



Analyse factorielle en référence à un modèle. Application à l'analyse de tableaux d'échange

Brigitte Escofier

► To cite this version:

Brigitte Escofier. Analyse factorielle en référence à un modèle. Application à l'analyse de tableaux d'échange. [Rapport de recherche] RR-0337, INRIA. 1984. inria-00076220

HAL Id: inria-00076220

<https://inria.hal.science/inria-00076220>

Submitted on 24 May 2006

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



CENTRE DE RENNES
IRISA

Institut National
de Recherche
en Informatique
et en Automatique

Domaine de Voluceau
Rocquencourt
BP 105
78153 Le Chesnay Cedex
France
Tél (3) 954 90 20

Rapports de Recherche

N° 337

**ANALYSE FACTORIELLE
EN RÉFÉRENCE À UN MODÈLE
APPLICATION À L'ANALYSE
DE TABLEAUX D'ÉCHANGES**

Brigitte ESCOFIER

Septembre 1984

Campus Universitaire de Beaulieu
Avenue du Général Leclerc
35042 - RENNES CÉDEX
FRANCE
Tél. : (99) 36.20.00
Télex : UNIRISA 95 0473 F

ANALYSE FACTORIELLE EN REFERENCE
A UN MODELE.
APPLICATION A L'ANALYSE DE TABLEAUX D'ECHANGES.

Brigitte ESCOPIER

Publication Interne n° 229
Juillet 1984
19 pages

RESUME :

On peut voir l'analyse des correspondances d'un tableau K_{IJ} comme l'analyse de son écart avec le tableau formé par le produit de ses deux marges, Ce dernier qui serait identique à K_{IJ} sous l'hypothèse d'indépendance constitue en quelque sorte un modèle.

La plupart des principes et des propriétés de l'analyse des correspondances se généralisent en prenant un modèle quelconque. Deux nuages de points représentent les écarts respectifs des lignes et des colonnes du tableau à celle du modèle. Ces nuages sont projetés sur leurs axes d'inertie et des formules de transition lient leur représentations.

Nous donnons quelques exemples de problèmes qui peuvent être étudiés en choisissant un modèle différent du modèle d'indépendance. Des modèles pour l'analyse de tableaux d'échange (tableaux carrés de diagonale indéterminée) sont étudiés. Quand le modèle cherche à ajuster le tableau de données, c'est la structure du tableau des résidus qui est analysée.

ABSTRACT

FACTOR ANALYSIS WITH A MODEL APPLICATION TO THE ANALYSIS OF EXCHANGE

TABLES

Brigitte ESCOFFIER

Correspondance Analysis of a data table K_{IJ} can be seen as the analysis of the distance between K_{IJ} and the table product of its two marginal distributions. This product would be equal to K_{IJ} should the hypothesis of independance prove to be true and could be considered as a model.

Most of the principles and properties of Correspondance Analysis can be generalized starting from any model. Two clouds of points represent the respective distances between the rows (or the columns) of the table and those of the model. Those clouds are projected on their inertia axes and their representations are linked together by transition formulas.

We give a few exemples of the problems that can be studied if we choose a different model from the independance one. Models for the analysis of exchange tables (a square table with undetermined diagonals) are studied in this article. When the model approaches the data table, the structure of the residual table is then analysed.



I - LA METHODE

I.1. Les données

Soit $K_{IJ} = \{K_{ij} ; i \in I, j \in J\}$ un tableau de nombres positifs de I lignes et J colonnes.

Soit $M_{IJ} = \{M_{ij}, i \in I, j \in J\}$ un tableau modèle de même dimension. Ce modèle peut être : le produit des marges de K_{IJ} , le produit des marges excepté sur la diagonale (modèle étudié pour les matrices d'échanges) ; il peut être construit à partir de variables extérieures par une sorte de régression ; il peut traduire des liaisons binaires si K_{IJ} est un tableau ternaire développé etc ...

Soient $P_I = \{P_i ; i \in I\}$ et $Q_J = \{Q_j ; j \in J\}$ deux tableaux de nombres positifs de dimensions respectives I et J . Ces tableaux peuvent être les marges de K_{IJ} , des combinaisons linéaires des marges de K_{IJ} et M_{IJ} , des valeurs extérieures ...

Notons avec les minuscules correspondantes k_{IJ} , m_{IJ} , p_I et q_J les tableaux obtenus en divisant chacun des tableaux K_{IJ} , M_{IJ} , P_I et Q_J par la somme de leurs termes.

Notons aussi k_I , k_J et m_I , m_J les marges de k_{IJ} et m_{IJ} :

$$k_{i.} = \sum_j k_{ij} \qquad k_{.j} = \sum_i k_{ij}$$

$$m_{i.} = \sum_j m_{ij} \qquad m_{.i} = \sum_j m_{ij}$$

I.2. Les nuages de points

A l'ensemble des lignes I , on associe un nuage de points noté $N(I)$. Le nuage se situe dans l'espace des mesures sur J , noté R_J , l'élément i est la différence entre les deux mesures :

$$\frac{k_{iJ}}{p_i} - \frac{m_{iJ}}{p_i} = \left\{ \frac{k_{ij}}{p_i} - \frac{m_{ij}}{p_i} ; j \in J \right\}$$

Si p_I est égale à la fois à la marge de k_{IJ} et de m_{IJ} , alors k_{ij}/p_i et m_{ij}/p_i sont des mesures de probabilité (ce sont les profils de la ligne i du tableau et du modèle).

L'espace R_J est muni de la métrique du χ^2 de centre q_J . La distance entre deux points i et i' s'écrit donc :

$$D^2(i, i') = \sum_j \left\{ \frac{k_{ij} - m_{ij}}{p_i} - \frac{k_{i'j} - m_{i'j}}{p_{i'}} \right\}^2 \frac{1}{q_j}$$

Deux points i et i' sont proches si les écarts entre la ligne du tableau et celle du modèle, pondérés par p_I se ressemblent pour tout j . Un point i est éloigné de l'origine si la ligne i s'écarte beaucoup de celle du modèle.

Le poids p_i est affecté au point i .

Le nuage $N(J)$ est défini symétriquement.

I.3. Les facteurs

Les facteurs sur I et sur J sont, par définition, les projections des nuages $N(I)$ et $N(J)$ sur leurs axes d'inertie dans l'ordre d'inertie décroissante. Ces axes sont calculés à l'origine qui n'est généralement pas le centre de gravité du nuage.

I.4. Les centres de gravité

Les coordonnées des centres de gravité de $N(I)$ et de $N(J)$ notés w_I et w_J valent respectivement :

$$w_I(j) = \sum_i p_i \frac{(k_{ij} - m_{ij})}{p_i} = k_j - m_j$$

$$w_J(i) = \sum_j p_j \frac{(k_{ij} - m_{ij})}{p_j} = k_i - m_i$$

Si les marges sur J du tableau et du modèle sont égales entre elles, le nuage des lignes est centré. De même, si les marges sur I sont égales, le nuage des colonnes est centré.

Les distances entre les points du nuage ne seront bien représentées par leurs projections sur les premiers axes d'inertie que si le centre de gravité n'est pas trop éloigné de l'origine. Il faut donc que les marges du modèle ne soient pas trop différentes de celles du tableau.

1.5. Calcul des facteurs. Dualité des analyses

Rappelons que les facteurs et les moments d'inerties d'un nuage s'obtiennent en diagonalisant une matrice qui s'écrit $XX'D$ où : X est la matrice des coordonnées des points du nuage, X' sa transposée, D la matrice diagonale des poids des éléments et M la matrice (diagonale ici) de la métrique de l'espace.

Ainsi les facteurs sur J et les moments d'inertie de $N(J)$ sont vecteurs propres et valeurs propres de la matrice de terme général :

$$a_{jj'} = \sum_i \frac{(k_{ij} - m_{ij})(k_{ij'} - m_{ij'})}{p_i q_j}$$

On montre aisément la dualité des analyses des nuages $N(I)$ et $N(J)$. Leurs moments d'inertie sont égaux deux à deux et des formules de "transition" lient les facteurs sur I et sur J , ce qui permet de déduire les uns des autres. Notons F_δ et G_δ les facteurs d'ordre δ sur I et J et λ_δ le moment d'inertie associé, les relations de transition s'écrivent :

$$F_\delta(i) = \frac{1}{\sqrt{\lambda_\delta}} \sum_{j \in J} \left(\frac{k_{ij} - m_{ij}}{p_i} \right) G_\delta(j)$$

$$G_\delta(j) = \frac{1}{\sqrt{\lambda_\delta}} \sum_{i \in I} \left(\frac{k_{ij} - m_{ij}}{q_j} \right) F_\delta(i)$$

Cette dualité permet une représentation simultanée des lignes et des colonnes. Les relations entre les positions des éléments de I et de J peuvent se traduire ainsi : un point i est situé du côté des points j auxquels il s'associe plus dans le tableau que dans le modèle ($k_{ij} > m_{ij}$) et à l'opposé de ceux auxquels il s'associe moins dans le tableau que dans le modèle ($k_{ij} < m_{ij}$) et réciproquement (pour j relativement aux points i).

1.6. Cas particulier de l'analyse des correspondances

Le modèle égal au produit des marges de k_{ij} : $m_{ij} = k_{i.} k_{.j}$ correspond à l'hypothèse d'indépendance. Dans ce cas $m_I = k_I$ et $m_J = k_J$. Si, de plus $p_I = k_I$ et $q_J = k_J$, alors on retrouve exactement l'analyse des correspondances du tableau k_{IJ} . En effet le point i est représenté dans R_J par le point de coordonnée :

$$\left\{ \frac{k_{ij}}{k_{i.}} - k_{.j} ; j \in J \right\}$$

(La ligne de référence du modèle est la marge de k_{IJ} et ne dépend pas de i .) C'est le profil de la ligne i en prenant comme origine le centre de gravité du nuage de ces profils. La métrique sur R_J est la métrique du χ^2 de centre $k_{.j}$ et le poids affecté à i est $k_{i.}$. Le nuage $N(I)$ est donc exactement celui de l'A.F.C.

Dans le cas plus général où on a seulement $m_I = k_I = p_I$ et $m_J = k_J = q_J$, les résultats peuvent être obtenus par l'A.F.C. du tableau de terme général $k_{ij} - m_{ij} + k_{i.} k_{.j}$. Les nuages associés à l'A.F.C. de ce tableau sont, en effet, les nuages que nous avons définis. Ils représentent les différences des profils des lignes ou des colonnes.

1.7. Éléments supplémentaires

Des lignes ou des colonnes peuvent être mises en éléments supplémentaires, i.e., projetées sur les axes d'inertie des nuages sans intervenir dans le calcul de ces axes. Les valeurs des facteurs pour ces éléments sont obtenus par les formules de transition.

En analyse des correspondances les lignes et les colonnes n'apparaissent que par leur profil. L'effectif total du tableau d'intervient pas, les formules écrites avec K_{IJ} et k_{IJ} sont équivalentes. Ce n'est pas le cas ici si P_I (ou P_J) n'est pas égal aux deux marges k_I et m_I (resp. k_J et m_J). Dans le programme l'effectif total des tableaux (par lequel tous les éléments sont divisés) est calculé uniquement sur les éléments principaux.

Si une ligne i_0 est la somme de I_0 lignes, avec $I_0 \subset I$ à la fois dans les tableaux k_{IJ} et m_{IJ} et dans le tableau P_I , elle se projette au centre de gravité des points i de I_0 et son poids est la somme des poids des éléments de I_0 . De même pour les colonnes.

I.8. Aides à l'interprétation

Elles sont analogues à celle de l'analyse des correspondances :

I.8.1. Contribution à l'inertie

La contribution d'un élément i à l'inertie d'un facteur F_δ est égale au quotient

$$CTR(i) = p_i F_\delta^2(i) / \lambda_\delta$$

De même la contribution de j à l'inertie du facteur G_δ est :

$$CTR(j) = q_j G_\delta^2(j) / \lambda_\delta$$

Si $k_J \neq m_J$, le centre de gravité w_I n'est pas à l'origine, sa projection sur l'axe δ est :

$$F_\delta(w_I) = \sum_i p_i F_\delta(i)$$

L'inertie d'un facteur est la somme de deux inerties, celle de w_I et celle autour de w_I . La contribution de w_I est :

$$CTR(w_I) = F_\delta^2(w_I) / \lambda_\delta$$

De même si $k_I \neq m_I$ on calcule :

$$G_{\delta}(w_j) = \sum_j q_j G_{\delta}(j)$$

$$CTR(w_j) = G_{\delta}^2(w_j) / \lambda_{\delta}$$

La projection et la contribution du centre de gravité montrent le décentrage du facteur et son importance dans la détermination de ce facteur.

I.8.2. Qualité de représentation

L'inertie du point i dans R_I est égale au produit de son poids p_i par le carré de sa distance à l'origine :

$$\text{Inertie de } i = p_i \sum_j \left(\frac{k_{ij} - m_{ij}}{p_i} \right)^2 \frac{1}{q_j}$$

Sa qualité de représentation sur le facteur F_{δ} est le quotient de $p_i F_{\delta}^2(i)$ par son inertie, c'est aussi le cosinus carré de l'angle entre l'axe δ et le vecteur Oi . La qualité de représentation de j et des centres de gravité (s'ils sont distincts de l'origine) se définissent de la même façon. Les qualités de représentation s'ajoutent sur les axes.

I.9. Formule de reconstitution des données

Connaissant tous les facteurs F_{δ} et G_{δ} et les inerties λ_{δ} , la différence entre le tableau k_{ij} et le modèle m_{ij} peut être reconstituée exactement :

$$k_{ij} - m_{ij} = p_i q_j \left(\sum_{\delta} F_{\delta}(i) G_{\delta}(j) / \sqrt{\lambda_{\delta}} \right)$$

I.10. Equivalence distributionnelle

Les distances et les facteurs sont inchangés lorsque 2 lignes ou 2 colonnes confondues dans les nuages sont remplacées dans tous les tableaux par leur somme. Pour deux lignes i et i' la condition s'écrit :

$$\frac{k_{ij} - m_{ij}}{p_i} = \frac{k_{i'j} - m_{i'j}}{p_{i'}} \quad \text{pour tout } j.$$

II - APPLICATIONS

II.1. Comparaison de 2 ou plusieurs tableaux

Le moyen le plus efficace de comparer deux tableaux de même dimension n'est pas de faire jouer à l'un le rôle de k_{IJ} et à l'autre celui de M_{IJ} . En effet, K_{IJ} et M_{IJ} ne jouent pas le même rôle, (si on les intervertit, les nuages $N(I)$ et $N(J)$ restent identiques, mais le sens relatif des facteurs F_Δ et G_Δ est inversé, car i est situé du côté des j pour lesquels $k_{ij} > m_{ij}$.) D'autre part, si les marges des deux tableaux sont différentes, on ne compare pas les profils des lignes (ou des colonnes) mais les lignes (pondérées par p_i). Le problème du choix de P_I et q_J est alors délicat.

Pour comparer les profils des lignes I des T tableaux, on analyse le nuage des écarts de ces profils à leur moyenne. Pour cela, on juxtaposera en lignes les 2 (ou T) tableaux et on compare ce tableau à $I \times T$ lignes et J colonnes à un modèle ainsi défini :

Soit k_{ijt} le terme général de la ligne i du tableau t , k_{ij} le terme général de la somme des T tableaux, $k_{i.t}$ la somme de la ligne i du tableau t . Le terme général du modèle s'écrit $k_{ij} \cdot k_{i.t} / k_{i.}$. Ce modèle a les mêmes marges $k_{i.t}$ et $k_{.j}$ que le tableau. On pose $P_{it} = k_{i.t} / k_{i.}$ et $q_j = k_{.j} / k_{i.}$. A chaque ligne (i,t) est associée, alors, la différence $k_{ijt} / k_{i.t} - k_{ij} / k_{i.}$ entre le profil de la ligne i du tableau t et la moyenne des profils des lignes i de tous les tableaux (ou profil de la ligne i de leur somme).

Pour comparer les profils des colonnes, on procède de la même façon en juxtaposant les tableaux en colonne. Cette méthode est détaillée et utilisée dans (2) et (3).

II.2. Les tableaux d'échange ou de flux

II.2.1. Les données

Les tableaux d'échange sont des tableaux carrés, non symétriques dont la diagonale est soit non définie, soit de nature différente du reste du tableau. Les tableaux de trafic interzones, d'échanges inter-industriels, d'import-export, sont des tableaux de ce type.

L'analyse des correspondances de ces tableaux est possible à condition de remplir la diagonale. Mais les résultats sont le plus souvent décevants. Une diagonale nulle éloigne deux zones ayant des échanges mutuels importants. Une diagonale représentant les échanges intérieurs peut quelquefois être définie, mais les effectifs sont alors souvent importants et de forts éléments diagonaux mettent à part des zones dont chacune s'associe avec elle-même. Burtchy (1) propose de compléter la diagonale par des méthodes de reconstitution de données manquantes, ce qui minimise dans l'analyse l'influence de la diagonale. Les résultats sont acceptables, mais la méthode est lourde et la reconstitution est artificielle. Il propose aussi des "analyses du triple" avec des distances spécifiques où les échanges mutuels ont un rôle particulier. Mais les résultats sont moins riches, car il n'y a pas de représentation simultanée.

II.2.2. Modèle 1

II.2.2.1. Le modèle

En analyse des correspondances, quelle que soit la valeur k_{ii} affectée à la diagonale, le terme $k_{i.} k_{.i}$ apparaît dans le produit des marges. La diagonale du tableau d'échange étant indéterminée, son écart avec celle du modèle ne peut que jouer un rôle perturbateur. Pour éliminer l'influence de cette diagonale, nous mettons la même valeur, 0, sur le tableau k_{IJ} et sur le modèle m_{IJ} . Pour les éléments non diagonaux, nous gardons le modèle d'indépendance.

$$m_{ij} = k_{i.} k_{.j} \quad \text{si } i \neq j$$

$$m_{ii} = k_{ii} = 0$$

Les marges de m_{IJ} sont différentes de celles de k_{IJ} et valent :

$$m_{i.} = \sum_{j \neq i} k_{i.} k_{.j} = k_{i.} (1 - k_{.i})$$

$$m_{.j} = k_{.j} (1 - k_{j.})$$

On pose :

$$p_i = k_{i.}$$

$$q_j = k_{.j}$$

II.2.2.2. Les nuages de points

La ligne i est représentée par la différence entre deux mesures. La première, le profil de la ligne i de k_{IJ} , est une mesure de probabilité ; la seconde la marge $k_{.j}$ - sauf pour $j=i$ où l'on a zéro - ne l'est pas :

$$i \rightsquigarrow \begin{cases} \frac{k_{ij}}{k_{i.}} - k_{.j} & \text{si } j \neq i \\ 0 & \text{si } j = i \end{cases}$$

Sauf pour $j=i$ où l'on a zéro, on a, comme en analyse des correspondances, la différence entre le pourcentage de trafic de i allant vers j et le pourcentage du trafic global allant vers j . Le nuage $N(I)$ n'est pas centré, les coordonnées de son centre de gravité valent :

$$w_I(j) = k_{.j} k_{j.}$$

Son inertie qui vaut $\sum_j (k_{j.})^2 k_{.j}$ est toujours assez faible par rapport à l'inertie totale du nuage et influe peu sur les facteurs dans tous les exemples traités.

II.2.2.3. Distance

La distance entre deux points i et i' s'écrit :

$$D^2(i, i') = \sum_{j \neq i, i'} \left(\frac{k_{ij}}{k_{i.}} - \frac{k_{i'j}}{k_{i'.}} \right)^2 \frac{1}{k_{.j}} + \left(\frac{k_{ii'}}{k_{i.}} - k_{.i'} \right)^2 \frac{1}{k_{.i'}} + \left(\frac{k_{i'i}}{k_{i'.}} - k_{.i} \right)^2 \frac{1}{k_{.i'}}$$

Pour une matrice d'échange dont les lignes représentent les origines, et les colonnes, les arrivées, le premier terme concerne les trafics vers les zones (ou pays) tiers. Ce terme est nul si le pourcentage des trafics d'origine i et i' vers chacune de ces zones sont égaux.

Les deux derniers termes concernent les échanges mutuels. La distance augmente si le pourcentage du trafic de i qui est dirigé vers i' , s'écarte de celui qui est calculé sur l'ensemble des zones et réciproquement.

II.2.2.4. Représentation simultanée

L'interprétation de la représentation simultanée est tout à fait analogue à celle de l'analyse des correspondances : une zone origine i est proche des zones destinations $j \neq i$ vers lesquelles le pourcentage de son trafic est supérieur au pourcentage global de trafic vers j (et réciproquement). Les formules de transition s'écrivent :

$$F_{\Delta}(i) = \sum_{j \neq i} \frac{1}{\sqrt{\lambda_{\Delta}}} \left(\frac{k_{ij}}{k_{i.}} - k_{.j} \right) G_{\Delta}(j)$$

$$G_{\Delta}(j) = \sum_{i \neq j} \frac{1}{\sqrt{\lambda_{\Delta}}} \left(\frac{k_{ij}}{k_{.j}} - k_{i.} \right) F_{\Delta}(i)$$

II.2.3. Modèle 2

Le modèle 2 est construit pour que la ligne i soit représentée par la différence entre deux mesures de probabilité. Il est défini par :

$$\begin{cases} m_{ij} = \frac{k_{i.} \cdot k_{.j}}{1 - k_{.i}} & \text{si } i \neq j \\ m_{ii} = 0 \end{cases}$$

On pose $p_I = k_I$ et $q_J = k_J$. La marge sur I de m_{IJ} est égale à k_J mais sa marge sur J n'est pas égale à k_J .

Un point i est représenté par la différence entre la répartition de son trafic vers les pays tiers et la répartition du trafic global-calculé après avoir supprimé la destination i . Celle-ci, qui n'existe pas dans les trafics d'origine i intervenait dans les calculs de pourcentage du premier modèle. Dans ce second modèle, un point i peut être situé à l'origine, ce qui n'était pas possible dans le premier modèle.

Les rôles de I et de J ne sont pas symétriques, la représentation de J est essentiellement une aide à l'intégration de celle de I . (Le nuage $N(J)$ est centré). Les rôles de I et de J peuvent être inversés.

II.2.4. Exemple

Nous avons traité une matrice de trafic domicile-travail entre 16 zones de la banlieue Sud de Paris. La matrice initiale extraite de (5) comportait des éléments sur la diagonale concernant les déplacements intra-zones. Nous ne nous intéressons qu'aux trafics inter-zones et nous caractérisons chaque zone par ses déplacements de - ou vers - les autres zones, ce qui revient à annuler la diagonale. Les modèles introduits dans le paragraphe précédent donnent des résultats très proches, nous ne détaillons que le premier.

Nous comparons nos résultats à ceux de l'analyse des correspondances de la matrice initiale à diagonale chargée et à ceux de la même matrice à diagonale nulle.

II.2.4.1. Centres de gravité

Les centres de gravité des deux nuages sont situés pratiquement à l'origine et leur contribution à l'inertie des 5 premiers facteurs ne dépasse jamais trois pour mille. L'influence du décentrage est donc négligeable.

II.2.4.2. Plan des deux premiers facteurs

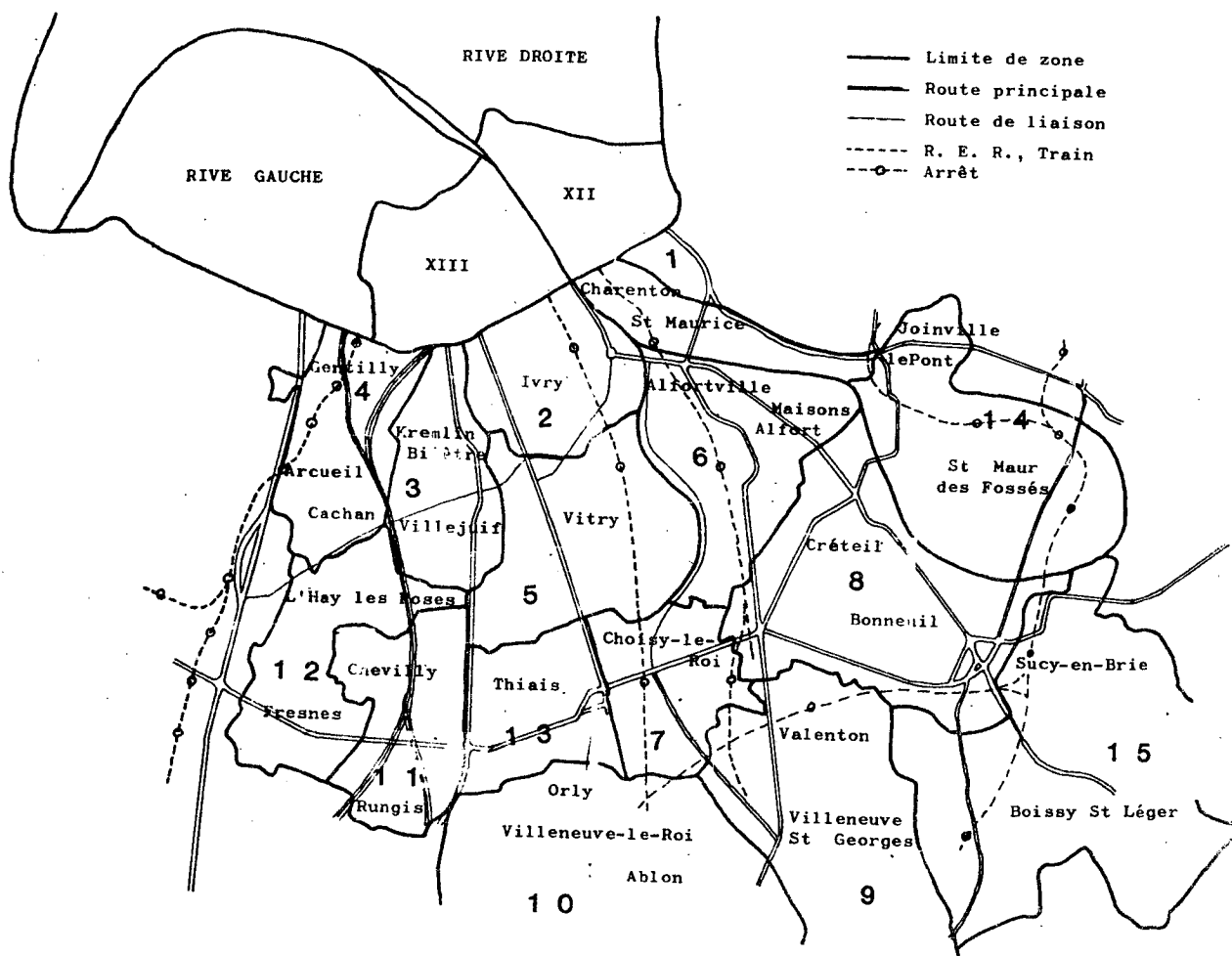
Les pourcentages d'inertie extraits par les deux premiers facteurs sont de 35% et de 17%, le premier plan représente donc plus de la moitié de l'inertie.

A une extrémité du premier facteur on trouve 5 zones situées toutes à l'est de la région étudiée. Entre ces 5 zones (1, 6, 8, 14, 15) traversées par de grands axes de circulation, les échanges sont importants. En tant qu'origine des trafics, ils sont plus groupés qu'en tant qu'arrivée. La zone d'arrivée 15 en particulier s'éloigne un peu de ce groupe. C'est la plus éloignée de Paris et le trafic vers elle ne vient pas seulement de l'axe vers Paris mais aussi de la zone limitrophe 9 qui n'appartient pas au même groupe.

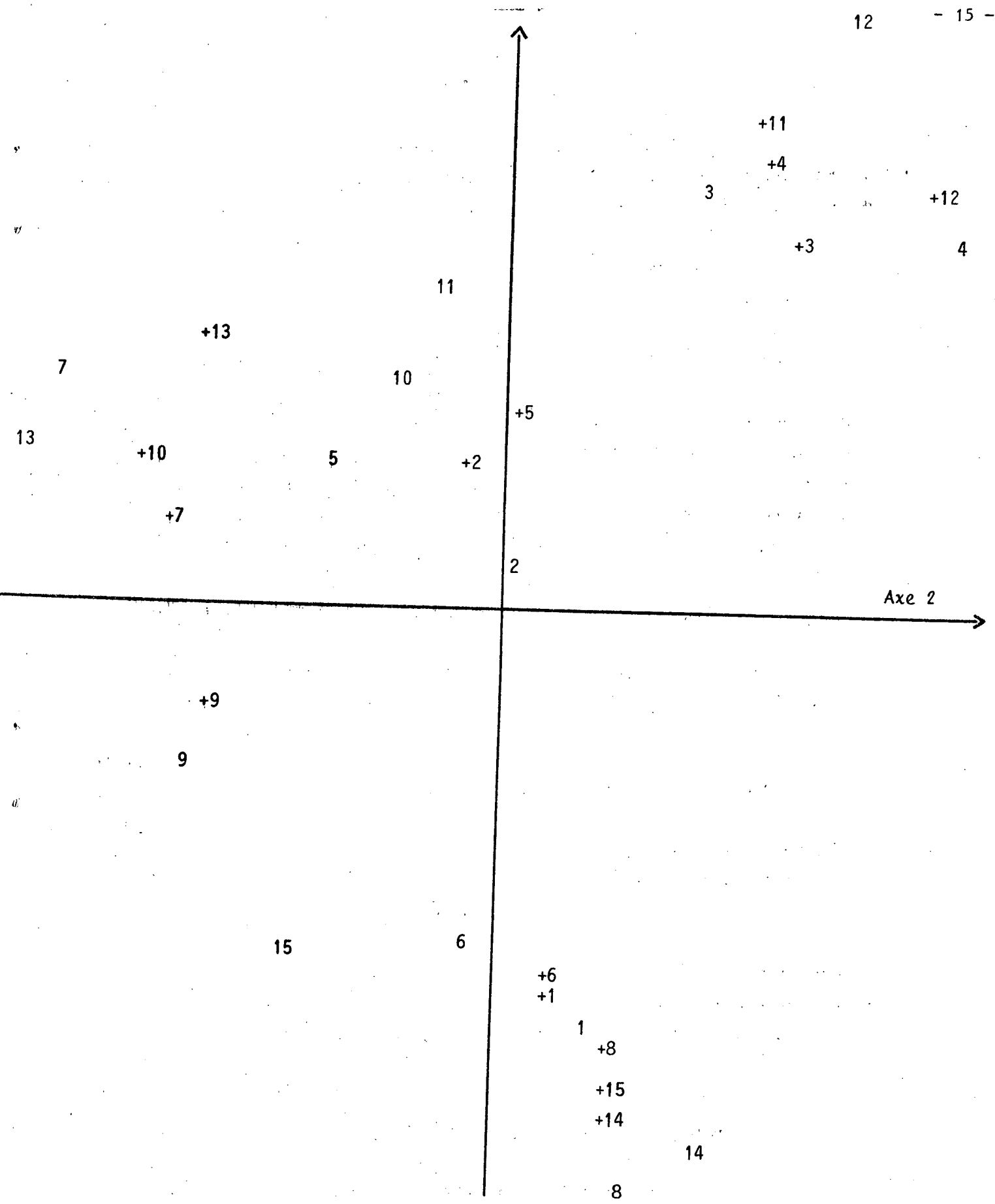
A l'opposé, la région Ouest avec laquelle la région Est entretient peu d'échanges. Elle comprend 4 zones (3, 4, 12, 11). Le phénomène noté pour la zone 15 se produit de manière encore plus marquée pour la zone 11 : elle s'éloigne beaucoup du groupe en tant qu'arrivée car les déplacements vers cette zone éloignée de Paris proviennent beaucoup de la zone limitrophe centrale 10.

Enfin, au centre s'étalant approximativement de l'Est à l'Ouest, apparaissent les zones de la région centrale. Notons la position proche de l'origine de la zone 2 très proche de Paris et centrale qui attire le trafic de presque toute les autres zones.

Les zones en tant qu'origine et en tant qu'arrivée sont généralement assez proches. Nous avons noté deux exceptions les zones 15 et 11 auxquelles s'ajoute la zone 10 située elle aussi à la périphérie. Pour ces trois zones, il y a une forte asymétrie du trafic origine et destination.



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
INSEE	301	302	303	304	310	311	312	313	401	402	403	404	493	506	608	
DE	Charenton	Ivry	Kremlin	Gentilly	Vitry	Alfort	Choisy	Bonneuil	Valenton	Orly	Rungis	Fresnes	Thiais	Joinville	Sucy	TOTAL
VERS																
1	6238	269	45	14	204	1824	57	250	70	76	16	36	0	403	189	9691
2	270	11268	1113	257	2483	1450	530	708	166	878	166	205	281	457	174	20406
3	34	585	11353	1001	1493	32	143	62	133	207	327	549	226	133	0	16278
4	0	106	1389	10695	425	100	99	220	27	111	215	1037	26	152	117	14719
5	186	667	894	281	11263	1009	1577	148	123	1021	154	265	860	314	90	18852
6	713	258	134	75	632	16420	595	1675	563	250	29	0	118	507	297	22266
7	0	181	78	41	763	148	5590	24	396	964	104	38	745	25	87	9184
8	51	81	68	0	133	1094	109	9235	107	92	0	28	39	1831	491	13359
9	31	34	34	28	34	316	271	148	6161	628	0	0	59	83	228	8055
10	14	108	492	177	353	104	528	209	568	6461	315	408	551	191	130	10609
11	0	21	160	83	81	33	23	20	64	248	1455	110	106	21	0	2425
12	0	53	310	260	156	0	0	0	0	82	481	3889	131	0	0	5362
13	0	66	21	0	151	40	421	24	43	248	26	0	1498	25	0	2563
14	327	43	0	63	206	801	42	1362	0	40	54	90	35	17045	774	20882
15	0	0	0	26	26	20	28	159	591	102	0	0	0	403	5624	6979
	7864	13740	16091	13001	18403	23391	10013	14244	9012	11408	3342	6655	4675	21590	8201	181630



Plan des deux premiers facteurs { Origine +1
Destination 1

II.2.4.3. Facteur 3 et Facteur 4

Le troisième facteur est déterminé par les trafics des régions 9 et 14 vers les régions 15 et 8. Les deux premières représentent 60% de l'inertie du nuage des lignes et les deux dernières 57% de l'inertie du nuage des colonnes.

Le quatrième facteur qui extrait encore 11% de l'inertie est un facteur inverse en ce sens que des régions origines se trouvent presque à l'opposé de leur image en tant qu'arrivée. Il est déterminé essentiellement par 4 zones de la région Est regroupées sur les deux premiers facteurs. Il traduit un phénomène interne à cette région où les deux zones extrêmes 1 et 8 ont des échanges très importants avec les deux zones médianes 6 et 14 alors que les échanges 1-8 et 6-14 sont faibles. A une extrémité du facteur on trouve les zones origines 1 et 8 associées aux zones destinations 6 et 14 et à l'autre extrémité leurs opposés.

II.2.4.4. Comparaison avec l'A.F.C. de la matrice initiale

Seul le plan des deux premiers facteurs ressemble un peu aux résultats précédents. Les facteurs suivants sont totalement différents.

Sur le premier plan, une première différence très notable est le rapprochement des deux images origine et destination de toutes les zones. Les formules de transition avec une matrice à diagonale très chargée expliquent ce phénomène. L'asymétrie très nette de cette matrice de trafic n'apparaît pas du tout dans les résultats.

Les 16 zones sont réparties sur une parabole approximativement d'Est en Ouest. Les 2 zones 4 et 14 à très fort trafic interne sont chassées aux deux extrémités de la parabole et les 3 régions Est-Ouest et Centre visibles sur le graphique précédent ne sont pas réellement séparées ici.

II.2.4.5. Comparaison avec l'A.F.C. de la matrice à diagonale nulle

Le deuxième facteur est un facteur "inverse" analogue au 4ème de l'analyse avec le modèle à diagonale nulle. On sait que dans l'analyse d'un tableau symétrique les facteurs inverses risquent d'apparaître très tôt lorsque

la diagonale est trop faible. Bien que le tableau ne soit pas symétrique, cet effet apparaît ici. Pour voir les grandes tendances traduites dans le premier plan des deux autres analyses, il faut regarder le plan 1-3. Les 2 images d'une même zone qui étaient rapprochées par la diagonale chargée sont systématiquement éloignées par la diagonale nulle. A ce phénomène s'ajoute l'augmentation de la distance entre deux zones dont les échanges mutuels sont importants, ce qui rend les résultats beaucoup plus délicats à interpréter que dans l'analyse avec modèle.

II.2.4.6. Les autres modèles

Les deux autres modèles recentrant respectivement $N(I)$ et $N(J)$ donnent des résultats quasiment identiques au premier. Ils ont moins d'intérêt car plus complexes et non symétriques en I et J . Nous les avons introduits craignant un effet perturbateur du décentrage qui n'apparaît jamais dans les exemples que nous avons traités.

II.2.4.7. Conclusion

Les résultats obtenus sont très satisfaisants et leur interprétation est très claire. Dans les deux analyses factorielles la diagonale apparaît comme un élément perturbateur. Quand elle est trop chargée, elle symétrise les résultats et a une influence prépondérante dans leur structure. Quand elle est nulle, elle fait apparaître très vite des facteurs inverses et l'interprétation est délicate.

La méthode est beaucoup moins lourde que la reconstitution de la diagonale proposé par Burtschy. La complexité du programme est équivalente à celle de l'A.F.C. classique dont elle dérive.

II.3. Autres applications

Un autre modèle introduit dans une optique différente a été utilisé et s'est montré efficace (4). Il consiste à remplacer dans la construction du produit des marges, l'une d'elles par des valeurs données. Ceci permet en particulier de traiter des sous-tableaux avec la marge du tableau entier et de supprimer certains éléments sans modifier la métrique. Les éléments supprimés peuvent être : des réponses manquantes dans un questionnaire, des éléments de grande inertie etc ...

Pour les tableaux ternaires, on peut construire différents modèles basés sur des liaisons binaires. Cette analyse permet de comparer différents modèles, l'inertie totale des nuages mesure l'écart du tableau et modèle et les facteurs permettent d'étudier sa structure.

D'autres applications avec un modèle dérivé de variables extérieures par des méthodes de régression sont à l'étude.

B I B L I O G R A P H I E

- (1) BURTSCHY B.,
Analyse des Matrices d'Echanges. Thèse de 3ème cycle.
Université Paris 6, 1981.
- (2) DROUET D., ESCOFIER B.,
Comparaison de Plusieurs Tableaux de Fréquence.
Cahiers de l'Analyse des Données, 1983, n° 4.
- (3) DROUET D.,
Comparaison de Tableaux de Fréquence. Application à l'Etude de l'Evolution des Causes de Décès en Bretagne.
Thèse de 3ème cycle, Université Paris 6, 1983.
- (4) ESCOFIER B.,
Analyse des Correspondances avec Marge Modifiée.
Publication ISUP à paraître. Rapport IRISA n° 146.
- (5) KAZMIERCZAK J.B.
Migrations Interurbaines dans la Banlieue Sud de Paris.
Cahiers de l'Analyse des Données, 1978, n° 2.

Imprimé en France

par

l'Institut National de Recherche en Informatique et en Automatique

