

STATISTIQUE ET ANALYSE DES DONNÉES

THIERRY FOUCART

Application de la méthode RAS

Statistique et analyse des données, tome 10, n° 1 (1985), p. 65-82

http://www.numdam.org/item?id=SAD_1985__10_1_65_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 LA METHODE RAS

Thierry Foucart

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

Résumé: *La méthode RAS est une procédure classique d'ajustement d'un tableau de contingence à des marges fixées. Nous l'avons développée et appliquée pour étudier la suite des tableaux donnant la répartition des cantons du Languedoc Roussillon suivant la CSP donnés par les recensements de 1954,62,68 et 75. Ce texte contient une description des développements proposés et les résultats obtenus, mais correspond à une démarche purement exploratoire.*

Abstract: *The algorithm RAS is a classical method to adjust a table to fixed margins. We generalize it to describe a sequence of boards, each of them being adjusted to the margins of the next one and use this approach to study commun data.*

Mots clés: *Algorithme RAS, ajustement des marges.*

Indices de classification STMA: *06-110.*

1 - LA METHODE

1.1 L'algorithme RAS [3]

L'algorithme RAS est une des nombreuses méthodes d'ajustement d'un tableau de contingence à des marges fixées. Il est particulièrement utilisé dans l'étude des tableaux d'échanges, où il donne en général de bons résultats, et dans les sondages, puisqu'il permet de caler un tableau de contingence calculé sur l'échantillon à des marges connues indépendamment de l'enquête.

La méthode est basée:

- sur un tableau initial A_{IJ} de marges A_I et A_J . Ce tableau initial peut être par exemple celui que l'on a obtenu lors du dépouillement de l'enquête.

- sur les marges fixées C_I et L_J . Ces marges résultent d'hypothèses effectuées par des spécialistes, comme dans le cas de tableaux d'échanges, ou sont déterminées par une enquête exhaustive.

L'algorithme est très simple dans son principe. Nous supposons que les marges fixées et le tableau initial sont de somme égale à 1. Cette hypothèse est implicite dans toute la suite du texte (sauf cas particulier).

(i) on ajuste le tableau A_{IJ} à la marge C_I en posant :

$$\forall i \in I \quad \forall j \in J \quad a_{i,j}(1)(1) = a_{i,j} \times C_i/a_i .$$

où a_i est le terme marginal de rang i du tableau A_{IJ} . On obtient ainsi le tableau $A_{IJ}(1)$

(ii) on ajuste le tableau $A_{IJ}(1)$ à la marge L_J en posant :

$$\forall i \in I \quad \forall j \in J \quad a_{i,j}(2) = a_{i,j}(1) \cdot L_j/a_{.j}(1)$$

où $a_{.j}(1)$ est le terme marginal de rang j du tableau $A_{IJ}(1)$. On obtient ainsi le tableau $A_{IJ}(2)$.

On recommence ensuite la procédure en considérant le tableau $A_{IJ}(2)$ comme tableau initial et ainsi de suite jusqu'à la convergence de l'algorithme.

On sait que, sous certaines conditions, cette convergence est assurée, le tableau limite étant donné par :

$$B_{IJ} = R_I A_{IJ} S_J$$

où R_I et S_J sont des matrices diagonales. On a donc :

$$\forall i \in I \quad \forall j \in J \quad b_{i,j} = r_i a_{i,j} s_j$$

On sait aussi que le tableau B_{IJ} minimise l'information discriminante de Kullback [1]

$$I = \sum_{i,j} b_{i,j} \text{Log} (b_{i,j} / a_{i,j})$$

qui est elle-même équivalente à la distance du χ^2 entre les deux tableaux à un terme du 2e ordre près.

Par ailleurs, la méthode de calcul conduit à des coefficients r_i et s_j positifs: les termes $b_{i,j}$ sont donc positifs, et nuls si et seulement si le terme initial est nul. Cette méthode conserve les zéros et ne fait pas apparaître de terme négatif.

1.2 Application de l'algorithme RAS à une suite de tableaux.

Soit $(N_{IJ}(t))$ $t = 1, k$ la suite de tableaux de probabilités à étudier. On note évidemment $N_I(t)$ et $N_J(t)$ les marges du tableau $N_{IJ}(t)$. Considérons le modèle ci-dessous :

$$\forall t = 1, k-1 \quad N_{IJ}(t+1) = R_I(t+1)N_{IJ}(t) S_J(t+1) + E_{IJ}(t)$$

où $R_I(t+1)$ et $S_J(t+1)$ sont les matrices diagonales telles que les marges du tableau $R_I(t+1) N_{IJ}(t) S_J(t+1)$ soient égales à $N_I(t+1)$ et

$N_j(t+1)$. Le tableau $E_{IJ}(t)$ est de terme général:

$$\forall i \in I \quad \forall j \in J \quad e_{i,j}(t) = n_{i,j}(t+1) - r_i(t+1) n_{i,j}(t) s_j(t+1)$$

Il est de marges $E_i(t)$ et $E_j(t)$ nulles et chaque terme peut être considéré comme un résidu.

Le modèle proposé est log linéaire. Les coefficients $r_i(t)$ et $s_j(t)$ représentent les facteurs d'évolution des modalités i et j de I et J respectivement. Trois cas se présentent:

$$\begin{aligned} r_i(t) > 1 & \quad (s_j(t) > 1) & \text{accroissement} \\ r_i(t) < 1 & \quad (s_j(t) < 1) & \text{diminution} \\ r_i(t) = 1 & \quad ; s_j(t) = 1 & \text{stagnation} \end{aligned}$$

Lorsqu'on étudie un tableau d'échanges, on parle fréquemment d'attraction et de répulsion. Cela ne signifie pas que les termes marginaux correspondants augmentent, diminuent ou restent stables. On a en effet :

$$\begin{aligned} \forall i \in I \quad n_{i.}(t+1) / n_{i.}(t) &= r_i(t) \sum_{j \in J} n_j^i(t) s_j(t) \\ \forall j \in J \quad n_{.j}(t+1) / n_{.j}(t) &= s_j(t) \sum_{i \in I} n_i^j(t) r_i(t) \end{aligned}$$

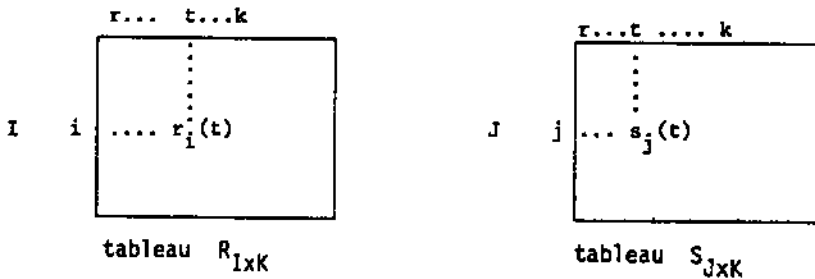
Supposons pour fixer les idées que les distributions conditionnelles sur J sachant i et i' sont égales :

$$j \in J \quad n_j^i(t) = n_j^{i'}(t)$$

On voit bien que l'accroissement des termes marginaux ne dépend plus que des coefficients $r_i(t)$ et $r_{i'}(t)$, qui mesurent donc un facteur d'évolution particulier à chaque ligne toutes choses égales par ailleurs. On peut évidemment tenir le même discours au sujet des coefficients $s_j(t)$.

1.3 Analyse des résultats

Comme nous l'avons dit dans l'introduction, nous ne faisons qu'explorer le modèle suggéré. Nous cherchons tout d'abord les propriétés des matrices $R_I(t)$ et $S_J(t)$. On peut les écrire sous la forme d'un tableau :

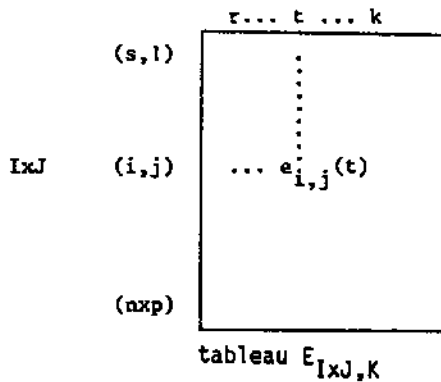


Il s'agit de tableaux à termes positifs: on sait les décrire par l'analyse en composantes principales ou l'analyse des correspondances.

La deuxième étude indispensable est celle des résidus contenus dans les tableaux $E_{IJ}(t)$. Plusieurs approches sont possibles pour analyser la suite de ces tableaux [2]:

- la méthode STATIS permet d'étudier les erreurs de reconstruction suivant les lignes d'une part et les colonnes d'autre part. Les tableaux étudiés concernant les mêmes individus I et les mêmes variables J, on pourra rechercher une certaine dualité dans les analyses.

- on peut également décrire les erreurs de reconstruction en effectuant l'analyse en composantes principales du tableau obtenu par juxtaposition des $E_{i,j}(t)$:



Ces deux analyses sont en fait complémentaires: la première met en évidence les lignes et les colonnes systématiquement mal reconstituées et la seconde particularise les couples (i, j) pour lesquels les effectifs $n_{i,j}(t)_{t=1, k}$ sont mal reconstruits.

Une persistance des erreurs dans le temps peut apparaître. Si cela se produit, cela signifie que le processus de transformation du tableau est mal décrit par le modèle.

1.4 Modèle de rang 1

Nous allons supposer l'existence d'une suite de tableaux de probabilités suivant un modèle de rang 1. On a alors, en supposant que les matrices $R_I(t)$ et $S_J(t)$ sont constantes:

$$N_{IJ}(t+1) = R_I N_{IJ}(t) S_J + E_{IJ}(t)$$

Pour un terme:

$$n_{i,j}(t+1) = r_i n_{i,j}(t) s_j + e_{i,j}(t)$$

Le modèle choisi implique que le tableau $R_I N_{I,J}(t) S_J$ est un tableau de probabilité.

Des équations précédentes on déduit:

$$n_{i,j}(t+k) = r_i^k n_{i,j}(t) s_j^k + \sum_{\ell=0}^{k-1} r_i^\ell e_{i,j}(t+k-\ell) s_j^\ell$$

S'il existe un produit $r_i \times s_j$ strictement supérieur à 1, le terme dépasse 1 pour une valeur suffisante de h . On a donc:

$$\forall i \in I \quad \forall j \in J \quad r_i s_j < 1$$

Le tableau $R_I N_{IJ}(t+1) S_J$ étant un tableau de probabilité, la somme de ses termes est égale à 1 :

$$\sum_{(i,j) \in I \times J} r_i n_{i,j}(t+1) s_j = 1 = \sum_{(i,j) \in I \times J} n_{i,j}(t+1)$$

Si l'un des produits $r_i \times s_j$ est inférieur strictement à 1, l'égalité ci-dessus est impossible et par conséquent:

$$\forall i \in I \quad \forall j \in J \quad r_i s_j = 1$$

$$\begin{array}{ll} \text{Cela implique évidemment:} & \forall i \in I \quad r_i = 1 \\ & \forall j \in J \quad s_j = 1 \end{array}$$

d'où:

$$\forall i \in I \quad \forall j \in J \quad n_{i,j}(t+h) = n_{i,j}(t)$$

En conclusion, il ne peut exister qu'un nombre fini de tableaux de probabilité se déduisant les uns des autres par un modèle de rang 1. Cela incite à étudier la suite des tableaux d'effectifs et non de probabilités, en prenant en compte donc dans le calcul des matrices $R_I(t)$ et $S_J(t)$ la somme des tableaux.

2 - APPLICATION: TABLEAUX CANTONS x CSP

Comme dans les autres études effectuées dans le cadre de ce GRECO, nous avons appliqué la procédure développée dans le paragraphe précédent à l'analyse des 4 tableaux donnant la répartition d'un certain nombre de cantons du Languedoc-Roussillon suivant leur CSP en 1954,62,68,75. Rappelons que le nombre de cantons est égal à 42 et que les CSP sont au nombre de 9. Le nombre de tableaux étant de 4, nous avons calculé les trois matrices $R_I(t)$ et $S_J(t)$ sur les tableaux de probabilités.

	AN21	AN32	AN43		AN21	AN32	AN43
A5	0.9340	0.9418	0.8982	M9	1.0561	1.0582	1.0027
A7	0.9931	0.9131	0.8804	I8	0.8736	0.7832	0.8041
C3	0.8461	0.8805	0.8435	J6	0.9296	0.9157	0.9637
C8	0.9825	1.0555	0.9155	J9	1.0329	1.2031	1.2976
C9	1.0088	1.0404	0.9405	K1	1.0751	1.0664	1.0802
D3	0.9335	0.9758	1.0153	K2	1.0143	0.9651	1.3033
D7	0.7176	0.6813	0.6966	K3	0.8766	0.7932	0.6565
D8	0.6303	0.6128	0.7240	K4	1.1553	1.1632	1.0361
E1	0.6689	0.7783	0.6648	K5	1.0487	1.2553	1.5844
F1	1.1671	1.4367	1.4543	K6	1.1192	1.1821	1.4858
F2	1.0255	1.0018	0.8920	K7	1.0266	0.9760	1.0027
F9	1.0033	0.9660	1.1058	K8	1.2311	1.2299	1.2840
G1	1.0418	1.2780	1.0011	M2	0.9435	0.9658	0.9518
G4	0.8690	0.8553	0.8628	L3	1.3366	1.0328	0.6439
G6	0.7700	0.6931	0.7746	M4	1.0293	1.0526	0.8588
G7	0.9438	0.9535	0.9089	M6	1.0548	0.8090	0.7611
G9	0.8464	0.7560	0.6611	M9	0.8104	1.0075	1.2074
H3	0.8586	0.7663	0.7905	N4	1.8485	0.8615	0.7990
H4	0.9649	1.0876	1.0002	N7	1.0120	0.9394	0.8145
H6	1.0583	1.0575	1.0674	N8	0.9669	0.8468	0.8045
H7	1.1020	1.0103	0.9119	O4	1.1868	1.2212	1.3162

coefficients r_i

2.1 Analyse en composantes principales des coefficients r_i

Les variables étant identifiées par AN21, AN32, AN43 correspondant au passage de 1958 à 1962, de 1962 à 1968 et de 1968 à 1975, examinons tout d'abord les moyennes, variances et corrélations.

On constate que la moyenne des coefficients diminue dans le temps avec un écart-type relativement élevé (coefficient de variation de 20%). La matrice des corrélations entre les trois variables contient des valeurs assez fortes qui diminuent lorsque l'écart entre les tableaux augmente. On remarque la valeur 0.79 du coefficient de corrélation linéaire entre les variables AN32 et AN43 définies par les suites des coefficients $r_i(2)$ et $r_i(3)$ pour $i=1,42$. Cette valeur très élevée laisse penser que les valeurs (centrées réduites) restent très proches les unes des autres pour chaque canton.

Toutes les corrélations sont positives. On a évidemment un effet taille et le premier vecteur propre est constitué de termes tous positifs, la 1ère valeur propre expliquant 70% de l'inertie totale. L'évolution des variables dans le temps n'est guère significative et n'apparaît que sur le 2e axe dont la contribution se limite à 25% (rappelons que le nombre de variables est égale à 3).

L'effet taille étant caractérisé par le 1er axe, on trouve sur la

partie droite de la figure 1 les cantons dont les coefficients $r_i(1)$, $r_i(2)$ et $r_i(3)$ sont supérieurs à la moyenne. Réciproquement, les cantons placés dans la partie gauche du même schéma ont des coefficients inférieurs à la moyenne. Enfin, l'axe 2 met en évidence des cantons dont les coefficients $r_i(1)$ ont pris une valeur particulière (passage du tableau 1 au tableau 2).

En résumé, les cantons placés à droite attirent la population et ceux qui sont à gauche la repoussent. On ne peut rien en déduire sur les termes marginaux correspondants dont l'évolution dépend du profil du canton et des coefficients s_j .

Pour comparer aux différentes études effectuées sur ces données, nous avons formé quatre groupes de cantons:

groupe I: il rassemble ceux qui attirent la population. Il s'agit de:

K5 (les Matelles), K6 (Mauguio), O4 (Perpignan): ce sont des cantons classés par les géographes "d'origine agricole régénérés".

F1 (Nîmes), J9 (Castries), K8 (Montpellier): cantons du type "rural-urbain"

groupe II: ce sont les cantons que la population a tendance à quitter:

D8 (La Grand'Combe), E1 (St Ambroix), K3 (Ganges), G9 (Sumène): cantons à caractère industriel

D7 (Génolhac): à caractère rural industriel

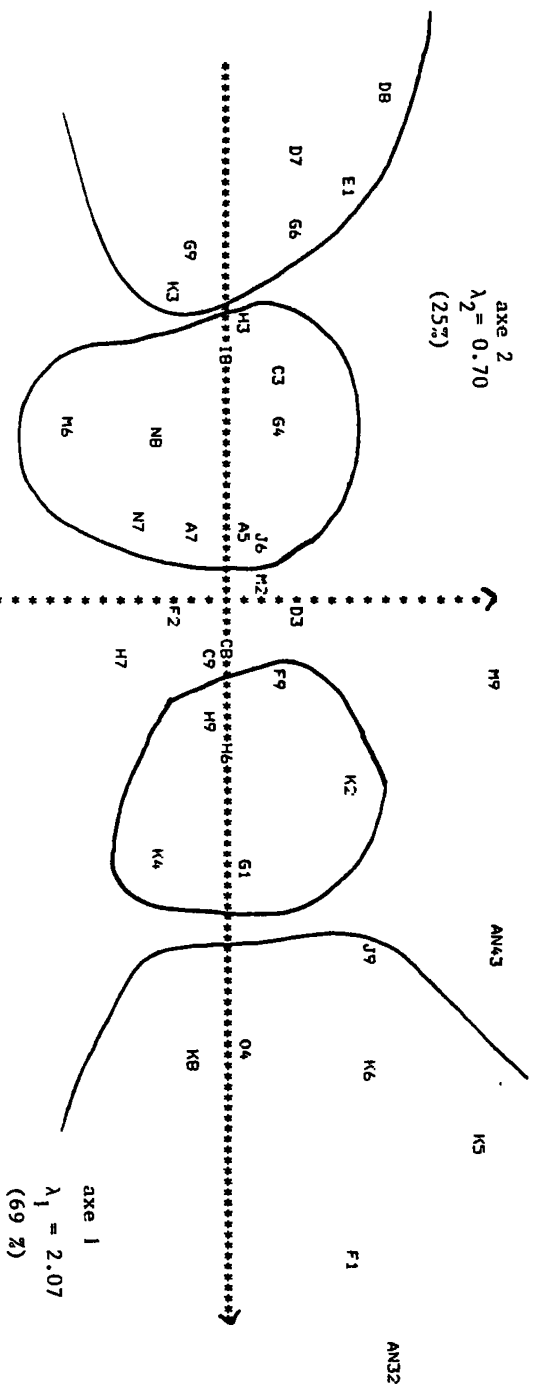
G6 (St André de Valborgne): agricole en déclin

groupe III: cantons dont l'attraction est relative (moins forte que celle du groupe I)

H4 (Agde), K4 (Lunel), H6 (Béziers), H9 (Montagnac): cantons d'origine agricole régénérés.

G1 (Vauvert), K2 (Frontignan): de type rural-urbain

F9 (Uzès): canton agricole en déclin



MOYENNES	ECARTS-TYPES	VARIANCES
AN21	.999843	.0374892
AN32	.975943	.0111064
AN43	.968279	.0534037

CORRELATIONS

*AN21 / AN32 / AN43 /
AN21/ 1.00000
AN32/ 0.4930 1.00000
AN43/ 0.2777 0.7934 1.00000

Figure 1 : plan principal 1x2 (ACP du tableau K1k)

AN4 AN21

groupe IV: légère tendance à l'émigration

H3 (Le Vigan), C3(Quillan): cantons industriels

I8 (St Pons): rural industriel

G4 (Lasalle), A5 (Castelnaudary), M6 (Langogne): agricole en déclin

N7 (Argelès sur mer), N.8 Arles sur Tech): non agricole

A7 (Fanjeaux): d'origine agricole régénéré

J6 (Lodève): agricole moyen

Les cantons qui ne figurent pas dans cette liste sont en général proches du centre de gravité du nuage sauf évidemment M3,M9,N4 qui se différencient par le coefficient $r_i(1)$.

Pour conclure cette analyse, nous allons comparer les résultats avec ceux de l'A.F.C. du tableau moyen déjà effectuée. On voit que les groupes I et III rassemblent surtout des cantons d'origine agricole régénérés ou du type "rural-urbain" alors que les groupes II et IV sont constitués en majorité de cantons de type "industriel", "rural-industriel" ou "agricole en déclin". L'axe 1 obtenu par l'A.C.P. des coefficients $r_i(t)$ correspond à l'axe 2 déterminé par l'A.F.C. du tableau moyen et dont l'interprétation est fournie par l'étude de la trajectoire de la marge relative aux cantons (tab1, tab2, tab3, tab4). Citons le texte (déjà au rédigé moment de la rédaction de ce paragraphe): "on constate une diminution de la population relative des cantons de type agricole en déclin, à dominante exploitant, rural industriel et agricole moyen, au profit des cantons de type rural-urbain."

La différence la plus caractéristique entre les deux analyses concerne le canton F9 (Uzès) classé dans le groupe III bien qu'il soit du type agricole en déclin. Il semble que sa part dans la population totale décroît en dépit d'un coefficient r_i supérieur à la moyenne (cf remarque du paragraphe 1.2).

Nous n'avons pu établir de relations entre la taille des coefficients r_j et la trajectoire des cantons donnée dans la figure 3 du texte sur l'A.F.C.

2.2 Etude des coefficients s_j

Nous avons effectué l'A.C.P. du tableau des coefficients $s_j(t)$ afin de pouvoir comparer les résultats obtenus avec ceux de l'A.F.C..

En réalité, la taille du tableau permet d'étudier ces coefficients par une simple lecture. La particularité des CSP Exploitants Agricoles (EA) et Ouvriers Agricoles (OA) est évidente: les coefficients s_j correspondants sont très nettement inférieurs à 1 et aux autres valeurs du tableau. En revanche, les CSP Professions Libérales (OL), Cadres Moyens (CM) et Employés (EM) paraissent très attractives surtout au cours de la dernière période où les coefficients prennent des valeurs supérieures à 1.5. Parmi les autres, on note la stabilité des Artisans-Commerçants (AC); la diminution du coefficient s_7 caractérise l'évolution de la CSP Ouvriers (OU) et le coefficient s_8 montre l'irrégularité de la CSP des Services (SE) surtout pendant la deuxième période.

Là encore, il est difficile d'établir une relation entre les coefficients s_j et les trajectoires des CSP obtenues par l'A.F.C. du tableau moyen.

On notera que la moyenne de ces coefficients évolue irrégulièrement en raison surtout des variations importantes de la dernière période. Par contre, les coefficients de corrélation sont encore élevés: 0.885 dans la dernière période. Il apparaît donc une certaine stabilité des coefficients centrés réduits.

L'A.C.P. n'apporte évidemment pas grande information supplémentaire. L'axe 1 met en évidence l'effet taille et l'on retrouve à gauche du schéma les CSP correspondants aux coefficients les plus grands (PL, CM, EM) et à droite évidemment les CSP EA et OA. Le deuxième axe est entière-

	AN21	AN32	AN43
EA	0.8074	0.7484	0.6454
OA	0.7681	0.6281	0.3789
AC	0.9896	0.9956	1.0132
PL	1.4080	1.5361	2.1085
CM	1.3584	1.4224	1.5666
EM	1.2028	1.3863	1.5402
OU	1.3074	1.2332	0.9660
SE	1.0682	1.2396	1.2454
CP	1.6202	0.8325	1.1525

	MOYENNES	ECARTS-TYPES
AN21	1.16912	.27022
AN32	1.11647	.301069
AN43	1.20186	.454946

CORRELATIONS

	*AN21 / AN32	AN32 / AN43
AN21/	1.00000	
AN32/	0.5283	1.00000
AN43/	0.6577	0.8849

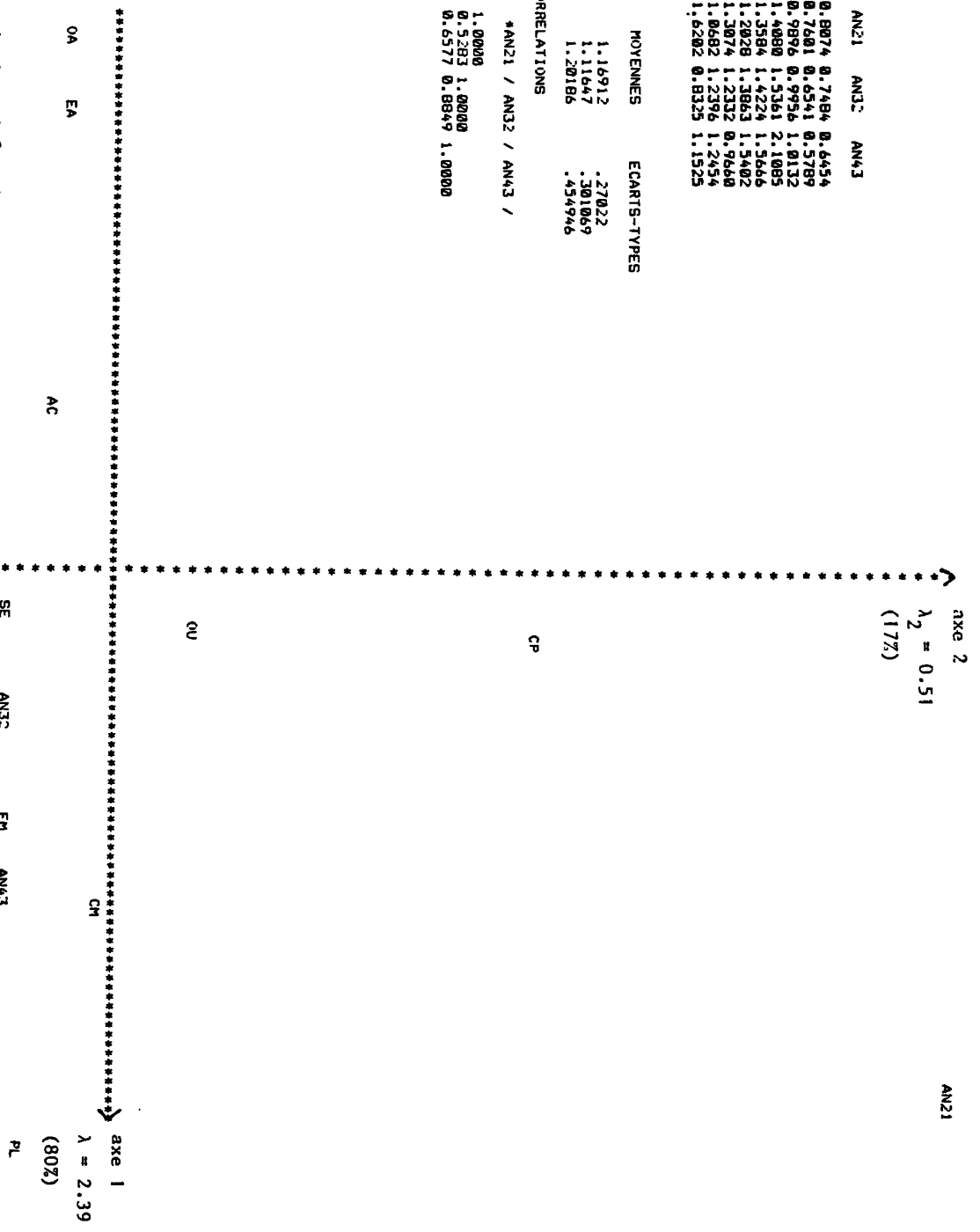


Figure 2 : plan principal 1x2 (ACP du tableau S_{IK}²)

ment déterminé par la CSP Services (SE) et le troisième par la CSP Ouvriers (OU). L'A.C.P. particularise donc ces deux derniers coefficients dont les valeurs sont plus ou moins orthogonales aux autres. Il faut aller jusqu'au 5e axe obtenu par l'A.F.C. du tableau moyen pour que la CSP Services contribue de façon significative à l'inertie. Il existe donc sur ce point une différence entre les deux analyses.

En conclusion, ces deux A.C.P. nous paraissent compléter ceux de l'A.F.C. du tableau moyen. On en retrouve les résultats principaux mais elles signalent des lignes et des colonnes ignorées de l'A.F.C.. Il faut évidemment chercher une explication à ces particularités, explication qui peut découler d'ailleurs du modèle choisi: il est en effet possible que, pour les CSP Ouvriers et Services (OU et SE) et le canton Uzès (F9), le modèle RAS soit inadapté: cela peut être vérifié par l'étude des résidus.

2.3 Validation du modèle RAS

Nous nous proposons d'étudier maintenant les tableaux $E_{IJ}(t)$ définis par :

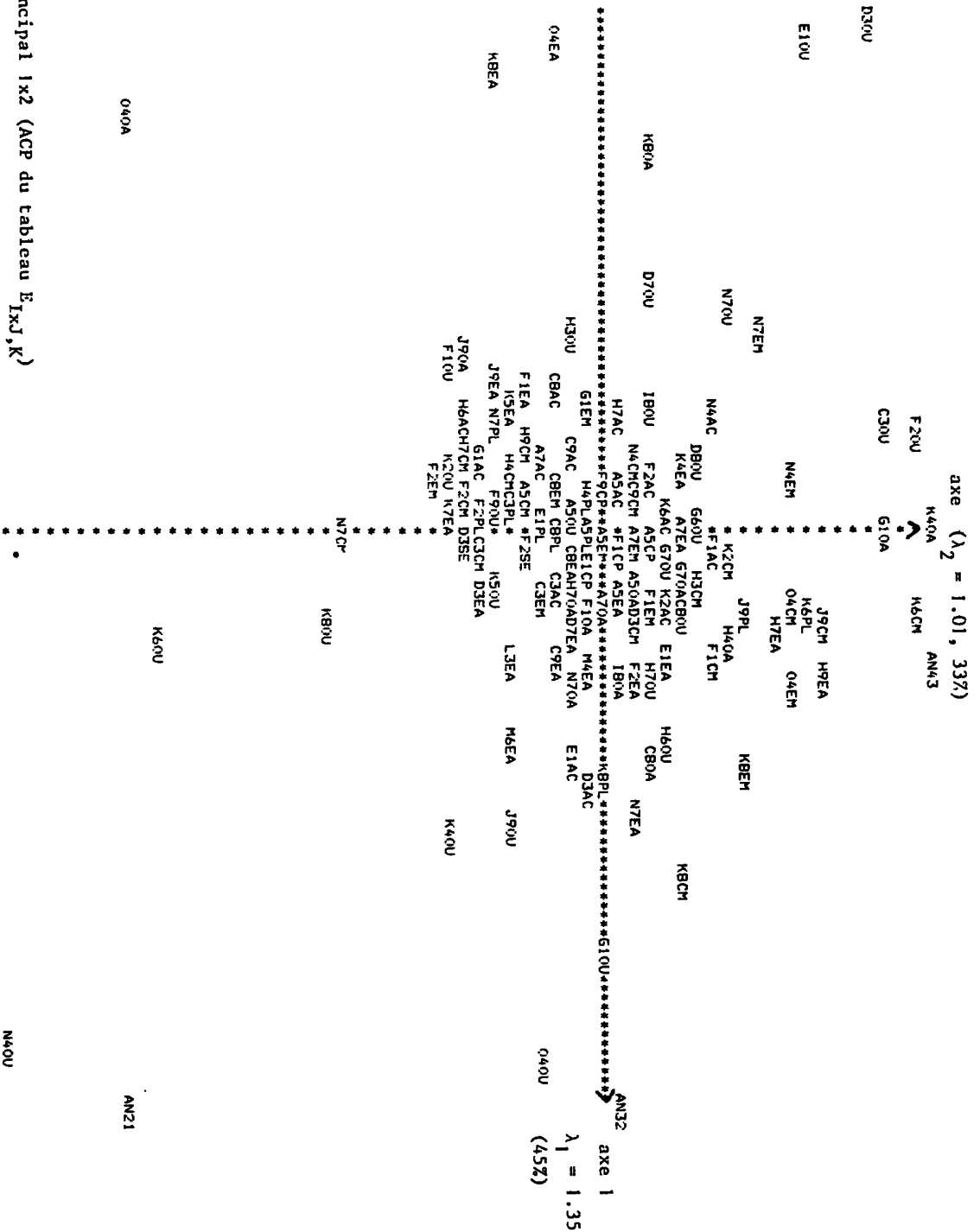
$$\forall i \in I \quad \forall j \in J \quad e_{i,j}(t) = n_{i,j}(t+1) - r_i(t) n_{i,j}(t) s_j(t)$$

Par construction, ces tableaux sont de marges (et de somme) nulles. L'erreur la plus grande en valeur absolue est de 0.00714, ce qui peut être considéré comme élevé puisque les calculs sont effectués sur les tableaux de probabilité.

On peut étudier les erreurs absolues ou relatives. Le choix est difficile car dans le premier cas, on néglige les erreurs commises sur les petits termes et dans le second, les erreurs relatives peuvent être très importantes. Nous avons étudié ici les erreurs absolues.

L'A.C.P. que nous avons effectuée porte sur le tableau obtenu par linéarisation des $E_{IJ}(t)$; il comporte donc 3 colonnes et 42x9 lignes

Figure 3 : plan principal 1x2 (ACP du tableau E_{1xJ, K'})



correspondant chacune à un couple (canton, CSP). Les trois variables définies par la liste des erreurs effectuées dans la reconstruction des trois derniers tableaux ont été centrées et réduites sans que cela ait eu d'ailleurs un effet important, leur variance étant voisines. La matrice de corrélations est la suivante:

	AN21	AN32	AN43
AN21	1		
AN32	0.311	1	
AN43	-0.015	0.179	1

Compte tenu du nombre d'observations sur lesquelles ces coefficients sont calculés (378), les valeurs obtenues ne sont pas négligeables. Il existe en particulier une certaine corrélation entre les erreurs commises dans les reconstructions du tableau 2 et du tableau 3 (1962 et 1968). Les erreurs les plus grandes sont représentées par les points les plus à l'extérieur du schéma. Du côté gauche, on trouve les couples suivants:

(D3,OU), (D7,OU), (E1,OU), (H3,OU), (KB,EA), (KB,OA), (O4,EA), (O4,OA), (N7,OU)

Du côté positif de l'axe 2 :

(C3,OU), (F2,OU), (G1,OA), (K4,OA), (K6,CM)

Du côté positif de l'axe 1 :

(O4,OU), (G1,OU), (KB,LM), (N7,EA), (K4,OU), (J9,OU), (D3,AC), (KB,PL), (KB,EM), (CB,OA), (E1,AC), (H6,OU), (M6,EA)

Ces listes montrent que les erreurs les plus grandes en valeur absolue concernent souvent les mêmes CSP et les mêmes cantons: sur les 32 couples de cette liste, la CSP Ouvriers est citée 15 fois ce qui représente largement plus de 1/9, proportion normale en cas d'indépendance des erreurs, et les cantons N7 (Argelès sur mer) 4 fois, KB (Montpellier) 6 fois (1/8 et 1/5 au lieu de 1/42). La corrélation observée entre les erreurs de reconstruction des tableaux de 1962 et 1968 est donc due essentielle-

ment à 1 CSP et 2 cantons, dont les évolutions sont mal représentées par le modèle RAS.

La méthode STATIS appliquée à ces trois tableaux de données aurait vraisemblablement conduit à des résultats analogues.

2.4 Conclusion

Nous considérons que le modèle RAS donne des résultats satisfaisants sur ces données. Il représente bien, dans la plupart des cas, l'évolution des tableaux et l'étude des erreurs permet de déterminer les cantons et les CSP qui obéissent à une évolution d'une autre nature. L'analyse des matrices R et S des coefficients de passage débouche sur la formation de groupes différents de ceux que l'on obtient par l'A.F.C. du tableau moyen, mais il existe un lien évident entre les deux partitions. Cette différence s'explique par le fait que l'A.F.C. consiste à représenter aux mieux les distances entre profils alors que les coefficients r_i et s_j sont indépendants des lois conditionnelles (selon les hypothèses du modèle). Cette différence est particulièrement flagrante lorsque l'on examine le canton Uzès (F9).

Il est intéressant aussi d'examiner l'évolution des coefficients de passage: ils varient en effet de façon régulière et assez nette ou encore restent stables.

On peut regretter que l'analyse ait été effectuée sur les tableaux de probabilités. Cela revient à diviser les coefficients r_i et s_j par les effectifs des tableaux de répartition, éliminant ainsi l'effet de l'évolution générale de la population. Toutefois, l'évolution constatée sur les coefficients calculés contredit complètement l'hypothèse d'un modèle de rang 1.

3.1 CONCLUSION GENERALE

La méthode que nous avons employée dans ce texte pour étudier une suite de tableaux de contingence ne relève pas stricto sensu de l'analyse des données puisqu'elle fait appel à un modèle d'évolution. On peut évidemment envisager d'autres types de modèles suivant le problème étudié: la quasi-symétrie par exemple est plus adaptée pour l'étude de tableaux de migrations dont on veut prévoir une marge à un horizon fixé alors que dans le cas d'une suite de tableaux d'échanges, on cherche plutôt à reconstruire un tableau dont les marges sont connues ou résultent d'hypothèses exogènes. Rappelons également l'existence d'autres méthodes d'ajustement d'un tableau à des marges fixées [1] , [2] .

L'esprit dans lequel la méthode RAS a été appliquée ici est donc différent de celui de l'analyse des données: on détermine des coefficients suivant un modèle que l'on cherche à vérifier à posteriori. Si l'on juge que le modèle est acceptable, on peut effectuer des prévisions sous contraintes de marges et vérifier les hypothèses qui ont conduit aux choix des marges en comparant les coefficients d'ajustement aux précédents.

Les deux approches: modélisation et analyse factorielle, paraissent donc ici plus complémentaires que concurrentes.

BIBLIOGRAPHIE

- [1] Caussinus H. Quelques points de vue sur l'analyse des tableaux d'échanges
Annales de l'INSEE n° 22-23, Avril Septembre 1976, Paris, p 83-99

- [2] Nijkamp P. Paelinck J.H.P. Some methods for updating input-output tables.
Revue belge de Statistique, d'Informatique et de Recherche Opérationnelle. Vol 16 n° 14, 1976 p 17-41.

- [3] Thionet P. Construction et reconstruction des tableaux statistiques. Annales de l'INSEE n° 22-23, Avril Sept 1976, Paris, p 5-27.