



**HAL**  
open science

## Analyse factorielle lissée et analyse factorielle des différences locales

Brigitte Escofier, Habib Benali

► **To cite this version:**

Brigitte Escofier, Habib Benali. Analyse factorielle lissée et analyse factorielle des différences locales. [Rapport de recherche] RR-0964, INRIA. 1989. inria-00075595

**HAL Id: inria-00075595**

**<https://inria.hal.science/inria-00075595>**

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.

# INRIA

UNITÉ DE RECHERCHE  
INRIA-RENNES

Institut National  
de Recherche  
en Informatique  
et en Automatique

Domaine de Voluceau  
Rocquencourt  
B.P.105  
78153 Le Chesnay Cedex  
France  
Tél. (1) 39 63 55 11

## Rapports de Recherche

N° 964

*Programme 5*

### **ANALYSE FACTORIELLE LISSÉE ET ANALYSE FACTORIELLE DES DIFFÉRENCES LOCALES**

**Brigitte ESCOFIER, Habib BENALI**

**Janvier 1989**



### ANALYSE FACTORIELLE LISSEE ET ANALYSE FACTORIELLE DES DIFFERENCES LOCALES

Brigitte ESCOFIER et Habib BENALI

#### Résumé

Nous proposons ici deux méthodes qui permettent de tenir compte, dans une analyse en composantes principales, d'une structure de proximité ou de contiguïté définie sur l'ensemble des lignes d'un tableau de données. La première, l'analyse lissée permet d'analyser les tendances générales des données en éliminant les fluctuations locales. La deuxième, l'analyse des différences locales permet, au contraire, d'analyser ces fluctuations en éliminant les variations générales liées aux positions spatiales. Nous comparons cette dernière à l'analyse locale (cf. Lebart 1984) et à l'analyse en composantes principales pour des unités statistiques corrélées (cf. Caussinus 1980). Des exemples d'applications illustrent ces techniques.

#### SMOOTH FACTORIAL ANALYSIS AND FACTORIAL ANALYSIS OF LOCAL DIFFERENCES

#### Summary

In this paper, two methods are proposed which allow to take into account a proximity or contiguity structure defined on the set of the data matrix rows in a principal component analysis. With the first method, called smooth factorial analysis, one can study general trends in the data by removing local fluctuations. On the contrary, the second method, called factorial analysis of local differences, is concerned with the analysis of these fluctuations. We compare the latter to local analysis (Lebart 1984) and to Principal Component Analysis for correlated statistical units (Caussinus 1980).

**ANALYSE FACTORIELLE LISSEE  
ET ANALYSE FACTORIELLE  
DES DIFFERENCES LOCALES**

**Brigitte ESCOPIER**

**Habib BENALI**

**Publication Interne n° 444**

**Décembre 1988**



**INIA**

## 1. LA PROXIMITE ET LE GRAPHE DE CONTIGUITE

### 1.1 Quelques exemples de proximité ou de contiguïté

Donnons d'abord quelques exemples courants qui nous permettront de concrétiser les problèmes et de dégager des objectifs.

Le terme de contiguïté évoque d'abord une proximité spatiale que l'on rencontre dès que les lignes du tableau représentent, soit des zones géographiques, soit des zones d'une image, soit encore des régions d'un espace euclidien de dimension quelconque défini par exemple par un ensemble de variables.

Les séries chronologiques multidimensionnelles induisent aussi des structures de contiguïté : chaque moment peut être relié à son (ou ses) suivant(s). Ce type de structure se retrouve dès qu'il existe un ordre sur l'ensemble des lignes d'un tableau, ou sur des sous-ensembles de ces lignes, notamment lorsque ces lignes représentent les points de discrétisation d'une courbe.

Un cas limite, particulièrement simple, nous guidera dans la prise en compte en analyse factorielle de cette notion de contiguïté. C'est celui induit par une partition de l'ensemble des lignes, où les voisins d'un élément sont les éléments qui appartiennent à la même classe que lui. Nous étudions d'abord ce cas limite et nous le généralisons à un graphe quelconque.

### 1.2 Graphe de contiguïté

Nous définissons la contiguïté sur un ensemble  $I$  par un graphe dont les sommets sont les éléments de  $I$  et les arêtes relient un sommet  $i$  à ses voisins. L'ensemble des voisins de  $i$ , qui est contenu dans  $I$  est noté  $V(i)$ . On note  $A$  la matrice carrée, de dimension égale au cardinal de  $I$ , associée au graphe. Son terme général,  $a_{ij}$  (positif ou nul), est le "poids" de l'élément  $j$  dans le voisinage de  $i$ . Si  $j$  n'est pas voisin de  $i$ , ce poids est nul. Si on veut faire jouer le même rôle à tous les voisins de  $i$ , le poids de chacun d'entre eux vaut 1, le graphe est alors un graphe non pondéré. Le graphe n'est pas nécessairement symétrique.

Cette définition de la contiguïté par un graphe est assez générale pour recouvrir la plupart des situations auxquelles on peut être confronté. On note  $n_i = \sum_j a_{ij}$  la somme des poids des voisins de  $i$ ,  $n_i$  n'est autre que le nombre de ces voisins si le graphe n'est pas pondéré et on note  $T$  la matrice diagonale de dimension  $n$  dont le coefficient  $t_{ij}$  vaut  $1/n_i$ .

## 2 LES DONNEES

Nous traitons ici le cas de tableaux de données traités habituellement par l'analyse en composantes principales, c'est-à-dire des tableaux croisant des individus et des variables numériques. Nous exposerons ultérieurement la généralisation des méthodes proposées à d'autres types de tableaux.

On note  $I$  l'ensemble des individus et  $n$  leur nombre. On note  $J$  l'ensemble des variables et  $p$  leur nombre. On note  $X$  le tableau de données, son terme général  $x_{ij}$  représente la valeur de la variable  $j$  pour l'individu  $i$  et  $x_j$  représente la variable  $j$ . La plupart du temps, les variables sont centrées et réduites, pour simplifier les notations,  $X$  désigne la matrice des données ainsi transformées.

### 3. CAS D'UNE PARTITION, ANALYSE INTER ET ANALYSE INTRA

Lorsque la structure de contiguïté est définie par une partition de l'ensemble de  $I$ , la matrice du graphe est composée de blocs de 1 (au croisement d'une classe avec elle-même) et de blocs de 0 (un croisement d'une classe avec une autre classe).

Dans ce cas, tenir compte de cette structure de  $I$  revient à étudier le comportement des différentes classes pour l'ensemble des variables.

Deux aspects complémentaires de cette étude sont :

- l'analyse du comportement moyen de chaque classe
- l'analyse des différences de comportement à l'intérieur d'une même classe.

La matrice de covariance (de corrélation si les variables sont réduites) se décompose en variance inter-classe (variance des moyennes des classes) et variance intra-classe (variance de chaque classe autour de sa moyenne). Etudier le comportement moyen de classes, c'est analyser la variance inter, tandis que l'étude du comportement à l'intérieur des classes est l'analyse de la variance intra. Nous appelons respectivement "analyse inter" et "analyse intra" ces deux analyses.

#### 3.1. Analyse inter

##### 3.1.1. Nuage des individus

A une partition de  $I$  en  $K$  classes notées  $I_k$  correspond une décomposition de nuage d'individus en  $K$  sous-nuages disjoints.

Le barycentre d'une classe  $I_k$ , dont l'effectif est noté  $n_k$  a pour coordonnées :

$$y_{kj} = \sum_{i \in I_k} (x_{ij} / n_k) = \sum_{i \in V(i)} (x_{ij} / n_k) \quad \text{pour tout } i \text{ de } I_k$$

Analyser la variance inter revient à placer chaque individu au barycentre de sa classe, donc à faire l'Analyse en Composantes Principales (A.C.P.) de ces classes en leur affectant le poids  $n_k$  ; ou bien encore à faire l'A.C.P. d'un tableau  $Y$  de même dimension que  $X$  dans lequel  $x_{ij}$  est remplacé par la valeur moyenne des voisins de  $i$ .

### 3.1.2. Nuage des variables

Les colonnes du tableau Y définissent de nouvelles variables. Etudions la transformation du nuage des variables lorsque l'on passe du tableau X au tableau Y.

Nous avons considéré que les variables initiales étaient centrées ; cette transformation conserve le centrage, par contre les nouvelles variables ne sont pas réduites (dans le cas extrême où une variable initiale a la même valeur moyenne pour chaque classe, la nouvelle variable a une variance nulle et doit être supprimée du tableau).

A la décomposition de l'ensemble I en K classes correspond une décomposition de l'espace  $R^n$  en K sous espaces orthogonaux notés  $R^{n_k}$  :

$$I = \bigcup_k (I_k) \quad R^n = \bigcup_k (R^{n_k})$$

La variable indicatrice de la classe  $I_k$  qui vaut 1 sur les individus de cette classe et 0 sur les autres individus appartient à  $R^{n_k}$  (c'est la "première bissectrice" de  $R^{n_k}$ ). Les nouvelles variables  $y_j$  sont les projections des variables  $x_j$ , sur le sous-espace E engendré par les variables indicatrices des q classes. En effet, en notant  $1_k$  la variable indicatrice de la classe k, et  $x_j$  le vecteur représentant la variable j, on a :

$$\langle x_j, 1_k \rangle = (1/n) \sum_{i \in I_k} x_{ij} \quad \text{et} \quad \|1_k\|^2 = n_k/n$$

D'où :

$$y_j = \sum_k \langle x_j, 1_k \rangle (1_k / \|1_k\|^2)$$

## 3.2. Analyse intra

Cette technique a été utilisée notamment par Carlier (1985) et Le FOLL (1972).

### 3.2.1. Nuage des individus

Si l'on s'intéresse à l'analyse des différences de comportement entre les individus d'une même classe, il faut supprimer du nuage d'individus associés à X la dispersion inter-classe et ne conserver que la dispersion intra-classe. Pour cela, on translate chaque sous-nuage  $N(I_k)$  pour que son centre de gravité  $g_k$  coïncide avec l'origine des axes. Un individu i de la classe  $I_k$  est représenté par la différence avec la moyenne de sa classe, point dont les coordonnées sont :

$$z_{ij} = x_{ij} - (1/n) \sum_{i' \in I_k} x_{i'j} = x_{ij} - y_{ij}$$

Dans ce nuage deux individus sont proches s'ils s'éloignent de la même façon de la moyenne de leur classe respective.

### 3.2.2. Nuage des variables

L'ensemble des coordonnées des points de ce nouveau nuage regroupées dans un tableau noté  $Z$  définit de nouvelles variables.

Cette transformation consiste tout simplement à centrer les variables sur chaque classe de la partition. Comme pour l'analyse inter, les variables définies dans le tableau  $Z$  sont centrées mais non réduites (si une variable initiale est constante sur chaque classe, la nouvelle variable a une variance nulle et doit être supprimée des données).

Puisque  $Z = X - Y$  et que la transformation qui permet de passer des variables de  $X$  à celles de  $Y$  est une projection orthogonale sur  $E$  (i.e. le sous-espace engendré par la première bissectrice de chacun des  $R^{nk}$ ) la transformation qui permet de passer des variables de  $X$  à celles de  $Z$  est une projection orthogonale sur l'orthogonal de  $E$ . On retrouve ici une généralisation de la propriété classique de l'A.C.P. : le centrage des variables se traduit dans le nuage d'individus par un recentrage du nuage et dans le nuage des variables par une projection sur le sous-espace orthogonal à la première bissectrice.

### 3.3. Analyse des covariances ou des corrélations

Pour l'analyse inter, comme pour l'analyse intra, une alternative se pose : analyser les variables définies respectivement dans les tableaux  $Y$  et  $Z$  sans les réduire ou en les réduisant. Les deux solutions sont possible et se justifient suivant les cas. Les analyses exactes des variances inter et intra correspondent à la première solution. Dans ce cas les résultats des analyses de  $X$ ,  $Y$ ,  $Z$  sont directement comparables, notamment les valeurs propres, puisque l'inertie des nuages définis par  $X$  est la somme des inerties des nuages définis par  $Y$  et  $Z$ .

Si les variables sont réduites, l'influence de chacune des variables dans l'analyse est équilibré, ce qui est souvent intéressant, mais les analyses de  $X$ ,  $Y$ ,  $Z$  concernent des objets différents. Par exemple, une variable dont les valeurs sont très proches sur les barycentres des classes (resp. à l'intérieur d'une même classe) aura une importante artificiellement grande dans l'analyse réduite de  $Y$  (resp.  $Z$ ).

## 4. GENERALISATION A UN GRAPHE QUELCONQUE

Dans le cas d'un graphe quelconque définissant une structure de proximité sur les éléments de  $I$ , on peut souhaiter éliminer les variations locales pour analyser plus facilement les tendances générales. Dans ce cas, nous proposons l'analyse lissée qui est une généralisation de l'analyse inter. A l'opposé, on peut souhaiter analyser uniquement les variations locales. Nous proposons alors l'analyse des différences locales qui généralise l'analyse intra.



#### 4.1. Analyse factorielle lissée (A.F.L.)

Le terme de lissage est emprunté aux séries chronologiques. Le but principal poursuivi en analyse factorielle lissée est de dégager les grandes tendances, en éliminant l'influence des fluctuations locales. Son principe est très simple, il consiste à remplacer chaque point  $i$  représentant une ligne du tableau par le centre de gravité (pondéré par les  $a_{ij}$ ) de ses voisins. Il est clair que ce lissage diminue l'influence des différences locales, il peut servir à améliorer la fiabilité dans deux circonstances au moins : lorsque les individus ne sont pas définis a priori mais correspondent à un découpage arbitraire (ce qui peut se produire pour des zones géographique par exemple) ; lorsque les mesures sont entachées d'erreurs et que ces mesures ont des valeurs proches pour des individus voisins au sens du graphe. Cette analyse peut aussi être utilisée avant une étape de classification sur facteurs pour "forcer" la classification et construire des classes plus contiguës que celles auxquelles aboutirait une analyse classique.

##### 4.1.1 Nuage des individus

Chaque point  $i$  du nuage  $N(l)$  défini dans l'A.C.P. de  $X$  est remplacé par le barycentre de ses voisins. On étudie alors la dispersion entre ces nouveaux points. Les coordonnées du point représentant l'individu  $i$  s'écrivent :

$$y_{ij} = \sum_{i' \in l} (a_{i'i} x_{i'j} / n_i)$$

Dans le cas d'un graphe non pondéré, cette expression se réduit à :

$$y_{ij} = \sum_{i' \in V(i)} (x_{i'j} / n_i)$$

L'ensemble de ces coordonnées définit un tableau  $Y$  qui s'écrit matriciellement :

$$Y = ATX$$

Généralement le nuage d'individus n'est pas centré. En effet, la moyenne de la colonne  $j$  s'écrit :

$$(1/n) \sum_{i \in l} y_{ij} = (1/n) \sum_{i,i} a_{i'i} (x_{i'j} / n_i) = (1/n) \sum_{i'} (\sum_i (a_{i'i} / n_i)) x_{i'j}$$

Cette moyenne est nulle (comme celle de la variable  $x_j$ ) lorsque le coefficient affecté à  $x_{i'j}$  est indépendant de  $i'$  ce qui se produit notamment dans le cas d'un graphe de partition et dans le cas d'un graphe où  $n_i$  est constant. Dans les autres cas, la moyenne de  $y_j$  peut être différente de 0. Dans beaucoup d'applications pratiques, le nombre et les poids des voisins des éléments varient peu d'un élément à l'autre et cette moyenne est peu différente de 0.

Pour avoir une bonne représentation de ce nuage d'individus, il faut centrer le nuage et donc appliquer une A.C.P. centrée au tableau  $Y$ .

Dans le cas d'un graphe de partition, on retrouve exactement l'analyse inter. Si le graphe est un graphe complet, tous les points du nuage sont confondus avec le barycentre tandis que dans le cas d'un graphe réduit à l'identité, l'analyse lissée se confond avec l'A.C.P. de  $X$ .

#### 4.1.2. Nuage des variables

Les colonnes du tableau Y définissent de nouvelles variables, que nous appelons "variables lissées". Dans le cas de l'analyse inter, à la transformation du nuage d'individus (recentrage des sous-nuages) correspond une transformation duale du nuage des variables (projection sur un sous-espace). En dehors de ce cas, les variables lissées ne sont pas des projections orthogonales sur un sous-espace. Cependant, dans le cas où le graphe de contiguïté n'est pas pondéré, les variables lissées se déduisent des variables initiales par une transformation non linéaire appelée "projection sur un graphe". Notons  $1_i$ , la variable indicatrice des voisins de l'élément  $i$ , alors :

$$\langle x_j, 1_i \rangle = (1/n) \sum_{i \in V(i)} (x_{ij})$$

$$\| 1_i \|^2 = n_i / n$$

D'où :

$$y_{ij} = \langle x_j, 1_i \rangle / \| 1_i \|^2$$

#### 4.1.3. Projection des individus et des variables du tableau initial

Comme dans toute A.C.P., il est possible d'introduire des individus et des variables en éléments supplémentaires. Il est particulièrement intéressant d'introduire dans l'analyse lissée les éléments du tableau initial X. Ces projections peuvent même être, dans certains cas, les résultats essentiels de l'analyse lissée.

En effet, on souhaite souvent garder exactement les individus initiaux pour lesquels on a la plupart du temps des informations extérieures aux données traitées, tout en les représentant sur des axes de dispersion générale moins dépendants des variations locales que les axes issus de l'A.C.P. de X. La projection en supplémentaires des individus initiaux dans l'analyse lissée répond exactement à ce problème. D'autre part, la comparaison des projections des individus d'origine et des individus lissés permet d'étudier, dans une certaine mesure, l'effet du lissage.

La projection des variables initiales permet de visualiser leur position et de les comparer aux variables d'origine. Notons que dans le cas d'un graphe de partition (analyse inter) cette projection ne présente pas d'intérêt car les variables lissées et les variables initiales se confondent sur tous les axes puisque le nuage des variables lissées est une projection du nuage d'origine. Dans la plupart des applications pratiques ces projections sont assez proches car, d'une part, les lissages choisis modifient souvent assez peu les variables et, d'autre part, les différences entre variables initiales et variables projetées ne sont pas mises en évidence sur les axes d'inertie des variables lissées.

## 4.2. Analyse factorielle des différences locales (A.F.D.L.)

Le problème étudié est l'opposé du précédent. On cherche à analyser les différences locales, en se dégageant des tendances générales liées à la contiguïté, comme par exemple un gradient nord-sud qui peuvent déterminer les premières composantes principales dans une A.C.P. classique. Le principe de l'analyse des différences locales est exactement le complémentaire de celui de l'analyse lissée. Il consiste à remplacer chaque point représentant une ligne du tableau X par ses différences avec le barycentre de ses voisins. Une analyse des différences locales permet notamment de mettre en évidence des "anomalies" régionales, des accidents dans des séries chronologiques. Elle permet aussi de regrouper des lignes qui diffèrent entre elles, mais dont le point commun est de s'écarter de la même manière de leur voisinage respectif. Une telle analyse peut, comme l'analyse lissée, peut être utilisée comme première étape d'une classification basée sur les différences locales.

Des solutions différentes sont proposées dans (5) et (2) pour résoudre des problèmes assez proches.

### 4.2.1. Nuage des individus

Chaque individu  $i$  est représenté par un point qui représente sa différence avec le barycentre de ses voisins. Ses coordonnées s'écrivent :

$$z_{ij} = x_{ij} - y_{ij}$$

L'ensemble de ces coordonnées définit un tableau Z qui s'écrit matriciellement

$$Z = X - Y = X - ATX$$

Deux individus sont proches s'ils s'éloignent de la même façon de leurs voisins (dans le nuage d'origine ces mêmes points peuvent être très éloignés). Il résulte de la relation entre les tableaux Y et Z, que ce nuage n'est centré que lorsque le nuage d'individus défini dans l'analyse lissée l'est (graphe de partition ou somme des poids des voisins constants). Pour avoir une bonne représentation de ce nuage d'individus, il faut le centrer et donc appliquer une A.C.P. centrée au tableau Z.

Dans le cas d'un graphe de partition, on retrouve exactement l'analyse intra. Dans le cas d'un graphe complet, le nuage n'est pas modifié, l'analyse des différences locales se confond avec l'A.C.P. de X tandis que dans le cas d'un graphe réduit à l'identité tous les points du nuage sont confondus.

### 4.2.2. Nuage des variables

Les colonnes du tableau Z définissent de nouvelles variables que nous appelons différences locales. Ces variables ne se déduisent des variables initiales par une projection sur un sous-espace de  $R^n$  que dans le cas d'un graphe de partition.

### 4.2.3. Projection des variables et des individus du tableau initial

Des individus et des variables peuvent être introduit en éléments supplémentaires,, mais les projections du tableau initial X ont beaucoup moins d'intérêt que dans l'analyse lissée.

Si dans l'analyse lissée, la projection des individus initiaux est souvent un des résultats essentiels, ceci est rarement le cas en analyse des différences locales. En effet, si "l'individu lissé" a souvent peu de signification concrète et joue plutôt le rôle d'intermédiaire de calcul, "l'individu local" qui traduit l'écart au voisinage garde pleinement son identité propre et s'interprète facilement : la proximité d'individus dans l'analyse locale traduit le même type "d'anomalies" par rapport au voisinage, anomalies que l'on peut relier à des connaissances extérieures.

La position des individus initiaux, qui peuvent être très éloignés des individus locaux n'est guère utile, ni pour l'interprétation d'axes qu'ils ne déterminent pas, ni pour leur étude propre. D'autre part, si la comparaison des positions des individus lissés et des individus initiaux (qui se réfèrent toutes deux à des notions globales) a un sens, celle des positions des individus locaux et des individus initiaux qui se réfèrent à des notions très différentes est beaucoup plus délicate.

La plupart du temps, les projections des variables initiales (très différentes des variables de différences locales) ne présentent guère d'intérêt. Comme pour l'analyse lissée, dans le cas d'un graphe de partition, ces projections sont confondues avec celles des variables des différences locales puisque le nuage analysé est une projection du nuage d'origine sur un sous-espace.

## 4.3. ANALYSE DES COVARIANCES OU DES CORRELATIONS

L'analyse lissée et l'analyse des différences locales aboutissent à des A.C.P., l'une du tableau Y et l'autre du tableau  $Z = X - Y$ . On peut analyser ces deux tableaux, soit en réduisant les variables, soit en ne les réduisant pas.

En dehors du cas de la partition, l'inertie des nuages définis par X n'est pas la somme des inerties des nuages définis par Y et Z, les barycentres des nuages d'individus associés à X et à Y ne sont d'ailleurs même pas toujours confondus. L'analyse lissée et l'analyse des différences locales ne sont donc pas, dans le cas général, les analyses de deux parties exactement complémentaires de l'inertie de X. De ce fait, les résultats de ces différentes analyses, et notamment les valeurs propres, sont à comparer avec précaution, même si les A.C.P. de Y et de Z sont des A.C.P. non normées. Cette propriété étant de toutes façon perdue, on appliquera généralement une A.C.P. normée.

## 5. QUELQUES GRAPHES PARTICULIERS

### 5.1. Cas d'un graphe pondéré

Si le graphe n'est pas pondéré, tous les voisins d'un individu (y compris l'individu lui-même) ont la même influence dans le lissage. Si un individu a beaucoup de voisins, ses caractéristiques peuvent être "couvertes" par celles de ses voisins, ce qui peut être un inconvénient. D'autre part, dans le cas d'une proximité géographique (ou temporelle) par exemple, on peut souhaiter que des zones interviennent dans le lissage avec une importance qui décroît avec la distance. Pour résoudre ces problèmes, un poids différent de 1 peut être attribué à chaque arête  $(i, i')$  du graphe. Ce poids est fixé par l'utilisateur en fonction du problème traité.

### 5.2. Cas d'un graphe non symétrique (orienté)

Un graphe orienté exprime une hypothèse de liaisons non symétrique entre l'ensemble des individus. Donnons deux exemples où il est utile d'introduire un tel graphe :

- hypothèse Markovienne d'ordre  $r$  sur des données évolutives, le graphe associé est de la forme suivante

$$a_{(i,t)(i,t')} = 1 \text{ si } 0 \leq t - t' \leq r \\ = 0 \text{ sinon}$$

- hypothèse sur le sens de la pollution due à un département sur l'écologie des départements voisins, ce graphe est de la forme

$$a_{ij} = 1 \text{ si } i \text{ pollue } j \\ = 0 \text{ sinon}$$

### 5.3. Cas d'un graphe issu de variables instrumentales

La méthodologie présentée précédemment s'applique quand le graphe est construit de façon artificielle. Le graphe peut être déduit de distance ou de similarités sur l'ensemble des individus définis par un groupe de variables "instrumentales". L'analyse factorielle lissée renforce les liaisons qui passent par ces variables, tandis que l'analyse des différences locales neutralise leurs effets. Ces analyses ont à la base les mêmes idées introduites par Rao (7) sur les variables instrumentales et développées depuis dans d'autres contextes.

## 6. COMPARAISON DE L'ANALYSE DES DIFFERENCES LOCALES AVEC D'AUTRES TECHNIQUES

Plusieurs analyses factorielles ont été proposées dans le but d'analyser des liaisons locales entre les variables. La plus connue, dont les programmes avaient été implantés dans la bibliothèque Spad, est l'analyse locale introduite par L. Lebart (4) et Aluja(5). L'analyse en composantes principales des unités statistiques corrélées a été introduite par Aragon et Caussinus (3) selon un point de vue différent. Ces techniques peuvent être regroupées dans un contexte plus général, celui où on introduit des poids éventuellement nuls ou négatifs aux couples d'individus. Nous allons comparer ces techniques à l'analyse des différences locales au niveau de la matrice de covariances diagonalisée, de la représentation des individus et à celui de la représentation des variables. Nous concluons par la mise en oeuvre et l'interprétation des résultats.

### 6.1. Les matrices de covariances

L'analyse locale est introduite à partir de la matrice de covariance. Rappelons que le terme général  $V_{jj'}$  de cette matrice  $V = (1/n) X'X$  s'exprime comme une somme, soit sur l'ensemble des individus (1), soit sur celui des couples d'individus (2) :

$$(1) \quad V_{jj'} = (1/n) \sum_i (x_{ij} - \bar{x}_j) (x_{ij'} - \bar{x}_{j'})$$

$$= (1/n) \sum_i x_{ij} x_{ij'} \quad \text{si les variables sont centrées}$$

$$(2) \quad V_{jj'} = (1/2n^2) \sum_{i, i'} (x_{ij} - \bar{x}_{i'j}) (x_{ij'} - \bar{x}_{i'j'})$$

En analyse locale, la contiguïté est définie par un graphe non pondéré. La variance locale  $L$  est définie à partir de la deuxième expression en supprimant de la somme les couples qui ne sont pas reliés par une arête :

$$L_{jj'} = (1/m) \sum_i \sum_{i' \in V(i)} (x_{ij} - \bar{x}_{i'j}) (x_{ij'} - \bar{x}_{i'j'})$$

où  $m$  est le double du nombre d'arêtes.

Dans l'analyse des différences locales, la matrice de covariance  $W$  est définie par les nouvelles variables. Elle se déduit la première expression en remplaçant les moyennes générales des variables par les moyennes locales  $y_{ij}$  :

$$W_{jj'} = (1/n) \sum_i (x_{ij} - y_{ij}) (x_{ij'} - y_{ij'}) - y_j y_{j'}$$

Le dernier terme, le produit des moyennes, découle du centrage.

Dans les deux cas, seules les différences locales (entre les points reliés par le graphe) augmentent les covariances : la covariance entre  $j$  et  $j'$  est positive si  $j$  et  $j'$  varient dans le même sens, *par rapport à leurs voisins*.

Mais ces matrices, et donc les analyses sont différentes. En effet, en développant  $L_{jj'}$ , on obtient :

$$\begin{aligned} L_{jj'} &= (1/m) \left( \sum_i x_{ij} x_{ij'} n_i - \sum_j x_{ij} y_{ij'} n_i - \sum_i y_{ij} x_{ij'} n_i + \sum_i \sum_{i \in V(i)} x_{ij} x_{ij'} \right) \\ &= (1/m) \sum_i n_i (2 x_{ij} x_{ij'} - x_{ij} y_{ij'} - y_{ij} x_{ij'}) \end{aligned}$$

En développant  $W_{jj'}$ , on obtient :

$$W_{jj'} = (1/n) \sum_i (x_{ij} x_{ij'} - x_{ij} y_{ij'} - y_{ij} x_{ij'} + (y_{ij} - \bar{y})(y_{ij'} - \bar{y}'))$$

Les deux expressions sont assez semblables.

Cependant, dans  $L$ , chaque individu  $i$  intervient avec un poids égal au nombre de ses voisins, les individus ayant un voisinage riche interviendront donc plus que ceux qui ont un voisinage pauvre ; alors que dans  $W$  tous les individus interviennent avec le même poids. D'autre part, dans  $W$ , la covariance entre les variables initiales se cumule avec la covariance entre les variables lissées ; alors que dans  $L$ , la première intervient deux fois.

Matriciellement, on a :

$$L = (1/m) X' T^{-1} (X - Y) + (1/m) (X' - Y') T^{-1} X$$

$$W = (1/n) (X - Y) (X' - Y') - Y' U' U Y$$

où  $U$  est un vecteur dont les  $n$  coordonnées valent  $1/n$ .

### Cas d'un graphe de partition

Nous avons vu que dans le cas d'un graphe de partition, la variance  $W_{jj'}$  se confond avec la variance intra. Dans la variance locale  $L_{jj'}$ , on obtient une expression analogue, mais chaque individu a un poids proportionnel au nombre de ses voisins : les individus des classes nombreuses interviennent avec une importance plus grande que ceux des classes faibles. La variance ne se confond avec la variance intra que dans le cas où toutes les classes ont le même nombre d'éléments.

### Graphe pondéré et non symétrique

Bien que les programmes proposés ne le permettent pas actuellement, l'analyse locale pourrait être généralisée sans difficulté particulière à un graphe pondéré en mettant dans l'expression (2) de la variance, des poids aux arêtes.

### Covariance et corrélations

Dans les deux cas, il est possible de considérer une matrice de corrélation au lieu d'une matrice de covariance.

## **6.2. Représentation des individus**

Dans l'analyse locale, les individus du tableau initial  $X$  n'apparaissent plus. Mais on voit facilement que l'analyse locale est équivalent à une A.C.P. d'un tableau  $A$  dont les lignes sont les couples d'individus reliés par une arête du graphe ; la valeur de la variable  $j$  pour le couple  $(i, i')$  est la différence  $x_{ij} - x_{i'j}$ .

Il est tout à fait possible de proposer une représentation des individus dans cette analyse en mettant le tableau  $X$  en supplémentaire dans l'A.C.P. de  $A$ . Ce qui revient à projeter le nuage d'origine sur des axes différents des axes d'inertie (les axes de projection maximisent la somme des carrés des distances entre les couples de points reliés par une arête du graphe). C'est d'ailleurs ainsi que l'A.C.P. sur des unités statistiques corrélées (6) est présentée : les auteurs proposent d'affecter des poids aux couples de points individus et de chercher des axes qui maximisent la somme ainsi pondérée des carrés des distances des couples de points. Le principe général de cette analyse est donc équivalent à celui de l'analyse locale.

Dans l'analyse des différences locales, les individus "locaux" sont représentés directement et on dispose non seulement de leur qualité de représentation, mais aussi de leur contribution à l'inertie, qui est un indice d'aide à l'interprétation essentiel. La représentation des individus que nous proposons ici s'apparente beaucoup plus à celle obtenu en mettant le tableau  $X$  en individus supplémentaires dans l'A.C.M. de  $Z$ . Nous avons souligné au paragraphe 4.2.3. l'intérêt limité de ces représentations.

## **6.3. Représentation des variables**

L'analyse locale étant équivalente à une A.C.P. d'un tableau  $A$  dont les lignes sont les arêtes du graphe, on peut définir des variables "locales" dont le domaine de définition est l'ensemble des arêtes et non l'ensemble des individus. Ces variables peuvent être représentées dans l'analyse, mais leur domaine de définition n'étant pas l'ensemble classique, l'interprétation est plus délicate que celle des variables de l'analyse des différences locales.

Pour comparer ces variables "locales" aux variables initiales, il faut que les domaines de définition soient identiques. Le tableau  $A$  peut être complété par des lignes nulles pour les couples qui ne sont pas reliés par une arête; l'ensemble de définition est alors l'ensemble des



couples d'individus. Les variables initiales sont représentées par des variables "générales" définies par le graphe complet. Notons que les variables locales sont des projections orthogonales des variables générales sur le sous-espace engendré par les arêtes du graphe et que dans l'analyse locale les deux types de variables ont leur projections confondues. Pour les comparer les auteurs proposent de projeter en élément supplémentaires les variables locales dans l'analyse des variables générales. Notons que les variables de différences locales (tableau Z) et les variables lissées (tableau Y) peuvent être introduites dans l'A.C.P. du tableau initial.

#### **6.4. Mise en oeuvre et interprétation des résultats**

La mise en oeuvre de l'analyse des différences locales consiste à appliquer un programme classique d'A.C.P. à un tableau obtenu par une transformation simple des données tandis que l'analyse locale nécessite un programme spécifique. Dans les deux cas, il est bien entendu nécessaire d'introduire le graphe de contiguïté. L'interprétation des résultats de l'analyse des différences locales est absolument analogue à celle d'une A.C.P. classique et, contrairement à l'analyse locale, peut s'appuyer sur les indices de qualité de représentation et de contribution à l'inertie tant des variables que des individus ainsi que sur la dualité entre les variables et les individus.

## **6 EXEMPLES D'APPLICATION**

Nous utilisons pour l'ensemble des traitements les programmes de la bibliothèque SPAD (Système Portable pour l'Analyse des Données). La démarche méthodologique est la suivante :

- a) A.C.P. sur le tableau (X, Y ou Z)
- b) Classification hiérarchique sur les composantes principales issues de a)
- c) Explication des classes de la hiérarchie obtenue en b).

Nous illustrons nos propos sur deux exemples :

- le premier exemple traite des données de la production agricole française pour l'année 85
- le second exemple traite de l'évolution de cette production entre les années 1959 et 1985.

### **Exemple 1**

- Les données :

Il s'agit de 13 productions agricoles, d'origine végétale et animale, mesurée en francs constants sur les 89 départements français.

production	
végétale	animale
céréales pommes de terre légumes fruits vin	porc veau gros bovins ovins-caprins lait volaille.+oeufs

Tabl.

- Le graphe :

Le graphe considéré est le graphe de contiguïté géographique des départements (deux départements sont contigus s'ils ont une frontière commune).

- Les objectifs :

Certains départements ont un rôle "moteur" sur la production agricole des départements voisins. On souhaite mesurer cette influence, en donnant une image globale des régions de France caractérisées par certains types de productions; techniquement nous appliquons l'analyse factorielle lissée au tableau de données X. Notre second objectif est de "neutraliser" sur chaque production l'effet du graphe. Il s'agit là d'analyser les différences locales entre l'ensemble des départements. Ceci permet de voir quels sont les départements qui s'écartent (ou se ressemblent) par rapport à leurs voisins. Nous appliquerons pour cette étude l'analyse factorielle des différences locales.

A - A.F.L. du tableau X

On donne pour les trois premiers axes factoriels les taux d'inertie et les "variables lissées" les plus discriminantes, avec leur contribution.

	Axe1	Axe2	Axe3
Taux d'inertie	40.49	19.60	15.01
Gros-bovins	13%	-	-
Veau	15%	-	-
Porc	18%	-	-
Lait	20%	-	-
Volailles	17%	-	-
Ovins-caprins	-	13%	-
Céréales	-	31%	-
Pommes de terre	-	24%	-
Légumes	-	-	24%
Fruits	-	-	40%
Vin	-	-	20%

Tab2.

L'inertie totale des trois premiers facteurs est de 75.01%. Cette inertie est supérieure à celle donnée par les trois premiers axes de l'A.C.P. du tableau X.

#### Premier facteur "Production animale"

Cet axe oppose les départements de l'Ouest, en particulier la Bretagne, aux autres départements de France, en particulier le Sud. L'Ouest est caractérisé par une production animale (Porc, volaille, lait, gros bovins et veau), et une faible production de fruits, légumes et vin qui caractérise plutôt les départements du Sud de la France.

#### Second facteur "Production végétale et ovins caprins"

Ce facteur met principalement en relief les départements du Centre et du Limousin produisant des ovins-caprins qui s'opposent aux départements du Sud et du Sud-Ouest qui, eux, se caractérisent pour l'essentiel par une production de fruits, légumes et vin.

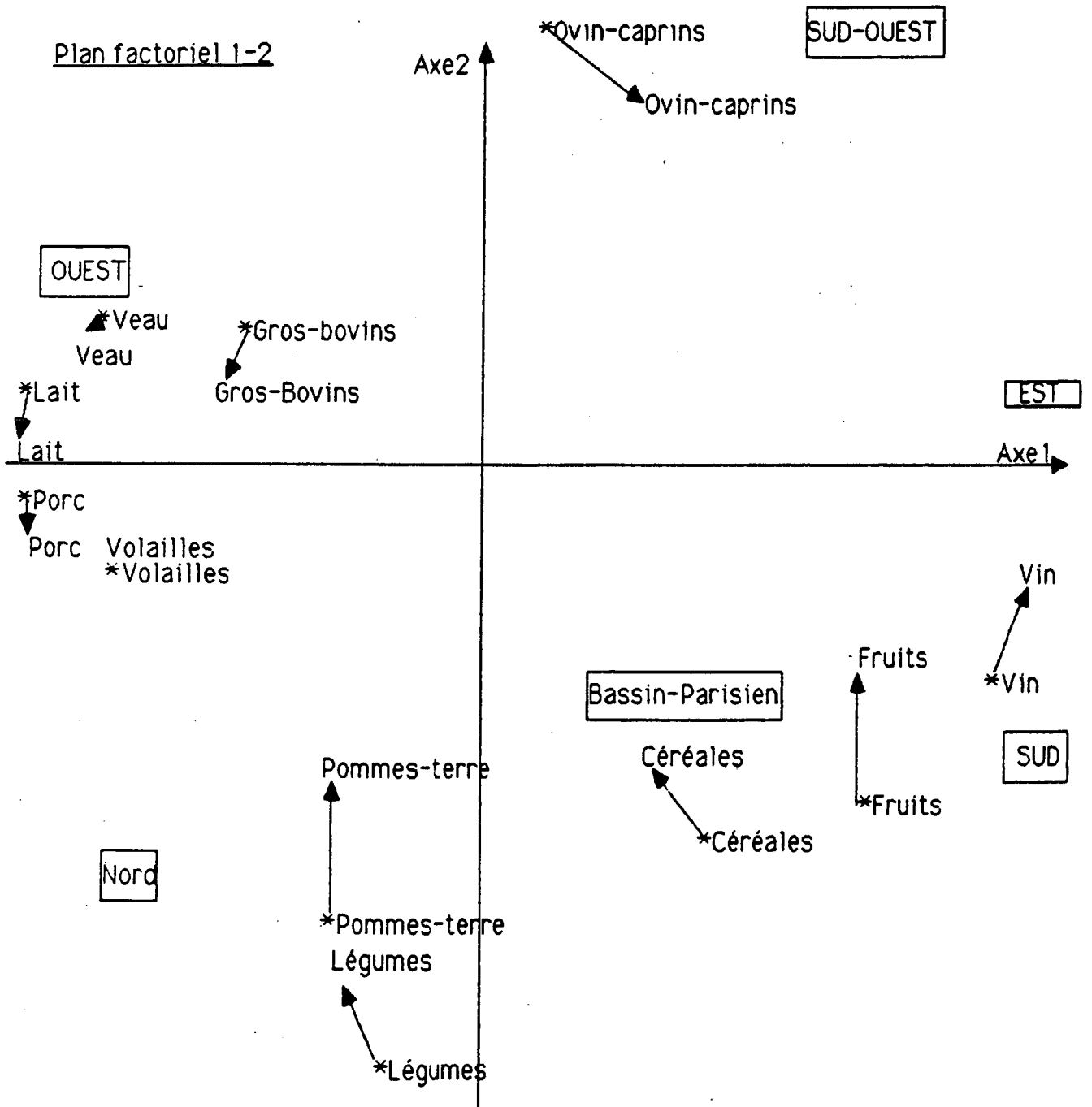


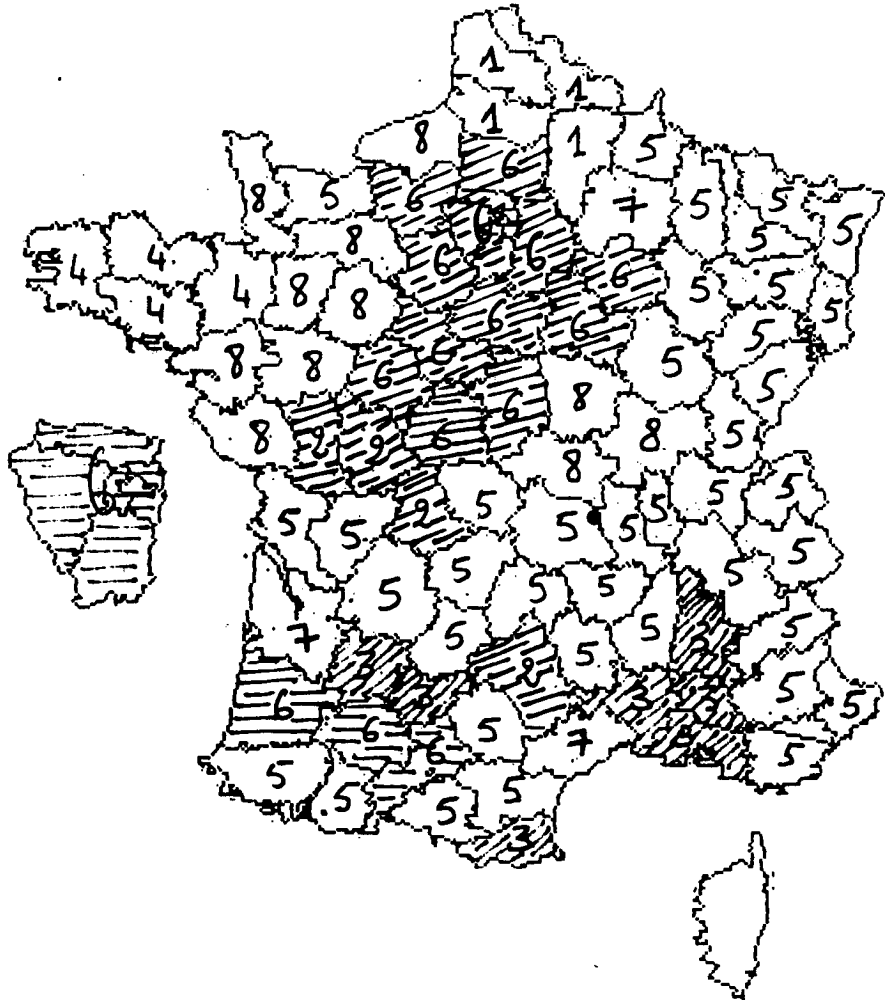
Fig1.

### Premier plan factoriel

Il est intéressant de noter les positions relatives des différentes variables sur ce plan, et plus particulièrement celles des variables d'origine (en bas des flèches sur la fig.1) et des variables lissées. Une variable est d'autant plus contiguë au graphe qu'elle s'éloigne de la variable d'origine; ce résultat est confirmé par deux indices de mesures de contiguïté, celui de Geary est un autre indice que nous avons mis au point.

### Interprétation de la C.A.H

On a effectué une C.A.H à partir des huit coordonnées factorielles des départements, totalisant une inertie de 98.7 %. Une forte variation de l'indice de la hiérarchie donne une partition en huit classes.



classe 1	classe 2	classe 3	classe 4	classe 5	classe 6	classe 7	classe 8
4	4	7	4	41	15	3	11

Tab3.

Description des classes

Nous donnons pour l'ensemble des classes les productions les plus caractéristiques, leurs moyennes dans la classe, ainsi que dans la population générale. La figure 2 donne l'appartenance des départements aux classes.

Classe 1

	potomme de terre	légumes	céréales
moy/cl	230.50	133.27	587.17
moy/pop	29.26	46.27	229.41

Le Nord, le Pas de Calais, la Somme et l'Aisne composent cette classe. On observe une forte production de pommes de terre, de légumes et de céréales.

Classe 2

	ovins - caprins
moy/cl	154.20
moy/pop	24.31

La production d'ovins - caprins est la seule caractéristique des départements de cette classe.

Classe 3

	céréale	lait	gros bovin
moy/cl	595.12	63.92	85.59
moy/pd	229.41	124.51	156.91

Les départements de cette classe se trouvent pour la grande majorité dans le Bassin Parisien. Les autres départements se localisent au Sud-Ouest [ fig ]. Leur spécificité réside dans la production de céréales. On note aussi une sous production, par rapport à la moyenne nationale, de gros-bovins et de lait.

#### Classe 7

	vin
moy/cl	842.87
moy/pd	88.70

Les départements de cette classe sont : la Marne, la Gironde et l'Herault. L'essentiel de leur production est le vin.

#### Classe 8

	gros bovin	lait
moy/cl	369.98	329.91
moy/pd	124.51	157.01

La majorité des départements du pays de la Loire et de la Bourgogne se trouvent dans cette classe . Ils se caractérisent par une production de gros bovins et de lait.

L'analyse factorielle lissée donne une image de l'agriculture française, tenant compte à la fois de l'homogénéité de la production et des proximités géographiques. Les classes de départements obtenues sont beaucoup plus contiguës que celles de l'ACP classique.

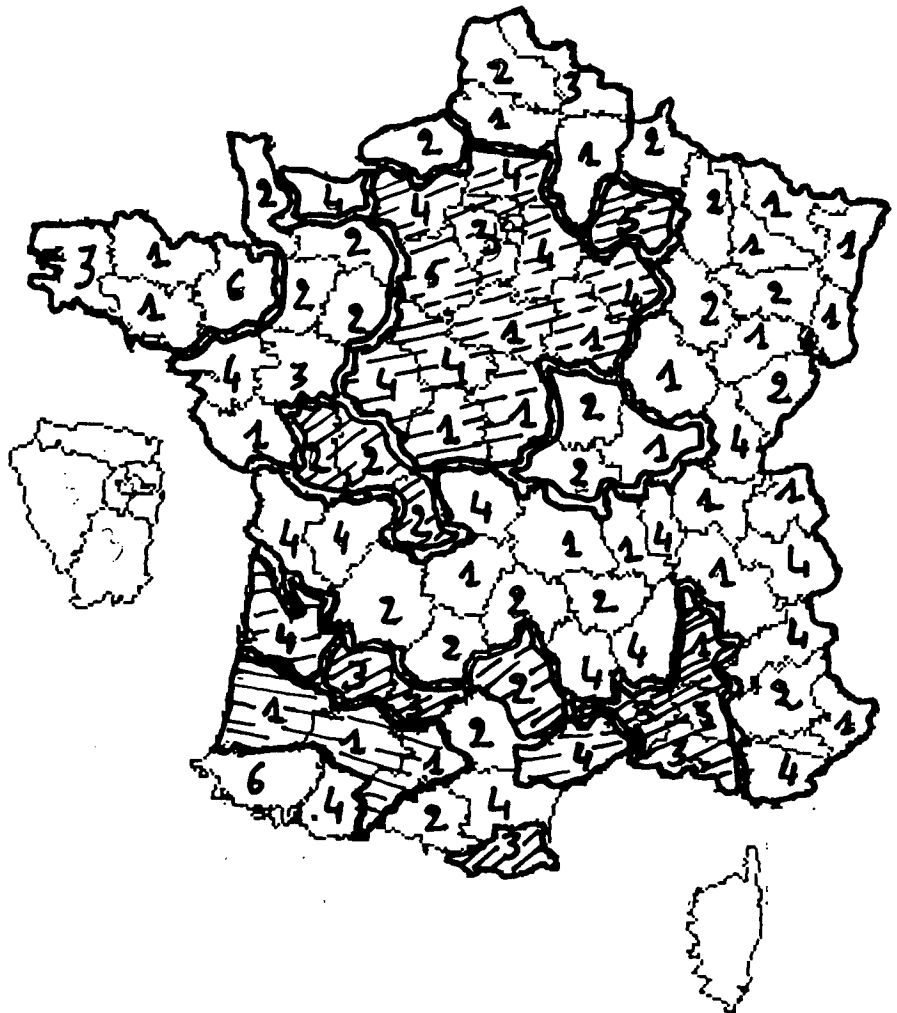
#### B - A.F.D.L du tableau X

On souhaite dans cette analyse, neutraliser l'effet du graphe des proximités géographiques. Ceci permet d'éliminer les grandes tendances observées précédemment. On donnera une image des différences locales de la production agricole dans chacune des classes mises en relief dans la partie A.

Les variables se trouvant à la base des flèches ( représentées sur le premier plan factoriel ) sont les variables d'origines. Les flèches représentent l'intensité de la liaison d'une variable avec le graphe de contiguïté géographique. La variance locale des variables est plus faible que leur variance initiale. On observe alors une contraction du nuage des variables vers l'origine. Certaines, comme la production de volailles plus oeufs et céréales, apparaissent comme indépendantes dans l'analyse lissée - où l'on observe uniquement les tendances globales - . Elles présentent des liaisons dans cette analyse locales ( fig2 )

Le groupe des départements de l'Est observé dans l'analyse A et ayant une faible production agricole ne se retrouve plus dans l'analyse des différences locales. Néanmoins, on observe dans cette analyse, une classe ( classe 4 ) contenant des départements ayant une faible production agricole par rapport à leur voisins. Il est intéressant de noter dans ce cas que parmi les départements de l'Est, le Jura, la Savoie, les Hautes Alpes et le Var se placent parmi les départements français de plus faible production agricole.

Une autre remarque intéressante concerne la classe 2 de l'analyse A qui se caractérise principalement par une production d'ovins-caprins. Les mêmes départements de cette classe se retrouvent avec les mêmes types de caractéristiques quand on les compare à leurs voisins ( fig2. ), montrant ainsi que les départements de l'Aveyron, des Deux sèvres, de la Vienne et de la Haute Vienne ont axé uniquement leur production sur l'ovin-caprin





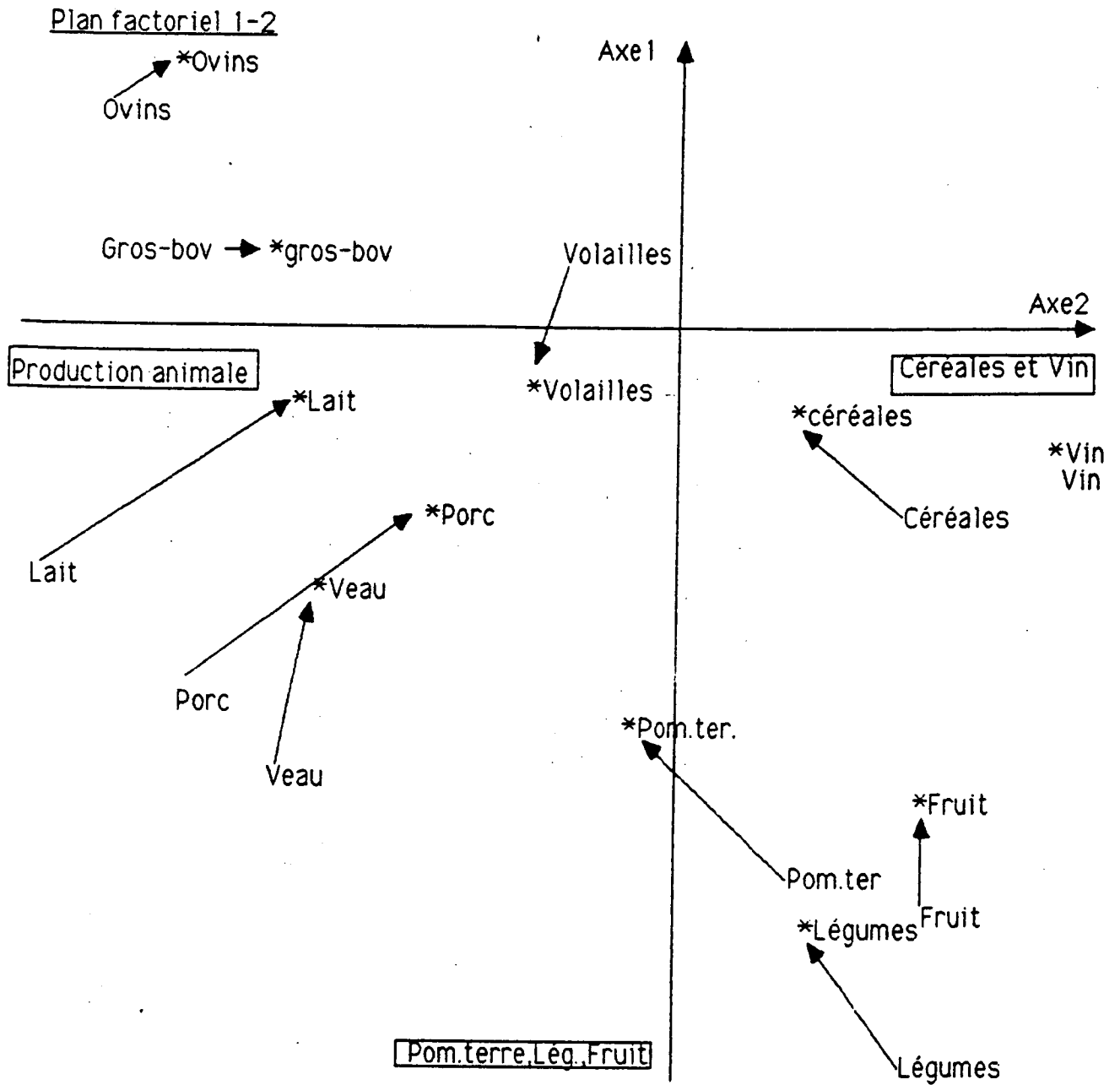


Fig2.

Classe 1	volaille+oeuf	céréale	pomme de terre
moy/clas	42.7	53.66	9.68
moy/pop	-0.61	-8.05	-0.17

Classe 2	ovins-caprins	lait	céréale	légume	vin	pomme de terre
moy/cla	18.17	37.77	-81.66	-17.81	-73.81	-12.57
moy/pop	-0.71	0.10	-8.05	0.75	1.29	-0.17

Classe 3	fruit	légume
moy/clas	73.53	65.67
moy/pop	-0.12	0.76

Classe 4	vin	lait	gros-bovins	volaille+oeuf	veau	porc	ovins-caprins
moy/c	65.1	-43.25	-48.16	-36.94	-14.2	-22.68	-11.33
moy/p	1.29	0.1	-1.62	-0.61	-0.68	1.14	-0.71

Classe 5	céréale	vin
moy/clas	512.75	347.5
moy/pop	-8.04	1.29

Classe 6	veau	lait	porc	volaille+oeuf
moy/clas	131.1	155.1	146.55	-113.85
moy/pop	-0.68	0.1	1.14	-0.61

## Exemple 2

Cet exemple traite de l'évolution de la production agricole entre 1959 et 1985. On cherche à détecter des périodes de production sur l'ensemble des départements.

### Les données:

On dispose du cube de données  $X_{IJT}$ , où:

I représente l'ensemble des 89 départements, J est l'ensemble des 13 productions présentées précédemment, et T représente la période 1959 à 1985.

Le tableau analysé est le tableau  $X_{(I \times T)J}$

### Le graphe

Le graphe que nous considérons dans cette étude est un graphe temporel définie ainsi:

$$a_{(i,t)}(i,t') = 1 \quad \text{si } |t-t'| = 1$$

$$a_{(i,t)}(i,t') = 0 \quad \text{sinon}$$

Ce graphe permet de lisser les trajectoires des départements. Ce lissage a un double avantage: d'une part, il permet de réduire les biais pour un même département d'une année à l'autre; (biais dus à des accidents climatiques ou autres), et d'autre part de tenir compte de façon explicite du phénomène temporel sur nos données.

### Les objectifs

Notre objectif premier, comme on l'a souligné en introduction, est de détecter des périodes 'types' de production. On cherche ainsi à caractériser dans le temps des ensembles de départements ayant les mêmes profils agricoles.

Nous présentons uniquement les résultats de l'étape tamis qui nous a permis de faire une très bonne synthèse entre les trajectoires des départements et secteurs de la production agricole. La lecture des plans factoriels était difficile à cause de la superposition d'un ensemble de points considérables. L'étape de la classification des trajectoires a permis de résoudre ce problème. On obtient ainsi 12 classes de départements(x année), que l'on caractérise par les différents types de productions.

Pour chacune des classes obtenues, on donne la moyenne des variables caractéristiques dans la classe et dans l'ensemble des départements(x années).

Bretagne et Nord**Manche 1960-1984**

Ille et Vilaine 1967-1984

Ille et Vilaine 1960-1966

Côtes du Nord 1960-1972

Morbihan 1962-1978

Morbihan 1979-1984

Côtes du Nord 1973-1984

Finistère 1974-1984

Finistère 1960-1973

**Nord 1960-1984**

	veau	lait	porc	gros bovin
moy/cl	193	633	228	214
moy/po	43	130	67	97

	pom. terre	porc	vol.	veau	lait
moy/cl	79	287	188	103	308
moy/po	20	67	58	43	130

	porc	vol.	lait	veau	gros bovin
moy/cl	819	558	600	148	241
moy/po	67	58	130	43	97

	pom. terre	leg.	porc	vol.	lait	g. bov
moy/cl	133	216	343	182	327	206
moy/po	67	47	67	58	130	87

Dans ces classes, on observe l'évolution agricole des départements bretons, du Nord et de la Manche. Pour l'essentiel, on a le même type de production sur l'ensemble de la période étudiée. Les départements de la Manche et du Nord présentent une stabilité agricole sur toute la période, comme le montrent les résultats de la classe 1 et de la classe 4. Remarquons que seule l'Ille et Vilaine intégrée dans le groupe Côtes du Nord et Morbihan se distingue sur la période 1967 à 1984 par une surproduction orientée vers le veau, le lait, le porc et les gros bovins, avec une baisse sensible de la production de pommes de terre. Le département du Finistère entre les période 1960 et 1973 a la même production que le département du Nord. Il rejoint par la suite les départements bretons ( classe 3), et intensifie ainsi sa production de porc, de volaille+oeufs, de lait, de veau et de gros bovins. Il diminue par contre sa production de pommes de terre et de légumes.

Le Sud

Gironde 1961-1984  
 Aude 1960-1984  
 Herault 1960-1984  
 Gard 1960-1962

	vin
moy/cl	657
moy/pop	79

Bouches du rhône 1961-1984  
 Vaucluse 1963-1984

	fruits	légumes	vin
moy/cl	226	247	177
moy/po	28	47	79

Marne 1960-1968  
 Côte d'or 1960-1963 et 66 73  
 Charentes Maritimes 1960-1984  
 Rhône 1972-1984

	vin
moy/cl	323
moy/po	79

Marne 1969-1984

	vin	céréales
moy/cl	823	684
moy/po	79	126

Les productions de vin, fruits et légumes, sont les variables les plus caractéristiques des départements du sud de la France. On observe dans la plus part de ces départements une stabilité dans la production agricole. C'est le cas :

- de la Gironde, de l'Aude, et de l'Herault dans la classe 1 avec une très forte production de vin.

- Des Bouches du Rhône et du Vaucluse avec une production de fruits et Légumes importante et une production de vin ( plus modeste que les départements précités ).

Le département de la Marne a intensifié sa production de vin et de céréales à partir de 1968, où l'on observe une rupture( classe 3 ). Il passe ainsi d'une production moyenne de vin ( intermédiaire entre les classes 1 et 2 ) à une très forte production de vin et de céréales ( classe 4 ).

Ile de France

	Céréales	Volaile
Gers 1969-1984		
Isère 1968-1984		
Eure et Loire 19660-1963	moy/clas 245	70
Loir et Cher 1964-1972 et 75-78	moy/pop 126	58
Yonne 1968-1977		
Eure 1968-1980		
Cher 1967-1982		
Cher 1983-1984		
Eure 1981-1984		
Yonne 1978-1984		
Loir et Cher 1979-1984		
Eure et Loir 1964-1984		
Oise 1973-1984		
Seine et Marne 1967-1984		
Loiret 1969-1984		
Seine et Marne 1961-1966		
Aisne 1960-1983		
Somme 1960-1965		
somme 1966-1984		

	céréales
	moy/clas 577
	moy/pop 126

	céréales	pomme de terre
	moy/clas 333	66
	moy/pop 126	20

	pomme de terre	céréales	lait	porc	Légume
moy/cl	117	350	271	190	102
moy/pI	20	126	130	67	47

L'ensemble des départements dans ces différentes classes ont une production de céréales. On observe par exemple dans les départements de la classe 2 une intensification de cette production après les années 80 ( en général ). Il s'agit là pour l'essentiel des départements de l'île de France. La Somme est le seul département à avoir diversifié sa production après 1965 ( classes 4 et 5 ). Sur la deuxième période on observe une stabilisation de sa production de céréales, une intensification de sa production de pomme de terre, de lait, de porc et de légumes. Ce département se rapproche par cette orientation agricole des départements du Nord et de la Manche.

Sud ( fruit )

Lot et Garonne 1960-1984  
 Tarn et Garonne 1963-1984  
 Drôme 1961-1984

	fruit	légume	volaille
moy/clas	118	82	87
moy/pop	28	47	58

Gard 1962-1984  
 Vaucluse 1960-1962  
 Pyrénées orientales 1960-1984

	fruit	vin	légume
moy/clas	143	390	117
moy/pop	28	79	47

L'ensemble de ces départements, indépendamment du vaucluse, ont une production stable de fruits et légumes sur la période étudiée.

Centre-Poitou charente-acquaine

Charente 1960-1962  
 Vienne 1960-1967  
 Dordogne 1960-1984  
 Pyrénées atlantiques 1960-1983  
 Gers 1960-1962 et 65-68  
 Tarn 1960-1976  
 Corrèze 1960-1984  
 Creuse 1960-1976  
 Isère 1960-1984  
 Haute loire 1960-1984  
 Puy de dôme 1960-1984  
 Haute vienne 1961-1964

	veau	ovins-caprins
moy/clas	96	23
moy/pop	43	16

Vienne 1976-1984  
 Haute vienne 1977-1984  
 Aveyron 1973-1984  
 Deux Sèvres 1979-1984

	ovins-caprins	gros-bovins
moy/clas	128	180
moy/pop	16	97

Nièvre 1978-1984  
 Saône et Loire 1968-1984  
 Deux sèvres 1965-1967 et 70-78  
 Allier 1967-1984

	gros-bovins	ovins-caprins	volailles
moy/cl	298	42	86
moy/pl	97	16	58

Aveyron 1960-1972  
 Tarn 1977-1984  
 Lot 1975-1984  
 Haute vienne 1965-1976  
 Vienne 1971-1975

	ovins-caprins	veau
moy/cl	74	70
moy/pl	16	43

Ariège 1960-1984  
 Hautes pyrénées 1960-1984  
 Lozère 1960-1984  
 Hautes alpes 1960-1984  
 Lot 1960-1984  
 Alpes de hautes provences 1960-1984

	ovins-caprins
moy/clas	27
moy/pop	16



Globalement, l'ensemble des départements de ces différentes classes ont une production d'Ovins-Caprins, avec tantôt une production de gros bovins tantôt une production de veau.

### Pays de loire

Calvados 1960-1984					
Orne 1960-1984					
Mayenne 1966-1977					
Sarthe 1962-1979					
Loire atlantique 1972-1981					
Vendée 1960-1973					
Vendée 1974-1984					
Sarthe 1980-1984					
Mayenne 1978-1984					

		gros-bovins	lait
moy/clas		250	285
moy/pop		97	130

		gros-bovins	volailles	lait	veau	porc
moy/cl		398	218	353	64	119
moy/p		97	58	130	43	67

La Sarthe, la Vendée et la Mayenne ont une rupture intéressante dans la production agricole. Ces départements ont sur la deuxième période de production les mêmes caractéristiques que les pays bretons. D'autres départements du pays de la Loire comme le Calvados et l'Orne gardent une stabilité dans la production de gros bovins et de lait.

Les autres départements ( sur les différentes périodes ) ne figurant pas dans ces classes ont une sous-production agricole touchant globalement tous les secteurs.

## 7 CONCLUSION

L'introduction de la notion de proximité par un graphe pondéré, éventuellement non symétrique, est très souple et permet de traduire la plupart des situations concrètes. Les deux méthodes proposées, l'analyse lissée et l'analyse des différences locales sont faciles à appliquer : elles ne nécessitent qu'une transformation simple du tableau des données avant une A.C.P. classique. L'interprétation des résultats est absolument analogue à celle d'une A.C.P. classique. La projection en éléments supplémentaires des individus et des variables du tableau initial enrichit l'analyse et permet de mesurer l'influence des transformations du tableau de données.

Dans le cas particulier d'une proximité induite par une partition de l'ensemble des individus en classes, l'analyse lissée se confond avec l'analyse interclasses tandis que l'analyse des différences locales se confond avec l'analyse intraclasses.

**Bibliographie**

- [1] Benali H. et Escofier B. 1988 Smooth factorial analysis and factorial analysis of local differences. Analysis of Multiway Data Matrices Edit. R.Coppi and S.Bolasco Amsterdam
- [2] Carlier A. 1985 Applications de l'analyse factorielle de l'évolution et de l'analyse intrapériode. Statistique et Analyse des données Vol.10
- [3] Caussinus H. et Aragon Y. 1980 Une analyse en composantes principales pour des unités statistiques corrélées. Data analysis and Informatics North Holland
- [4] Lebart L. 1984 Correspondance Analysis of graph structure. Bulletin technique du centre de statistique et informatique appliquées. Vol. 2 p.5-19
- [5] Aluja T. et Lebart L 1985 Factorial analysis upon a graph Bulletin technique du centre de statistique et informatique appliquées. Vol. 3 p.4-34
- [5] Foll Y. 1972 L'analyse des évolutions. Actes du petit séminaire sur l'analyse Factorielle Paris
- [7] Rao C.R. 1964 The use and interprétation of Principal Component Analysis in applied research. Sankhya A 26,4 p.329-358

## LISTE DES DERNIERES PUBLICATIONS INTERNES

- PI 438 - A PROPOS DE LA RESOLUTION D'UN SYSTEME LINEAIRE DANS UN CORPS FINI : ALGORITHMES ET MACHINES PARALLELES  
Hervé LE VERGE, Patrice QUINTON, Yves ROBERT, Gilles VILLARD  
22 Pages, Novembre 1988.
- PI 439 - ALPHA DU CENTAUR : A PROTOTYPE ENVIRONMENT FOR THE DESIGN OF PARALLEL REGULAR ALGORITHMS  
Pierrick GACHET, Patrice QUINTON, Christophe MAURAS, Yannick SAOUTER  
20 Pages, Novembre 1988.
- PI 440 - CONSTRUCTION METHODIQUE D'UN ALGORITHME REPARTI DE DETECTION DE LA TERMINAISON  
Jean-Michel HELARY, Michel RAYNAL  
18 Pages, Décembre 1988.
- PI 441 - LES GRAPHES A MOTIF  
Didier CAUCAL  
46 Pages, Décembre 1988.
- PI 442 - CAUSAL TREES  
Philippe DARONDEAU, Pierpaolo DEGANO  
44 Pages, Décembre 1988.
- PI 443 - TROIS IMPLANTATIONS DU RECUPERATEUR DE MEMOIRE DE LA MACHINE MALI  
Michel LE HENAFF, Hervé SANSON  
118 Pages, Décembre 1988.
- PI 444 - ANALYSE FACTORIELLE LISSEE ET ANALYSE FACTORIELLE DES DIFFERENCES LOCALES  
Brigitte ESCOFIER, Habib BENALI  
34 Pages, Décembre 1988.

