

REVUE DE STATISTIQUE APPLIQUÉE

N. BRU

S. DOSSOU-GBETE

Estimations non-paramétriques des courbes d'abondance de civelles d'Anguille dans l'estuaire de l'Adour

Revue de statistique appliquée, tome 48, n° 1 (2000), p. 69-83

http://www.numdam.org/item?id=RSA_2000__48_1_69_0

© Société française de statistique, 2000, tous droits réservés.

L'accès aux archives de la revue « Revue de statistique appliquée » (<http://www.sfds.asso.fr/publicat/ras.htm>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques
<http://www.numdam.org/>

ESTIMATIONS NON-PARAMÉTRIQUES DES COURBES D'ABONDANCE DE CIVELLES D'ANGUILLE DANS L'ESTUAIRE DE L'ADOUR

N. Bru^{(1),(2)} et S. Dossou-Gbete⁽²⁾

(1) INRA, Saint-Pée sur Nivelle, France.

(2) Université de Pau et des Pays de l'Adour, Pau, France.

RÉSUMÉ

L'objectif de ce travail est de décrire les principales caractéristiques de l'évolution des captures de civelles d'Anguille dans l'estuaire de l'Adour afin de réaliser une classification d'une série de saisons de pêche. La méthodologie utilisée combine des techniques d'estimation non-paramétrique de fonctions avec celles issues de l'Analyse en Composantes Principales (ACP) de données fonctionnelles.

Mots-clés : *Estimation non-paramétrique de courbes, ACP, civelles d'Anguille.*

ABSTRACT

The aim of this work is the description of the main characteristics of the evolution of glass-eels catches into the Adour estuary, in order to propose a classification of several fishing seasons. The method used in this paper combines some tools of non-parametric estimation of functions with ones from Principal Components Analysis of functional data.

Keywords : *Non-parametric estimation of curves, PCA, glass-eels.*

1. Introduction

La pêche de la civelle contribue largement à la survie de l'activité de pêche professionnelle sur l'Adour. La pérennité de cette activité ne peut être assurée que par la considération des exigences économiques combinée avec la nécessaire protection de la ressource en diminution constante. Dans ce contexte, la gestion de la biomasse de civelles présentes dans l'Adour et la maîtrise de l'exploitation du stock revêtent un caractère primordial. La présente étude constitue une première étape dans la mise en place de mécanismes qui permettront de réaliser ces objectifs. Elle se fonde sur des données de captures de civelles collectées auprès des pêcheurs professionnels de l'Adour exploitant le stock de civelles et dont l'activité est répertoriée

officiellement. Elles ont été regroupées par le laboratoire INRA-IFREMER de la station d'hydrobiologie de Saint Pée sur Nivelle.

Pour décrire l'évolution des captures sur chaque saison de pêche, nous avons considéré qu'elle peut être modélisée par la composition additive de deux termes : une fonction déterministe du temps qui décrit l'évolution moyenne des captures au cours d'une saison de pêche et un terme aléatoire résultant d'une déviation aléatoire par rapport à cette moyenne. Cet article présente une méthode d'estimation de la fonction moyenne en deux étapes. Une première étape fournit une estimation d'un effet moyen général à partir d'un lissage des données observées à l'aide d'une technique de régression non-paramétrique. La seconde étape consiste à estimer l'effet différentiel spécifique à chaque saison de pêche par une approximation de faible rang réalisée par une technique de type ACP (Besse *et al.* 1997).

L'exposé de ce travail s'organise comme suit. Dans la section 2, nous présentons les données utilisées. La section 3 contient le modèle statistique. Nous détaillons les estimations des composantes de ce modèle dans le cas où les données analysées sont des courbes discrétisées dans la section 4. La section 5 décrit la démarche d'interprétation des résultats. La section 6 est dédiée à l'application de la méthode aux séries de captures de civelles. L'interprétation des résultats d'un point de vue biologique est abordée en section 7. La section 8 fait office d'annexe.

2. Présentation des données de captures de civelles

2.1. Provenance des données

Les prises réalisées par les pêcheurs au cours de chacune de leur sortie sont consignées dans des carnets de pêche individuels. C'est à partir de ces données qu'ont été calculés des indicateurs de niveau moyen de captures désignés par CPUE (Captures Par Unité d'Effort). Leur calcul prend en compte le nombre d'engins de pêche utilisés (effort de pêche) et le nombre de sorties de pêche effectuées par pêcheur. Les CPUE s'expriment en kilogrammes. Ce sont des données journalières depuis 1984/1985 jusqu'à 1992/1993, soit 9 saisons de pêche qui couvre une période de 5 mois, du premier novembre d'une année au 31 mars de l'année suivante.

Les captures réalisées par les professionnels sont effectuées dans la couche superficielle de la colonne d'eau, jusqu'à 1.5 m de profondeur. De ce fait, l'intensité des captures est fonction de la disponibilité de la ressource civelles qui résulte à la fois de l'accessibilité de la ressource et de sa vulnérabilité à l'engin de pêche. Les CPUE sont des indicateurs de l'évolution de cette disponibilité à différents moments d'une saison de pêche.

2.2. Remarques préliminaires sur les courbes de captures

La figure 1 représente les courbes de CPUE journalières de civelles d'Anguille sur les 9 saisons de pêche étudiées sur une période de 144 jours successifs à partir du 1-er novembre, date d'ouverture de la saison de pêche.

Nous pouvons noter que les allures des courbes présentent une très grande variabilité à travers l'échantillon et sont très bruitées. On observe aussi des différences

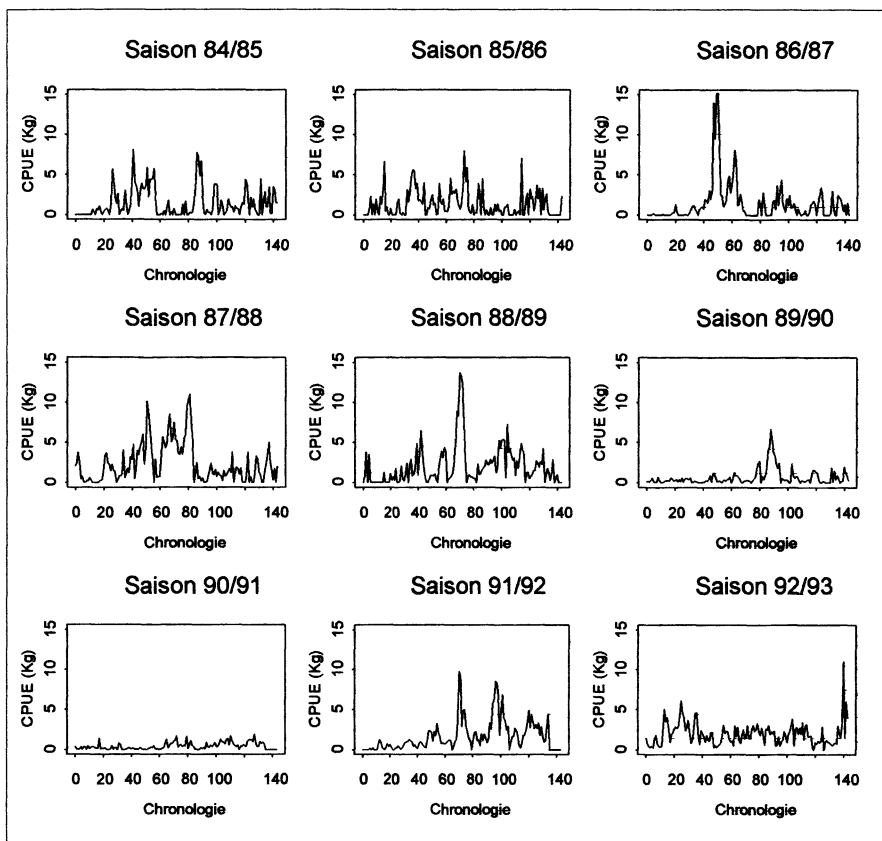


FIGURE 1
Courbes de CPUE sur les 9 saisons de pêche

sensibles entre le niveau moyen des CPUE d'une saison à l'autre. Néanmoins, on peut remarquer que dans chaque saison, les principaux pics de captures se situent entre le 50-ième jour et le 120-ième de la saison, soit de décembre à février.

Les points d'observation (à savoir les jours) peuvent être différents d'une saison de pêche à l'autre. Il existe plusieurs raisons à cela. En particulier, les données relatives à certaines dates sont manquantes pour certaines saisons, par exemple les saisons de pêche 1990/91 et 1991/92. Une des raisons à cette situation réside dans le fait que la législation en vigueur sur l'Adour impose que le dimanche soit jour de «relève» (suspension provisoire) pour la pêche à la civelle. Cependant, nous trouvons parfois sur les données de CPUE des valeurs non nulles certains dimanches. Il existe donc une grande incertitude sur les valeurs nulles de CPUE.

3. Modèle statistique pour la description de l'évolution des CPUE

Nous avons à notre disposition un échantillon de n ($n = 9$) saisons de pêche observées à des points de discrétisation temporels successifs t_{ij} , $j = 1, \dots, p_i$, en nombre fini, qui sont des jours de chaque saison. Soit $x_i(t_{ij})$ la valeur de CPUE correspondant à la date t_{ij} , $i = 1, \dots, n$ et $j = 1, \dots, p_i$. Nous supposons que les données observées sont compatibles avec le modèle additif suivant :

$$x_i(t_{ij}) = \mu_i(t_{ij}) + \varepsilon_{ij} \quad i = 1, \dots, n, \text{ et } j = 1, \dots, p_i. \quad (1)$$

On peut considérer sans perte de généralité que les t_{ij} appartiennent à l'intervalle $[0, 1]$ qui symbolise une saison de pêche, ainsi $0 \leq t_{i1} < t_{i2} < \dots < t_{ip_i} \leq 1$. Pour simplifier, nous noterons parfois dans la suite $x_i(t_{ij}) = x_{ij}$, $i = 1, \dots, n$, et $j = 1, \dots, p_i$.

L'effort d'analyse se concentre sur les composantes systématiques μ_i , que l'on souhaite estimer et interpréter en tant que caractéristique propre à chaque saison de pêche.

Les traitements réalisés dans la suite à partir du modèle (1) supposent que :

- (i) Les effets moyens μ_i sont des fonctions lisses qui reflète un phénomène écologique «régulier» qui est la disponibilité de la ressource «civelle» présente dans l'estuaire de l'Adour. C'est pourquoi nous considérons que nos données sont de nature fonctionnelle (Ramsay et Silverman 1997). Pour prendre en compte la nature fonctionnelle du paramètre μ_i à estimer et sa régularité, on impose que μ_i appartient à un sous-espace d'un espace de Hilbert H de fonctions régulières définies sur $[0, 1]$. Pour traduire le fait que la variation de la disponibilité d'une saison à l'autre résulte d'un petit nombre q de «causes» sous-jacentes, on considère que les μ_i sont contenus dans un sous-espace affine V de H , de dimension finie et faible q . On désignera par V_0 le sous-espace vectoriel directeur de V également de dimension q . Il en résulte que les fonctions μ_i se décomposent comme suit :

$$\mu_i = \alpha + \nu_i, \quad (2)$$

avec $\alpha \in V$ et $\nu_i \in V_0$. Ainsi, le terme α représente la tendance moyenne générale des captures (ou effet moyen général) sur l'ensemble des n saisons de pêche et ν_i l'effet différentiel propre à la i -ème saison de pêche qui permet de la «particulariser» par rapport à l'effet moyen général α . Pour des raisons d'identifiabilité, nous imposons la contrainte $\sum_{i=1}^n \nu_i = 0$ sur $[0, 1]$ qui implique

$$\text{que } \alpha = \frac{1}{n} \sum_{i=1}^n \mu_i = \bar{\mu}_n.$$

- (ii) Les spécialistes du phénomène considèrent que l'intervalle de temps séparant deux saisons de pêche successives est suffisamment grand (à savoir 7 mois entre la fin de l'une et le début de la suivante) pour qu'il n'y ait aucun lien entre les captures d'une année et celles de la précédente. On suppose alors que les termes d'erreur non-observables ε_{ij} sont, pour i fixé, les réalisations

de variables aléatoires non corrélées de moyenne nulle et de variance finie σ^2 indépendante de i . On a donc :

$$\begin{cases} E(\varepsilon_{ij}) = 0, & i = 1, \dots, n, \text{ et } j = 1, \dots, p_i, \\ E(\varepsilon_{ij}\varepsilon_{i'j'}) = \sigma^2\delta_{i,i'}\delta_{j,j'}, & i, i' = 1, \dots, n, j = 1, \dots, p_i \text{ et } j' = 1, \dots, p_{i'}. \end{cases}$$

où $\delta_{a,b}$ est le symbole de Kronecker égal à 1 si $a = b$ et 0 sinon.

4. Estimations des courbes moyennes μ_i

Compte tenu de la relation (2), afin que le modèle (1) ci-dessus soit complètement spécifié, il nous faut estimer V_0 ainsi que sa dimension q , $\alpha \in V$ et $(\nu_i)_{i=1, \dots, n} \subset V_0$ pour caractériser entièrement μ_i . Pour construire une estimation de l'effet déterministe fonctionnel μ_i , il est nécessaire de prendre en compte à la fois le caractère régulier des fonctions μ_i et le fait que les données sont observées sur un nombre fini de jours non-identiques d'une saison à l'autre. Ceci nous a amené à considérer une procédure d'estimation en deux étapes :

1. On réalise une première évaluation des courbes μ_i en tant que fonction lisse par une méthode de régression non-paramétrique sur les données discrétisées $\mathbf{x}_i = (x_{ij})_{j=1, \dots, p_i}$ qui est en fait le vecteur d'observation de CPUE correspondant aux données de la i -ème saison de pêche. La courbe ainsi obtenue est une estimation préliminaire de l'effet moyen fonctionnel μ_i et sera notée $\tilde{x}_i$.
2. On réalise ensuite l'estimation des composantes du modèle par la minimisation du critère moindres-carrés fonctionnels suivant :

$$\frac{1}{n} \sum_{i=1}^n \|\tilde{x}_i - \mu_i\|_{L^2}^2, \quad (3)$$

où, $\|\cdot\|_{L^2}$ est la norme usuelle de l'espace de Hilbert $L^2([0, 1])$. En partant de l'hypothèse que μ_i s'écrit sous la forme (2), on montre que cette étape se ramène à la décomposition spectrale de l'opérateur nucléaire $\Gamma_{\tilde{X}}$ de noyau

$$\Gamma_{\tilde{X}}(s, t) = \frac{1}{n} \sum_{i=1}^n \tilde{x}_i(s) \tilde{x}_i(t) \text{ où } \tilde{x}_i \text{ résulte de l'étape 1 ci-dessus après centrage.}$$

Cette approche permet de considérer dans un même cadre la situation où les discrétisations des courbes sont identiques et celle où les discrétisations diffèrent d'une courbe à l'autre. Elle offre par ailleurs l'avantage de réduire, grâce à l'étape d'estimation préliminaire, les bruits dûs à d'éventuelles erreurs de mesure. La méthode d'estimation qui est présentée dans la seconde étape s'inspire de la démarche utilisée pour estimer les composantes du modèle à effets fixes en analyse multivariée (Caussinus 1986, Fine 1989). Cette conception rejoint aussi celle de Kneip (1994) qui propose une estimation de μ_i par minimisation du critère des moindres-carrés fonctionnels sur une estimation préliminaire de μ_i correspondant à la projection des observations sur une base de fonctions.

Les estimations des composantes du modèle sont alors (cf. Annexe), pour q fixé :

$$\begin{cases} \hat{\alpha} = \frac{1}{n} \sum_{i=1}^n \tilde{x}_i, \\ \hat{\nu}_i = P_{\hat{V}_0} \tilde{x}_i \quad i = 1, \dots, n, \end{cases} \quad \text{sur } [0, 1], \quad (4)$$

où $P_{\hat{V}_0}$ est le projecteur orthogonal sur le sous-espace $\hat{V}_0$ engendré par les vecteurs propres de $\Gamma_{\tilde{X}}$ associés aux q premières valeurs propres, lorsque celles-ci sont rangées en ordre décroissant et sont répétées un nombre de fois égal à leur multiplicité.

En assimilant $\Gamma_{\tilde{X}}$ à un opérateur de covariance, les vecteurs propres f_k peuvent être considérés comme les facteurs principaux résultant de l'ACP fonctionnelle d'ordre q sur les données centrées lissées $\tilde{x}_i$. Par conséquent,

$$\hat{\nu}_i = \sum_{k=1}^q c_{ik} f_k \quad i = 1, \dots, n. \quad (5)$$

Les composantes principales $c_k = (c_{ik})_{i=1, \dots, n}$ sont liées aux facteurs principaux par la relation suivante : $c_{ik} = \int_0^1 \tilde{x}_i(t) f_k(t) dt$. Pour q fixé, l'estimation $\hat{\mu}_i$ de μ_i est alors donnée par le calcul suivant : $\hat{\mu}_i = \hat{\alpha} + \hat{\nu}_i$, $i = 1, \dots, n$.

Nous désignerons dans la suite par L la technique de régression non-paramétrique que nous appellerons également «lissage» et qui fournit $\tilde{x}_i = Lx_i$ à l'étape 1. L'algorithme qui permet de mettre en oeuvre la méthode exposée précédemment est le suivant :

- | |
|--|
| <ol style="list-style-type: none"> 1. Lisser les données par régression non-paramétrique : $\tilde{x}_i = Lx_i \quad i = 1, \dots, n$, 2. Calculer la moyenne empirique des données lissées : $\bar{x} = \frac{1}{n} \sum_{i=1}^n \tilde{x}_i = \hat{\alpha}$, 3. Centrer les données lissées : $\tilde{\tilde{x}}_i = \tilde{x}_i - \bar{x} \quad i = 1, \dots, n$, 4. Calculer les q premiers vecteurs propres de l'opérateur : $\Gamma_{\tilde{\tilde{X}}}$ avec $\Gamma_{\tilde{\tilde{X}}} = \frac{1}{n} \sum_{i=1}^n \tilde{\tilde{x}}_i \otimes \tilde{\tilde{x}}_i$. |
|--|
- (6)

Dans la procédure (6), on obtient $\hat{\alpha}$ dès l'étape 2. L'étape 4 nous permet de construire $\hat{V}_0$ et ainsi de calculer ensuite $\hat{\nu}_i$ d'après la relation contenue dans (4).

Le choix de la technique de régression L est un problème pratique qui dépend fortement des données. Cette question a été largement abordée dans la bibliographie. Les techniques les plus courantes sont les ajustements, suivant un critère des moindres-carrés, par la méthode des noyaux, par des polynômes, des B-splines après

spécification des noeuds (Eubank 1988, Besse *et al.* 1997), ou par des polynômes trigonométriques (Ramsay et Silverman 1997). Le lissage des données, comme le dit Besse (1994), doit prendre en considération les caractéristiques des séries de données.

5. Interprétation des résultats de la procédure d'estimation

Nous nous intéressons dans la suite à une discussion des variations des courbes $\mu_i(t)$, $t \in [0, 1]$, au cours d'une saison de pêche i fixée. Malgré la terminologie d'ACP employée dans ce qui précède, les résultats de l'estimation des effets fonctionnels μ_i à partir des données bruitées x_i ne peuvent pas être interprétés comme dans le contexte classique de l'ACP. L'interprétation des variations des effets μ_i se fera en considérant chacune de ses composantes : l'effet moyen général α et l'effet différentiel ν_i propre à l'unité statistique considérée. Les conclusions obtenues par la suite sur les variations du phénomène au cours d'une saison de pêche sont donc fondées sur une interprétation des variations de α et de celles des ν_i à travers l'analyse des facteurs f_k , $k = 1, \dots, q$, qui fixent les propriétés des ν_i d'après la relation (5). Afin d'interpréter le rôle des facteurs sur l'effet moyen général α , nous comparons leurs actions respectives sur la courbe moyenne à travers l'analyse des courbes $\alpha + cf_k$ et $\alpha - cf_k$, $k = 1, \dots, q$, où c désigne un réel fixé arbitrairement > 0 . Cette technique rejoint une idée de Silverman (1995).

6. Traitement des CPUE de civelles

L'objectif de cette section est de comparer les différentes estimations des courbes moyennes μ_i obtenues sur chaque saison de pêche, sous l'hypothèse que μ_i se décompose sous la forme (2).

6.1. Estimation préliminaire de la courbe moyenne par lissage des courbes de CPUE

Pour permettre aux valeurs estimées de conserver le caractère de positivité des valeurs de CPUE de départ, nous avons privilégié un lissage à l'aide de noyau. L'étape de lissage permet d'apprécier la tendance générale de chaque série de captures au cours d'une saison de pêche en faisant abstraction des valeurs extrêmes, que ce soit les valeurs ponctuellement nulles ou très fortes qui sont difficiles à expliquer dans l'état actuel des connaissances du phénomène. De manière alternative, on pourrait projeter les observations sur une base de B-splines en imposant des contraintes de positivité sur les coefficients en suivant en cela une idée de Kelly et Rice (1990).

Nous avons testé la régression par noyau gaussien pour différentes valeurs de la fenêtre de lissage choisies entre 0.007 et 0.1, tandis que la discrétisation de chaque saison de pêche est de 144. L'estimation fonctionnelle $\tilde{x}_i$ de x_i ainsi obtenue s'écrit alors :

$$\tilde{x}_i(t) = \sum_{j=1}^{p_i} W_h(t, t_{ij}) x_i(t_{ij}), \quad i = 1, \dots, n, \text{ et } t \in [0, 1],$$

avec, $W_h(t, t_{ij}) = \frac{K\left(\frac{t - t_{ij}}{h}\right)}{\sum_{l=1}^{p_i} K\left(\frac{t - t_{il}}{h}\right)}$ où K désigne le noyau gaussien et h est la fenêtre de lissage.

Nous présentons dans la suite les résultats obtenus pour une fenêtre de lissage de 0.1.

6.2. Estimation de l'effet moyen général α

La courbe présentée ci-dessous représente l'estimation $\hat{\alpha}$ de α .

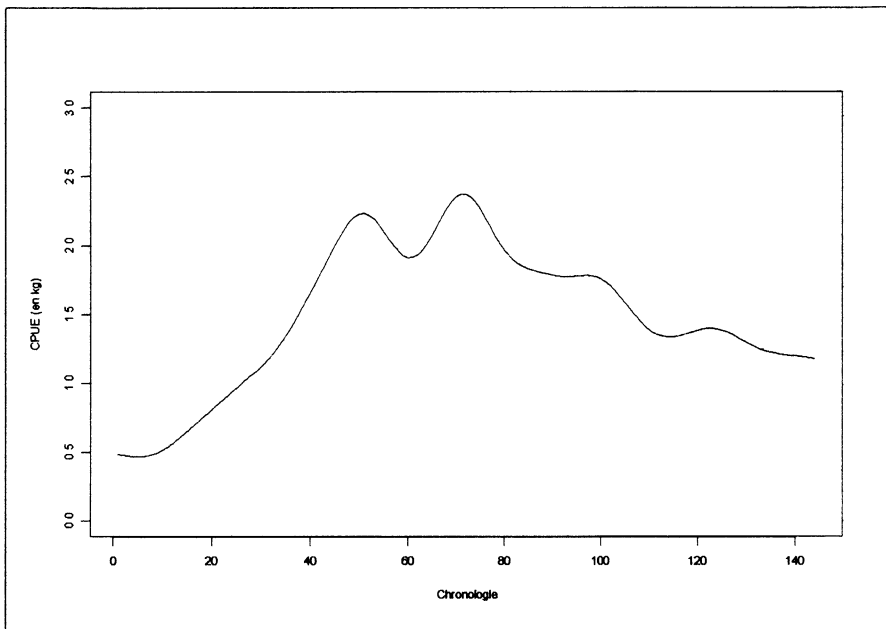


FIGURE 2
Estimation de l'effet moyen général α

Nous pouvons remarquer la présence de 4 pics : deux de forte intensité en milieu de saison, un léger au 2/3 et le quatrième en fin de saison.

6.3. Etude des facteurs principaux

Le pourcentage d'inertie expliquée par les trois premières valeurs propres est de 82 %. La représentation dans l'espace V_0 construit à l'aide d'une base à 3 éléments ($q = 3$) est à notre avis satisfaisante.

Les ν_i s'écrivant en fonction des facteurs f_k , $k = 1, \dots, q$, il s'agit d'interpréter ces facteurs pour comprendre les variations des ν_i . Sur un même graphique sont représentées la courbe α décrivant l'effet moyen général en trait continu, les courbes $(\alpha + 1.5f_k, k = 1, \dots, 3)$ en pointillés et $(\alpha - 1.5f_k, k = 1, \dots, 3)$ en traits brefs alternés avec des pointillés.

Effet du premier facteur

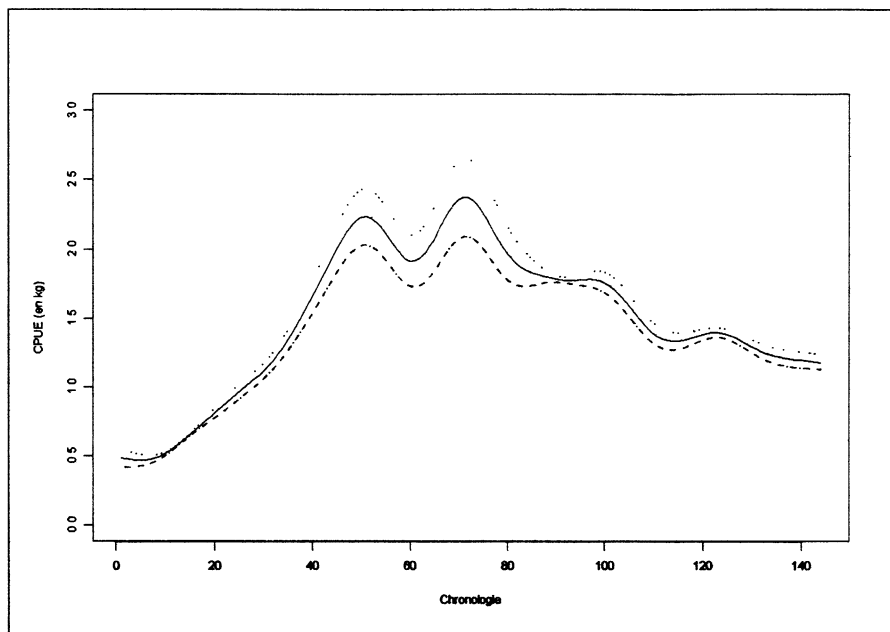


FIGURE 3

Effet du premier facteur sur l'effet moyen général :
 α (trait plein); $\alpha + 1.5f_1$ (pointillé); $\alpha - 1.5f_1$ (trait + pointillé)

Le premier facteur est une fonction de signe constant qui agit sur le niveau de la courbe moyenne : c'est un facteur de niveau. L'écart entre les courbes à la moyenne est plus important en milieu de saison : si on le multiplie par un nombre de même signe que les composantes du facteur, le milieu de la saison a été nettement meilleur que la moyenne et inversement.

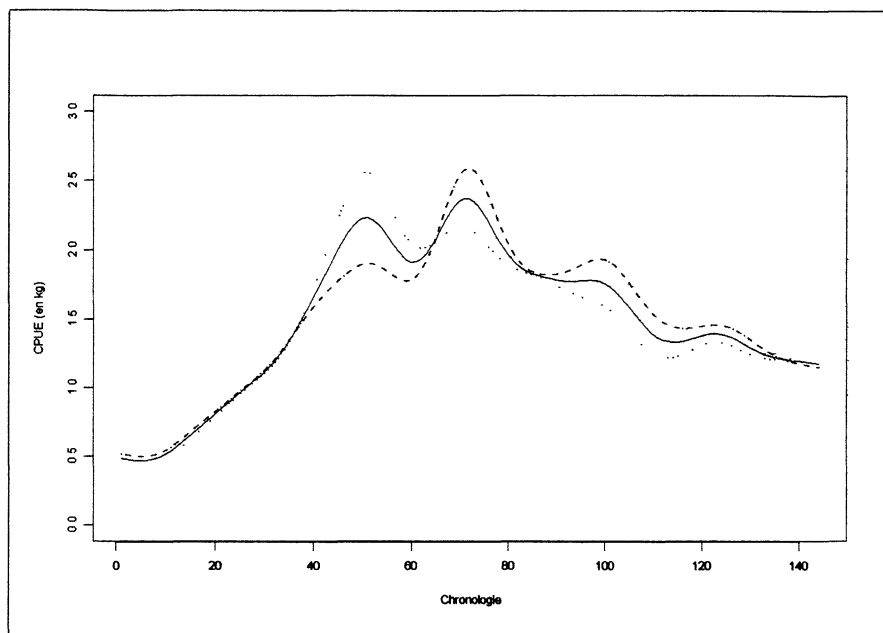
Effet du deuxième facteur

FIGURE 4

Effet du second facteur sur l'effet moyen général :
 α (trait plein); $\alpha + 1.5f_2$ (pointillé); $\alpha - 1.5f_2$ (trait + pointillé)

Le deuxième facteur permet de discuter de l'évolution de saisons de pêche qui ont commencé avec une intensité de captures de l'ordre de la moyenne. Il semble indiquer un effet de compensation au cours d'une saison : si la saison commence bien c'est-à-dire avec une intensité de captures plus forte que la moyenne au niveau du premier pic alors elle se termine mal, et vice versa.

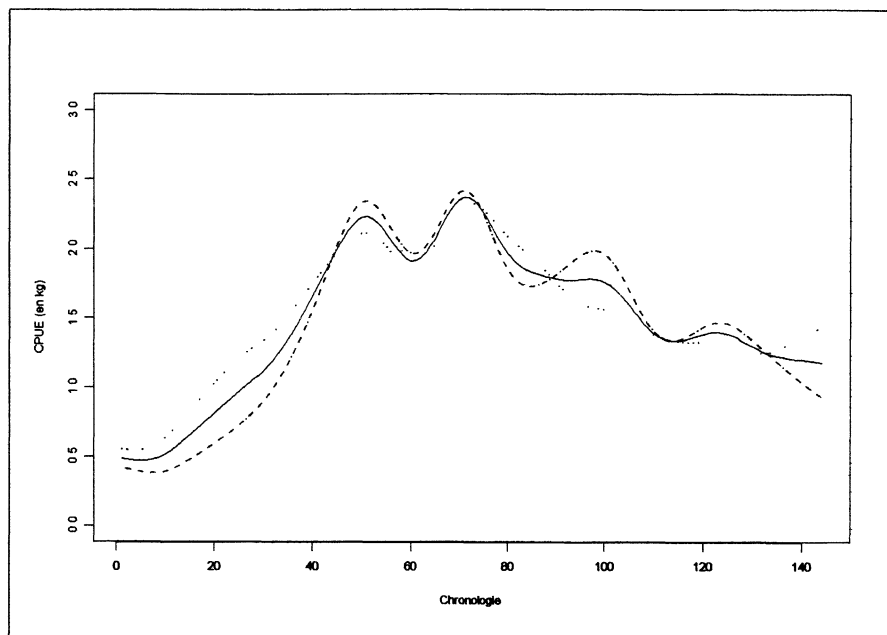
Effet du troisième facteur

FIGURE 5

Effet du troisième facteur sur l'effet moyen général :
 α (trait plein); $\alpha + 1.5f_3$ (pointillé); $\alpha - 1.5f_3$ (trait + pointillé)

Le troisième facteur met en relief les périodes extrêmes d'une saison de pêche : le début et la fin de la saison. Globalement, si la saison commence mal, elle se termine dans des conditions de captures relativement fortes et inversement.

6.4. Etude des composantes principales

Les composantes principales correspondant aux 3 premiers facteurs principaux sont les suivantes :

TABLEAU 1

Valeurs des 3 premières composantes principales pour chaque saison de pêche

	1984/85	1985/86	1986/87	1987/88	1988/89	1989/90	1990/91	1991/92	1992/93
c_1	-0.0165	-0.0096	0.0202	0.1089	0.0577	-0.0787	-0.0916	0.0048	0.0047
c_2	0.0428	0.0011	0.0919	0.0157	-0.0628	-0.0062	-0.0071	-0.0557	-0.0197
c_3	0.0071	0.0227	-0.0393	0.0198	-0.0287	-0.0020	-0.0039	-0.0446	0.0689

7. Discussion

7.1. Localisation temporelle des flux de civelles

L'estimation de l'effet moyen général α nous montre l'apparition en moyenne de 4 pics durant la saison de pêche. Ces pics pourraient correspondre à plusieurs flux (ou vagues) d'arrivée de civelles d'intensités différentes, au nombre généralement de 4 semble-t-il pour l'estuaire de l'Adour. Guérault *et al.* (1985) signalent l'arrivée de 5 flux de civelles dans la Vilaine. En partant de cette hypothèse, nous remarquons que les flux de plus grande intensité se localisent en milieu de saison. Ces flux ne sont pas nettement différenciés, ce qui signifie que l'on peut avoir des civelles provenant d'arrivées différentes dans un même endroit. Cette idée est confirmée par de Casamajor *et al.* (1996) à partir d'analyses de variance effectuées sur les caractéristiques biométriques des individus en migration dans l'estuaire de l'Adour durant la saison de pêche 96/97.

Cette analyse reflète certainement une caractéristique du phénomène bien connue en dynamique de population que l'on appelle la «variabilité saisonnière des arrivées de civelles». Les arrivées de civelles se font toute l'année, mais les plus fortes abondances sont signalées d'octobre à avril, les mois où la pêche est autorisée (Elie et Rochard 1994). Cependant le pic principal de migration apparaît plus ou moins tôt. Les variations sont dues aux arrivées par vagues successives au cours de la saison.

7.2. Classification des saisons de pêche de 84/85 à 92/93

Le premier facteur permet de caractériser les années qui ont été globalement «bonnes» si la première composante principale est positive, ou «mauvaises» si la première composante principale est négative. Il permet ainsi de caractériser la tendance générale de l'intensité des captures au cours de la saison de pêche. Le deuxième facteur semble caractériser les années pour lesquelles la bonne période a été «précoce» dans la saison de pêche, si le signe de la seconde composante est positif, ou «tardive» si la seconde composante est négative. Il permet d'avoir une idée de la localisation des bonnes périodes durant la saison de pêche. Guérault *et al.* (1991) parlent également de remontées tardives de civelles. Le troisième facteur fixe les propriétés, en termes d'intensité des captures, du début et de la fin de chaque saison de pêche. Castelnau *et al.* (1994) font état de telles comparaisons sur les saisons de pêche depuis 1990, et Guérault *et al.* (1985) parlent également de «variations d'abondance». Prouzet *et al.* (1994) s'intéressent également à l'évolution des captures de civelles dans l'estuaire de l'Adour avec un suivi annuel basé sur les données des carnets de pêche.

En rassemblant toutes les remarques ci-dessus, nous avons pu «trier» les saisons de pêche.

TABLEAU 2
Classification des saisons de pêche

		f ₁		
		Bonne	Moyenne	Mauvaise
f ₂	Précoce	86/87	84/85	
	Moyenne	87/88	85/86	90/91
	Tardive	88/89	91/92 92/93	89/90

Nous retrouvons ainsi des conclusions identiques à celles obtenues par de Casamajor *et al.* (1996) qui ont utilisé une analyse de variabilité mensuelle des captures de civelles de l'Adour pour caractériser les saisons de pêche.

8. Annexe

On désigne par $\mathcal{F}_{0q}(V)$ l'ensemble des sous-espaces vectoriels de dimension q de V . Résoudre (3) se ramène à la minimisation du terme suivant :

$$\sigma_q(V_0, \alpha, \nu) = \frac{1}{n} \sum_{i=1}^n \left\| \tilde{x}_i - \nu_i \right\|_{L^2}^2 + \left\| \bar{x}_n - \alpha \right\|_{L^2}^2,$$

avec $\bar{x}_n = \frac{1}{n} \sum_{i=1}^n \tilde{x}_i$ et $\tilde{x}_i = \tilde{x}_i - \bar{x}_n$.

LEMME 1. — *L'appartenance*

$$(\hat{V}_0, \hat{\alpha}, \hat{\nu}) \in \arg \min \{ \sigma_q(V_0, \alpha, \nu), \alpha \in V, \nu_i \in V_0, V_0 \in \mathcal{F}_{0q}(V) \}$$

équivalent à :

- (i) $\hat{\alpha} \in \arg \min \left\{ \left\| \bar{x}_n - \alpha \right\|_{L^2}^2, \alpha \in V \right\}$ si et seulement si : $\hat{\alpha} = \bar{x}_n$.
- (ii) $(\hat{V}_0, \hat{\nu}) \in \arg \min \left\{ \frac{1}{n} \sum_{i=1}^n \left\| \tilde{x}_i - \nu_i \right\|_{L^2}^2, \nu_i \in V_0, V_0 \in \mathcal{F}_{0q}(V) \right\}$ si et seulement si :

$$\left\{ \begin{array}{l} \hat{V}_0 \in \arg \max \left\{ \frac{1}{n} \sum_{i=1}^n \left\| P_{V_0} \tilde{x}_i \right\|_{L^2}^2, V_0 \in \mathcal{F}_{0q}(V) \right\}, \\ \hat{\nu}_i = P_{\hat{V}_0} \tilde{x}_i, i = 1, \dots, n, \end{array} \right.$$

où P_{V_0} est le projecteur orthogonal sur V_0 .

Preuve. — Le résultat (i) est trivial. On s'attache à la démonstration de (ii). Pour tout sous-espace V_0 donné dans $\mathcal{F}_{0q}(V)$ et pour toute suite $(\nu_i)_{i=1, \dots, n} \subset V_0$,

il vient que :

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n \left\| \tilde{x}_i - \nu_i \right\|_{L^2}^2 &= \frac{1}{n} \sum_{i=1}^n \left\| \tilde{x}_i - P_{V_0} \tilde{x}_i \right\|_{L^2}^2 + \frac{1}{n} \sum_{i=1}^n \left\| P_{V_0} \tilde{x}_i - \nu_i \right\|_{L^2}^2 \\ &= \frac{1}{n} \sum_{i=1}^n \left\| \tilde{x}_i \right\|_{L^2}^2 - \frac{1}{n} \sum_{i=1}^n \left\| P_{V_0} \tilde{x}_i \right\|_{L^2}^2 + \frac{1}{n} \sum_{i=1}^n \left\| P_{V_0} \tilde{x}_i - \nu_i \right\|_{L^2}^2. \end{aligned}$$

D'où le résultat. $\square$

LEMME 2. — Soit $\hat{V}_0$ le sous-espace vectoriel engendré par une suite orthonormée de q fonctions propres de l'opérateur $\Gamma_{\tilde{X}}$ associées aux q premières valeurs propres rangées en ordre décroissant. Alors :

$$\hat{V}_0 \in \arg \max \left\{ \frac{1}{n} \sum_{i=1}^n \left\| P_{V_0} \tilde{x}_i \right\|_{L^2}^2, V_0 \in \mathcal{F}_{0q}(V) \right\}.$$

Preuve. — On a $\left\| P_{V_0} \tilde{x}_i \right\|_{L^2}^2 = \text{tr} \left(P_{V_0} \tilde{x}_i \otimes \tilde{x}_i \right)$ et $\Gamma_{\tilde{X}} = \frac{1}{n} \sum_{i=1}^n \tilde{x}_i \otimes \tilde{x}_i$ opérateur autoadjoint et compact. Pour conclure, on utilise un résultat classique d'analyse fonctionnelle. $\square$

Références

- [1] BESSE Ph. (1994), Models for Multivariate Data Analysis. *Compstat 84*, Dutter and Grossmann (eds.), Physica-Verlag, 271–285.
- [2] BESSE Ph., CARDOT H. et FERRATY F. (1997), Simultaneous non-parametric regressions of unbalanced longitudinal data. *Computational Statistics and Data Analysis*, 24, 255–270.
- [3] CASAMAJOR (DE) M.-N., BRU N. et PROUZET P. (1996), *La civelle d'Anguille (Anguilla anguilla L.) dans le bassin de l'Adour : Productions et influence des conditions environnementales sur sa capturabilité*. Rapport intermédiaire IFREMER, 41 p.
- [4] CASTELNAUD G., GUERULT D., DESAUNAY Y. et ELIE P. (1994), Production et abondance de la civelle en France au début des années 90. *Bull. Fr. Pêche Piscic.*, 335, 263–287.
- [5] CAUSSINUS H. (1986), Models and Uses of Principal Component Analysis – A Comparison emphasizing graphical displays and metric choices. *Multi-dimensional Data Analysis*, J. De Leeuw et al. (eds), DSWO Press, Leiden, 149–170.
- [6] ELIE P. et ROCHARD E. (1994), Migration des civelles d'Anguilles (*Anguilla anguilla* L.) dans les estuaires, modalités du phénomène et caractéristiques des individus. *Bull. Fr. Pêche Piscic.*, 335, 81–98.

- [7] EUBANK R.L. (1988), *Spline smoothing and nonparametric regression*. Dekker, New York.
- [8] FINE J. (1989), *Contribution à l'analyse en Composantes Principales – Développements sur les Modèles Fonctionnels, Propriétés Asymptotiques*. Doctorat d'Etat, Laboratoire de Statistique et Probabilités, Université Toulouse III, Toulouse.
- [9] GUERAULT D., BEILLOIS P., DESAUNAY Y. et DOREL (1985), Variations de l'abondance de la civelle au travers des données de production secteur Loire et Vilaine. *CIEM*, GM 1985/M : 26, 7 p.
- [10] GUERAULT D., DESAUNAY Y., BEILLOIS P. et GRELLIER P. (1991), Etude des montées tardives de civelles et des conditions de colonisation des bassins versants de Loire et Vilaine. *IFREMER*, Nantes, 18 p.
- [11] KELLY C. et RICE J. (1990), Monotone Smoothing with Application to Dose-Response Curves and the Assessment of Synergism. *Biometrics*, 46, 1071–1085.
- [12] KNEIP A. (1994), Nonparametric estimation of common regressors for similar curve data. *The Annals of Statistics*, 22, 1386–1427.
- [13] PROUZET P., MARTINET J.-P. et CUENDE F.-X. (1994), Les pêches estuariennes du bassin de l'Adour de 1985 à 1991. *Repère Océan*, IFREMER (6), 31–48.
- [14] RAMSAY J.O. et SILVERMAN B.W. (1997), *Functional Data Analysis*. Springer-Verlag.
- [15] SILVERMAN B.W. (1995), Incorporating parametric effects into functional principal components analysis. *J. R. Statist. Soc B*, 57, 673–689.