCHOIS :
PRESENTATION GENERALE ET PREMIERS RESULTATS**

Olivier Gavart (olivier.gavart@univ-pau.fr)
Marc Artzrouni (marc.artzrouni@univ-pau.fr)
Laboratoire de Mathématiques Appliquées
UPRES A 5033
Avenue de l'Université
64000 Pau

Summary

Fishing and natural mortality rates F and M for the Bay of Biscay anchovy are estimated (between 1987 and 1995) using yearly anchovy biomass estimates obtained by the Daily Egg Production Method (DEPM) and monthly catches. Two methods (based on different hypotheses) are considered, one of which leads to a multiple decrement technique generally used in human demography. We give some results, and we estimate uncertainties on M and F based on the errors made on catch and biomass estimates. Possible improvements are discussed, that would incorporate additional data on fishing efforts.

Résumé

On estime les taux de mortalité par pêche F et de mortalité naturelle M de l'anchois du Golfe de Gascogne de 1987 à 1995, d'après les estimations mensuelles de biomasse obtenues par la méthode de production journalière d'œufs (DEPM) et les captures mensuelles. Pour cela, on présente deux méthodes (issues d'hypothèses différentes) dont l'une conduit à une technique utilisée principalement en démographie humaine, dite 'd'extinction multiple'. On donne quelques résultats puis on estime les incertitudes concernant M et F provenant des erreurs d'estimation des captures d'une part et des estimations de biomasse d'autre part. On conclut sur les améliorations à apporter aux estimations de M et F par l'intégration de données sur les efforts de pêche.

1. Introduction

Le but de ce travail est d'estimer les taux de mortalité naturelle M et par pêche F de l'anchois du Golfe de Gascogne (*Engraulis Encrasicolus*) pour la période allant de juin 1987 à 1995. Les méthodes généralement utilisées jusque là relevaient de l'analyse séquentielle des structures démographiques révélées par les captures (VPA), techniques introduites par J.A Gulland (1955). On rappelle que pour utiliser ces méthodes on doit, pour une cohorte donnée, se fixer une valeur constante de M et une estimation de F pour le dernier groupe d'âge. On en déduit F pour l'avant dernier groupe d'âge ainsi de suite jusqu'à l'année du recrutement. D'après R. Jones (1961) ou P.K Tomlinson (1971), le fait de 'remonter' la cohorte assure la convergence de F vers sa vraie valeur. Cette méthode nécessite cependant que l'exploitation soit la principale cause de mortalité (Mesnil, 1980). Or les caractéristiques biologiques de l'anchois ne vérifient pas ces hypothèses. En effet, d'une part la mortalité naturelle de l'anchois est élevée et surtout très variable, et d'autre part son cycle court (2 à 3 ans) ne permet pas à la convergence de F de se mettre en place. D'où notre choix de traiter ce problème directement à partir des estimations annuelles d'abondance obtenues par la méthode de production journalière des œufs (appelée DEPM pour Daily Eggs Production Method) et des données mensuelles de captures.

Nous allons exposer dans un premier temps les techniques utilisées pour calculer M et F puis surtout les méthodes utilisées pour estimer les incertitudes commises sur M et F . Enfin, nous donnerons quelques résultats numériques avant de conclure.

2. Position du problème

Les estimations de biomasse reproductrice, c'est-à-dire de biomasse totale car les anchois sont matures dès leur première année, sont donc faites par DEPM au moment de la ponte (en mai-juin). Cette biomasse totale est répartie en trois groupes : les anchois d'âge 1, ceux d'âge 2 et les âges '3 et +'. Par simplification, nous considérerons que ce sont des estimations au 1^{er} juin. Les recrues issues de la ponte de juin ne sont exploitables qu'à partir de septembre (compte tenu de leur taille et des moyens de pêche utilisés). Nous allons suivre l'évolution des poissons cohorte par cohorte entre le 1^{er} juin d'une année et le 1^{er} juin de l'année suivante. Pour la cohorte que l'on décide d'étudier, nous disposons des estimations :

- De la biomasse "de départ" au 1er juin (que nous appellerons $B(0)$).
- De la biomasse "d'arrivée" au 1er juin suivant (que nous appellerons $B(12)$).
- Des captures mensuelles associées à cette cohorte.

Pour fixer les idées, si on travaille sur les poissons d'âge 1 en juin 91 (estimés à 670.5×10^6) il faudra considérer les captures d'âge 1 de juin à décembre 91 puis les captures d'âge 2 de janvier à mai 92. On devra alors 'retomber' sur l'estimation $B(12)$ de la biomasse totale d'âge 2 au 1^{er} juin 92.

3. Méthode

Le pas de temps considéré pour l'étude d'une cohorte sera le mois, il sera désigné par i , i variant de 1 à 12. Ainsi $B(i)$ désignera la biomasse (en nombre) à la fin du mois i , $C(i)$ les captures du mois i , M_i et F_i les taux mensuels de mortalité naturelle et par pêche respectivement. Ces taux sont supposés constants pour un mois donné mais peuvent varier d'un mois à l'autre.

Plusieurs méthodes ont été utilisées pour calculer M et F à partir des données ($B(0)$, $B(12)$, et les captures). Toutes ces méthodes ont en commun une hypothèse d'indépendance : on suppose, par définition de M et de F , qu'à tout instant t le risque de disparaître associé à une cause ne dépend pas de la présence d'autres risques associés à d'autres causes. Cette hypothèse se traduit par une équation que l'on retrouve par exemple dans le modèle exponentiel de Ricker :

$$B(t) = B(0) \times \exp^{\int_0^t (-M(s) - F(s)) ds}$$

Cette équation, très familière aux halieutes est donc une simple traduction de cette hypothèse d'indépendance.

Ensuite, les méthodes se différencient par l'hypothèse supplémentaire que l'on choisit :

3.1. Soit on suppose que pour chaque mois le rapport des taux instantanés mensuels est constant :

$$\frac{F_i}{M_i} = K_i$$

où K_i est la constante de proportionnalité.

La connaissance des taux M_i et F_i permet de retrouver M et F par les relations simples :

$$M = \sum_{i=1}^{12} M_i \quad \text{et} \quad F = \sum_{i=1}^{12} F_i$$

En fait, cette hypothèse de proportionnalité est toujours vérifiée puisqu'on suppose que sur chaque mois ces taux sont constants, ils sont donc a fortiori proportionnels puisque la constante de proportionnalité K_i est indexée par le mois et peut donc changer chaque mois.

Cette hypothèse conduit à une méthode dite 'd'extinction multiple'.

3.2. Soit on suppose que le rapport des densités de disparition est constant.

Cette dernière hypothèse conduit à une approximation assez grossière permettant de trouver M et F . Cette hypothèse revient en fait à supposer que chaque mois, les décroissances dues à chacune des causes sont linéaires.

En notant $MN(i)$ le nombre de poissons qui meurent de mort naturelle le mois i , $mn(i)$ le taux de mortalité naturelle le mois i et $C(i)$ les captures, on obtient pour le premier mois, sous l'hypothèse faite ci-dessus :

$$mn(1) = \frac{MN(1)}{\left(B(0) - \frac{MF(1)}{2} \right)}$$

Cette relation permet de calculer $B(1)$ en fonction de $mn(1)$. En renouvelant ce raisonnement jusqu'à $B(12)$, on obtient une relation donnant $B(12)$ en fonction de la seule inconnue $mn(i)$. En supposant que $mn(i)$ est constante pour tout i (c'est à dire que

la mortalité naturelle est constante toute l'année) on peut calculer la valeur de $mn(i)=mn$ qui permet de 'retomber' sur la valeur de $B(12)$ dont on connaît une estimation.

$$B(12) = B(0) \times (1 - mn)^{12} + \left(\frac{mn}{2} - 1 \right) \times \sum_{k=1}^{12} (MF(k) \times (1 - mn)^{12-k})$$

On décide de choisir la première méthode, qui conduit à une technique dite "d'extinction multiple". Les résultats numériques obtenus par les deux méthodes (en 3.1 et 3.2) sont très proches, on ne s'attardera donc pas sur une discussion concernant la justification de ce choix.

Les hypothèses sous-tendant chacune des méthodes proposent de traiter de l'interférence entre les deux causes de mortalité (Artzrouni et Le Bras, 1980). Notons toutefois que l'hypothèse de constance des rapports des taux instantanés mensuels est simple et permet de retomber sur une équation elle aussi bien connue des halieutes appelée l'équation des captures :

$$C(i) = \frac{F_i}{M_i + F_i} \times (1 - e^{-M_i - F_i}) \times B(i)$$

D'autres méthodes ont été testées, notamment en essayant de se fixer un profil de mortalité naturelle durant l'année ; l'hypothèse de constance étant peut-être discutable. Mais on se retrouvait devant un problème d'influence (relative) a priori sur les résultats. En effet, le choix, souvent subjectif, du profil de mortalité naturelle pouvait avoir une forte influence sur les résultats numériques concernant M et F. Puisque le bien-fondé d'imposer tel type de profil était difficilement justifiable, on supposera donc pour l'instant que le taux de mortalité naturelle reste constant au cours de l'année pour une même cohorte.

Quoiqu'il en soit, on ne s'attardera pas sur une discussion concernant le choix de la méthode utilisée, qui serait d'un intérêt purement théorique : on constate en effet, d'après les résultats numériques obtenus, une grande robustesse des résultats par rapport à la méthode choisie, que ce soit avec les différentes hypothèses permettant de traiter l'interférence des 2 causes de mortalité, ou que ce soit en imposant des profils de mortalité naturelle très différents.

Ces résultats ont été obtenus à l'aide de programmes réalisés pour chacune des méthodes précitées. Ces programmes permettent de calculer M et F à partir des 14 entrées : $B(0)$, $B(12)$ et les 12 captures mensuelles intermédiaires.

4. Discussion sur les incertitudes

Les résultats précédemment obtenus sont certes intéressants mais ils auraient pu être approximativement trouvés à l'aide d'un calcul rapide fait sur l'année considérée avec comme pas de temps l'année toute entière et en utilisant les captures totales de l'année. Ce qui nous intéresse beaucoup plus, c'est de connaître l'influence numérique qu'ont les incertitudes concernant les données sur les résultats. Précisons immédiatement qu'une résolution analytique de ce problème est impossible. En effet, les données entachées d'erreurs (les captures ainsi que $B(0)$ et $B(12)$) sont utilisées de façon complexe par le programme chargé de calculer les taux de mortalité. Ainsi, nous n'aurons pas d'autre recours que d'utiliser des méthodes de type Monte-Carlo pour avoir une idée des incertitudes concernant M et F.

Les incertitudes concernant les données à prendre en compte sont de deux ordres.

4.1 Incertitudes sur les estimations de biomasse par DEPM ($B(0)$ et $B(12)$).

Pour prendre en compte ces incertitudes, nous nous sommes servis des premières conclusions du groupe de travail étudiant ce problème (Anonyme, 1997a). Ainsi, nous faisons l'hypothèse que $B(0)$ et $B(12)$ sont des variables aléatoires qui suivent des lois normales. Leurs moyennes sont les valeurs utilisées dans les calculs du chapitre 3 et leurs coefficients de variation sont donnés par le groupe de travail avec les estimations de $B(0)$ et $B(12)$ et sont généralement voisins de 0.2. Les lois normales suivies par $B(0)$ et $B(12)$ sont indépendantes car ce sont des estimations issues de campagnes différentes.

Puisqu'il est impossible d'obtenir une relation analytique de M et F à partir de $B(0)$ et $B(12)$, nous avons simulé aléatoirement un grand nombre (n) de fois les lois normales suivies par les variables aléatoires donnant $B(0)$ et $B(12)$. Puis nous avons appliqué la méthode d'extinction multiple avec les deux valeurs obtenues par cette simulation, et ce n fois, tout en supposant pour l'instant que les données de captures étaient exactes. On obtient ainsi n valeurs de M et n valeurs de F dont on peut tirer les caractéristiques statistiques de base (moyenne, médiane, coefficient de variation, intervalle de confiance...) ou des représentations graphiques (histogrammes, diagramme de Tukey, box plot...).

4.2 Incertitudes concernant les captures.

Pour connaître ces incertitudes, il est nécessaire de savoir comment sont évaluées ces captures. Elles le sont en trois étapes, et à chacune de ces étapes on précise la provenance des incertitudes et la manière de les prendre en compte (Anonyme, 1997b).

1- Les anchois débarqués sont d'abord répartis par taille communautaire. Il existe 4 tailles communautaires : Sachant qu'on appelle moule le nombre d'anchois par kilo, la taille T1 correspond aux moules inférieurs à 30 (les plus gros poissons), la taille T2 aux moules entre 31 et 50, T3 entre 51 et 84, et T4 supérieurs à 85. On estime que cette répartition est relativement fiable, on la supposera sans erreur.

2- Ensuite, pour certains ports de référence et pour certaines périodes, les informations récoltées concernant les répartitions par moule des débarquements nous ont permis d'obtenir un moule moyen correspondant à chaque taille communautaire. On en déduit ainsi le nombre d'anchois par taille communautaire et par mois. On supposera que l'erreur commise sur l'estimation du moule moyen est de plus ou moins 5 à 10 % suivant la période et la taille communautaire considérées. On simulera donc la valeur de ce moule moyen par une variable aléatoire suivant une loi uniforme de paramètres appropriés.

3- Enfin, la lecture des otolithes sur des prélèvements donne la répartition par âge pour un moule moyen associé à une taille communautaire et une période données. Les pourcentages de ces clés de répartitions contiennent des incertitudes. On suppose que chacun des pourcentages de ces clés suivent une loi uniforme avec une incertitude de plus ou moins 5 à 10 % suivant les périodes de l'année. Nous veillerons à ce que les sommes sur les âges de ces pourcentages fassent bien 100%. Nous prendrons aussi en compte le fait suivant : lorsque par exemple une erreur est commise sur le pourcentage des poissons d'âge 0 elle doit se répercuter sur le pourcentage des poissons d'âge 1.

A l'issue de ces trois étapes on obtient, pour un mois donné, la répartition par âge des captures. Nous devons donc simuler, pour les douze mois qui nous intéressent les incertitudes indiquées aux points 2 et 3 pour obtenir un nouveau vecteur de douze captures.

Nous pouvons désormais prendre en compte toutes les incertitudes sur les données. On simule n fois $B(0)$, $B(12)$ et les douze captures mensuelles correspondantes et on applique à ces 14 données le programme donnant M et F à l'aide de la méthode d'extinction multiple. On obtient ainsi n valeurs de M et n de F pour lesquelles on peut calculer les caractéristiques statistiques que l'on veut (moyenne, médiane, coefficient de variation...). Pour les exemples numériques suivants, nous avons choisi $n=2000$.

5. Premiers résultats, conclusion, perspectives

Nous n'allons pas donner une liste exhaustive des résultats, nous nous contenterons du tableau suivant qui donne les taux de mortalité M et F de 1987 à 1992 suivant l'âge (soit d'âge 1 à âge 2, soit d'âge 2 à âge 3), résultats obtenus par la méthode d'extinction multiple en supposant les données exactes. Les autres méthodes donnent des résultats similaires.

Années	Agés	M	F
87-88	1-2	0.524	0.409
87-88	2-3	1.802	0.138
88-89	1-2	1.858	0.575
88-89	2-3	2.284	0.612
89-90	1-2	-	-
89-90	2-3	1.397	0.326
90-91	1-2	2.185	0.777
90-91	2-3	1.996	1.873
91-92	1-2	0.302	0.862
91-92	2-3	2.120	0.751

Ce tableau confirme juste ce qui a été dit en introduction sur la mortalité naturelle (élevée et variable) : Pour la période considérée et les différentes cohortes, M varie suivant les années entre 0.302 (période 91-92, âge 1) et 2.284 (en 88-89, âge 2) alors que F varie de 0.138 (87-88, âge 2) à 1.873 (en 90-91, âge 2).

Quant aux incertitudes, nous donnerons un exemple effectué sur la cohorte d'âge 1 en 91, pour la période 91-92. L'estimation de $B(0)$ est de 670.5×10^6 et le coefficient de variation associé est de 0.2 donc $B(0)$ suit une loi normale de moyenne 670.5×10^6 et d'écart-type $0.2 \times 670.5 \times 10^6$. De même, $B(12)$ suit une loi normale de moyenne 209.3×10^6 et d'écart-type $0.2 \times 209.3 \times 10^6$. Par ailleurs, pour un mois donné, le vecteur des n captures a un coefficient de variation d'environ 0.05 qui est stable suivant le mois choisi.

Dans un premier temps, si on simule les captures uniquement, le vecteur des n valeurs de M a un coefficient de variation de 0.13 (Fig. 1). Ensuite, si on ne simule que les incertitudes sur $B(0)$ et $B(12)$, ce vecteur a alors un coefficient de variation de 0.42 (Fig. 2). Enfin, si on simule toutes les erreurs à la fois, le coefficient de variation sur les résultats concernant M est 0.53 (Fig. 3). Ces résultats concernant les incertitudes sur M provenant des incertitudes sur les données sont illustrés par les histogrammes suivants qui représentent les résultats obtenus pour M dans chacun des trois cas précités. Notons que les coefficients de variation ainsi obtenus varient d'une année à l'autre et suivant la cohorte étudiée. En effet, M est encore plus sensible à des variations sur les données lorsque les effectifs étudiés sont peu nombreux ou lorsque M est petit.

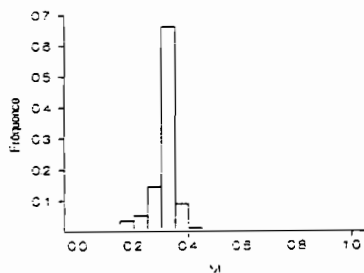


Fig.1

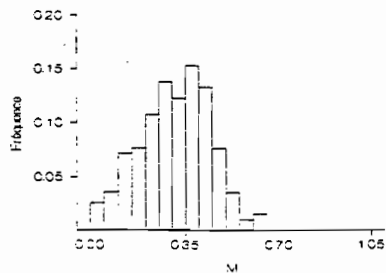


Fig. 2

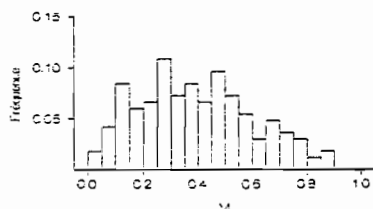


Fig.3

Comme on pouvait s'y attendre, le degré d'incertitude sur les résultats est très important, il traduit simplement l'évidence de la difficulté à estimer le nombre de poissons dans la mer et l'âge des poissons qui y sont pêchés. Ces incertitudes sont telles que pour la cohorte d'âge 1 en 1989 et d'âge 2 en 1990 le vecteur de 14 données donne une valeur de M négative, c'est-à-dire que la différence entre les estimations de populations de départ et d'arrivée est inférieure aux captures effectuées sur ce stock au cours de cette période.

Quelle que soit la procédure suivie, ce problème ne sera résolu qu'en reconnaissant la grande incertitude qui pèse sur les données, ce qui nous conduit à recevoir les résultats obtenus avec une certaine prudence.

On remarque par ailleurs que les moyennes et les médianes associées aux résultats où l'incertitude est prise en compte sont très proches des résultats obtenus à partir des données supposées sans erreurs bien que les coefficients de variations soient élevés.

Pour tenter de réduire les incertitudes et de résoudre le problème d'un M négatif, il faudra par exemple introduire des relations faisant intervenir l'effort de pêche. En effet, ces données étant disponibles pour certains ports de référence, on pourra les utiliser pour obtenir des informations complémentaires sur le rapport de $B(0)$ et $B(12)$ par exemple.

Références

ANONYME (1997a) - Premier rapport intermédiaire au contrat européen 95/PROP/018 -1 9p.

ANONYME (1997b) - Premier rapport intermédiaire au contrat européen 95/PROP/018 -2 4p.

GULLAND (J.A) - Estimation of growth and mortality in commercial fish populations - *Fish. Invest.* ,Londres, Ser.2, 18 (9) : 46p., 1955.

JONES (R.) - The assessment of the long term effects of changes in gear selectivity and fishing effort. *Mar. Res. Scoll.* , 2 : 19p. , 1961.

LE BRAS (H.), ARTZROUNI (M.) - Interférence, indifférence, indépendance. *Populations*, 6, 21p. , 1980.

MESNIL (B.) - Théorie et pratique de l'analyse des cohortes. *Rev. Trav. Inst. Pêches marit.*,44 (2) :34p. , 1980.

TOMLINSON (P.K) - A generalization of the Murphy catch equation. - *J. Fish. Res. Bd. Canada*, 27 : 5p. , 1970.

STRUCTURATION SPATIALE DE LA RESSOURCE HALIEUTIQUE : EXEMPLE D'UNE PECHERIE DEMERSALE A VANUATU (OCEANIE)

Espérance Cillaurren ⁽¹⁾, Monique Simier ⁽¹⁾ et Laurence Blanc ⁽²⁾

(1) *Laboratoire Halieutique et Ecosystèmes Aquatiques, Centre IRD,
B.P. 5045, 34032 Montpellier cedex 1 - France.*

(2) *Laboratoire d'Hydrobiologie Marine et Continentale,
Université Montpellier 2, Case c. 093, Place Eugène Bataillon,
34095 Montpellier cedex 5 - France.*

Résumé

Sur l'archipel de Vanuatu (Océanie), la pêcherie artisanale exploite une ressource démersale constituée essentiellement de onze espèces de poissons. Nous mettons en évidence ici par une Analyse Factorielle des Correspondances (AFC) la forte structure verticale de ces espèces. Puis deux autres méthodes multivariées sont utilisées pour étudier en détail les influences géographique et bathymétrique sur l'abondance. Une approche multi-tableaux (AFC de Foucart), traite le jeu de données comme une matrice à trois dimensions (espèces, îles, années) et montre une déstructuration progressive du gradient bathymétrique du Nord-Ouest au Sud-Est de l'archipel. Une approche par variables instrumentales (ACPVI) décompose la variabilité totale du tableau (espèces en colonnes, îles, années en lignes) par un modèle d'analyse de variance, montrant la faible variabilité géographique des espèces des strates profonde et superficielle. Elle permet aussi de modéliser l'abondance spécifique à partir des axes des différentes ACPVI.

Mots-clés

Vanuatu, pêche artisanale, poissons démersaux, répartition spatiale, AFC, analyse factorielle des correspondances, STATIS, méthodes multitableaux, ACPVI, variables instrumentales.

Summary

On the Vanuatu archipelago (Oceania), the artisanal fishery harvests a demersal resource mainly composed by eleven fish species. By a Factorial Correspondence Analysis (COA) we show the strong vertical structure of these species. Then two other multivariate methods are applied in order to study in details the geographical and bathymetric influences on the abundance. A multitable approach (Foucart's COA) considers the data set as a 3-way table (species, islands, deepness) and shows a progressive destructuring of the bathymetric gradient from the North-West to the South-East of the archipelago. An instrumental variables approach (PCAIV) splits the total variability of the table (species in columns, islands, years in rows) by an ANOVA model, showing the weak geographical variability of species from the deepest and superficial strata. This approach also allows building species abundance models from the different PCAIV axes.

Key-words

Vanuatu, artisanal fishery, demersal fish, spatial distribution, COA, factorial correspondence analysis, STATIS, multitable methods, PCAIV, instrumental variables.

Introduction

En halieutique, la prise en compte de l'espace dans la modélisation de la dynamique des populations est relativement récente (Lebreton, 1997). En effet, ce n'est qu'à partir de la seconde moitié du vingtième siècle que la diffusion de la ressource ou l'hétérogénéité de sa distribution dans l'espace ont été intégrées dans les modèles halieutiques. Cette démarche est encore considérée comme difficilement applicable par les acteurs sociaux.

Pour une pêcherie artisanale comme celle qui prévaut à Vanuatu, archipel océanien, et à l'instar des autres états insulaires du Pacifique, les activités sont fortement inféodées aux caractéristiques de l'espace, qu'il s'agisse de l'accès aux lieux de pêche ou de la disponibilité de la ressource. Les populations de poissons ciblés par cette pêche sont des espèces démersales qui vivent entre 50 et 500 m de fond au-dessus des pentes récifales externes. Leur faible caractère migratoire handicape leur extension, également limitée par la structure de l'archipel constitué d'îles hautes séparées par des chenaux profonds. L'hypothèse émise sur la localisation de la ressource exploitée est donc à étudier ; ses implications pour la régulation de l'effort de pêche selon les lieux et dans la durée sont évidentes, compte tenu des risques de surexploitation localisée.

Nous nous proposons ici de comparer l'application de trois méthodes d'analyse multivariée à un jeu de données provenant de cette pêcherie. Dans un premier temps nous situerons le contexte de la récolte des données et décrirons les méthodes utilisées (analyse des correspondances simple, analyse des correspondances de Foucart et analyse sur variables instrumentales). Puis nous présenterons les résultats de ces analyses. Les conclusions porteront sur l'apport comparé de ces méthodes pour l'étude de la ressource.

1. Matériel et méthodes

1.1. Données

Les données ont été collectées au sein d'une pêcherie artisanale développée à Vanuatu, archipel situé au Nord de la Nouvelle-Calédonie (Figure 1). Le pays est caractérisé par l'éclatement de son espace, résultat d'une activité sismique intense. Des îles ou groupes d'îles apparaissent ainsi, séparés par des chenaux dont la profondeur est supérieure à 1000 m et qui sont autant de barrières à la migration des espèces démersales étudiées.

Ces espèces vivent au-dessus des pentes récifales externes, à des profondeurs variant entre 50 et 500 m. La pêche se pratique de manière très uniforme par des petites embarcations - 5 à 8 m - équipées de moulinets à main sur lesquels sont montées des lignes. Onze espèces constituent l'essentiel - environ 80% - des prises. Deux familles principales sont représentées : les *Lutjanidae* (avec le genre *Etelis* et le genre *Pristipomoides* et deux espèces isolées) et les *Serranidae* (avec le genre *Epinephelus*).

Une base de statistiques de pêche a été constituée pendant dix ans, de 1982 à 1991, à partir d'enquêtes journalières portant sur les sorties de pêche. 10 000 sorties de pêche et 100 000 longueurs-poissons appartenant aux onze principales espèces capturées ont été enregistrées. Un volet du système d'enquête a permis, à travers le suivi des débarquements, de sélectionner une base de données destinée à étudier les caractéristiques biologiques et écologiques de la ressource. La capture par unité d'effort ou C.P.U.E. en kg par jour-ligne est considérée ici comme un indice de l'abondance de la ressource à un moment et sur un lieu donné. Divers tableaux de données peuvent être construits à partir de cette base, afin d'étudier la répartition spatio-temporelle des espèces étudiées.

Dans le travail présenté ici, nous nous sommes intéressés à l'étude de la répartition spatiale des espèces. Six îles ou groupes d'îles ont été définis (Figure 1), pour six strates de profondeur, allant de 125 à 375 m. Un tableau rectangulaire a été constitué, dont les onze colonnes sont les espèces étudiées, et les trente-six lignes des combinaisons île x profondeur. Chaque case de ce tableau contient une valeur de CPUE transformée en logarithme.

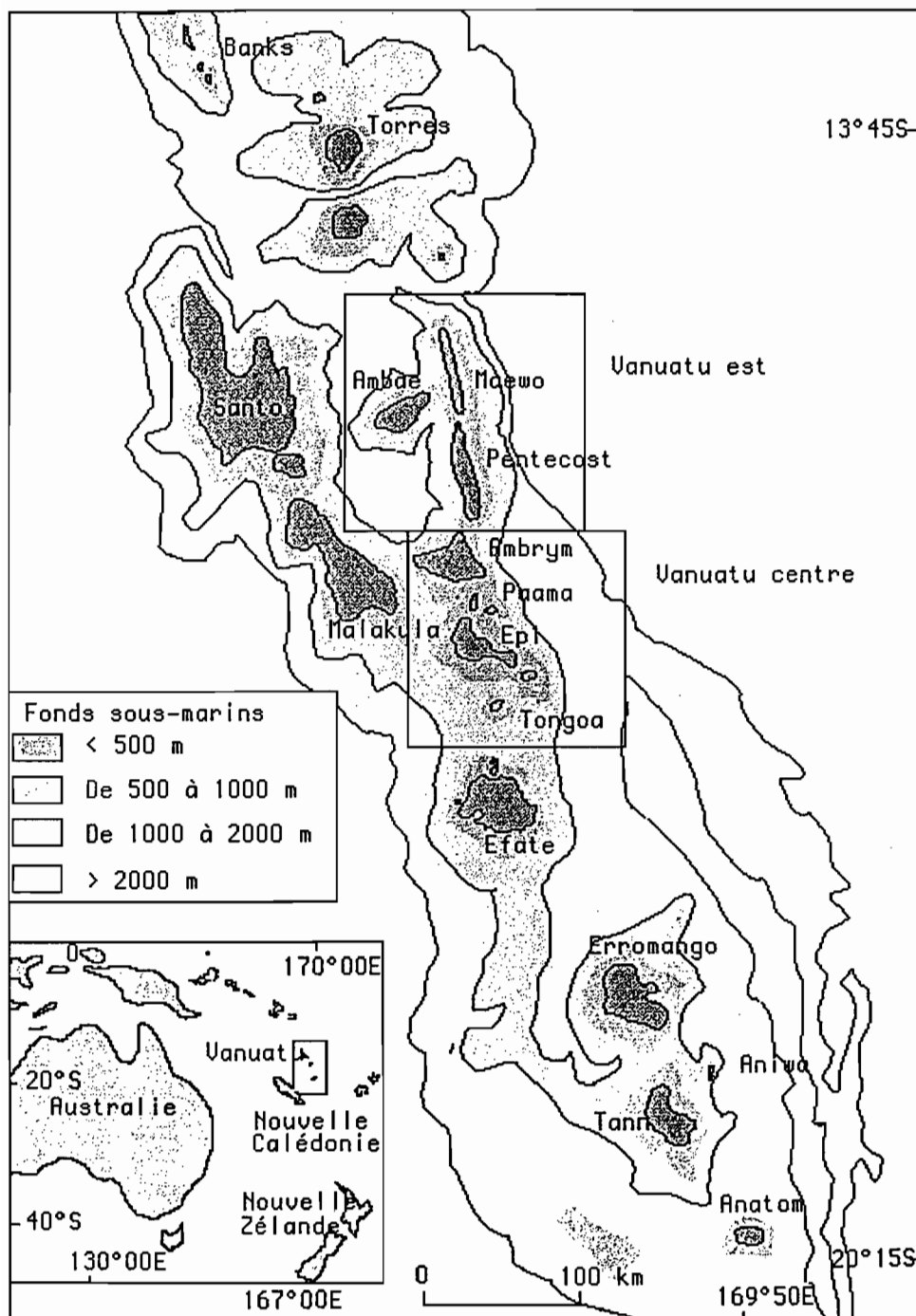


Figure 1 : L'archipel du Vanuatu. Situation géographique et profil bathymétrique. Localisation des 6 zones d'exploitation définies dans le cadre de cette étude : Santo, Vanuatu Est, Malekula, Vanuatu Centre, Efate, Tanna.

1.2. Méthodes

1.2.1. Analyse Factorielle des Correspondances simple

Afin de comparer les profils de répartition des espèces plutôt que leurs abondances absolues, nous avons fait le choix d'appliquer au tableau de données à 36 lignes et 11 colonnes une Analyse Factorielle des Correspondances (AFC) de préférence à une Analyse en Composantes Principales (ACP).

L'Analyse des Correspondances simple permet, par définition, de mettre en « correspondance » deux facteurs qualitatifs dont les modalités s'expriment respectivement en ligne et en colonne. Ici la situation est particulière dans le sens où le facteur définissant les lignes résulte lui-même de la combinaison de deux facteurs qui sont l'île et la profondeur. Pour décomposer leurs rôles respectifs, nous utilisons ensuite deux approches dont nous comparerons les avantages et les inconvénients.

1.2.2. Approche Multi-tableaux : Analyse des Correspondances de Foucart, 1978

La première approche envisagée consiste à considérer le tableau comme un cube de données à 3 indices : Espèces / Îles / Profondeurs (Figure 2).

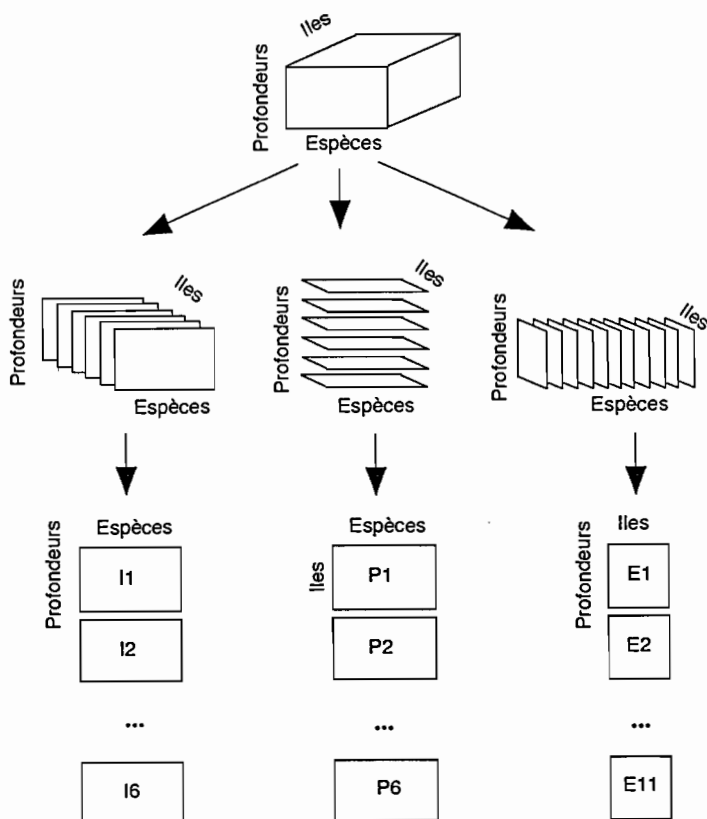


Figure 2 : L'approche multi-tableaux par l'AFC de Foucart. Le cube de données à trois indices (espèces, îles, profondeurs) peut être abordé de trois points de vue : un tableau espèces x profondeurs par île, un tableau espèces x îles par profondeur ou un tableau îles x profondeurs par espèce.

Apparentées aux méthodes STATIS décrites par Lavit *et coll.*, 1994, deux méthodes permettent de traiter ce type de situation. La première, fondée sur l'Analyse Factorielle des Correspondances, a été proposée par Foucart, 1978 et 1983. L'Analyse Triadique Partielle (ATP), introduite en écologie aquatique par Thioulouse et Chessel, 1987 et reprise récemment par Blanc *et coll.*, 1998, est fondée quant à elle sur une logique d'Analyse en Composantes Principales. Afin de rester dans une logique d'AFC, nous avons choisi ici l'AFC de Foucart.

Le principe de ces méthodes est de traiter le cube comme une série de tableaux à deux entrées. La première étape consiste à rechercher la structure commune à tous les tableaux, le compromis, et à l'analyser. Dans un second temps, sa reproductibilité est étudiée à travers les projections des tableaux initiaux sur les axes issus de l'analyse du compromis.

Le cube de données peut être abordé de trois manières et le choix de la présentation des tableaux fait au début de l'analyse est fondamental car il reflète la question posée par l'expérimentateur. En effet chaque point de vue correspond à un objectif bien précis. Ainsi, en prenant un tableau espèces x profondeurs par île, on s'intéressera à la répartition verticale moyenne des espèces et à son éventuelle variabilité d'une île à l'autre. Avec un tableau espèces x îles par profondeur, on décrira la répartition géographique moyenne des espèces et sa variabilité selon la profondeur. Enfin, présenter un tableau île x profondeur pour chaque espèce nous amènera à décrire la répartition moyenne de la ressource dans l'espace, toutes espèces confondues, et à étudier la variabilité spécifique autour de cette moyenne.

1.2.3. Approche par les Variables Instrumentales : Analyse en Composantes Principales sur Variables Instrumentales

L'approche par les Analyses sur Variables Instrumentales (Rao, 1964 ; Sabatier *et coll.*, 1989, Lebreton *et coll.*, 1991), combine l'ACP ou l'AFC d'un tableau de variables à expliquer (échantillons-espèces) avec une régression linéaire sur un second tableau de variables explicatives (échantillons-variables instrumentales). Elle a été utilisée récemment par Pech et Laloë (1997) pour l'analyse de la structuration spatio-temporelle de rendements de pêche.

Dans le cas où les variables instrumentales sont qualitatives, l'analyse sur variables instrumentales consiste à décomposer la variance totale du tableau de données selon le principe de l'analyse de variance (Figure 3). Dans le cas qui nous intéresse, le tableau de données à expliquer, Y, à 11 colonnes-espèces et 36 lignes est associé à 2 variables explicatives à 6 modalités chacune qui sont l'île et la profondeur ; le plan est équilibré : toutes les combinaisons sont également représentées. L'analyse de variance des 11 variables décompose le tableau Y en la somme des tableaux $Y_{île}$ et Y_{prof} qui prennent en compte les effets principaux, et d'un tableau $Y_{île,prof}$ correspondant à l'interaction entre île et profondeur. Dans notre cas, il n'y a pas de répétitions : chaque combinaison île-profondeur définit une seule observation et il est donc impossible de distinguer l'interaction de la résiduelle.

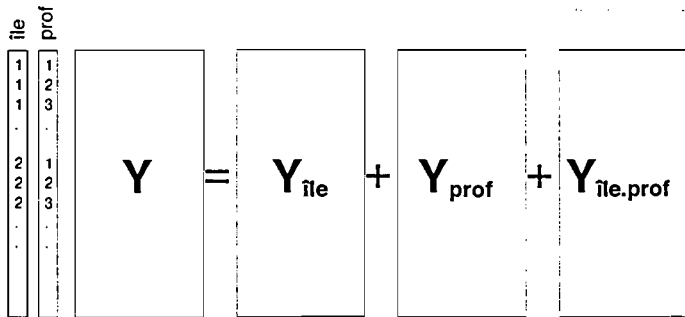


Figure 3 : L'approche par les variables instrumentales. Le tableau de données centrées-réduites Y, associé à deux facteurs de variation (île et profondeur), peut se décomposer selon le principe des analyses de variance en un tableau de moyennes par île ($Y_{île}$), un tableau de moyennes par profondeur (Y_{prof}) et un tableau d'interaction île.profondeur, équivalent à la résiduelle s'il n'y a pas de répétitions. L'ACP de chacun de ces tableaux est l'ACPVI selon respectivement l'île, la profondeur et l'interaction.

Le programme utilisé pour cette analyse (Vey, 1998) permet de décomposer les effets dus à 3 facteurs qualitatifs et de reconstituer les variables de Y à partir des axes des ACPVI. Il ne permet pas d'effectuer d'AFCVI, par conséquent, nous présenterons ici les résultats d'une ACPVI. Dans notre cas, il y a assez peu de différence entre les résultats de l'une et de l'autre, en raison de la forte stabilité des données transformées en logarithme.

2. Résultats

2.1. AFC simple

Dans une première étape, nous avons réalisé une AFC du tableau à 36 lignes et 11 colonnes. Les deux premiers axes prennent en compte respectivement 57 et 15% de l'inertie du tableau (Figure 4a). Cette analyse ne considère pas explicitement l'appartenance d'un relevé à une île ou à une profondeur, mais le calcul des points-moyens par île (Figure 4b) ou par profondeur (Figure 4c) et leur superposition sur la projection des lignes sur les axes en facilite l'interprétation.

L'axe 1 est prépondérant. Il exprime clairement le gradient bathymétrique (Figure 4b). Ce gradient correspond à une succession d'espèces (Figure 4d) des couches supérieures (125 / 175 m) puis intermédiaires (225 / 275 m), puis profondes (325 / 375 m).

L'axe 2 prend en compte seulement 15% de l'inertie et son interprétation est moins évidente (Figure 4c). La variabilité autour des points-moyens par île apparaît ici bien plus forte que pour la profondeur. On observe cependant une différenciation entre Santo-Malekula (au Nord-Ouest), Vanuatu Est et Centre et Efate, et enfin Tanna (au Sud-Est).

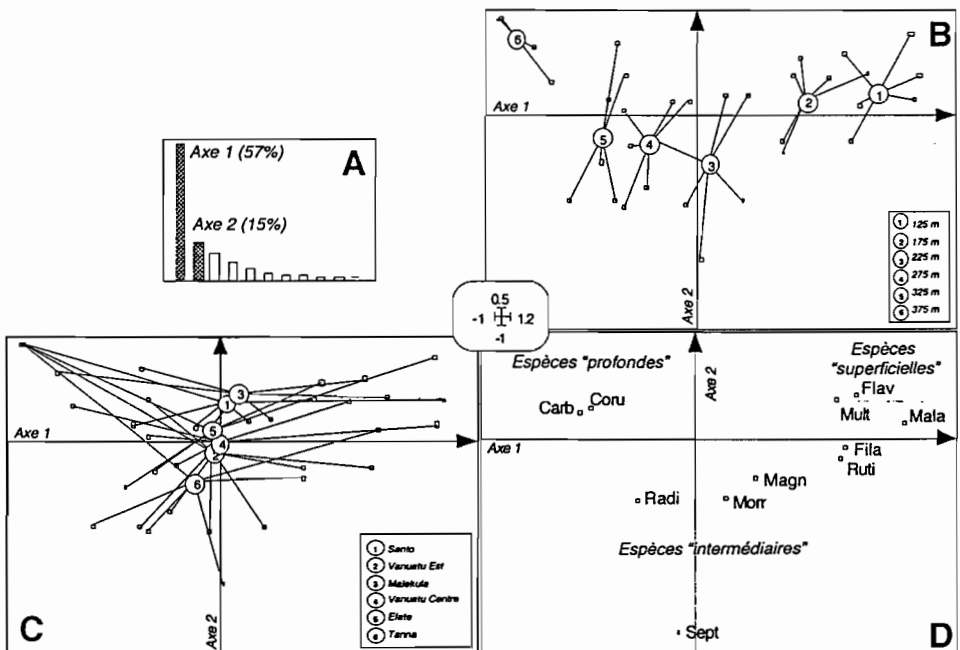


Figure 4 : Résultats de l'AFC du tableau à 36 lignes et 11 colonnes. A : histogramme des valeurs propres. B : Projection des 36 lignes sur le plan défini par les axes 1 et 2. Les points-moyens par profondeur sont superposés. C : Projection des 36 lignes sur le plan défini par les axes 1 et 2. Les points-moyens par île sont superposés. D : Projection des 11 colonnes-espèces sur le plan factoriel 1-2. Trois groupes d'espèces peuvent être identifiés en fonction de la profondeur. Légende des espèces : Carb : *Etelis carbunculus*, Coru : *Etelis coruscans* ; Radi : *Etelis radius* ; Fila : *Pristipomoides filamentosus* ; Flav : *Pristipomoides flavipinnis* ; Mult : *Pristipomoides multidens* ; Magn : *Epinephelus magniscuttis* ; Morr : *Epinephelus morrhua* ; Sept : *Epinephelus septemfasciatus* ; Mala : *Lutjanus malabaricus* ; Ruti : *Aphaeus rutilans*.

2.2. AFC de Foucart

L'AFC nous ayant permis de mettre en évidence la prépondérance du gradient bathymétrique, nous choisissons dans l'approche multi-tableaux de considérer un tableau espèces x profondeurs par île, afin d'étudier la variabilité géographique du gradient bathymétrique moyen.

L'analyse d'inertie du tableau compromis permet de retrouver sur le premier axe, représentatif de 87% de l'inertie totale (Figure 5a), le gradient de profondeur déjà observé sur l'AFC simple (Figure 5b).

L'analyse de la stabilité de cette structuration verticale moyenne d'une île à l'autre se fait en projetant les lignes (Figure 5c) et les colonnes (Figure 5d) des tableaux initiaux sur les axes du compromis. Une représentation graphique multifenêtrée par tableau (donc par île) en facilite l'interprétation.

Sur la projection des lignes (Figure 5c), on observe une déstructuration du gradient bathymétrique du Nord-Ouest (Santo-Malekula) vers le Sud-Est (Tanna). Ce phénomène est à relier à la contribution apportée par les espèces de la couche intermédiaire à l'axe 2 (Figure 5b), espèces plus abondantes dans le Sud-Est de l'archipel.

La projection des espèces (Figure 5d) permet de noter quelques anomalies par rapport au compromis, comme la présence en profondeur à Malekula de *E. morrhua*, espèce de la couche intermédiaire, ou bien la forte abondance de *E. radiosus*, généralement localisée dans la couche 325 m, jusqu'à 375m à Vanuatu Centre, ou encore la présence de *E. magniscuttis* dans les couches plus superficielles à Tanna.

2.3. ACPVI

Nous avons réalisé l'ACPVI du tableau selon les effets principaux île et profondeur et l'interaction île-profondeur. Les résultats sont présentés sur la figure 6.

Comme en ACP normée classique, les axes s'interprètent à l'aide des cercles de corrélations avec les modalités du facteur considéré. Ainsi, dans l'ACPVI selon l'île (Figure 6a), deux axes sont interprétables (50 et 26%) et sont définis respectivement par Tanna et Vanuatu Est. Dans l'ACPVI selon la profondeur (Figure 6b), on retrouve une fois encore le gradient de profondeur prépondérant sur l'axe 1 (76%).

Pour chaque ACPVI, les espèces sont projetées à l'intérieur de cercles de corrélation définis par des couples d'axes, comme en ACP normée classique (Figure 6c,d,e). Chaque espèce est représentée par 2 points, reliés par un segment de droite. Le point intérieur donne la corrélation entre la variable totale (avant modèle) et les axes concernés. Le point extérieur donne la corrélation entre la variable modélisée et les axes. Elevé au carré, le rapport entre les distances à l'origine de ces deux points, donne un R^2 , indicateur de la part de variation expliquée par le facteur considéré (île, profondeur ou interaction), pour la variable étudiée (espèce). On peut de cette manière quantifier l'influence des facteurs géographique et bathymétrique sur les variations d'abondance de chaque espèce : ainsi *E. septemfasciatus* est une espèce dont l'abondance varie à la fois selon la profondeur (0.34) et selon l'île (0.35), tandis que *E. carbunculus* est une espèce à variabilité essentiellement verticale (0.83 pour la profondeur et 0.05 seulement pour l'île). En faisant la moyenne des R^2 spécifiques pour les 11 espèces, on obtient un R^2 global, qui permet de classer les effets par ordre d'influence : ici on confirme la prépondérance de l'effet profondeur (0.53) par rapport à l'effet île (0.14), les 0.32 restants étant imputables à l'interaction, impossible à distinguer de la résiduelle. En règle générale, les espèces des couches supérieure et inférieure sont bien modélisées par le facteur profondeur, à la différence des espèces de la couche intermédiaire (Figure 6d).

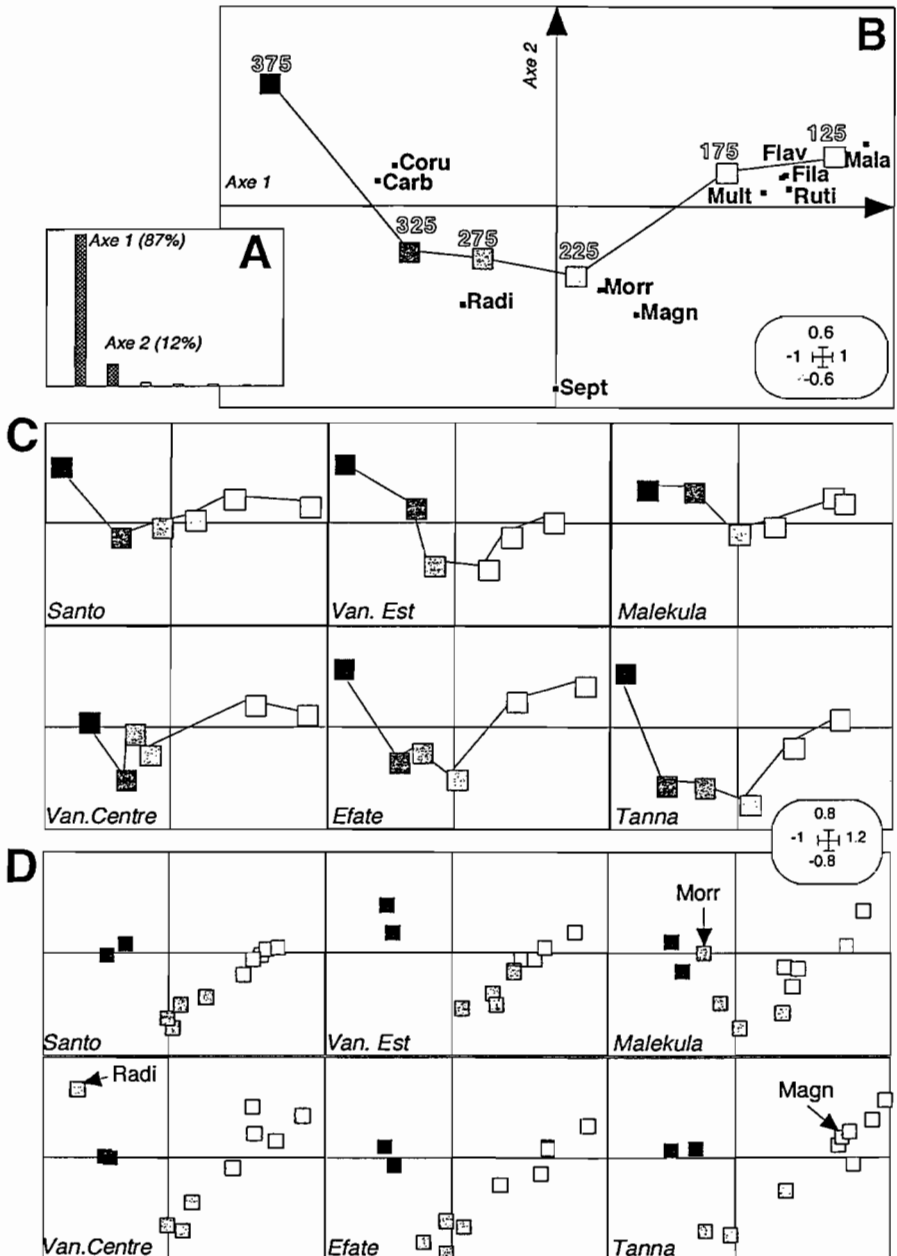


Figure 5 : Résultats de l'AFC de Foucart. A : Histogramme des valeurs propres. B : Projection des lignes (profondeurs) et des colonnes (espèces) sur les axes 1 et 2 de l'analyse du compromis. C : Projection des lignes (profondeurs) des 6 tableaux initiaux sur les axes 1 et 2 du compromis ; multifenêtrage par île. Les profondeurs sont figurées par des grisés allant du plus clair (125m) au plus foncé (375m). D : Projection des colonnes (espèces) des tableaux initiaux sur les axes 1 et 2 du compromis. En grisé clair, les espèces « superficielles », en grisé moyen les espèces « intermédiaires », en grisé foncé les espèces « profondes ». Seuls les points remarquables sont identifiés par leur code.

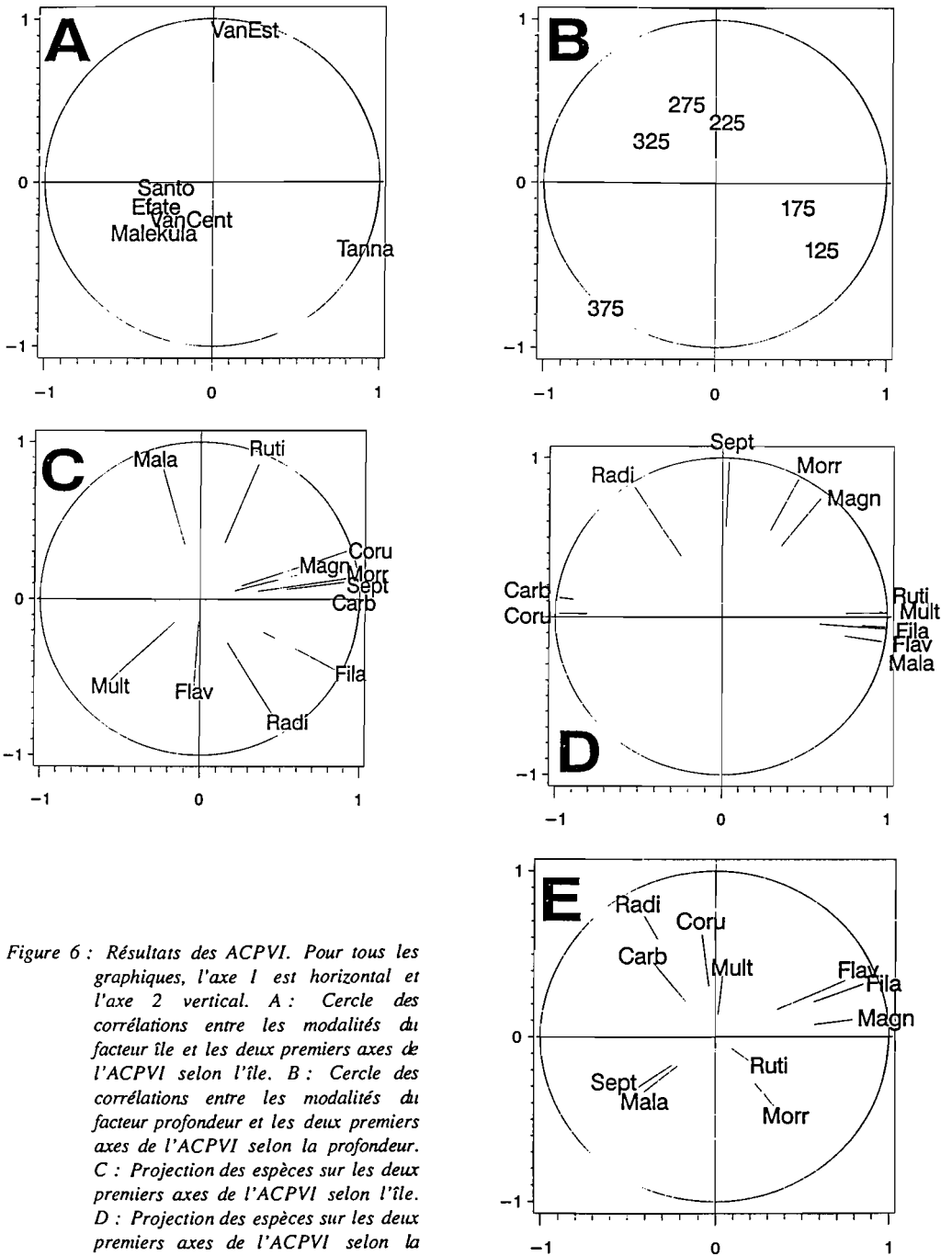


Figure 6 : Résultats des ACPVI. Pour tous les graphiques, l'axe 1 est horizontal et l'axe 2 vertical. A : Cercle des corrélations entre les modalités du facteur île et les deux premiers axes de l'ACPVI selon l'île. B : Cercle des corrélations entre les modalités du facteur profondeur et les deux premiers axes de l'ACPVI selon la profondeur. C : Projection des espèces sur les deux premiers axes de l'ACPVI selon l'île. D : Projection des espèces sur les deux premiers axes de l'ACPVI selon la profondeur. E : Projection des espèces sur les deux premiers axes de l'ACPVI selon l'interaction île*profondeur.

Au vu de ces projections, nous allons maintenant tenter de reconstituer les variations des deux espèces précitées (*E. septemfasciatus* et *E. carbunculus*) à partir des axes des ACPVI.

Pour *E. septemfasciatus*, un modèle fondé sur l'axe 1 de l'ACPVI selon l'île (Figure 7a) rend compte de la variabilité assez importante entre îles mais pas du gradient de profondeur. Un modèle fondé sur l'axe 2 de l'ACPVI selon la profondeur (Figure 7b) permet de reconstituer ce gradient, mais le reproduit à l'identique pour les 6 îles. Ce second modèle sous-estime considérablement l'abondance de *E. septemfasciatus* à Vanuatu Est et Tanna. L'addition de ces deux modèles (île et profondeur) et de l'interaction île-profondeur reconstituée à partir des axes 1 et 2 de l'ACPVI correspondante (Figure 7c), permet de reproduire plus fidèlement la variabilité de *E. septemfasciatus* (Figure 7d).

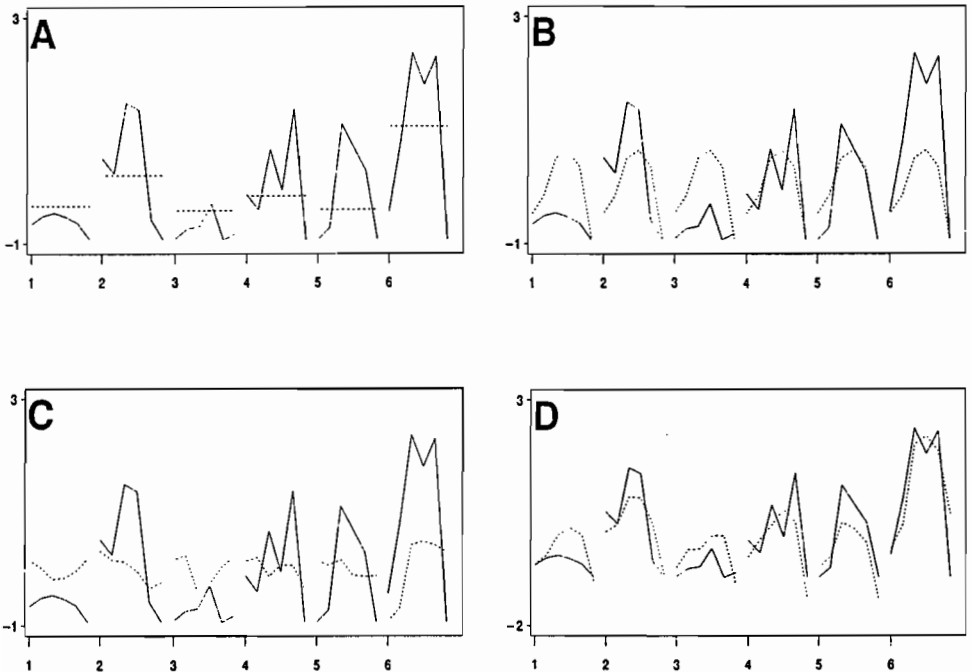


Figure 7 : Reconstitution des variations de CPUE (en log(kg/j.l.)) pour l'espèce *E. septemfasciatus* à partir des axes des ACPVI. A : Axe 1 de l'ACPVI selon l'île. B : Axe 2 de l'ACPVI selon la profondeur. C : Axes 1 et 2 de l'ACPVI selon l'interaction. D : Addition des 3 précédents modèles : Axe 1 (île) + Axe 2 (profondeur) + Axes 1 et 2 (île*profondeur). Sur chaque graphe, le gradient bathymétrique par île est représenté par un trait plein pour les valeurs observées et un pointillé pour la reconstitution. 1 : Santo ; 2 : Vanuatu Est ; 3 : Malekula ; 4 : Vanuatu Centre ; 5 : Efate ; 6 : Tanna.

L'abondance de *E. carbunculus* augmente avec la profondeur, de manière assez similaire quelle que soit l'île. La reconstitution de sa variabilité à partir d'un modèle fondé sur l'axe 1 de l'ACPVI selon l'île (Figure 8a) ne présente donc aucun intérêt. A l'inverse, un modèle fondé sur l'axe 1 de l'ACPVI selon la profondeur (Figure 8b) reconstitue assez fidèlement la variabilité de cette espèce. On n'apporte que peu d'information supplémentaire en additionnant l'effet île, l'effet profondeur et leur interaction, ici encore reconstituée à partir des axes 1 et 2 de l'ACPVI correspondante (Figure 8c). On constate cependant que ce dernier modèle (Figure 8d) améliore sensiblement la reconstitution pour la zone Sud-Est de l'archipel : Vanuatu Centre, Efate et Tanna.

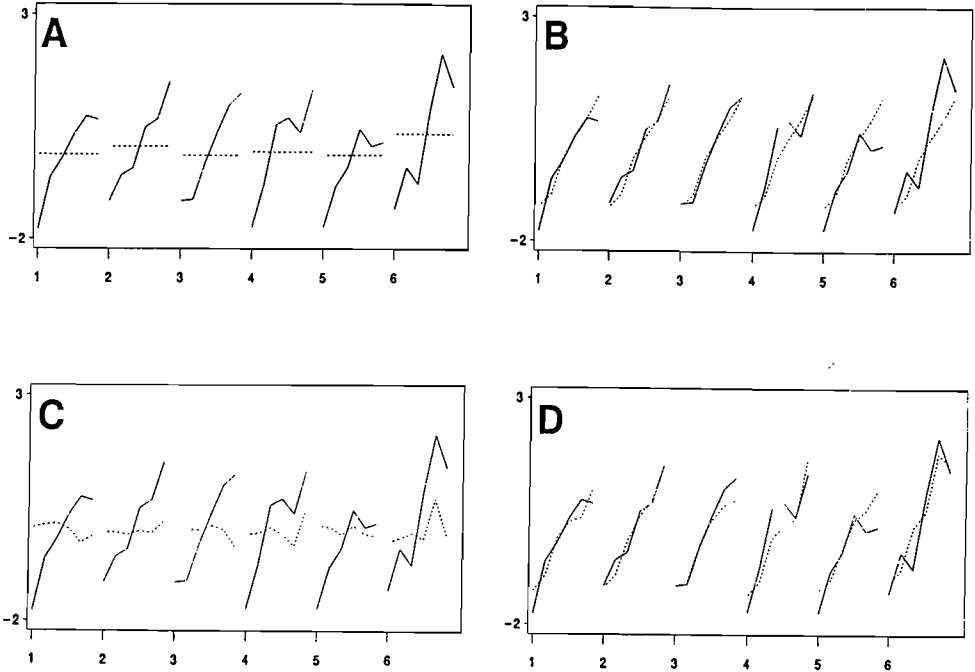


Figure 8 : Reconstitution des variations de CPUE (en log(kg/j.l.)) pour l'épèce *E. carbunculus* à partir des axes des ACPVI. A : Axe 1 de l'ACPVI selon l'île. B : Axe 1 de l'ACPVI selon la profondeur. C : Axes 1 et 2 de l'ACPVI selon l'interaction. D : Addition des 3 précédents modèles : Axe 1 (île) + Axe 1 (profondeur) + Axes 1 et 2 (île*profondeur). Sur chaque graphe, le gradient bathymétrique par île est représenté par un trait plein pour les valeurs observées et un pointillé pour la reconstitution. 1 : Santo ; 2 : Vanuatu Est ; 3 : Malekula ; 4 : Vanuatu Centre ; 5 : Efate ; 6 : Tanna.

Conclusion

En conclusion, nous discuterons des avantages respectifs des trois méthodes employées ici et de leur apport pour l'étude de la ressource dans le cadre de cette étude

L'AFC simple permet d'extraire la structure prédominante des données. Dans le cas présent, nous avons mis en évidence la structuration de la présence des espèces en fonction de la profondeur en identifiant ainsi des habitats dans la dimension verticale.

Dans le cas où un double niveau de structuration spatiale et/ou temporelle de la ressource (ici bathymétrie et géographie) permet de considérer le jeu de données comme un cube, une analyse multi-tableaux apporte une dimension dynamique à l'étude. Elle permet d'aborder trois questions différentes selon le point de vue dans lequel on se place par rapport au cube de données. Nous avons choisi ici de privilégier une étude de la structuration verticale moyenne des espèces et de sa reproductibilité dans l'espace géographique. L'image moyenne renvoyée par l'analyse du compromis est celle d'une structuration des espèces avec la profondeur. La projection sur ce compromis des situations propres à chaque île permet d'observer une déstructuration progressive du gradient bathymétrique du Nord-Ouest vers le Sud-Est de l'archipel où les espèces inféodées à la couche intermédiaire sont particulièrement abondantes, et également de mettre en évidence des localisations marquées de certaines espèces.

A la différence de l'AFC de Foucart, où le compromis est une moyenne simple des tableaux initiaux, l'Analyse Triadique Partielle, version ACP de ce type de méthode, utilise des

pondérations privilégiant la ressemblance entre tableaux. Dans le cas que nous avons traité ici, la différence est peu sensible, du fait de la forte ressemblance des îles entre elles. Dans d'autres cas, outre la question du choix entre ACP et AFC, cette différence devra entrer en ligne de compte dans le choix de la méthode. L'ATP sera particulièrement indiquée dans la situation d'un tableau « atypique » par rapport aux autres, dont l'influence sur la définition du compromis se trouvera ainsi limitée.

Les analyses sur variables instrumentales permettent, quant à elles, de décomposer la variabilité d'un tableau selon plusieurs facteurs explicatifs, qualitatifs ou quantitatifs, et selon leurs interactions, s'apparentant en cela aux régressions linéaires et analyses de variance. Ces méthodes permettent de quantifier la part de variation imputable aux différents facteurs explicatifs (dans notre étude la bathymétrie et la géographie). Ainsi nous avons mis en évidence la permanence des espèces de la strate profonde et de la strate superficielle, avec une faible variation géographique de leurs habitats. A l'inverse, l'abondance des espèces de la strate intermédiaire varie davantage selon l'île. Ce type d'analyse permet également de reconstituer les données à partir des axes des ACPVI, ce qui présente un intérêt évident de modélisation.

Un avantage de l'approche par variables instrumentales par rapport aux méthodes multi-tableaux est qu'elle permet de prendre en compte plus de deux variables explicatives à la fois. Son inconvénient majeur résulte de la régression linéaire sous-jacente : si les variables de milieu sont trop nombreuses en regard de la quantité d'observations, l'analyse n'a plus aucun sens. Ce n'est heureusement pas le cas ici.

En raison de contraintes techniques, nous avons présenté ici les résultats d'une ACP sur Variables Instrumentales. Une AFCVI, portant sur l'effet île, fournirait des résultats strictement équivalents à l'AFC de Foucart que nous avons réalisée avec un tableau par île. De même, une AFCVI sur la variable profondeur équivaldrait à une AFC de Foucart avec un tableau par profondeur. Il n'existe pas d'équivalent en AFCVI à l'analyse multi-tableaux qui choisirait le point de vue d'un tableau par espèce.

Les Analyses Factorielles Simples et de Foucart présentées ici ont été réalisées avec le logiciel ADE-4 (Thioulouse *et coll.*, 1997), en libre diffusion sur internet à l'adresse <http://pbil.univ-lyon1.fr/ADE-4/ADE-4.html>. La mise en oeuvre de l'ACPVI utilise le macro-programme *species.sas* développé sous macro-langage SAS par Vey (1998), et accessible par ftp anonyme sur <ftp.mpl.ird.fr> dans le répertoire /pub/acpvi.

Remerciements

Nous remercions Francis Laloë de l'IRD Montpellier pour ses conseils avisés dans la réalisation de ce travail.

Références bibliographiques

- BLANC, L., D. CHESSEL, et S. DOLEDEC (1998). Stabilité temporelle des structures spatiales. I - Analyses d'une série de tableaux de relevés faunistiques totalement appariés. *Bulletin Français de la Pêche et de la Pisciculture* : 348, 1-21.
- FOUCART, T. (1978) Sur les suites de tableaux de contingence indexés par le temps. *Statistique et Analyse des données* : 2, 67-84.
- FOUCART, T. (1983) Une nouvelle approche de la méthode STATIS. *Revue de Statistique Appliquée* : 31, 61-75.
- LAVIT, Ch., Y. ESCOUFIER, R. SABATIER, et P. TRAISSAC (1994) The ACT (Statist method). *Computational Statistics and Data Analysis* : 18, 97-119.
- LEBRETON, J.D., R. SABATIER, G. BANCO, et A.M. BACOU (1991) Principal component and correspondence analyses with respect to instrumental variables: an overview of their role in studies of structure-activity and species- environment relationships. In : *Applied Multivariate Analysis in SAR and Environmental Studies*. Devillers, J. & Karcher, W. (Eds.) Kluwer Academic Publishers. 85-114.

- LEBRETON, J.D. (1997) L'espace en dynamique des populations. Quelques considérations statistiques. *Tendances nouvelles en modélisation pour l'environnement. journées du Programme Environnement, Vie et Sociétés du CNRS.* (F. Blasco, coord.) CNRS. Elsevier ed. : 3-18.
- PECH, N. et F. LALOË (1997) Use of Principal Component Analysis with Instrumental Variables (PCAIV) to analyse fisheries catch data. *ICES Journal of Marine Science* : 54, 32-47.
- RAO, C.R. (1964) The use and interpretation of principal component analysis in applied research. *Sankhya, A* : 26, 329-359.
- SABATIER, R., J.D. LEBRETON, et D. CHESSEL (1989) Principal component analysis with instrumental variables as a tool for modelling composition data. In : *Multiway data analysis.* Coppi, R. & Bolasco, S. (Eds.) Elsevier Science Publishers B.V., North-Holland. 341-352.
- THIOULOUSE, J. et D. CHESSEL (1987) Les analyses multi-tableaux en écologie factorielle. I De la typologie d'état à la typologie de fonctionnement par l'analyse triadique. *Acta Ecologica, Ecologia Generalis* : 8, 4, 463-480.
- THIOULOUSE, J., D. CHESSEL, S. DOLEDEC, et J.M. OLIVIER (1997) ADE-4: a multivariate analysis and graphical display software. *Statistics and Computing* : 7, 75-83.
- VEY, P. (1998) SPECIES SAS, Macro SAS qui réalise une ACPVI. *Centre ORSTOM de Montpellier. Laboratoire halieutique et écosystèmes aquatiques (HEA)*, 1-33.

**MODELE DE MORTALITE EXPONENTIEL ETENDU
ET ESTIMATION SEMI-PARAMETRIQUE
APPLICATION A L'EVALUATION DE LA PRODUCTION
DES OEUF D'ANCHOIS DANS LE GOLFE DE GASCOGNE⁽¹⁾**

B. Truong-van

Laboratoire de Statistique et Probabilités UMR CNRS 5583

Université Paul Sabatier

31062 Toulouse cedex 4 France

⁽¹⁾La présente étude a été subventionnée par la Communauté Européenne Contrat 95/009

Abstract

The exponential model used in the Daily Egg Production Method for evaluating the biomass of the anchovies in the Gascogne gulf does not fit the data. Based on a deep analysis of the problem and the data, we propose an improvement of this model as well as new estimation methods (maximum of likelihood technique and semi-parametric approach). Numerical results are presented and discussed.

1- INTRODUCTION

Pour évaluer la biomasse des anchois dans le golfe de Gascogne par la méthode de production journalière des oeufs (DEPM) (cf. [1]), une étape primordiale est d'estimer à partir des densités (poids par unité de surface) des oeufs d'anchois prélevés lors des campagnes d'échantillonnage, la densité moyenne α des oeufs d'âge nul dans toute l'aire de ponte. Jusqu'à présent, l'estimation de α et du taux de mortalité z_0 s'obtient en ajustant une seule courbe exponentielle. Un simple examen visuel ainsi que les critères statistiques montrent l'inadéquation totale de ce modèle aux données. Une reformulation du problème nous conduit à proposer une extension consistant à remplacer l'amplitude constante α de l'exponentiel par une variable aléatoire A . Le problème consiste alors à estimer la densité de probabilité de A et le taux de mortalité z_0 . Nous introduisons à ce propos deux méthodes semi-paramétriques en même temps que nous considérons la technique du maximum de vraisemblance pour estimer à la fois la densité de probabilité de A et le paramètre taux de mortalité. A notre connaissance, les propriétés des estimateurs fondés sur ces méthodes ne sont pas connues, aussi nous présentons une étude de leurs propriétés asymptotiques qui permettent de cerner la précision des estimateurs proposés. Les résultats d'estimation obtenus sont bien plus satisfaisants que ceux donnés par le modèle exponentiel simple. Conjugés aux résultats de simulation, ils confirment la pertinence du modèle exponentiel étendu.

2- L'INADEQUATION DU MODELE DE MORTALITE EXPONENTIELLE

D'après le plan d'échantillonnage retenu, chaque station est visitée une seule fois, et un seul prélèvement d'oeufs est effectué à chaque station. Les données, représentées dans la Figure 1, sont alors des densités $y(t_x) = P_x(t_x)/s_x, x \in X$, où $P_x(t_x)$ est le poids des oeufs d'âge t_x collectés à la station x d'aire s_x (cf.[2]). Actuellement, on ajuste aux données $y(t_x)$ le modèle de mortalité exponentielle (EM) suivant :

$$y(t_x) = \alpha e^{-\alpha t_x} + \varepsilon(t_x), x \in X, \quad (1)$$

où les $\varepsilon(t_x)$ sont des termes d'erreur.

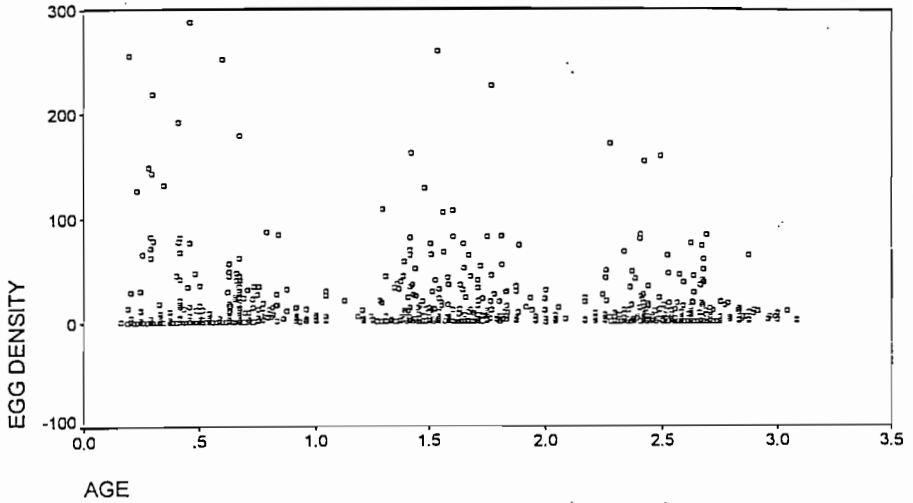


Figure 1 : Densité d'oeufs d'anchois en fonction de l'âge (en unité jour).

Le problème majeur est l'inadéquation évidente du modèle EM aux données, révélée par un simple examen visuel du nuage de points $y(t_x), x \in X$ (cf. Figure 1) et confirmée par tous les critères statistiques de l'ajustement.

De toute évidence, le modèle EM n'est applicable qu'aux situations où les données sont de la forme $y(t_l) =: \frac{P(t_l)}{s}, l = 1, \dots, n$, où les $P(t_l)$ sont les poids des oeufs de la même ponte dans une même surface s fixe (la même dans le temps), collectés à des instants successifs $t_l, l = 1, \dots, n$. Une reformulation du problème s'impose.

3- MODELE DE MORTALITE EXPONENTIELLE ETENDU

La situation expérimentale considérée ici correspond non pas à une seule équation telle que (1) mais en fait à une famille d'équations de la forme

$$y(t_{x,c}) = \alpha_{x,c} e^{-\alpha_{x,c} t_{x,c}} + \varepsilon(t_{x,c}); c \in C, x \in X_c,$$

où C est l'ensemble des niveaux de densité d'abondance i.e. des différents niveaux de production journalière des oeufs (d'âge nul) rapportée à des surfaces découpées a priori dans l'aire totale de ponte par le plan d'échantillonnage, et X_c est l'ensemble des stations correspondant à un niveau c .

Nous considérons que les valeurs de l'ensemble $\{\alpha_{x,c}; c \in C, x \in X_c\}$ sont celles d'une variable aléatoire A qui prend en compte des aléas à la fois nombreux et divers (pour les détails de la reformulation, cf. [3]).

On est alors conduit à l'extension suivante du modèle EM, que nous appellerons modèle de mortalité exponentielle étendu (EEM) :

$$Y(t_x) = A e^{-\alpha t_x} + \varepsilon(t_x), x \in X. \quad (2)$$

Le paramètre α qui représente la densité moyenne des oeufs d'âge nul dans toute l'aire de ponte n'est autre que l'espérance mathématique de A , autrement dit $\alpha = E(A)$.

Cette relation s'interprète simplement : Une estimation judicieuse de la densité moyenne α des oeufs d'âge nul est une moyenne pondérée et non point une simple moyenne arithmétique effectuée par un ajustement des densités observés au modèle EM, puisque la distribution de A n'est pas uniforme mais d'allure exponentielle, comme le montre la Figure 2. Les histogrammes de A que nous avons effectués pour différentes estimations du taux de mortalité z_0 sont tous similaires à la Figure 2 qui correspond à une estimation de z_0 égale à 0.1.

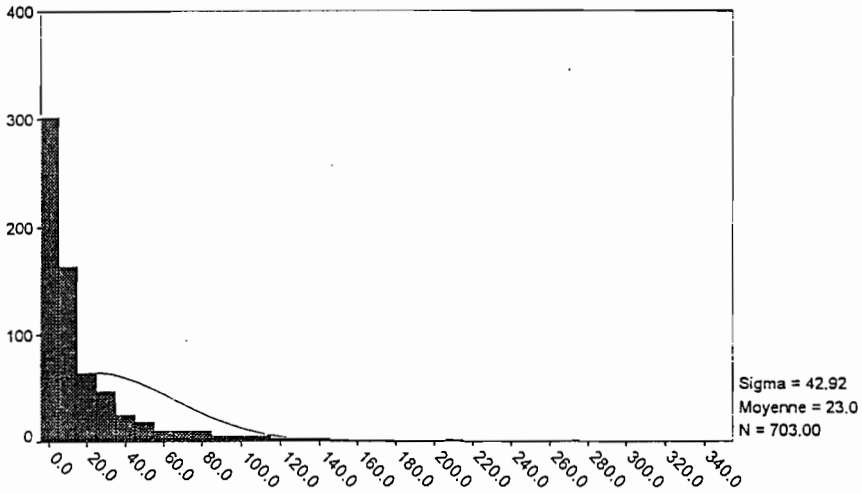


Figure 2 : Histogramme de A pour un taux de mortalité de 0.1.

Pour vérifier la pertinence du modèle EEM (2), nous avons aussi simulé le modèle

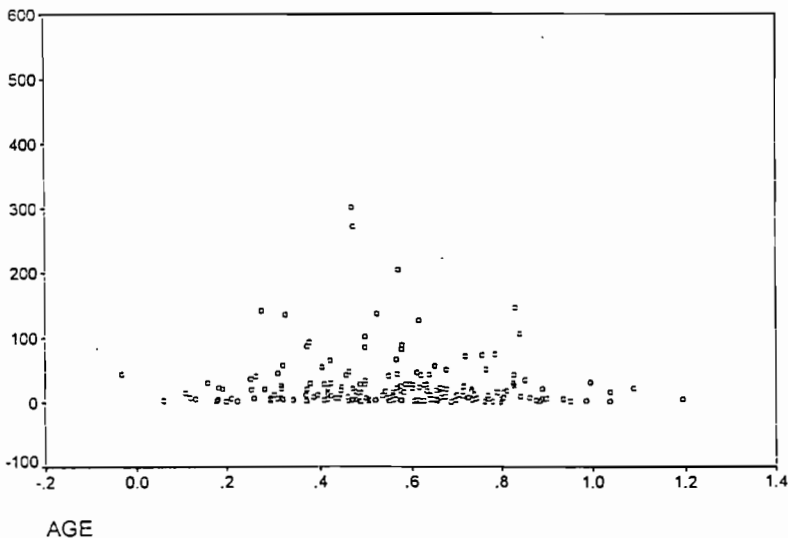


Figure 3 : Modèle simulé de densité d'oeufs pour un jour.

$Y(T) = Ae^{-z_0 T}$, où T est une variable gaussienne centrée sur l'âge de 0.6 jour et A une variable exponentielle, T et A sont indépendantes entre elles.

Nous reportons une simulation de ce modèle dans la Figure 3. On constate que le modèle $Y(T) = Ae^{-z_0 T}$ simule bien le comportement des densités observées autour d'un jour, les données de la Figure 1 étant sur 3 jours.

4- METHODES D'ESTIMATION

Le problème consiste à estimer α à partir des $y(t_x), x \in X$. Pour estimer $\alpha = E(A)$, on est conduit à estimer la densité de probabilité f_A de A . Nous proposons principalement trois méthodes que sont la régression semi-paramétrique (RSP), la minimisation de la variance de la densité d'abondance estimée (MVA) et le maximum de vraisemblance (MV). Les deux premières méthodes sont semi-paramétriques combinant l'estimation de la fonction f_A avec celle du paramètre z_0 , elles requièrent que l'ensemble $A(\Omega)$ des valeurs de A soit compact, condition remplie en pratique. De plus, il est raisonnable de supposer que A est indépendant des $\varepsilon(t_x)$ et que les $\varepsilon(t_x)$ sont indépendants, identiquement distribués, centrés et de variance $\sigma^2 > 0$.

4-1 Régression semi-paramétrique

La méthode générale est décrite dans [4], ici nous en présentons le cas particulier simple où la fonction f_A est estimée par des histogrammes et non par des noyaux. Supposons que $f_A = \sum_{j=1}^{J_0} p_{0j} I(C_{0j})$ où $p_{0j} = P(A \in C_{0j})$, les C_{0j} forment une partition de $A(\Omega)$ et $I(C)$ est l'indicatrice d'un ensemble C .

Comme $\hat{f}_{A,z} = \sum_{j=1}^{J_z} \hat{p}_{z,j} I(C_{z,j})$ est un estimateur-histogramme de f_A et que la distance quadratique est $Q_{G1}(z, \Pi) = \sum_{j=1}^J Q_{0j,z}(a_j)$, avec $Q_{0j,z}(a) = \sum_{x \in X_j} (y(t_x) - ae^{-z t_x})^2$, $\Pi = (C_j, a_j, j = 1, \dots, J), a_j \in C_j$ et $X_j = \{x \in X / y(t_x) \in C_j\}$, l'estimateur RSP se réduit alors à l'estimateur des moindres carrés sous contraintes suivant

$$(\hat{z}, \hat{\Pi}) = \arg \text{Min}_{(z, \Pi: z \geq \delta, n_j \geq \nu_j(n), j=1, \dots, J)} Q_{G1}(z, \Pi), \quad (3)$$

où n_j est l'effectif de la classe C_j , n est le cardinal de X , δ un réel positif et les $\nu_j(n)$ sont des entiers donnés tendant vers l'infini avec n .

L'algorithme pour résoudre (3) consiste :

(i) à partir d'une estimation initiale z de z_0 et d'une partition initiale $\Pi = (C_j, a_j, j = 1, \dots, J), a_j \in C_j$ du nuage de points des données, à effectuer pour chaque classe C_j , la régression simple :

$$y(t_x) = a_j e^{-z t_x} + \varepsilon(t_x), x \in X_j, \quad (4)$$

(ii) puis à itérer (i) avec une nouvelle estimation de z_0 et une partition plus fine du nuage des points respectant la contrainte sur les effectifs n_j des classes C_j , jusqu'à minimisation de $Q_{G1}(z, \Pi)$.

D'après [4], sous des conditions adéquates, le problème (3) a une solution et $n^{1/2}(\hat{z} - z_0)$ converge en loi vers la distribution normale $N(0, \nu)$ où $\nu = (2z_0)^3 (E(A^2))^{-2} \sigma^2 \sum_{j=1}^{J_0} a_{0j}^2$.

Pour chaque taux de mortalité estimé z , on a un estimateur $\hat{f}_{A,z}$ de f_A , d'où nous déduisons l'estimateur $\bar{A}(z) = \sum_{j=1}^{J_z} p_{0j} \hat{a}_j(z)$ de $E(A)$, avec $\hat{a}_j(z) = \arg \text{Min}_a Q_{0j,z}(a)$.

De ces résultats, on déduit (voir [4]) une estimation de la variance de z en remplaçant dans l'expression de ν , z_0 par z et a_{0j} par $\hat{a}_j(z)$, ainsi que l'estimation

$\widehat{Var}(\bar{A}(z)) = \frac{1}{2nz} \sum_{j=1}^{Jz} \frac{n_j}{n} MSE_j$ de $Var(\bar{A}(z))$, où MSE_j est la moyenne des carrés des résidus de la régression (4).

4-2 Minimisation de $Var(\bar{A}(z))$

Puisque l'objectif est d'obtenir un estimateur précis (i.e. efficace dans un certain sens) de $\alpha = E(A)$, il est naturel de considérer des estimateurs qui minimisent la variance de $\bar{A}(z)$. Nous définissons alors l'estimateur $\hat{z} = ArgMin_{z \geq 0} Var(\bar{A}(z))$, puis estimons f_A par $\hat{f}_{A, \hat{z}}$.

D'après [4], sous des conditions adéquates, pour z proche de z_0 ,

$\widehat{Var}(\bar{A}(z)) \simeq \frac{1}{n^2(2z)} \sum_{j=1}^{Jz} Q_{0jz}(a_j)$ de sorte que le critère MVA ne diffère de RSP que par le facteur $(2zn^2)^{-1}$.

4-3 Maximum de vraisemblance avec A exponentiel

Nous supposons ici que les termes $\varepsilon(t_x)$ dans (2) sont nuls et que A est exponentiel de paramètre $\theta_0 > 0$. D'après [4], les estimateurs MV $(\hat{z}, \hat{\theta})$ de (z_0, θ_0) sont définis par les équations de maximum de vraisemblance suivantes

$$\begin{cases} \frac{N_1}{\hat{\theta}} = \sum_{x \in \mathcal{V}} t_x y(t_x) e^{-z t_x} \\ \frac{N_0}{\hat{\theta}} = \sum_{x \in \mathcal{V}} y(t_x) e^{-z t_x} \end{cases}$$

De plus, ils sont asymptotiquement Normal $N(0, V)$, où

$$V = \frac{1}{N_0 N_2 - N_1^2} \begin{bmatrix} N_0 & -N_1 \theta_0 \\ -N_1 \theta_0 & \theta_0^2 N_2 \end{bmatrix}, \text{ et } N_j = \sum_i t_i^j.$$

Par ailleurs, on a $\bar{A} = \frac{1}{\hat{\theta}} = \frac{\sum_{x \in \mathcal{V}} y(t_x) e^{-z t_x}}{N_0}$ avec $Var(\bar{A}) = Var(\frac{1}{\hat{\theta}}) = \frac{Var(\hat{\theta})}{\theta_0^4}$. On en déduit une estimation de la variance de z et de \bar{A} , en remplaçant θ_0 par $\hat{\theta}$ dans l'expression de V , puis dans celle de $Var(\bar{A})$.

5- RESULTATS NUMERIQUES ET DISCUSSION

Les données traitées ici sont celles de la campagne d'échantillonnage de 1994 dans le golfe de Gascogne. Elles contiennent 1164 valeurs de densités d'oeufs d'anchois. La méthode actuellement utilisée consiste à effectuer une simple régression sur l'ensemble de ces données via l'équation de mortalité exponentielle (1). Les résultats obtenus avec cette méthode sont reportés à la ligne EMR1 du tableau ci-dessous.

Un problème préliminaire soulevé par l'application de la méthode actuelle provient de la présence de 461 valeurs de densités d'oeufs nulles, soit une proportion de 39.6%, dans les données. Ces valeurs nulles correspondent ou bien à des mesures nulles qui sont des prélèvements nuls à certaines stations d'échantillonnage, ou bien à des "données estimées nulles" qui sont des valeurs zéro ajoutées par les biologistes à l'étape de datation des oeufs d'anchois, la proportion entre les mesures nulles et les "données estimées nulles" étant imprécise voire inconnue. Ces évaluations de biologistes même experts, étant subjectives, soulèvent la question de l'objectivité et de l'interprétation des résultats statistiques obtenus avec la méthode actuelle. De plus, les valeurs nulles de densités d'oeufs ne peuvent être pris en compte par le modèle de mortalité exponentielle (1), puisque $0 = \exp(-\infty)$, c'est à dire, une densité d'oeufs nulle ne peut correspondre qu'à un âge infini. Pour résoudre ce problème, nous sommes amenés à stratifier simplement les données en 2 sous-populations P_0 et P_1 , P_0 étant celle des données nulles.

Nous avons ensuite appliqué à P_1 une régression simple utilisant le modèle (1), la méthode de régression semi-paramétrique et la procédure de maximum de la vraisemblance. Les résultats obtenus avec chacune de ces méthodes sont reportés respectivement dans les lignes EMR2, RSP et MV du tableau ci-dessous. Les résultats obtenus avec la méthode de minimisation de $\widehat{Var}(\bar{A}(z))$ étant similaires à ceux donnés par RSP ne sont pas reportés ici.

Méthode	\hat{z}	$\hat{\sigma}(\hat{z})$	$CV(\hat{z})$	\bar{A}	$\hat{\sigma}(\bar{A})$	$CV(\bar{A})$
EMR1	0.26	0.09	0.35	15.26	2.08	0.14
EMR2	0.25	0.09	0.36	25.00	3.28	0.13
SPR	0.2	0.037	0.18	23.07	0.34	0.01
MV	0.25	0.04	0.23	24.86	1.98	0.08

Les résultats du tableau ci-dessus appellent les commentaires suivants :

1) En appliquant la formule de stratification simple aux résultats (EMR2) de la régression simple avec le modèle de mortalité exponentielle (1) sur les données non-nulles, on retrouve bien l'estimation de A obtenue par la même méthode appliquée à l'ensemble de toutes les données (EMR1) puisque

$$0 \times \frac{461}{1164} + 25.00 \times \frac{703}{1164} = 15.10.$$

Autrement dit les 2 méthodes EMR1 et EMR2 conduisent statistiquement aux mêmes résultats. On peut noter au passage que ces résultats corroborent le fait que les densités d'oeufs nulles ne peuvent être prises en compte par le modèle de mortalité exponentielle (1). D'autres détails sont donnés dans [5].

2) Les résultats du tableau montrent clairement la meilleure performance des nouveaux estimateurs fournis par les procédures RSP et MV.

3) La meilleure performance des estimateurs RSP par rapport aux estimateurs MV semble d'être due aux hypothèses simplificatrices retenues pour établir les formules asymptotiques. Notre travail en cours comble cette lacune.

En conclusion, les résultats ci-dessus conjugués avec les études en simulation montrent la pertinence du modèle de mortalité exponentielle étendu.

Références

- [1] R. Lasker (1985) An Egg Production method for estimating spawning biomass of pelagic fish: NOAA Technical Report NMFS 36.
- [2] L. Motos, A. Uriarte, Metzuzals & P. Prouzet (1994) L'évaluation de la biomasse d'anchois dans le golf de Gascogne en 1994.
- [3] B. Truong-van (1997a) Some improvements on the mortality model and estimation techniques for the determination of the anchovy biomass in the Biscay-Gascogne bay. Preprint 1997 Laboratoire de Mathématiques Appliquées UPRES A 3055 CNRS Université de Pau France.
- [4] B. Truong-van (1997b) An extended mortality model for the Biscay bay anchovy data and a new estimation approach. Preprint 1997 Laboratoire de Mathématiques Appliquées UPRES A 3055 CNRS Université de Pau France.
- [5] B. Truong-van (1997c) Statistical Analysis of the Daily Egg Production Biomass Equation. Technical Report to the European Commission/ Study contract 95/009 October 23 1998.

ESTIMATION DE LA BIOMASSE DES ANCHOIS DANS LE GOLFE DE GASCOGNE UNE ETUDE STATISTIQUE DES ESTIMATEURS

B. Truong-van⁽¹⁾ & J.C. Turlot⁽²⁾

⁽¹⁾Laboratoire de Statistique et Probabilités UMR C55830 CNRS
Université Paul Sabatier 31062 Toulouse France

⁽²⁾IUT STID Université de Pau et des Pays de l'Adour
64000 Pau France

Cette étude est subventionnée par la Communauté Européenne Contract 95/009

Résumé

Dans la méthode dite de production journalière des oeufs, on exprime la biomasse d'une espèce de pélagiques comme une fonction rationnelle d'autres paramètres de la population. L'estimateur naturel \hat{B} de la biomasse dérivé directement de cette expression est biaisé. Pour la population des anchois dans le golfe de Gascogne-Biscay, nous présentons ici une étude statistique de la méthode actuellement utilisée pour la correction de biais de l'estimateur \hat{B} , en utilisant les données de la campagne d'échantillonnage de 1994. Nous donnons aussi une comparaison avec des estimateurs de la biomasse par des techniques de Jackknife.

Abstract

In the Daily Egg Production Method, the biomass of a pelagic population is expressed as a rational function of several population parameters. The natural estimate \hat{B} of the biomass based on this expression of the biomass is biased. For the anchovies in the Biscay-Gascogne gulf, we give a statistical analysis of the procedure currently used for correcting the bias of \hat{B} , when using the data of the 1994 sampling campaign. Then, alternative estimators, mainly those given by Jackknife techniques, are considered..

1 - INTRODUCTION

L'estimation de la biomasse B des anchois dans le golfe de Gascogne est réalisée avec la méthode de production journalière des oeufs. Elle est fondée sur la formule suivante (cf. Parker [5] ou Picquelle & Stauffer [6])

$$B = \frac{PW}{RFS} \quad (1)$$

où $P = cAP_0$, P_0 est la production journalière des oeufs (poids des oeufs d'âge nul produits par jour et par unité de surface), A l'aire totale de pont, W le poids moyen de femelle mature, R la fraction de population de femelles matures, F la fécondité par lot, S la fraction de femelles qui pondent par jour, c une constante de conversion d'unités.

Désignons par $\hat{\theta}$ un estimateur d'un paramètre $\theta \in \{B, P, R, S, F, W\}$, alors l'estimateur naturel retenu pour B est

$$\hat{B} = \frac{\hat{P} \hat{W}}{\hat{F} \hat{R} \hat{S}}. \quad (2)$$

Il est bien connu (cf. Cochran [1]) que cet estimateur est biaisé. A l'instar de l'approche habituelle en théorie de sondage, Parker [5] s'inspirant de Seber [7] (voir aussi les références

dans Picquelle & Stauffer [6]) a proposé la correction approchée (au second ordre) suivante b_2 du biais $b = E(\hat{B}) - B$, ainsi que l'expression de la variance de l'estimateur \hat{B} après correction du biais de \hat{B} :

$$b_2 = \hat{B}\{cv(\hat{R})^2 + ccv(\hat{F})^2 + ccv(\hat{S})^2 + COVS\}, \quad (3)$$

$$Var(\hat{B}) \simeq B^2\{cv^2(\hat{P}) + cv^2(\hat{W}) + cv^2(\hat{F}) + cv^2(\hat{S}) + COVS\}, \quad (4)$$

avec

$$COVS = ccv(P, W) + ccv(R, F) + ccv(R, S) + ccv(S, F) \\ - ccv(P, R) - ccv(P, F) - ccv(P, S) - ccv(W, R) - ccv(W, F) - cccv(W, S)$$

où $ccv(U, V) = \frac{Cov(U, V)}{E(U)E(V)}$ est le coefficient de variation croisée de deux variables aléatoires U et V , et où $cv(U)$ est le coefficient de variation habituelle de U , i.e. $cv(U) = \sqrt{ccv(U, U)}$.

Notons au passage que la formule originale de Parker [5] contient des incorrections (confusions entre les paramètres de la population B, P, R, S, F, W et leurs estimateurs $\hat{B}, \hat{P}, \hat{W}, \hat{F}, \hat{R}, \hat{S}$).

Les deux relations (3) et (4) sont actuellement utilisées pour estimer la biomasse B des anchois dans le golfe de Gascogne. Le but ultime à moyen terme de notre étude est de proposer des estimateurs de B qui réalisent le meilleur compromis entre le biais minimal et la variance minimale. La première tâche que nous avons menée et que nous présentons ici est une analyse statistique fine de la méthode actuelle de correction de biais et une première comparaison avec d'autres estimateurs, principalement ceux déduits des techniques de Jackknife et ce, en utilisant les données de la campagne d'échantillonnage de 1994 (données qui nous étaient communiquées par L. Motos de l'Instituto de Alimentario y Pesquero, ASTI, San Sebastain, Espagne). Pour cela, après avoir reformulé de façon adéquate les relations de correction de biais pour \hat{B} , nous les étudions en nous plaçant dans un cadre simplifié mais réaliste et en utilisant une approche paramétrique car une étude de ces relations dans sa généralité s'avère difficilement surmontable.

Nous montrons maintenant qu'au lieu de (3) et (4), la formule de correction de biais de \hat{B} peut être à la fois généralisée et simplifiée pour les données considérées :

(i) En effet, puisque P est estimé à partir des échantillons indépendants de ceux utilisés pour les autres paramètres R, S, F, W , nous réécrivons les équations (1) et (2) sous la forme :

$$B = PZ,$$

$$\hat{B} = \hat{P} \hat{Z}$$

où

$$Z = \frac{W}{RFS}$$

$$\hat{Z} = \frac{\hat{W}}{\hat{F} \hat{R} \hat{S}}.$$

De toute évidence, on a

$$E(\hat{B}) = E(\hat{P})E(\hat{Z}),$$

$$Var(\widehat{B}) = Var(\widehat{P})var(\widehat{Z}) + E^2(\widehat{P})Var(\widehat{Z}) + Var(\widehat{P})E^2(\widehat{Z}),$$

et $E(\widehat{P}) = cAE(\widehat{P}_0)$, $Var(\widehat{P}) = (cA)^2 Var(\widehat{P}_0)$, car A est supposé non-aléatoire.

L'étude de $E(\widehat{P})$, $Var(\widehat{P})$ étant présentée dans une autre communication (voir aussi Truong-van [8], [9]), nous nous restreignons ici à l'estimation de $E(\widehat{Z})$ et de $E(\widehat{Z}^2)$.

(ii) Posons $X = FRS$, $\widehat{X} = \widehat{F}\widehat{R}\widehat{S}$; $\widehat{f} = \frac{\widehat{F}-F}{F}$, $\widehat{r} = \frac{\widehat{R}-R}{R}$, $\widehat{s} = \frac{\widehat{S}-S}{S}$, $\widehat{w} = \frac{\widehat{W}-W}{W}$. Alors $\widehat{Z} - Z = \frac{\widehat{W}}{\widehat{X}} - \frac{W}{X} = Z(g(\widehat{w}, \widehat{f}, \widehat{r}, \widehat{s}) - 1)$ où $g(\widehat{w}, \widehat{f}, \widehat{r}, \widehat{s}) = \frac{1+\widehat{w}}{(1+\widehat{f})(1+\widehat{r})(1+\widehat{s})}$. Si pour $\widehat{\theta} \in \{\widehat{F}, \widehat{R}, \widehat{S}, \widehat{W}\}$, $E(\widehat{\theta} - \theta)^k \ll \theta^k$, $k \geq 3$ alors au second ordre en probabilité,

$$\widehat{Z} - Z \simeq Z\{\widehat{w} - (\widehat{f} + \widehat{r} + \widehat{s}) + (\widehat{f}^2 + \widehat{r}^2 + \widehat{s}^2) - \widehat{w}(\widehat{f} + \widehat{r} + \widehat{s}) + \widehat{f}\widehat{r} + \widehat{f}\widehat{s} + \widehat{r}\widehat{s}\}$$

$$E(\widehat{Z} - Z) \simeq Z(\beta + \eta)$$

où $\beta = \beta(\widehat{W}) - \beta(\widehat{F}) - \beta(\widehat{R}) - \beta(\widehat{S})$ et

$$\eta = \eta(\widehat{F}, \widehat{F}) + \eta(\widehat{R}, \widehat{R}) + \eta(\widehat{S}, \widehat{S}) - \eta(\widehat{W}, \widehat{F}) - \eta(\widehat{W}, \widehat{R}) - \eta(\widehat{W}, \widehat{S}) \\ + \eta(\widehat{F}, \widehat{R}) + \eta(\widehat{F}, \widehat{S}) + \eta(\widehat{R}, \widehat{S}).$$

avec $\beta(\widehat{\theta}) = \frac{E(\widehat{\theta}) - \theta}{\theta}$, $\eta(\widehat{\theta}, \widehat{\theta}') = \frac{E(\widehat{\theta} - \theta)(\widehat{\theta}' - \theta')}{\theta\theta'}$, $\theta, \theta' \in \{F, R, S, W\}$.

(iii) Les estimateurs "après correction de biais au second ordre" pour Z et B sont

$$\widehat{Z}_a = \widehat{Z}(1 + \widehat{\beta} + \widehat{\eta}), \quad (5)$$

$$\widehat{B}_a = \widehat{P}\widehat{Z}_a. \quad (6)$$

où $\widehat{\beta}$ et $\widehat{\eta}$ sont respectivement les estimateurs de β et η .

Notons que si $b(\widehat{Z})$ est le biais exact de \widehat{Z} et $\widehat{b}(\widehat{Z})$ un de ses estimateurs, alors au lieu de l'estimateur approché \widehat{Z}_a , on utilisera

$$\widehat{\widehat{Z}} = \widehat{Z} + \widehat{b}(\widehat{Z}), \quad (7)$$

$$\widehat{\widehat{B}} = \widehat{P}\widehat{\widehat{Z}}. \quad (8)$$

Les estimateurs $\widehat{F}, \widehat{R}, \widehat{S}, \widehat{W}$ que nous retenons ici sont ceux proposés par Picquelle & Stauffer [6]. Ils sont de la forme

$$\widehat{\theta} = \sum_{i=1}^n p_i \widehat{\theta}_i \text{ pour } \widehat{\theta} \in \{\widehat{F}, \widehat{S}, \widehat{W}, \widehat{R}\} \text{ et}$$

$$\widehat{R} = \sum_{i=1}^n \frac{m_i}{M} \widehat{R}_i,$$

où les $p_i = p_i(\widehat{\theta})$ sont des poids proportionnels au nombre de poissons adultes à chacune des n prises, $\widehat{F}_i, \widehat{S}_i, \widehat{W}_i, \widehat{R}_i$ sont respectivement (resp.) les estimateurs de F, S, W, R fondés sur le i ème échantillon, k_i est le nombre de femelles, m_i le poids (des mâles et femelles) du i ème échantillon, $K = \sum_{i=1}^n k_i$ and $M = \sum_{i=1}^n m_i$. Pour les données de la campagne 1994 (voir Tableau 1), $n = 27$

N°	R	W	F	S ₀	S ₁	S ₂	S	Wm
1	0.4279	29.6	15162	0.44	0.125	0.4375	0.2812	26.09
2	0.6533	38.39	20151	0	0.8	0	0.4	34.22
3	0.1765	31.67	16377	0.33	0	0.3	0.15	26.16
4	0.2196	24.32	11498	0.61	0.3636	0.1818	0.2727	19.2
5	0.464	29.41	14767	0.24	0.1364	0.2273	0.1818	21.75
6	0.3531	26.87	12962	0.52	0.2	0.0667	0.1333	21.22
7	0.6307	43.56	21737	0.54	0.0556	0.5	0.2778	36.3
8	0.4449	24.58	12051	0.4348	0.2121	0.2121	0.2121	22.04
9	0.6709	30.1	15490	0.7037	0.2727	0	0.1364	23.09
10	0.2551	35.62	18571	0.8261	0	0.25	0.125	31.06
11	0.6383	37.49	19191	0.7931	0.25	0.50	0.375	31.87
12	0.2976	23.18	11680	0.08	0.2333	0.3	0.2667	20.19
13	0.6915	25.84	13625	0.44	0.0667	0.2	0.1333	17.62
14	0.6713	22.81	11376	0.1923	0.2222	0.3333	0.2778	19.05
15	0.4602	25.15	13047	0.3462	0.3462	0.1923	0.2692	23.22
16	0.4754	29.27	14882	0.25	0.2222	0.2222	0.2222	22.47
17	0.5401	35.45	18462	0.12	0.4359	0.1026	0.2692	25.98
18	0.3374	24.59	12342	0.4615	0	0.5714	0.2857	21.13
19	0.4945	19.16	9529	0.25	0	0.6111	0.3056	16.97
20	0.6463	17.88	8872	0.04	0.1111	0.3704	0.2407	14.68
21	0.5786	15.58	7475	0.12	0.0833	0.2083	0.1458	13.99
22	0.4463	18.38	9049	0.04	0.04	0.16	0.1	18.36
23	0.4806	28.99	15083	0.2963	0.2963	0.1111	0.2037	22.78
24	0.6103	16.51	8094	0.2593	0.1304	0.3913	0.2609	13.45
25	0.4852	24.35	12160	0.25	0.16	0.48	0.32	18.55
26	0.5854	18.08	8927	0	0.1481	0.3704	0.2593	15.09
27	0.6473	23.86	11571	0.5357	0.1875	0.1875	0.1875	19.16

Tableau 1. Données de la campagne d'échantillonnage de 1994, où N° est le numéro de l'échantillon, et où pour chaque échantillon, R est la fraction de population de femelles matures, W est le poids moyen de femelle mature, F est la fécondité par lot, S₀ (resp. S₁, S₂) est la fraction de femelles qui pondent le jour 0 (resp. le jour 1, le jour 2) et S est une moyenne des S₀, S₁, S₂, Wm est le poids moyen des mâles.

et les estimations $\hat{F}_i, \hat{S}_i, \hat{W}_i$ et \hat{R}_i sont reportés dans le Tableau 2 ci-dessous.

i	\hat{F}_i	\hat{W}_i	\hat{S}_i	\hat{R}_i
1	496.83	0.9699	0.00921	0.0181
2	956.32	1.8219	0.01898	0.0360
3	314.58	0.6083	0.00288	0.0077
4	233.86	0.4946	0.00555	0.0072
5	533.95	1.0634	0.00657	0.0180
6	366.16	0.7590	0.00376	0.0129
7	908.78	1.8211	0.01161	0.0382
8	626.38	1.2776	0.01102	0.0157
9	1067.67	2.0747	0.00940	0.0271
10	482.64	0.9257	0.00325	0.0129
11	780.65	1.5250	0.0152	0.0336
12	409.13	0.8119	0.0093	0.0098
13	846.75	1.6059	0.0083	0.0228
14	372.77	0.7474	0.0091	0.0213
15	545.47	1.0515	0.0112	0.0169
16	538.11	1.0583	0.0080	0.0187
17	771.86	1.4821	0.0112	0.0252
18	195.24	0.3890	0.0045	0.0117
19	279.95	0.5629	0.0090	0.0136
20	300.74	0.6061	0.0081	0.0160
21	312.51	0.6514	0.0061	0.0130
22	337.42	0.6853	0.0037	0.0124
23	545.37	1.0482	0.0074	0.0189
24	338.39	0.6902	0.0109	0.0139
25	384.72	0.7704	0.0101	0.0158
26	332.87	0.6742	0.0097	0.0147
27	366.09	0.7549	0.0059	0.0211

Tableau 2. Estimations $\hat{F}_i, \hat{W}_i, \hat{S}_i$ et \hat{R}_i pour chaque échantillon i .

Pour estimer la variance des estimateurs $\hat{F}, \hat{S}, \hat{W}$ et \hat{R} , Picquelle & Stauffer [6] ont proposé :

$$\widehat{Var}(\hat{\theta}) = \frac{\sum_{i=1}^n k_i^2 (\hat{\theta}_i - \hat{\theta})^2}{(K/n)^2 n(n-1)}; \hat{\theta} \in \{\hat{S}, \hat{W}\}$$

$$\widehat{Var}(\hat{R}) = \frac{\sum_{i=1}^n m_i^2 (\hat{R}_i - \hat{R})^2}{(M/n)^2 n(n-1)},$$

$$\widehat{Var}(\widehat{F}) = \frac{\sum_{i=1}^n m_i^2 \left\{ \frac{(\widehat{F}_i - \widehat{F})^2}{n-1} + \frac{s_h^2}{N_h} + \widehat{Var}(\widehat{\beta}) (\widehat{W}_i^* - \widehat{W}_{(h)}^*) \right\}}{(M/n)^2 n}$$

où \widehat{W}_i^* est le poids moyen de femelles sans ovaires de l'échantillon i , $\widehat{W}_{(h)}^*$ le poids moyen de femelles hydratées sans ovaires, N_h le nombre de femelles hydratées, $\widehat{\beta}$ et s_h^2 sont resp. la pente estimée et la variance des résidus de la régression suivante :

$$F_{i,j} = \alpha + \beta W_{i,j}^* + \varepsilon_{i,j},$$

$F_{i,j}$ (resp. $W_{i,j}^*$) est le nombre d'oeufs par lot (resp. le poids sans ovaire) de la femelle mature j dans l'échantillon i .

D'après (8), l'étude statistique de \widehat{B} se ramène donc à celle de \widehat{Z}_a ou de \widehat{Z} . Il est clair que ces estimateurs restent biaisés. La première question est de savoir si le biais de ces nouveaux estimateurs est suffisamment petit pour être considéré comme négligeable. Au vu de (5), il apparaît difficilement concevable d'aborder cette question dans sa généralité, car $E(\widehat{Z})$ dépend des espérances mathématiques des estimateurs des covariances des $\widehat{F}, \widehat{R}, \widehat{S}, \widehat{W}$ pris deux à deux et la variance $Var(\widehat{Z})$ requiert des moments des $\widehat{F}, \widehat{R}, \widehat{S}, \widehat{W}$ supérieurs à deux. Aussi, nous considérons proposons dans le paragraphe suivant un cadre plus simplifié mais réaliste.

2- ETUDE STATISTIQUE DANS UN CADRE SIMPLIFIE

Nous supposons à l'instar de Parker [5] que les estimateurs de F, R, S, W sont sans biais. Le cadre simplifié que nous considérons est le suivant :

Nous exprimons d'abord l'estimateur \widehat{Z} sous la forme du ratio $\frac{\widehat{W}}{\widehat{X}}$ où $\widehat{X} = \widehat{F}\widehat{R}\widehat{S}$. Les régressions linéaires $\widetilde{W}_i = a + b\widetilde{X}_i + U_i$ que nous avons effectuées sur les données de 1994 donnent les résultats suivants :

R2 corrigé = 0.61; $\widehat{a} = -32.05, \widehat{\sigma}(\widehat{a}) = 16.92; \widehat{b} = 96.76, \widehat{\sigma}(\widehat{b}) = 15.51$ où \widehat{a} et \widehat{b} sont respectivement (resp.) les estimations de a et de b , et où $\widehat{\sigma}(\widehat{a})$ et $\widehat{\sigma}(\widehat{b})$ sont leur écart-type empirique respectif.

Ces résultats nous conduisent à adopter l'hypothèse suivante de modèle (H1) : $\widehat{X} = a + b\widehat{W} + U$ avec $E(U/\widehat{W}) = 0, Var(U) = \epsilon = O(1/n)$.

Sous l'hypothèse (H1), une formule exacte pour la correction du biais de \widehat{Z} est :

$$b(\widehat{Z}) = Z \left\{ \sum_{k=2}^q (-1)^k \{ (bZ)^{k-1} (bZ - 1) E(\widehat{w}^k) + \frac{E(U^k)}{X^*} \} + (-1)^q E\left\{ \widehat{x}^q \left(\widehat{w} - \frac{\widehat{x}(1+\widehat{w})}{1+\widehat{x}} \right) \right\} \right\} \tag{9}$$

en posant $\widehat{x} = \frac{\widehat{X}-X}{X}, \widehat{w} = \frac{\widehat{W}-W}{W}$.

Pour les détails de calcul, nous référons le lecteur à l'Annexe 1.

Il s'ensuit de (9) que pour évaluer $b(\widehat{Z})$ et donc pour contrôler la correction approchée au second ordre du biais, il nous faut déterminer tous les moments de \widehat{w} . Il est donc naturel d'adopter une hypothèse de distribution paramétrique de loi de probabilité pour \widehat{w} . Une inspection des distributions empiriques des $\widehat{F}_i, \widehat{R}_i, \widehat{S}_i, \widehat{W}_i$ justifie l'hypothèse (H2) de loi Gamma $G(\alpha, \lambda)$ pour $\widetilde{W}_i = p_i \widehat{W}_i, i = 1, \dots, n$. Pour les données considérées, l'estimation des paramètres par le maximum de vraisemblance donne

$$\hat{\lambda} = 17, \hat{\alpha} = 17.0443.$$

Sous les hypothèses (H1) et (H2), la correction du biais approché pour \hat{Z} est donnée par

$$b_2(\hat{Z}) = Z \left\{ \frac{\text{Var}(\hat{X})}{X^2} - \frac{\text{Cov}(\hat{W}, \hat{X})}{W X} \right\} \quad (10)$$

$$= bZ^2(bZ - 1)E(\hat{w}^2) + \frac{E(U^2)}{X^2},$$

avec $E(\hat{w}^k) = \frac{E(\hat{W}-W)^k}{W^k}, k = 0, 1, \dots$

L'expression (10) fait appel à des résultats sur la loi Gamma qui sont rappelés en Annexe 2.

Les résultats obtenus alors sur la correction exacte et approchée du biais $E(\hat{Z})$ sont reportés dans le Tableau 3. Ils montrent que pour les données considérées et sous les hypothèses (H1) et (H2), l'approximation $b_2(\hat{Z})$ est adéquate.

Ordre k de la correction	Correction de biais à l'ordre k	Correction cumulée
2	2.5094E-05	2.5094E-05
3	-1.6371E-07	2.05776E-05
4	3.6925E-07	2.51146E-05

Tableau 3. Correction de biais pour Z

On peut donc utiliser la formule (10) pour dériver un estimateur $\hat{b}_2(\hat{Z})$ de $b_2(\hat{Z})$, en remplaçant les quantités théoriques par leurs estimateurs respectifs :

$$\hat{b}_2(\hat{Z}) = \hat{Z} \left\{ \frac{\widehat{\text{Var}}(\hat{X})}{\hat{X}^2} - \frac{\widehat{\text{Cov}}(\hat{W}, \hat{X})}{\hat{W} \hat{X}} \right\}. \quad (11)$$

La question est maintenant de savoir si le nouvel estimateur $\hat{\hat{Z}} = \hat{Z} + \hat{b}_2(\hat{Z})$ a un biais suffisamment petit pour être négligeable et de déterminer sa variance. C'est l'objet du paragraphe ci-dessous.

Propriétés statistiques de $\hat{b}_2(\hat{Z})$:

Sous la forme (11), les propriétés de $\hat{b}_2(\hat{Z})$ requièrent des conditions de moments des estimateurs $\widehat{\text{Var}}(\hat{X}), \widehat{\text{Cov}}(\hat{W}, \hat{X})$. Pour surmonter cette difficulté, nous utilisons de nouveau l'hypothèse (H2) de loi $G(\alpha, \lambda)$ pour les $\hat{W}_i, i = 1, \dots, n$ permettant d'écrire $\hat{b}_2(\hat{Z})$ sous la forme $\hat{b}_2(\hat{Z}) = \frac{\hat{b}Z^2(\hat{b}Z-1)}{n\hat{\lambda}} + \frac{\sigma_U^2}{\hat{X}^2}$, où $\hat{b}, \hat{\sigma}_U^2$ estiment respectivement b et $\sigma_U^2 = \text{Var}(U)$.

On pose $g_1(b, Z, \lambda) = \frac{bZ^2(bZ-1)}{n\lambda}, g_2(\sigma_U^2, X) = \frac{\sigma_U^2}{X^2}$;

$\hat{b} = b + \delta\hat{b}, \hat{Z} = Z + \delta\hat{Z}, \hat{\lambda} = \lambda + \delta\hat{\lambda}, \hat{\sigma}_U^2 = \sigma_U^2 + \delta\hat{\sigma}_U^2, \hat{X} = X + \delta\hat{X}$,

et par un développement de Taylor de $g_1(\hat{b}, \hat{Z}, \hat{\lambda})$ (resp. de $g_2(\hat{\sigma}_U^2, \hat{X})$) au second ordre au voisinage de (b, Z, λ) (resp. de (σ_U^2, X)), on obtient en négligeant certains termes de covariance

$$E(\widehat{b}_2(\widehat{Z})) \simeq b_2(\widehat{Z}) + c_b \text{Var}(\widehat{b}) + c_Z \text{Var}(\widehat{Z}) + c_\lambda \text{Var}(\widehat{\lambda}) \\ + c_{bZ} \text{Cov}(\widehat{b}, \widehat{Z}) + c_{b\lambda} \text{Cov}(\widehat{b}, \widehat{\lambda}) + c_{Z\lambda} \text{Cov}(\widehat{\lambda}, \widehat{Z}) + \frac{\sigma_{\widehat{W}}^2}{X^2} b^2 \frac{\text{Var}(\widehat{W})}{X^2},$$

$$\text{Var}(\widehat{b}_2(\widehat{Z})) \simeq d_b \text{Var}(\widehat{b}) + d_Z \text{Var}(\widehat{Z}) + d_\lambda \text{Var}(\widehat{\lambda}) + \frac{\sigma_{\widehat{W}}^2}{X^4} (1 + 4b^2 \frac{\text{Var}(\widehat{W})}{X^2}) \\ + d_{bZ} \text{Cov}(\widehat{b}, \widehat{Z}) + d_{b\lambda} \text{Cov}(\widehat{b}, \widehat{\lambda}) + d_{Z\lambda} \text{Cov}(\widehat{\lambda}, \widehat{Z}),$$

où

$$c_b = \frac{Z^3}{n\lambda}, c_Z = \frac{3b^2Z - b}{n\lambda}, c_\lambda = \frac{bZ^2(bZ - 1)}{n\lambda^3}, \\ c_{bZ} = \frac{6bZ^2 - 2Z}{n\lambda}, c_{b\lambda} = \frac{-4bZ^3 + 2Z^2}{n\lambda^2}, c_{Z\lambda} = \frac{-6b^2Z^2 + 4bZ}{n\lambda^2}, \\ d_b = c_b^2, d_Z = c_Z^2, d_\lambda = c_\lambda^2, d_{bZ} = 2c_b c_Z, d_{b\lambda} = c_b c_\lambda, d_{Z\lambda} = c_Z c_\lambda.$$

Par conséquent, pour l'estimateur $\widehat{Z} = \widehat{Z} + \widehat{b}_2(\widehat{Z})$ de Z , on a

$$E(\widehat{Z}) \simeq Z + c_b \text{Var}(\widehat{b}) + c_Z \text{Var}(\widehat{Z}) + c_\lambda \text{Var}(\widehat{\lambda}) \\ + c_{bZ} \text{Cov}(\widehat{b}, \widehat{Z}) + c_{b\lambda} \text{Cov}(\widehat{b}, \widehat{\lambda}) + c_{Z\lambda} \text{Cov}(\widehat{\lambda}, \widehat{Z}) + \frac{\sigma_{\widehat{W}}^2}{X^2} b^2 \frac{\text{Var}(\widehat{W})}{X^2} \\ \text{Var}(\widehat{Z}) = \text{Var}(\widehat{Z}) + \text{Var}(\widehat{b}_2(\widehat{Z})) + 2\text{Cov}(\widehat{b}_2(\widehat{Z}), \widehat{Z}).$$

Ces deux équations montrent que le biais de l'estimateur \widehat{Z} est relativement petit, mais que sa variance est plus grande que $\text{Var}(\widehat{Z})$, quantité actuellement retenue pour évaluer la précision de la biomasse estimée. Il apparaît cependant difficile au vu de ces équations d'obtenir un estimateur "correct" de $\text{Cov}(\widehat{b}_2(\widehat{Z}), \widehat{Z})$, sans recourir à des calculs longs et complexes. Aussi, nous proposons dans le paragraphe suivant l'utilisation de techniques alternatives pour estimer Z .

3- ESTIMATEURS DE TYPE JACKKNIFE

Rappelons d'abord le principe de l'estimation jackknife générale : Il consiste à partitionner l'échantillon $(W_i, X_i)_{i=1, \dots, n}$ en l sous-échantillons de taille h puis à construire des pseudo-estimations suivantes de Z :

$$\widetilde{Z}_{(j)} = l\widehat{Z} - (l-1)\widehat{Z}_{(-j)}; j = 1, \dots, l,$$

où $\widehat{Z}_{(-j)}$ est l'estimateur obtenu en supprimant le sous-échantillon j . L'estimateur jackknife de Z est alors défini par

$$\widehat{Z}_{jack} = l^{-1} \sum_{j=1}^l \widetilde{Z}_{(j)}.$$

Sa variance estimée est donnée par

$$\hat{\sigma}_{jack}^2 = \frac{1}{l(l-1)} \sum_{j=1}^l (\tilde{Z}_{(j)} - \tilde{Z}_{jack})^2.$$

Si $E(\hat{Z})$ est de la forme $E(\hat{Z}) = Z + a/n + O(n^{-2})$, cet estimateur a pour propriété d'éliminer le biais d'ordre $1/n$. Les résultats des études par Miller [4] sur le choix de l ont montré que si $l = n$, l'estimateur jackknife a le plus petit biais et généralement une variance plus petite que \hat{Z} . Cette procédure revient à supprimer une observation (W_j, X_j) pour obtenir $\tilde{Z}_{(-j)}$. Notons que dans ce cas, la pseudo-estimation $\tilde{Z}_{(j)}$ constitue le jackknife usuel.

L'application de l'estimation jackknife sur les données de 1994 donne les résultats suivants :

Si $E(\hat{Z})$ est de la forme $E(\hat{Z}) = Z + a/n + O(n^{-2})$, cet estimateur a pour propriété d'éliminer le biais d'ordre $1/n$. Les résultats des études par Miller [4] sur le choix de l ont montré que si $l = n$, l'estimateur jackknife a le plus petit biais et généralement une variance plus petite que \hat{Z} . Cette procédure revient à supprimer une observation (W_j, X_j) pour obtenir $\tilde{Z}_{(-j)}$. Notons que dans ce cas, la pseudo-estimation $\tilde{Z}_{(j)}$ constitue le jackknife usuel.

L'application de l'estimation jackknife sur les données de 1994 donne les résultats suivants :

$$\begin{aligned} \hat{Z}_{jack} &= 0.01603 & \hat{b}_{jack} &= 0.00025 \\ \hat{\sigma}_{jack}^2 &= 3.7802 \times 10^{-6} & \hat{\sigma}_{jack} &= 1.994 \times 10^{-3}, \end{aligned}$$

où \hat{b}_{jack} est l'estimation du biais par la technique jackknife. Notons à titre de comparaison que $\hat{Z} = 0.016286$.

Une autre technique d'estimation d'un ratio en compétition avec le jackknife est le bootstrap. Efron [2] et Efron et Tibshirani [3] ont montré qu'elles sont voisines, plus précisément, le jackknife peut être vu comme une approximation linéaire de l'estimation bootstrap.

CONCLUSION

La variance estimée de \hat{Z} que nous avons obtenue sur les données de 1994 est plus grande que celle de \hat{Z}_{jack} . Cependant, son calcul a requis plusieurs approximations dont certaines ne sont pas contrôlables. Ce qui n'est pas le cas des techniques jackknife ou bootstrap. Cet argument joint aux bonnes propriétés statistiques de \hat{Z}_{jack} (voir Miller [4]) nous conduit à préconiser l'utilisation des techniques jackknife ou bootstrap pour estimer Z .

ANNEXE 1 Expression exacte de la correction du biais

Supposons que la variable aléatoire X admette pour densité de probabilité f et ait une moyenne μ et variance σ^2 .

Posons $U = \frac{X-\mu}{\sigma}$ et $V = \frac{X-\mu}{\mu} = \frac{\sigma}{\mu} U$. On suppose également que $E(\frac{1}{X})$ et $E(\frac{1}{1+V})$ existent. Ecrivant $\frac{1}{X}$ sous la forme $\frac{1}{X} = \frac{1}{\mu} (\frac{1}{1+V})$, on a

$$\frac{1}{1+V} = 1 + \sum_{k=1}^q (-1)^k V^k + \xi_{q+1}(V)$$

où $\xi_{q+1}(V) = \frac{1}{1+V} - (1 + \sum_{k=1}^q (-1)^k V^k) = \frac{(-V)^{q+1}}{1+V}$.

Si les moments $EV^k, k = 2, \dots, q$ existent alors

$$E\left(\frac{1}{1+V}\right) = 1 + \sum_{k=1}^q (-1)^k EV^k + E\xi_{q+1}(V)$$

En particulier, pour $q=2,3$ on a

$$E\left(\frac{1}{X}\right) = \frac{1}{\mu} \left(1 + \left(\frac{\sigma}{\mu}\right)^2 + E\xi_3(V)\right); E\xi_3(V) = -E\frac{V^3}{1+V} = -\left(\frac{\sigma}{\mu}\right)^3 E\frac{U^3}{1+\frac{\sigma}{\mu}U}$$

$$E\left(\frac{1}{X^2}\right) = \frac{1}{\mu} \left(1 + \left(\frac{\sigma}{\mu}\right)^2 + E\xi_4(V)\right); E\xi_4(V) = E\frac{V^4}{1+V} = \left(\frac{\sigma}{\mu}\right)^4 E\frac{U^4}{1+\frac{\sigma}{\mu}U}$$

Avec les notations du paragraphe 2, $\hat{x} = V$ et en utilisant l'hypothèse (H1), on déduit l'expression de $b(\hat{Z})$.

ANNEXE 2 Expression des moments de la loi Gamma(α, λ).

Rappelons que si X is distribué suivant la loi $G(\alpha, \lambda)$, avec $\lambda > 0, \alpha > 0$, sa densité de probabilité est

$$f(x) = \frac{\alpha(\alpha x)^{\lambda-1} e^{-\alpha x}}{\Gamma(\lambda)}, x > 0 \text{ et } E(X^m) = \frac{1}{\alpha^m} E\{(\alpha X)^m\} = \frac{\Gamma(m+\lambda)}{\alpha^m \Gamma(\lambda)} \text{ pourvu que } m + \lambda > 0.$$

Par conséquent,

$$\left\{ \begin{array}{l} \mu = E(X) = \frac{\Gamma(\lambda+1)}{\alpha \Gamma(\lambda)} = \frac{\lambda}{\alpha} \\ E(X^2) = \frac{\Gamma(\lambda+2)}{\Gamma(\lambda)} = \frac{(\lambda+1)\lambda}{\alpha^2} \\ \text{var}(X) = \frac{\lambda}{\alpha^2} \end{array} \right.$$

et

$$E\left(\frac{1}{X}\right) = \frac{\alpha}{\lambda - 1}$$

$$E\left(\frac{1}{X^2}\right) = \frac{\alpha^2}{(\lambda - 1)(\lambda - 2)}, \text{Var}\left(\frac{1}{X}\right) = \frac{\alpha^2}{(\lambda - 1)^2(\lambda - 2)} \text{ pour } \lambda > 2.$$

Références

- [1] W. G. Cochran 1977 Sampling techniques, 3rd edition, New York, Wiley
- [2] B. Efron 1990 Bootstrap methods : Another look at the jackknife. Ann. Statist. 7, pp. 1-26
- [3] B. Efron, R. J. Tibshirani 1993 An introduction to the bootstrap. Chapman & Hall
- [4] R. G. Miller 1974 The jackknife - a review. Biometrika 61, pp. 1-15.
- [5] K. Parker 1985 Biomass model for the egg production method. NOAA Technique report NMFS 36 pp.5-6.
- [6] S. Picquelle & G. Stauffer 1985 Parameter estimation for an egg production method of Northern anchovy biomass assessment. NOAA Technique report NMFS 36 pp.7-15.
- [7] G.A.F. Seber 1973 The estimation of animal abundance and related parameters. Hafner Press
- [8] B. Truong-van 1997a Some improvements on the mortality model and estimation techniques for the determination of the anchovy biomass in the Biscay-Gascogne bay. Rapport du Laboratoire de Mathématiques Appliquées UPRES A 3055 CNRS Université de Pau et des Pays de l'Adour France
- [9] B. Truong-van 1997b An extended mortality model for Biscay bay anchovy data and a new estimation approach. Laboratoire de Mathématiques Appliquées UPRES A 3055 CNRS Université de Pau et des Pays de l'Adour France

Méthodes de vraisemblance pour les captures successives. Applications aux saumons dans l'Adour-Nive

Benoît Truong-van⁽¹⁾, Célestin C. Kokonendji ⁽²⁾& Patrick Prouzet⁽³⁾

⁽¹⁾Laboratoire de Statistique et Probabilités UMR C55830
Université P. Sabatier F-31062 Toulouse

⁽²⁾Laboratoire de Mathématiques Appliquées UPRES 5033
Université de Pau et des Pays de l'Adour F-64000 Pau

⁽³⁾IFREMER Station d'Hydrobiologie
F-64310 Saint-Pée sur Nive

Abstract

We give conditions for the unimodality of the likelihood function under binomial models for the removal method, when the capture probabilities depend on two parameters. Then we apply the methods to the genuine problem of estimating the salmon biomass in the river Adour from data of captures by professional fisheries and from evaluations of the separation between Adour and Nive. An analysis and discussion on the results obtained are also given.

Résumé

Nous donnons des conditions d'unimodalité de la fonction de vraisemblance pour des modèles binomiaux utilisés dans les méthodes de captures séquentielles, quand les probabilités de capture dépendent de deux paramètres. Puis, nous appliquons ces procédures et une de leurs dérivées au problème réel d'estimation de la biomasse des saumons remontant le fleuve Adour à partir des données de pêcheries professionnelles, de pêches à la ligne, de frayères et des évaluations de la séparation Adour-Nive. Nous donnons ensuite une analyse et une discussion des résultats obtenus.

1 - Introduction

Nous considérons ici un problème d'estimation de la biomasse N des saumons d'Atlantique remontant le fleuve Adour situé dans le sud-ouest de la France, à partir des données des trois campagnes de pêches de 1988 à 1990 (Prouzet *et al.*, 1996). Pour chaque campagne de pêche, on dispose du nombre k de pêcheries (professionnelles) ($k = 5$ pour l'année 1988 et $k = 6$ pour les 2 années 1989 et 1990), des captures par pêcherie (x_i , $i = 1, \dots, k$), du nombre de pêcheurs par pêcherie, du nombre de coups de filet, des captures des pêches (amateurs) à la ligne ($y = y_n + y_g$ où y_n sont celles sur la Nive et y_g dans les Gaves après la k ème pêcherie), du nombre de frayères de grands salmonidés ($f = f_n + f_g$ où f_n concerne la Nive et f_g les Gaves de l'Adour), des proportions π_1 des castillons (S1) et π_2 des grands saumons (S2), et des proportions ρ_1 et ρ_2 des femelles des deux catégories de saumons (S1) et (S2). Les pêcheries professionnelles successives, $i = 1, \dots, k$, sont situées uniquement sur l'Adour et la séparation Adour-Nive se trouve entre la deuxième et la troisième pêcherie.

Le problème considéré ne peut pas être résolu par des méthodes classiques de captures séquentielles. Aussi, nous le traitons en plusieurs étapes. D'abord, nous supposons que le taux r

de la séparation Adour-Nive est connu (sa valeur estimée par les experts étant de $\frac{1}{4}$) et ne considérons que les données des k pêcheries. Avec ces hypothèses, nous nous ramenons à un problème connu de captures séquentielles. Là, de nombreuses méthodes sont disponibles dans la littérature (e.g., Pollock, 1991), mais la plupart d'entre elles sont fondées ou dérivent de la procédure de maximum de vraisemblance associée au modèle appelé binomial qui considère chaque x_i comme une observation d'une variable aléatoire X_i dont la loi conditionnelle à S_{i-1} est binomiale $\mathcal{B}(N_i, p_i)$ avec $N_i = N - S_{i-1}$, $S_0 = 0$ et $S_i = x_1 + \dots + x_i$, $i = 1, \dots, k$ et p_i la probabilité de capture de la pêcherie i . Pour le cas considéré ici, la fonction de vraisemblance pour $\mathbf{x} = (x_1, \dots, x_k)$ s'écrit

$$L_r(N, \mathbf{p}) = \frac{\Gamma(N+1)\Gamma\{(1-r)N_3+1\}}{\Gamma(N_3+1)\Gamma\{(1-r)N_3-S_k+S_2+1\}} \left(\prod_{i=1}^k \frac{1}{\Gamma(x_i+1)} \right) \times \left(\prod_{i=1}^2 p_i^{x_i} (1-p_i)^{N_{i+1}} \right) \left(\prod_{i=3}^k p_i^{x_i} (1-p_i)^{(1-r)N_3-S_i+S_2} \right) \quad (1)$$

où $\mathbf{p} = (p_1, \dots, p_k)$ est le vecteur des probabilités de capture.

Comme le nombre des observations est de 6 au maximum, il est statistiquement judicieux de se restreindre à un ou deux paramètres. Nous nous limitons donc au deux cas suivants :

(A) $p_i = p$ (probabilités de capture constantes)

(B) $p_i = g(i; p, \alpha)$ où g est une fonction de i (généralement décroissante) et dépendante de 2 paramètres $p = p_1$ et α .

Nous avons retenu les hypothèses (A) et (B) ci-dessus puis que d'une part, ce sont celles adoptées dans la quasi-totalité de la littérature (voir Seber 1973), avec pour fonction g dans le cas (B), l'une des deux fonctions suivantes :

(B1) $p_i = p\alpha^{(i-1)}$, $0 < \alpha < 1$, (décroissance exponentielle) et

(B2) $p_i = (i-1)\alpha + p$ (décroissance linéaire),

et que d'autre part, elles correspondent aux résultats d'analyse que nous avons effectuée sur les efforts de pêches des différentes pêcheries.

Aux fonctions (B1) et (B2), nous ajoutons une troisième

(B3) $p^{(i-1)\alpha+1}$, $0 < \alpha \leq 1$ (décroissance polynomiale).

Bien que le modèle binomial avec l'une des conditions (A) ou (B) ait été couramment utilisé depuis de longue date, les premiers résultats théoriques concernant l'existence et l'unicité des estimateurs fondés sur le maximum de vraisemblance ne sont établis qu'en 1994 par Bedrick pour le cas (A) seulement.

Aussi, nous donnons d'abord dans le paragraphe suivant, des résultats d'existence et d'unicité de ces estimateurs pour le cas (B). Puis, nous présentons une analyse des résultats obtenus en appliquant le modèle (1) sous les deux hypothèses (A) et (B3), aux données du problème considéré. Enfin, nous mentionnons quelques développements en cours.

2 - Existence et unicité des estimateurs

Le problème que nous étudions dans ce paragraphe, est l'existence et l'unicité de solution de

$$\max_{N, \mathbf{p}} L_r(N, \mathbf{p}). \quad (2)$$

Nous prenons $r = 0$ sans perte de généralité, car r est supposé connu dans (1).

Dans le cas (A) des probabilités de capture constantes, Bedrick (1994) en résolvant le problème (2) sous la forme

$$\max_N \max_p L_0(N, p), \quad (3)$$

aboutit aux résultats suivants :

Proposition *Sous l'hypothèse (A), on a alors*

$$\hat{p} = \arg \max_p \ln L_0(N, p) = S / (S - kN + \sum_{i=1}^k S_i) =: \hat{p}(N).$$

De plus,

$$(i) \arg \max_N \ln L_0(N, \hat{p}(N)) < \infty \Leftrightarrow \sum_{i=1}^k (i-1)x_i < (k + (k-1)S)/2$$

$$(ii) \text{ pour } N > S, \text{ l'équation } \frac{\partial}{\partial N} \ln L_0(N, \hat{p}(N)) = 0 \text{ a au plus une solution finie}$$

(iii) la fonction log-vraisemblance profilée $\ln L_0(N, \hat{p}(N))$ est unimodale en N , et fournit une solution unique à (3).

Pour le cas (B) des probabilités p_i décroissantes et dépendantes de deux paramètres p et α , l'approche (3) ne convient plus à la résolution de (2). Aussi, dans Truong-van (1998), nous avons adopté l'approche suivante

$$\max_p \max_N L_0(N, p). \quad (4)$$

En écrivant la fonction log-vraisemblance sous la forme

$$\ln L_0(N, p) = G(N) + a(p)N + b(p)$$

$$\text{avec } a(p) = \sum_{i=1}^k \ln(1 - p_i), \quad b(p) = \sum_{i=1}^k \{x_i \ln p_i - S_i \ln(1 - p_i)\} \text{ et}$$

$G(N) = \ln \Gamma(N+1) - \ln \Gamma(N-S+1)$, on remarque que $a(p) < 0$, $b(p) \in [-\infty, \infty]$ et que $G(N)$ est strictement concave en N . Ainsi, pour qu'il y ait une solution à (4), il faudrait que la vitesse de croissance de $G(N)$ soit moins rapide que la décroissance de la droite $aN + b$ pour tout p fixé.

On peut alors énoncer le résultat principal suivant dont le point-clé est de révéler (voir corollaire 1) que l'on ne peut avoir d'existence et d'unicité d'estimateurs de maximum de vraisemblance du modèle (1) que si les probabilités de capture dépendent d'au plus de deux paramètres.

Théorème (i) *Pour p tel que $b(p)$ est réel et $0 > a(p) \geq -a_s$, avec $a_s = 1 + \frac{1}{2} + \dots + \frac{1}{S}$,*

$$N_* = \arg \max_{N \geq S} \ln L_0(N, p) =: N_*(a(p))$$

existe de manière unique et vérifie $\frac{\partial}{\partial N} G(N_) = -a(p)$.*

$$(ii) \text{ Si } \frac{S}{k^2} \frac{K}{(1-K)^2} \sum_{j=0}^{S-1} (N_*(a) - j)^{-2} > 1 \text{ où } K = \exp(ak/k), \text{ alors}$$

$$H_*(a) = G(N_*(a)) + a(N_*(a)) + \sum_{i=1}^k S_i/k$$

admet un maximum unique.

En conséquence, on peut garantir l'existence et l'unicité de la solution du problème (2) sous

l'hypothèse (B), à l'aide du résultat suivant :

Corollaire 1 *Sous la condition $a_s \leq a \leq a_M, b_m \leq b \leq b_M$ avec $a_M < 0, -\infty < b_m \leq b_M < \infty$, le problème (4) possède une solution unique si et seulement si il existe un C^1 -difféomorphisme entre (a, b) et $(h_1(p), h_2(p))$ où h_1 et h_2 sont des fonctions adéquates.*

De plus, l'approche (4) permet également de résoudre le problème (2) dans le cas (A). Ainsi, le résultat de Bedrick (1994) est un cas particulier du théorème :

Corollaire 2 *Sous l'hypothèse (A) $p_i = p$ identiques, si S est suffisamment grand et si*

$$\left(\frac{k}{q}\right)^2 \left\{ \frac{1}{k} (N_*(a(q)) - M_0) + \frac{s}{(1-q)^2} \left(\frac{q}{k}\right)^2 \right\} > \frac{\partial}{\partial a} N_*(a(q))$$

avec $q = 1 - p$ et $M_0 = \sum_{i=1}^k S_i/k$, alors le problème (4) admet une solution unique

$$q_* = \text{Argmax}_q \ln L_0(N_*(a(q)), 1 - q).$$

Pour les détails et preuves, nous référons le lecteur à Truong-van (1998).

3-Applications et discussion

Les captures x_i des différentes pêcheries $i = 1, \dots, k$ sont présentées dans le Tableau 1 ci-dessous.

Pêcherie	captures en 1988		captures en 1989		captures en 1990	
	S1	S2	S1	S2	S1	S2
1	445	810	112	93	807	119
2	87	204	40	52	32	28
3	123	220	56	44	139	47
4			5	1	42	10
5	39	29	42	4	75	29
6	17	19	26	16	39	25
Capture totale	711	1282	281	210	1134	258

Tableau 1.

La quatrième pêcherie ne s'est installée qu'à partir de 1989.

Nous présentons dans les deux tableaux 2 et 3 ci-dessous les résultats obtenus avec le modèle binomial (1) pour un coefficient de séparation Adour-Nive $r = \frac{1}{4}$ et sous les deux conditions de probabilités de capture (A) et (B3). Le taux d'exploitation que sont les ratios $\frac{S_i}{\hat{N}}$, sont exprimés en pourcentage.

	1988		1989		1990	
	S1	S2	S1	S2	S1	S2
\hat{p}	0.50	0.54	0.22	0.38	0.48	0.25
\hat{N}	784	1403	397	239	1237	353
Taux d'exploitation	90.7%	91.4%	70.8%	87.9%	91.7%	73.0%

Tableau 2.
Résultats dans le cas (A) de probabilités de capture constantes.

	1988		1989		1990	
	S1	S2	S1	S2	S1	S2
\hat{p}	0.39	0.43	0.18	0.38	0.48	0.20
$\hat{\alpha}$	0.40	0.47	0.06	0.07	0.47	0.08
\hat{N}	1050	1836	516	250	1447	474
Taux d'exploitation	67.7%	69.8%	54.4%	84.0%	78.4%	54.5%

Tableau 3.
Résultats dans le cas (B3) de probabilités de capture à décroissance polynomiale.

Pour les estimations dans les deux tableaux 2 et 3, les écart-types empiriques estimés que nous avons déduits des formules asymptotiques du maximum de vraisemblance, conduisent à des coefficients de variation estimés inférieurs à 10%.

Le problème majeur soulevé par les résultats obtenus dans les tableaux 2 et 3 ci-dessus, avec le modèle (1)-(2) sous les deux hypothèses (A) et (B3), est que les taux d'exploitation obtenus, surtout ceux du tableau 2, sont trop élevés par rapport aux évaluations des experts de la pêche fondées sur le comptage des frayères et des captures de pêche (amateur) à la ligne qui sont en amont dans les Gaves. Ce problème nous a conduit aux analyses et études suivantes :

(1) Les procédures d'estimation ne peuvent être mises en cause, pourvu que les modèles soient adéquats, car les résultats de simulation de Broutin & Moutou (1997) ont corroboré la consistance des estimateurs déduits de ces procédures.

(2) Nous avons aussi étudié la sensibilité des résultats d'estimation en fonction du paramètre r de séparation Adour-Nive. Les résultats obtenus en faisant varier r entre 0.20 et 0.30 sont quasiment les mêmes que ceux obtenus dans les tableaux 2 et 3 avec $r = 0.25$. (Pour plus de détails, voir Truong-van (1998)).

(3) Une approche proposée par Prouzet *et al.* (1996) pour adapter le modèle binomial au problème considéré, consiste à restreindre la variation de la biomasse N dans des intervalles dont les extrêmes sont déterminées à partir de données de frayères et des pêches à la ligne en amont dans les Gaves.

Les résultats que nous avons obtenus avec cette approche et en considérant le modèle binomial avec l'hypothèse (B3) et avec des contraintes d'inégalités sur la biomasse N , sont résumés dans le Tableau 4 ci-dessous.

	1988		1989		1990	
	S1	S2	S1	S2	S1	S2
\hat{N}	1396	2539	1105	627	1693	407
Taux d'exploitation	51%	50%	25%	33%	66%	63%

Tableau 4.

Résultats obtenus avec le modèle binomial sous contraintes d'inégalités sur N et avec l'hypothèse (B3).

Les résultats obtenus dans le Tableau 4 sont conformes aux évaluations des experts. Cependant, les maxima obtenus avec cette approche sont très dépendants des bornes choisies des contraintes et ne permettent pas d'obtenir une estimation des variances des estimateurs par les formules asymptotiques associées à la vraisemblance, quand ces maxima sont situés sur les extrémités des contraintes.

Aussi, nous proposons dans nos travaux en cours les axes des développements suivants :

(i) Prise en compte des contraintes sur la biomasse à partir des données en amont de l'Adour et de la Nive sous forme de termes de pénalisation (principalement quadratique) de la vraisemblance, de façon d'une part à nous placer dans un "système fermé" et d'autre part à obtenir des formules asymptotiques pour les estimateurs, surtout leurs variances asymptotiques.

(ii) Une meilleure modélisation des probabilités de capture, puisqu'à partir d'un simple examen du Tableau 1, ces dernières se comportent en moyenne comme une fonction décroissante entre la 1ère et la 2ème pêcherie, puis croissante ou discontinue entre la 2ème et la 3ème, et décroissante très rapidement ensuite. Or, si on conserve la procédure de vraisemblance associée au modèle binomial, le nombre de paramètres est limité à deux par le théorème du paragraphe précédent. Pour surmonter cette difficulté, la solution que nous étudions actuellement est une pénalisation de la vraisemblance.

(iii) Une approche bayésienne d'abord pour la prise en compte de la séparation Adour-Nive, en supposant que le taux r est une variable aléatoire de loi beta d'espérance $\frac{1}{4}$, puis en supposant que la biomasse est une variable aléatoire dont l'étendue ou plus généralement dont la loi de probabilités est déterminée par les données de frayères et/ou de pêches à la ligne en amont. C'est une alternative à (i).

Bibliographie

- Bedrick, E.J. (1994). Maximum-likelihood estimation for the removal method, *The Canadian Journal of Statistics*, **22**(2), 285-293.
- Broutin, E. & Moutou, S. (1997). *Méthodes de vraisemblance pour l'estimation de la biomasse par des captures successives*, Rapport Technique CURS, DESS d'Ingénierie Mathématique et Outils Informatiques, Université de Pau.
- Pollock, K.H. (1991). Modelling capture, recapture, and removal statistics for estimation of demographic parameters for fish and wildlife populations: past, present, and future. *Journal of the American Statistical Association*, **86**(413), 225-238.
- Prouzet, P., Martinet, J.-P. & Cuendé, F.-X. (1996). *Rapport sur la pêche des marins pêcheurs dans l'estuaire de l'Adour en 1995*, Rapport IFREMER, Saint-Pée sur Nivelle.
- Seber, G. A. F. (1973) *The estimation of animal abundance and related parameters*. Hafner Press.
- Truong-van, B. *Estimation de la biomasse des saumons dans le bassin de l'Adour*. Rapport du contrat IFREMER-ADERA 1997-1998 et Annexe technique du 28/07/1998.

BIOMÉTRIE ET HALIEUTIQUE
C. DUBY, J.-P. GOUET, F. LALOË — N° 15