

CLASSIFICATION NEURONALE ET ANALYSE DES DONNEES « TRADITIONNELLE » : Quelques applications aux conditions de vie des ménages

M. COTTRELL^(*) et S. PONTHEUX^(**)

^(*)SAMOS/MATISSE - CNRS UMR

^(**)INSEE, Division « Conditions de vie des ménages »

1. Introduction

Depuis les années 70 au Royaume-Uni, plus récemment en France, la pauvreté est analysée en termes de conditions d'existence, parallèlement aux mesures plus classiques de la pauvreté monétaire, dans une perspective d'approche multi-dimensionnelle.

Les conditions d'existence couvrent des domaines *a priori* nombreux. Dickes (1994), en répertorie dix : le logement, les biens durables, la nourriture, l'habillement, les ressources financières, la santé, les relations sociales, les loisirs, l'éducation et le travail. Selon les travaux, le nombre de domaines retenus est plus ou moins grand. Le choix des domaines peut relever de deux logiques : une logique reposant sur l'hypothèse de rationalité des sujets, selon laquelle on retient les domaines pour lesquels les privations, manques ou absences sont censés diminuer ou disparaître lorsque les ressources monétaires augmentent (Mack & Lansley, 1984) ; une logique reposant plutôt sur la notion de « standard » (Townsend, 1989), qui conduit à considérer tous les domaines dès lors qu'ils concernent potentiellement tous les sujets, ou, comme le note Dickes (1994, p.184), « tous les ménages, quelle que soit leur composition ou leur situation quant au cycle de vie ».

Que l'on adopte l'une ou l'autre de ces logiques, et quels que soient les choix faits en matière des domaines retenus, le nombre d'indicateurs correspondant est nécessairement élevé. Le problème va donc être de traiter une grande quantité d'informations, de nature majoritairement qualitative. S'il est évidemment possible de décrire en détail la façon dont se combinent les différentes modalités dans chacun des domaines, il est toutefois malaisé d'en avoir une vision synthétique qui permettrait de classer les observations. On peut, à l'inverse, chercher à « résumer » cette information, par exemple par la construction de « scores » bâtis sur un

décompte des « manques », « absences » ou « privations » ; mais là, on perd la richesse du qualitatif... et la question du classement n'est pas résolue : à partir de quel score dira-t-on que les conditions de vie sont défavorables ? On propose ici d'aborder ce problème à travers trois exemples d'application, exploratoire, du traitement de ce type d'information à l'aide de l'une des techniques de « réseaux de neurones », le modèle - ou algorithme - de Kohonen, pour l'instant sur un nombre relativement petit de domaines des conditions d'existence (logement - en deux sous-domaines : confort élémentaire, problèmes particuliers -, qualité de l'environnement, biens durables et privations). Ces applications sont développées autour de deux questions :

- dans une optique de « description », comment se combinent, entre elles, les modalités des dimensions retenues, *i.e.* quelles sont les caractéristiques les plus fréquemment associées, et comment des ensembles de caractéristiques se situent-ils les uns par rapport aux autres ? On compare ici la classification des modalités obtenue par l'Analyse des Correspondances Multiples classique, et par une classification apparentée de Kohonen (KACM dans ce qui suit).

- dans une optique de « mesure », comment regrouper les observations, caractérisées uniquement par leur situation en termes de conditions de vie, de façon à obtenir un classement opérationnel ? Les observations sont classées tout d'abord selon les scores partiels calculés sur chacune des dimensions retenues, et on compare le produit d'une classification hiérarchique et celui d'une classification de Kohonen apparentée à l'analyse en composantes principales (KACP dans ce qui suit). Un dernier traitement consiste à construire, à l'aide d'une KACP, des classes d'individus caractérisés par leurs coordonnées issues d'une ACM ce qui permet d'utiliser toute la richesse qualitative de l'information.

La *section 2* est consacrée à une présentation de l'algorithme de Kohonen¹ Les programmes SAS utilisés pour les classifications de Kohonen ont été mis au point et développés par Patrick LETREMY (SAMOS/MATISSE, Université de Paris 1) La *section 3* fournit une rapide description des données utilisées. La *section 4* présente deux classifications alternatives des caractéristiques, et la *section 5* propose et compare plusieurs classements des observations.

¹ Les programmes SAS utilisés pour les classifications de Kohonen ont été mis au point et développés par Patrick LETREMY (SAMOS/MATISSE, Université de Paris 1)

2. L'algorithme de Kohonen

Il s'agit d'un algorithme original de classement qui a été défini par Teuvo Kohonen, dans les années 80, à partir de motivations neuromimétiques (Kohonen, 1984, 1995). Dans le contexte d'analyse des données, l'ensemble des entrées (ou inputs) est un ensemble fini et étant donnée une matrice de données, formée de N individus ou observations, décrits par un identificateur et p variables, l'algorithme regroupe les observations en classes, en respectant la topologie de l'espace des observations.

Cela veut dire qu'on définit a priori une notion de voisinage entre classes et que des observations voisines dans l'espace des variables (de dimension p) appartiennent (après classement) à la même classe ou à des classes voisines.

Les voisinages entre classes peuvent être choisis de manière variée, mais en général on suppose que les classes sont disposées sur une grille bi-dimensionnelle rectangulaire qui définit naturellement les voisins de chaque classe. On peut aussi considérer une topologie uni-dimensionnelle dite en ficelle, ou éventuellement un tore ou un cylindre.

Par exemple sur une grille, on pourra prendre des voisinages de rayon 3 (49 voisins), de rayon 2 (25 voisins) ou de rayon 1 (9 voisins). Les classes situées sur les bords ont éventuellement moins de voisins (Fig. 1.a). Pour une ficelle, les mêmes rayons donnent 7, 5 et 3 voisins (Fig. 1.b).

Figure 1.a – Voisinage sur une grille

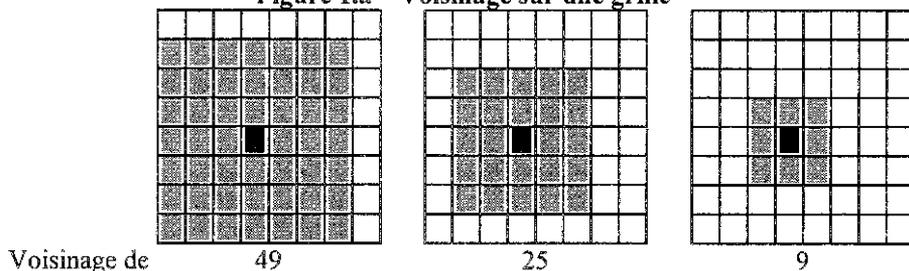


Figure 1.b – Voisinage sur une ficelle



2.1 Principe de l'algorithme de Kohonen

L'algorithme de classement est itératif. L'initialisation consiste à associer à chaque classe un *vecteur code* (ou *représentant*) de p dimensions choisi de manière aléatoire. Ensuite, à chaque étape, on choisit une observation au hasard, on la compare à tous les vecteurs codes, et on détermine la classe gagnante, c'est-à-dire celle dont le vecteur code est le plus proche au sens d'une distance donnée a priori. On rapproche alors de l'observation les codes de la classe gagnante et des classes voisines.

Cet algorithme est analogue à l'algorithme des centres mobiles (version stochastique), mais dans ce dernier cas, il n'y a pas de notion de voisinage entre les classes et on ne modifie à chaque étape que le *code* (ou *représentant*) de la classe gagnante.

2.2 Notations et définitions

On considère une table constituée de N observations (x_1, x_2, \dots, x_N) , où chaque individu est décrit par p variables quantitatives. La matrice $(N \times p)$ ainsi formée est la *matrice des données*. On peut choisir une loi de probabilité $P = (P_1, P_2, \dots, P_N)$ pour pondérer les N observations, mais dans la plupart des cas, on les pondère de façon uniforme, en prenant pour tout j de 1 à N , $P_j = 1/N$.

On se donne un réseau de Kohonen formés de n unités, rangées suivant une certaine topologie (en général une grille, ou bien une ficelle). Ces n unités sont munies d'un système de voisinages homogènes dans l'espace. Pour chaque unité i du réseau, on définit un voisinage de rayon r noté $V_r(i)$ et formé de l'ensemble des unités situées sur le réseau à une distance inférieure ou égale à r .

Chaque unité i est représentée dans l'espace R^p par un vecteur C_i (appelé vecteur poids par Kohonen) que nous désignerons par *vecteur code* ou par *représentant* de l'unité i (ou de la classe i). L'état du réseau au temps t est donné par $C(t) = (C_1(t), C_2(t), \dots, C_n(t))$.

Pour un état donné C et une observation donnée x , l'unité (ou classe) *gagnante* $i_0(C, x)$ est l'unité dont le vecteur code $C_{i_0}(C, x)$ est le plus proche de l'observation x au sens d'une certaine distance. On a donc :

$$i_0(C, x) = \underset{i}{\text{Arg min}} \|x - C_i\|.$$

Si l'on pose pour toute unité i , $?_i = \{x \in \{x_1, x_2, \dots, x_N\} / i_0(C, x) = i\}$, on dira que $?_i$ est la classe de numéro i et l'ensemble des classes $(?_1, ?_2, \dots, ?_n)$ forme une partition de l'ensemble des observations, appelée *partition de Voronoï*. Chaque classe est

représentée par le *vecteur code* correspondant. Chaque observation est représentée par le vecteur code le plus proche, exactement comme dans la méthode du plus proche voisin.

Alors pour un état C donné, le réseau définit une application $?_C$ qui à chaque observation x associe l'unité gagnante correspondante, c'est-à-dire le numéro de sa classe. Après convergence de l'algorithme de Kohonen, l'application $?_C$ respecte la *topologie* de l'espace des entrées, en ce sens que des observations voisines dans l'espace R^p se retrouvent associées à des unités voisines ou à la même unité.

L'algorithme de construction des vecteurs codes est défini de manière itérative comme suit :

- Au temps 0, les n vecteurs codes sont initialisés de manière aléatoire (on peut par exemple tirer au hasard n observations dans la table des données).
- Au temps t , l'état du réseau est $C(t)$, et on présente une observation tirée au hasard suivant la loi P , soit $x(t+1)$, on a alors :

$$\left\{ \begin{array}{l} i_0(C(t), x(t+1)) = \text{Arg min} \left\{ \|x(t+1) - C_i(t)\|, 1 \leq i \leq n \right\} \\ C_i(t+1) = C_i(t) - \varepsilon(t)(C_i(t) - x(t+1)), \quad \forall i \in V_{r(t)}(i_0) \\ C_i(t+1) = C_i(t+1), \quad \forall i \notin V_{r(t)}(i_0) \end{array} \right.$$

- * où $\varepsilon(t)$ est le *paramètre d'adaptation* ou de *gain*,
- * et où $r(t)$ est le rayon des voisinages au temps t .

Les paramètres importants sont

- la dimension p de l'espace des entrées,
- la topologie du réseau (grille, ficelle, cylindre, tore, etc.),
- le paramètre d'adaptation, positif, compris entre 0 et 1, constant ou décroissant,
- le rayon des voisinages, en général décroissant,
- la loi de probabilité des observations P .

L'étude de la convergence de cet algorithme est incomplète et pose des problèmes mathématiques difficiles (voir par exemple Cottrell & Fort, 1987 ; Cottrell, Fort & Pagès, 1995, 1997, 1998). Pour l'instant, l'essentiel des résultats

correspond à la dimension 1 (topologie en ficelle et observations de dimension $p = 1$). Des résultats d'organisation et de convergence en loi sont disponibles quand le paramètre d'adaptation est constant. Pour les résultats de convergence presque sûre après réorganisation, il faut que la suite des paramètres $\eta(t)$ vérifient des conditions de Robbins-Monro, classiques pour les algorithmes stochastiques, qui s'énoncent :

$$\sum_i \varepsilon_i = +\infty \quad \text{et} \quad \sum_i \varepsilon_i^2 < +\infty .$$

En d'autres termes, le paramètre doit être " petit ", mais " pas trop ".

Un des points qui rendent difficile l'étude théorique dans le cas général de l'algorithme de Kohonen, est qu'il n'existe pas de potentiel (ou énergie) associé lorsque les entrées sont munies d'une loi de probabilité P continue (cf. Erwin & al., 1992). Mais au contraire, dans le cas qui nous intéresse ici, c'est-à-dire quand l'espace des entrées est fini, muni d'une probabilité discrète, Ritter & al. (1992) ont montré que l'algorithme de Kohonen pour un rayon constant r de voisinage, est alors un algorithme du gradient stochastique qui minimise le potentiel :

$$V(C_1, C_2, \dots, C_N) = \sum_{i=1}^n \sum_{k \in \mathcal{V}_r(i)} \sum_{x_j \in \Gamma_k} \|x_j - C_i\|$$

Ce potentiel généralise la somme des carrés intra-classes. Ici on calcule la somme des carrés des distances de chaque observation non seulement à son vecteur code mais aussi aux vecteurs codes des classes voisines. L'étude mathématique de ce potentiel n'est pas simple non plus, car il n'est pas partout différentiable. En outre, il n'est pas possible d'éviter les minima locaux.

2.3 Classes et super-classes, étude des classes (KACP)

Un inconvénient de l'algorithme de base est que le nombre de classes doit être fixé a priori. Pour pallier cet inconvénient, on peut après la classification de Kohonen, pratiquer une classification de type hiérarchique sur les codes des classes de Kohonen, de manière à les regrouper en classes moins nombreuses.

Après convergence de l'algorithme, on a vu que les N observations sont classifiées en n classes selon la méthode du plus proche voisin, relativement à la distance choisie dans \mathcal{R}^p .

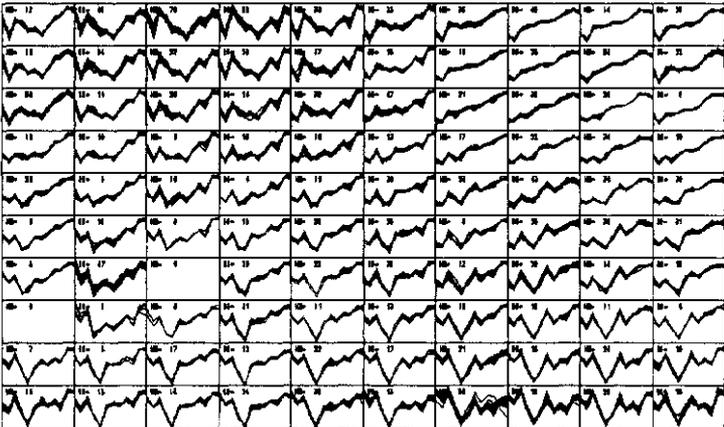
On peut alors construire une représentation graphique selon la topologie du réseau. Dans chaque cellule du réseau (grille, ficelle, etc.), on dessine les observations

associées ou on fait la liste de ces observations. Grâce à la propriété de conservation de la topologie, la représentation respecte les relations de voisinage. On obtient ainsi une carte de Kohonen où les grandes caractéristiques du nuage des données sont visibles après un petit nombre d'itérations (de l'ordre de 5 ou 6 N , en général).

Cette carte est un instrument d'analyse des données qui fournit des informations analogues à celle que donne une Analyse en Composantes Principales (ACP). Bien sûr, il n'y a pas de projection à proprement parler, la carte est grossière. Mais elle est unique, ce qui évite de devoir combiner les différentes projections planes qu'on obtient en ACP. La continuité d'une classe à ses voisins permet de bien comprendre l'évolution le long de la grille, et est facile à interpréter. C'est pour ces raisons que l'ensemble des techniques de classification et représentation fournies par l'algorithme de Kohonen appliqué à une matrice de données est désigné par le sigle KACP.

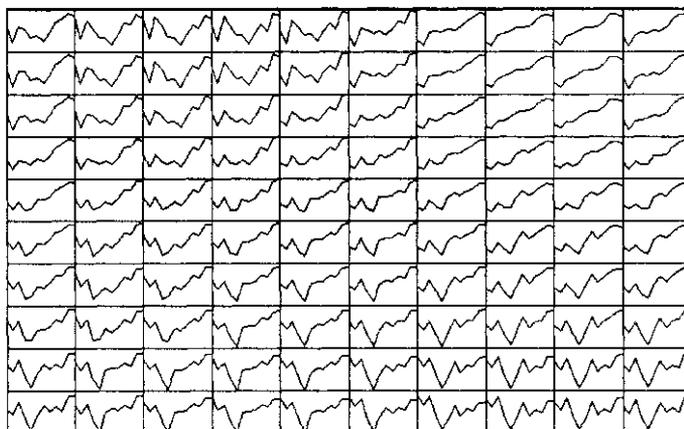
La figure 2 fournit une représentation du contenu de 100 classes, pour un cylindre 10 par 10, et des données (transformées) de consommation demi-horaire (Cottrell & Rousset, 1997). Dans ce cas, les observations qui sont des vecteurs de $p = 48$ points sont des courbes. Dans d'autres cas, on choisira des représentations en histogrammes, ou autres.

Figure 2 : Contenu des classes



Il est également intéressant de représenter sur le réseau les vecteurs codes, comme on le voit à la figure 3.

Figure 3 : Vecteurs codes



2.3.1 Distances entre les classes, super-classes

On peut mettre en évidence visuellement les *distances* entre les classes, qui sont artificiellement équidistantes dans les représentations ci-dessus. Pour cela, en suivant la méthode proposée par E. De Bodt & al. (1996), on dessine dans chaque cellule un octogone. Dans chacune des 8 directions principales, son sommet est d'autant plus proche du bord que la distance au voisin dans cette direction est petite. Ceci permet de faire apparaître les groupes de classes proches et donne une idée de la discrimination entre classes.

Comme le choix du nombre n de classes est arbitraire (et souvent élevé puisqu'on choisit couramment des grilles 8 par 8 ou 10 par 10), on peut réduire le nombre de classes, en les regroupant au moyen d'une *classification hiérarchique* classique sur les n vecteurs codes. On peut alors *colorier* les groupes de classes (appelés *super-classes*) pour les rendre visibles. On constate toujours que les super-classes ne regroupent que des classes contigües, ce qui s'explique par la propriété de respect de la topologie de l'algorithme de Kohonen. D'ailleurs, le non-respect de cette propriété serait un signe de manque de convergence de l'algorithme ou d'une structure particulièrement "repliée" du nuage des données.

Les figures 4 et 5 montrent les deux classements emboîtés (100 classes de Kohonen et 10 super-classes) et les distances inter-classes (Cottrell & Rousset, 1997).

Figure 4 : Classes et super-classes coloriées.

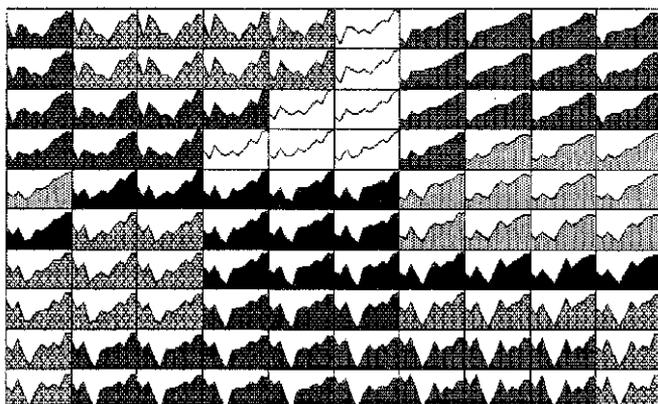
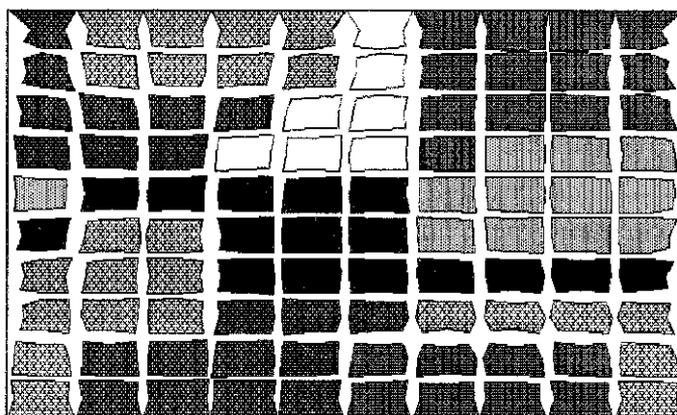


Figure 5 : Distances inter-classes



Pour décrire les super-classes, on calcule les statistiques de base des variables quantitatives utilisées, et on étudie la répartition des modalités des variables qualitatives non utilisées dans l'algorithme de classement de Kohonen (Cottrell & Rousset, 1997 ; Gaubert & al., 1995).

2.3.2 Cas particulier des réseaux uni-dimensionnels

Lorsqu'on utilise un réseau uni-dimensionnel, autrement dit une ficelle, en pratiquant ensuite éventuellement une classification hiérarchique sur les vecteurs-codes obtenus, on obtient alors naturellement une classification ordonnée, et on peut alors donner comme score aux individus de chaque classe (le numéro de la classe). Ce score pourra être interprété en ordre croissant ou en ordre décroissant, mais donnera en général une indication intéressante.

2.4 Analyse des relations entre variables qualitatives

On peut étendre l'usage de l'algorithme de Kohonen pour analyser les relations entre un nombre quelconque de variables qualitatives. On l'appelle alors KACM et il s'inspire de l'Analyse des Correspondances Multiples (pour une présentation de cette méthode accompagnée d'exemples, voir Cottrell, Letremy & Roy, 1993 ; Cottrell & Letremy, 1994 ; Cottrell & Ibbou, 1995).

On considère N individus et un certain nombre K de variables qualitatives. Chaque variable $k = 1, 2, \dots, K$ a m_k modalités. Chaque individu choisit une et une seule modalité pour chaque variable. Si M est le nombre total de modalités, chaque individu est représenté par un M -vecteur composé de 0 et de 1. Il n'y a qu'un 1 parmi les m_1 premières composantes, seulement un 1 entre la m_1+1 -ième et la (m_1+m_2) -ième, etc. Le tableau à N lignes et M colonnes ainsi formé est le tableau disjonctif complet, noté D .

Quand on ne cherche pas à classer les individus, mais seulement les modalités, on résume les données dans une table de Burt qui est un tableau de contingence généralisé. C'est une matrice symétrique de dimension $M \times M$ composée de $K \times K$ blocs, tels que le bloc B_{kl} (pour $1 \leq k \leq l \leq M$) est la table de contingence ($m_k \times m_l$) des variables k et l . Le bloc B_{kk} est une matrice diagonale, dont la diagonale est formée des nombres d'individus qui ont choisi les modalités 1, 2, ..., m_k , pour la variable k . Dans la suite, la table de Burt est notée B et elle se calcule par $B = D'D$.

2.4.1 L'algorithme KACM pour les modalités

On adapte l'algorithme de Kohonen à la table de Burt prise comme matrice des données B . Les lignes sont normalisées avec somme 1. A chaque étape, on tire une ligne normalisée au hasard selon la loi de probabilité donnée par la distribution empirique de la modalité correspondante. On définit l'unité gagnante selon la distance du χ^2 et on met à jour les vecteurs codes comme d'habitude.

Ces opérations (normalisation, pondération, usage de la distance du χ^2) sont équivalentes à la transformation de la matrice de Burt B en une matrice B^c dite matrice corrigée, où b_i et b_j sont les sommes par lignes et par colonnes de la matrice B . Il suffit alors d'appliquer l'algorithme de Kohonen (KACP) aux lignes de la matrice B^c :

$$b_{i,j}^c = \frac{b_{i,j}}{\sqrt{b_{i.}} \sqrt{b_{.j}}}$$

Après convergence, on obtient une classification organisée des modalités, telle que des modalités "liées" appartiennent à la même classe ou à des classes voisines. Dans ce cas, la méthode KACM fournit une technique alternative très intéressante à l'Analyse des Correspondances Multiples classique.

Un des principaux avantages de la méthode KACM est sa simplicité, sa rapidité et son faible temps de calcul. Elle produit une seule carte alors que les analyses classiques fournissent plusieurs représentations d'information décroissante. Elle est plus grossière, mais permet une interprétation rapide (pour plus de détails cf. Cottrell & Ibbou, 1995, Cottrell, de Bodt & Henrion, 1996).

2.4.2 Algorithme KACM pour les modalités et les individus

Si on désire classer et représenter simultanément les modalités des variables qualitatives et les individus, on s'inspire de la méthode factorielle classique, qui consiste à considérer le tableau disjonctif complet comme un grand tableau de contingence, où le facteur ligne est une variable "individu" à N modalités, et où le facteur colonne est une variable "descripteur" à M modalités. On utilise alors une version de l'algorithme adaptée au cas de deux variables. Dans ce cas, le temps calcul peut être assez important, mais fournit un classement simultané, unique, des modalités et des variables.

3. Les données

La source utilisée est la partie française du Panel Communautaire de Ménages (PCM), ici la troisième vague (1996). Cette source présente l'intérêt de fournir une information assez détaillée sur les revenus, l'activité et les conditions - matérielles - de vie (logement, environnement, biens durables, privations). On dispose également d'indicateurs d'aisance financière subjective (les ressources permettent-elles de vivre ? « très difficilement »... à « très aisément »).

Dans ce qui suit, les observations sont des ménages, décrits à la fois par leurs réponses aux questions sur les domaines des conditions matérielles (*cf. infra*) et par un ensemble de caractéristiques précisant : le type de ménage, l'âge moyen des adultes (personnes de 17 ans et plus), le nombre d'enfants de moins de 17 ans, le type d'habitat et la zone d'habitation, le degré d'aisance financière subjective, un score global de conditions de vie (obtenu par simple sommation des « mauvais » points) sur l'ensemble des items de conditions matérielles et un score partiel pour chacun des domaines, le revenu mensuel courant par unité de consommation et une indicatrice de pauvreté monétaire. On ne retient que les observations pour lesquelles cet ensemble d'informations est complet, soit 6458 ménages. La table 1 fournit les moyennes de ces caractéristiques pour l'ensemble des ménages.

Seules les variables relatives aux conditions matérielles sur les domaines du logement, de l'environnement, des biens durables et des privations sont prises en compte pour décrire les « conditions de vie », et pour classer les ménages ; les autres descripteurs permettent de comparer les classements obtenus. Sur les différents

domaines retenus, tous les items disponibles ont été conservés par souci de simplicité².

On a au total 10 items relatifs au logement (5 de description du confort et 5 relatifs à des problèmes dans le logement), 4 concernant la qualité de l'environnement, 6 biens durables et 6 items de privation, soit un ensemble de 26 variables dichotomiques³, donc 52 modalités (la liste détaillée des items et les fréquences associées pour l'ensemble des observations figure à l'annexe 1).

Selon les traitements, on utilise successivement :

- l'ensemble des items dichotomiques sous la forme d'un tableau de réponses ou d'un tableau disjonctif complet, pour la représentation des modalités,
- les scores obtenus sur chacun des domaines (« note » égale au nombre de « mauvais » points par domaine ; les notes vont donc de 0 à 5 pour le confort du logement, de 0 à 5 pour les problèmes du logement, de 0 à 4 pour le domaine de l'environnement, etc.), pour un premier type de classification des observations,
- les coordonnées des individus calculées après une ACM pour un second type de classification.

² On s'éloigne donc un peu de l'option de Dickes (1994), pour qui le choix des items doit viser à repérer les « manques perçus comme défavorables par la majorité », ce qui renvoie à un critère de « consensus », parfois ramené, faute d'enquêtes qui permettraient d'appliquer ce critère, à un contrôle par la fréquence. Lollivier & Verger (1997), avaient écarté certains items du domaine des biens durables.

³ Pour les biens durables, la réponse distingue les cas de non possession par manque de moyens des cas de non possession par choix (le sujet déclare ne pas « éprouver le besoin »). On a ici regroupé la non possession par « choix » avec la possession, sous l'hypothèse, simplificatrice, que dans ce cas, « ne pas avoir » n'est pas le signe d'un manque ou d'une privation.

Table 1 - Les descripteurs

		Variable	Fréquences (%)	Moyennes
Type de logement	maison isolée	LOGT 1	39.9	
	maison av. voisinage	2	21.4	
	immeuble <10 logements	3	13.4	
	collectif 10 logements et +	4	24.3	
	autre	5	1.0	
Zone d'habitation	commune rurale	TUR 0	27.7	
	commune <10000 h	1	10.9	
	de 10000 à <100000 h	2	19.5	
	de 100000 à <2000000 h	3	28.5	
	Agglomération parisienne	4	13.5	
Type de ménage (enfants : <25 ans)	personne isolée	TYM 0	25.1	
	couple sans enfant	1	26.3	
	couple avec enfant(s)	2	37.5	
	famille monoparentale	3	7.3	
	autre cas	4	3.8	
Nombre total de personnes	NBTOT			2.6
Nombre d'enfants <17 ans	NB17			0.6
Age moyen des personnes de 17 ans et plus	AGEM			46.7
Aisance déclare « vivre :	très difficilement	SLS 1	5.7	
	difficilement	2	12.3	
	assez difficilement	3	29.6	
	assez aisément	4	39.0	
	aisément et très aisément	5	13.3	
Revenu mensuel courant par u.c.(a)	REVUC			7650
Pauvreté monétaire	pauvre(b)	POOR 1	10.7	
	non pauvre	2	89.3	
Conditions de vie (score=somme des « mauvais points »)	Score total	TOT4		3.5
	score confort du logement	CLOGT		0.3
	score problèmes du logt.	PLOGT		0.5
	score environnement	ENVIR		0.7
	score biens durables	DURAB		0.4
score privations	PRIV			1.0

Source : Insee, PCM vague 3

Champ : ménages

(a) avec échelle modifiée (1 - 0.5 - 0.3)

(b) pour un seuil de pauvreté défini à 50 % du revenu médian par u.c.

Les fréquences détaillées pour l'ensemble des items figurent à l'annexe 1.

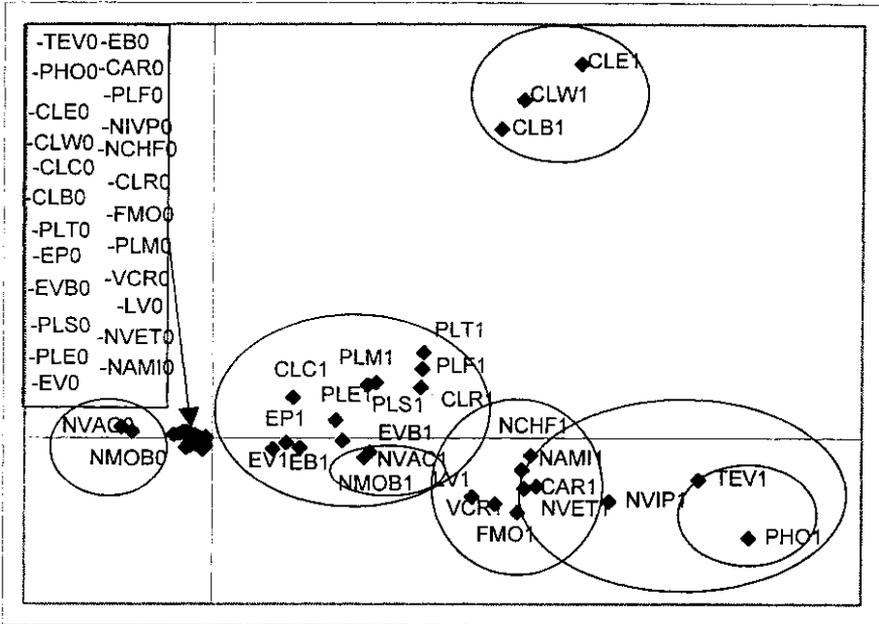
4. Classements des modalités

On va chercher ici à obtenir une représentation de la façon dont les différentes modalités des items retenus se combinent, et comment elles se situent les unes par rapport aux autres. Pour cela, on utilise successivement l'analyse des correspondances multiples puis une classification de Kohonen.

4.1. La représentation des modalités par l'ACM

Toutes les projections font apparaître un « paquet » des modalités « positives », et, selon les dimensions projetées divers groupes de modalités « négatives » (on ne présente ici que la projection des deux premières dimensions).

Graphique 1 - Analyse des correspondances multiples



Dans toutes les projections, un ensemble de modalités apparaît systématiquement isolé : il correspond à ce que l'on pourrait appeler absence de confort minimum du logement et combine : pas d'eau chaude courante, pas de baignoire ou douche, et pas de wc à l'intérieur du logement (CLE1-CLB1-CLW1). Les autres modalités s'organisent en quelques groupes associant souvent problèmes dans le logement et mauvaise qualité de l'environnement d'une part (respectivement modalités de type Pxx1 et Exx1), certaines privations (type Nxx1) et l'absence de biens durables. On note également la proximité de l'absence des biens durables les plus « modernes » d'un côté (four à micro-ondes - FMO1, magnétophone - VCR1 et lave-vaisselle - LV1) et de l'autre celle des plus classiques (téléphone - PHO1, et télévision - TEV1). Enfin, toutes les projections suggèrent un caractère particulier des privations quant aux départs en vacances et quant à la possibilité de remplacer des meubles usés ou cassés, et ce d'ailleurs tant du côté « négatif » que du côté « positif » (ces deux modalités - NVAC et NMOB - sont les seules à être un peu distantes du « paquet » mentionné plus haut).

L'ACM suggère finalement plus ou moins 4 groupes de modalités :

- un premier groupe de modalités de indiquant l'absence de manques, privations ou problèmes, dont émergent les cas particuliers de la possibilité de vacances et de remplacement des meubles usés ou cassés (NVAC0 et NMOB0 dans le graphique 1) ;

- un second groupe associant existence de problèmes dans le logement, mauvaise qualité de l'environnement (modalités commençant par P ou par E dans le graphique 1) et impossibilité de vacances ou de remplacement des meubles (NVAC1 et NMOB1) ;

- un - ou deux groupes, selon les projections -, constitués autour de l'absence de confort minimum du logement (CLE1, CLB1, CLW1), de l'absence du téléphone et de la télévision (PHO1 et TV1), et de l'impossibilité de consommer viande ou poisson au moins un jour sur deux (NVIP1) - cet ensemble de modalités n'apparaît clairement qu'à la projection des plans 3 et 4 -.

4.2 La représentation des modalités avec une KACM

On a classé les modalités successivement dans une grille de 10 x 10, puis dans une « ficelle » de 5 classes. L'intérêt de la grille est de fournir une carte de la position des modalités qui s'interprète très directement, celle de la ficelle de fournir une vue très synthétique des associations.

Graphique 2 - KACM des modalités

2.a. Position des modalités sur une grille de 10 x 10

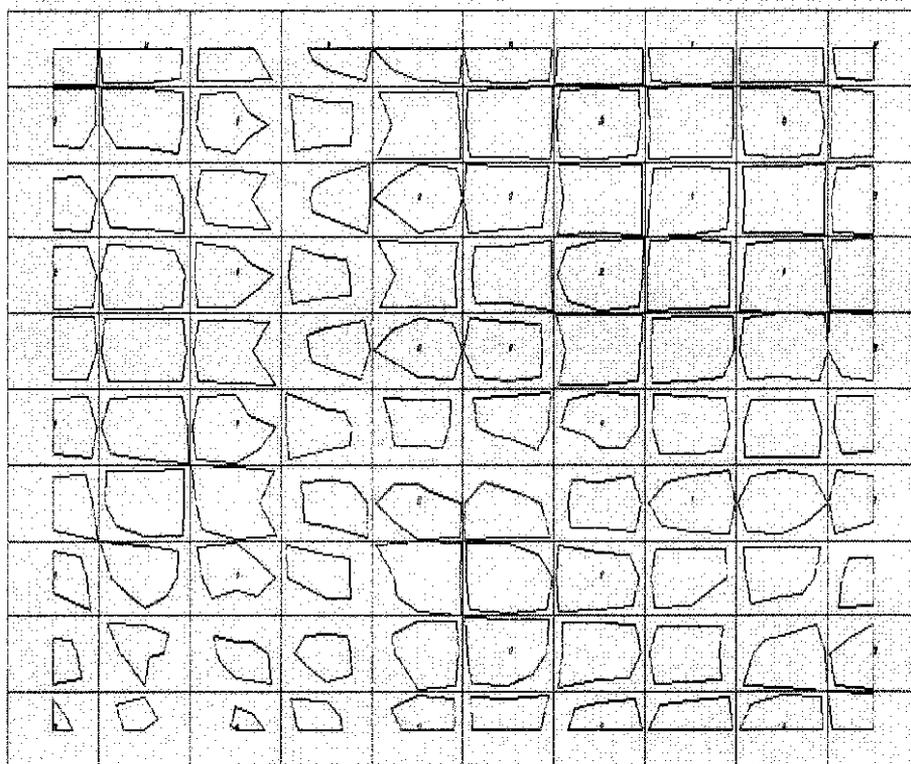
	EB1 EP1		LV0		FMO0 VCR0		CAR0		CLB0 CLW0 CLE0
EVB1		EV1				NCHF0 NVIP0		PLT0 PHO0 TEV0	
				NAMIO	NVETO		EVB0		CLC0 PLE0 PLS0
CLB1 PLF1		PLM1				EB0 EP0		CLR0	
				NMOB0	NVAC0				PLF0 PLM0
PLS1		PLE1				EV0			
				VCR1			NMOB1		NVAC1
PLT1		CLC1				LV1			
					FMO1				NAMI1 NVET1
CLB1 CLW1 CLE1		PHO1 TEV1		CAR1		NCHF1		NVIP1	

Classification des modalités en 100 classes au moyen de l'algorithme KACM. Les modalités terminées par 0 indiquent l'absence de problème, la détention ou l'absence de privation, celles qui sont terminées par 1 indiquent la présence de problèmes.

Les modalités « positives » se regroupent en haut et à droite de la grille, les « négatives » en bas et à gauche. La classe tout en bas à gauche correspond aux modalités extrêmes (pas de salle de bain, pas de toilettes, pas d'eau courante). Les problèmes liés à l'environnement sont regroupés en haut et à gauche. Si l'on cherche plutôt les classes « isolées », on retrouve le trio des modalités signalant l'absence de confort minimum du logement : elles sont regroupées dans une classe sans voisines,

par ailleurs complètement à l'opposé de la classe regroupant les modalités positives de ces items. Cette classe apparaît également très distante de toutes les autres. On observe également que les problèmes du logement pris un à un sont sans voisins immédiats, bien que pas trop éloignés les uns des autres. Enfin, absence de téléphone et absence de télévision sont sans voisins. Si on cherche maintenant à regrouper des classes voisines, l'examen des distances (graphique 2.b) suggère un ensemble constitué autour de l'absence du confort élémentaire et de biens durables très répandus (CLB1-CLE1-CLW1, PHO1-TEV1, et éventuellement CAR1) et éventuellement associé à des privations sur des besoins élémentaires (NCHF1, NVIP1) - en gros, le bas de la carte, qui grouperait des caractéristiques de conditions très défavorisées. Un autre ensemble pourrait se constituer avec les problèmes du logement et de l'environnement (partie gauche de la carte). Enfin, un gros ensemble pourrait regrouper les modalités « positives », à la frontière duquel se trouvent certaines caractéristiques « négatives » (notamment départs en vacances et remplacement de mobilier).

2.b. Distances entre les classes sur une grille de Kohonen



Représentation des distances entre les 100 classes. Les classes en haut et à droite (modalités « positives ») sont plutôt rapprochées les unes des autres. Les classes contenant les modalités « négatives » extrêmes (pas de salle de bain, pas de

toilettes, pas d'eau courante) sont au contraire assez séparées des autres.

Le classement sur une ficelle donne une vision beaucoup plus ramassée, d'autant qu'on a ici volontairement choisi un petit nombre de classes (graphique 2.c).

S'opposent très nettement presque toutes les modalités « positives », à l'exception là encore des vacances et du remplacement de mobilier - qui se trouvent associés ici à une variable d'environnement - (graphique 2.c., à droite) à trois classes de modalités « négatives », celle de gauche regroupant les modalités identifiées plus haut comme conditions très défavorisées, puis, en allant vers la droite, des modalités représentant principalement des problèmes du logement, et, encore à droite les problèmes d'environnement, associés au duo vacances-mobilier.

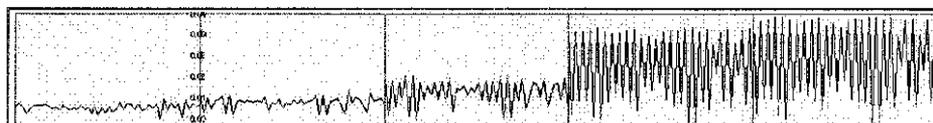
2.c. Position des modalités sur une ficelle de Kohonen (5 classes)

CLB1	CLC1	EB1	EV0	CLB0	EVB0
CLW1	CLR1	EP1	NMOB0	CLC0	FMO0
CLE1	PLF1	EV1	NVAC0	CLR0	VCR0
PLT1	PLM1	NMOB1		CLW0	PHO0
FMO1	PLE1	NVAC1		CLE0	TEV0
PHO1	PLS1			PLF0	CAR0
TEV1	EVB1			PLM0	LV0
CAR1	VCR1			PLE0	NAM10
NCHF1	LV1			PLS0	NCHF0
NVIP1	NAM11			PLT0	NVET0
	NVET1			EB0	NVIP0
				EP0	

Classification des modalités en 5 classes au moyen de l'algorithme KACM

L'examen des représentants des classes suggère une gradation nette (graphique 2.d), et on pourrait envisager, avec un nombre de classes un peu plus grand, de tester ce classement pour la construction d'une pondération des items.

2.d. Profil des représentants de classes sur la ficelle de Kohonen



Représentants (vecteurs-codes) des 5 classes

5. Classements des observations

On cherche maintenant à classer les observations. La première possibilité consiste à utiliser directement le score de « mauvais points » obtenu sur l'ensemble des items. Reste à déterminer le seuil (ou les seuils) permettant de constituer des classes.

Lollivier & Verger (1997) discutent de ce problème, et proposent d'utiliser le taux de pauvreté monétaire pour déterminer, à partir des fréquences cumulées, le score à partir duquel on dira les conditions de vie « défavorables », ou en d'autres termes le « seuil » de pauvreté des conditions de vie. Un inconvénient de cette méthode est évidemment qu'elle ne permet pas directement de confronter les évolutions au cours du temps des pauvretés monétaire et en termes de conditions de vie, puisque les deux « taux » sont, par construction, identiques ou très proches. Elle produit également un classement assez abrupt (les conditions de vie sont « bonnes » tant que le nombre de mauvais points ne dépasse pas le score s et défavorables lorsqu'il est égal à $s+1$) qui, s'il permet de repérer les situations de cumul, n'est pas complètement satisfaisant : en effet, même si on ne dispose pas d'éléments pour pondérer les items, tous ne se situent clairement pas au même « niveau » de gravité.

Par rapport à ce point de départ, l'objectif dans ce qui suit est d'explorer des classements indépendants d'un « seuil », qu'il soit donné par le taux de pauvreté monétaire ou n'importe quel autre indicateur exogène, et tenant compte du fait que différents domaines sont pris en compte. On a tout d'abord classé les observations à partir des scores partiels (par domaine). On a ensuite classé les observations à partir de leurs coordonnées dans une ACM, le point de départ étant cette fois-ci l'ensemble des réponses sur tous les items.

Pour toutes les classifications, on présente systématiquement :

- une représentation graphique le cas échéant,
- un ensemble de statistiques descriptives donnant la fréquence de chaque classe, sa position dans un regroupement obtenu avec une classification hiérarchique en 3 classes, les scores moyens de chacune de ces classes sur chaque dimension, et quelques autres statistiques descriptives : revenu moyen par u.c., indicatrice de pauvreté, et indicatrice d'aisance subjective.

L'annexe 2 fournit une statistique détaillée pour tous les items selon les différents classements.

5.1. Classes basées sur les scores par domaines

On part ici des 5 scores obtenus sur les 5 domaines retenus. On utilise successivement une classification hiérarchique puis une classification de Kohonen.

5.1.1 Classification hiérarchique classique

On a produit dix clusters, regroupés ensuite en 3 « super-classes ». La table 2 présente les statistiques associées.

Table 2- Dix classes résultant d'une classification hiérarchique sur les scores partiels

10 classes		%	3 « super-classes »		
			1	2	3
1		5.1	0	0	100
2		3.2	0	0	100
3		8.0	0	100	0
4		6.1	100	0	0
5		3.6	0	0	100
6		2.8	0	0	100
7		16.6	0	100	0
8		3.3	100	0	0
9		15.0	100	0	0
10		36.3	100	0	0
Total		100.0	60.7	24.6	14.8
Scores moyens par domaine et total	CLOGT		0.2	0.3	0.7
	PLOGT		0.4	0.3	1.2
	ENVIR		0.6	0.5	1.1
	DURAB		0.0	0.6	1.2
	PRIV		0.2	1.6	3.5
	TOT4		1.6	4.4	10.0
Revenu moyen par u.c.	REVUC		9071	5940	4664
Pauvreté monétaire	oui		4.03	13.7	33.1
	non		96	86.3	66.9
Aisance (l'indicateur va de « vivre difficilement » à « vivre très aisément »)	1		0.7	5.0	27.2
	2		5.5	18.9	29.7
	3		22.1	45.8	33.7
	4		52.0	25.4	8.4
	5		19.7	4.92	1.15

Ici, l'examen des descripteurs conduirait à identifier la classe 3 comme la moins favorisée (14,8 % des observations, dont un tiers de pauvres du point de vue monétaire), la classe 2 comme intermédiaire (24,6 % des observations, dont près de 14 % de pauvres au sens monétaire) ; la classe 1, *a priori* bénéficiant des conditions les plus favorables, regroupe 60 % des observations et les pauvres y sont très nettement sous-représentés (4 % de pauvres contre 10,7 % pour l'ensemble de l'échantillon). De la classe 1 à la classe 3, les scores par domaine ainsi que le score total sont croissants. Enfin, les trois classes se différencient également du point de vue de l'aisance subjective, nettement décroissante de la classe 1 à la classe 3.

5.1.2 Classification de Kohonen, KACP

A partir de la même information que ci-dessus (5 scores partiels), on produit, avec l'algorithme de Kohonen, 10 classes, qui sont ensuite regroupées en 3 super-classes. Compte tenu des données (scores sur chaque domaine), on a priori l'idée d'une progression « en ligne », et donc il nous a semblé plus intéressant de classer les observations sur une « ficelle ». On obtient le classement présenté au graphique 3. A la différence de la classification hiérarchique précédente, les observations sont directement classées selon un *continuum*, ce qui s'illustre d'ailleurs aussi dans le croisement entre classes et super-classes (table 3).

Graphique 3 – Classement des observations selon les scores partiels dans une ficelle de Kohonen

3.a. Distances entre les classes



Classement des ménages selon les scores partiels par dimension, dans une ficelle de Kohonen à 10 classes, regroupées en 3 super-classes. Cela permet d'attribuer une note entre 1 et 10 à chaque ménage.

3.b. Vecteurs moyennes des classes



Le graphique 3.b., qui représente les moyennes de classes montre nettement une progression le long de la ficelle, et indique un classement des ménages allant des « meilleures » conditions de vie (à gauche – super-classe 1) vers ceux ayant le plus de difficultés (à droite, super-classe 3). L'examen des statistiques descriptives associées confirme cette hiérarchie (table 3).

Table 3- Dix classes résultant d'une classification de Kohonen sur les scores partiels

10 classes		%	3 « super-classes »		
			1	2	3
1		30.8	100	0	0
2		9.6	100	0	0
3		11.9	100	0	0
4		5.5	100	0	0
5		10.7	100	0	0
6		9.1	100	0	0
7		4.8	0	100	0
8		6.1	0	100	0
9		4.6	0	0	100
10		6.8	0	0	100
Total		100.0	77.6	10.9	11.5
Scores moyens par domaine et total	CLOGT		0.1	0.5	1.1
	PLOGT		0.2	2.2	1.0
	ENVIR		0.5	1.5	0.8
	DURAB		0.1	0.3	2.0
	PRIV		0.7	1.3	2.8
	TOT4		2.3	6.7	8.5
Revenu moyen par u.c.	REVUC		8267	6484	4587
Pauvreté monétaire	oui		6.7	14.5	34.0
	non		93.3	85.5	66.0
Aisance (de « vivre très difficilement » à « vivre très aisément »)	1		2.8	7.2	23.5
	2		9.8	17.4	24.9
	3		27.9	37.3	34.2
	4		43.7	30.7	15.3
	5		15.8	7.4	2.16

Les proportions des trois classes finales sont légèrement différentes de celles obtenues par la classification hiérarchique : on aurait ici 11,5% des observations dans des conditions « défavorables », 10,9% dans des conditions « intermédiaires » et près de 78% dans les conditions les plus favorables. Les classes sont moins « tranchées » du point de vue de la pauvreté monétaire que dans l'analyse précédente, et les classes 2 et 3 (ici les moins favorisées) se distinguent par les scores des problèmes du logement et de l'environnement (plus élevés pour la classe 2), du confort du logement, des biens durables et des privations (plus forts dans la classe 3). L'aisance décroît de la classe 1 à la classe 3.

5.2. Classes basées sur les coordonnées des observations après une ACM

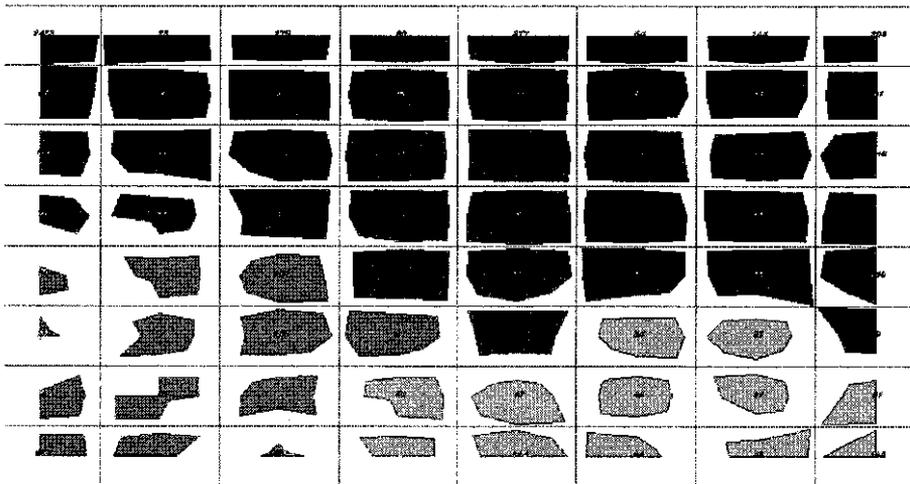
Les classes que l'on va constituer maintenant ne reposent plus sur des scores, mais sur les coordonnées des observations, calculées après une ACM classique. Il s'agit donc au départ d'un changement de base dans l'espace des réponses et donc d'une transformation des réponses de toutes les observations sur tous les items. L'idée est ici de « restituer » la richesse qualitative de l'information initiale dans la construction des classes, richesse évidemment perdue dans l'étape précédente où les observations n'étaient repérées que par leurs scores sur les différents domaines. On utilise l'algorithme de Kohonen pour classer les observations successivement dans une grille puis sur une ficelle. Chaque classement est ensuite regroupé en 3 super-classes finales.

5.2.1 KACP des coordonnées et grille

Une classification de Kohonen apparentée à l'analyse en composantes principales permet de constituer une grille (ici de 8 X 8). Les 64 classes obtenues sont regroupées en 10 ou en 3 super-classes (on ne présente dans les statistiques descriptives associées, pour une raison évidente de place, que ces deux niveaux de regroupement). Le graphique 4 montre la carte ainsi obtenue.

Graphique 4 – Grille de Kohonen avec KACP à partir des coordonnées des observations

4.a – Distances entre les 64 classes



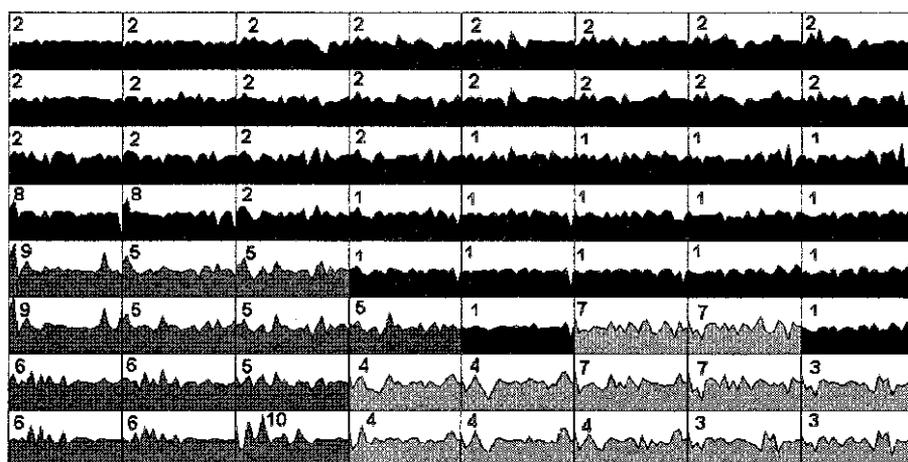
Classement des ménages (représentés par leurs coordonnées calculées après une analyse des correspondances multiples classique) sur une grille de Kohonen de

taille 8 par 8. Les tons de gris mettent en évidence les 3 super-classes obtenues par regroupement.

On retrouve ici une représentation assez proche de celle obtenue par KACM sur les modalités, et on obtient *in fine* une super-classe assez ramassée (partie haute de la carte), et deux super-classes qui occupent les parties gauche et droite du bas de la carte.

Les profils des classes (graphique 4.b) font apparaître des différences sensibles plutôt des niveaux entre la super-classe A et les deux super-classes B et C, et plutôt des profils à proprement parler entre les super-classes B et C. On constate en outre que le regroupement en 10 super-classes se trouve contenu dans le regroupement en 3 super-classes, ce qui illustre la propriété de voisinage, qui constitue l'un des « plus » de ce type de classification.

4.b – Profils des classes, regroupement en 10 ou 3 super-classes



Classement des ménages (représentés par leurs coordonnées calculées après une analyse des correspondances multiples classique) sur une grille de Kohonen de taille 8 par 8. Les 64 classes ont été regroupées en 10 (numérotées), puis en 3 super-classes (tons de gris). La super-classe A correspond au haut de la carte, la B est en bas à gauche, la C en bas à droite.

Les trois super-classes finales représentent des proportions sensiblement différentes de celles obtenues avec les classifications précédentes : de la super-classe A à la super-classe C⁴, on a respectivement 15,2 %, 14,2 % et 70,6 % des observations

⁴ Les numéros des super-classes n'ont a priori pas de sens hiérarchique : sur une ficelle, où les classes sont ordonnées « en ligne », les super-classes le sont également

(table 4). Si le groupe dont les conditions sont les « moins défavorables » apparaît nettement (super-classe A), les situations sont moins tranchées entre les deux autres classes, même si elles s'ordonnent du point de vue des revenus moyens avec $C > B$; cela pourrait illustrer l'effet de la prise en compte de l'information à un niveau très détaillé, et donc plus riche que celle fournie par les scores partiels.

La comparaison des classes B et C est particulièrement intéressante : au niveau des scores partiels, ces deux super-classes se différencient principalement sur le domaine du logement (scores moyens plus élevés dans la super-classe C, tant du point de vue « confort » - CLOGT - que du point de vue « problèmes » - PLOGT -), et les biens durables et privations (qui concernent plus intensivement la super-classe B). Du point de vue du score global, de la pauvreté monétaire et de l'aisance, la classe B apparaît moins « favorisée » que la classe C.

Table 4- Dix super-classes et trois super-classes résultant d'une KACP des coordonnées sur une grille

10 super-classes, dans une grille 8 x 8		%	3 « super-classes »		
			A	B	C
	1	18.3	100	0	0
	2	50.4	100	0	0
	3	4.6	0	100	0
	4	5.5	0	100	0
	5	4.8	0	0	100
	6	6.4	0	0	100
	7	4.1	0	100	0
	8	1.9	100	0	0
	9	1.9	0	0	100
	10	2	0	0	100
Total		100.0	70.6	14.2	15.2
Scores moyens par domaine et total	CLOGT		0.1	0.3	1.2
	PLOGT		0.4	0.7	0.9
	ENVIR		0.6	0.9	0.7
	DURAB		0.1	1.2	0.6
	PRIV		0.6	2.9	1.2
	TOT4		2.2	7.8	5.3
Revenu moyen par u.c.	REVUC		8378	5240	6518
Pauvreté monétaire	oui		6.1	24.6	19.0

par construction ; ce n'est pas le cas des super-classes issues d'une grille. C'est pourquoi ici on les désigne par des lettres, pour éviter une interprétation trop rapide.

	non		93.9	75.4	81.0
Aisance (de « vivre très	1		2.0	21.2	8.0
difficilement » à	2		9.6	26.1	12.2
« vivre	3		27.6	36.8	32.3
très aisément »)	4		44.6	13.8	36.8
	5		16.1	2.1	10.7

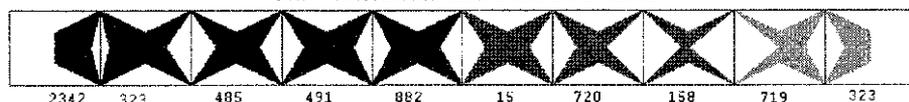
On a donc ici un classement qui suggère une dissociation entre les aspects « niveau de vie » et les aspects « conditions de vie », et que nous interprétons comme la manifestation du caractère particulier des éléments manquants au confort du logement (*cf.* l'analyse des modalités). De ce point de vue, ce pourrait être la classification la plus intéressante de celles obtenues jusqu'ici sur le plan qualitatif.

5.2.2 KACP des coordonnées et ficelle

Ici, on a classé les observations sur une « ficelle », c'est-à-dire implicitement que l'on fait l'hypothèse d'une structure unidimensionnelle approximative de la distribution des observations. Les 10 classes de Kohonen sont ensuite regroupées, comme précédemment, en 3 super-classes (graphique 5).

Graphique 5 – Ficelle de Kohonen avec KACP à partir des coordonnées des observations

5.a – Distances entre les 10 classes



Classement des ménages (représentés par leurs coordonnées calculées après une analyse des correspondances multiples) sur une ficelle à 10 classes et regroupement en 3 super-classes grisées. Les nombres correspondent aux effectifs des classes. Ce classement permet d'attribuer aux ménages une note de 1 à 10

5.b – Profils des 10 classes



Table 5- Dix classes résultant d'une classification KACP des coordonnées sur une ficelle

10 classes		%	3 « super-classes »		
			1	2	3
1		36.3	100	0	0
2		5.0	100	0	0
3		7.5	100	0	0
4		7.6	100	0	0
5		13.7	100	0	0
6		0.2	0	100	0
7		11.1	0	100	0
8		2.4	0	100	0
9		11.1	0	0	100
10		5.0	0	0	100
Total		100.0	70.0	13.8	16.1
Scores moyens par domaine et total	CLOGT		0.2	0.6	0.3
	PLOGT		0.4	0.6	0.7
	ENVIR		0.4	0.7	1.7
	DURAB		0.3	0.5	0.6
	PRIV		0.8	1.4	1.9
	TOT4		2.5	4.7	6.6
Revenu moyen par u.c.	REVUC		7996	6469	7160
Pauvreté monétaire	oui		9.1	11.7	16.7
	non		90.9	88.4	83.3
Aisance (de « vivre très difficilement » à « vivre très aisément »)	1		4.0	4.2	14.2
	2		10.6	16.1	16.8
	3		27.4	41.4	29.6
	4		42.6	33	28.5
	5		15.4	5.39	10.9

Les super-classes 1 à 3 représentent respectivement 70 %, 14 % et 16 % des observations ; c'est avec ce classement que l'on obtient la plus forte proportion de la (ou des) classes la (les) plus défavorisée(s) du point de vue des conditions de vie.

C'est aussi la classification dans laquelle les classes se différencient le moins du point de vue du revenu par u.c. moyen, bien que le taux de pauvreté soit croissant de la classe 1 à la classe 3 (ce qui indique une plus forte hétérogénéité des revenus par u.c. ; notamment, la classe 3 combine un revenu moyen supérieur à celui de la classe 2, et un taux de pauvreté nettement plus élevé). Le score total est croissant de la super-classe 1 à la super-classe 3. Comme dans la classification précédente, on

trouve (entre les super-classes 2 et 3) une opposition entre conditions de logement (score le plus élevé dans la classe 2), et conditions dans les autres domaines (en particulier environnement et privations pour la classe 3).

5.3. Synthèse des résultats

A l'issue de ces classements, on constate donc à la fois que la proportion de ménages dont les conditions de vie peuvent être qualifiées, du point de vue des domaines retenus, comme « défavorables » apparaît assez variable, et que les autres caractéristiques des classes obtenues - en particulier la pauvreté monétaire - ne présentent pas systématiquement une hiérarchie comparable.

La table 6 présente une synthèse des différentes classifications : elle reprend la proportion des classes obtenues, et indique les principales surreprésentations observables sur les différents descripteurs, auxquels on a ajouté les caractéristiques relatives au type de logement, à la zone d'habitation et au type de ménage.

L'examen des descripteurs en termes de surreprésentation fait apparaître des différences sensibles entre les caractéristiques des classes selon la méthode de classification :

- les classifications pour lesquelles on utilise en entrée les scores partiels font apparaître en général une gradation assez claire dans les conditions de vie, qui se rangent, par construction, du « moins défavorable » au « moins favorable » ; toutefois, la classification Kohonen fait apparaître des nuances que la classification hiérarchique traditionnelle ne restitue pas entre les différents domaines.

- les classifications basées sur les coordonnées des observations après ACM fournissent de notre point de vue des résultats très intéressants : en particulier, le manque de confort élémentaire dans le logement (éventuellement associé aux problèmes du logement) distingue presque toujours une classe, qui du point de vue du score total, du taux de pauvreté, ou même de l'aisance subjective n'apparaît pas comme la plus « défavorisée ».

Avec la construction de plus de deux classes, et surtout avec la possibilité d'utiliser l'information à un niveau assez détaillé, les classifications font apparaître des associations ou des particularités qui ne se révèlent pas avec l'utilisation du score unique. Ici, c'est le cas notamment du domaine confort du logement. On peut avancer l'idée que la grande diffusion d'un confort minimum du logement fait de ceux qui n'en bénéficient pas une classe nettement à part, même s'ils ne sont pas systématiquement « très » défavorisés aussi selon d'autres critères.

Pour les autres caractéristiques (type de logement, zone d'habitation, type de ménage), on constate également des sur-représentations nettes, qui sont plus ou moins accentuées mais qui se distribuent de façon assez cohérente quelle que soit la

technique de classement utilisée : en général, les conditions les moins favorables sont associées à des fréquences plus élevées qu'en moyenne de logements de type collectif (modalités 3 et 4 de la variable LOGT). La taille des unités urbaines et leur caractère rural ou urbain (variable TUR) ne fait pas, en général, apparaître de différences massives entre les classes obtenues ; il est toutefois intéressant de noter que dans la dernière classification, la classe 3, très sous-représentée en zone rurale et nettement sur-représentée dans les villes de grande taille et en agglomération parisienne (respectivement modalités 0, puis 3 et 4) est également celle où le score partiel ENVIR relatif à la (mauvaise) qualité de l'environnement (bruit, pollution, etc.) est le plus élevé par rapport à la moyenne.

Enfin, en ce qui concerne les types de ménages (variable TYM), deux catégories se distinguent du côté des classes les moins favorisées en termes de conditions de vie : d'une part, les personnes vivant seules (modalité 0), et d'autre part les familles monoparentales (modalité 3). Par contre, la différenciation semble faible sur la taille du ménage (NBTOT) et l'âge moyen des adultes (AGEM), un peu plus sensible quant au nombre d'enfants (NB17).

Table 6 – Synthèse des classements

	Score base		Classes sur les scores partiels						Kcap sur les coordonnées						
			Class. hiérarchique			Kcap			Grille+super-classes(*)			Ficelle+super-classes			
	0	1	1	2	3	1	2	3	A	B	C	1	2	3	
Distribution %	89.2	10.2	60.7	24.6	14.8	77.6	10.9	11.5	70.6	14.2	15.2	70.0	13.8	16.1	
Rapport des moyennes et proportions par classe aux moyennes et proportions dans l'ensemble															
CLOGT	0.7	3.2	0.5	1.2	2.5	0.5	1.8	3.9	0.3	1.0	4.4	0.7	2.3	1.1	
PLOGT	0.8	2.9	0.8	0.6	2.5	0.4	4.3	2.0	0.7	1.5	1.7	0.8	1.2	1.5	
ENVIR	0.9	2.2	0.9	0.8	1.7	0.8	2.2	1.2	0.9	1.4	1.1	0.6	1.1	2.6	
PRIV	0.7	3.7	0.2	1.6	3.4	0.7	1.2	2.7	0.6	2.9	1.2	0.7	1.3	1.9	
DURAB	0.7	3.2	0.1	1.7	3.4	0.4	0.8	5.6	0.4	3.4	1.7	0.8	1.3	1.8	
TOT4	0.7	3.2	0.4	1.2	2.9	0.7	1.9	2.4	0.6	2.2	1.5	0.7	1.3	1.9	
REVUC	1.0	0.6	1.2	0.8	0.6	1.1	0.8	0.6	1.1	0.7	0.9	1.0	0.8	0.9	
POOR	1	0.7	3.2	0.4	1.3	3.1	0.6	1.4	3.2	0.6	2.3	1.8	0.9	1.1	1.6
	2	1.0	0.7	1.1	1.0	0.7	1.0	1.0	0.7	1.1	0.8	0.9	1.0	1.0	0.9
SLS	1	0.5	5.4	0.1	0.9	4.8	0.5	1.3	4.1	0.4	3.7	1.4	0.7	0.7	2.5
	2	0.8	2.3	0.4	1.5	2.4	0.8	1.4	2.0	0.8	2.1	1.0	0.9	1.3	1.4
	3	1.0	1.1	0.7	1.5	1.1	0.9	1.3	1.2	0.9	1.2	1.1	0.9	1.4	1.0
	4	1.1	0.2	1.3	0.7	0.2	1.1	0.8	0.4	1.1	0.4	0.9	1.1	0.8	0.7
	5	1.1	0.1	1.5	0.4	0.1	1.2	0.6	0.2	1.2	0.2	0.8	1.2	0.4	0.8
LOGT	1	1.0	0.6	1.1	0.9	0.7	1.1	0.6	0.7	1.1	0.7	0.9	1.1	0.9	0.7
	2	1.0	1.0	1.0	1.1	1.0	1.0	1.1	1.0	1.0	1.0	1.0	1.0	1.1	1.0
	3	0.9	1.4	0.9	1.0	1.3	0.9	1.5	1.4	0.9	1.2	1.4	0.9	1.1	1.2
	4	1.0	1.4	0.9	1.0	1.3	0.9	1.3	1.2	0.9	1.4	1.0	0.9	1.0	1.4
	5	1.0	1.3	1.0	1.1	1.1	0.9	1.1	2.0	0.8	0.7	2.3	1.0	1.2	0.8
TUR	0	1.0	0.9	1.0	1.1	1.0	1.0	0.9	1.1	1.0	0.9	1.1	1.1	1.0	0.6
	1	1.0	0.8	1.1	1.0	0.8	1.0	0.9	0.8	1.0	0.9	1.0	1.0	0.8	1.0
	2	1.0	1.0	0.9	1.1	1.0	1.0	1.0	1.0	1.0	1.1	1.0	1.0	1.1	1.0
	3	1.0	1.2	1.0	0.9	1.1	1.0	1.0	1.1	1.0	1.1	0.9	0.9	1.0	1.3
TYM	0	0.9	1.5	0.8	1.2	1.4	0.9	1.0	1.6	0.9	1.2	1.4	0.9	1.3	1.1
	1	1.1	0.5	1.1	0.9	0.6	1.1	0.8	0.6	1.1	0.7	0.9	1.0	0.8	0.9
	2	1.0	0.8	1.1	0.8	0.8	1.0	1.1	0.6	1.1	0.9	0.8	1.1	0.8	0.9
	3	0.9	1.9	0.6	1.3	2.0	0.9	1.1	1.9	0.9	1.9	0.8	0.8	1.3	1.5
NB17	0	1.0	0.9	1.0	1.1	1.0	1.0	0.9	1.3	1.0	1.0	1.1	1.1	0.9	0.6
	1	1.1	0.5	1.1	0.9	0.6	1.1	0.8	0.6	1.1	0.7	0.9	1.0	0.8	0.9
	2	1.0	0.8	1.1	0.8	0.8	1.0	1.1	0.6	1.1	0.9	0.8	1.1	0.8	0.9
	3	0.9	1.9	0.6	1.3	2.0	0.9	1.1	1.9	0.9	1.9	0.8	0.8	1.3	1.5
NBTOT	1.0	1.0	1.0	0.9	1.0	1.0	1.1	0.9	1.0	1.0	0.9	1.0	0.9	1.0	
AGEM	1.0	1.0	1.0	1.0	1.0	1.0	0.9	1.0	1.0	1.0	0.9	1.0	1.0	1.0	
NB17	1.0	1.1	1.0	0.9	1.1	1.0	1.3	0.9	1.0	1.1	0.9	1.0	0.9	1.0	

Perspectives

Au total, l'intérêt des classifications réside dans la possibilité de construire plus de deux classes, qui sont « coupées » indépendamment du recours à un indicateur exogène ou à l'arbitraire de l'auteur. Une part d'arbitraire subsiste néanmoins dans le choix du nombre de classes (ici, on a choisi la taille de la grille et celle des ficelles, ainsi que le nombre de super-classes, ce qui pèse évidemment sur la part des observations que l'on obtient *in fine* dans chaque classe), mais il est toujours possible de produire des variantes dans lesquelles on retiendra un nombre plus ou moins grand de classes.

Ces premiers résultats invitent à tester mieux le caractère opérationnel des classements obtenus ; une première direction d'approfondissement consistera à utiliser la même méthode en utilisant des données plus riches (enquêtes Conditions de vie des ménages). Il sera également intéressant de reconduire les applications à diverses dates d'enquêtes, afin de comparer au cours du temps les évolutions respectives des « taux » de pauvreté en termes monétaires et en termes de conditions de vie.

Par ailleurs, on peut envisager d'introduire des variables supplémentaires permettant, par analogie avec les échelles d'équivalence employées pour la mesure de la pauvreté monétaire, de « relativiser » les situations en tenant compte, pour la construction des classes, des contextes (habiter en zone rurale ou urbaine par exemple), et de la composition des ménages (taille, âges, enfants, etc.). Enfin, si l'on considère que l'absence d'un système de pondération des items constitue un problème fondamental, on pourrait tester l'utilisation d'une classification des modalités pour bâtir une pondération.