

**UNIVERSITE PARIS 1 PANTHEON SORBONNE**

*CENTRE DE RECHERCHE*

*S.A.M.O.S*

*STATISTIQUE APPLIQUEE ET MODELISATION STOCHASTIQUE*

**Cellulage de données d'enquêtes de  
consommation par une méthode neuronale**

**François GARDES Patrice GAUBERT  
Patrick ROUSET**

**Prépublication du SAMOS n°69  
Décembre 1996**

**90, rue de Tolbiac - 75634 PARIS CEDEX 13**

# Cellulage de données d'enquêtes de consommation par une méthode neuronale

F. GARDES, LAMIA, Université de Paris I  
P. GAUBERT, METIS, Université de Paris I  
P. ROUSSET, SAMOS, Université de Paris I

## I. Introduction

L'objet de cette étude est la mise en évidence de groupes d'individus homogènes du point de vue de leurs choix de consommation. L'intérêt d'un tel cellulaire est au moins double.

D'une part il permet de constituer des pseudo-panels lorsque l'on dispose d'enquêtes successives sur des individus distincts issus d'une même population mère. La constitution de groupes dans chaque enquête à l'aide d'une même procédure de cellulaire permet de créer de nouveaux « types » statistiques que l'on peut étudier longitudinalement.

D'autre part il peut permettre un appariement d'enquêtes réalisées sur des échantillons distincts d'une même population et apportant chacune des informations différentes de ce qu'apportent les autres. Si l'on dispose d'informations pour l'ensemble de la population, donc pour l'ensemble des individus composant les échantillons ci-dessus, la constitution de groupes d'individus homogènes du point de vue de ces informations doit permettre de raccorder les informations fournies par les enquêtes partielles.

On peut, par exemple, appliquer ces méthodes à l'appariement de moitiés d'enquêtes de consommation faites pour la même période auprès d'échantillons distincts (chaque sous-population étant interrogée sur les consommations d'une moitié de la nomenclature), appariement nécessaire à la reconstitution de profils de consommation complets. On notera, cependant, qu'un tel mode de cellulaire ne correspond pas à la méthode proposée par Deaton, méthode qui oblige à suivre, d'une enquête à l'autre, les mêmes générations (donc à comparer d'une enquête à l'autre des individus-types nés au même moment).

Lorsque l'on travaille avec des pseudo-panels, on a l'habitude de constituer des cellules en croisant quelques variables dont la théorie indique qu'elles ont une influence sur le phénomène étudié. Par exemple, dans l'étude des lois de consommation des ménages, les variables servant à constituer les cellules sont principalement la cohorte d'âge et le niveau d'éducation combinés avec la distribution du revenu<sup>1</sup>.

Ici il s'agit d'utiliser une méthode de classification particulière, un algorithme neuronal conçu par Kohonen<sup>2</sup> dont on met en évidence les résultats appliqués aux informations fournies par deux enquêtes de consommation, comparativement à ce que donnerait une classification traditionnelle.

La section 2 présente brièvement la méthode neuronale, la section 3 présente les enquêtes utilisées puis décrit les résultats obtenus.

Dans une étape ultérieure le pseudo-panel ainsi constitué sera utilisé pour estimer des élasticités-revenu pour les principales fonctions de consommation afin de les comparer aux résultats obtenus sur un cellulaire traditionnel des mêmes enquêtes.

---

<sup>1</sup> Voir Gardes, Langlois et Richaudeau (1995).

<sup>2</sup> T. Kohonen *Self-Organization and Associative Memory*, Springer-Verlag, 1989.

## 2. Présentation de la méthode

L'idée de base du réseau de neurones, tel qu'il est utilisé dans la présente étude, est la construction d'une structure uni- ou bi-dimensionnelle à l'aide d'un algorithme d'auto-organisation des individus<sup>3</sup>. Ce traitement possède des éléments communs avec les méthodes traditionnelles de classification et quelques différences essentielles.

### 2.1 Généralités sur l'analyse neuronale et l'algorithme de Kohonen

Les points principaux concernent la notion de distance, la conception de la structure par rapport à laquelle on définit la notion de voisinage, le phénomène d'apprentissage spécifié dans ce contexte. Le réseau est, habituellement, une structure uni- ou bi-dimensionnelle: les unités sont disposées en ligne ou en forme de grille, la notion de voisinage n'ayant pas la même extension dans l'un et l'autre cas.

#### 2.1.1 Distances

Chaque individu de la population à classer est représenté par un vecteur défini dans  $R^p$ , si l'on dispose de  $p$  variables les caractérisant. A chaque neurone de la structure est également associé un vecteur de poids appartenant au même espace. On attribue l'individu au neurone dont la distance euclidienne calculée sur ces variables est la plus faible. On prend soin de normer les variables définissant le vecteur d'entrées ainsi que les poids (variables centrées réduites).

Dans l'application présentée ici (structures budgétaires) on a utilisé la distance du  $\chi^2$ , afin d'éviter que les variables de faible dimension relative ne soient écrasées dans le calcul des distances et qu'elles puissent, le cas échéant, jouer leur rôle de variables discriminantes.

#### 2.1.2 Dimension de la structure et voisinage

La dimension doit être définie selon la spécificité du phénomène traité. Il est vraisemblable que plusieurs variables interviennent simultanément pour définir des comportements de consommation différents; ainsi une structure en grille (c'est à dire à deux dimensions) est souvent mieux adaptée à la classification que nous recherchons.

---

<sup>3</sup> Pour une présentation plus complète on peut se reporter à F. Blayo & M. Verleysen *Les réseaux de neurones artificiels*, PUF, 1996 et M. Cottrell & P. Letrémy « *Classification et analyse des correspondances au moyen de l'algorithme de Kohonen: application à l'étude de données socio-économiques* », prépublication du SAMOS, Université de Paris I, 1995.

De façon générale, si  $n$  est le nombre d'individus et  $p$  le nombre de variables servant à la classification, on se donne un réseau de  $K$  unités, chacune étant définie par un poids  $W_i$ , vecteur de  $R^p$  pour l'unité  $i$ <sup>4</sup>.

Pour démarrer le processus les poids sont tirés au hasard. Chaque individu est ensuite confronté à la grille, le tirage de l'individu étant fait selon une loi uniforme; on note  $x_j$  cet individu représenté par un vecteur dans  $R^p$ .

L'unité  $i_c$  du réseau à laquelle cet individu  $j$  est attribué est celle pour laquelle on obtient

$$\|x_j - W_{i_c}\| = \underset{i}{\text{Min}} \|x_j - W_i\|$$

Le poids de cette unité est recalculé pour tenir compte de l'ajout d'un individu. Ici intervient la définition du voisinage qui indique que des unités voisines dans la structure doivent avoir des poids proches à la fin de l'apprentissage et que deux individus qui possèdent des caractéristiques voisines doivent être classés dans la même unité ou des unités voisines. Par conséquent les poids des unités voisines sont également recalculés.

### 2.1.3 Apprentissage et convergence

Le fonctionnement de l'algorithme repose sur un phénomène d'apprentissage que l'on doit paramétrer dans la réalisation de la classification.

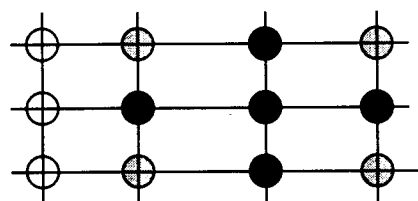
Autour de chaque unité on définit un voisinage plus ou moins grand, c'est à dire un nombre plus ou moins grand d'unités affectées par le classement d'un individu dans l'unité gagnante et dont on recalcule les poids. Ainsi pour une structure en ligne on peut avoir un voisinage à quatre unités (rayon 2)



ou à deux unités (rayon 1)



et de même pour une structure en grille



<sup>4</sup> Ce vecteur de poids indique donc la situation du nœud  $i$  dans l'espace des  $p$  variables de classification.

on peut considérer comme voisins d'une unité les quatre unités les plus proches (en gris foncé) ou étendre la zone d'adaptation à un nombre plus grand (8, incluant les nœuds en gris clair, ou 24, par exemple).

On fait décroître le rayon de voisinage au cours du processus de classification pour obtenir la convergence. On estime que la structure est stable lorsque l'on a, en moyenne, tiré 5 fois chaque individu; on peut après un premier passage utilisant une définition du voisinage à 24 unités (pour une grille), réaliser deux passages avec 8 voisins et terminer le traitement avec deux passages à quatre voisins.

Le processus d'adaptation s'écrit, en appelant  $t$  l'étape à laquelle le poids d'une unité a été défini, et  $i_c$  l'unité sélectionnée

$$W_i(t+1) = W_i(t) + \alpha(t) (x(t+1) - W_i(t)) \text{ pour } i \in V_{r(t)}(i_c)$$

avec  $V$  représentant l'ensemble des voisins de l'unité sélectionnée,  $r$  le rayon choisi;  $\alpha$  est le paramètre d'adaptation, il décroît au cours du processus comme une fonction inverse du nombre d'itérations. Cette modification du poids s'applique à l'unité sélectionnée elle-même comme à ses voisins.

Pour les unités qui n'appartiennent pas à  $V$ , on ne change rien.

L'algorithme converge en général avec un nombre d'itérations représentant 4 à 5 fois le nombre d'individus à classer. On peut alors établir une carte représentant l'organisation des individus et leurs caractéristiques, en classant tous les individus.

Dans l'application présentée ici<sup>5</sup>, les structures budgétaires sont utilisées pour construire la classification; on obtient ainsi des profils de consommation pour chaque classe dont on analyse les caractéristiques à l'aide d'un certain nombre de variables socio-démographiques.

## ***2.2 Une méthode classique: la classification hiérarchique ascendante***

On procède à une classification réalisée selon une technique de classification hiérarchique. Cette technique présente des similitudes avec la technique neuronale exposée plus haut, le calcul de distance à l'aide d'un vecteur de variables et la minimisation de la distance intra; il y a, au moins, une différence essentielle: la notion de voisinage dans la constitution des classes est absente des techniques de classification traditionnelles.

Il est possible toutefois de calculer des distances entre les groupes constitués et d'indiquer que certains classes sont plus proches entre eux qu'ils ne le sont avec d'autres.

Le principe de la classification est toujours de placer dans un même groupe des « individus » ayant une certaine similitude, alors que des groupes différents seront définis par une certaine dissimilarité. Pour réaliser ceci de nombreuses méthodes ont été utilisées, pour tenir compte des particularités de la distribution des observations à classer; des biais peuvent apparaître dans les différentes méthodes

---

<sup>5</sup> Le programme de classification mettant en Œuvre l'algorithme de Kohonen, ainsi que les programmes de représentations graphiques mettant en évidence les caractéristiques des classes obtenues, ont été écrits par Patrick Rousset, chercheur au SAMOS, Paris I.

selon les particularités rencontrées<sup>6</sup>. La méthode retenue ici est celle qui fournit en général les meilleurs résultats, celle de la variance minimum de Ward.

La distance entre deux classes est définie pour les classes  $K$  et  $L$  dont les effectifs sont, respectivement,  $N_K$  et  $N_L$ , par

$$D_{KL} = \frac{\|\bar{x}_K - \bar{x}_L\|^2}{\left(\frac{1}{N_K} + \frac{1}{N_L}\right)}$$

les  $x$  étant les vecteurs moyens de chaque cluster. Elle est équivalente à la variance « between » obtenue à partir de la variance « within » des classes:

si on écrit la variance du cluster  $K$

$$W_K = \sum_{i \in C_K} \|x_i - \bar{x}_K\|^2$$

on peut définir la variance « between » de l'ensemble constitué par la réunion des classes  $K$  et  $L$  par

$$B_{KL} = W_M - W_K - W_L$$

$$\text{pour } C_M = C_K \cup C_L$$

La classification est construite hiérarchiquement en déterminant, à chaque étape, la réunion de deux unités de l'étape précédente (individus ou classes) dans un cluster dont la variance intra est minimisée, sur l'ensemble des variables utilisées, par rapport à toute autre partition possible.

Un défaut de cette méthode est qu'elle tend à égaliser les effectifs des classes, faisant ainsi disparaître les classes à petits effectifs; elle est par ailleurs très sensible à la présence d'observations atypiques.

### 3. Application aux enquêtes canadiennes

On constitue des classes de ménages à l'aide d'un réseau de neurones et l'on compare les résultats obtenus avec ce qu'aurait fourni une classification classique.

#### 3.1 Présentation des enquêtes

On a utilisé ici deux enquêtes de consommation réalisées par Statistiques Canada en 1986 et 1992 auprès d'environ 10000 ménages non suivis d'une enquête à l'autre. Ces enquêtes ont déjà servi à la mise en évidence de comportements de consommation<sup>7</sup>.

La structure de consommation des ménages est connue dans une nomenclature à 20 fonctions (voir Annexe I). Les deux enquêtes ont été réunies afin de constituer des classes composées d'individus appartenant à l'une ou à l'autre des années; de cette façon il sera possible d'observer l'évolution dynamique de groupes de ménages ayant des structures de consommation similaires.

On constate que les classes obtenues contiennent de façon à peu près égale des ménages de l'enquête 1986 et des ménages de l'enquête 1992: il ne semble pas y

<sup>6</sup> Pour une présentation synthétique à partir de simulations voir « *Introduction to Clustering Procedures* » in *SAS/STAT User's Guide*, vol. 1, version 6, Cary, 1992.

<sup>7</sup> Cf. Gardes et alii.

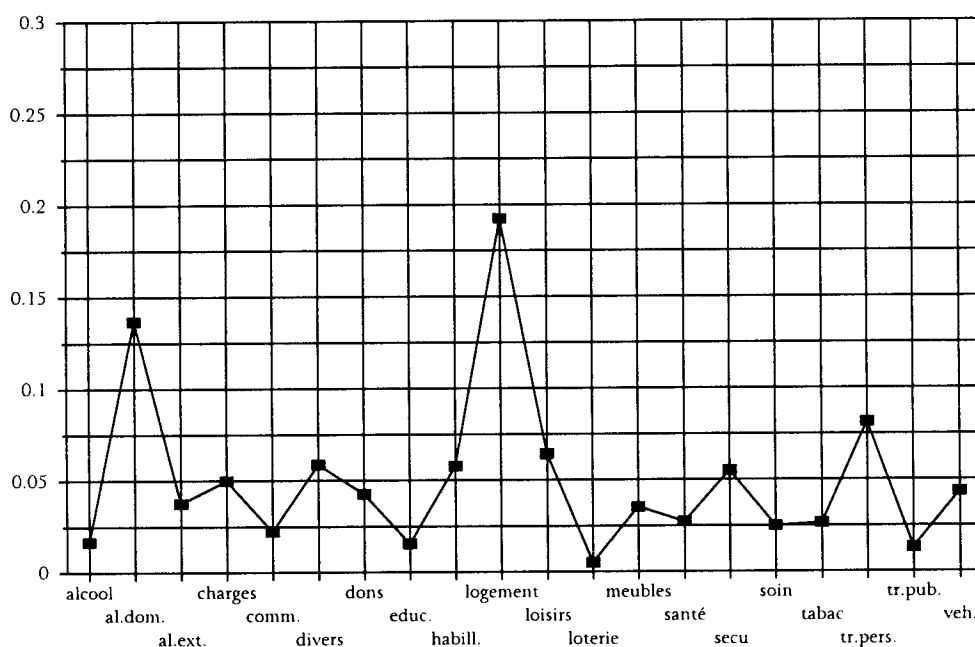
avoir pour les groupes constitués un effet temporel spécifique qui compliquerait les analyses transversales.

On a représenté le profil de la consommation de l'ensemble de la population. Ce profil moyen pour 1992 (graphe 1) fait apparaître quelques grands postes, notamment l'alimentation à domicile (légèrement inférieure à 15 %) et le logement (voisin de 20 %), les autres postes se situant entre 2 et 6 %.

Graphe 1

PROFILS DE CONSOMMATION

1992



3.2 Les classes obtenues et leur regroupement

Les individus des deux enquêtes ont été groupés en 64 classes au moyen d'un algorithme de Kohonen. Une analyse de classes sur les 64 classes a permis de constituer des groupes de classes voisines, ayant par conséquent des similitudes dans le profil de consommation.

3.2.1 Contenu des classes

On peut énoncer plusieurs caractéristiques de la classification Kohonen:

- les effectifs des classes
- la répartition entre les deux enquêtes
- la spécificité des profils de consommation



### 3.2.2 Les groupes de classes

Les groupes de classes ont des caractéristiques assez bien différenciées. Sur les 10 groupes obtenus on remarque

- *les effectifs*: ils se situent soit autour de 600 à 700 individus, soit à plus de 1200 individus, à l'exception d'un qui n'en compte qu'un peu plus de 400. Il s'agit d'un groupe qui ne rassemble que quatre classes ayant un profil bien différencié, que l'on explicitera plus loin.

- *les groupes* sont constitués, pour la plupart, à part égale d'observations de 1986 et de 1992, il ne semble pas y avoir d'effet d'année important dans la classification

- *les profils*: les groupes sont bien identifiés en termes de structure de consommation, lorsque l'on les compare au profil moyen de consommation de la population entière.

Neuf postes de consommation déterminent principalement la différenciation des groupes de classes de ménages<sup>8</sup>:

- i. un groupe fortement dominé par le *logement*: le groupe 5, la part de ce poste (0.38) étant très supérieure à ce qu'elle est dans la population entière

- ii. un groupe, le groupe 9, dont le logement occupe une place importante (0.26), ainsi que les *transports publics*

- iii. Deux groupes dominés par *l'achat de véhicule*:

- l'un (groupe 1) ne se distingue de la moyenne de la population que par cette composante (0.17) alors que les autres postes s'en trouvent réduits de façon quasi-homothétique

- et le groupe 2 dont l'achat de véhicule est dominant (36 % alors que la part moyenne n'atteint pas 5 %), la part du logement étant simultanément réduite, ce qui correspond à une composition moyenne du groupe où la catégorie des propriétaires est plus fortement représentée (71.4 % au lieu de 60)

- iv. Les groupes dominés par une composante autre

- *alimentation à domicile*: groupe 7 (0.20), marqué également par le poids du poste divers, essentiellement constitué des charges d'emprunts

- *dépenses de sécurité* (assurance chômage, maladie, retraite): groupe 3, la part occupée par ce poste est le double de ce qu'il est pour l'ensemble

- *dons*: groupe 10 (0.25 au lieu de 0.04)

- *loisirs*: groupe 8 (0.13 au lieu de 0.06)

- *tabac*: groupe 4 (0.12 au lieu de 0.02)

- *éducation*: groupe 6 (0.1 au lieu de 0.01).

On constate que ces regroupements sont fortement contrastés par rapport au profil moyen de consommation. Il s'agit, à chaque fois, du profil dominant du groupe, les classes qui le composent ayant des ressemblances puisqu'elles sont voisines dans la grille Kohonen, mais présentant des particularités propres plus ou moins accentuées. Les différenciations sont plus marquées sur les 64 classes<sup>9</sup>.

<sup>8</sup> Voir Annexe II les profils des groupes.

<sup>9</sup> Voir Annexe III les profils de classes et leur composition selon les différentes variables socio-démographiques.

### 3.3 Caractérisation des cellules

Une première approche des facteurs qui déterminent les profils de consommation identifiés est fournie par la représentation graphique de variables n'ayant pas servies à la classification, puis par une analyse statistique des groupes de classes à partir de ces variables.

#### 3.3.1 Représentations graphiques

On peut mettre en évidence quelques grandes caractéristiques des classes obtenues; on utilise pour cela des variables socio-démographiques issues de l'enquête ainsi qu'une typologie en termes de pauvreté-richesse définie à l'aide d'un critère synthétique<sup>10</sup>.

Brièvement, ce critère combine 3 critères simples:

- la distribution du revenu (représentée par les quartiles)
- la dépense totale
- la dépense alimentaire,

chaque individu étant placé dans une catégorie de pauvreté-richesse selon la place qu'il occupe, pour chacun de ces critères, dans la classe à laquelle il appartient lorsque l'on croise les tranches d'âge (six tranches), les niveaux d'éducation au nombre de trois et les cinq régions .

Les caractéristiques socio-démographiques utilisées sont

- l'âge du chef de famille
- la taille du ménage mesurée en unités de consommation selon l'échelle d'Oxford
- le niveau d'éducation (selon une nomenclature en trois postes: primaire, secondaire et post-secondaire<sup>11</sup>)
- le mode de détention du logement (propriétaire ou locataire)

ainsi que la dépense totale du ménage.

Les moyennes de variables quantitatives et les effectifs moyens des catégories (pauvres-riches, propriétaires, niveaux d'éducation) sont calculés pour chaque cellule Kohonen.

Les graphiques en secteurs (Cf. Annexes) fournissent une première représentation des caractéristiques de chaque cellule, à comparer avec celui des profils de consommation placés dans la grille Kohonen.

Les profils de consommation de la grille Kohonen sont proches dans un voisinage donné puisque cette grille a été construite à l'aide des coefficients budgétaires.

Simultanément on peut remarquer que les caractéristiques socio-démographiques font apparaître dans la grille des sous-ensembles ayant une certaine homogénéité;

---

<sup>10</sup> Cf. Gardes, F., P. Gaubert et S. Langlois « *La pauvreté...* ».

<sup>11</sup> Dans les graphiques le niveau d'éducation est repéré selon une nomenclature en 5 postes où l'on distingue en particulier les études post-secondaires, entreprises d'une part et sanctionnées par un diplôme d'autre part, du niveau diplôme universitaire.

ils se répartissent dans la grille différemment de ce que produit le classement sur les comportements de consommation.

Par exemple la répartition de la population selon le critère de pauvreté-richesse (5 classes) indique que, dans la grille à 64 classes, les classes comportant une forte proportion de riches et quasi-riches sont placées plutôt aux extrémités de la diagonale angle supérieur droit-angle inférieur gauche, les pauvres et quasi-pauvres occupant la zone centrale.

Dans le même temps la caractéristique propriétaire-locataire semble distribuée selon une opposition plus simple, sur cette diagonale, les premiers autour de l'angle inférieur gauche, les seconds à l'opposé; on retrouve ainsi une situation bien connue de pauvres qui peuvent être locataires ou propriétaires de leur logement. Ceci contribue à expliquer que les profils de consommation de ces deux ensembles soient différents.

On peut faire des constatations de même type pour les autres caractéristiques et, par conséquent, c'est la combinaison de ces caractéristiques qui doit être analysée pour interpréter les différentes zones de la grille telles qu'elles apparaissent, par exemple, dans les 10 groupes déterminés par l'analyse hiérarchique.

### 3.3.2 Analyse discriminante

On peut mettre en évidence une relation entre les profils de consommation et les variables socio-démographiques à l'aide d'une analyse canonique discriminante appliquée aux classes Kohonen pour expliquer ces 10 groupes.

L'analyse canonique discriminante permet, pour une variable de classification et des variables numériques mesurées sur des individus d'une population, de construire des variables canoniques dont on évalue l'influence sur la variable de classification; les variables canoniques sont des combinaisons linéaires des variables numériques, cette combinaison fournissant la signification particulière de chaque variable canonique.

Comme pour une analyse en composantes principales les variables canoniques sont ordonnées: la première est celle qui fournit le coefficient de corrélation multiple, entre les classes et les variables d'origine, le plus élevé; la deuxième est la combinaison, non corrélée avec la première, ayant la corrélation avec les classes la plus élevée après la précédente, etc.

Les variables doivent être distribuées de façon normale dans chaque classe, ce que l'on peut tester. On teste également la significativité de chaque variable d'origine, la nullité de la corrélation de chaque variable canonique et des suivantes. On peut ainsi

- définir les variables canoniques significatives
- interpréter, à partir des variables d'origine, chaque variable canonique
- indiquer ce qui différencie chaque classe des autres en obtenant, au point moyen de chacune, la valeur des variables canoniques.

Les résultats obtenus<sup>12</sup> fournissent une mesure de la combinaison des variables socio-démographiques qui détermine le regroupement des classes d'individus dans

---

<sup>12</sup> Voir le détail Annexe IV.

les 10 grands groupes; ils confirment les impressions fournies par l'observation des graphiques:

- les variables utilisées sont toutes significatives au seuil de 5 % (le niveau supérieur d'éducation est à la limite) avec un coefficient de corrélation multiple très correct;

- les corrélations canoniques fournissent quatre variables canoniques significatives, le test du rapport de vraisemblance indique que les suivantes ne sont pas différentes de 0

- l'étude de la structure des variables canoniques fournit une interprétation assez claire de ces variables:

- . la première est essentiellement composée du mode de détention du logement, de la taille de la famille et du niveau de richesse

- . la deuxième est déterminée par l'âge, le niveau primaire d'éducation et la pauvreté

- . la troisième, moins explicite que les trois autres, résulte de l'âge (négativement) et plutôt de la richesse pour un niveau d'éducation relativement plus élevé

- . la dernière variable significative est la résultante de la taille de la famille et de la pauvreté.

- les moyennes des variables pour chacun des 10 groupes fournissent la caractérisation de chacun

- . le premier constitué par des propriétaires relativement jeunes et aisés;

- . le deuxième se différencie du précédent par un vieillissement sensible et une richesse plus grande, toujours pour des propriétaires avec un niveau d'instruction plus faible que celui de la population entière;

- . le troisième est fortement marqué par le fait d'être locataire et la jeunesse des individus qui le constituent, ceux-ci appartenant en général aux classes aisées; on peut relier ceci à la caractéristique principale du profil de consommation, les dépenses de sécurité, les ménages utilisant leur position aisée pour se prémunir contre les risques (chômage, maladie);

- . les groupes 5 et 9, fortement dominés par les dépenses de logement se trouvent déterminés, naturellement, par la caractéristique de détention du logement, les premiers plutôt pauvres les seconds plutôt riches;

- . le groupe 6 est constitué de ménages aisés, jeunes et à niveau d'éducation plutôt inférieur à la moyenne, sa consommation ayant une forte composante éducation, ce qui n'est pas choquant;

- . le groupe 8 à dominante loisirs est plutôt constitué par des individus jeunes et aisés;

- . le dixième est constitué de propriétaires âgés, à niveau de ressources moyen ou élevé et dont la famille est d'une taille inférieure à la moyenne; ceci correspond à la caractéristique majeure de leur profil de consommation, les dons c'est à dire les transferts à des personnes extérieures au ménage.

### 3.4 Classification classique

#### 3.4.1 Constitution des classes

On peut obtenir une classification de la population à l'aide d'une analyse hiérarchique ascendante<sup>13</sup>, en se donnant un nombre de classes égal à celui retenu pour opérer un regroupement des classes Kohonen, soit 64 puis 10 classes. La méthode utilisée est la méthode de Ward.

Les résultats obtenus<sup>14</sup> font apparaître

- des groupes de tailles assez inégales
- des profils plus ou moins bien différenciés faisant apparaître les facteurs les plus importants de la séparation des comportements de consommation, mais les distinctions entre les groupes semblent être davantage d'ordre quantitatif, et par conséquent moins riches d'enseignements sur des types de consommations.

On peut constater principalement que

- sur les 10 classes, quatre ont un effectif supérieur 1000 voire à 2000 individus, les autres évoluant plutôt entre 200 et 400 ménages;
- sur ces quatre cellules importantes, trois (les groupes 1, 4 et 6) ont un profil peu différent de celui de la population moyenne, avec simplement une composante un peu plus élevée (alimentation à domicile pour les deux premiers, logement pour le dernier);
- pour les sept autres on trouve deux facteurs majeurs:
  - . le logement qui est la caractéristique majeure pour deux classes (2 et 7) dont on ne voit pas clairement ce qui les différencie;
  - . l'achat de véhicules: pour deux classes (3 et 5), le premier marquant une composante logement en retrait;
  - . les trois derniers ont chacun une caractéristique bien marquée: le groupe 9 avec les loisirs, le groupe 10 avec les dons et le groupe 8 avec l'alimentation à domicile.

Ainsi les groupes bien repérés sont en général ceux dont l'effectif est relativement faible alors que les groupes plus nombreux sont peu différenciés entre eux ou très proches du profil moyen de la population entière.

Par ailleurs, certains comportements spécifiques, notamment ceux qui sont marqués par une importance relativement forte des dépenses de sécurité ou d'éducation, ne ressortent pas de ce regroupement.

#### 3.4.2 Analyse du regroupement des 64 classes en 10 groupes

On peut appliquer à la classification obtenue ici le même traitement analytique que ce qui a été réalisé pour la classification neuronale.

On obtient là encore des résultats moins explicites, ce qui est une conséquence d'une classification en classes moins nettement séparés que ceux qui ont été obtenus à partir des classes Kohonen.

---

<sup>13</sup> Analyse de « clusters ».

<sup>14</sup> Voir annexe V.

Les variables utilisées pour l'analyse sont toujours significatives mais le coefficient de corrélation multiple est inférieur à celui qui a été obtenu précédemment. Les corrélations canoniques ne font apparaître que trois variables conservées par le test du rapport des vraisemblances. Leur interprétation est un peu plus délicate dans la mesure où chacune tend à combiner, de façon différente, à peu près toutes les variables d'origine.

#### 4. Conclusion

Un premier résultat obtenu est la pertinence du regroupement obtenu à l'aide de la méthode neuronale, dans la mise en évidence de comportements de consommation spécifiques pour des groupes dont la constitution résulte en fait d'une combinaison complexe de variables explicatives, vraisemblablement différente de celle qui peut être obtenue par le simple croisement de variables qualitatives.

L'originalité des regroupements que nous avons opérés tient au fait qu'ils ont été définis directement par les profils de consommation, et non par la similarité des variables socio-économiques caractérisant les ménages. Il est donc essentiel que l'on compare ce regroupement à ceux qu'entraîneraient des méthodes plus traditionnelles.

Un développement possible pour approfondir le rôle de ces caractéristiques est l'analyse, à l'aide de modèles qualitatifs, de leur influence marginale sur la probabilité de se trouver dans l'un des groupes, c'est à dire de posséder un profil de consommation donné, selon les caractéristiques socio-démographiques possédées.

Une autre utilisation serait la mesure transversale des comportements de consommation à l'aide de ces groupements opérés sur des enquêtes successives, puisque l'on a constaté que les classes obtenues fournissaient une représentation correcte des deux années regroupées. Ce cellulaire a l'avantage de réduire la variance intra des cellules, puisqu'elles sont plus homogènes que celles que l'on peut définir par un cellulaire classique à la Deaton (dans la mesure où le regroupement des individus s'opère par la similarité de leurs profils de consommation). Le problème est de savoir si les types ainsi définis sont stables et peuvent donc être comparés entre deux enquêtes (le suivi de cohortes selon la méthode de Deaton ne posant pas ce problème puisque les cellules sont constituées par des tirages indépendants dans la même population mère en deux périodes successives). Le fait d'effectuer le cellulaire neuronal sur l'ensemble des enquêtes qui seront analysées longitudinalement permet d'assurer cette comparabilité des cellules entre les enquêtes.

## Bibliographie

- F. Blayo & M. Verleysen *Les réseaux de neurones artificiels*, PUF, 1996
- M. Cottrell & P. Letrémy « *Classification et analyse des correspondances au moyen de l'algorithme de Kohonen: application à l'étude de données socio-économiques* », prépublication du SAMOS, Université de Paris I, 1995.
- M. Cottrell & I. Ibbou « *Multiple Correspondence Analysis of a crosstabulations matrix using the Kohonen algorithm* », prépublication du SAMOS, Université de Paris I, 1995.
- F. Gardes, P. Gaubert et S. Langlois « *Pauvreté et convergence des structures de consommation au Canada* » à paraître dans *The Canadian Review of Sociology and Anthropology*.
- F. Gardes, S. Langlois et D. Richaudeau « *Cross-sections versus Time Series Elasticities of Canadian Consumption* » *Economics Letters*, juin 1996.
- T. Kohonen *Self-Organization and Associative Memory*, Springer-Verlag, 1989.

## ANNEXES

I Nomenclature des fonctions de consommations

II Profils des groupes obtenus à partir des classes Kohonen

III Profils et composition des classes selon quelques variables socio-économiques

IV Analyse canonique discriminante sur les classes Kohonen

V Résultats de la classification classique



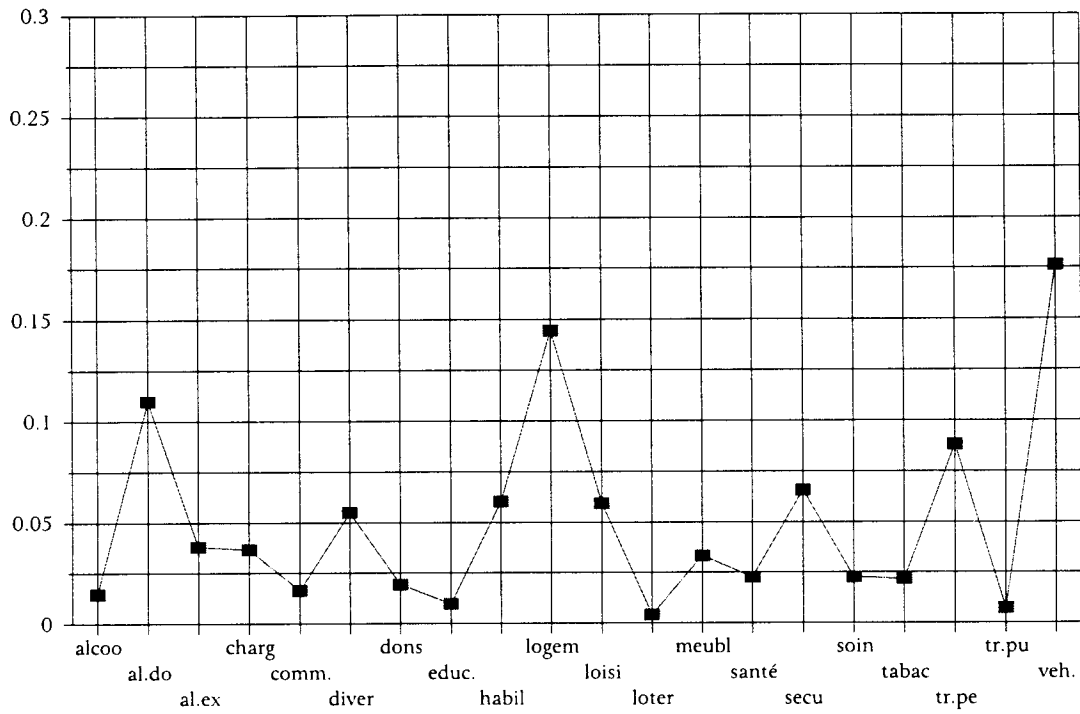
## ANNEXE I

Nomenclature des fonctions de consommation  
utilisée dans les enquêtes canadiennes 1986 et 1992

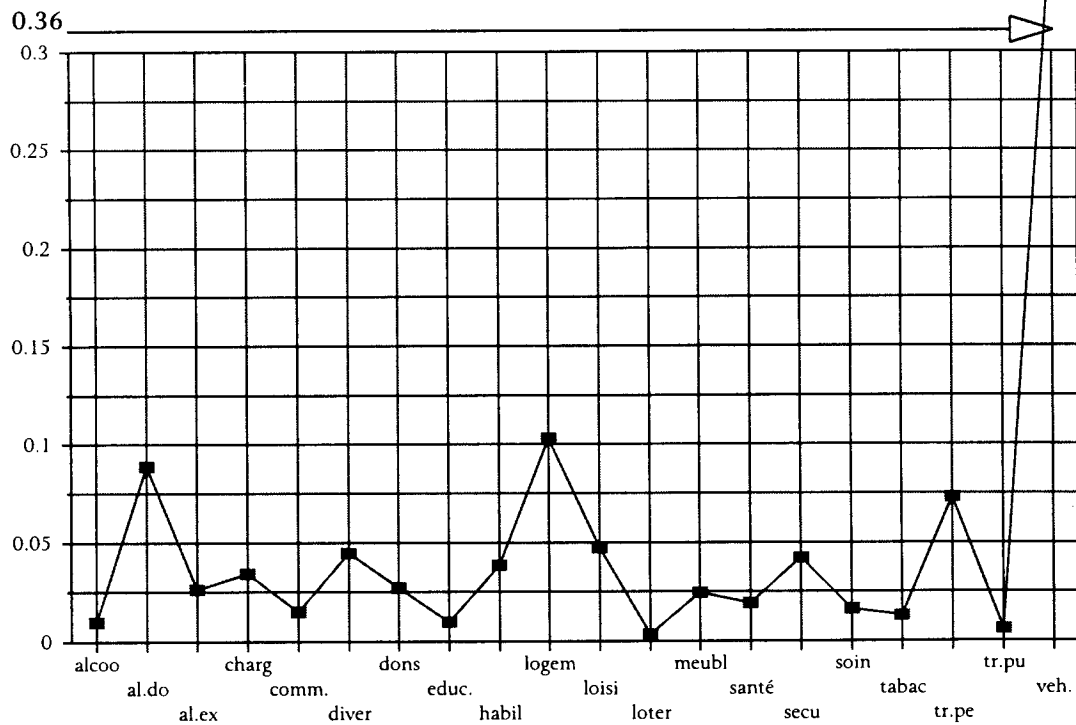
Fonction	Description du contenu
Alcool	Boissons servies dans les établissements licenciés et achetées en magasin
Alimentation domicile	à Aliments achetés dans un magasin, sauf les biens isolés comme une fonction distincte
Alimentation l'extérieur	à Aliments achetés au restaurant
Charges	Charges liées à l'entretien ménager (gardes d'enfants, animaux domestiques, produits d'entretien)
Communication	Dépenses de téléphone
Divers	Intérêts des prêts personnels, cotisations
Dons	Dons à des personnes n'appartenant pas au ménage et à des institutions
Education	Frais de scolarité
Habillement	Acquisitions et dépenses liées
Logement	Loyers et entretien, fonctionnement
Loisirs	Acquisitions et utilisations de matériels liés aux loisirs, spectacles
Loterie	
Meubles	
Santé	Frais directs (produits médicaux et soins) et assurance médicale
Sécurité	Assurance-vie, assurance chômage, cotisations aux caisses de retraite
Soins	Produits de toilette et soins personnels
Tabac	
Transport personnel	Dépenses liées à l'utilisation de véhicules personnels
Transport public	
Achats de véhicules	

ANNEXE II  
PROFILS DES GROUPES - 1992

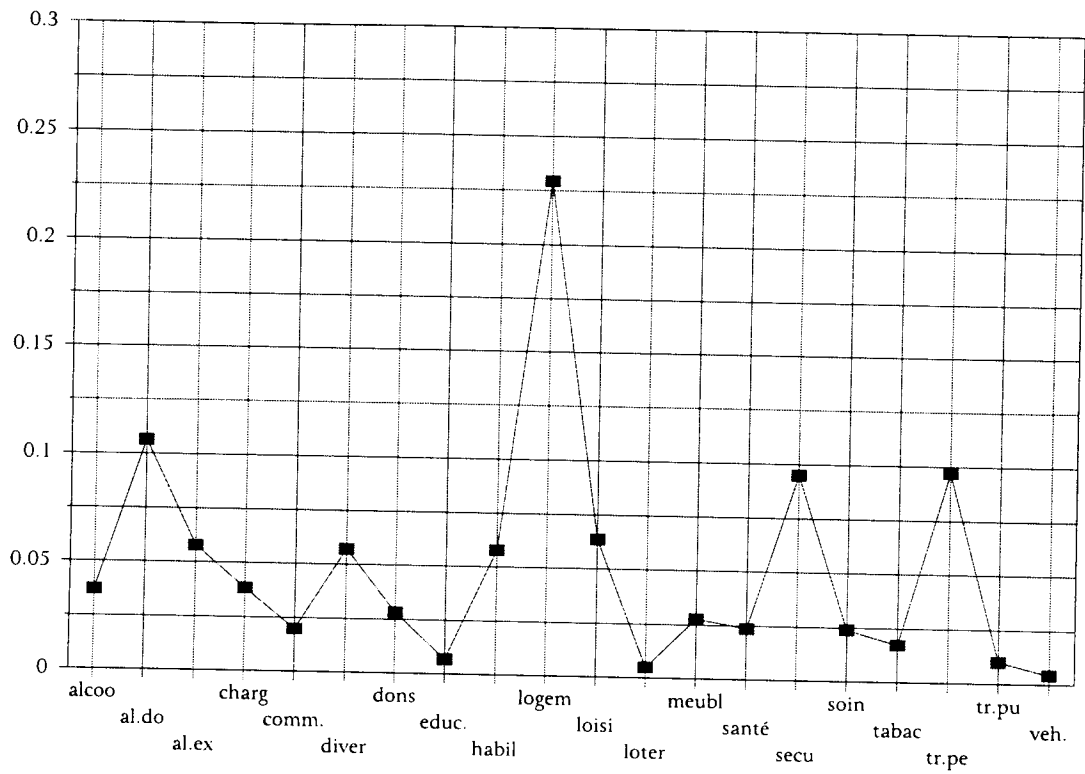
Groupe 1



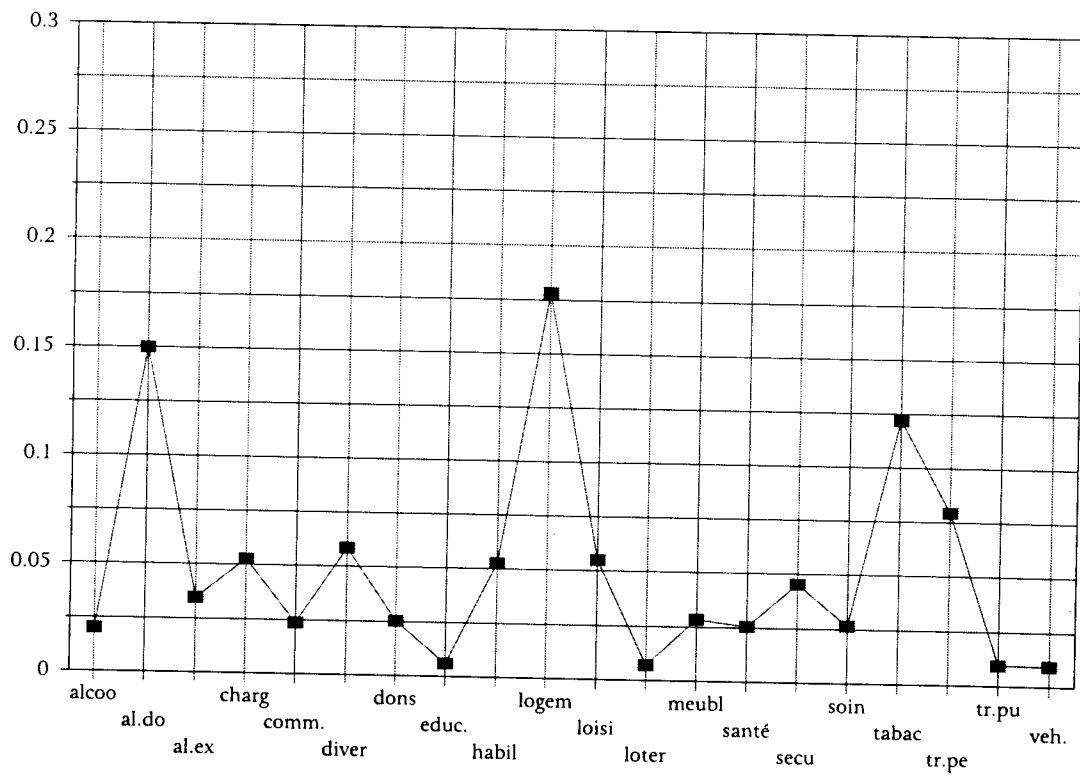
Groupe 2



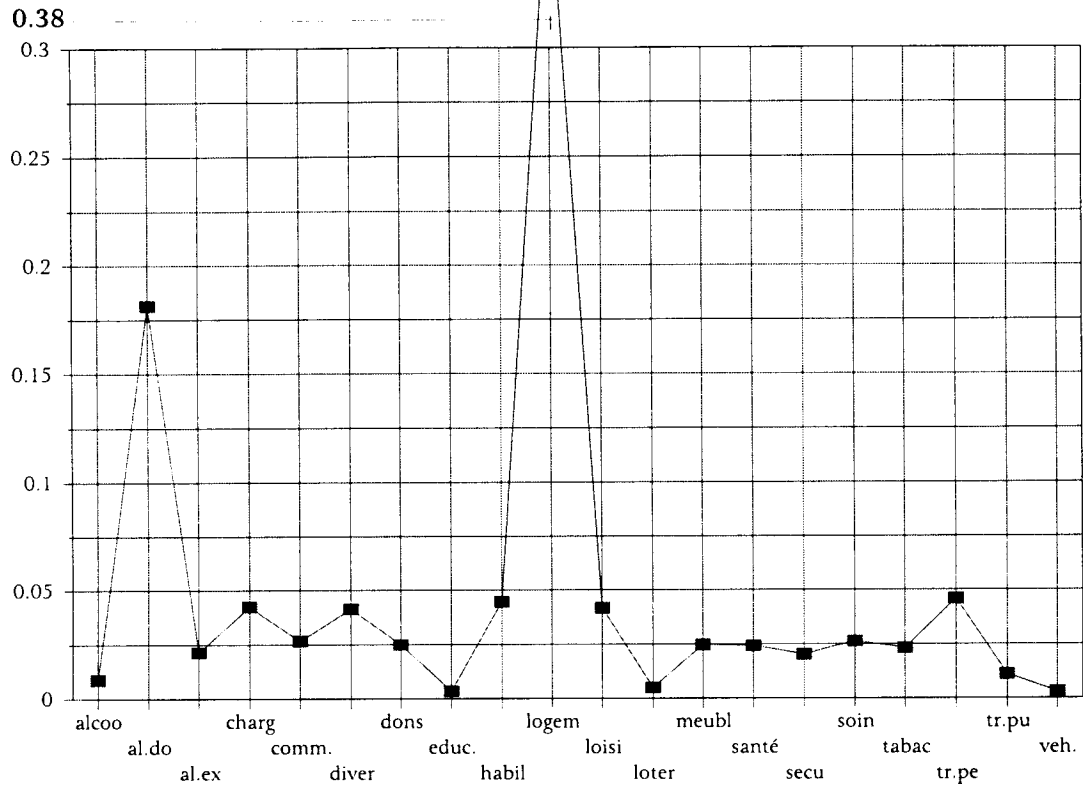
### Groupe 3



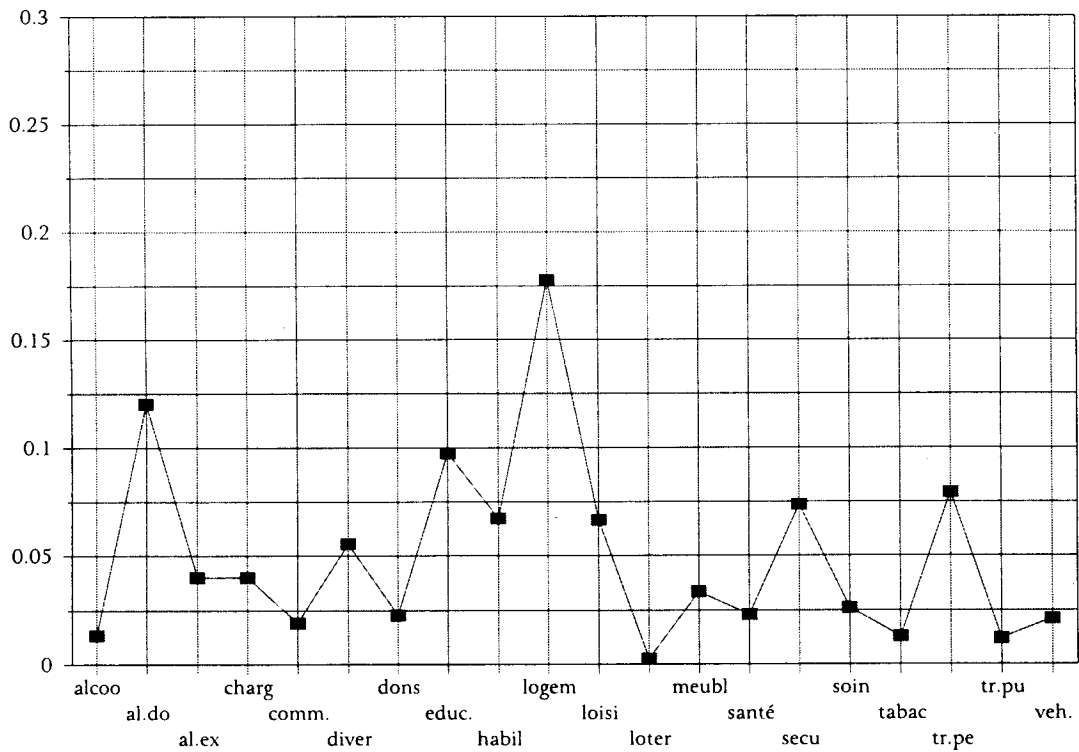
### Groupe 4



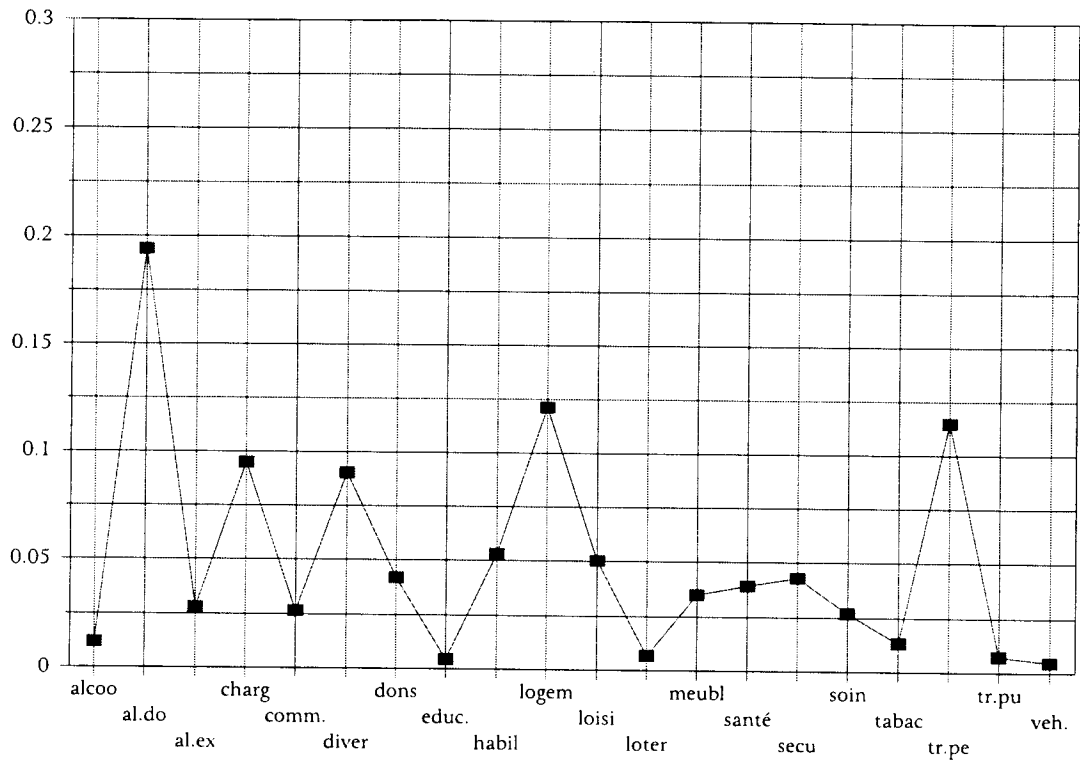
### Groupe 5



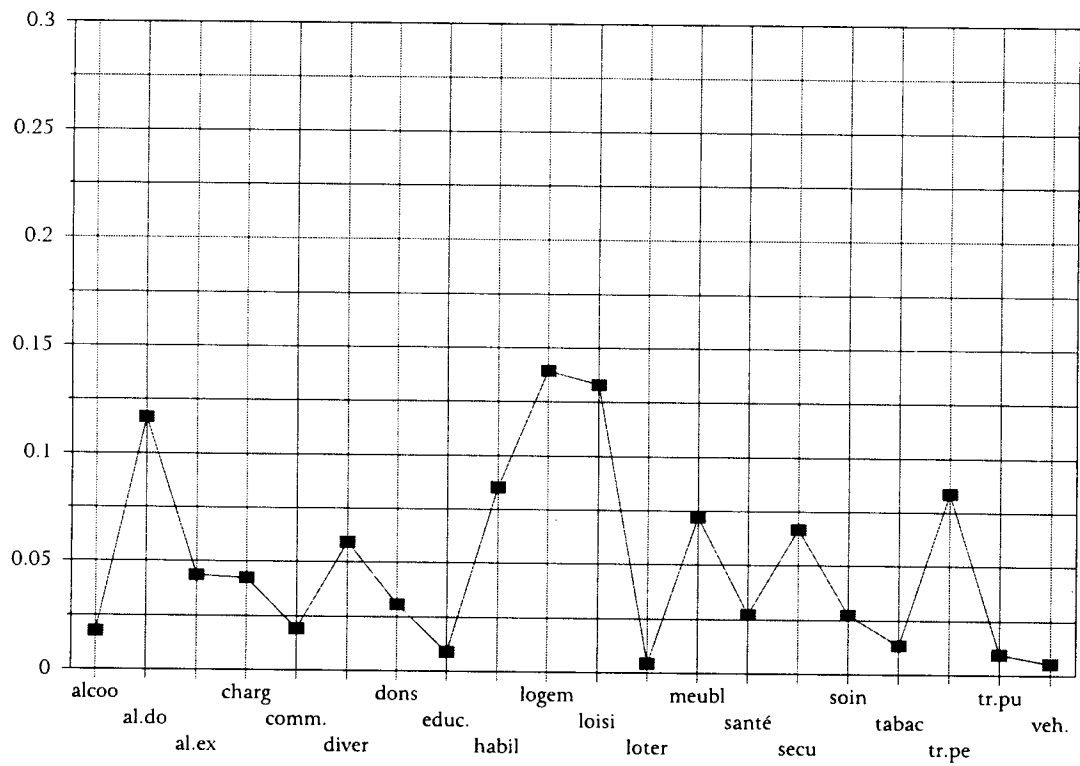
### Groupe 6



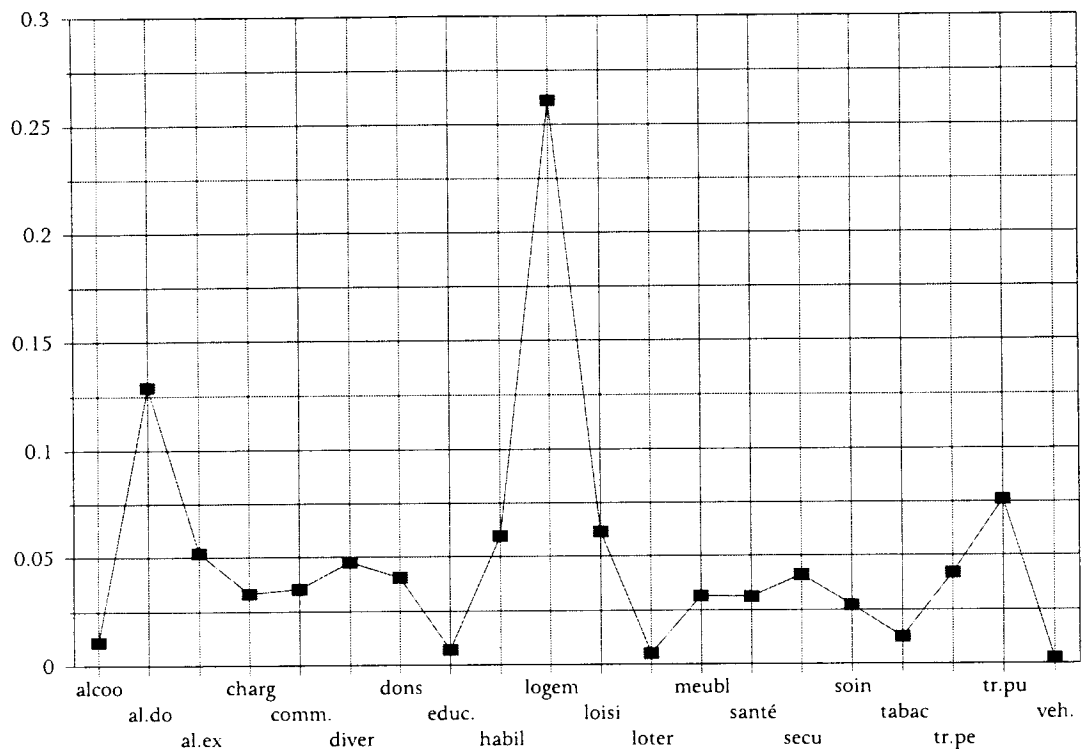
### Groupe 7



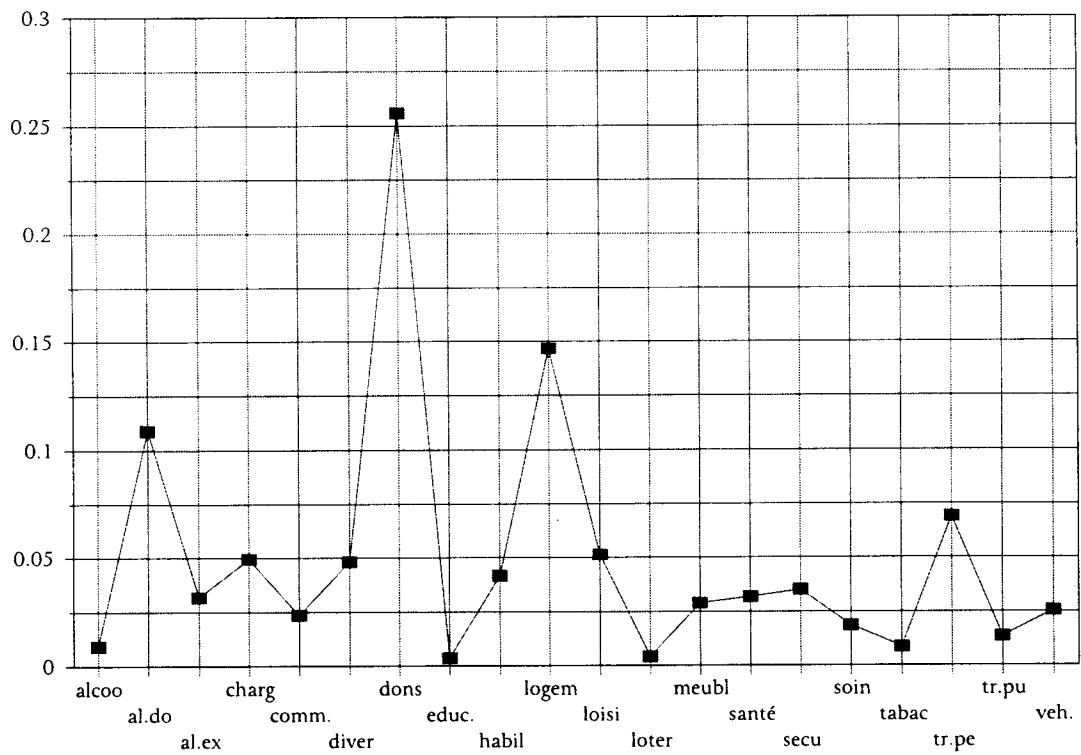
### Groupe 8



### Groupe 9



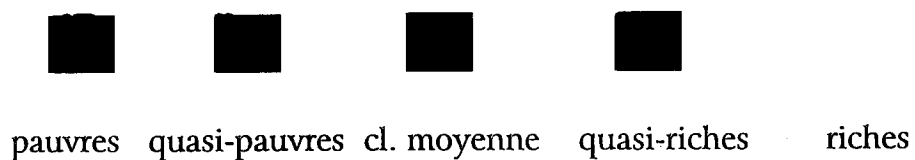
### Groupe 10



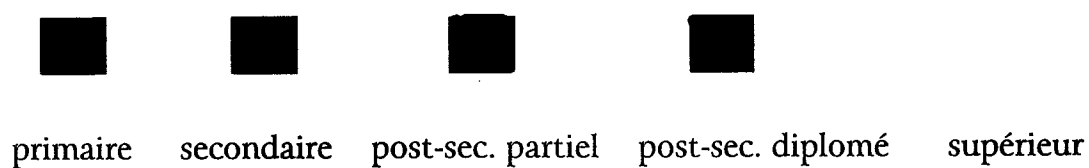
ANNEXE III  
PROFILS ET COMPOSITION DES CLASSES



### Critère de pauvreté-richesse



### Niveaux d'éducation

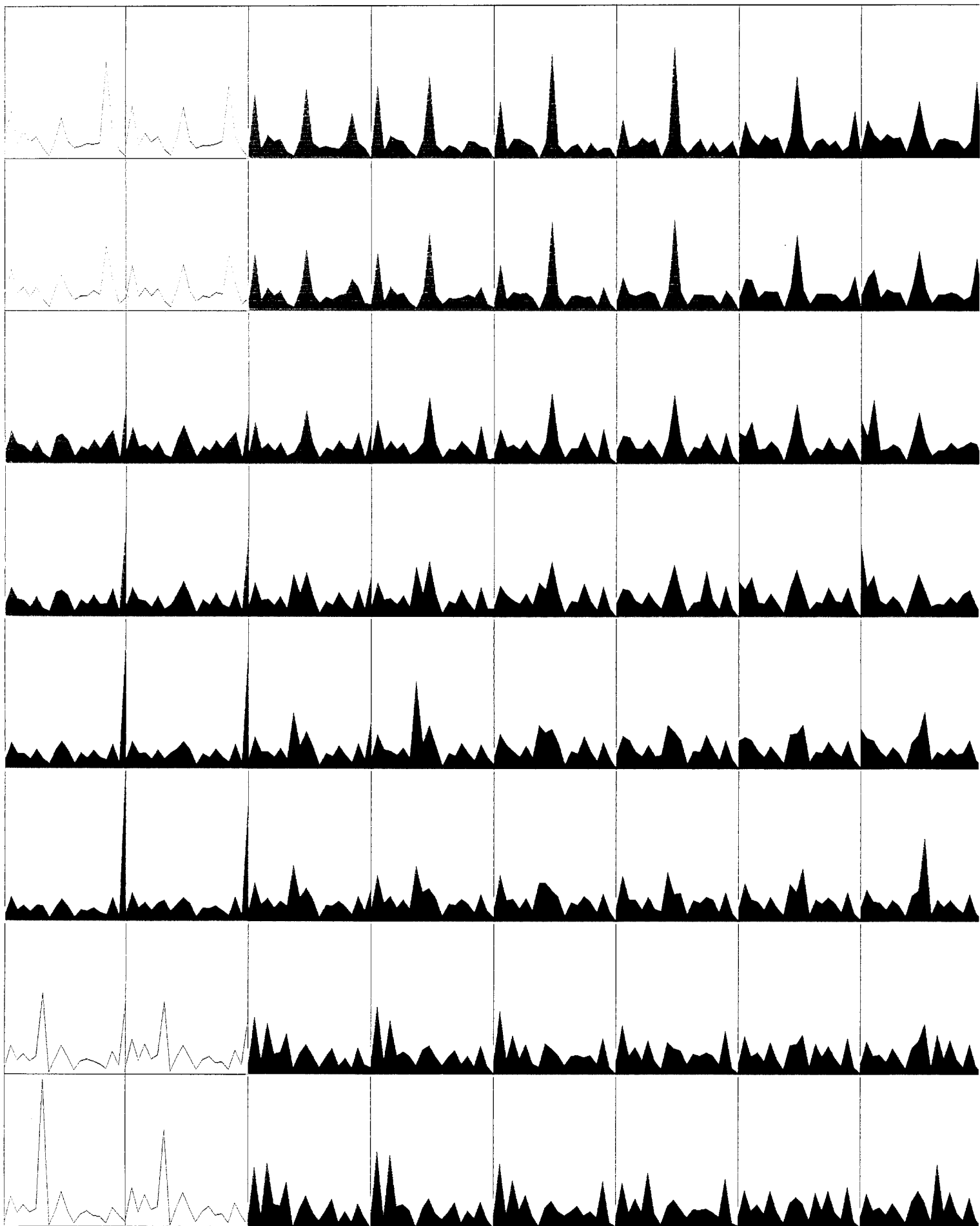


### Tranches d'âge



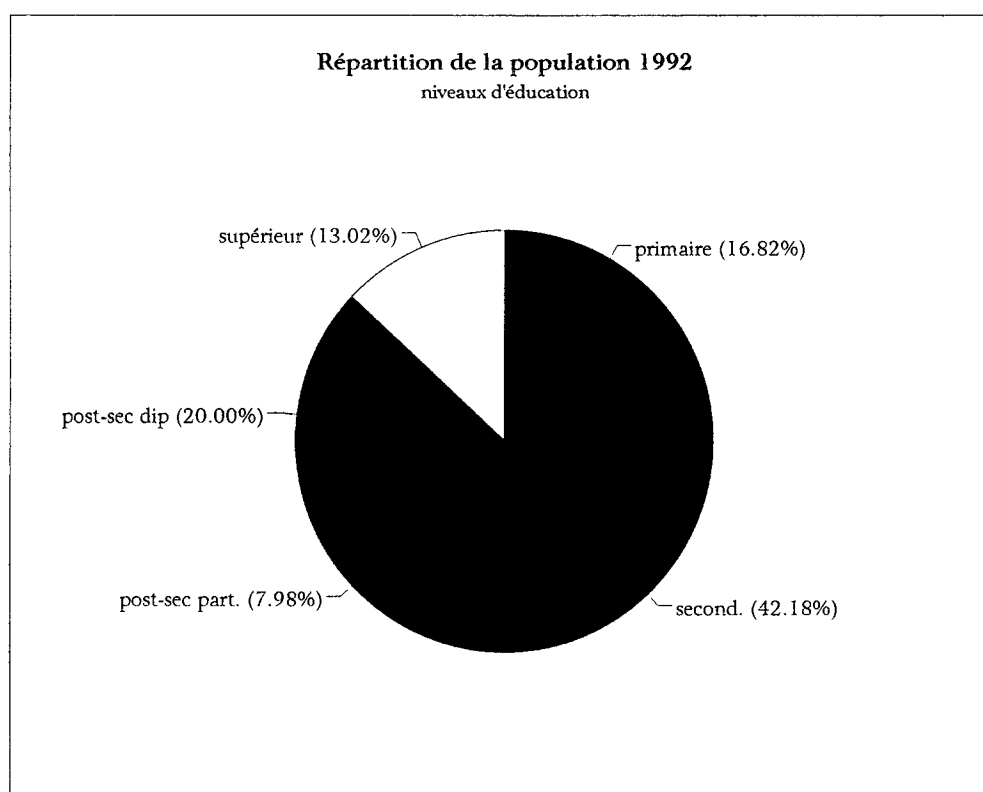
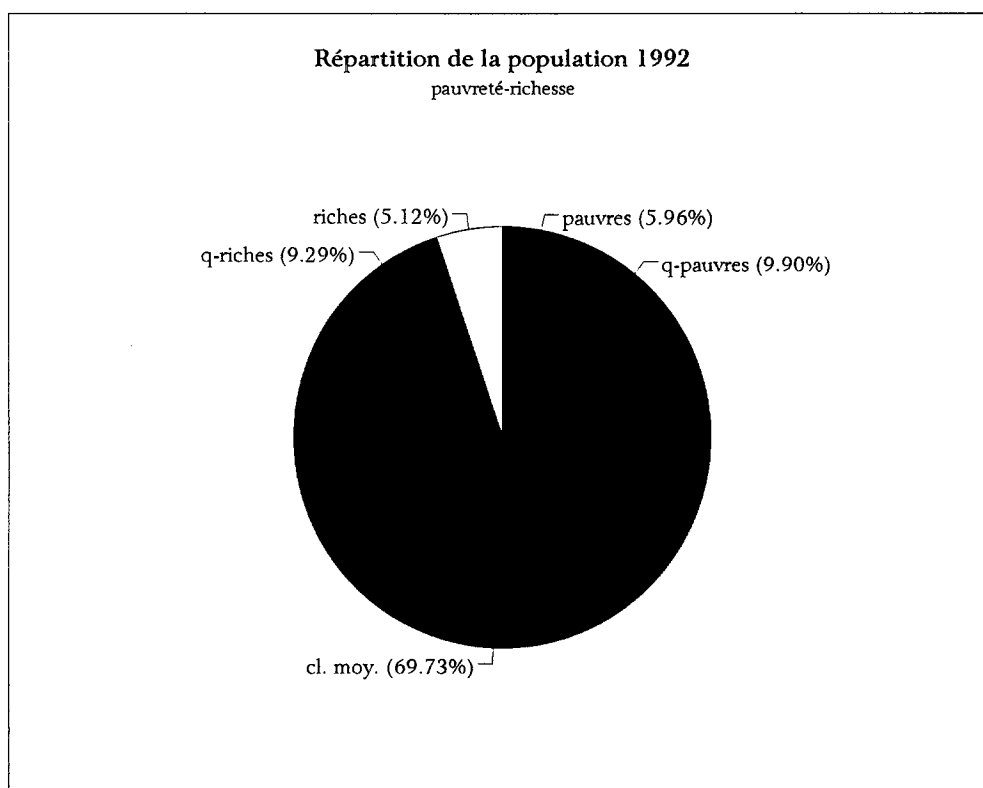
### Détention du logement



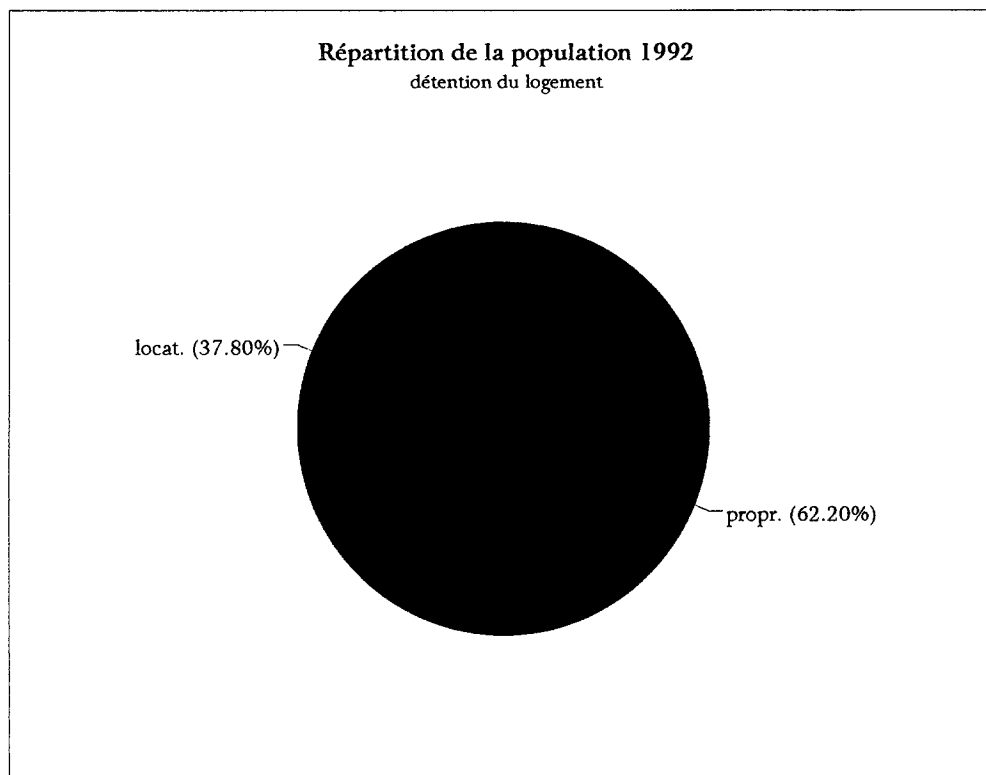
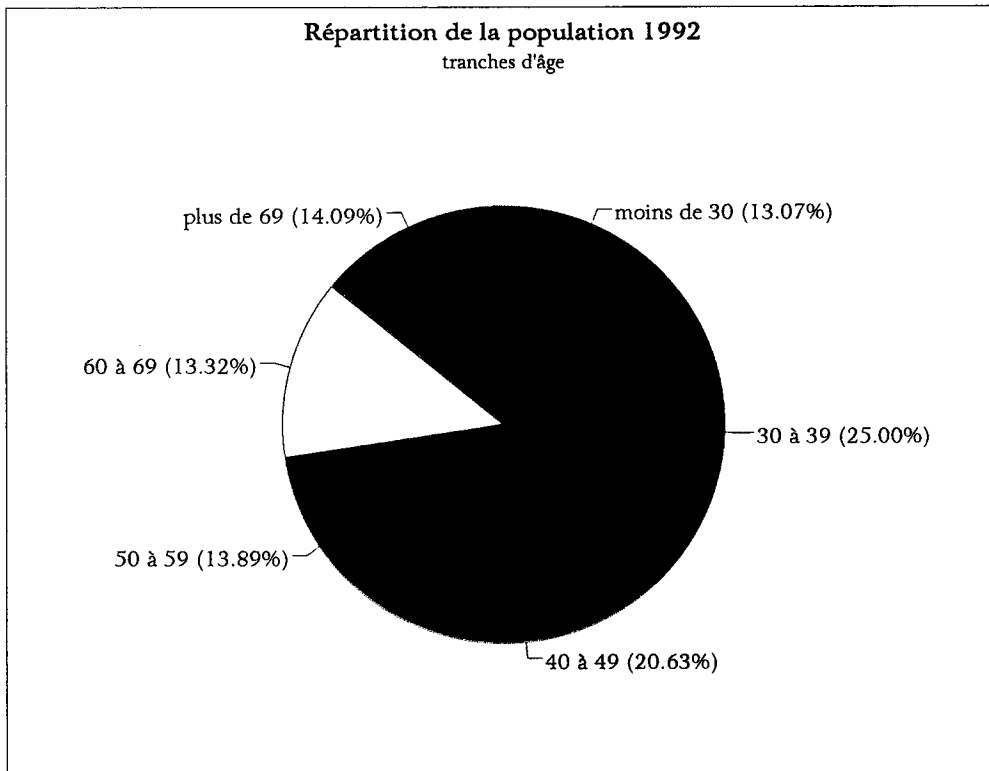


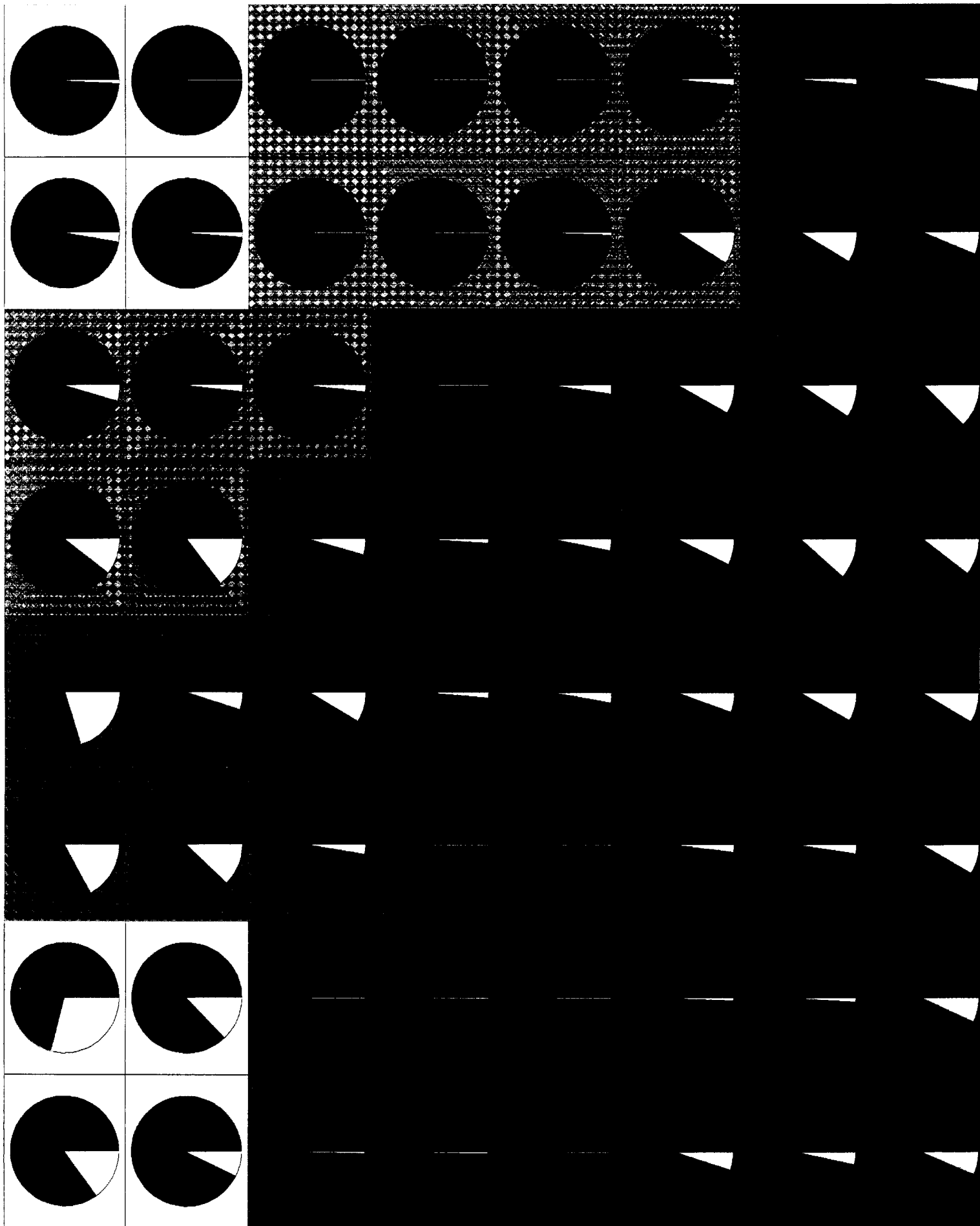
PROFILS DE CONSOMMATION DES CLASSES KOHONEN

## ENSEMBLE DE LA POPULATION

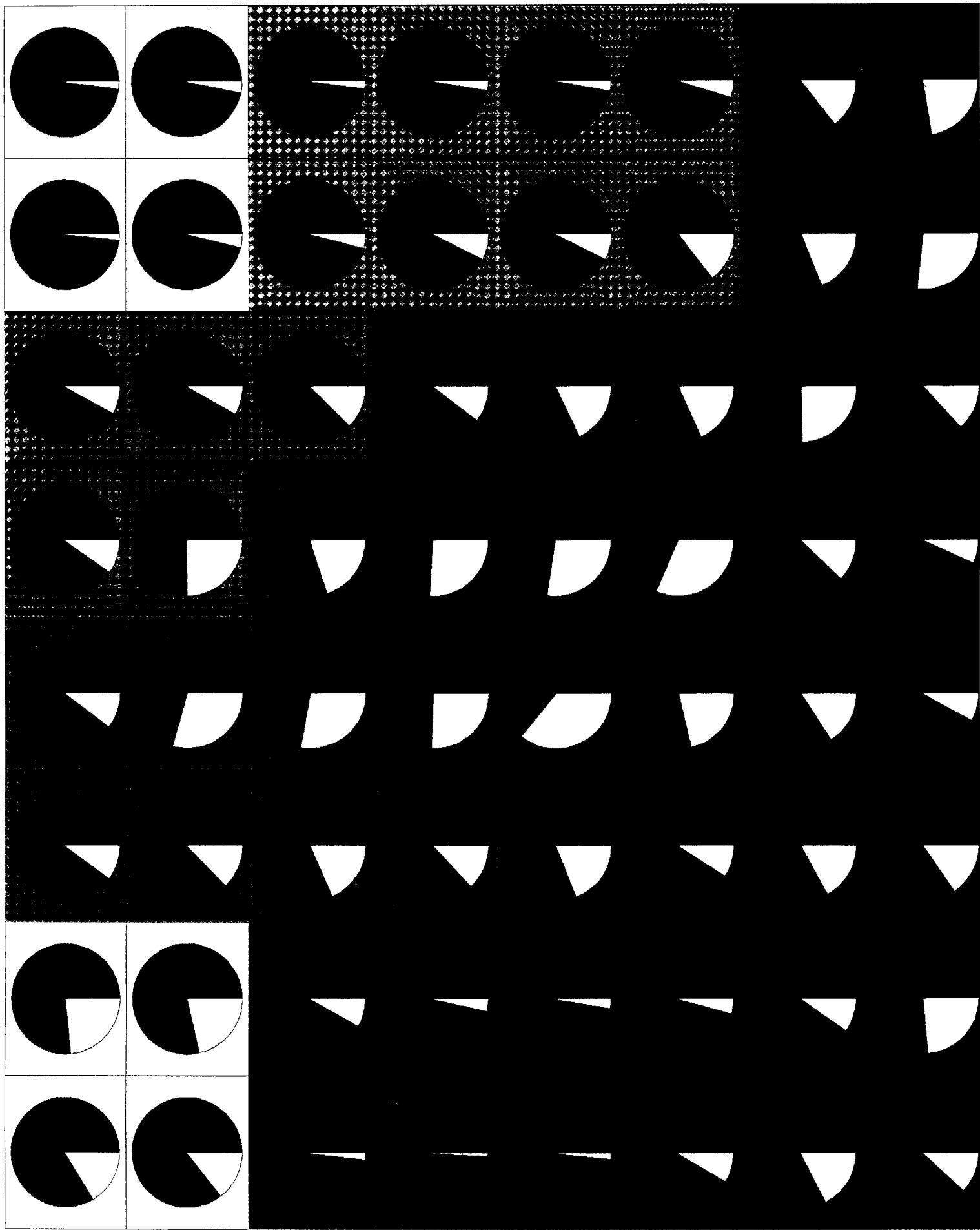


## ENSEMBLE DE LA POPULATION

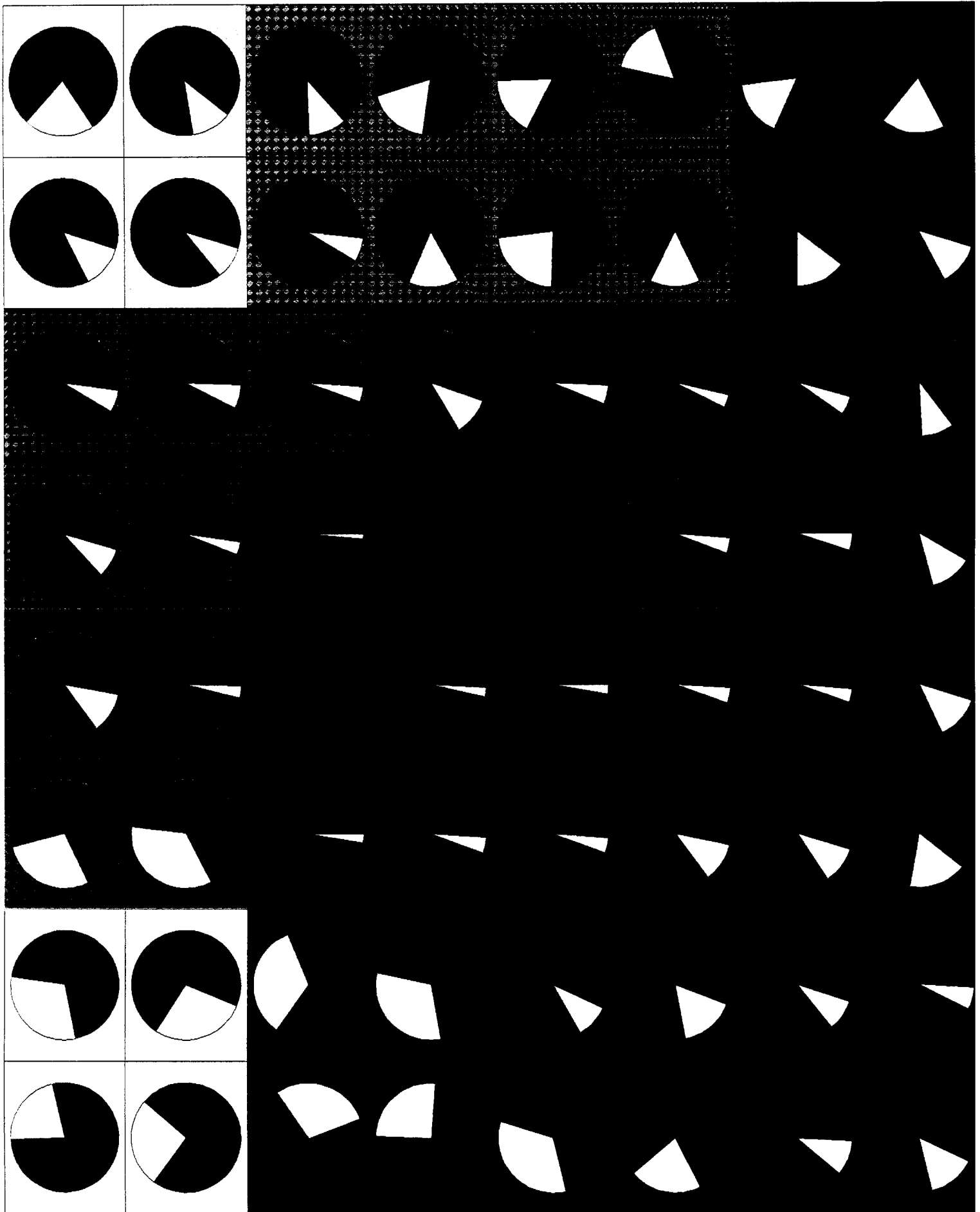




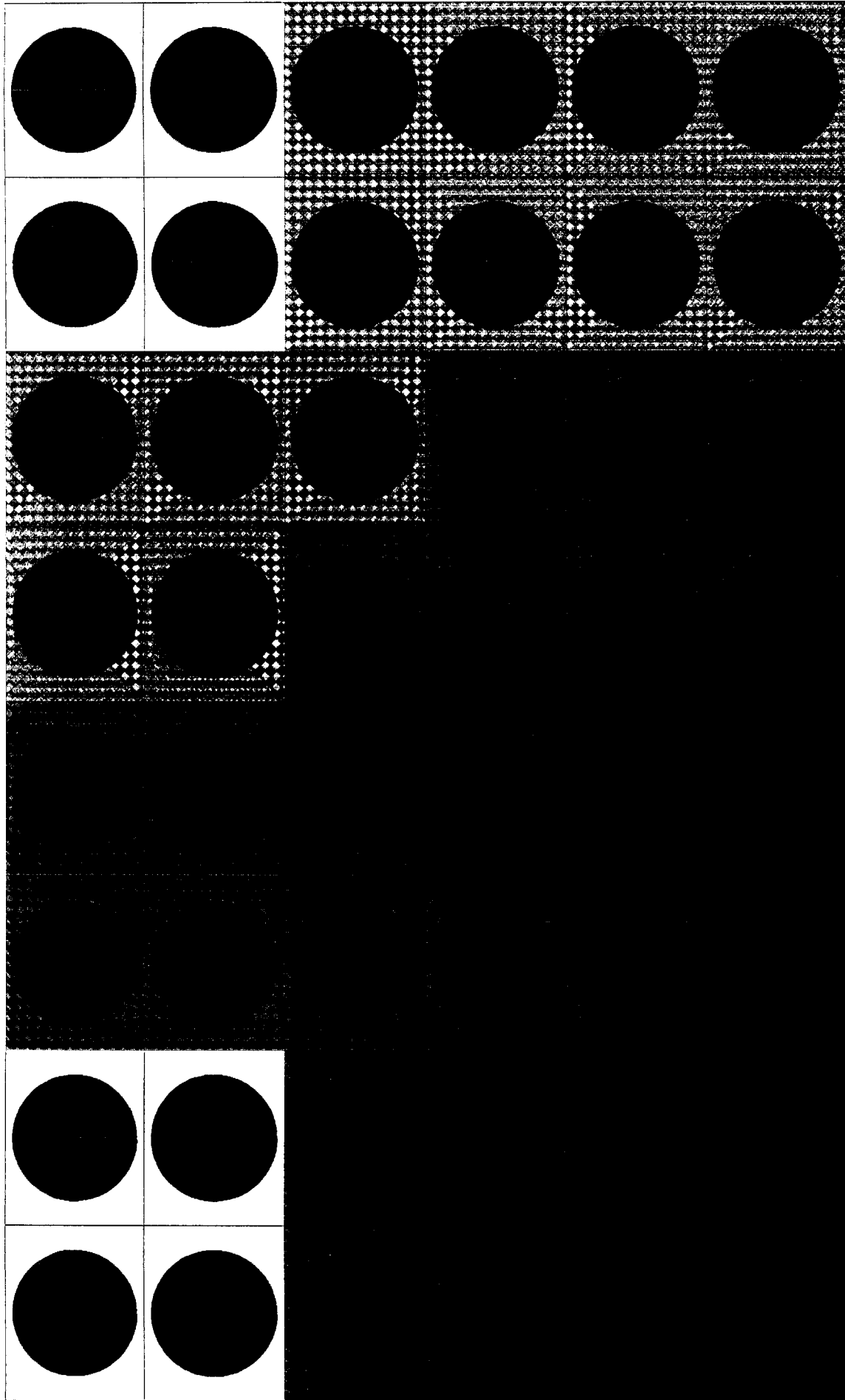
CRITERE PAUVRETE-RICHESSE



NIVEAUX D'EDUCATION



TRANCHES D'AGE



TYPE DE DETENTION DU LOGEMENT



TYPE DE DETENTION DU LOGEMENT  
ANNEXE IV

Analyse canonique discriminante des classes Kohonen

A. Statistiques descriptives sur les 10 groupes de classes

Table 1  
Fréquences en %

Variable	Modal.	1	2	3	4	5	6	7	8	9	10	Ens.
pauvreté	pauvres	1.8	0.2	0.9	6.6	18.1	3.2	12.7	1.8	4.5	0.6	6.0
	q-pauvres	4.4	1.1	5.0	14.3	19.3	9.1	15.8	6.2	11.7	2.8	9.9
	cl. moy.	73.0	65.9	70.0	72.8	58.3	79.7	66.7	73.2	72.5	69.5	69.7
	q-riches	12.9	19.0	16.6	4.9	2.3	5.8	3.5	12.5	5.8	15.4	9.3
	riches	7.9	13.8	7.4	1.5	2.0	2.2	1.4	6.3	5.4	11.8	5.1
âge	<29 ans	14.9	12.0	19.6	13.8	12.2	18.9	4.9	14.6	16.2	2.4	13.1
	30-39 ans	31.9	23.7	33.2	24.2	22.2	40.1	14.2	28.6	15.8	7.5	25.0
	40-49 ans	28.2	16.6	22.2	21.4	15.3	26.8	16.9	25.8	18.8	11.6	20.6
	50-59 ans	16.8	13.8	14.6	17.3	10.5	11.2	14.7	13.5	14.1	12.9	13.9
	60-69 ans	5.3	21.0	6.9	13.5	13.5	2.4	22.5	11.3	14.9	26.8	13.3
	>69 ans	2.9	12.8	3.5	10.0	26.2	0.5	26.8	6.3	20.1	38.8	14.1
éducation	primaire	9.1	17.7	6.6	23.2	25.2	4.6	29.8	9.7	16.5	24.3	16.8
	secondaire	45.8	41.1	37.9	53.3	48.8	30.6	44.8	40.3	36.4	38.4	42.2
	post-sec.	6.8	9.3	9.2	7.7	7.7	10.3	5.8	9.3	7.6	5.1	8.0
	post-sec.dip	22.5	18.4	26.6	13.3	12.6	30.4	13.7	25.9	18.4	16.9	20.0
	supérieur	15.9	13.5	19.6	10.0	5.8	24.1	5.9	14.7	21.2	15.4	13.0
région	Iles-N.Bruns.	22.2	20.3	18.2	26.6	18.1	26.5	33.2	20.3	13.9	15.5	22.5
	Québec	23.1	22.1	22.1	21.0	24.6	18.1	19.2	21.3	16.9	10.9	20.5
	Ontario	22.0	20.7	26.3	21.3	25.4	22.2	18.0	20.8	32.0	22.7	22.7
	Man.Sas.Al.	23.1	26.0	21.9	22.7	18.6	25.8	22.1	28.1	21.4	36.5	24.0
	Col.Brit.	9.5	10.9	11.5	8.5	13.2	7.5	7.5	9.6	15.8	14.4	10.3
logement	locataires	25.4	27.6	44.7	48.1	70.5	31.3	13.9	30.6	63.6	27.9	37.9
	propriétaires	74.6	72.4	55.3	51.9	29.5	68.7	86.1	69.4	36.4	72.1	62.2

Moyennes des variables quantitatives

Variabes	1	2	3	4	5	6	7	8	9	10	Ens.
revenu (milliers \$)	47.2	38.3	40.3	28.2	20.2	46.1	28.9	44.3	28.2	35.5	35.4
dépense totale par tête	21.3	24.9	20.5	13.9	12.6	17.5	12.8	19.7	17.2	21.6	17.5
âge	42.1	49.2	41.3	46.7	51.6	38.5	56.2	44.2	50.2	61.7	47.7
taille ménage(Oxford)	2.54	2.09	1.91	2.04	1.67	2.61	2.03	2.27	1.63	1.64	2.06
nombre d'enfants	3.24	2.59	2.34	2.51	2.07	3.52	2.47	2.86	1.90	1.83	2.56
effectif du groupe	618	615	1310	944	1124	925	1398	1180	462	534	9132

## B. Principaux résultats de l'analyse canonique discriminante

### Univariate Test Statistics

F Statistics, Num DF= 9 Den DF= 54							
Variable	Total STD	Pooled STD	Between STD	R-Squared	RSQ/ (1-RSQ)	F	Pr > F
MAGE	8.0831	5.4109	6.6343	0.615905	1.6035	9.6211	0.0001
MUNI	0.4574	0.3098	0.3727	0.606789	1.5432	9.2590	0.0001
MNEDU1	0.1182	0.0866	0.0908	0.539518	1.1716	7.0298	0.0001
MNEDU3	0.0367	0.0344	0.0192	0.249667	0.3327	1.9964	0.0576
MPROP	0.2114	0.1083	0.1946	0.775051	3.4455	20.6728	0.0001
MDTOT	12104	6945	10725	0.717836	2.5440	15.2642	0.0001
MP1	0.0986	0.0807	0.0674	0.426376	0.7433	4.4598	0.0002
MP2	0.0869	0.0653	0.0653	0.515593	1.0644	6.3863	0.0001
MP4	0.0755	0.0517	0.0611	0.598310	1.4895	8.9369	0.0001
MP5	0.0601	0.0485	0.0419	0.443277	0.7962	4.7773	0.0001

Average R-Squared: Unweighted = 0.5488321 Weighted by Variance = 0.7178357

### Multivariate Statistics and F Approximations

S=9 M=0 N=21.5

Statistic	Value	F	Num DF	Den DF	Pr > F
Wilks' Lambda	0.00493154	4.1660	90	315.4635	0.0001
Pillai's Trace	3.21292506	2.9425	90	477	0.0001
Hotelling-Lawley Trace	10.61990644	5.1002	90	389	0.0001
Roy's Greatest Root	4.65862215	24.6907	10	53	0.0001

### Canonical Discriminant Analysis

				Eigenvalues of INV(E)*H = CanRsq/(1-CanRsq)				
Canonical Correl.	Adjusted Correl.	Approx Standard Error	Squared Canonical Correl.	Eigenval.	Diff.	Propor.	Cumul.	
1	0.907347	0.875750	0.022265	0.823279	4.6586	2.2901	0.4387	0.4387
2	0.838533	0.780966	0.037401	0.703137	2.3686	0.6384	0.2230	0.6617
3	0.796070	0.752457	0.046146	0.633727	1.7302	0.4689	0.1629	0.8246
4	0.746843	0.731775	0.055715	0.557775	1.2613	0.9662	0.1188	0.9434
5	0.477356	0.315584	0.097279	0.227869	0.2951	0.0917	0.0278	0.9712
6	0.411175	.	0.104688	0.169065	0.2035	0.1421	0.0192	0.9903
7	0.240443	0.062506	0.118704	0.057813	0.0614	0.0306	0.0058	0.9961
8	0.172717	.	0.122230	0.029831	0.0307	0.0202	0.0029	0.9990
9	0.102125	.	0.124674	0.010430	0.0105	.	0.0010	1.0000

Test of H0: The canonical correlations in the current row and all that follow are zero

	Likelihood Ratio	Approx F	Num DF	Den DF	Pr > F
1	0.00493154	4.1660	90	315.4635	0.0001
2	0.02790573	3.1974	72	287.3864	0.0001
3	0.09400219	2.5438	56	258.4137	0.0001
4	0.25664496	1.8308	42	228.592	0.0028
5	0.58034909	0.9617	30	198	0.5285
6	0.75161964	0.7497	20	166.7811	0.7695
7	0.90454705	0.4355	12	135.2248	0.9466
8	0.96005047	0.3570	6	104	0.9044
9	0.98957041	0.2793	2	53	0.7574

Total Canonical Structure

	CAN1	CAN2	CAN3	CAN4
MAGE	-0.039681	0.854238	-0.290161	-0.248732
MUNI	0.603942	-0.415042	-0.081647	0.548646
MNEDU1	-0.200609	0.755204	-0.156249	0.196413
MNEDU3	-0.059556	-0.482784	0.207449	0.077713
MPROP	0.921842	0.213303	0.164244	0.182340
MDTOT	0.674217	-0.658853	-0.087927	-0.206974
MP1	-0.327767	0.486971	-0.175420	0.404151
MP2	-0.446772	0.469934	-0.084262	0.550404
MP4	0.322157	-0.371642	0.294509	-0.798552
MP5	0.182706	-0.170112	-0.036637	-0.832734

Total-Sample Standardized Canonical Coefficients

	CAN1	CAN2	CAN3	CAN4
MAGE	0.0816454	0.8235801	-2.1655775	-1.0350914
MUNI	0.1354477	-0.0610556	-0.5320657	0.0988416
MNEDU1	0.8318723	-0.4708264	0.0365936	0.8480933
MNEDU3	-0.1750613	0.0581225	-0.1272292	0.0785463
MPROP	1.3699400	1.0899432	1.7097099	0.3405969
MDTOT	1.4157367	-1.5245145	-2.6465152	0.0123102
MP1	-0.0491558	0.7141572	-0.5017009	-0.2034585
MP2	0.1003762	-0.5020609	0.2339749	-0.0948655
MP4	0.3305922	-0.2212004	1.0725595	-0.9641742
MP5	-0.0817183	0.4526023	0.2078257	-0.2555235

Class Means on Canonical Variables

CLUS	CAN1	CAN2	CAN3	CAN4
1	2.037137	-1.563278	-0.615233	0.149518
2	2.409510	-0.575743	-0.662051	-1.376030
3	-1.203272	-0.768546	2.412103	-1.034863
4	-1.524319	-0.375137	0.487224	1.023402
5	-3.432353	0.496635	-1.274220	0.268602
6	1.245556	-1.315249	-0.716495	1.299391
7	1.159204	2.911310	0.773070	0.859457
8	1.005777	-0.640147	0.540809	-0.059571
9	-3.431471	-0.362669	-0.796341	-0.583073
10	1.597402	1.660662	-1.879670	-2.442049

## ANNEXE V

### Résultats de la classification classique

#### Canonical Discriminant Analysis

##### Univariate Test Statistics

F Statistics, Num DF= 9 Den DF= 54

Variable	Total STD	Pooled STD	Between STD	R-Squared	RSQ/ (1-RSQ)	F	Pr > F
MAGE	8.8977	6.7329	6.6402	0.509202	1.0375	6.2250	0.0001
MUNI	0.4272	0.3447	0.2970	0.441878	0.7917	4.7503	0.0001
MNEDU1	0.1487	0.1052	0.1175	0.570649	1.3291	7.9746	0.0001
MNEDU3	0.0316	0.0280	0.0188	0.324267	0.4799	2.8792	0.0075
MPROP	0.2542	0.1534	0.2204	0.687649	2.2015	13.2092	0.0001
MDTOT	12408	7588	10696	0.679385	2.1190	12.7140	0.0001
MP1	0.1235	0.0962	0.0895	0.480004	0.9231	5.5385	0.0001
MP2	0.0884	0.0585	0.0731	0.624703	1.6646	9.9873	0.0001
MP4	0.1012	0.0876	0.0633	0.357483	0.5564	3.3383	0.0026
MP5	0.0773	0.0664	0.0490	0.367675	0.5815	3.4888	0.0018

Average R-Squared: Unweighted = 0.5042895 Weighted by Variance = 0.6793848

##### Multivariate Statistics and F Approximations

S=9 M=0 N=21.5

Statistic	Value	F	Num DF	Den DF	Pr > F
Wilks' Lambda	0.00963496	3.4446	90	315.4635	0.0001
Pillai's Trace	2.89983366	2.5195	90	477	0.0001
Hotelling-Lawley Trace	9.03822994	4.3406	90	389	0.0001
Roy's Greatest Root	4.24152237	22.4801	10	53	0.0001

##### Canonical Discriminant Analysis

	Canonical Correlation	Adjusted Canonical Correlation	Approx Standard Error	Squared Canonical Correlation
1	0.899564	0.864885	0.024037	0.809216
2	0.849926	0.815104	0.034978	0.722374
3	0.699403	0.601664	0.064359	0.489164
4	0.632493	0.576543	0.075587	0.400047
5	0.455951	0.282344	0.099796	0.207891
6	0.396988	.	0.106132	0.157600
7	0.276741	0.192717	0.116339	0.076586
8	0.170839	0.094188	0.122311	0.029186
9	0.088149	.	0.125009	0.007770

Eigenvalues of INV(E)\*H  
 = CanRsq/(1-CanRsq)

	Eigenvalue	Difference	Proportion	Cumulative
1	4.2415	1.6396	0.4693	0.4693
2	2.6020	1.6444	0.2879	0.7572
3	0.9576	0.2908	0.1059	0.8631
4	0.6668	0.4043	0.0738	0.9369
5	0.2625	0.0754	0.0290	0.9659
6	0.1871	0.1041	0.0207	0.9866
7	0.0829	0.0529	0.0092	0.9958
8	0.0301	0.0222	0.0033	0.9991
9	0.0078	.	0.0009	1.0000

Test of H0: The canonical correlations in the current row and all that follow are zero

	Likelihood Ratio	Approx F	Num DF	Den DF	Pr > F
1	0.00963496	3.4446	90	315.4635	0.0001
2	0.05050185	2.5294	72	287.3864	0.0001
3	0.18190591	1.7179	56	258.4137	0.0027
4	0.35609463	1.3403	42	228.592	0.0923
5	0.59353737	0.9194	30	198	0.5914
6	0.74931302	0.7581	20	166.7811	0.7601
7	0.88949761	0.5099	12	135.2248	0.9055
8	0.96327049	0.3274	6	104	0.9213
9	0.99222967	0.2075	2	53	0.8132

Total Canonical Structure

	CAN1	CAN2	CAN3
MAGE	0.123522	0.608852	-0.095252
MUNI	0.434505	-0.330361	-0.065627
MNEDU1	-0.078845	0.813709	-0.244721
MNEDU3	-0.151003	-0.530154	0.304347
MPROP	0.877969	0.177212	0.183010
MDTOT	0.654375	-0.667380	-0.069616
MP1	-0.371389	0.588678	-0.313887
MP2	-0.539710	0.668871	-0.149944
MP4	0.450323	-0.389525	0.208384
MP5	0.481925	-0.337958	-0.073105

Total-Sample Standardized Canonical Coefficients

	CAN1	CAN2	CAN3
MAGE	-0.500629372	-0.710352986	-0.188867690
MUNI	1.088267978	-1.364105402	1.323587302
MNEDU1	1.170930727	0.789128905	-0.650526940
MNEDU3	-0.254265699	0.104807934	0.389410468
MPROP	1.201925118	1.593977283	0.658874562
MDTOT	-0.438541101	0.059293605	-3.936304238
MP1	-0.272765175	0.677488326	-0.496829720
MP2	-0.403945305	0.762083398	-0.348032289
MP4	0.337449693	0.147519014	2.340848516
MP5	1.085138609	-0.172802107	0.055180774

27

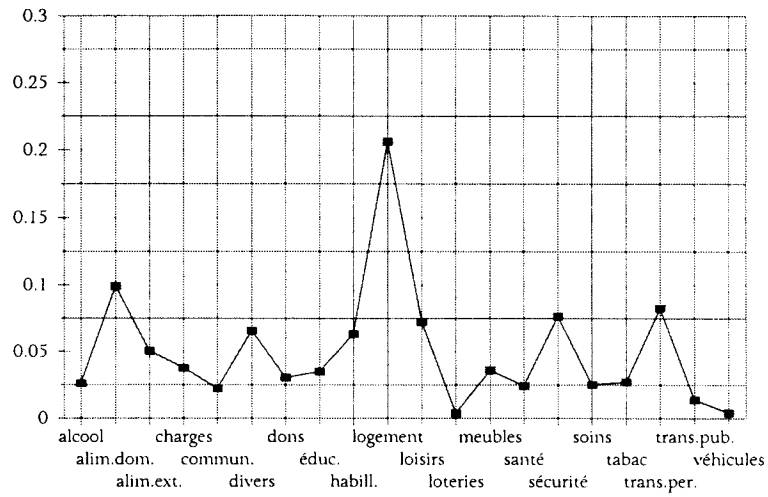
Class Means on Canonical Variables

CLU10	CAN1	CAN2	CAN3
1	-0.275215963	-1.057546937	1.015963386
2	-4.243884076	0.186223346	-0.253077198
3	1.678635664	-1.504985040	-1.531601530
4	-2.358916864	0.153880358	-0.830014646
5	3.061476680	0.201613452	-1.171804110
6	1.266784654	0.094792551	0.875128990
7	-2.100548207	-0.276814657	0.276954817
8	0.774978849	4.653461896	-0.293230475
9	1.234208374	-1.014766399	-1.173406743
10	1.911682637	-0.001135414	0.297922899

# PROFILS

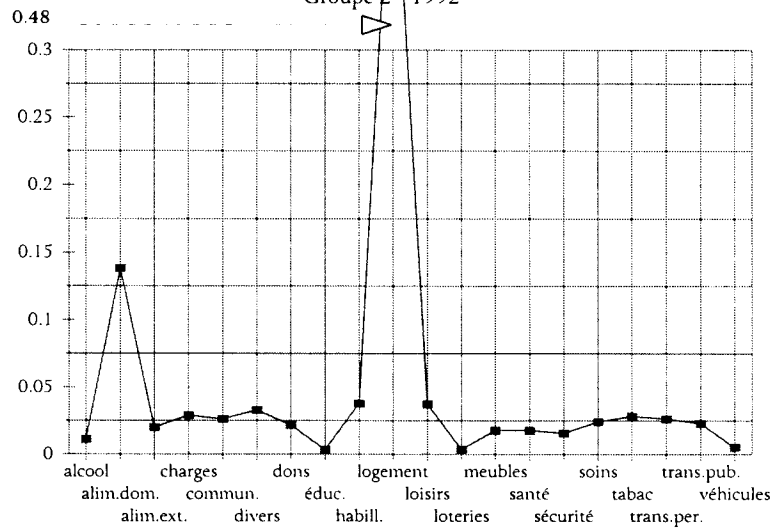
## PROFILS DE CONSOMMATION

Groupe 1 - 1992



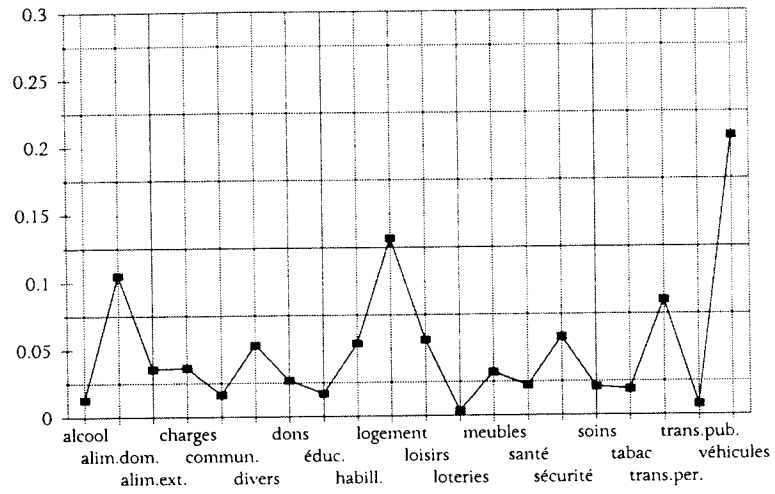
## PROFILS DE CONSOMMATION

Groupe 2 - 1992



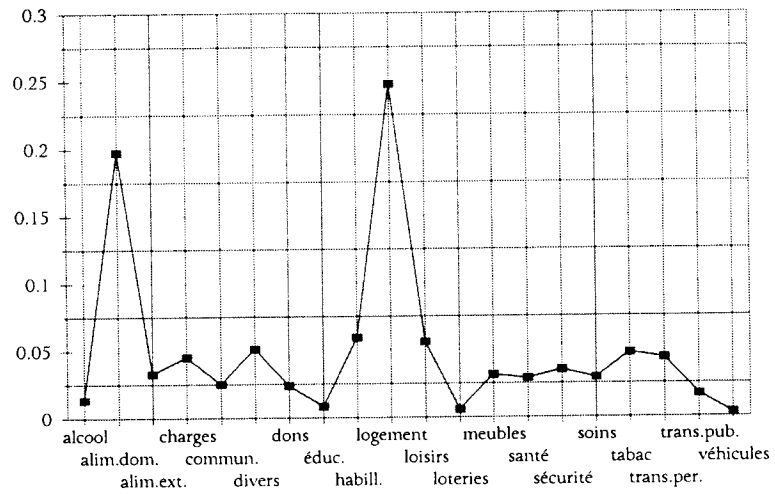
## PROFILS DE CONSOMMATION

Groupe 3 - 1992



## PROFILS DE CONSOMMATION

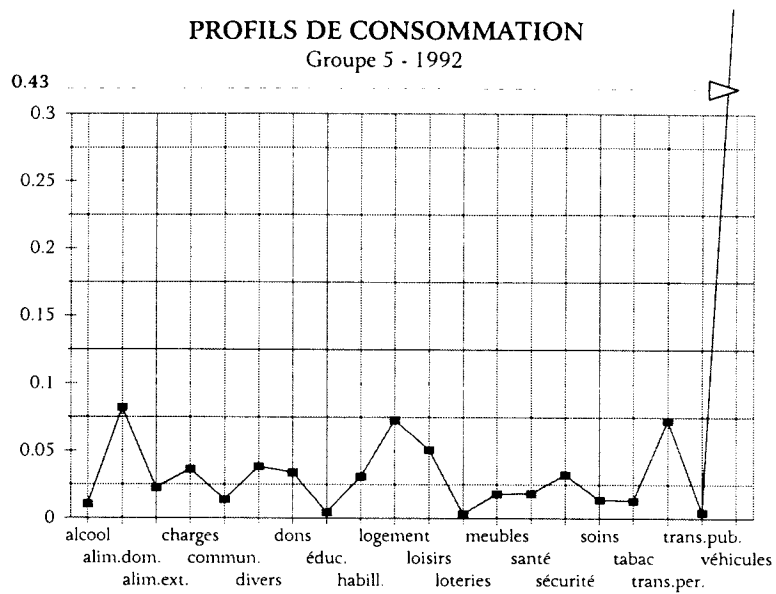
Groupe 4 - 1992





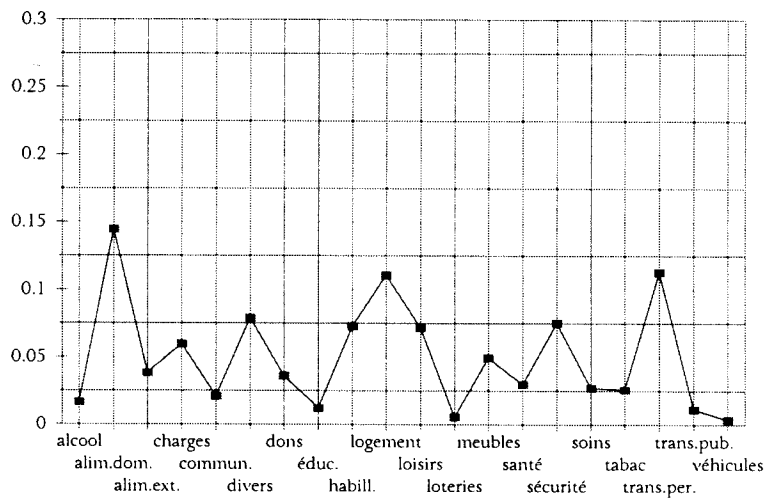
### PROFILS DE CONSOMMATION

Groupe 5 - 1992



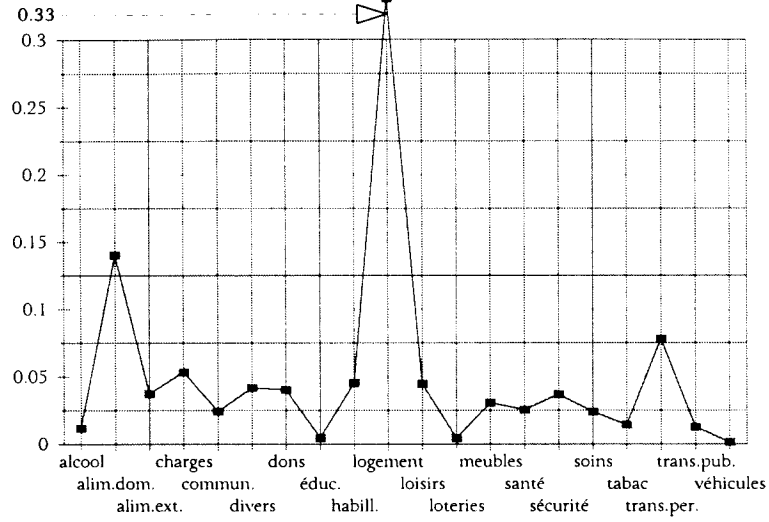
### PROFILS DE CONSOMMATION

Groupe 6 - 1992



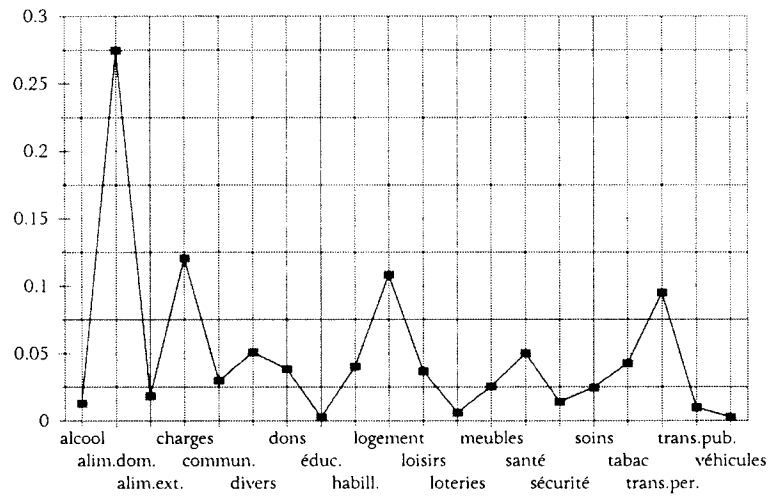
### PROFILS DE CONSOMMATION

Groupe 7 - 1992



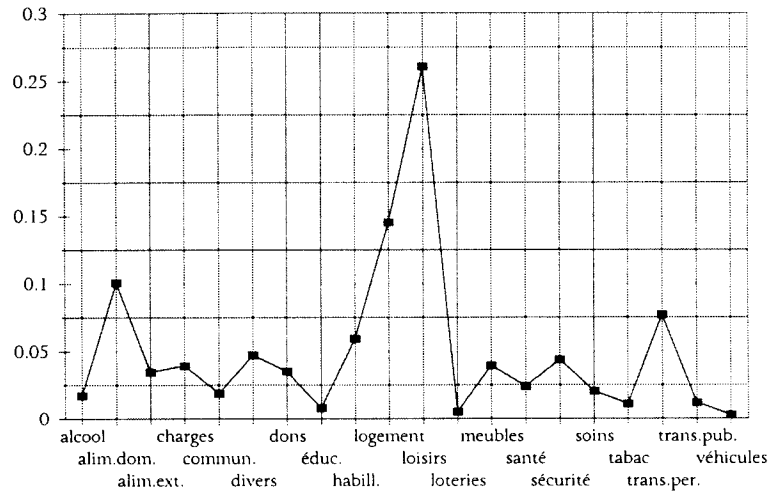
### PROFILS DE CONSOMMATION

Groupe 8 - 1992



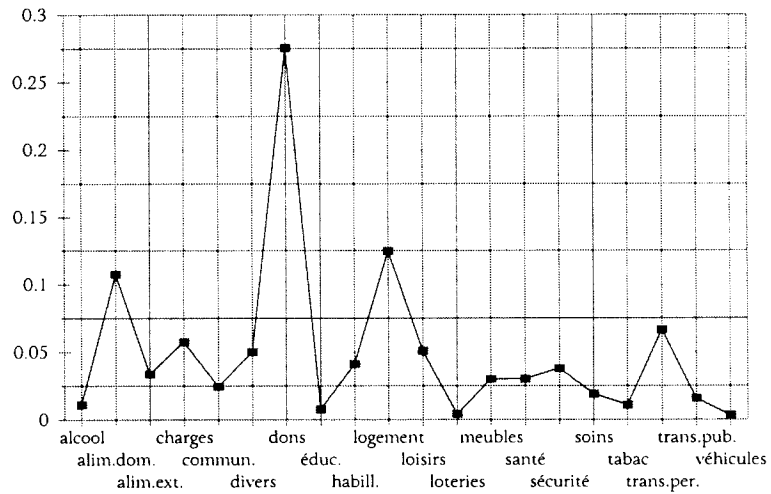
## PROFILS DE CONSOMMATION

Groupe 9 - 1992



## PROFILS DE CONSOMMATION

Groupe 10 - 1992



## TABLE DES MATIERES

1. Introduction	1
2. Présentation de la méthode	2
2.1 Généralités sur l'analyse neuronale et l'algorithme de Kohonen	2
2.2 Une méthode classique: l'analyse hiérarchique ascendante	4
3. Application aux enquêtes canadiennes	5
3.1 Présentation des enquêtes	5
3.2 Les classes obtenues et leur regroupement	6
3.3 Caractérisation des cellules	8
3.4 Classification classique	11
4. Conclusion	12
Bibliographie	13
Annexes	14
Annexe I Nomenclature des fonctions de consommation	15
Annexe II Profils des groupes	16
Annexe III Profils et composition des classes	22
Annexe IV Analyse canonique discriminante des classes Kohonen	30
Annexe V Résultats de la classification classique	33

## Prépublications du SAMOS

1991

- 1 - Catherine BOUTON, Marie COTTRELL, Jean-Claude FORT, Gilles PAGES  
Self-organization and convergence of the Kohonen algorithm. 21p.
- 2 - Ellen SAADA  
Seuil critique dans un réseau de neurones complètement connecté par des liens inhibiteurs. 6p.
- 3 - Xavier GUYON, Cécile HARDOUIN  
The chi-square coding test for nested Markov random field hypotheses. 11p.
- 4 - Xavier GUYON  
Variations et identifications de champs gaussiens sur  $R^2$ . 6p.
- 5 - Xavier GUYON, Hans R. KUNSCH  
Asymptotic comparison of estimators in the Ising model. 22p.
- 6 - Xavier GUYON  
Méthodes de pseudo-vraisemblance et de codage pour les processus ponctuels de Gibbs. 19p.

1992

- 7 - Collectif - SAMOS  
Résumés des exposés : Séminaires 1990/91 et 1991/92. 42p.

Hors Série

Actes du Congrès Satellite du Congrès Européen de Mathématiques  
Aspects Théoriques des Réseaux de Neurones  
Paris, 2 et 3 Juillet 1992, M.Chaleyat-Maurel, M.Cottrell, J.C.Fort (Editeurs)

- 8 - Gilles PAGES  
Mosaïques de Voronoï, algorithmes de quantification de l'espace et intégration numérique. 21p.

1993

- 9 - Claude BOUZITAT, Gilles PAGES  
Française des Jeux, biographie non autorisée. 21p.
- 10 - Jean-Claude FORT, Gilles PAGES  
Sur la convergence p.s. de l'algorithme de Kohonen généralisé. 7p.
- 11 - Jian-Feng YAO  
Détection de défauts de cuir et leur classification. 11p.
- 12 - Actes de la Rencontre Franco-Suisse 93 : Laboratoire d'Econométrie de Genève et SAMOS  
112p.
- 13 - Morgan MANGEAS, Corinne MULLER  
Réseaux de Neurones et Prévision de Séries Temporelles :  
Premier Rapport CERD SOAD/SAMOS. 33p.
- 14 - Morgan MANGEAS  
Réseaux de Neurones et Prévision de Séries Temporelles :  
Deuxième Rapport CERD SOAD/SAMOS. 71p.
- 15 - Marie COTTRELL et Jean-Claude FORT

Bases Mathématiques pour les Réseaux de Neurones Artificiels  
Cours COMETT-NEURAL. 94p.

- 16 - Marie COTTRELL et Jean-Claude FORT  
Mathematical Bases for the Artificial Neural Networks  
Cours COMETT-NEURAL 94p.
- 17 - Marie COTTRELL  
Les Mathématiques des Réseaux de Neurones  
Cours donné à Neuro-Nîmes 1992.
- 18 - Marie COTTRELL  
Cours NSI 93 (Saint Jean du Gard)  
Apprentissage et Chaîne de Markov, Analyse de Données et Réseaux de Neurones. 76p.
- 19 - Marie COTTRELL, Patrick LETREMY et Elisabeth ROY  
Analysing a Contingency Table with Kohonen Maps : a Factorial Correspondence Analysis. 7p.
- 20 - Marie COTTRELL, Bernard GIRARD, Yvonne GIRARD, Morgan MANGEAS et Corinne MULLER  
Neural Modeling for Time series: a statistical Stepwise Method for weight elimination. 21p.
- 21 - Xavier GUYON, Philippe JOLIVALDT  
Description de l'ensemble de bon choix de modèle par le critère du minimum de contraste pénalisé. Application à l'identification de modèle. 30p.
- 22 - Xavier GUYON, Philippe JOLIVALDT, José R. LEON  
Schémas de discrétisation d'un processus gaussien pour la simulation et l'estimation. 41p.
- 23- Morgan MANGEAS  
Réseaux de Neurones et Prévion de Séries Temporelles :  
Troisième Rapport CERD SOAD/SAMOS
- 24- Jean-Claude FORT, Gilles PAGES  
Réseaux de neurones: des méthodes connexionistes d'apprentissage. 19p.
- 25- Fabienne COMTE, Eric RENAULT  
Non causality in continuous time models. 36 p.
- 26 - Fabienne COMTE, Eric RENAULT  
Long memory continuous time models. 48 p.
- 27 - Xavier GUYON, Cécile HARDOUIN, Jian-Feng YAO  
Test de Différences de Contrastes et Somme Pondérés de Chi-deux, 24 p.
- 28 - Jean-Claude FORT, Gilles PAGES  
A non linear Kohonen algorithm, 6 p.

## 1994

- 29 - Jean-Claude FORT, Gilles PAGES  
About the a.s. convergence of the linear Kohonen algorithm with a generalized neighbourhood function,  
6 p.
- 30- Marie COTTRELL, Bernard GIRARD, Yvonne GIRARD, Morgan MANGEAS, Corinne MULLER  
SSM: A Statistical Stepwise Method for Weight Elimination, 6p.
- 31- Marie COTTRELL, Jean-Claude FORT, Gilles PAGES  
Two or three things that we know about the Kohonen algorithm, 10 p.
- 32- Marie COTTRELL, Jean-Claude FORT, Gilles PAGES

Comments about "Analysis of the Convergence Properties of Topology Preserving Neural Networks" by Zhen-Ping Lo, Yaoki Yu and Behnam Bavarian pp 207-220, Vol 4, N° 2, March 1993. 5 p.

- 33 - Jean-Claude FORT, Gilles PAGES  
Convergence d'algorithmes stochastiques: le théorème de Kushner & Clark revisité, 33 p.
- 34 - Claude BOUZITAT, Gilles PAGES  
Tant qu'il y aura des routes... 12 p.
- 35 - Xavier GUYON  
Modèle d'équation différentielle stochastique linéaire échantillonnée à temps discret. 13 p.
- 36 - Carlo GAETAN  
A stochastic algorithm for maximum likelihood of Gibbs point processes. 14 p.
- 37 - Smail IBBOU, Patrice GAUBERT, Christian TUTIN  
Prix des logements et prix du sol en Ile de France. 34 p.
- 38 - Catherine BOUTON, Gilles PAGES  
Convergence in distribution of the multidimensional Kohonen algorithm with 0 neighbour. 30 p.
- 39 - Samuel BAYOMOG, Xavier GUYON, Cécile HARDOUIN, Jian-Feng YAO  
Test de différence de contrastes et somme pondérée de Chi-deux. 24 p.

## 1995

- 40 - Claude BOUZITAT, Gilles PAGES  
Pour quelques images de plus..., 12 p.
- 41 - Jean-Claude FORT, Gilles PAGES  
About the Kohonen algorithm: Strong or Weak Self-organisation? 15 p.
- 42 - Marie COTTRELL, Patrick LETREMY  
Classification et analyse des correspondances au moyen de l'algorithme de Kohonen: application à l'étude de données socio-économiques. 10 p.
- 43 - Joël CHADOEUF, Xavier GUYON, Jian-Feng YAO  
Sur l'ergodicité de l'estimation par Restauration-Estimation de modèles incomplètement observés. 11 p.
- 44 - Jean-Gabriel ATTALI, Gilles PAGES  
Approximation of functions by perceptrons, a new approach. 11 p.
- 45 - Marie COTTRELL, Bernard GIRARD, Yvonne GIRARD, Corinne MULLER, Patrick ROUSSET  
Daily electrical power curves : classification and forecasting using a Kohonen map. 8 p.
- 46 - Fabienne COMTE, Cécile HARDOUIN  
Regression on log-regularized periodogram for fractional models at low frequencies. 19 p.
- 47 - Fabienne COMTE, Cécile HARDOUIN  
Regression on log-regularized periodogram under assumption of bounded spectral densities: the non fractional and the fractional cases. 14 p.
- 48 - Patrick ROUSSET  
Prévision des courbes demi-horaires au moyen d'une classification de Kohonen. 25 p.
- 49 - Smail IBBOU et Marie COTTRELL  
Multiple Correspondence analysis of a crosstabulations matrix using the Kohonen algorithm. 6 p.
- 50 - Philippe JOLIVALDT

Schémas de discrétisation pour la simulation et l'estimation d'un CAR(2): une étude expérimentale. 22 p.

51 - Philippe JOLIVALDT

Utilisation de méthodes implicites pour la simulation et l'estimation de modèles CAR(2) . 14 p.

1996

52- Samuel BAYOMOG

Estimation of a Markov field dynamic. 14 p.

53-.Morgan MANGEAS et Jian-feng YAO

Sur l'estimateur des moindres carrés des modèles auto-régressifs fonctionnels. 19 p.

54- Marie COTTRELL, Florence PIAT, Jean-Pierre ROSPARS

A Stochastic Model for Interconnected Neurons. 17 p.

55- Marie COTTRELL, Jean-Claude FORT, Gilles PAGES

Two or three mathematical things about the Kohonen algorithm. 31 p.

56- Marie COTTRELL, Bernard GIRARD, Patrick ROUSSET

Forecasting of curves using a Kohonen classification. 14 p.

57- Jean-Claude FORT, Gilles PAGES

Quantization vs Organization in the Kohonen S.O.M. 5 p.

58- Eric de BODT, Marie COTTRELL, Michel LEVASSEUR

Réseaux de neurones en finance. 33 p.

59- Marie COTTRELL, Eric de BODT, Emmanuel HENRION, Ismaïl IBBOU, Annick WOLFS, Charles Van WYMEERSCH

Comprendre la décision à l'aide d'une carte de Kohonen. Une étude empirique. 16 p.

60 Marie COTTRELL, Eric de BODT

Understanding the leasing decision with the help of a Kohonen map. An empirical study of the Belgian market. 5p.

61 Marie COTTRELL, Eric de BODT, Philippe GREGOIRE

The relation between interest rate shocks and the initial rate structure: an empirical study using a Kohonen map. 16p.

62 Marie COTTRELL, Eric de BODT, Philippe GREGOIRE

A kohonen map representation to avoid misleading interpretation. 8p.

63 Marie COTTRELL, Eric de BODT

Analyzing shocks on the interest rate structure with Kohonen map. 6p.

64 Marie COTTRELL, Eric de BODT, Philippe GREGOIRE

Simulating interest rate structure evolution on a long term horizon. A kohonen map application. 5p.

65 Fabienne COMTE, Cécile HARDOUIN

Log-regularized periodogram regression. 22p.

66 Jian Feng YAO

Simulation et optimisation sous contrainte par une dynamique de Metropolis. 6p.

67 Morgan MANGEAS, Jian-feng YAO

On least squares estimation for nonlinear autoregressive processes. 16p.

68 Carlo GAETAN, Xavier GUYON

Simulation des modèles de Gibbs spatiaux par chaîne de Markov. 28p.



69-François GARDES, Patrice GAUBERT, Patrick ROUSSET  
Cellulage de données d'enquêtes de consommation par une méthode neuronale. - 41p.