

STATISTIQUE ET ANALYSE DES DONNÉES

THIERRY FOUCART

Application de l'analyse factorielle des correspondances

Statistique et analyse des données, tome 10, n° 1 (1985), p. 54-64

<http://www.numdam.org/item?id=SAD_1985__10_1_54_0>

© 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/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/>

APPLICATION DE L'ANALYSE FACTORIELLE DES CORRESPONDANCES

Thierry Foucart

Département de Mathématiques et d'Informatique
Université d'Orléans
Orléans La Source (45)

Résumé: *Après une présentation rapide de la méthode que nous employons ici, nous développons l'interprétation des résultats que nous avons obtenus en tentant de mettre en évidence les avantages et les inconvénients de l'approche choisie.*

Abstract: *The method used here to describe a sequence of boards is based upon Correspondence Factor Analysis. First we will explain particular procedures we have had to use, and second we will present the result obtained with this method, in the analysis of the common data.*

Mots clés: *Analyse factorielle des correspondances, éléments supplémentaires.*

Indices de classification STMA: 06-020

1 - PRESENTATION DE LA METHODE D'ANALYSE [1]

Nous avons utilisé pour étudier cette suite de tableaux de contingence l'analyse factorielle des correspondances de la façon suivante:

- 1) Calcul du tableau moyen des effectifs
- 2) Analyse des correspondances de ce tableau
- 3) Projection sur le référentiel obtenu des lois conditionnelles en lignes et en colonnes et des lois marginales de chaque tableau.

Les métriques utilisées dans les projections sont les métriques du χ^2 définies par les lois marginales du tableau moyen.

Chaque canton (comme chaque CSP) est donc représenté par 5 points: le point défini par la ligne du tableau moyen, les points définis par les lignes des 4 tableaux. Ces points caractérisent des lois conditionnelles: les effectifs n'interviennent donc pas directement. La projection en éléments supplémentaires des lois marginales revient à projeter les centres de gravité des différents nuages, puisque la loi marginale sur I du k^e tableau est la moyenne des lois conditionnelles sur I définies sur ce tableau et pondérées par les termes marginaux correspondants dans la loi marginale sur J.

Il est clair que cette méthode permet d'observer les évolutions des lois conditionnelles et marginales relativement aux caractères mis en évidence par l'analyse du tableau moyen: l'optimisation concerne la représentation de ces caractères et non celles des trajectoires.

Le tableau 1 (page suivante) indique la disposition des différents tableaux de lois conditionnelles par rapport au tableau moyen

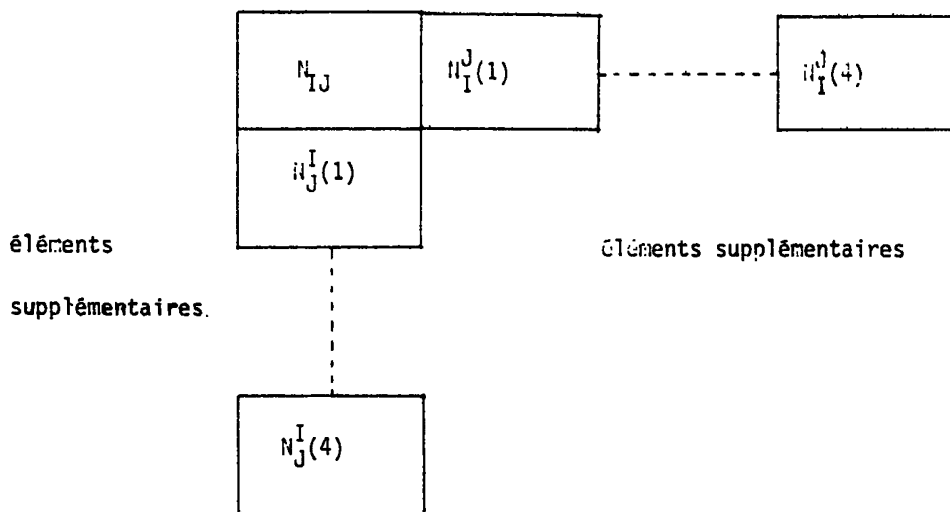


tableau 1: disposition des tableaux de lois conditionnelles par rapport au tableau moyen:

VALEURS PROPRES (1E LIGNE)
CONTRIBUTION A L'INERTIE (2E LIGNE)

0.130	0.046	0.024	0.010	0.003	
59.919	21.294	10.998	4.551	1.327	INERTIE TOTALE .21726

POUR CHAQUE AXE:
1E COLONNE: COORDONNEE
2E COLONNE: CORRELATION A L'AXE (AU CARRE)
3E COLONNE: CONTRIBUTION A L'INERTIE EXPLIQUEE PAR L'AXE

		AXES PRINCIPAUX											
		AXE 1		AXE 2		AXE 3		AXE 4					
COMPROMIS													
EA	**	0.488	0.722	36.9*	-0.301	0.274	39.4*	-0.018	0.001	0.3*	-0.028	0.002	1.6*
OA	**	0.450	0.539	23.0*	0.395	0.417	50.1*	0.128	0.044	10.1*	-0.016	0.001	0.4*
AC	**	-0.082	0.093	0.6*	0.007	0.001	0.0*	-0.036	0.018	0.6*	0.248	0.845	72.6*
PL	**	-0.222	0.148	1.0*	0.178	0.095	1.8*	-0.442	0.589	21.1*	-0.167	0.084	7.2*
CM	**	-0.151	0.261	1.2*	0.045	0.023	0.3*	-0.222	0.560	13.6*	-0.078	0.069	4.0*
EM	**	-0.190	0.333	2.3*	0.149	0.107	4.0*	-0.188	0.328	12.3*	-0.075	0.052	4.7*
OU	**	-0.381	0.842	33.9*	-0.067	0.026	2.9*	0.146	0.124	27.2*	-0.036	0.007	3.9*
SE	**	-0.195	0.188	1.0*	0.115	0.065	1.0*	-0.273	0.370	10.7*	0.107	0.056	3.9*
CP	**	-0.082	0.034	0.1*	-0.111	0.063	0.5*	-0.219	0.245	4.0*	0.090	0.041	1.6*

tableau 2 : Résultats numériques de l'A.F.C.

2 - INTERPRETATION DES RESULTATS

2.1 Structure et tendances.

Nous avons étudié les résultats de l'analyse telle qu'elle est décrite précédemment en plusieurs étapes. Dans un premier temps, nous nous limitons à l'étude du tableau moyen pour mettre en évidence les principaux caractères des données étudiées. Une tendance générale peut être dégagée en étudiant les projections des lois marginales qui sont représentées sur le même schéma (figure 1).

Une structure très forte apparaît dans ces données: les deux premiers axes contribuent pour plus de 80% à l'inertie totale, et le premier pour plus de 60%. Le plan principal 1x2 contient donc la presque totalité de l'information que l'interprétation des axes va préciser (cf tableau 2)

L'axe 1 sépare visiblement les professions agricoles (ouvriers agricoles OA, exploitants agricoles EA) des CSP des secteurs secondaire et tertiaire (ouvriers OU, employés EM, Professions Libérales PL, etc...). Les trois CSP OA, EA, OU contribuent pour plus de 95% de l'inertie expliquée par l'axe 1. Corrélativement, on peut séparer les cantons en deux: cantons ruraux à prédominance agricole (professions agricoles fortement représentées par rapport à la moyenne) et à prédominance non agricole (proportion élevée d'ouvriers). Cette interprétation est confirmée lorsque l'on étudie les cantons placés à droite et à gauche de la figure 1: nous les avons regroupés suivant la classification donnée par les géographes (F. Auriac, M.C. Bernard, 1974) pour 1968 sur la totalité des cantons ruraux du Languedoc-Roussillon. Cette classification correspond tout à fait à l'A.F.C. du tableau moyen que nous avons effectuée ici.

L'axe 2 oppose les ouvriers agricoles (salariés) aux exploitants agricoles (non salariés). Une hypothèse peut être avancée: une exploitation agricole peut être individuelle ou familiale (peu d'ouvriers agricoles évidemment) ou avoir besoin d'une main-d'oeuvre nombreuse. On retrouve effectivement cette distinction dans les

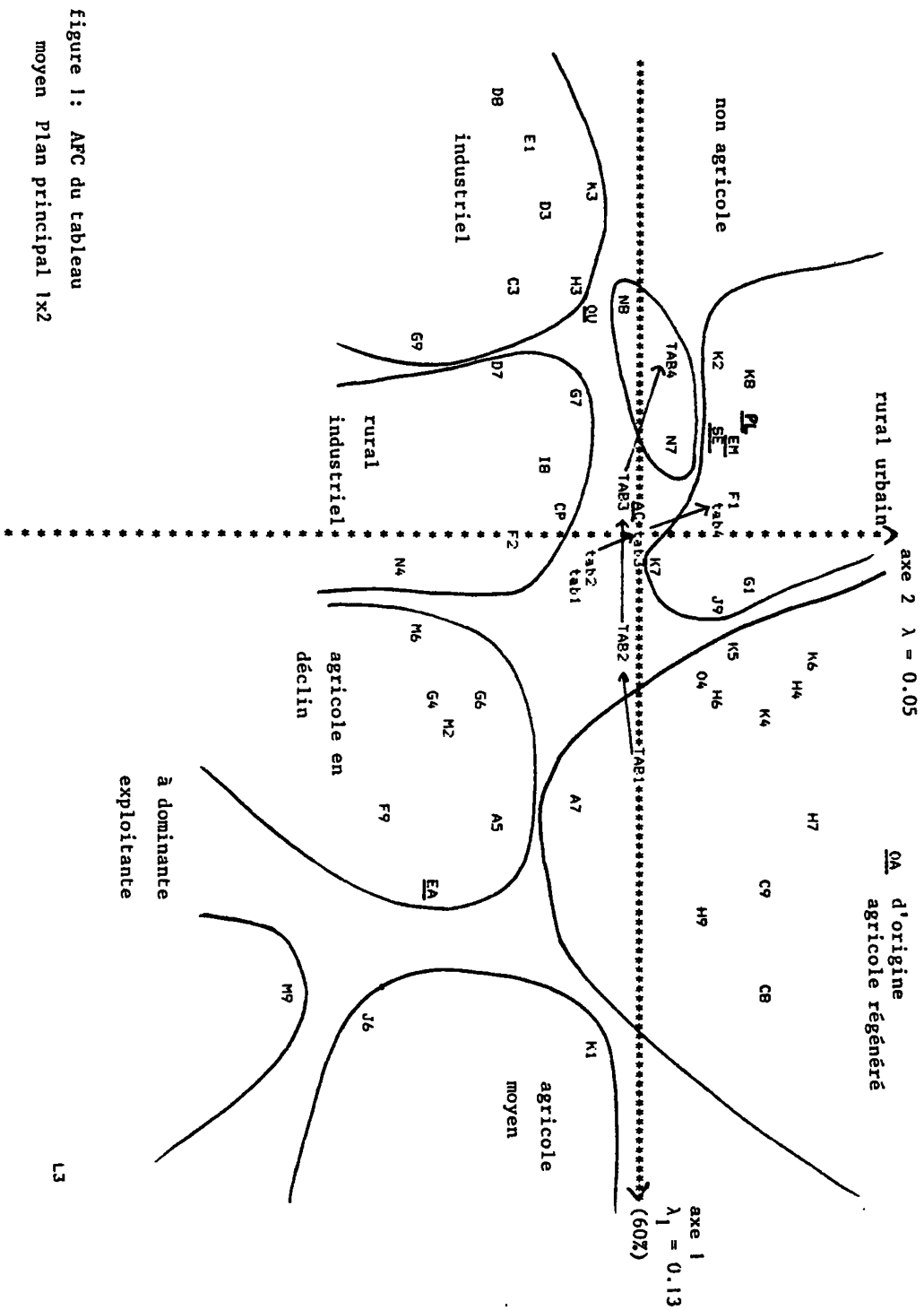


figure 1 : AFC du tableau moyen Plan principal 1x2

L3

M4

cantons suivant le côté où ils sont représentés sur le schéma: du côté positif des axes 1 et 2, on observe les cantons de vignoble regroupés dans la classe des cantons "d'origine agricoles, régénéré". Du côté négatif de l'axe 2, ce sont les cantons des types suivants: à dominante exploitant, agricole moyen et agricole en déclin.

Les axes 3 et 4 sont beaucoup moins significatifs (15% de l'inertie totale à eux deux). Ils sont toutefois faciles à interpréter: l'axe 3 sépare les ouvriers et les ouvriers agricoles des autres CSP, et l'axe 4 les artisans et les commerçants.

Ces quatre axes sont calculés par l'AFC du tableau moyen: ils caractérisent des propriétés communes à chaque tableau de la suite, avec évidemment des nuances à préciser. La représentation des marges sur la figure 1 permet d'interpréter en fonction de ces axes les tendances définies par les répartitions marginales. On constate qu'elles n'évoluent pas de la même façon: les points TAB1, ... , TAB4 suivent à peu près l'axe 1 tandis que les points tab1, ... , tab4 s'échelonnent le long de l'axe 2. Les premiers correspondent à la répartition des CSP dans la population totale: l'orientation de la trajectoire montre l'augmentation des classes des secteurs secondaire et tertiaire, en pourcentage bien entendu. Les deuxièmes caractérisent la distribution de la population dans les cantons indépendamment de la CSP. On constate une diminution de la population relative des cantons des types: agricole en déclin, à dominante exploitant, rural industriel et agricole moyen au profit des cantons de type rural urbain. On peut expliquer ces évolutions par une double migration: la première concerne les cantons cités précédemment et la seconde les CSP, les migrants changeant de catégorie socio-professionnelle.

2.2 Infrastructures

Le paragraphe précédent contient une description de la structure de la suite des tableaux comprise comme l'ensemble des propriétés qui leur sont communes, et des tendances générales définies par les suites des marges. Nous abordons maintenant ce que nous appelons ici l'étape des infrastructures par analogie avec d'autres méthodes: la

figure 2 représente les trajectoires des CSP au cours des 4 recensements et la figure 3 celles des cantons.

Seuls les points caractérisant les ouvriers agricoles (OA) et les exploitants agricoles (EA) sont stables sur le plan 1x2. Les autres trajectoires s'orientent du côté positif de l'axe 2. Cela signifie que les répartitions des ouvriers agricoles et des exploitants agricoles dans les différents cantons ont peu varié entre 1954 et 1975. Il n'en est pas de même des autres CSP dont les trajectoires sont toutes orientées vers le côté positif de l'axe 2. Rappelons qu'il s'agit de pourcentages et que la population des cantons n'intervient pas dans ces trajectoires.

L'explication que nous proposons est la suivante: la diminution de la part relative des professions du secteur primaire dans l'ensemble des CSP n'affecte pas leur répartition géographique: l'exode rural concerne tous les cantons étudiés. Ce n'est pas le cas des CSP des secteurs secondaire et tertiaire, dont l'importance augmente en particulier dans certains cantons. La conclusion est analogue à la précédente mais précise l'évolution de chaque CSP.

Cette interprétation est confirmée par l'étude des trajectoires des cantons. Là encore, il s'agit de lois conditionnelles et chaque point caractérise la répartition socio-professionnelle du canton correspondant exprimée en pourcentages: les effectifs du canton n'interviennent pas dans sa trajectoire. On peut constater sur la figure 3 que la quasi-totalité des cantons évoluent de la même façon. En rappelant qu'il s'agit de cantons ruraux, nous pouvons expliquer ce phénomène à l'aide des interprétations précédentes: on trouve dans chaque canton un pourcentage de plus en plus important de travailleurs des secteurs secondaire et tertiaire, en conséquence de l'émigration rurale. La diminution en valeur absolue de la population d'un canton n'empêche pas cette évolution puisqu'elle est due au départ des travailleurs du secteur primaire, l'augmentation non plus puisqu'elle se fait au profit des CSP des secteurs secondaire et tertiaire.

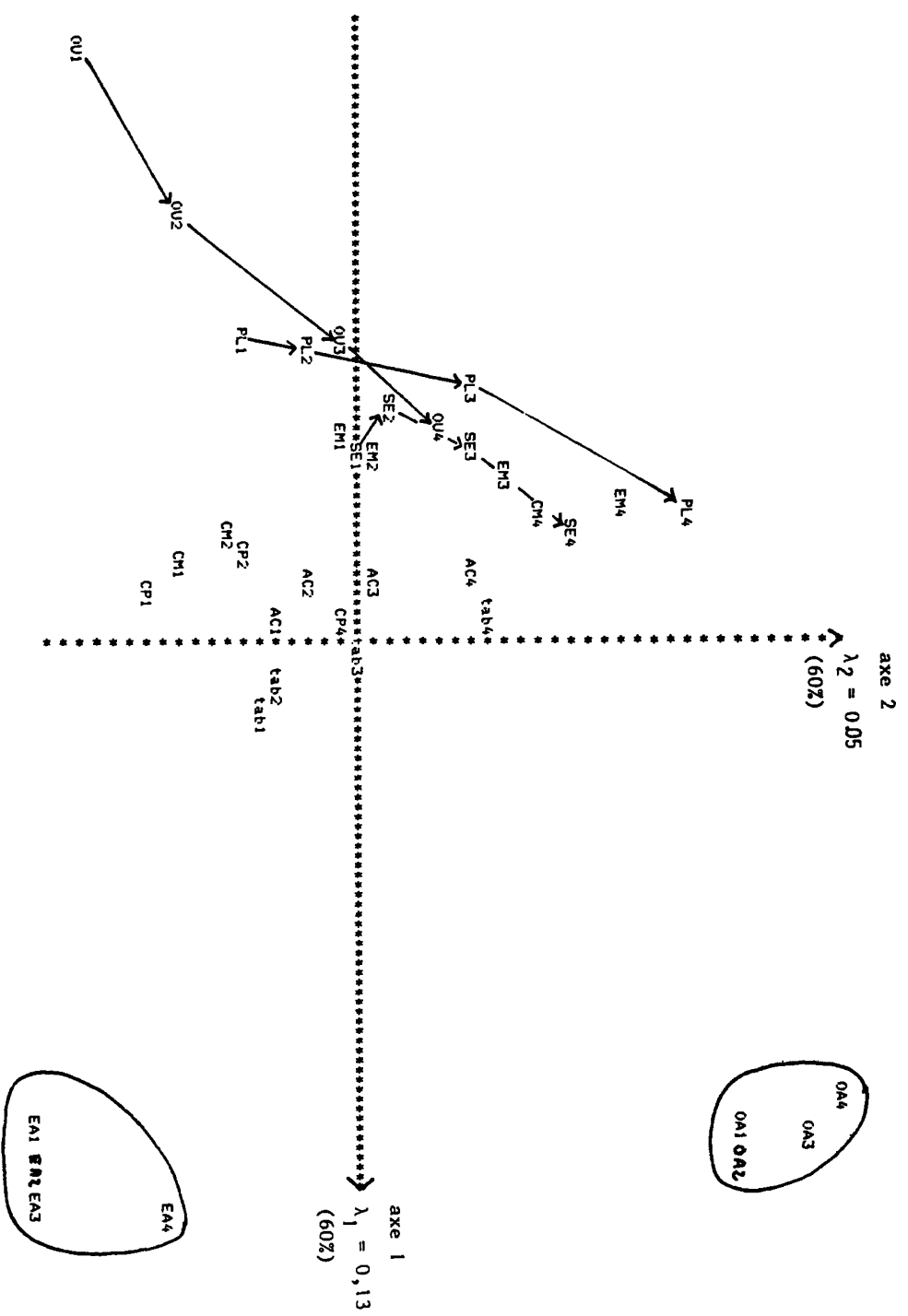


figure 2: trajectoires des GSP
(plan principal 1x2 du tableau moyen)

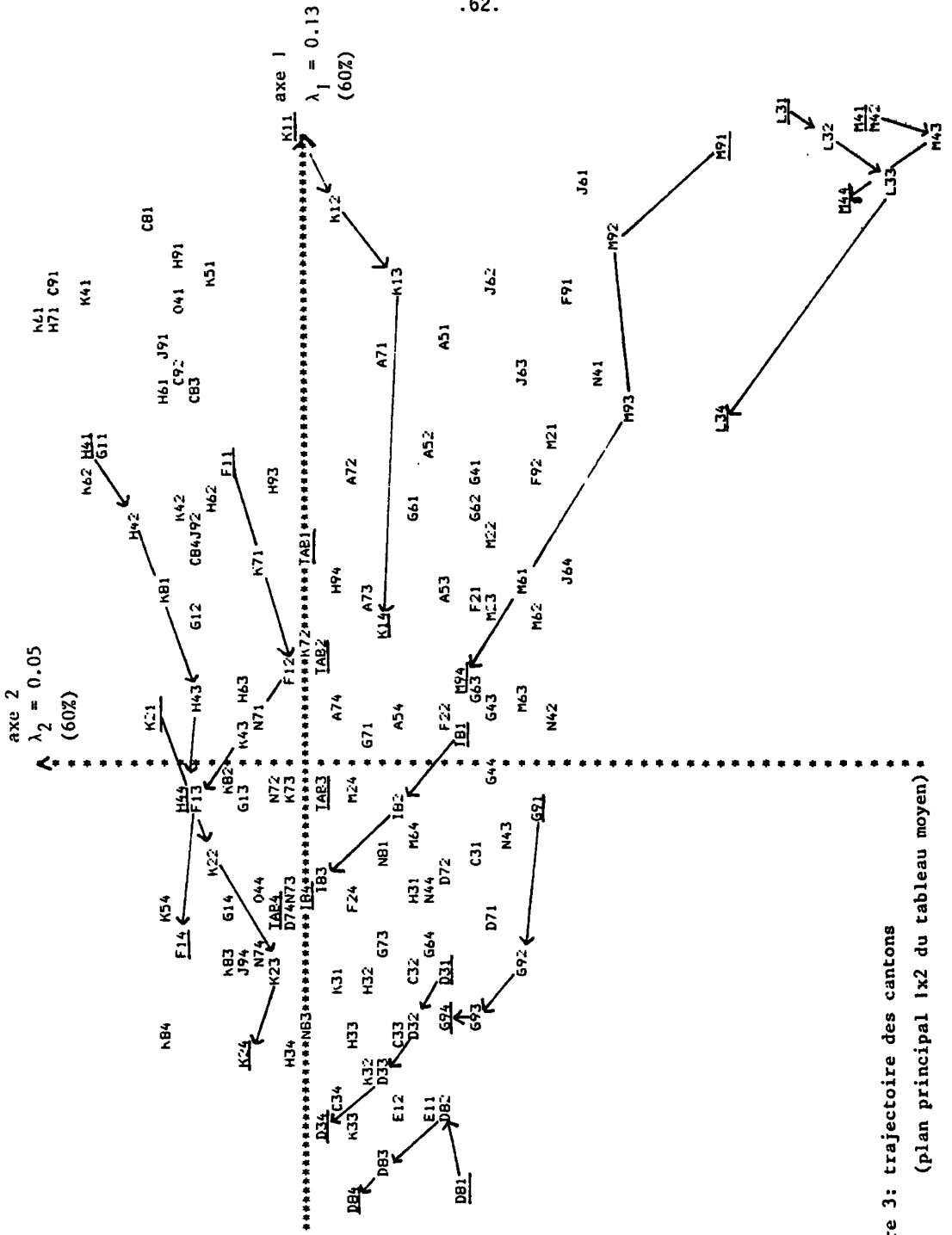


figure 3: trajectoire des cantons
(plan principal 1x2 du tableau moyen)

3 - CONCLUSION

Cette analyse met bien en évidence les conclusions auxquelles étaient arrivés les géographes par la comparaison des différents plans factoriels établis par l'analyse des correspondances de chaque tableau (F. Auriac, M.C. Bernard, 1974). Il apparaît qu'au cours des dernières années, la structure socio-professionnelle des cantons ruraux de la région Languedoc-Roussillon a profondément évolué. Le caractère purement agricole de la population tend à disparaître, soit par diminution du nombre de travailleurs du secteur primaire, soit par augmentation du nombre des travailleurs des secteurs secondaire et tertiaire. Par contre, nous voyons un lien entre l'évolution des effectifs agricoles et celles des autres groupes alors que l'analyse en composantes principales effectuée sur les tableaux des différences des effectifs ne la met pas en évidence [2]

L'analyse en composantes principales sur le tableau des différences ne nous paraît pas adaptée à une telle recherche. La notion de coefficient de corrélation linéaire elle-même ne correspond pas à la nature du problème: en effet, nous avons mis en évidence une double évolution, la première concernant les cantons et la seconde les CSP. Le coefficient de corrélation linéaire ne peut mesurer un lien entre deux variables que dans la mesure où les individus restent les mêmes, ce qui serait le cas dans notre étude si l'évolution des structures socio-professionnelles avait lieu à l'intérieur de chaque canton. Nous avons vu que la nature du lien entre les deux évolutions ne correspond pas à ce modèle. Un autre problème est de définir un critère pour représenter les trajectoires sur un référentiel non pas calculé à partir des tableaux de contingence mais permettant de les interpréter indépendamment des facteurs principaux. Il est concevable en effet que des évolutions existent sans interférence avec la structure de la suite définie comme précédemment par l'ensemble des propriétés qui sont communes aux tableaux. Une des

difficultés de l'analyse telle que nous l'avons présentée est que la tendance apparaît systématiquement sur tous les axes: cela rend difficile le choix des facteurs à interpréter (dans l'étude ci-dessus, nous nous sommes limités arbitrairement aux deux premiers axes). D'autres approches ont été développées dans le cas de variables quantitatives. Elles nous paraissent difficilement utilisables pour étudier des tableaux de contingence dans lesquels les notions d'individus et de variables sont délicates à définir.

BIBLIOGRAPHIE

[1] FOUCART, T., "Analyse Factorielle de Tableaux Multiples", Masson, Paris, 1984

[2] AURIAC, F., BERNARD, M.C., LOCHARD, E., "Le changement social dans les campagnes languedociennes. L'espace géographique, 1975, n° 4, p. 239-250.