ICES Journal of Marine Science, 54: 32–47. 1997

Use of Principal Component Analysis with Instrumental Variables (PCAIV) to analyse fisheries catch data

Nicolas Pech and Francis/Laloë

Pech, N., and Laloë, F. 1997. Use of Principal Component Analysis with Instrumental Variables (PCAIV) to analyse fisheries catch data. – ICES Journal of Marine Science, 54: 32–47.

Principal Component Analysis with respect to Instrumental Variables (PCAIV) is a statistical tool for exploratory analysis combining both principal component analysis and multivariate regression analysis. This tool is used to analyse mean fortnightly catches obtained by Senegalese fishermen in two ports from 1975 to 1991. The aim of the study is to identify significant sources of variation and to present separately the impact of each of them. These descriptions are used to characterize the initial data.

Key words: principal component analysis, multivariate analysis of variance, instrumental variables, selection of models.

Received 28 June 1995; accepted 22 May 1996.

N. Pech and F. Laloë: Centre ORSTOM, HEA, B.P. 5045, 34032 Montpellier cedex 1, France.

L'Analyse en Composantes Principales sur Variables Instrumentales (ACPVI) est un outil statistique d'analyse exploratoire faisant intervenir l'analyse en composantes principales et l'analyse de regression multivariée. Cet outil est utilisé ici pour l'analyse d'estimations bimensuelles de rendements de pêche réalisés par les pêcheurs artisans Sénégalais dans deux ports de 1975 à 1991. L'objet de notre étude est d'identifier des sources de variation influentes en présentant séparément l'impact de chacune d'entre elles. Ces descriptions seront ensuite utilisées afin de former une synthese des données initiales.

Mots clés: analyse en composantes principales, analyse de variance multivariée, variables instrumentales, sélection de modèles.

Introduction

The Centre de Recherches Océanographiques de Dakar-Thiaroye (CRODT) of the Institut Sénégalais de Recherche Agricole (ISRA) has been collecting data for at least 20 years on the artisanal fishery along the coast of Senegal, using a consistent sampling design (Gérard and Greber, 1985; Laloë, 1985). The objective of this system is to obtain fishing effort and catch data used for stock assessment purposes. In this design, data are collected within strata defined by combinations of gears, fortnights[1] and ports of landing.

While stock assessments are generally done on a single species basis, questions concerning biological and technical interactions (the latter include the "effort allocation problem", see Laurec *et al.*, 1991) require multispecies approaches. To that end, data from individual fishing trips are usually analysed using multivariate methods and cluster analysis to build up typologies of fishing units or typologies of "métiers" (Murawski *et al.*, 1983; Biseau and Gondeaux, 1988). In our experience (Gérard and Greber, 1985; Laloë and Samba, 1990; Samba and Laloë, 1991; Ferraris and Samba, 1992), such analyses of these kinds of data clearly indicate the existence, even within the use of a particular gear, of different "tactics" (Laloë and Samba, 1990) or "métiers" (Laurec *et al.*, 1991) or "technotopes" (Fay, 1994). In addition, the fishermen may take information from the "environment" into account in order to decide which "métier" to use (Garrod, 1973; Hilborn, 1985; Allen and MacGlade, 1986; Laurec *et al.*, 1991; Laloë and

[1]Fortnight is defined here as "half a month", thus there are 24 fortnights in a year.

1054–3139/97/010032+16 $25.00/0/jm960180

Table 1. Fish species mainly caught by hand-line in Senegalese fishery, with code numbers used on the figures in this paper. Ouolof is one of the national languages of Senegal.

Code	Scientific name	English name	Ouolof name
1	*Pomatomus saltatrix*	Bluefish	Ngott
2	*Pagrus caeruleostictus*	Blue spotted seabream	Kibaro naar
3	*Decapterus rhonchus*	False scad	Diaï
4	*Epinephelus aeneus*	White grouper	Thiof
5	*Euthynnus alletteratus*	Little tuny	Oualass
6	*Pagellus bellotti*	Red pandora	Youfouf
7	*Arius latisculatus*	Rough-head sea catfish	Dakak
8	*Alectis alexandrinus*	Alexandria pompano	Yawal
9	*Trichiurus lepturus*	Largehead hairtail	Tallar
10	*Epinephelus guaza*	Dusky sea perch	Kauthieu
11	*Pseudot. senegalensis*	Cassava croaker	Feute
12	*Rhinobato* spp.	Guitarfish	Thiaukher
13	*Argyrosomus regius*	Meagre	Seukhebi
14	*Sphyrna* spp.	Sharks	Gaïndé Guédj
15	*Epinephelus goreensis*	Dungat grouper	Doï
16	*Lagocephalus laevigatus*	Smooth puffer	Boun foki
17	*Sarda sarda*	Atlantic bonito	Oual
18	*Dentex canariensis*	Canary dentex	Kibaro ngokh
19	*Diversus ienaplenus*	Various	Ndiakhas
20	*Coryphaena hippurus*	Common dolphinfish	Ndiakhssine
21	*Istiophorus albicans*	Atlantic sailfish	Dieunou dong
22	*Mustelus mustelus*	Smooth hound	Mâne
23	*Brotula barbata*	Bearded brotula	Mori
24	*Octopus vulgaris*	Common octopus	Yaranka
25	*Dentex macrophtalmus*	Large eye dentex	Mbagne mbagnère

Samba, 1991); the term environment here refers to the environment experienced by the fish as well as that experienced by the fishermen (Fréon, 1986; Cury and Roy, 1988, 1991; Samba and Laloë, 1991).

While the time series of mean catches for a single species may simply reflect changes in abundances as is usually assumed, these changes may also be due to many other sources of variation and interactions. Therefore, we need tools to partition out these sources of variation. In this paper, we present an application of principal component analysis with respect to instrumental variables (PCAIV: Rao, 1964; Inzenman, 1980; Sabatier *et al.*, 1989; Lebreton *et al.*, 1991) to partition the sources of variation in the Senegalese landings data for handlines. This method combines features of the more familiar methods of multivariate regression analysis and principal component analysis (PCA). The partitions identified herein correspond to inter-annual variation, intra-annual variation, variation of port of landing and the interactions between all of these. Our approach emphasizes the use of graphics.

Table 2. Decomposition of the inertia according to orthogonal subspaces induced by instrumental variables. This variability is expressed in terms of inertia (see Appendix 1) which is the multivariate expression of the variance.

Source of variation	Degree of freedom	Inertia	Inertia df
Port	1	1.87	1.87
Year	16	2.98	0.19
Fortnight	23	4.89	0.21
Port × year	16	1.33	0.08
Port × fortnight	23	2.25	0.10
Year × fortnight	368	6.56	0.02
Year × port × fortnight	368	5.12	0.01
Total	815	25	0.03

Materials and methods

Data

For our analysis we used data extracted from the CRODT data base (Ferraris *et al.*, 1993), consisting of mean catches of fish from handlines on daily trips and landed in one of two ports (Saint-Louis and Kayar). These catch data are given by species and by fortnight from 1975 to 1991. The 25 species that we considered are listed in Table 1. In our analysis the data were in matrix form Z, with 816 rows and 25 columns, where each column (or variable) represents a species and each of the 816 rows (2 ports × 17 years × 24 fortnights) contains the catch by species. Based on empirical evidence of skew, we transformed the data with the logarithmic function ($Y=\log(Z+1)$). Moreover, Y has been centred and scaled in columns, so that the mean of each column is zero with variance equal to one.

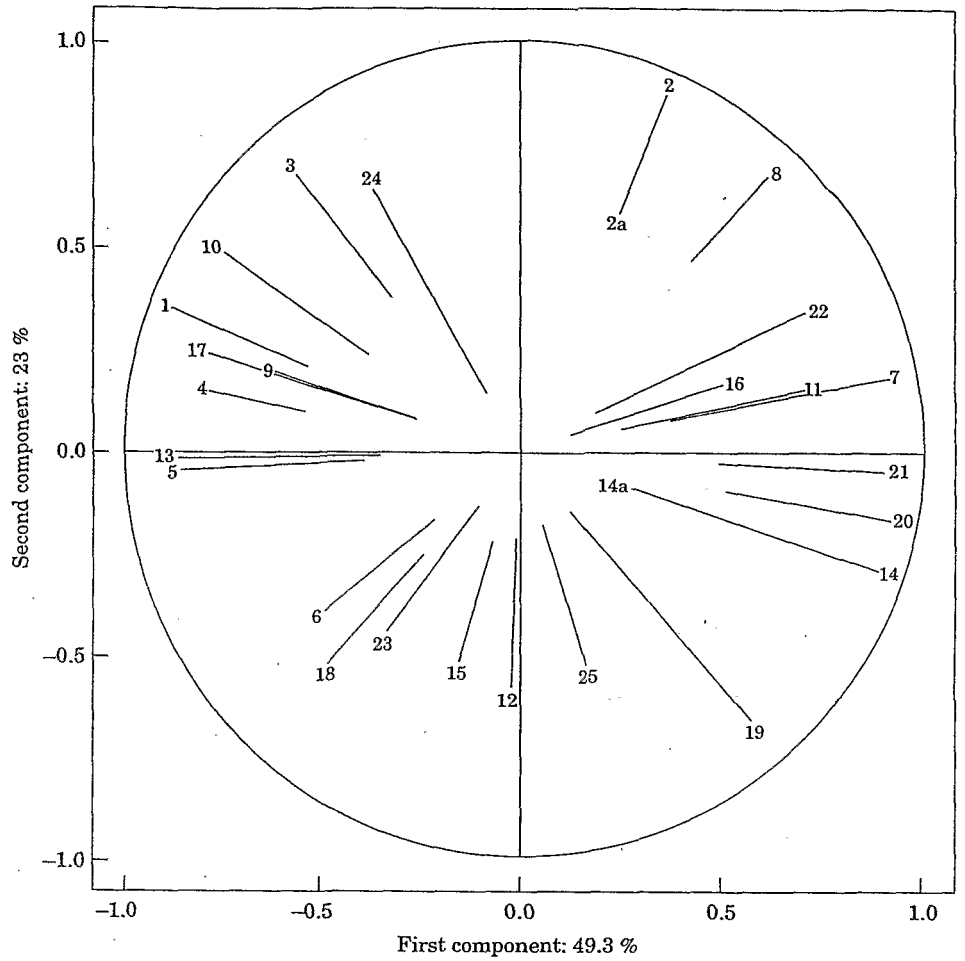

Figure 1. Principal component analysis of Y with respect to Fortnight, representation of species on the principal plane, within the correlation circle.

Method

For species l the data consist of a 816 element vector, Y^l (column l of Y), with each element y^l_{ijk} of Y^l being the transformed mean catch for port i during year j and for fortnight k.

Such data are usually represented in terms of an analysis of variance model (see for example Draper and Smith, 1981 or Arnold, 1981). This type of model is univariate (ANOVA) when applied to one species (Y^l), and multivariate (MANOVA) when applied to all of the species (Y). The model contains three main factors (Port, P; Year, A and Fortnight, F), three two-way interactions terms (Port × Fortnight, PF; Port × Year, PA; Year × Fortnight, AF) and one three-way interaction term (Year × Fortnight × Port, AFP). There is one observed catch for each combination of the factors (Port, Year, Fortnight) so that the design is balanced and the main factors and interaction terms are orthogonal.

Let us consider Y^l and the complete ANOVA model based on the three previously defined factors (i.e. including main and interaction terms). Because there is no replication the model is saturated and hence the residual term is null. Such a model allows us to decompose Y^l (see Appendix 1) into additive terms, each of these being linked to a factor or an interaction between factors:

$$Y^l = Y^l_P + Y^l_A + Y^l_F + Y^l_{PA} + Y^l_{PF} + Y^l_{AF} + Y^l_{AFP} \qquad (1)$$

For example, Y^l_F contains the fortnight effect of Y^l. Moreover, due to the orthogonality we have a similar decomposition for the variance of Y^l.

$$1 = \mathrm{var}(Y^l) = \mathrm{var}(Y^l_P) + \mathrm{var}(Y^l_A) + \mathrm{var}(Y^l_F) + \cdots + \mathrm{var}\,(Y^l_{AFP}) \qquad (2)$$

Due to the initial scaling of the columns of Y, $\mathrm{var}(Y^l) = 1$. Thus, for example, that part of the variance

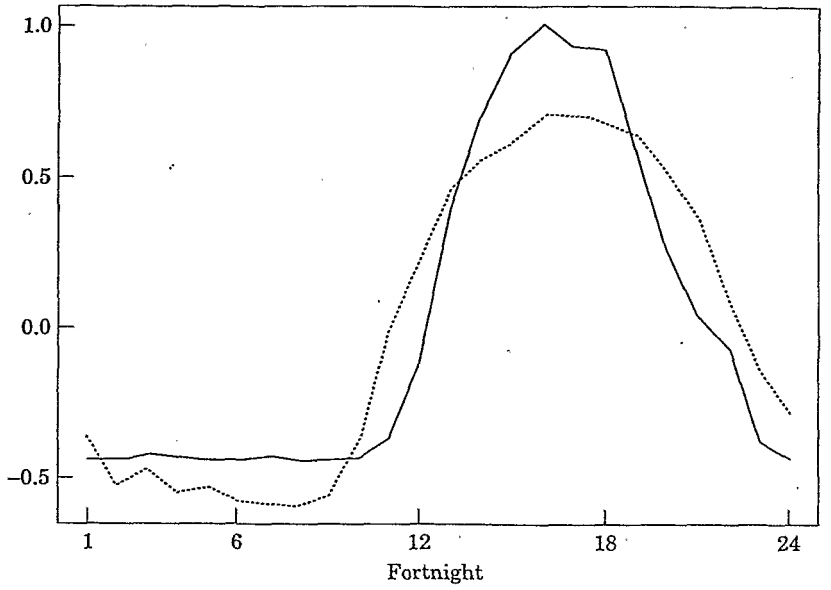

Figure 2. Fitted values from the first component of seasonal effect of *Istiophorus albicans*, —— seasonal effect, – – – fitted values.

of Y^l explained by the fortnight factor is expressed as $var(Y^l_F)$, the value of which will be between zero and one.

Considering all the species, the saturated MANOVA model generalizes the univariate decomposition presented above to,

$$Y=Y_P+Y_A+Y_F+Y_{PA}+Y_{PF}+Y_{AF}+Y_{AFP}. \quad (3)$$

Y therefore consists of a 816×25 matrix which is decomposed in a sum of 7 fitted matrices (see Appendix 1). Each of these contains the effects of the 25 species relative to a factor or interaction. Note that by considering the l^{th} column of each of those matrices, we find again the decomposition in (1).

Similar to (2), we have an additive decomposition for the variability of Y. This variability is now expressed in terms of inertia (see Appendix 1) which is the multivariate expression of the variance. The inertia of Y (I_Y) is defined as the sum of the variances over the columns, i.e. $I_Y=\Sigma^{25}_{l=1}\ var(Y^l)$. Thus, analogous to (2), we obtain,

$$25=\sum_{l=1}^{25} var(Y^l)=\sum_{l=1}^{25} var(Y^l_P)+\cdots+\sum_{l=1}^{25} var(Y^l_{AFP})$$

or

$$25=I_Y=I_P+I_A+I_F+I_{PA}+I_{PF}+I_{AF}+I_{AFP}. \quad (4)$$

The latter Equation (4) may be expressed as the term by term sum of the 25 Equations (2). For example, $I_F=\Sigma^{25}_{l=1}\ var(Y^l_F)$ expresses the part of the inertia of Y explained by the fortnight factor. Furthermore, if each of the seven sources of variation has no systematic effect at all, then all the ratios of inertia to the corresponding degree of freedom have the same expectation. Hence, it is useful to consider those ratios in order to describe the impact of the various sources of variation.

Being saturated, such a model is not explanatory because it contains as many parameters as we have data. Nevertheless, it does allow us to link the catches to the qualitative or instrumental variables (Port, Year, Fortnight or their interactions) used in the sampling design. The method used to study the relationship between catches and instrumental variables was principal component analysis with respect to instrumental variables, or PCAIV (Rao, 1964; Sabatier *et al.*, 1989).

Principal Components Analysis (PCA) is a useful tool for description of global linear correlations between variables (see, for example, Biseau and Gondeaux, 1988). However, particularly in the case of data collected according to a sampling design, it may be interesting to present an analysis of the correlations of the variables of interest conditional on the instrumental variables. PCAIV is suitable method for this purpose. In practice, a PCAIV on several variables of interest relative to an instrumental variable consists of carrying out a PCA on the fitted variables of interest after the regression on the instrumental variable (Sabatier *et al.*, 1989). This analysis can be done using any software that has both general linear models and PCA (e.g. SAS, S-PLUS, Genstat).

As in (3), we decomposed Y into a sum of seven fitted matrices. Each of them is linked to a factor or interaction between factors identified here as instrumental variables. Such a decomposition may be useful for

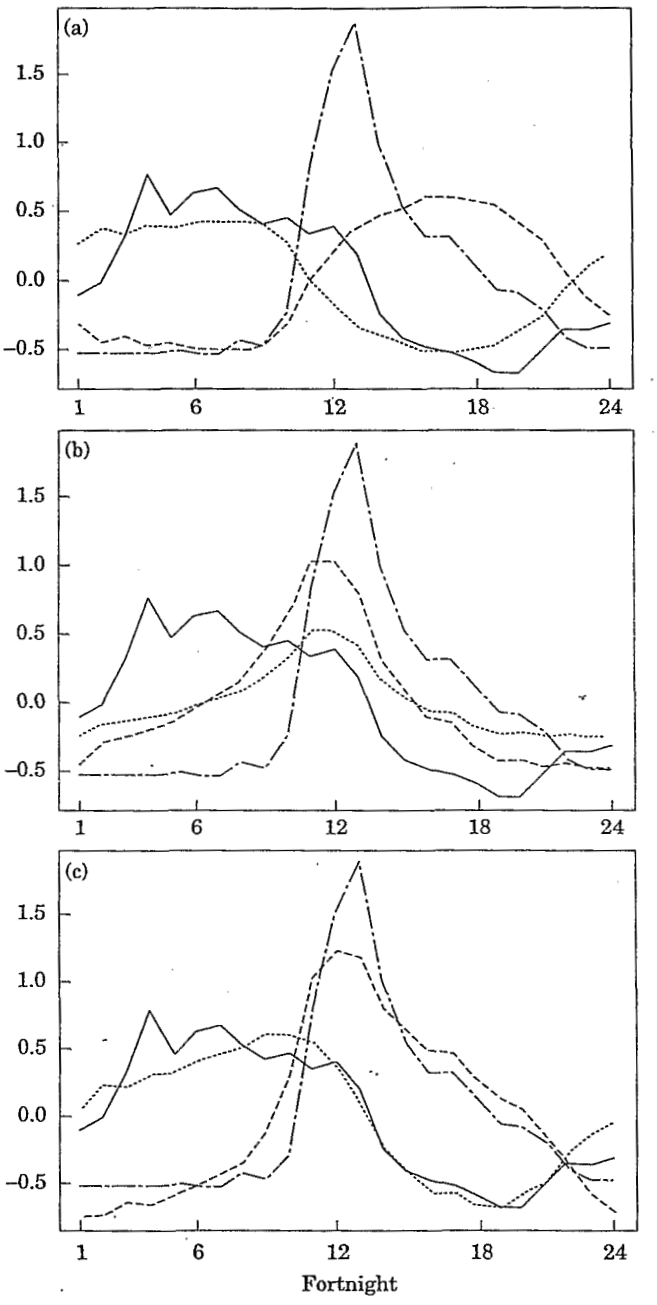

Figure 3. (a) Fitted values from using the first component of seasonal effect of *Alectis alexandrinus* (species 8) and *Epinephelus guaza* (species 10). (b) Fitted values from using the second component of seasonal effect of *Alectis alexandrinus* (species 8) and *Epinephelus guaza* (species 10). (c) Fitted values from using the two first components of seasonal effect of *Alectis alexandrinus* (species 8) and *Epinephelus guaza* (species 10). For each graph, —— seasonal effect of species 10, · · · · fitted values for species 10, – · – · seasonal effect of species 8, – – – fitted values for species 8.

interpretation purposes. For example, Y_F contains the seasonal effects with respect to the yields and, due to the orthogonality, those effects may be discussed independently of other sources of variation. Hence, PCA of each of these arrays will allow us to describe the relations between yields of the species for each of the seven sources of variations defined by the saturated model.

Graphical outputs are very useful for the presentation and interpretation of results. We shall focus on fitted values of the observations from multiple regression on

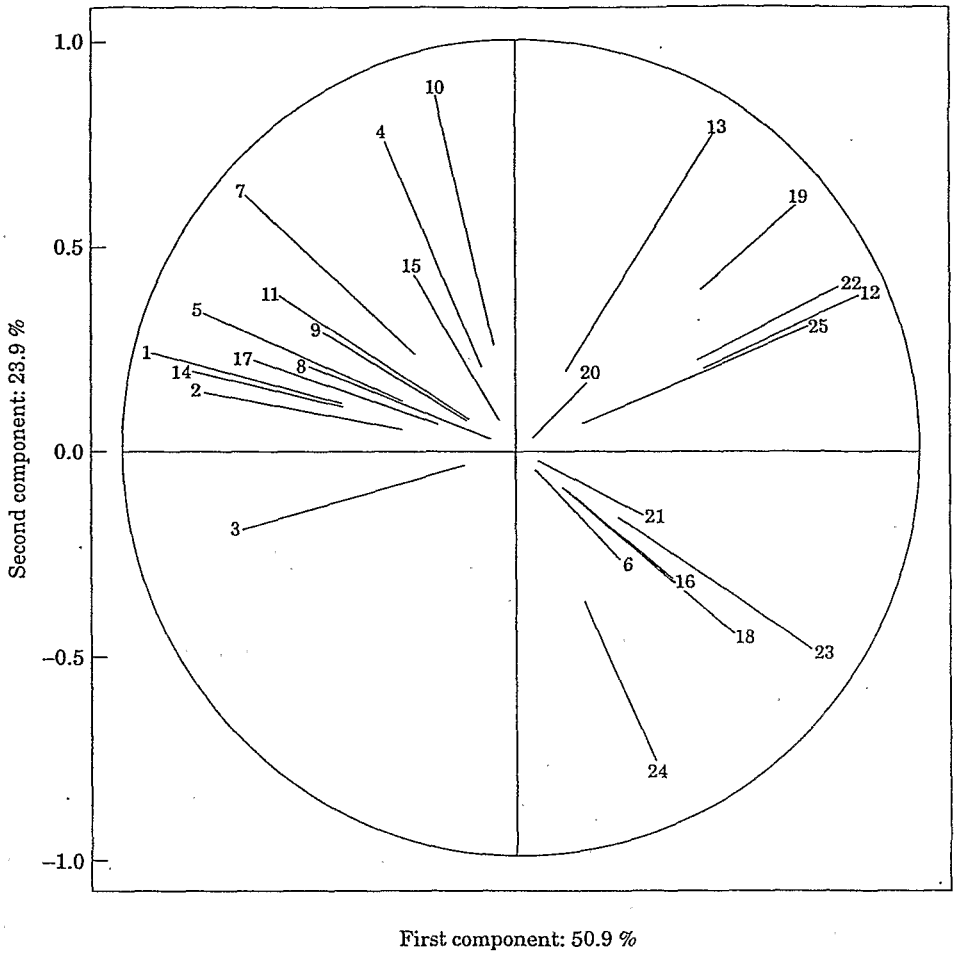

Figure 4. Principal component analysis of Y with respect to Year, representation of species on the principal plane, within the correlation circle.

principal components (Persat and Chessel, 1989) and correlation of the variables of interest with the principal planes. These fitted values will be described in connection with the presentation of results of the PCA of Y_F (i.e. the PCAIV of Y with respect to the factor Fortnight).

Results

Decomposition of inertia

The decomposition of the inertia of Y according to Equation (4) is presented in Table 2. We can distinguish two groups here. The first one (Port, Year, Fortnight and interactions Port × Fortnight and Port × Year) characterizes terms with low degree of freedom (df), strong inertia and high ratio of inertia to df. The second group (interactions Year × Fortnight and Port × Year × Fortnight) comprises terms with high degrees of freedom, whose ratio of inertia to df is low. This latter group accounts for about 47% of the total inertia.

PCAIV of Y for Fortnight

Representation of the variables in the principal plane

Each PCAIV concerns a fitted matrix whose columns are not reduced. Hence, their variances, which belong to [0,1], express the importance of the considered factor. Figure 1 presents the variables in the plane of the first two principal components for PCAIV for Y Fortnight (i.e. the PCA of Y_F). Species are indicated by numbers (1 to 25, see Table 1 for species names) with an associated line segment of length s_l, where $l=1, \ldots, 25$. The ratio of $(d_l - s_l)/d_l$, where d_l is the distance from species number l to the centre of the circle, is equal to the square root of the variance explained by the instrumental variable (e.g. for Fortnight, $\mathrm{var}(Y^l_F)^{1/2}$). Therefore the representation

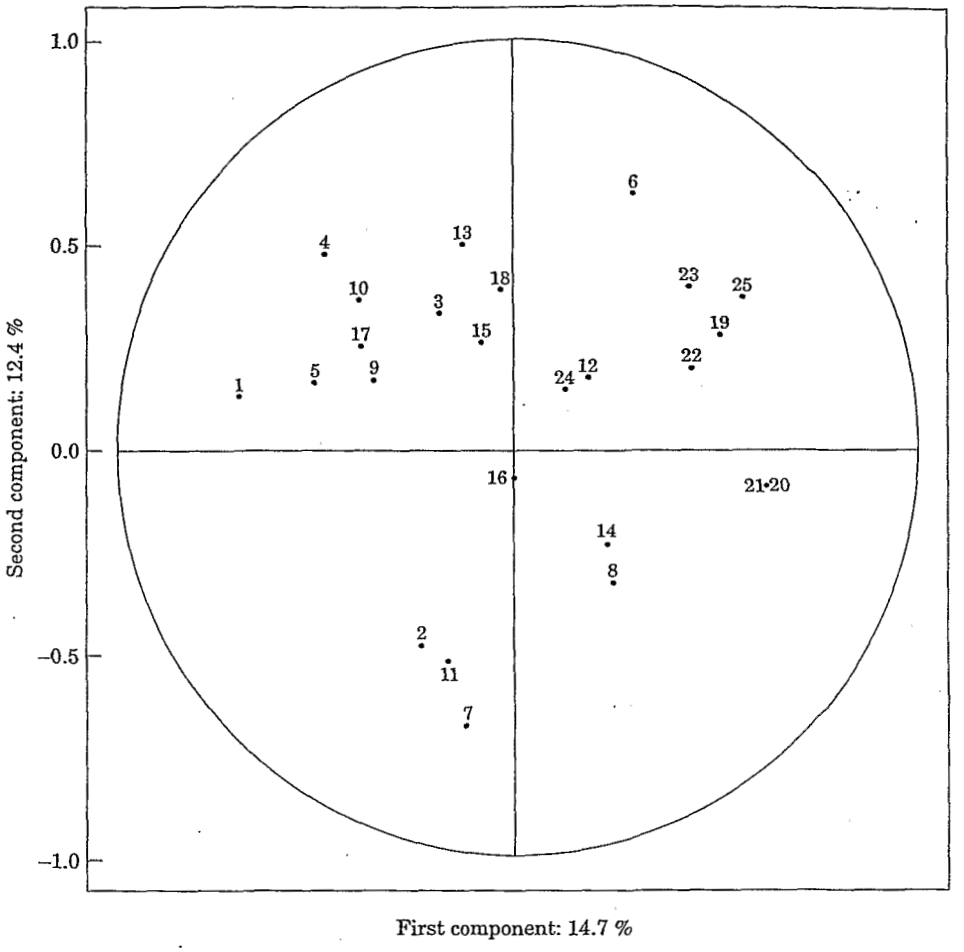

Figure 5. Principal component analysis of Y, representation of species on the principal plane, within the correlation circle.

of the variables by the principal plane in Figure 1 contains information about:

(1) the quality of the representation by the principal plane with points closer to the circumference of the circle being better represented (i.e. higher correlation) and,
(2) the influence of the instrumental variable being considered which is measured by the length s with shorter lengths indicating a higher proportion of the variance being explained.

For example, let us consider (Fig. 1) species 14 (*Sphyrna* spp.) and 2 (*Pagrus caeruleostictus*). The positions of items 14 and 2 indicate that the seasonal effects of those species are well represented by the principal plane. However, extremity 2a is nearer to 2 than 14a to 14. This difference corresponds to the differing importance of fortnight variabilities. Indeed, the factor fortnight explains about 45% of the variance of the *Pagrus caeruleostictus*, and only 9% for the *Sphyrna* spp. Hence, consideration of both extremities of the segment allows us to make an analysis taking into account the quantitative influence of the factor on each of the species.

On the whole (Fig. 1), many of the species have important fortnight effects. We distinguish an opposition between cold season species (1, 3, 4, 5, 10, 13) and warm season species (7, 8, 20, 21). A positive correlation with the second component is interpreted here as indicating a seasonal effect which extends past the cold season (species 10, *Epinephelus guaza*) or anticipates the warm season (species 8, *Alectis alexandrinus*). This interpretation may be illustrated by looking at the fitted values from using principal components.

Fitted values from principal components

We obtained fitted values from principal components by applying a multiple regression of Y^l_F on the first principal components. This allows one to substitute a smoothed image for Y^l_F taking into account the structure of all the data (Persat and Chessel, 1989).

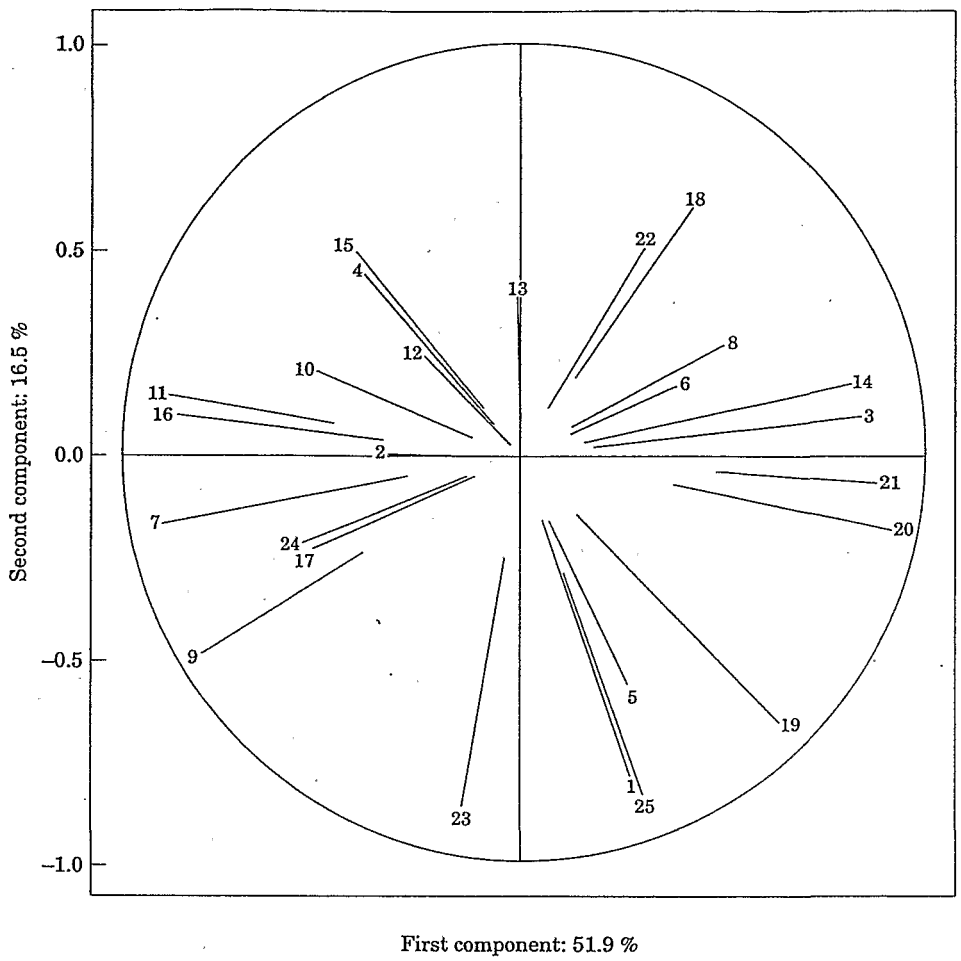

Figure 6. Principal component analysis of Y with respect to interaction Port × Fortnight, representation of species on the principal plane, within the correlation circle.

The seasonal effect for *Istiophorus albicans* (species 21) is highly correlated with the first principal component (Fig. 1). Therefore model fitting with this component is very efficient (Fig. 2) and allows a clear interpretation of this component and of the seasonal effect for *Istiophorus albicans*. This is also true for the seasonal effects for the other species that are highly correlated (positively or not) with the first component. This first principal component is characteristic of the succession of seasons, a warm and rainy season from July to October (fortnights 13–20) and a cold and dry season from December to May (fortnights 23, 24 and 1–10) – June (fortnights 11, 12) and November (fortnights 21, 22) being "inter-seasonal". The cold and dry season is also characterized by the presence of an upwelling phenomenon (Rébert, 1983).

The seasonal effects of mean catches for *Alectis alexandrinus* (species 8) and *Epinephelus guaza* (species 10) are combinations of the two first principal components (Fig. 3a, b, c); the first species is mainly caught during the warm season, with high values observed in the inter-seasonal month of June (fortnight 11 and 12). The second species is mainly caught during the cold season, also with high values in June. This characteristic is taken into account by the second component which presents a peak during June and July (Fig. 3b). We may note on Figure 3a and b that the contributions of first component are in opposition and that the contributions of the second are quite similar.

Species whose code number or item is not close to the correlation circle (Fig. 1) are not well correlated with a combination of the first two components. For such species (for example species 12, 15, 23, 25), a useful model fit would require more than two components.

PCAIV of Y for Year

The principal plane (Fig. 4) explains about 75% of the inertia of Y_A. The variables whose year effect is strong are generally well fitted by the model. We can distinguish three groups of variables: species (23, 24) (with an

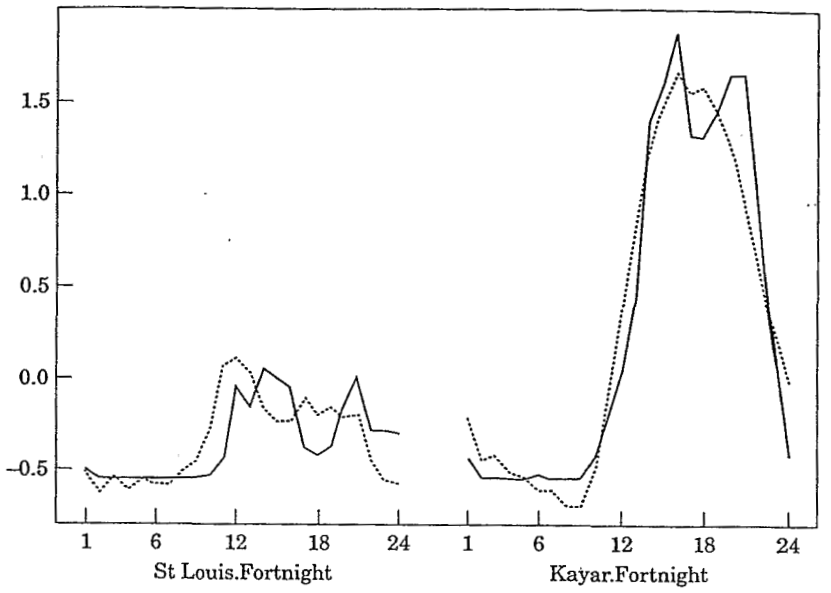

Figure 7. Fitted values from using the interaction Port × Fortnight by aid of the first component for *Coryphaena hippurus*, including main effects. —— interaction, ····· fitted.

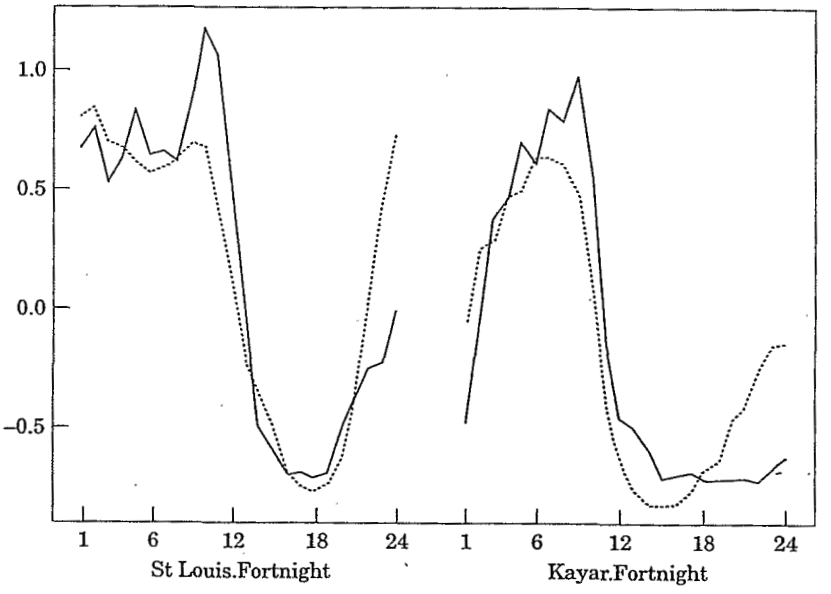

Figure 8. Fitted values from using the interaction Port × Fortnight by aid of the two first components for *Pomatomus saltatrix*, including main effects. —— interaction, ····· fitted.

increasing year effect) are in opposition to species (1, 5, 7, 14) whose year effect is decreasing. These two groups are orthogonal to species (12, 13, 19, 22), whose year effect first increases and then decreases.

If we now consider the general PCA of Y (Fig. 5), and the PCA of Y_F and Y_A, we may observe one of the interesting insights from PCAIV. On the general PCA for example, species 1 (*Pomatomus saltatrix*) and 7 (*Arius latisculatus*) appear to be quite orthogonal ($r = -0.06$). Considering the PCA of Y_F (Fig. 1), shows that their seasonal effects are opposed (one cold season species versus warm season species), while PCA of Y_A (Fig. 4) shows that their inter-annual variabilities are actually positively correlated.

PCAIV of Y for Port

The qualitative variable Port having two modalities and therefore only generates a subspace of dimension 1. That

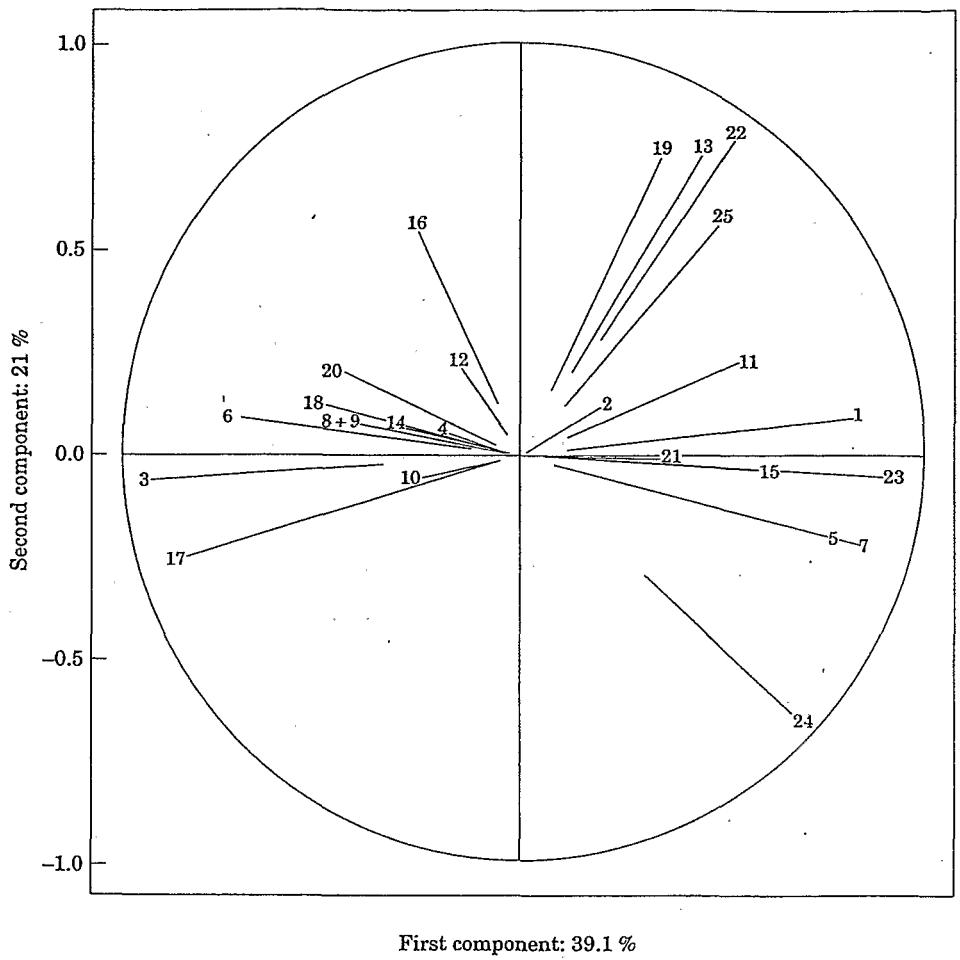

Figure 9. Principal component analysis of Y with respect to interaction Port × Year, representation of species on the principal plane, within the correlation circle.

is why we do not show the representation in the principal plane. We can distinguish St Louis's species (7, 1) from Kayar's species (6, 20, 21, 23, 25).

PCAIV of Y for Port × Fortnight

The influence of this factor on species is variable (Fig. 6). It is difficult to interpret the interactions without taking into account the main effects. As an example, consider the following representations for species 20 (*Coryphaena hippurus*) and species 1 (*Pomatomus saltatrix*). The global fitting of species 20 was constructed as the regression on first component of PCA on Fortnight, Port and Port × Fortnight (Fig. 7). For species 1, we observe that interaction Port × Fortnight is principally correlated with the second principal component. So, we may obtain fitted values (Fig. 8) for this species with first component of PCA of Y_F, the only component of PCA of Y_P, and the two first components of PCA of Y_{PF}.

We see that the interactions may reflect different situations. For *Coryphaena hippurus*, the catches are mainly made at Kayar during the warm season; catches are quite small during the cold season in the two ports. For *Pomatomus saltatrix*, the interaction highlights a possibly more interesting situation, with catches made during a longer period of the cold season in St Louis. This is in agreement with available knowledge on the migratory pattern of that species (Champagnat *et al.*, 1983).

PCAIV of Y for Year × Port

The influence of this interaction is weak (Fig. 9). Only species 24 (*Octopus vulgaris*) has a significant interaction. Indeed, the exploitation of this species is quite recent and takes place mainly at Kayar.

PCAIV for Y for Year × Fortnight and Port × Fortnight × Year

Inertias corresponding to these effects are strong, but with a low ratio inertia/df (Table 2). Contrary to previous PCAIV, inertia is spread out over the principal

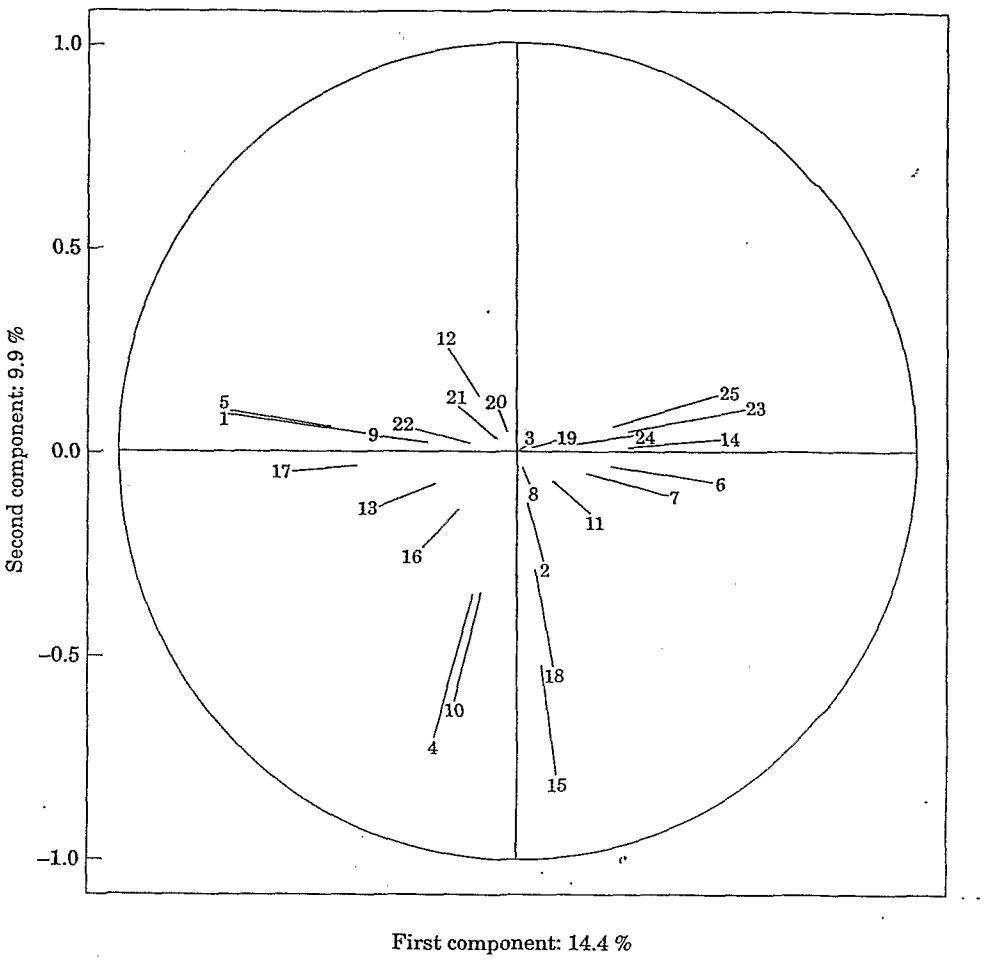

Figure 10. Principal component analysis of Y with respect to interaction Year × Fortnight, representation of species on the principal plane, within the correlation circle.

components and no one or two of the components dominate enough to represent typical behaviours (Figs 10 and 11).

A global model

In preceding sections, partial models for fitted values obtained from the principal components were considered. Taking into account a small number of principal components, we may fit the model in a quite satisfying way with each term of the decomposition as given in Equation (3). It is important to look at the data i.e. at the matrix Y itself. This may be done by pooling the partial results and by summing the results for each fitted matrix. For example, using the one component for the Port effect and three components for the six other sources of variation (Table 3). We present this model fit with species 1 (*Pomatomus saltatrix*) in Figure 12b. Such a model fit may be done for each species. This kind of model should be considered to be a non-parametric model because it was constructed as a linear combination of a limited number of smoothed series (i.e. principal components).

As an alternative to the above, a parametric model may be obtained by fitting multivariate linear models. The best-fitting model can be selected among a greater number of possibilities using a criterion derived from the Akaike's information criterion (Hurvitch and Tsai, 1989; Sakamoto *et al.*, 1986) and adapted for multivariate models (Bedrick and Tsai, 1994; see Appendix 2). Among 165 possible linear models, we selected the following expression:

$$Y = \text{Port} + P(A) + H3 + \text{Port} \times P(A) + \text{Port} \times H3 + P(A) \times H3 + \text{Port} \times P(A) \times H3 + \varepsilon$$

where H3 is a set of six sine and cosine functions on the fortnight number of respective periods 24, 12 and 8;

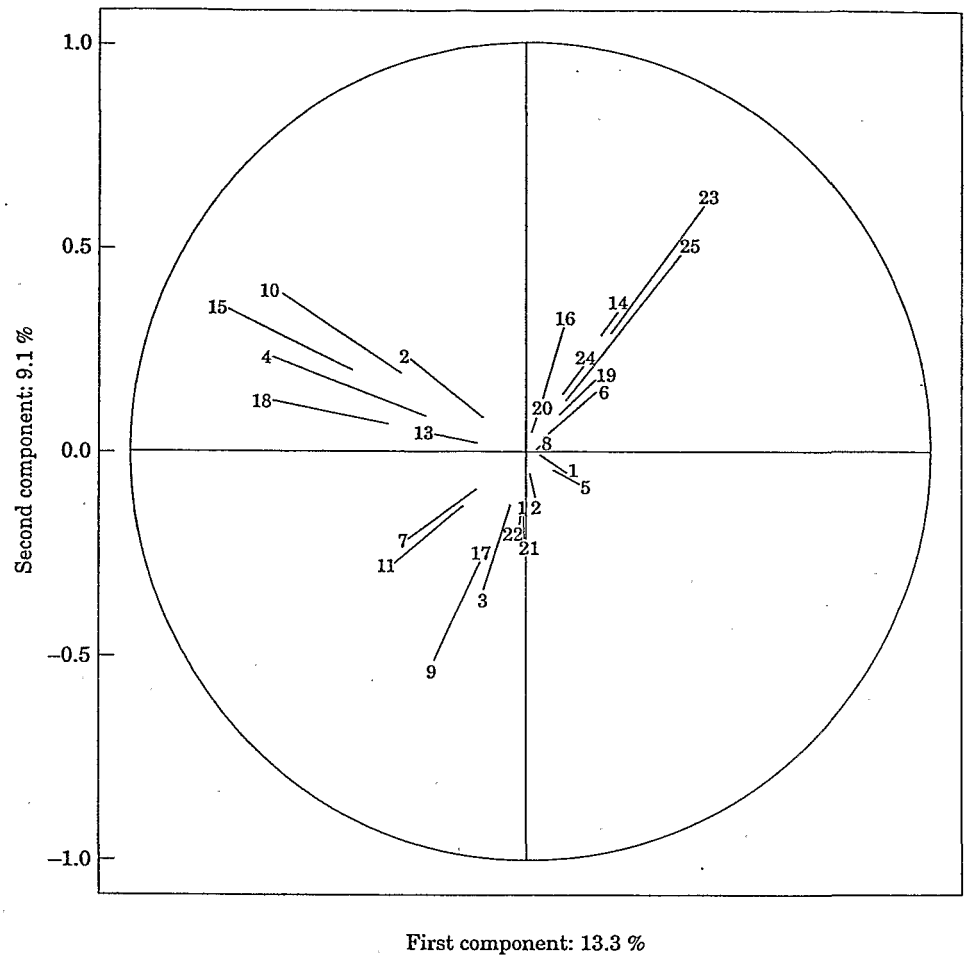

Figure 11. Principal component analysis of Y with respect to interaction Port × Year × Fortnight, representation of species on the principal plane, within the correlation circle.

Table 3. Decomposition of the inertia according to orthogonal subspaces induced by principal components. This variability is expressed in terms of inertia (see Appendix 1) which is the multivariate expression of the variance.

Source of variation	Number of components	Inertia	Inertia/initial inertia
Port	1	1.87	1
Year	3	2.42	0.81
Fortnight	3	4.13	0.84
Port × year	3	0.92	0.69
Port × fortnight	3	1.87	0.83
Year × fortnight	3	2.06	0.31
Year × port × fortnight	3	1.47	0.29

P(A) is a polynomial of degree 6 on the year number. This model has 97 degrees of freedom, and the decomposition of sum of squares is given in Table 4. However, this model must be not considered as completely optimal because assumptions of homogeneity of variances and normality do not hold. Fitted values from this model are also given for *Pomatomus saltatrix* on Figure 12b.

The parametric and non-parametric versions of the analysis are compared by considering Tables 3 and 4, and Figure 12b. Furthermore, values obtained by the two methods are similar for each species, as shown by the quite high correlation values (Table 5).

The versatility of the two methods is illustrated in Figure 12 where we present the original data set, the fitted values described above and the results of two partial models, of potential interest for *Pomatomus saltatrix*. In Figure 12c we show the fitted values using one component for the Port effect and three components for the two other main effects and interaction Port × Fortnight. We also present in Figure 12c the fitted values obtained from the linear model:

$$Y = \text{Port} + P(A) + H3 + \text{Port} \times H3 + \varepsilon$$

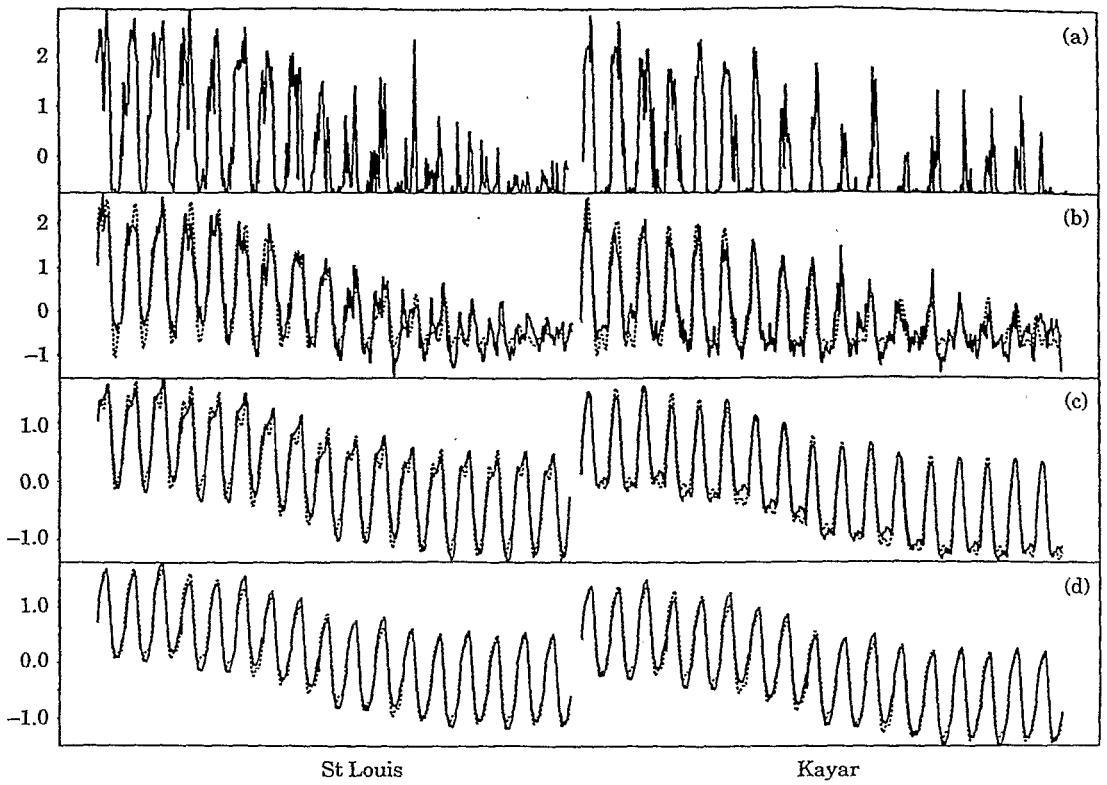

Figure 12. Yields of *Pomatomus saltatrix*. Fitted values from principal components (continuous lines) and fitted values with linear models (dotted lines). (a) Observed yields. (b) Fitted yields from PCAIV and the linear model with all the sources of variations (see text). (c) Fitted yields from PCAIV and the linear model with the three main effects and interaction Port × Fortnight. (d) Fitted yields from PCAIV and the linear model with the three main effects only.

Table 4. Decomposition of the inertia according to orthogonal subspaces induced by the selected model. This variability is expressed in terms of inertia (see Appendix 1) which is the multivariate expression of the variance.

Source of variation	Df	Inertia	Inertia/df	Inertia/initial inertia
Port	1	1.87	1.87	1
P(A) (year)	6	2.43	0.40	0.82
H3 (fortnight)	6	4.56	0.76	0.93
Port × *P(A)*	6	0.99	0.17	0.74
Port × H3	6	2.02	0.33	0.89
P(A) × H3	36	2.22	0.06	0.33
P(A) × port × H3	36	1.48	0.04	0.28
Residuals	718	9.43	0.01	
Total	815	25	0.03	

Figure 12c includes the previously discussed (see Fig. 8) results on the difference between intra-annual yield patterns in Kayar and Saint-Louis, together with the inter-annual trend of those yields.

The results given in Figure 12d were obtained by dropping the components relative to the interaction Port × Fortnight in the model fitting procedure for the principal components and the term Port × H3 in the linear model.

Both methods give quite similar results in the three cases. Results presented in Figure 12 may be used to illustrate the nature of what is taken into account from the original data when using one source or a combination of sources of variation.

Table 5. Correlation for each species between adjusted variables for both models (first column); between adjusted variables for the parametric model and initial data (second column); between adjusted variables for the non parametric model and initial data (third column).

Species	Cor(Yp, Ynp)	Cor(Yp, Y)	Cor(Ynp, Y)
1	0.92	0.91	0.89
2	0.86	0.86	0.79
3	0.83	0.82	0.76
4	0.83	0.81	0.86
5	0.80	0.76	0.88
6	0.88	0.88	0.82
7	0.92	0.81	0.78
8	0.88	0.79	0.73
9	0.84	0.80	0.74
10	0.73	0.70	0.79
11	0.86	0.77	0.71
12	0.76	0.77	0.66
13	0.77	0.66	0.66
14	0.86	0.86	0.77
15	0.60	0.62	0.83
16	0.75	0.65	0.69
17	0.65	0.64	0.75
18	0.70	0.66	0.74
19	0.85	0.77	0.79
20	0.91	0.80	0.74
21	0.94	0.89	0.84
22	0.87	0.78	0.77
23	0.86	0.88	0.83
24	0.73	0.95	0.72
25	0.85	0.83	0.82

Our intention is not to choose one model over the other. Indeed, the non-parametric model can be considered to be a more parsimonious summary of the data than the parametric model. However, the components of the non-parametric model are linear combinations of the yields fitted using instrumental variables. Hence, they form a summary of the influences of those factors. Such a summary may be considered an "*ad hoc* smoothed" transformation of the factors. So, principal components cannot be considered regressor variables in the usual sense and we cannot consider the non-parametric model as to be a classical multiple regression model with 19 independent variables.

Discussion and conclusion

Descriptions made by means of PCAIV helped identify the most important sources of variation among those defined by the sampling design. Such sources of variation have been described by fitting the model with the major factors with the aid of principal components.

The initial data set Y may be described in a satisfying way by using a few components in the model. A similar model fit may also be obtained by selecting a parametric model from a family of candidate models. Both appear to be equivalent in the sense that principal data characteristics are taken into account.

Such models may be considered with some criticism with respect to violations of assumptions usually required for classical inference (independence of the residuals, homogeneity of their variances). However, because they capture some major characteristics of our initial data set, we can think that they have been formulated according to the principle expressed by Lebreton *et al.* (1992):

> "We approach data analysis in this spirit: we want to find an useful model that correctly represents the biologically important structure that is real in the data. We may be unable to ferret out the correct form of the more subtle structure in the data. In this case, we believe it is appropriate to *sweep* this residual structure into the model error component."

The analysis presented in this paper is exploratory, not explanatory. It does not result in a model of population dynamics nor of fleet dynamics. We have only tried to give a parsimonious synthesis of the spatio-temporal variability. This variability represents many sources of variation in environmental conditions of fish species and fishermen and from interactions between such sources. Hence, further models are needed and the results presented here should be considered as possible frameworks for analysis of outputs of simulation models. The use of "multi-species–multi-fleet" models appears to be necessary in the context of many fisheries (Garrod, 1973; Gulland and Garcia, 1984; Hilborn, 1985) and such models have been used in the Senegalese case (Laloë and Samba, 1990, 1991; Lefur, 1995). Model outputs have to be compared with available data sets for tuning and validation purposes, and the methods presented here offer tools for describing the salient features of the available data.

Acknowledgements

We thank Drs D. Toure and A. Samba of the CRODT at Dakar. We also thank Drs R. Sabatier and J. D. Lebreton and two anonymous referees. We are very grateful to assistant editor S. J. Smith for very valuable help on many aspects of this work.

References

Allen, P. M., and Mac Glade, J. M. 1986. Dynamics of discovery and exploitation, the case of Scotian shelf groundfish fishery. Canadian Journal of Fisheries and Aquatic Science, 43: 1187–1200.

Arnold, S. F. 1981. The theory of linear models in multivariate analysis. John Wiley & Sons, New York, USA.

Bedrick, E. J., and Tsai, C. L. 1994. Model selection for multivariate regression in small samples. Biometrics, 50: 226–231.

Biseau, A., and Gondeaux, E. 1988. Apport des méthodes d'ordination en typologie des flotilles. Journal du Conseil International pour Exploration de la Mer, 44: 286–296.

Champagnat, C., Caverivière, A., Conand, C., Cury, P., Durand, J. R., Fontana, A., Fonteneau, A., Fréon, P., and Samba, A. 1983. Pêche, biologie et dynamique du tassergal (*Pomatomus saltator*, Linaeus, 1766) sur les côtes sénégalo mauritaniennes. Trav. Doc. ORSTOM Paris, 168: 279pp.

Cury, P., and Roy, C. 1988. Migration saisonnière du thiof (*Epinephelus aenus*) au Sénégal: influence des upwellings sénégalais et mauritaniens. Oceanologica Acta, 11: 25–36.

Cury, P., and Roy, C. (eds) 1991. Pêcheries Ouest Africaines. Variabilité, instabilité, changement. Orstom, Paris, 525pp.

Draper, N., and Smith, H. 1981. Applied regression analysis. John Wiley & Sons, New York, USA.

Fay, C. 1994. Organisation sociale et culturelle de la production de pêche: morphologie et grandes mutations. *In* La pêche dans le delta central du Niger, pp. 191–207. Ed. by J. Quensière. I. E. R/ORSTOM/KARTHALA, Paris.

Ferraris, J., and Samba, A. 1992. Variabilité de la pêche artisanale Sénégalaise et statistique exploratoire. Seminfor 5, ORSTOM, september 1991. Montpellier: 169–190.

Ferraris, J., Fonteneau, V., and Sy Bo, A. 1993. Structuration de la base de donnees "pêche artisanale" et chaîne de traitement informatique. Arch. C.R.O.D.T., 39pp+annexes.

Fréon, P. 1986. Réponses et adaptation des stocks de cupleidés d'Afrique de l'ouest à la variabilité du milieu et de l'exploitation. Analyse et réflexion à partir de l'exemple du Sénégal. Thèse doctorat d'état. Université Aix Marseille II.

Garrod, D. J. 1973. Management of multiple resources. Journal of the Fisheries Research Board of Canada, 30: 1977–1985.

Gérard, M., and Greber, P. 1985. Analyse de la pêche artisanale au cap vert: description et étude critique du système d'enquête. Doc. Scient. Cent. Rech. Océano. Dakar Thiaroye, 98: 77pp.

Gulland, J., and Garcia, S. 1984. Observed patterns in multispecies fisheries. *In* Exploitation of marine communities, pp. 155–190. Ed. by R. M. May. Dahlem Konferenzen, Springer Verlag.

Hilborn, R. 1985. Fleet dynamics and individual variation: why some people catch more fish than other. Canadian Journal of Fisheries and Aquatic Science, 42: 2–13.

Hurvitch, C. M., and Tsai, C. L. 1989. Regression and time series model selection in small samples. Biometrika, 76: 297–307.

Inzenman, A. J. 1980. Assessing dimensionality in Multivariate regression. *In* Handbook of statistics, Vol. 1, Analysis of variance, pp. 571–591. Ed. by P. R. Krishnaiah. North-Holland, Amsterdam.

Laloë, 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 C.R.O.D.T., Doc. Sci. Cent. Rech. Océano. Dakar Thiaroye, 100: 36pp.

Laloë, F., and Samba, A. 1990. La pêche artisanale au Sénégal: ressources et stratégies de pêches. Collection études et thèses, ed. ORSTOM, Paris.

Laloë, F., and Samba, A. 1991. A simulation model of artisanal fisheries of Senegal. ICES Marine Symposium, 193: 281–286.

Laurec, A., Biseau, A., and Charuau, A. 1991. Modelling technical interactions. ICES Marine Symposium, 193: 225–234.

Lebreton, J. D., Sabatier, R., Banco, G., and Bacou, A. M. 1991. Principal component and Correspondences 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, pp. 85–114. Ed. by J. Devillers and W. Karcher. Kluwer.

Lebreton, J. D., Burnham, K. P., Clobert, D., and Anderson, D. R. 1992. Modelling survival and testing biological hypotheses using marked animals: a unified approach with case studies. Ecological Monographs, 62: 67–118.

Lefur, J. 1995. Modeling adaptative fishery activities facing fluctuating environment: An AI approach. AI Applications, 9: 85–97.

Murawski, S. A. 1984. Mixed species yield per recruit analyses accounting for technological interactions. Canadian Journal of Fisheries and Aquatic Science, 41: 897–916.

Persat, H., and Chessel, D. 1989. Typologies de distributions en classes de taille: interêt dans l'étude des populations de poisons et d'invertébrés. Acta Oecologica, 10: 175–195.

Rao, C. R. 1964. The use and interpretation of Principal Component Analysis in applied research. Sankhya Series, 26: 329–358.

Rébert, J. P. 1983. Hydrobiologie et dynamique des eaux du plateau continental sénégalais. Doc. Scient. Cent. Rech. Oceano. Dakar-Thiaroye, 89: 99pp.

Sabatier, R., Lebreton, J. D., and Chessel, D. 1989. Principal component analysis with instrumental variables as a tool for modelling composition data. *In* Multiway data analysis, pp. 341–352. Ed. by R. Coppi and S. Bolasco. Elsevier, Amsterdam.

Sakamoto, Y., Ishiguro, M., and Kitagawa, G. 1986. Akaike information criterion statistics. KTK Scientific Publishers, Tokyo, Japan.

Samba, A., and Laloë, F. 1991. Upwelling Sénégalo-Mauritanien et pêche au tassergal (*Pomatomus saltator*) sur la côte nord du Sénégal. *In* Pêcheries Ouest Africaines. Variabilité, instabilité, changement, pp. 307–310. Ed. by P. Cury and C. Roy. ORSTOM, Paris.

Appendix 1: the decomposition of matrix Y and the inertia

Let us consider Y^l. The complete model based on the three previously defined factors (i.e. including main and interaction terms) may be written as an addition of effects:

$$Y^l_{ijk}=\mu^l+p^l_i+a^l_j+f^l_k+pa^l_{ij}+pf^l_{ik}+af^l_{jk}+afp^l_{jki}+\varepsilon_{ijk}$$

where i=1, 2 refers to the port; j=1, . . ., 17 refers to the year; k=1, . . ., 24 refers to the fortnight.

With usual notations, estimators of parameters are (cf. Draper and Smith, 1981, p. 446)

$$\hat{\mu}^l=Y\ldots$$
$$\hat{p}^l_i=Y_{i\cdot\cdot}-Y\ldots$$
$$\hat{a}^l_j=Y_{\cdot j\cdot}-Y\ldots$$
$$\hat{f}^l_k=Y_{\cdot\cdot k}-Y\ldots$$
$$\widehat{pa}^l_{ij}=Y_{ij\cdot}-Y_{i\cdot\cdot}-Y_{\cdot j\cdot}+Y\ldots$$
$$\widehat{pf}^l_{ik}=Y_{i\cdot k}-Y_{i\cdot\cdot}-Y_{\cdot\cdot k}+Y\ldots$$
$$\widehat{af}^l_{jk}=Y_{\cdot jk}-Y_{\cdot j\cdot}-Y_{\cdot\cdot k}+Y\ldots$$
$$\widehat{paf}^l_{ijk}=Y_{ijk}-Y_{ij\cdot}-Y_{i\cdot k}-Y_{\cdot jk}+Y_{i\cdot\cdot}+Y_{\cdot j\cdot}+Y_{\cdot\cdot k}-Y\ldots$$

(note that $\hat{\mu}^l=Y\ldots$ equals zero because Y^l has been centred.)

Now, each of the seven matrices in decomposition

$$Y=Y_P+Y_A+Y_F+Y_{PA}+Y_{PF}+Y_{AF}+Y_{AFP}$$

is obtained by the estimations of the effects given above.

For example, the element of line ijk and column l of matrix Y_F

$$y^l_{F_{ijk}}=\hat{f}^l_k=y^l_{..k},j$$

The total inertia is obtained from:

$$I_Y=\sum_l\left[\sum_{ijk}\frac{(Y^l_{ijk}-Y^l_{...})^2}{816}\right]$$

Inertia for each source of variation is obtained in a similar way. For example we have for fortnights:

$$I_F=\sum_l\left[\sum_{ijk}\frac{(Y^l_{...k}-Y^l_{...})^2}{816}\right]$$

Note that the additivity of inertias (cf. Equation 4) stems from that of the sum of squares decomposition in a balanced design.

Appendix 2: the AICc criterion

Let us consider the multivariate regression model:

$$Y=XB+U$$

where $Y_{n\times p}$ corresponds to p response variables on each of n individuals, $X_{n\times m}$ is a known matrix of covariate values, and $B_{m\times p}$ is a matrix of unknown regression parameters. The rows of the error matrix $U_{n\times p}$ are assumed to be independent, with identical $N_p(0,\Sigma)$. Maximum likelihood estimators for B and Σ are $\hat{B}=(X'X)^{-1}X'Y$, and $\hat{\Sigma}=Y'(I-X(X'X)^{-1}X')Y/n$.

The AICc value for model (1) is then defined as:

$$AICc=n\log|\hat{\Sigma}|+dp(n+m)$$

where $d=n/(n-(m+p+1))$.