

L'ANALYSE EN COMPOSANTES PRINCIPALES  
SUR VARIABLES INSTRUMENTALES :  
APPLICATION AUX RENDEMENTS DE PECHE  
DE DEUX PORTS DE LA COTE SENEGALAISE

Nicolas Pech<sup>a</sup>

## I - INTRODUCTION

La collecte de données relatives à l'activité de la pêche artisanale Sénégalaise se déroule suivant un plan d'échantillonnage stratifié à plusieurs niveaux (Laloë, 1985; Gérard et Greber, 1985). Nous disposons ainsi, sur plusieurs années, d'un ensemble d'estimations de rendements de pêche. Ces derniers concernent plusieurs espèces capturées au moyen de divers engins de pêche et débarquées dans différents ports. Ces rendements sont *a priori* soumis à un ensemble de sources de variation soulignées par le plan d'échantillonnage (variabilités inter et intra annuelles, port, engins de pêche) ou pas (influences environnementales (Cury et Roy, 1988); adaptabilité des pêcheurs (Laloë et Samba, 1989)). Notre objet va être la recherche d'une description de nos données qui soit la plus concise et la plus interprétable possible. Pour cela, nous utiliserons les sources de variation dégagées par le plan d'échantillonnage, et présenterons un outil d'analyse des données : l'analyse en composantes principales sur variables instrumentales (Rao, 1965; Sabatier et Lebreton, 1989).

## II - DONNEES ET METHODES

Le jeu analysé ici consiste en des estimations de rendements de pêche concernant 25 espèces débarquées dans deux ports de la côte Sénégalaise (Saint-Louis et Kayar). Ces rendements sont donnés par quinzaine de 1975 à 1991. On dispose ainsi d'un tableau ou matrice  $Y$  de 816 lignes et 25 colonnes. Chaque colonne (ou variable) contient les rendements d'une espèce, et chaque ligne (ou unité d'observation) les rendements de 25 espèces pour une année, une quinzaine, et un port donnés.

---

<sup>a</sup>Centre ORSTOM, HEA, B.P 5045, 34032 Montpellier Cedex 1.

En plus des variables d'intérêt consignées dans la matrice  $Y$ , on peut relever sur les mêmes unités d'observation un ensemble de variables dites de structure : les indicatrices des ports, des années et des quinzaines.

Considérons  $Y^l$ , la colonne  $l$  du tableau  $Y$  (préalablement centré et réduit par colonne). On peut alors représenter  $Y^l$  en utilisant les variables de structure au moyen d'un modèle d'analyse de variance (Draper et Smith, 1981). L'estimation des paramètres par l'estimateur des moindres carrés conduit à décomposer  $Y^l$  en :

$$Y^l = \widehat{Y}_P^l + \widehat{Y}_A^l + \widehat{Y}_Q^l + \widehat{Y}_{AP}^l + \widehat{Y}_{AQ}^l + \widehat{Y}_{PQ}^l + \widehat{Y}_{APQ}^l + \widehat{\epsilon}^l \quad (1)$$

où  $\widehat{Y}_P^l, \widehat{Y}_A^l, \widehat{Y}_Q^l$  sont les effets simples estimés Port, Année et Quinzaine et  $\widehat{Y}_{AP}^l, \widehat{Y}_{AQ}^l, \widehat{Y}_{PQ}^l, \widehat{Y}_{APQ}^l$  sont respectivement les interactions estimées Année  $\times$  Port, Année  $\times$  Quinzaine, Port  $\times$  Quinzaine, et Année  $\times$  Port  $\times$  Quinzaine.

Géométriquement chacun des vecteurs de la somme est obtenu en projetant  $Y^l$  sur le sous-espace vectoriel engendré par l'effet correspondant. Le plan étant équilibré, les vecteurs entrant dans la décomposition (1) sont deux à deux orthogonaux. D'autre part, le modèle est ici saturé, donc  $\widehat{\epsilon}^l = 0$ . L'orthogonalité a pour conséquence l'additivité des variances de chacun des effets considérés.

Le modèle d'analyse de variance que nous avons écrit pour une variable se généralise pour les 25 espèces en un modèle d'analyse de variance multidimensionnel (Arnold, 1981) :

$$(Y^1, \dots, Y^p) = (\widehat{Y}_P^1, \dots, \widehat{Y}_P^p) + (\widehat{Y}_A^1, \dots, \widehat{Y}_A^p) + \dots + (\widehat{Y}_{APQ}^1, \dots, \widehat{Y}_{APQ}^p) \quad (2)$$

On décompose ainsi  $Y$  en une somme de 7 matrices, chacune contenant les effets des 25 variables associés à un effet simple, une interaction d'ordre 2, ou une interaction d'ordre 3. L'inertie de  $Y$  (qui vaut 25, les variables étant réduites) se décompose aussi en :

$$I_Y = I_P + I_A + \dots + I_{APQ} \quad (3)$$

comme indiqué dans le tableau 1.

L'analyse exploratoire du tableau de données  $Y$  peut consister à en faire l'ACP (figure 1). Seulement, son interprétation n'est pas toujours simple. En effet, considérons deux espèces dont les rendements sont chacun la somme de deux effets simples annuels et saisonniers. Supposons que

leurs effets saisonniers soient proches et leurs effets annuels opposés. La représentation dans le plan principal fera apparaître une corrélation entre ces deux variables, mais il ne sera pas forcément aisé d'en déterminer la cause. Une solution à ce type de problème est alors de faire l'ACP (centrée mais qui n'est plus normée de chacun des 7 "tableaux effets" définis dans l'équation (2)). Les deux espèces apparaîtront corrélées dans l'ACP de  $\widehat{Y}_Q$  (i.e. quant à leurs effets saisonniers) et opposées dans l'ACP de  $\widehat{Y}_A$  (ie quant à leurs effets annuels). L'ACP de  $\widehat{Y}_A$  est appelée ACPVI de Y relativement à l'effet annuel (ACPVI Y/A). Cela revient à faire une analyse des corrélations entre variables relativement à un facteur sans interférer (au sens de l'orthogonalité entre sous espaces) avec les autres facteurs définis par les variables de structure.

### III - RESULTATS

Les 7 ACPVI associées à chacun des facteurs ont été réalisées. Dans la représentation par le plan principal (figure 2 pour l'ACVI Y/Quizaine), chaque espèce est représentée par son code. La distance au cercle de corrélation permet de juger de la qualité de la reconstitution, comme pour l'ACP. La distance au centre du cercle est en partie soulignée. La proportion de la part non soulignée de ce segment est égale à la racine carrée de la variance due au facteur. Ainsi, on peut comparer le comportement saisonnier des variables en tenant compte de l'influence du facteur sur chacune d'entre elles. Les représentations dans les plans principaux permettent ainsi des regroupements d'espèces relativement à plusieurs des effets définis. Ces regroupements sont interprétés à la lumière de la reconstitution de certaines variables par les axes des ACPVI correspondantes (figure 3). Ils permettent une caractérisation de plusieurs comportements types: oppositions Kayar/Saint-Louis, saison froide/saison chaude, tendance interannuelle à la baisse/tendance interannuelle à la hausse, interactions Saint-Louis×saison chaude/Kayar ×saison chaude,...

### IV - DISCUSSION

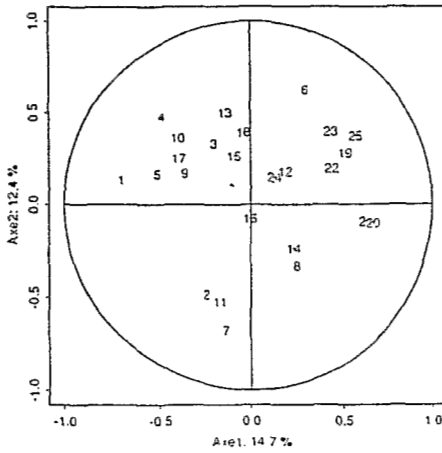
Nous pouvons alors nous demander si ces "comportements types" ne peuvent pas être utilisés pour décrire notre jeu de données initial au moyen d'un modèle statistique, c'est à dire une description des données sous la forme d'un effet systématique plus un résumé sur la nature et l'ampleur des variations inexplicées (Mac Cullagh et Nelder, 1989). Nous avons construit plusieurs modèles statistiques appartenant à la famille des modèles

linéaires multidimensionnels où la résiduelle estimée n'est plus nulle, et où les variables explicatives sont définies à partir d'un ensemble de restrictions des 3 effets principaux. En utilisant un critère (AICc) de sélection de modèles multidimensionnels construit à partir du critère d'Akaike (Bedrick et Tsai, 1994), le modèle (99 ddl) qui minimise le critère (figure 4) est alors sélectionné. Ce modèle constitue notre synthèse du tableau initial.

## BIBLIOGRAPHIE

- ARNOLD (S.F.), 1981 - *The theory of linear models in multivariate analysis*. John Wiley & Sons, New-York, USA.
- BEDRICK (E.J.), & TSAI (C.L.), 1994 - Model selection for multivariate regression in small samples. *Biometrics*, 50 : 226-231.
- CURY (P.), & ROY (C.), 1988 - Migration saisonnière du thiof (*Epinephelus aenus*) au Sénégal: influence des upwellings Sénégalais et mauritaniens. *Oceanol. Acta*, 11 : 25-36.
- DRAPER (N.), & SMITH (H.), 1981 - *Applied Regression Analysis*. John Wiley & Sons, New-York, USA.
- GERARD (M.), & GREBER (P.), 1985 - Analyse de la pêche artisanale au Cap Vert: description et étude critique du système d'enquête, Centre ORSTOM de Dakar Thiaroye, *Doc. Scient. Cent. Rech. Océano*. 98, 77p.
- LALOE (F.), 1985 - Etude de la précision des estimations de captures et prises par unité d'effort obtenues à l'aide du système d'enquêtes de la section "pêche artisanale" du CRODT, Centre ORSTOM de Dakar Thiaroye, *Doc. Scient. Cent. Rech. Océano*. 100, 36p.
- LALOE (F.), & SAMBA (A.), 1989 - *La pêche artisanale au Sénégal : ressources et stratégies de pêches*. Collection études et thèses, Paris, ORSTOM.
- MC CULLAGH (P.), & NELDER (J.A), 1989 - *Generalized linear Models*. Second edition, Chapman & Hall, New-York. 511 p.
- RAO (C.R.), 1965 - The use and interpretation of Principal Component Analysis in applied research. *Sankhya Ser. A*, 26 : 329-358.

SABATIER (R.), LEBRETON (J.D.), & CHESSEL (D.), 1989 - Principal component analysis with instrumental variables as a tool for modelling composition data. *Multiway data analysis* : 341-352.



Source de variation	ddl	Inertie	Inertie/ddl
Port	1	1.87	1.87
Année	16	2.98	0.186
Quinzaine	23	4.89	0.212
Port×Année	16	1.33	0.083
Port×Quinzaine	23	2.25	0.098
Année×Quinzaine	368	6.56	0.018
Année×Port×Quinzaine	368	5.12	0.014
Total	815	25	0.031

Tableau 1 : Tableau de décomposition de l'inertie de Y selon les sous-espaces engendrés par les variables de structure

Figure 1 : Analyse en Composantes principales de Y, représentation des variables dans le plan principal

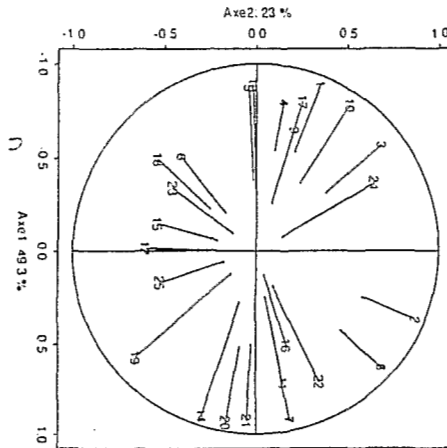


Figure 2 : ACPVI de Y relativement à l'effet quinzaine, représentation des variables dans le plan principal

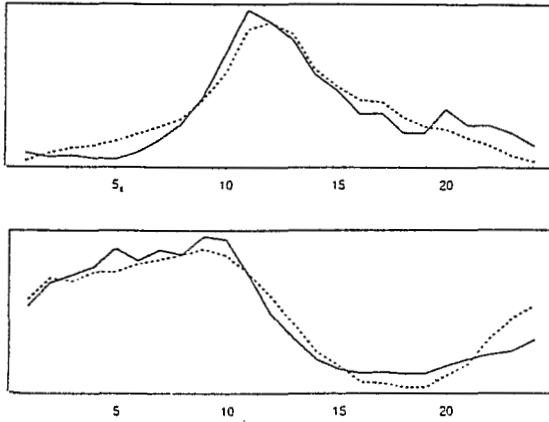


Figure 3 : ACPVI de Y relativement à l'effet quinzaine, reconstitution par les deux premiers axes des comportements saisonniers du pagre à points bleus et du tassergal

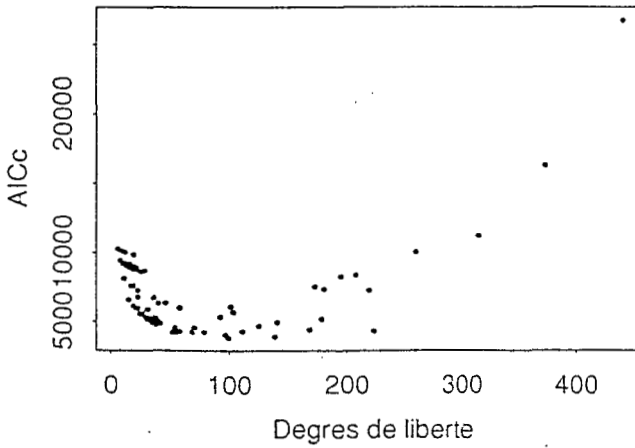


Figure 4 : Tracé du critère d'akaike en fonction du degré de liberté pour les différents modèles définis