# STATISTIQUE ET ANALYSE DES DONNÉES

## CHRISTINE LAVIT

# Application de la méthode STATIS

Statistique et analyse des données, tome 10, nº 1 (1985), p. 103-116

<a href="http://www.numdam.org/item?id=SAD\_1985\_\_10\_1\_103\_0">http://www.numdam.org/item?id=SAD\_1985\_\_10\_1\_103\_0</a>

© Association pour la statistique et ses utilisations, 1985, tous droits réservés.

L'accès aux archives de la revue « Statistique et analyse des données » implique l'accord avec les conditions générales d'utilisation (http://www.numdam.org/legal.php). 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.



Article numérisé dans le cadre du programme Numérisation de documents anciens mathématiques http://www.numdam.org/ Statistiques et Analyse de Données 1985 Vol. 10 n° 1 pp. **103 - 116** 

## APPLICATION DE LA METHODE STATIS

Christine LAVIT
Unité de Biomêtrie
9, Place Viala
34060 MONTPELLIER CEDEX
France

Résumé: STATIS est une méthode d'Analyse de Données permettant d'extraire l'information contenue dans un ensemble de tableaux individus x variables sous forme de graphiques: résumé global par un nuage de points-tableaux, position compromis des individus dans un système d'axes interprétables à l'aide des variables, évolution de chaque individu autour de sa position compromis dans ce même système d'axes.

Abstract: Statis is a DATA Analysis method which extracts information from the data - when it is a set of individuals x variables matrices - and gather it by means of graphic output. We get an overall description on a plot where each point represents a matrix, a plot of compromise individuals on two or three axes which can be explained by the variables and, in addition in these axes, we can follow each individual's evolution around its compromise position.

Mots clés : Analyse de données - Analyse en Composantes Principales.

Indices de classification STMA : 06-010, 06-070

## O - INTRODUCTION

Lorsque les données se présentent sous la forme d'un ensemble de tableaux : individus x variables quantitatives la méthode STATIS permet d'en extraire l'information sous forme de graphiques:résumé global, position compromis des individus dans un système d'axes interprétables à l'aide des variables, évolution de chaque individu autour de sa position compromis dans ce même système d'axes. Si on préfère suivre l'évolution des variables, on utilisera la méthode duale.

Ainsi les trois premiers objectifs donnés dans l'introduction d'Yves Escoufier sont atteints. Dans la plupart des exemples c'est l'objectif 4 que l'on recherche : mais le temps, qui n'est pas pris en compte dans la mise en oeuvre de STATIS, n'est qu'un élément d'interprétation des résultats. L'objectif 5 est satisfait lorsqu'on applique STATIS aux différences entre tableaux mais alors l'interprétation des trajectoires est difficile car elle n'est pas directe.

### I - DESCRIPTION RAPIDE DE LA METHODE

On part d'un ensemble de K tableaux individus x variables  $x_1, \ldots, x_K$  concernant les mêmes individus pondérés par les éléments diagonaux d'une matrice de poids D, sur lesquels sont mesurées des variables quantitatives. Ces variables peuvent être différentes d'un tableau à l'autre.

Pour suivre l'évolution des individus on va travailler sur les tableaux  $W_kD$ , produits scalaires entre individus :  $W_kD = X_k \ ^t X_k D$ . On prendra comme compromis le tableau  $\sum\limits_{l=1}^{K} \alpha_l \ ^t W_lD$  qui maximise :

où le produit scalaire entre matrices est l'opérateur "trace". Si les variables sont différentes d'un tableau à l'autre on travaille sur les tableaux normés  $\frac{W_k D}{\prod W_k D \prod}$ . Les coefficients de la combinaison linéaire donnant le compromis sont les éléments (tous de même signe) du premier vecteur propre de la matrice C des produits scalaires entre les  $W_k D$ . En fait la norme du compromis étant égale à la racine carrée de la première valeur propre de C on ramêne cette norme à la moyenne des normes des  $W_k D$ . On peut dire que le compromis est un

résumé "majoritaire" des K études.

Analyse globale: la diagonalisation de C fournit une représentation des tableaux analogue à celle des variables en Analyse en Composantes Principales. Si les variables sont centrées (respectivement centrées et réduites) par tableau, le produit scalaire entre  $W_kD$  et  $W_1D$  est égal à la somme des covariances (respectivement des corrélations) de toutes les variables de  $X_k$  avec toutes les variables de  $X_1$ . La norme d'un point ou la proximité de deux points sur le graphique s'interprétent donc en termes de covariances (ou de corrélations) entre variables des matrices de données.

Analyse fine : l'Analyse en Composantes Principales du tableau  $X = (\sqrt{\alpha_1} \ X_1 \ \vdots \ \sqrt{\alpha_K} X_K)$ 

donne la représentation des individus et une représentation des variables des différents tableaux.

Il reste à trouver une représentation graphique permettant de comparer les positions respectives des individus tels qu'ils sont décrits par les différents tableaux ou de comparer leurs évolutions. Pour cela on représente les individus dans un système d'axes communs - ceux du compromis WD = X X'D qui sont les axes principaux de l'ACP de X- de la manière suivante : soit Y la matrice dont les colonnes sont les vecteurs propres du compromis, normés à la valeur propre, et A la matrice diagonale de ses valeurs propres décroissantes, les coordonnées des individus décrits par le tableau k sont les lignes de :

W.DYA-1.

Pour un individu donné, sa position compromis est le barycentre (avec les poids  $\alpha_k$ ) des points de sa trajectoire car WDYA<sup>-1</sup> = Y. Or, l'ACP de X nous permet d'expliquer la position des points compromis à l'aide des variables. Cela situe "en gros" l'individu.

Autour de cette position compromis, la trajectoire fait très nettement apparaître les écarts à la moyenne : pour un individu donné, chaque point k de la trajectoire situe cet individu par rapport à l'individu moyen du tableau  $X_k$ , avec les mêmes règles que pour les points compromis, et le segment joignant deux points successifs de la trajectoire reflète, en direction et en longueur, la variation de l'individu entre l'état k et l'état k+1 par rapport à la tendance générale, c'est à dire par rapport à l'écart entre

l'individu moyen du tableau  $X_k$  et l'individu moyen du tableau  $X_{k+1}$ . Comme dans tout plan principal, les variations sont mieux traduites sur les trajectoires "extérieures" du graphique que sur les trajectoires "moyennes" du centre.

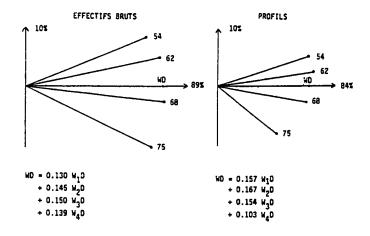
## II - APPLICATION DE STATIS A L'ETUDE DE L'EVOLUTION DES CANTONS RURAUX

Nous considérons les cantons ruraux en tant qu'individus, les catégories socioprofessionnelles comme des variables et les recensements sont les tableaux.

La méthode STATIS a été appliquée d'abord sur les effectifs bruts provenant de la répartition des actifs suivant les catégories socioprofes-sionnelles retenues, ensuite sur les profils des cantons, c'est-à-dire les répartitions en pourcentage des diverses catégories socioprofessionnelles par canton. Ces deux analyses apportent des informations différentes et complémentaires : dans l'analyse des effectifs bruts on met en évidence les variations en nombre alors que dans l'analyse des pourcentages la taille des cantons n'intervient plus et seuls apparaissent les changements de profils.

Dans les deux cas on a travaillé sur les variables centrées et réduites par tableau, sans normer les tableaux  $W_{\rm L}D$ .

## 1. Analyse globale



Quand on analyse les effectifs bruts, les points représentant les recensements se succèdent chronologiquement de façon régulière en étant légèrement de plus en plus éloignés de l'origine. Cela signifie qu'il y a une évolution régulière des effectifs dans le temps et que les corrélations entre catégories socio-professionnelles pour un recensement donné augmentent légèrement d'année en année. Peut-être est-ce dû tout simplement au découpage du secteur tertiaire en cinq catégories alors que le secteur secondaire n'est représenté que par une catégorie : en effet pour la plupart des cantons les effectifs du secteur tertiaire augmentent d'année en année et les corrélations entre les diverses catégories tertiaires augmentent naturellement.

Quand on travaille sur les profils, les points se succèdent encore dans l'ordre chronologique mais 1975 est à part. Il se sépare très nettement des autres points et se rapproche de l'origine. Cela signifie que les profils des cantons changent notablement entre 1968 et 1975, et que les corrélations entre catégories socioprofessionnelles données en pourcentage, sont plus lâches en 1975 que pour les autres recensements : en fin de période, les cantons perdent leurs spécificités. Remarquons que le compromis s'éloigne dans ce cas d'une moyenne équipondérée et que 1975 est moins bien représentée que les autres années dans les graphiques de l'analyse fine.

Après cette vision très générale des recensements nous allons entrer de plus en plus dans les détails.

## 2. Analyse fine

Sur les graphiques donnés par l'ACP de  $X = (\sqrt{\alpha_1}X_1 : ... : \sqrt{\alpha_4}X_4)$  on explique la répartition des cantons compromis par leurs catégories socioprofessionnelles dominantes.

## a) Evolution des catégories socioprofessionnelles

Pour les effectifs bruts on remarque sur le graphique 1 que pour chacune des catégories socioprofessionnelles agricoles -exploitants et sa-lariés- les différents recensements sont très corrêlés. L'étude plus approfondie des cantons nous en donnera l'explication : les cantons agricoles bien qu'en récession gardent pour la plupart des effectifs

agricoles relativement importants et aucun canton non agricole en 1954 ne le devient aprés. Les catégories socioprofessionnelles tertiaires et surtout les "ouvriers" ne sont pas stables. Cela signifie qu'il y a des mutations importantes : les cantons à prédominance ouvrière en 1954 ne le sont plus dans la décennie suivante, et par contre, certains cantons ruraux proches des grandes villes gagnent non seulement en tertiaire mais aussi en population ouvrière au cours des années.

Dans l'analyse des profils (graphiques 3 et 5) les catégories socioprofessionnelles sont plus stables. Prenons le cas des ouvriers : la plupart des cantons à dominante ouvrière sont en récession et, par conséquent, les variations du pourcentage de cette catégorie socioprofessionnelle sont moins accentuées que pour les effectifs bruts.

## b) Répartition des cantons compromis

Dans l'analyse des effectifs bruts, le premier axe est essentiellement un axe de taille qui discrimine les cantons suivant leur effectif global : les petits cantons se situent du côté positif de l'axe 1 et les gros cantons du côté négatif. Cet axe, qui n'apporte finalement pas beaucoup d'information, joue un rôle important dans la répartition graphique des cantons et cela est dû aux données analysées : en effet les tailles des cantons étant trés disproportionnées, les quatre gros cantons - Montpellier, Argelès sur Mer, Perpignan, Béziers - expliquent à eux seuls une grande partie de l'axe.

Le second axe oppose les catégories socioprofessionnelles "exploitants et salariés agricoles" à la catégorie "ouvrier". On retrouve, par conséquent, les cantons agricoles comme Lézignan-Corbières ou Fournels en bas du graphique et les cantons plutôt ouvriers comme St Ambroix ou Alès en haut.

Il faut noter également que parmi les catégories socioprofessionnelles qui caractérisent le premier axe, le tertiaire est prépondérant.

Dans l'analyse des profils, le premier axe oppose les catégories socioprofessionnelles agricoles, et plus particulièrement les "exploitants agricoles" aux autres catégories. Le second axe oppose les "salariés" aux "exploitants agricoles" d'une part, les "ouvriers" et les employés" aux "cadres moyens" et "autres catégories" d'autre part. Le troisième axe oppose seulement les "ouvriers" aux "salariés agricoles".

Ainsi les cantons agricoles se projettent du côté négatif de l'axe 1, la séparation entre les cantons à fort pourcentage d'exploitants agricoles et ceux à fort pourcentage de salariés se faisant sur les axes 2 et 3.

Un canton se projetant du côté négatif des axes 2 et 3 est certainement un canton ouvrier etc.

### c) Trajectoires des cantons

Un deuxième niveau dans le détail est atteint lorsqu'on représente la trajectoire suivie par chaque canton de 1954 à 1975 dans le système d'axes qui a permis de repérer les cantons compromis.

Une fois que l'on a donné un nom aux axes (en termes de catégories socioprofessionnelles) la lecture d'une trajectoire est facile. Les variables étant centrées pour chaque tableau, ce sont les écarts par rapport au canton moyen qui se lisent sur une trajectoire.

Le tableau suivant donne l'évolution du profil du canton moyen :

	EA	OA	AC	PL	CM	EM	OU	SE	CP
1954	32.11%	20.47	10.94	0.98	3.50	4.39	23.76	2.38	1.48
1962	27.96	16.80	10.80	1.38	4.74	5.35 🚚	28.05	2.56	2.36
1968	23.42	11.80	10.79	2.03	6.34	7.34 -	32.90	3.22	2.16
1975	16.88	7.62	11.25	4.30	9.80	10.92	32.55	4.16	2.52

La part des actifs agricoles recule, de plus de 50% à moins de 25%, au profit de toutes les autres catégories, à l'execption des "artisans et commerçants" dont la part reste stable.

Lorsqu'il ya beaucoup d'individus, l'enchevêtrement des trajectoires est un obstacle à une lecture aisée. C'est pourquoi sur la partie du graphique n° 2 correspondant aux petits cantons nous n'avons conservé que des cantons caractéristiques d'une typologie basée sur les effectifs bruts :

- M4 (commeF9, G4, G6, J6, K1, L2, M3, M9 et A7, C9) sont des petits cantons agricoles au départ (par leurs positions sur l'axe 2) sans grand changement

d'effectif global (stabilité sur l'axe 1) qui perdent des agriculteurs (remontée le long de l'axe 2) et se tertiarisent peu. En effet l'axe 3, non représenté ici, met en évidence les variations du secteur tertiaire : un accroissement supérieur à la moyenne dans ce secteur se traduit par une montée très forte vers le côté positif de l'axe, une récession ou un accroissement trop faible par rapport à la moyenne se traduit par une descente vers le côté négatif de l'axe. Or les cantons du type de M4 chutent fortement sur l'axe 3.

- D8 (comme G9) ont le même type d'évolution que M4 mais ils sont ouvriers (côté positif de l'axe 2).
- K3 (comme C3, D7, H3, I8, M6, N8) cantons d'effectifs moyens sauf I8 et D7 plus petits (car situés du côté négatif de l'axe 1) et ouvrier en 1954 (côté positif de l'axe 2) se dépeuplent fortement (la trajectoire suit l'axe 1) perdent leurs ouvriers (descente le long de l'axe 2) sans développer le secteur tertiaire (descente le long de l'axe 3 surtout en fin de période). Le cas du gros canton E1 est semblable.
- C8 (comme A5, H4, H7, H9) très agricoles (côté négatif de l'axe 2) se dépeuplent plus ou moins (leurs trajectoires suivent l'axe 1) et ne développent pas ou peu (pour A5 et H7) leur secteur tertiaire (descente plus ou moins brutale le long de l'axe 3).
- K5 (comme F1, J9, K6 et K2) sont des cantons plutôt agricoles au départ (côté négatif de l'axe 2) sauf K2 qui est ouvrier (côté positif de l'axe 2) qui se développent relativement bien (trajectoires allant vers le côté négatif de l'axe 1) par un gain d'ouvriers (montée le long de l'axe 2) et un accroissement du secteur tertiaire après 1968 (descente nette sur l'axe 3 de 1954 à 1962 puis peu accentuée entre 1962 et 1968 suivie d'une forte remontée après 1968).
- G7 (comme N4 et F2, K7) sont relativement stables. Le sens de la trajectoire sur l'axe 3 montre que F2 et K7 (et non G7 et N4) se tertiarisent.
- K4 se développe (côté négatif de l'axe 1) par gain en ouvriers (la trajectoire monte le long de l'axe 2 jusqu'en 1968) et non en tertiaire (descente sur l'axe 3).
- L'évolution des gros cantons est plus délicate à déterminer car les trajectoires sont très sensibles aux variations par rapport à la tendance moyenne. Sur le troisième axe les trajectoires de K8, 04, H6 et D3 sont ascendantes : ces cantons connaissent un développement important de leur secteur tertiaire

(accompagné d'un fort accroissement de population pour K8 et 04 : la trajectoire se déplace vers le côté négatif de l'axe 1). D3 reste à majorité ouvrière (côté positif de l'axe 2) mais l'augmentation de sa population ouvrière d'un recensement à l'autre n'est pas suffisante, comparée aux variations du pourcentage d'ouvriers du profil moyen : sa trajectoire chute sur l'axe 2. Entre 1962 et 1968, la direction oblique de la trajectoire de N7 traduit le fait que ce canton perd moins d'agriculteurs que la moyenne.

L'étude des trajectoires des cantons décrits par leurs effectifs bruts permet d'énoncer les conclusions suivantes : une grande partie de ces cantons déclinent ; aucun canton ne gagne des agriculteurs (ce qui se traduirait par une trajectoire dirigée vers les côtés négatifs des axes 1 et 2) ; les cantons ouvriers en 1954 déclinent en perdant leur population ouvrière (trajectoires dirigées vers le côté positif de l'axe 1 et fuyant le côté positif de l'axe 2) ; et enfin le développement du secteur tertiaire sauve une partie des petits cantons d'un total déclin, soit accompagné d'un fort accroissement de la population (trajectoires obliques dirigées vers le côté négatif de l'axe 1 et le côté positif de l'axe 3) soit - et c'est l'objet du paragraphe suivant - en pourcentage.

Les trajectoires des cantons décrits par leurs profils (graphiques 4 et 6) sont de trois types.

La plupart des trajectoires vont plus ou moins directement vers le centre du graphique qui correspond à un profil moyen. Cela signifie que les cantons perdent leurs spécificités, décrites par leurs positions excentrées sur le graphique en 1954, en se tertiarisant pour la plupart. Nous avons déjà constaté cela lors de l'analyse globale. C'est le cas par exemple de C8 :

	EA	OA	AC	PL	CM	EM	OU	SE	CP
1954	36.22%	39.60	10.55	0.36	2.34	2.37	6.22	1.61	0.73
1962	34.04	36.57	9.79	0.63	3.24	3.47	9.65	1.28	1.34
1968	30.61	32.43	9.76	1.28	3.61	4.60	15.05	1.58	1.08
1975	23.64	26.00	9.46	2.00	4.55	7.46	20.73	4.91	1.27

Jusqu'en 1968, C8 ne suit pas la tendance générale décrite par l'évolution du profil moyen : il ne perd pas suffisamment d'agriculteurs et ne se tertiarise pas assez. Entre 1962 et 1968 l'orientation de sa trajectoire vers les côtés négatifs des axes 2 et 3 reflète le gain important en ouvriers. Après 1968 il suit la tendance générale (sa trajectoire se dirige vers le centre du graphique) tout en restant plutôt agricole avec beaucoup d'ouvriers agricoles (trajectoire restant du côté négatif des axes 1 et 2, et du côté positif de l'axe 3).

Certains cantons changent complètement de profil et leurs trajectoires traversent largement le graphique. C'est le cas des cantons peri-urbains K5 (ou 04) qui étaient essentiellement agricoles en 1954 et qui deviennent tertiaires en 1975.

K5 :	EA	0A	AC	PL	CM	EM	ou	SE	CP
1954	37.07%	34.92	7.37	0.54	2.76	3.84	9.06	2.30	2.15
1962	28.25	28.17	6.75	1.50	3.92	6.33	16.08	3.42	5.58
1968	19.97	17.27	7.15	4.52	8.64	8.70	25.30	4.05	4.39
1975	8.24	7.85	10.35	15.52	15.71	13.41	20.88	3.64	4.41

Enfin les trajectoires des cantons en déclin "tournent le dos" au centre du graphique. Ils se dépeuplent et ne se régénèrent pas par le tertiaire. C'est le cas de D8 qui tombe de 977 actifs en 1954 à 480 actifs en 1975, ou de K4 en dernière période.

D8 :	EA	OA	AC	ΡL	CM	EM	OU	SE	CP
1954	7.37%	1.12	3.58	0.31	4.09	4.09	76.46	2.35	0.61
1962	9.95	2.29	6.32	1.08	5.24	4.70	66.94	2.42	1.08
1968	4.88	2.61	5.75	1.05	5.23	9.58	67.07	2.79	1.05
1975	1.04	3.13	7.29	2.71	7.08	6.67	69.17	1.46	1.46

En conclusion, l'examen des trajectoires donne une idée de l'allure générale de chaque canton, décrit son évolution par rapport à l'évolution du canton moyen et met en évidence les cas particuliers dont l'explication doit être complétée par un retour aux données.

## REFERENCES

ESCOUFIER, Y., "Exploratory Data Analysis when Data are Matrices", Recent Development in Statistical Inference and Data Analysis, 1980, K. Matusita Editor, pp. 45-53.

GLACON, F., "Analyse conjointe de plusieurs matrices de données", Thèse, Université Scientifique et Médicale de Grenoble, France, 1981.

LAVIT, C., ROUX, C., "Analyse conjointe de plusieurs tableaux de données par la méthode STATIS", Rapport technique n° 8402, Unité de Biométrie, Montpellier, France, 1984.

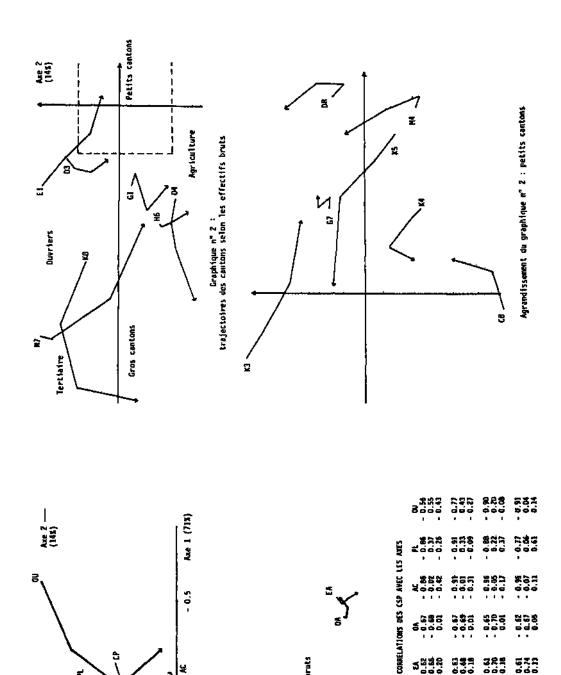
LAVIT, C., PEREZ HUGALDE, C., "Application de la méthode STATIS à des données économiques : évolution des secteurs agricoles et non agricoles des provinces espagnoles de 1960 à 1979",Quatrièmes Journées Internationales Analyse des Données et Informatique, 9-11 octobre 1985, Versailles, France.

LAVIT, C., PERNIN, M.O., "Multivariate and longitudinal data on growing children: solved by STATIS, Third International symposium on Data Analysis, 10-12 juin 1985, Bruxelles, Belgique.

L'HERMIER DES PLANTES, H., THIEBAUT, B., "Etude de la pluviosité au moyen de la méthode STATIS", Revue de Statistique Appliquée, 1977, Vol. XXV, n°2, pp. 57-81.

PLACE, M.C., "Contribution algorithmique à la mise en œuvre de la méthode STATIS", Thèse, Université des Sciences et Techniques du Languedoc, Montpellier II, France, 1980.

ROBERT, P., ESCOUFIER, Y., "A unifying tool for linear multivariate statistical methods: the RV-coefficient", Appl. Stat., C, 25 (3), 1976, pp. 257-265.



Graphique nº 1 CSP effectifs bruts

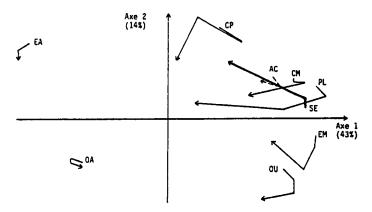
373

355

ž

533

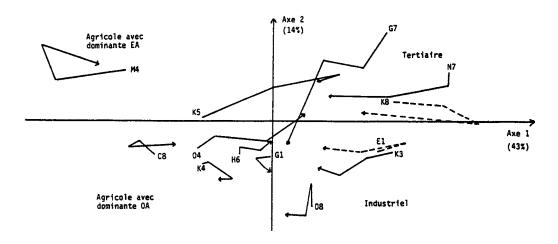
Ĭ



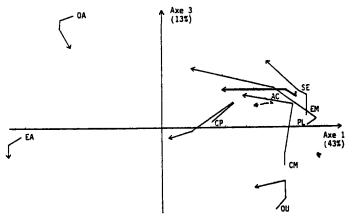
Graphique n° 3 : CSP profils

## CORRELATIONS DES CSP AVEC LES AXES

		EA	OA	AC	PL	OU
1954	axe 1	- 0.77	- 0.44	0.63	0.81	0.66
	axe 2	0.56	- 0.32	0.30	0.23	- 0.37
1962	axe 1	- 0.80	- 0.51	0. <b>66</b>	0.90	0.70
	axe 2	0.52	- 0.28	0.21	0.17	- 0.44
1968	axe i	- 0.80	- 0.51	0.61	0.68	0.70
	axe 2	0.51	- 0.31	0.25	0.08	- 0.53
1975	axe 1	- 0.80	- 0.47	0.52	0.17	0.56
	axe 2	0.44	- 0.31	0.26	0.12	- 0.59



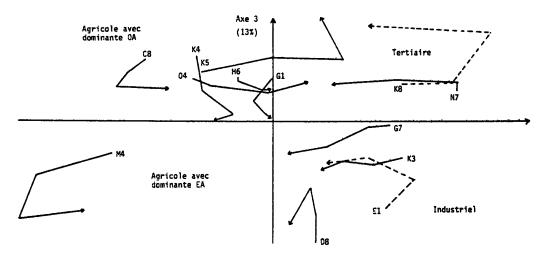
'Graphique n° 4 : trajectoires des cantons suivant leurs profils



Graphique n° 5 : CSP profils

## CORRELATIONS DES CSP AVEC LES AXES

		EA	OA	AC	PL.	Oυ
1954	axe 1	- 0.77	- 0.44	0.63	0.81	0.66
	axe 3	- 0.08	0.82	0.23	0.02	- 0.61
1962	axe 1	- 0.80	- 0.51	0.66	0.90	0.70
	axe 3	- 0.13	0.79	0.17	0.06	- 0.52
1968	axe 1	- 0.80	- 0.51	0.61	0.68	0.70
	axe 3	- 0.17	0.74	0.18	0.30	- 0.40
1975	axe 1	- 0.80 - 0.22	- 0.47 0.59	0.52	0.17 0.45	0.56



Graphique n° 6 : trajectoires des cantons suivant leurs profils