

# STATISTIQUE ET ANALYSE DES DONNÉES

J. Y. LAFAYE

## Application du modèle INDSCAL

*Statistique et analyse des données*, tome 10, n° 1 (1985), p. 83-102.

[http://www.numdam.org/item?id=SAD\\_1985\\_\\_10\\_1\\_83\\_0](http://www.numdam.org/item?id=SAD_1985__10_1_83_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/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.

NUMDAM

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

APPLICATION DU MODELE INDSCAL

J.Y. LAFAYE

Université de Haute-Bretagne

Résumé : Dans le cadre des travaux du GRECO CNRS ANADI, qui se propose de comparer les différentes approches statistiques possibles dans l'étude de données évolutives nous présentons ici, une utilisation du modèle INDSCAL appliqué à l'analyse d'une suite de tableaux de contingence.

Les données concernant les répartitions des C.S.P. dans 42 cantons ruraux du Languedoc-Roussillon, mesurées lors des recensements de 1954, 1962, 1968 et 1975. Une analyse antérieure, publiée dans le bulletin de la société languedocienne de géographie, portant sur les trois premières années est citée en bibliographie qui nous servira de référence.

Abstract : According to the frame of GRECO CNRS ANADI work, that aims at comparing several statistical approaches in the study of evolutive data, we present here, an application of the INDSCAL model to the analysis of a series of contingency tables.

The data refer to the distribution of the S.P.C. among 42 Languedoc-Roussillon rural counties. These data have been collected for the 1954, 1962, 1968 and 1975 population census. A prior analysis, published in the "Bulletin de la société languedocienne de géographie" dealing with the first three years is given in the bibliography and will be used as a reference.

Mots clés : INDSCAL, positionnement multidimensionnel

Indices de classification STMA : 06-090, 06-030.

A - METHODE

Après un rappel méthodologique, nous présenterons quelques remarques sur les options particulières choisies pour l'application du modèle INDSCAL à

l'étude des données géographiques qui nous intéressent ; nous détaillerons enfin les résultats obtenus.

AO - Modèle INDSCAL, hypothèses et méthodologie :

a) Généralités

Le modèle INDSCAL permet le positionnement multidimensionnel d'un ensemble d'objets (stimuli), conformément à la perception qu'en ont un certain nombre de juges. Cette perception est résumée dans des matrices de produits scalaires représentant les ressemblances entre stimuli, et sont fournies, directement ou non, par chacun des juges.

La résolution s'effectue sous les contraintes suivantes, qui constituent les hypothèses du modèle INDSCAL :

- pour percevoir les distances entre stimuli, tous les juges se réfèrent à une même représentation euclidienne dans un système  $p$  axes orthogonaux,
- dans son appréciation des distances, chaque juge pondère les axes d'une manière qui lui est propre.

b) Notations, Définitions

On suppose données  $m$  matrices :  $(W_i)_{i=1,m}$  de produits scalaires entre  $n$  objets. Les coordonnées des  $n$  stimuli sur les  $p$  axes de la configuration commune forment le tableau  $X$  ( $n \times p$ ). Enfin, à chacun des  $m$  juges, est associée une métrique diagonale de matrice  $M_i$  sur  $\mathbb{R}^p$ .

Si  $(E_i)_{i=1,m}$  est une suite de matrices carrées ( $n \times n$ ) rassemblant les termes d'écart introduits sous forme additive, le modèle INDSCAL s'écrit, de façon matricielle :

$$W_i = X M_i^t X + E_i \quad (i=1,m) \quad (1)$$

pour deux stimuli  $j$  et  $k$ , l'équation entre scalaires s'exprime par :

$$W_{ijk} = \sum_{t=1,p} x_{jt} m_{it} x_{kt} + e_{ijk} \quad \begin{matrix} (i=1,m) \\ (j=1,n) \\ (k=1,n) \end{matrix} \quad (2)$$

ou  $m_{it}$  représente le  $i$ ème terme diagonal de la matrice  $M_i$ .

c) Remarques, Généralisations

- Il est clair que les données initiales peuvent, en pratique, prendre des formes bien moins restrictives que celles stipulées ci-dessus. Par exemple, des

matrices de distances euclidiennes peuvent être transformées en matrices de produits scalaires par la transformation de TORGERSON. De même, des matrices de distances quelconques peuvent devenir euclidiennes après addition d'une constante convenable, enfin de simples matrices de similarité ou dissimilarité peuvent induire des matrices de distances.

- Tel qu'il est stipulé en (1), le modèle INDSCAL n'est pas totalement identifiable, ainsi est il nécessaire de préciser certaines contraintes supplémentaires.

- On suppose la configuration commune centrée sur chacun des  $p$  axes, et également que les matrices résiduelles  $E_i$  sont de marges nulles. En conséquence, les matrices  $W_i$  vérifient elles aussi, la nullité des marges (ceci est d'ailleurs en accord avec la transformation de TORGERSON).

- Les poids donnés aux axes et leur variance ne peuvent être choisis de façon indépendante. On suppose donc que la configuration commune est formée d'axes de variance unitaire.

- La méthode INDSCAL peut être considérée de différentes manières

- Analyse factorielle de suite de tableaux :

En effet, les axes de la configuration commune permettent une représentation de type interstructure, en faisant figurer chaque juge  $i$  avec la coordonnée  $m_{it}$

De même, les intrastructures peuvent être représentées dans un référentiel rigoureusement commun ; Chacune des  $m$  structures se déduisant, selon le modèle, de la configuration commune par composition d'affinités sur les différents axes (affinité de rapport  $\sqrt{m_{it}}$  pour le juge  $i$  et l'axe  $t$ ).

- Positionnement multidimensionnel :

L'objectif de la méthode INDSCAL est voisin de celui des analyses PROCUSTE. Dans les deux cas, on peut considérer qu'à chaque juge est associée une configuration des objets dans un espace euclidien et qu'il s'agit d'estimer, sous différentes hypothèses, les transformations qui font passer de l'une à l'autre. Simplement, INDSCAL, plus restrictif, interdit les rotations que le modèle PROCUSTE envisage.

A1 - Modèle INDSCAL, Résolution :

La résolution du modèle INDSCAL consiste en la recherche des matrices  $X$  et  $(M_i)_{i=1,m}$  réalisant conjointement le minimum de la quantité :

$$\sum_i \text{tr}(E_i E_i) = \sum_i \|E_i\|^2$$

avec  $\|E_i\|^2 = \sum_{jk} e_{ijk}^2$

Une transformation simple de (2) permet de résoudre le problème par une méthode itérative de décomposition de tableaux à trois indices (Méthode NILES : "Non linear Iterative LEast Squares").

De façon plus générale, le système (2) s'écrit :

$$W_{ijk} = \sum_{t=1,p} x_{jt} m_{it} y_{kt} + e_{ijk} \quad (3)$$

La symétrie des matrices  $W_i$  et  $M_i$  assure que la solution définitive peut être choisie de manière à réaliser la condition :

$$y_{kt} = x_{kt} \quad (\forall k, \forall t)$$

stipulée par le modèle.

Nous donnons dans la suite l'exposé de la méthode NILES, et le détail des calculs.

Méthode NILES :

Il s'agit d'une méthode d'estimation par relaxation.

Partant d'une solution initiale quelconque fixant X et Y ; on estime la matrice  $M = (m_{it})_{\substack{i=1,m \\ t=1,p}}$

définissant les m métriques, de façon à minimiser la norme des résidus (Régression classique).

On utilise cette première solution, ainsi que la matrice Y de départ pour estimer la configuration X.

L'itération suivante estime Y à partir des estimations précédentes de M et X, et ainsi de suite.

L'algorithme converge vers un optimum local.

La dernière étape consiste à réaliser effectivement l'égalité de X et Y puis à procéder à une ultime estimation de M.

On obtient la solution définitive en normant les axes et en modifiant les poids en conséquence.

(Diviser les coordonnées sur un axe  $X_j$  par  $\sigma_j$  revient à multiplier pour chaque juge i le poids  $m_{ij}$  par  $\sigma_j^2$ . Cette transformation ne modifie naturellement pas les résidus).

Calculs pratiques :

Bien que la méthode s'applique de façon semblable quelle que soit la nature de la matrice estimée (poids ou coordonnées) nous donnons ici le détail des calculs qui éclairent les permutations réalisées.

a) Estimation de M :

Le système (2) s'écrit :

$$w_{is} = \sum_{t=1,p} m_{it} g_{st} + e_{is} \quad \text{avec : } \begin{array}{l} i=1,m \\ j=1,n \\ k=1,n \\ s = (j-1)n + k \\ s = 1,n^2 \end{array}$$

$$W = MG' + E \quad g_{st} = y_{kt} x_{jt}$$

$$(m \times n^2) \quad (m \times p)(p \times n)^2 \quad (m \times n^2)$$

L'estimation  $\hat{M}$  de M s'obtient par projection orthogonale des lignes de W sur le sous-espace engendré par les colonnes de G.

La matrice du projecteur est :

$$G(G'G)^{-1} G'$$

Il vient :

$$\hat{W} = \hat{M}G' = WG(G'G)^{-1} G'$$

$$\hat{M} = WG(G'G)^{-1}$$

$$(m \times p) \quad (m \times n^2)(n^2 \times p)(p \times p)$$

Explicitons un élément particulier de  $\hat{M}$  pour  $x=1,m$  et  $y=1,p$  on a :

$$[\hat{M}]_{xy} = \sum_{s=1,n^2} w_{xs} [G(G'G)^{-1}]_{sy}$$

$$= \sum_{s=1,n^2} w_{xs} \sum_{t=1,p} g_{st}(G'G)^{-1}_{ty}$$

· Si :  $s = (j - 1), + k$   $j=1,n$   
 $k=1,n$

$$[\hat{M}]_{xy} = \sum_{j=1,n} \sum_{k=1,n} w_{xjk} \sum_{t=1,p} (y_{kt} x_{jt})(G'G)^{-1}_{ty}$$

b) Estimation de X :

De façon similaire il vient :

$$W_{js} = \sum_{t=1,p} x_{jt} g_{st} + e_{js}$$

où  $j=1,n$   
 $s=1,nm$  avec  $s = (k-1)m + i$   
 et  $g_{st} = m_{it} y_{kt}$

$$W = XG' + E$$

(n x nm) (n x p)(p x nm) (n x nm)

et  $\hat{X} = W G(G'G)^{-1}$

(n x p) (n x nm)(nm x p)(p x p)

Soit pour  $x=1,n$   
 $y=1,p$

$$[\hat{X}]_{xy} = \sum_{s=1,nm} w_{xs} [G(G'G)^{-1}]_{sy}$$

$$= \sum_{s=1,nm} w_{sx} \sum_{t=1,p} g_{st} (G'G)^{-1}_{ty}$$

et en décomposant  $s$  :

$$[\hat{X}]_{xy} = \sum_{k=1,n} \sum_{i=1,m} w_{ixk} \sum_{t=1,p} (m_{it} y_{kt}) (G'G)^{-1}_{ty}$$

c) Estimation de  $Y$  :

$$W_{ks} = \sum_{t=1,p} y_{kt} g_{st} + e_{ks}$$

ou  $k=1,n$   
 $s=1,mn$   $s = (i-1)n + j$   
 $g_{st} = x_{yt} m_{it}$

$$W = YG' + E$$

(n x nm) (n x p)(p x mn) (n x mn)

Soit :

$$Y = W G(G'G)^{-1}$$

(n x p) (n x mn)(mn x p)(p x p)

$$[\hat{Y}]_{xy} = \sum_{i=1,m} \sum_{j=1,n} w_{ijx} \sum_{t=1,p} x_{jt} m_{it} (G'G)^{-1}_{ty}$$

Un programme BASIC opérationnel sur MICRAL 9050 B sous CP/M est disponible pour résoudre le modèle INDSCAL.

#### A2 - Application à l'étude des cantons du Languedoc-Roussillon :

Dans ce qui suit, nous détaillerons les particularités et les développements du modèle INDSCAL liés à la nature des données sur lesquelles nous entendons l'appliquer.

Deux approches fondamentales et complémentaires peuvent être envisagées, que nous étudierons successivement.

Dans l'une et l'autre étude les juges correspondent aux différentes années des recensements.

L'étude 1 se réfère à la structure des C.S.P., à ses constantes et à son évolution temporelle. L'étude 2 se fixe le même objectif, mais, cette fois, relativement à la structure des cantons.

Usuellement dans le modèle INDSCAL, les juges sont censés proposer directement une matrice de distance ou de similarité entre les stimuli.

La situation dans le cas présent est plus riche dans la mesure où les données brutes fournissent déjà une représentation euclidienne des stimuli dans l'espace des variables-cantons pour l'étude 1, dans celui des variables-C.S.P. pour l'étude 2. Il sera tiré parti de cet état de choses pour compléter notre interprétation des résultats classiques.

Dans toutes les méthodes statistiques d'analyse de données évolutives l'un des problèmes majeur est de proposer un référentiel commun judicieux, pour représenter les données instantanées. Le modèle INDSCAL apporte un outil d'analyse dans la limite des hypothèses qu'il stipule.

Nous supposerons conformément à ce modèle, que les distances mutuelles entre stimuli, dépendent de caractères stables au long des années, seules l'importance relative de ces caractères dans l'appréciation des distances étant sujette à évolution. La résolution du modèle fournit à la fois les caractères essentiels et les pondérations contingentes.

#### A3 - Particularisation du modèle :

L'application du modèle INDSCAL aux données socio-professionnelles nécessite le choix d'une distance entre stimuli.

Notre but est de comparer entre elles les répartitions des différentes C.S.P. à travers chaque canton, et d'opérer de façon duale en étudiant la répartition des différents cantons pour chaque C.S.P. La métrique du CHI-2 apparaît comme naturelle entre profils, elle permettra en outre de comparer directement nos résultats avec les méthodes d'analyse de suites de tableaux dérivées de l'analyse factorielle des correspondances.



Soit  $N_{IJ}$  le tableau des effectifs  $I =$  ensemble des 42 cantons  
 $J =$  ensemble des 9 C.S.P.

On définit alors, pour l'année  $k$  :

- $P_J^I(k)$  : le tableau des lois conditionnelles sur  $J$  sachant  $i \in I$
- $P_I^J(k)$  : le tableau des lois conditionnelles sur  $I$  sachant  $j \in J$
- $D_{P_i}^J(k)$  : (resp.  $D(k)$ ) les matrices diagonales des poids des éléments de  $I$  (resp.  $J$ ).
- $D(k)$  : (resp.  $D(k)$ ) les métriques du CHI-2 sur  $J$  (resp.  $I$ ).

Dans l'étude 1, l'analyse porte sur la suite de tableaux de produits scalaires entre les profils centres concernant les C.S.P. :

$W(k)$   $k=1,4$  avec :

$$W(k) = {}^t [ P_I^J - P_I^J \frac{D(k)}{P_i} \mathbf{1} ] \frac{D(k)}{1/P_J} [ P_I^J - P_I^J \frac{D(k)}{P_i} \mathbf{1} ]$$

-ou  ${}^t A$  signifie transposée de  $A$  et ou  $\mathbf{1}$  représente le vecteur colonne de dimension convenable dont toutes les composantes sont égales à l'unité.

Dans l'étude 2, on a de façon symétrique pour les profils centres concernant les cantons :

$Z(k)$   $k=1,4$  avec :

$$Z(k) = {}^t [ P_J^I - P_J^I \frac{D(k)}{P_J} \mathbf{1} ] \frac{D(k)}{1/P_i} [ P_J^I - P_J^I \frac{D(k)}{P_J} \mathbf{1} ]$$

Enfin, le modèle INDSCAL suppose un certain nombre de normalisations. Elles concernent le choix de l'origine du repère commun, et la variance de chacun des axes. Il est usuel de considérer comme origine le centre de gravité du nuage des stimuli. Ceci pose naturellement la question de leur pondération. Une première optique consiste à considérer les pondérations des stimuli comme variables selon les juges (matrices diagonales  $D(k)$  et  $D(k)$  pour le juge  $k$ ).

Ceci peut être réalisé en travaillant sur les opérateurs  $W(k) \frac{D(k)}{P_i}$  et  $Z(k) \frac{D(k)}{P_J}$ ,

mais n'est pas compatible avec le modèle INDSCAL proprement dit.

Nous choisirons une seconde optique qui réalise l'équipondération des stimuli. Le nuage des représentations sera donc centré pour la métrique identité, ce qui

revient à opérer préalablement à l'étude, un double centrage des matrices  $W(k)$  et  $Z(k)$ . La variance de chaque axe sera alors choisie unitaire.

Un dernier choix est laissé à l'utilisateur qui peut souhaiter contrôler le poids de chacun des juges dans l'analyse. Ceci est réalisé par la normalisation des matrices de produits scalaires. La norme utilisée est celle, usuelle, de Hilbert-Schmidt. La norme initiale des matrices dans l'étude 1 est approximativement constante, l'équipondération des juges est donc ainsi, automatiquement réalisée. En ce qui concerne l'étude 2, le poids de l'année 1954 excédant celui des autres années de façon importante, nous choisirons d'imposer la contrainte de normalisation.

#### A3 - Arguments d'interprétations :

Nous considérons pour chacune des études trois arguments graphiques. Chaque représentation est réalisée dans le système des axes communs.

- les stimuli selon la configuration commune
- les juges, avec les poids de chaque axe comme système de coordonnées
- les variables, représentées suivant leur corrélation avec les axes.

Ces résultats graphiques seront interprétés à la lumière des indices habituels de mesure de la qualité de représentation.

## B - RESULTATS - INTERPRETATION

#### B1 - Analyse marginale :

L'étude directe des marges des tableaux d'effectifs annuels, permet de situer notre analyse qui s'inscrit dans un contexte de croissance de la population active (environ +10 %). Cette croissance est largement hétérogène, il se produit en fait une forte concentration, la population se regroupant de façon générale dans quelques grands centres des départements du Gard de l'Hérault ou des Pyrénées Orientales, et délaissant l'ensemble des cantons de l'Aude et de Lozère.

Sur l'ensemble des cantons, on observe une baisse continue des C.S.P. agricoles (EA,OA), parallèlement à une croissance en continue accélération des C.S.P. tertiaires, tandis que, après une forte augmentation, l'effectif des ouvriers non agricoles, stagne depuis 1968. L'effectif des artisans commerçants subit une hausse modérée.

#### B2 - Etude n° 1 - les C.S.P. comme stimuli :

Dans cette étude, la représentation des stimuli rend compte des distances

mutuelles (au sens du CHI-2) entre les profils. Un profil-C.S.P. représente la distribution des effectifs de cette C.S.P. parmi l'ensemble des cantons. En conséquence, 2 points seront proches si dans de nombreux cantons, les effectifs relatifs sont voisins.

L'analyse montre que trois axes sont nécessaires pour représenter correctement la configuration commune aux quatre années. La qualité globale de reconstruction des matrices de produits scalaires est largement satisfaisante ( $\cos^2=0,9$ ). C'est en 1975 que la reconstruction des données est la moins bonne ( $\cos^2=0,8$ ).

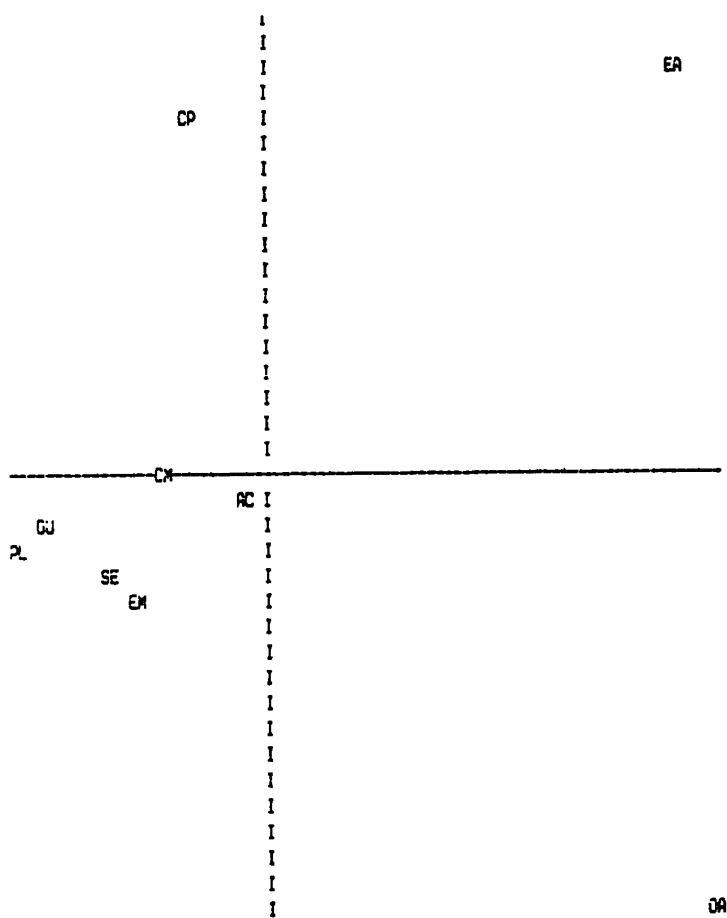


Figure 1 : Configuration commune des C.S.P. pour les 4 années  
(Plan 1 x 2)

Il apparaît que les 3 axes de la configuration commune sont non significativement corrélés entre eux, relativement aux stimuli ; Cela signifie que l'orthogonalité des axes stipulée par le modèle (métriques diagonales), correspond pour les juges à une notion de non corrélation.

	axe 1	axe 2	axe 3
axe 1	1,00		
axe 2	+0,02	1,00	
axe 3	+0,22	-0,03	1,00

Corrélation entre les axes relativement aux stimuli.

Il en résulte que les représentations graphiques faisant intervenir les variables initiales et leurs corrélations avec les axes pourront se lire comme des cercles de corrélation, les axes formant approximativement une base orthonormée pour la covariance. De la même façon, les poids donnés aux axes par les juges peuvent s'interpréter en termes de variance expliquée.

Cette remarque sera valable également pour l'étude n° 2.

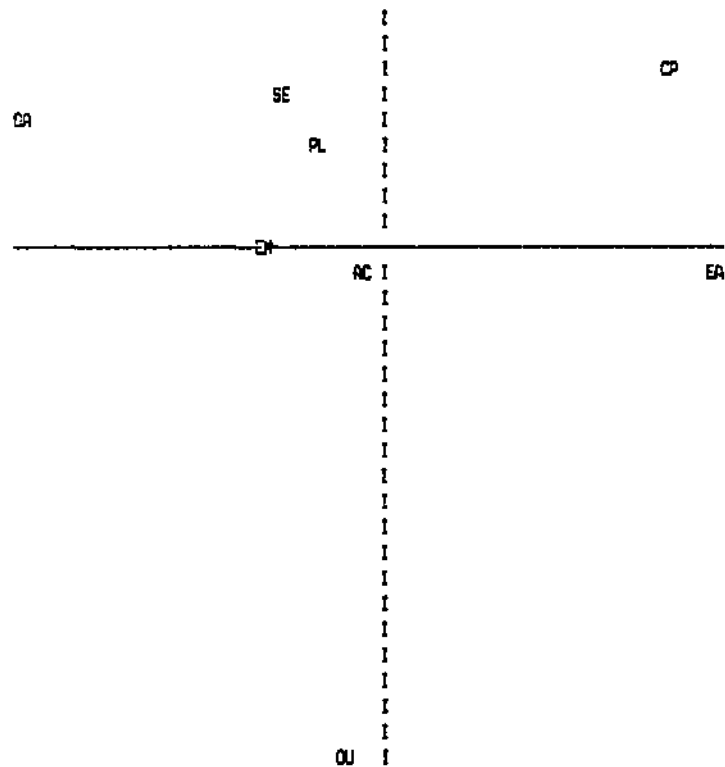


Figure 2 : Configuration commune des C.S.P. pour les 4 années  
(Plan 2 x 3)

Une structure très stable en trois principaux groupes, organise l'ensemble des C.S.P., la principale opposition se réalise entre l'agricole et le non agricole (importance relative de l'axe 1 : 82 %), a un second niveau, l'agricole se divise en exploitants et ouvriers (imp. relat. axe 2 : 12 %), tandis que le troisième axe isole les ouvriers non agricoles (imp. relat. axe 3 : 6 %).

La représentation graphique des poids données aux axes par les juges permet de distinguer des évolutions dans ce schéma de base.



Figure 3 : Poids donnés aux axes suivant les années  
(Plan 1 x 2)

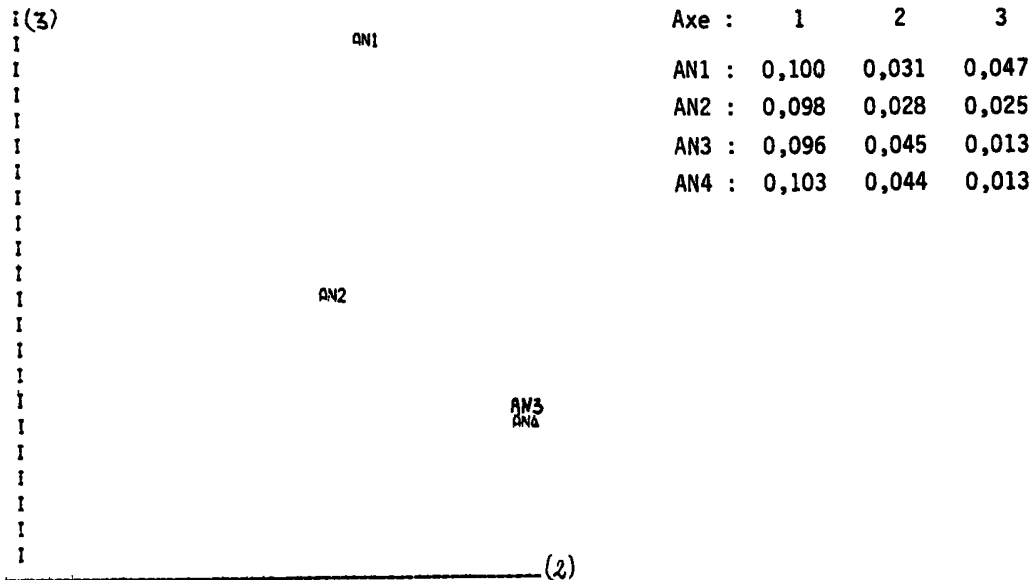


Figure 4 : poids donnés aux axes suivant les années  
(Plan 2 x 3)

Un nouvel éclairage peut être apporté sur les interprétations précédentes, en effectuant une analyse "duale", étudiant les corrélations des variables-cantons avec les trois axes précédents. Les variables-cantons, pour l'année k, les lignes du tableau  $P(k)$ .

On obtient naturellement quatre graphiques de corrélation. Les graphiques relatifs aux trois premières années se trouvent être excessivement voisins, par contre le quatrième, présente des modifications, témoins d'un changement de structure et cause de la qualité moindre de la reconstruction de  $W(4)$ , déjà notée. Ces modifications restent néanmoins marginales, car seulement liées à nombre restreint de cantons.

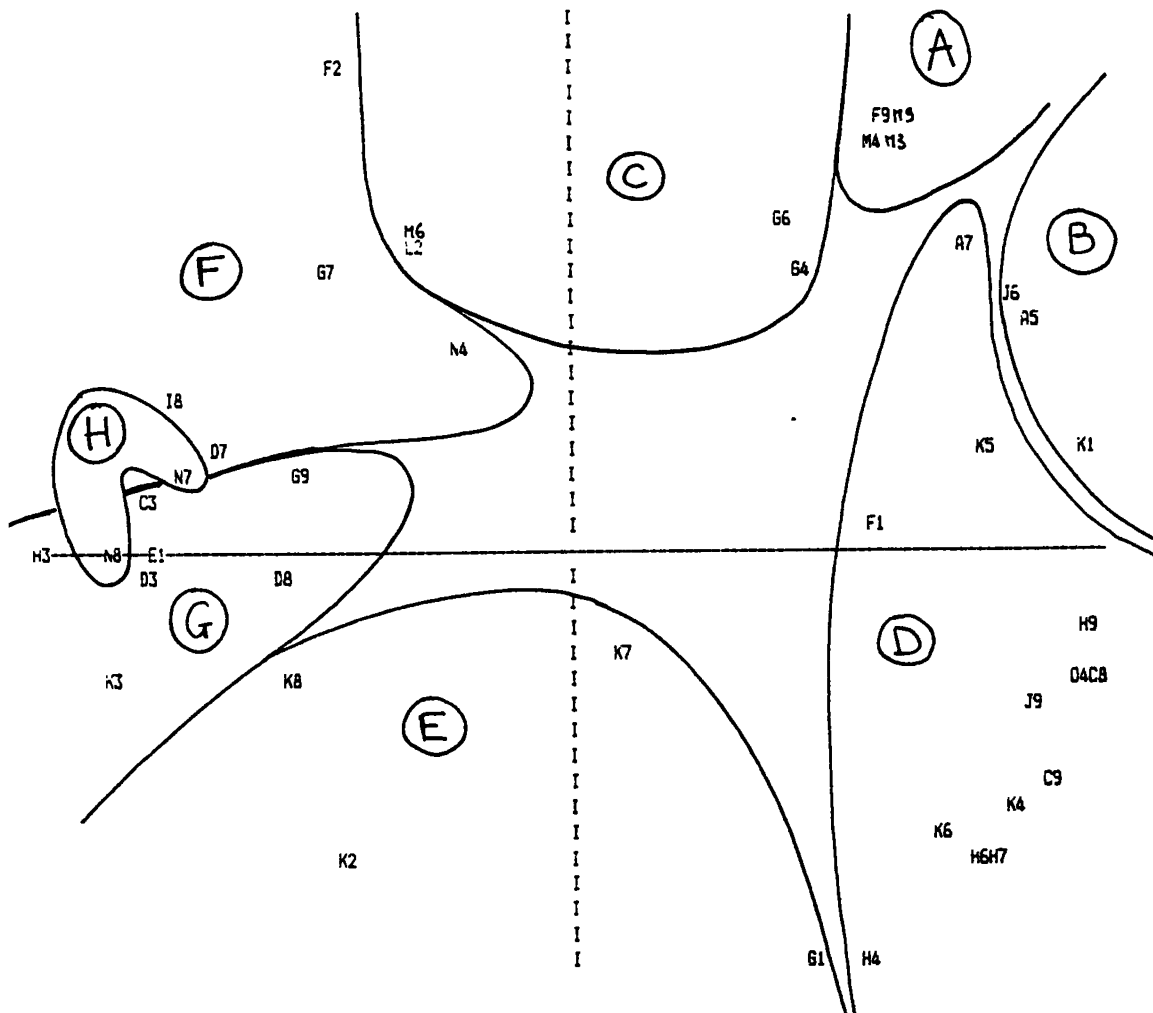


Figure 5 : Corrélations des variables-cantons avec les axes communs (années 1954, axe 1 horiz. axe 2 Vertic.)

On remarque la coïncidence très forte existant entre ce graphique et celui proposé dans [1]. En particulier, on retrouve tout à fait la classification des cantons en huit groupes qui y est établie, et qui est reproduite ici.

Groupes :

- |                                    |                                |
|------------------------------------|--------------------------------|
| A : agricole dominante exploitante | B : agricole "moyen"           |
| C : agricole en déclin             | D : origine agricole, régénère |
| E : rural-urbain                   | F : rural-industriel           |
| G : industriel                     | H : non agricole               |

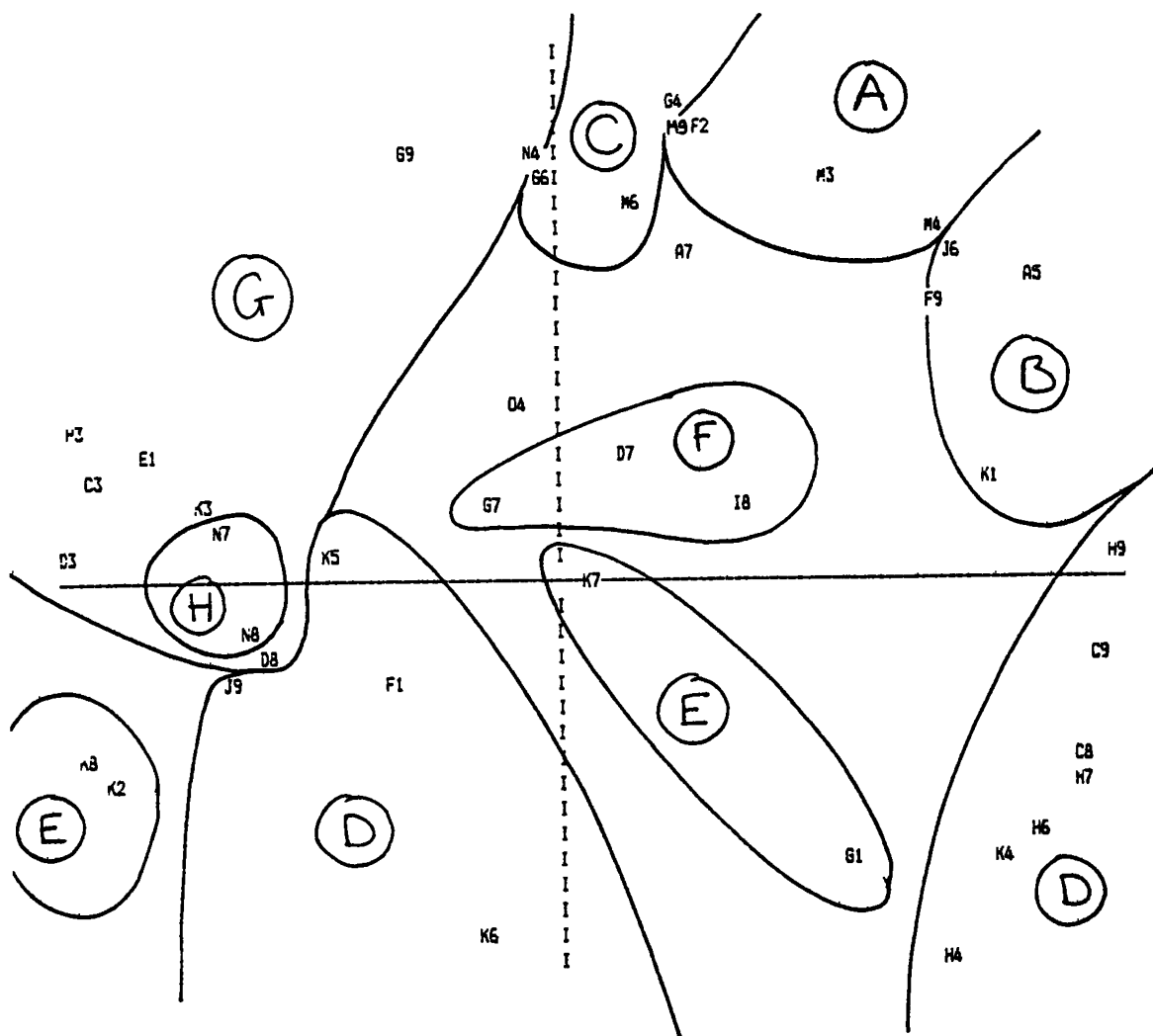


Figure 6 : Corrélations des variables-cantons avec les axes communs  
(année 1975, axe 1 horiz. axe 2 vertic.)

On retrouve sur ce graphique la confirmation des évolutions enregistrées dans l'étude directe des C.S.P. Les axes gardent néanmoins globalement la même signification, qui nous sert de référence pour évaluer les modifications existantes. En 1975, certains cantons ont une brusque évolution, l'interprétation de certaines classes prises dans leur ensemble varie :

Ainsi, le groupe F perd son caractère pertinent (déplacement vers le centre), les cantons N4 et F2 participant alors nettement au groupe agricole en déclin.

De même, la classe D éclate, certains cantons (O4,A7) devenant agricole en déclin (gr C), d'autres (J9,K5,F1,K6) se déplaçant vers la zone des groupes de type non agricole (grs G,H,F).

Il en va de même pour le groupe E, qui, rural-urbain à l'origine, se scinde en deux sous-groupes : type urbain-industriel (K8,K2) et type agricole moyen ou régénéré (K7,G1).

D'autre part, le nombre des cantons fortement corrélés avec l'axe 3 (ouvrier) diminue sensiblement (ex : G9,D7,K8,J9,K4,H9). (cf. Fig. 7)

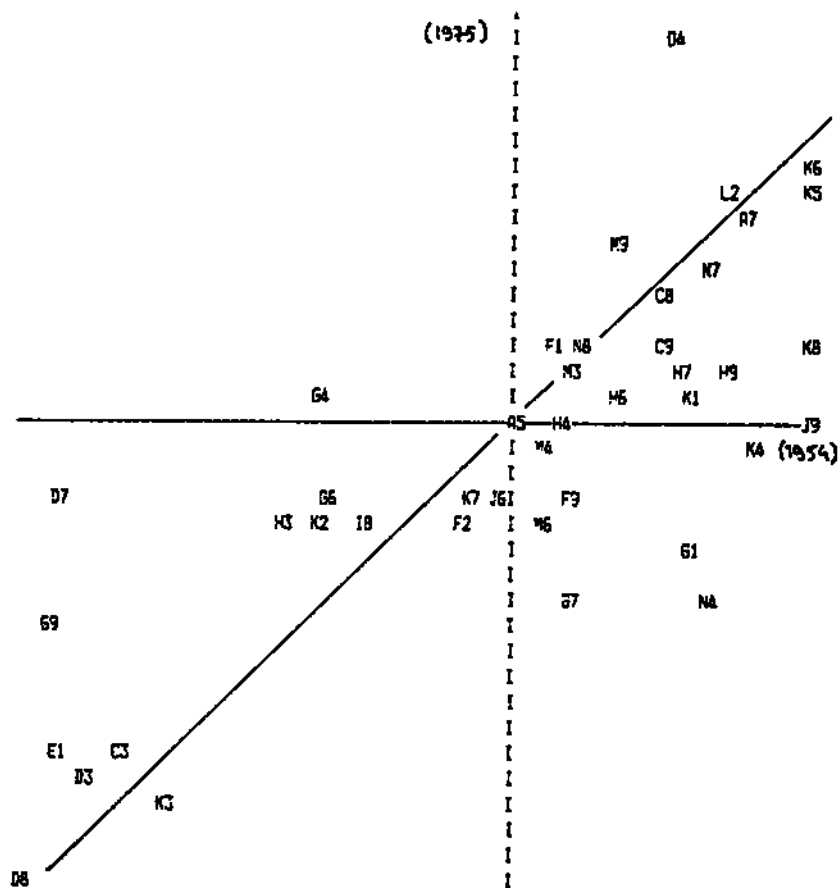


Figure 7 : Liaison entre les corrélations des variables-cantons avec l'axe 3 en 1954 et en 1975.

(nb : axes corrélés entre relativement aux stimuli)



B3 - Étude n° 2 - Les cantons comme stimuli :

Comme dans l'étude 1, la métrique choisie est celle du CHI-2 entre les profils-cantons. Ceux-ci représentent la distribution des effectifs d'un canton parmi les 9 C.S.P.

La configuration commune comporte 2 axes, qui suffisent pour représenter de façon satisfaisante la structure des cantons (reconstruction globale des  $Z(k)$ , ( $\cos^2=0,87$ ). L'année 1975 est la moins bien reconstruite ( $\cos^2=0,76$ ) ; Une dimension supplémentaire apparaît en fait, un troisième axe, non pertinent aux recensements antérieurs permet d'obtenir une reconstruction significativement meilleure de 1975 ( $\cos^2=0,87$ ).

Cette fois encore, les axes sont non significativement corrélés relativement aux stimuli ( $\text{cor}(1;2)=0,15$ ).

En ce qui concerne les poids accordés aux axes suivant les années, ceux-ci sont équivalents pour le premier, mais le second voit son influence décroître (cf. Fig. 9). L'importance relative de l'axe 1 est de 87 %.

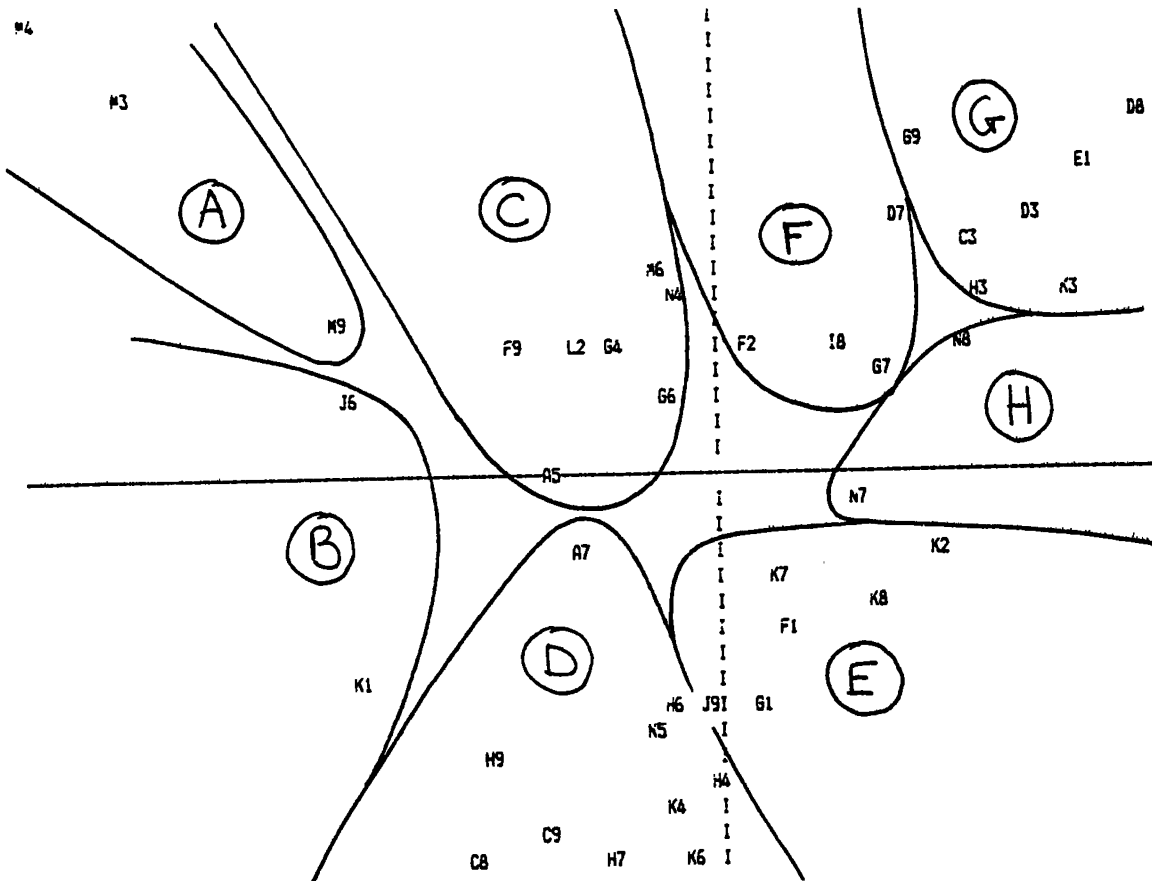


Figure 8 : Configuration commune des cantons pour les 4 années  
(axe 1 : horiz. axe 2 : vertic.)

.99.

		Axe :	1	2
I				
I				
I	AN1	AN1 :	0,200	0,115
I				
I	AN2	AN2 :	0,226	0,093
I		AN3 :	0,232	0,064
I		AN4 :	0,213	0,045
I	AN3			
I	AN4			
I				
I				
I				
I				

---

Figure 9 : Poids donnés aux axes suivant les années  
(axe 1 : horiz. axe 2 : vertic.)

On retrouve de façon évidente à la fois la structure mise en évidence dans l'analyse précédente et surtout celle fournie par l'analyse des correspondances et la classification automatique dans l'article de la Soc. Languedocienne de Géographie. Les huit groupes ont été reportés sur la figure 8.

Nous étudierons comme ci-dessus, les corrélations des variables-C.S.P. avec les axes communs suivant les années. (Ces variables sont les colonnes du tableau des profils :  $P(k)$  pour l'année  $k$ ). Les résultats des quatre années figurent sur un même graphique, le troisième caractère du code C.S.P. signifiant le rang de l'année. (cf. Fig. 10)

On observe trois pôles fixes (EA,OA,OU), autour desquels, durant les 20 années s'organisent les cantons. Ces corrélations confirment l'interprétation donnée aux huit groupes. Au contraire, trois C.S.P. évoluent dans le temps de façon très cohérente (CM,PL,EM, et SE). Ces quatre variables initialement corrélées positivement à OU, et opposées tant à EA qu'à OA, évoluent vers une orthogonalité à ces trois pôles. Les C.S.P. tertiaires tendent à se répartir de façon uniforme parmi les cantons.

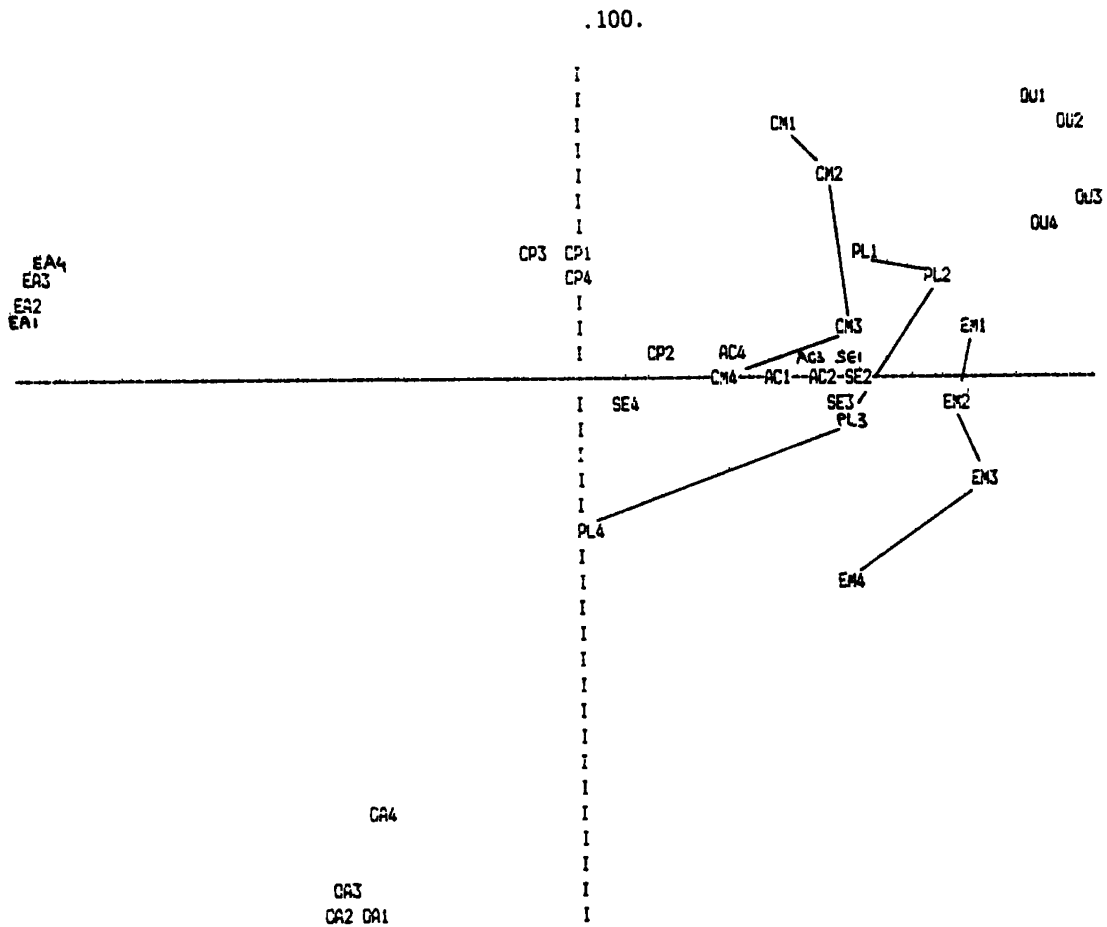


Figure 10 : Corrélations des variables-C.S.P. avec les axes communs  
(axe 1 : horiz. axe 2 : vertic.)

L'analyse du troisième axe, qui émerge en 1975, montre que, si celui-ci reste très lié à l'axe 1 précédent, il se trouve également précisément positivement corrélé avec ces quatre C.S.P. (principalement EM, CM, PL). Celles-ci tendraient donc à former un type à part entière, au même titre que OU, AE et OA ; Ce type nouveau s'affirme par opposition au groupe à forte densité d'exploitants agricoles.

## C - CONCLUSION

Les résultats apportés par la méthode utilisée ici sont cohérents avec ceux obtenus par plusieurs autres techniques statistiques d'analyse de données, même si elles ne se réfèrent pas nécessairement à la métrique du CHI-2. Il existe donc une structure forte organisant les cantons et les C.S.P.

La finesse des interprétations, inhabituelle dans les applications usuelles du modèle INDSCAL provient de la richesse des données, et en particulier de la connaissance que l'on a d'une représentation euclidienne a-priori, à la fois en ce qui concerne les C.S.P. et les cantons.

Le modèle INDSCAL appliqué à des matrices de distances ou de similarité cherche uniquement des représentations synthétiques. Il est inapte pour l'étude des cas particuliers. Dans le cas présent, l'étude des corrélations entre les variables et les axes, a permis, non seulement, de préciser leur interprétation, mais aussi de particulariser les trajectoires des variables.

Cette sorte de dualité, intéressante sur le plan pratique, paraît, ce qui est regrettable, difficile à formaliser.

D'un autre point de vue, il serait intéressant de compléter le modèle algébrique par un modèle probabiliste, qui permette de juger du caractère significatif ou non, des valeurs ou des variations des poids accordés par les juges à un axe. Dans cette voie, il serait aisé d'analyser les termes résiduels pour en tirer des enseignements plus fins sur la qualité des représentations, et la validité du modèle.

## D - BIBLIOGRAPHIE

- [1] F. AURIAC, MC BERNARD (1974)  
"Composantes et types socio-professionnels des campagnes du Languedoc-Roussillon"  
Bulletin de la Société Languedocienne de Géographie  
Tome 8 - Fascicule 1 - pp : 5-17
- [2] F. AURIAC, MC BERNARD (1975)  
"Le changement social dans les campagnes Languedociennes"  
L'espace géographique - n° 4 - pp : 239-250

[ 3 ] D. CARROLL, JJ. CHANG (1970)

"Analysis of individual differences in multidimensional scaling via an n-way generalization of Eckart-Young decomposition"

Psychometrika Vol. 35 - N° 3 - pp : 283-311

[ 4 ] H. WOLD (1966)

"Non linear estimation by iterative least squares procedures"

WILEY - NEW YORK