

BRUYNOOGHE

**La distance aux équipements urbains en
Languedoc-Roussillon : analyse de correspondance
et classification automatique pour un
ensemble de 1561 communes**

Les cahiers de l'analyse des données, tome 3, n° 1 (1978),
p. 35-46

http://www.numdam.org/item?id=CAD_1978__3_1_35_0

© Les cahiers de l'analyse des données, Dunod, 1978, tous droits réservés.

L'accès aux archives de la revue « Les cahiers de l'analyse des données » implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques

<http://www.numdam.org/>

LA DISTANCE AUX ÉQUIPEMENTS URBAINS
EN LANGUEDOC-ROUSSILLON :
ANALYSE DE CORRESPONDANCE
ET CLASSIFICATION AUTOMATIQUE
POUR UN ENSEMBLE DE 1 561 COMMUNES
[COMMUNES]

Le présent article montre sur un exemple de géographie* l'efficacité d'un nouvel algorithme rapide de classification des grands ensembles de données ; algorithme décrit dans ce même cahier (cf pp 7-33).

1 Le problème géographique

Dans les travaux géographiques antérieurs, les hiérarchies urbaines et les zones d'influence des villes, étudiées à partir d'indicateurs mesurant la densité des services urbains ou l'intensité de leur attraction (cf enquête PIATTIER), se traduisent par des découpages spatiaux qui explicitement, ou même implicitement, se réfèrent à un point de vue urbain.

Le point de vue ici choisi est diamétralement opposé : nous mesurons la distance physique à parcourir, au minimum, pour atteindre divers équipements. C'est là un premier paramètre de la distance vécue, telle que les habitants d'une région la subissent à travers la présence ou l'absence d'équipements, la proximité ou l'éloignement par rapport à la ville. Nous nous proposons d'étudier la situation géographique des communes selon la proximité ou l'éloignement à divers équipements urbains, des plus élémentaires (bureau de poste), aux plus rares et aux plus composites (ceux qu'offre une grande ville seule). Ainsi, le point de vue adopté est "rural" et non pas "urbain".

Le problème géographique posé consiste donc uniquement en une recherche des différenciations spatiales des communes selon les distances aux équipements existants qui se combinent, souvent contradictoirement, selon les lieux.

La classification doit permettre d'établir une description cartographique de la région de programme Languedoc-Roussillon qui soit un constat de la situation actuelle du point de vue de l'accessibilité à divers équipements urbains, commune par commune et non pas une représentation de l'adéquation ou de l'inadéquation des besoins aux équipements actuels qui ne pourraient être affirmées ou mesurées qu'à partir d'enquêtes auprès des populations concernées.

2 Les données et leur codage

2.a Les 14 distances kilométriques

Pour chaque commune de la région Languedoc-Roussillon, ont été retenues les distances (km) à 14 types d'équipements urbains dont on avait pu constater, après l'étude déjà citée (cf Note 2 *in fine* ; & DUPONT 1975), qu'ils étaient significatifs de plusieurs niveaux.

(1) Université d'Aix-Marseille II ; et C.R.E.T. ; Centre de recherche en économie des transports ; Avenue Gaston Berger ; 1310C Aix en Provence.

* sur les conditions précises du présent travail cf Note *in fine*.

- un niveau élémentaire avec 2 équipements : bureau de poste et arrêt quotidien d'un moyen de transport collectif (car ou train) ;
- un second degré manifesté par 2 autres équipements : pharmacie et école maternelle ;
- 4 équipements (dentiste, cinéma, guichet de banque, C.E.G. ou C.E.S.) assurant la transition vers les gros bourgs, voire les petits centres ;
- 2 équipements - scolaire (lycée) et hospitalier (hôpital ou clinique) caractérisant la ville ;
- 4 distances permettant de hiérarchiser la gamme des services urbains : distance à la "petite ville" (celles définies comme telles, à deux ou trois exceptions près, dans les travaux de l'Atlas régional du Languedoc-Roussillon) ; distance à la ville moyenne ; distance à la grande ville ; distance au chef-lieu de département (1).

2. b Le codage logique des données

Le codage retenu tient compte de la substance à traiter. Les 14 distances présentent de grandes variabilités de proportion : le rapport entre les distances à deux équipements peut varier du simple au décuple, voire au centuple. Pour réduire à l'unité le tableau des données, équilibrer l'importance des différents équipements et obtenir des résultats indépendants de toute modélisation de la distance vécue telle que se la représentent les usagers des équipements urbains, on a employé un codage logique en substituant à chaque échelle de mesure une partition en classes consécutives de poids voisins : au lieu de noter une valeur continue, on ne considère que l'appartenance à une classe (cf J.P. Benzécri; l'Analyse des Données, TII A n° 2, § 1.3, p 23).

Pour chaque équipement, on a construit l'histogramme de l'ensemble des valeurs des distances de chaque commune à l'équipement considéré et on a divisé l'intervalle de variation en 5 intervalles consécutifs contenant environ le cinquième des communes.

Le tableau en 14 variables de valeurs absolues kilométriques est donc transformé en un tableau booléen de 70 variables logiques, chaque variable distance ayant été transformée en 5 modalités : distance très faible, faible, moyenne, élevée, très élevée, notées e^- , e^- , e , e^+ , e^{++} pour chaque équipement e .

(1) Définition des villes petites, moyennes et grandes (cf Atlas régional du Languedoc-Roussillon) :

PETITES VILLES :

- . Aude : Castelnaudary, Lézignan-Corbières, Limoux, Quillan (Ax-les-Thermes, Lavelanet) ;
- . Gard : Beaucaire, Le Vigan, Pont-St-Esprit, Sommières, Uzès (Orange) ;
- . Hérault : Bédarieux, Clermont-l' Hérault, Ganges, Lodève, Lunel, Pézenas, Saint-Pons, (Mazamet) ;
- . Lozère : Langogne, Marvejols, Mende, St-Chely (Le Puy, Millau) ;
- . Pyrénées-Orientales : Céret, Ille-sur-Têt, Prades, Puigcerda.

VILLES MOYENNES :

- . Alès, Carcassonne, Narbonne, Sète (Arles, Castres, Rodez).

GRANDES VILLES :

- . Béziers, Montpellier, Nîmes, Perpignan (Avignon, Toulouse)

Les villes dont les noms sont mis entre parenthèses sont extérieures à la région de programme Languedoc-Roussillon).

Voici sur le tableau n° 1 ci-dessous, les valeurs des bornes pour chaque équipement, exprimées en kilomètres.

	e	Borne	e ⁻	Borne	e ⁻	Borne	e	Borne	e ⁺	Borne	e ⁺	Borne
Arrêt transport	TR	0		0		3		5		9		30
Poste	PT	0		0		3		5		8		22
Maternelle	MA	0		0		3		7		12		50
Pharmacie	PH	0		0		4		6		10		33
Dentiste	DT	0		3		7		11		17		50
Cinéma	CI	0		3		7		11		17		50
Collège	CC	0		4		7		11		16		50
Banque	BA	0		4		7		11		17		50
Lycée	LY	0		9		14		20		32		60
Hôpital	HP	0		10		17		24		38		55
Petite ville	PV	0		8		12		17		24		65
Ville moyenne	VM	0		17		27		40		57		120
Grande ville	GV	0		24		41		61		89		215
Chef-lieu	CF	0		23		36		49		66		143

Tableau n° 1 - Classes logiques et bornes

3 Justification du codage et méthode d'analyse factorielle

On pourrait songer à l'utilisation d'un "modèle gravitaire" pour définir l'accessibilité aux équipements et analyser le tableau des "accessibilités gravitaires" plutôt que le tableau des distances physiques. On s'intéresserait alors à la "distance perçue" par les usagers des différents équipements urbains et non plus à la "distance vécue" telle que les habitants la subissent.

L'accessibilité gravitaire, définie comme une fonction décroissante de la distance à un équipement, de forme puissance ou exponentielle ($d^{-\beta}$; $\exp(-\alpha d)$) est une première représentation de la distance perçue, l'une des plus simples puisqu'elle ne différencie pas les multiples composantes de la population susceptible de fréquenter les divers équipements urbains.

L'analyse factorielle en composantes principales du tableau des "accessibilités gravitaires" diffère de celle du tableau des distances physiques et présuppose une modélisation du comportement des usagers. De plus, l'estimation du paramètre de la fonction d'accessibilité à chaque équipement peut sembler délicate en l'absence d'enquêtes quant à la fréquentation des divers équipements urbains.

Par contre, l'analyse factorielle des correspondances des données codées sous forme logique disjonctive est indépendante de toute modélisation *a priori* du comportement des usagers et de toute représentation de la distance perçue, puisque le codage logique des distances aux équipements, qui substitue à chaque échelle de mesure une partition en classes consécutives équiprobables, est invariant par une transformation monotone des distances physiques : les inégalités que vérifient les accessibilités à un équipement ne sont pas sensibles au mode de leur calcul.

Le codage par rang qui substituerait également à chaque distance une partition en classes consécutives équiprobables, présente également cette propriété d'invariance et nous nous proposons de l'expérimenter ultérieurement. Nous nous contentons ici de soumettre le tableau des données, codées sous forme logique disjonctive, à l'analyse factorielle des correspondances.

4 Expérimentation et méthode d'interprétation

La classification hiérarchique des 1561 communes, selon le critère de la variance, a été construite à partir du tableau des facteurs issus de l'analyse factorielle des correspondances, afin de ne retenir de l'information initiale qu'une structure pertinente, en éliminant toute fluctuation aléatoire éventuelle ou liée à l'imprécision du découpage en cinq classes.

Les sept premiers facteurs ont été retenus pour construire la hiérarchie sur l'ensemble des communes afin de prendre en compte *a priori* une part notable de l'information contenue dans le tableau des données.

Pour construire en toute rigueur la classification à partir du tableau des facteurs, on n'aurait dû retenir que les facteurs interprétables, cependant étant donné le nombre élevé de communes à analyser, seule la représentation graphique des classes issues de l'analyse hiérarchique permettait l'interprétation des axes factoriels autres que le premier qui définit un gradient "urbain-rural" et dont l'interprétation était immédiate. L'exemple du septième facteur est typique et la dénomination des classes qu'il oppose de façon évidente ne pouvait être perçue qu'après leur représentation cartographique et dès lors s'éclairait la disposition des variables le long du septième axe.

Le tableau des contributions des classes aux facteurs et la représentation cartographique sont indispensables pour l'interprétation de chacun des axes puisque sont mises en évidence, d'une part, les classes qui contribuent le mieux à l'explication et, d'autre part, leur cohérence spatiale.

De cette manière, le géographe retrouve la dimension spatiale inhérente à sa démarche habituelle et la structure d'ensemble organisée selon les facteurs résulte autant des relations spatiales que des relations entre variables.

Il n'est donc pas aussi aisé de fixer *a priori* le nombre de facteurs à retenir, sans classification préalable et sans représentation cartographique. Seules l'analyse et l'interprétation géographiques permettent de justifier *a posteriori* le nombre de facteurs retenus.

Du niveau de coupure de l'arbre hiérarchique dépendent les possibilités d'interprétation d'un axe factoriel. Il est bon d'avoir toute facilité de sectionner l'arbre à divers niveaux jugés pertinents pour le géographe, la pertinence n'étant pas seulement d'ordre statistique. Après avoir défini 120 classes initialement et retenu 8 grandes classes à interpréter, il a fallu subdiviser quelques-unes de celles-ci afin d'explicitier certains facteurs.

Un aller-retour permanent est donc nécessaire entre examen des plans factoriels et étude du contenu des classes déterminées par coupures de l'arbre.

C'est par la complémentarité de l'analyse factorielle, de la classification hiérarchique et de la représentation cartographique qui se conjuguent, que l'analyse d'un vaste ensemble de données géographiques peut être conduite.

5 Une classification des communes en Languedoc-Roussillon [2]A - L'espace urbanisé

- La classe 220 comprend tous les types de communes urbanisées, des gros bourgs aux grandes villes. Une analyse des sous-classes qui la composent permet de hiérarchiser les niveaux urbains.

- La classe 223 s'articulant par quelques bourgs à la classe précédente, recouvre tout l'espace *péri-urbain équipé*. Absent de Lozère et de la re-tombée du Massif Central et des Pyrénées, ce type possédant les équipements de base, proche des équipements intermédiaires, est bien représenté par les bourgs viticoles de la plaine languedocienne ou roussillonnaise, puis dans les *interstices d'un maillage urbain dense*.

- La classe 221 représente un *espace péri-urbain de petits centres agricoles ou industriels*, qui émerge d'autant plus nettement que pèsent peu villes grandes et moyennes. S'éloignant de la plaine languedocienne, elle s'illustre sur le piémont du Massif Central, dans les Pyrénées ou en Lozère.

B - L'espace rural proche des équipements urbains

- La classe 229 définit un *espace tributaire des villes grandes ou moyennes*, par des communes qui, sous-équipées, sont dépendantes du noyau urbain central : Carcassonne, Perpignan, Béziers, Montpellier, Nîmes, Alès, Avignon...

- La classe 227 éloignée des villes grandes ou moyennes, elle est à *proximité relative de la petite ville* dont elle est tributaire, et souligne les axes urbains secondaires de l'espace régional : dorsale sur la limite des départements du Gard et de l'Hérault, axes audois, noyaux lozériens ou pyrénéens...

C - L'espace rural éloigné des équipements urbains

- La classe 212 constitue des communes de *relatif éloignement des villes* grandes, moyennes ou petites qui s'affirment autour des villes moyennes et petites de l'arrière pays et se fractionnent (aspect ponctuel) dans l'axe urbain de la plaine.

- La classe 222 s'éloigne des villes grandes ou moyennes et *ne se réfère plus qu'aux petites villes*.

D - L'espace rural lointain

Avec la classe 232, comble quelques lacunes de la plaine et ne dessine une lointaine influence urbaine qu'en zones faiblement urbanisées (Margeride, Causses, montagnes audoises)...

E - L'espace rural profond

Il n'est plus centré comme dans les cas précédents. Deux niveaux d'éloignement de tous les équipements ou services urbains se distinguent :

- La classe 225 concerne les *hautes garrigues*,

classe 228 couvre les *montagnes* où l'accessibilité urbaine est des plus défavorables.

6 Interprétation des résultats de l'analyse factorielle des correspondances

Le tableau n° 2 des coordonnées des classes et sous-classes confirme ce qui vient d'être dit sur la nécessité de recourir à la classification

et à la cartographie des résultats pour interpréter les facteurs.

Ce tableau condense bien, par ailleurs, les résultats obtenus. On imagine sans difficulté les degrés supplémentaires de finesse d'analyse si l'on désire redescendre à des niveaux plus bas soit sur l'ensemble de l'arbre hiérarchique, soit sur telle ou telle branche, puisque le programme de classification édite toutes les contributions des variables aux classes et des classes aux facteurs à partir de 120 noeuds initiaux.

Sous-classes	Légende carte	5 classes supérieures	Coordonnées des classes sur les facteurs							
			1	2	3	4	5	6	7	
220		A	617		518	-446	410			
223			705	-307	581	-363		-299		
221			430	-315	337			622	675	
229		B	685	348		465			-382	
227			251	617		574				
212		C	-460		-621	-486			-317	
222					-920	-484				
207		D		-953		464				
211			232		-595		447		332	-359
217				-653	-765		510			
225		E	-277	227	268		-568			
202			228	-895		299		-442		
210				-1114		346				

Tableau des coordonnées des principales classes avec leur représentation sur la carte.

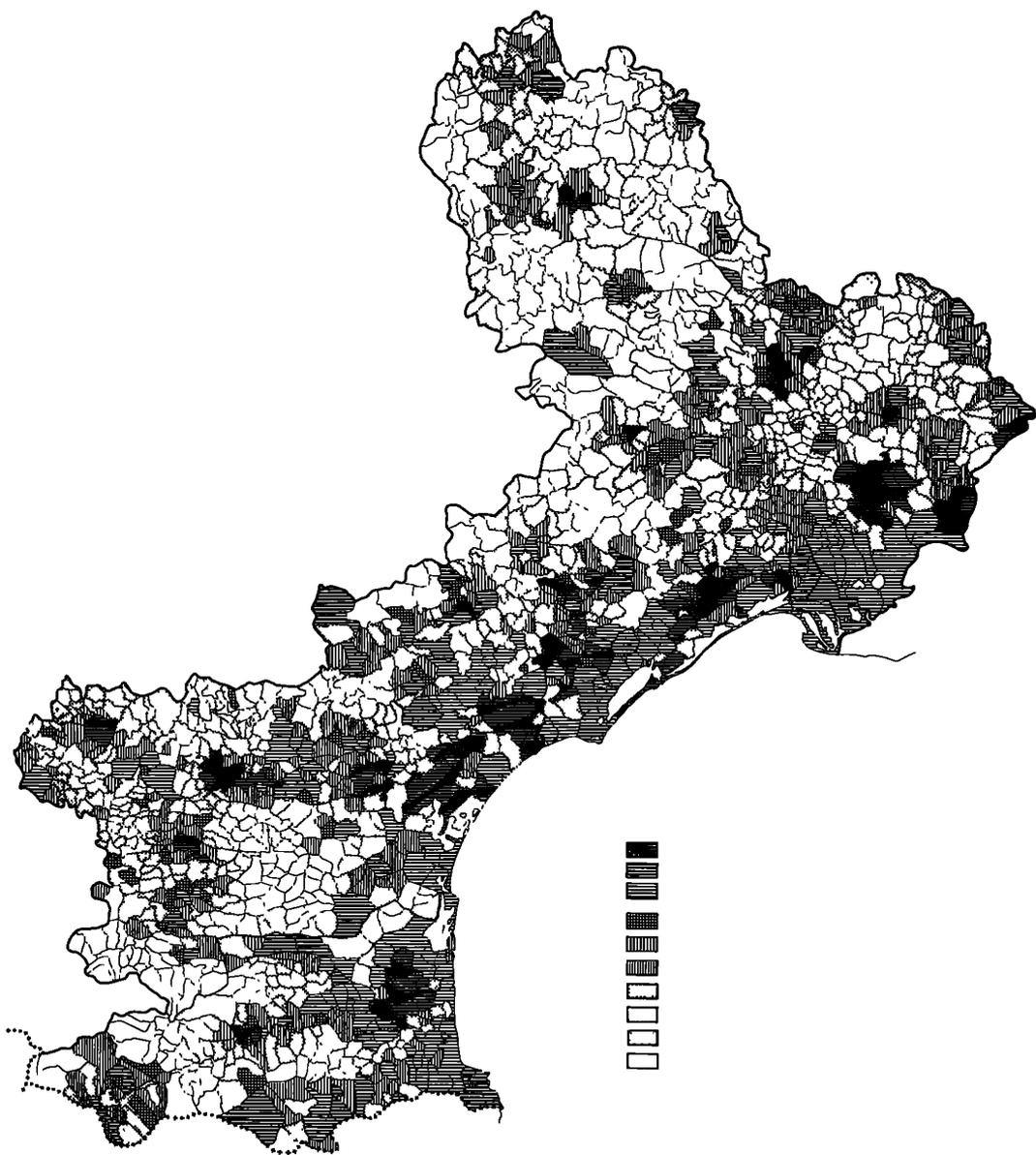
Le premier facteur définit un gradient "urbain-rural" et fournit une mesure synthétique de la distance à l'urbain.

Le second facteur fondé essentiellement sur l'opposition aux quatre équipements de niveau intermédiaire, différencie l'espace péri-urbain proche tributaire des équipements des villes, de l'espace rural encore défini par l'éloignement aux centres urbains et aux équipements qu'ils offrent, à l'exclusion de l'espace rural profond d'une part, et des communes urbanisées, d'autre part. Les classes 223 et 221 représentent des communes qui, par leur éloignement relatif des villes moyennes, grandes ou petites ont tendance à se rapprocher sur le second axe de l'espace rural lointain dans la mesure où elles ne sont pas tributaires des centres urbains.

Par effet Guttman, le troisième facteur oppose les deux types de classes extrêmes du premier facteur, au type intermédiaire. De nouveau, par effet Guttman, on note l'opposition sur le quatrième axe des distances faibles et élevées aux équipements intermédiaires, différenciées sur le second axe, à leurs distances moyennes.

Dans la définition du cinquième facteur, n'interviennent que les

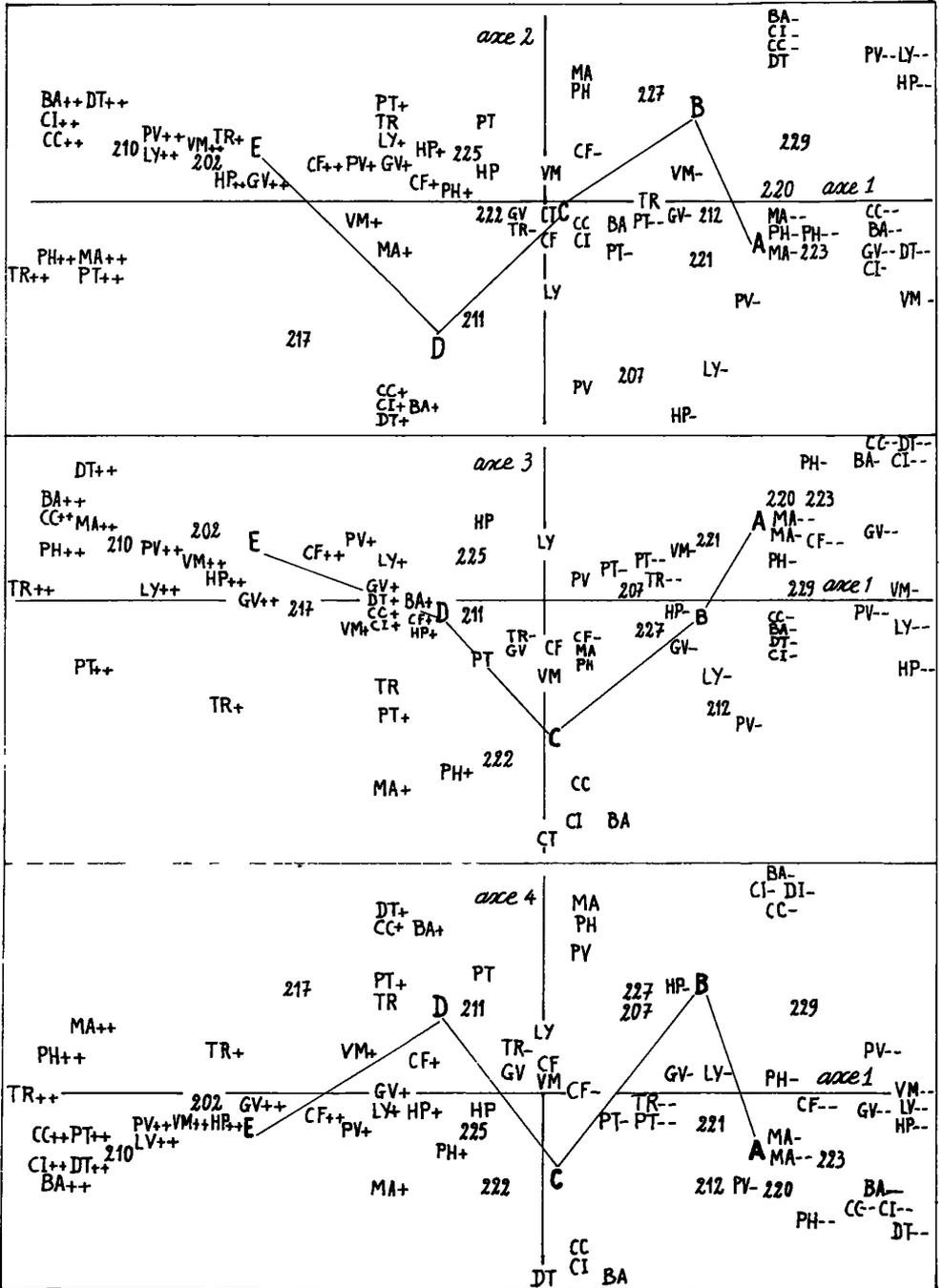
[COMMUNES]



Représentation cartographique des principales classes de communes.

L'explication des légendes est donnée au tableau de la page ci-contre. A la fin de l'article, on trouvera sur deux pages à une plus grande échelle la carte de la région divisée en deux parties empiétantes.

Analyse factorielle du tableau 1561 x 70 : 1561 communes par 70 modalités logiques : les notations du graphique sont rappelées à la page ci-contre



villes et espaces urbains (classe 220) opposés à l'espace rural profond (classes 225 et 202).

Le sixième facteur opère une différenciation dans le péri-urbain où il fait émerger en positif les communes proches de petits centres (classe 221) ou celles qui, dans l'espace rural lointain (classe 232) esquissent également une auréole de petit centre.

Le septième facteur, quant à lui, différencie à son tour les deux types précédents que regroupait le sixième facteur : la représentation cartographique révèle l'opposition de l'espace péri-urbain des petits centres agricoles ou industriels à une lointaine auréole urbaine par référence au chef-lieu, à la grande ville ou à la ville moyenne.

En consultant les graphiques ci-contre on prendra garde d'abord que le signe + signifie *éloignement*, non *présence* ! : ainsi BA++ signifie, grand éloignement au guichet de Banque ; DT-- signifie proximité au Dentiste ; PT-, proximité modérée à la Poste ; MA+ éloignement modéré à l'École Maternelle ; PH désigne une situation moyenne relativement à la Pharmacie ; et de même CC pour Collège ; LY pour Lycée, HP pour Hôpital etc. Les principales classes A,B,C,D,E (Urbain ; puis rural de proche à profond) ont été portées en éléments supplémentaires, ainsi que quelques sous-classes importantes désignées par leur numéro (e.g. 223, 229, 227 etc.). On observera que la ligne ABCD, en V sur le plan 1 x 3 (effet Guttman) est en S sur le plan 1 x 2 ; et en M (nombre maximum de traversées du 1° axe) dans le plan 1 x 4. Si l'interprétation a été surtout guidée par l'examen sur la carte géographique des classes de communes ; c'est que les associations de modalités de distances aux équipements, ne nous sont pas toutes bien connues. Prenons quelques exemples : on savait *a priori* (cf § 2.a) que les quatre équipements {guichet de Banque ; Dentiste ; Cinéma ; Collège} s'associent communément ; et l'on voit en effet dans le plan 1 x 2 les groupements : {BA--;DI--;CI--;CC--} ; {BA-;DT-;CI-;CC-} ; {BA+ ; DI+;CI+;CC+} ; {BA++;DT++;CI++;CC++} ; le groupe {PV--,LY--,HP} était attendu, lui aussi, car Lycée et Hôpital avaient été choisis comme des caractéristiques de la Ville ; il est remarquable que les points PH-- et MA-- présence de Pharmacie et d'école Maternelle ne se séparent guère du 1° jusqu'au 7° axe ; de même les points PT-- et TR-- (Poste et arrêt de Train ou car) ne se séparent pas. Mais l'opposition apparue sur le 2° axe (plan 1 x 2 ; Fl < 0) entre {BA++;CI++;DT++;CC++} d'une part et {TR++;PH++;MA++;PT++} de l'autre, i.e. entre l'éloignement maximum à deux classes d'équipement de base dont les premiers répartis suivant un réseau moins serré que les seconds, apprend déjà plus. Et l'on gagnerait sans doute à considérer toutes les associations de modalités intermédiaires (e.g. tel x+ avec tel y-...) en faisant une classification hiérarchique sur l'ensemble des 70 modalités d'éloignement.

Deux remarques suffiront à conclure :

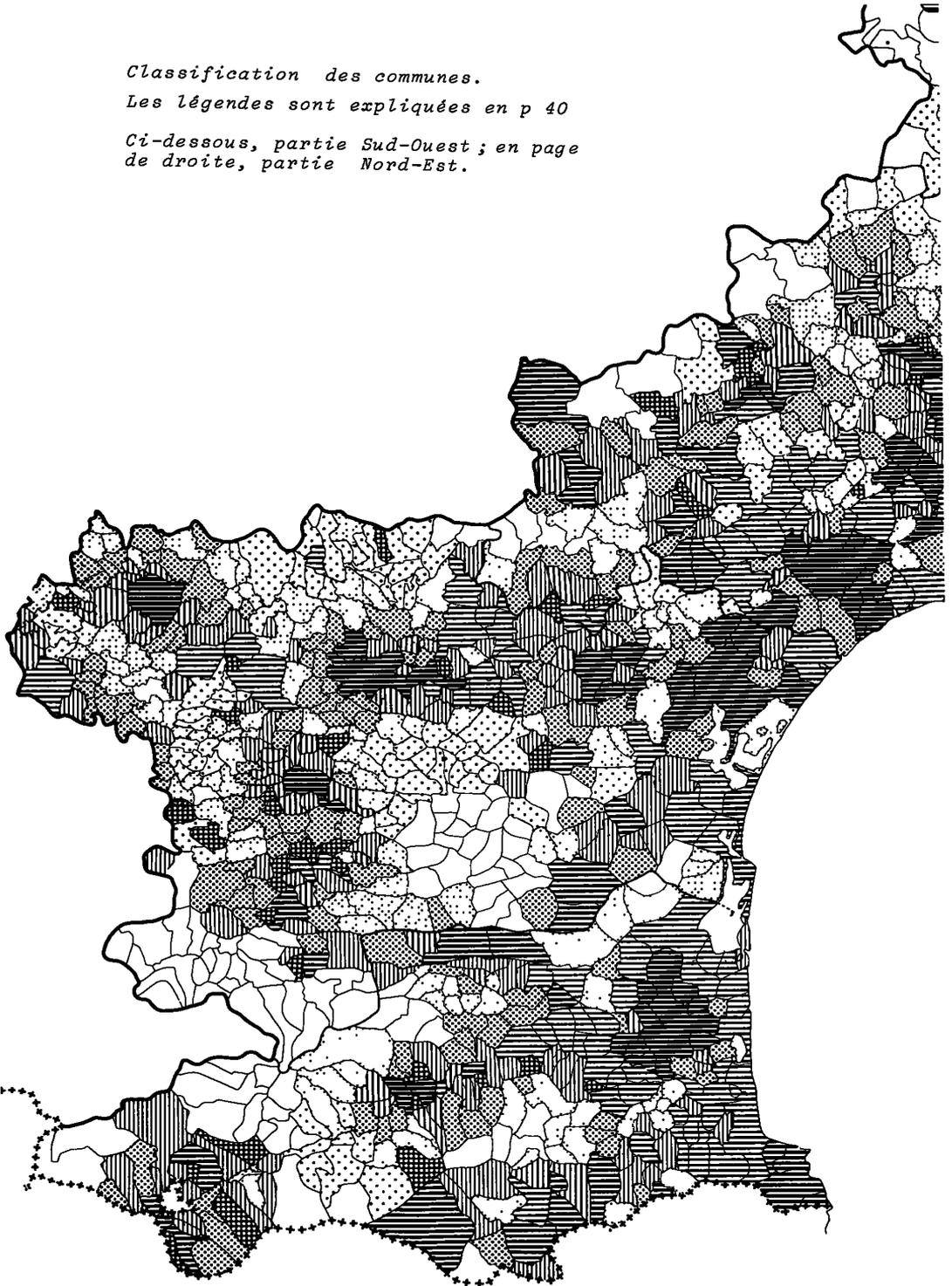
Habitué à voir dans l'espace une organisation urbaine longitudinale, les géographes n'étaient pas prêts à voir surgir avec autant de netteté la complexité du fait urbain gardois ou bien la disposition urbaine audoise profondément différente de celle du département de l'Hérault.

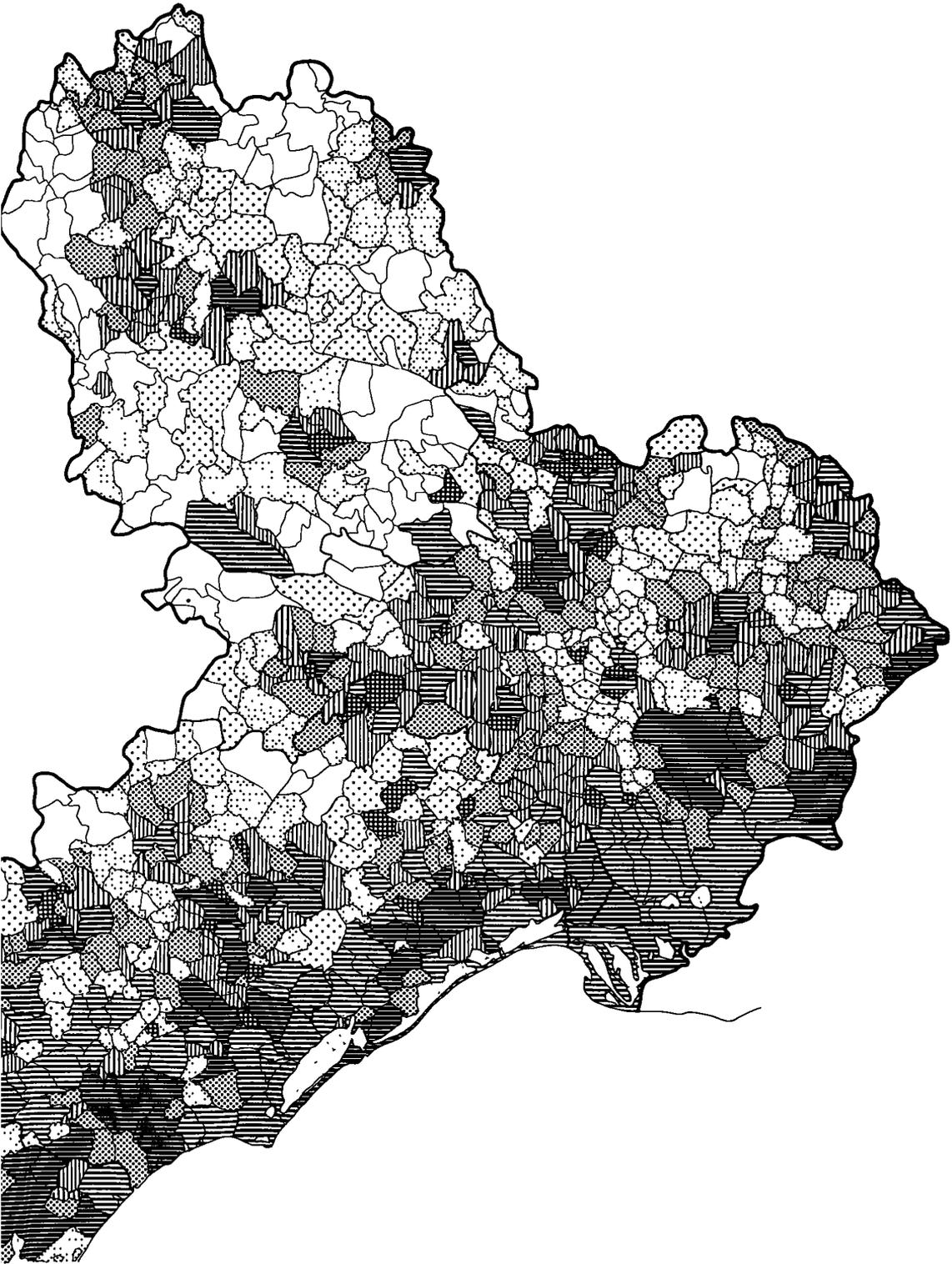
La nouvelle méthode de classification ascendante hiérarchique utilisée allie, à un temps de machine réduit, donc un coût modeste, de grandes facilités de lecture des résultats et permet ainsi l'analyse d'aires géographiques relativement vastes.

Classification des communes.

Les légendes sont expliquées en p 40

*Ci-dessous, partie Sud-Ouest ; en page
de droite, partie Nord-Est.*





BIBLIOGRAPHIE

AURIAC F., BERNARD M.C., FERRAS F.; VIGOUREUX M. (Groupe DUPONT) (1975). La distance à la ville: essais d'analyses factorielles appliquées aux cas de Grenoble et Montpellier. *L'ESPACE Géographique*, t. IV, n° 4, 225-238.

AURIAC F., BRUYNNOGHE M. (1976). Une classification sur plus de 1500 communes: la distance à la ville en Languedoc-Roussillon. Actes du 4° Colloque sur l'analyse des données en géographie. *Cahiers de Géographie de Besançon* (à paraître).

NOTE 1 : *Le travail présenté ici, résulte d'une collaboration interdisciplinaire: l'algorithme de classification par la "méthode des graphes réductibles" a été conçu par Michel BRUYNNOGHE (Université d'Aix-Marseille II, Centre de Recherche en Economie des Transports). Les applications géographiques sont le fruit d'un travail d'équipe à laquelle ont participé, à l'initiative du Professeur R. DUGRAND (Directeur de l'E.R.A. 242 du CNRS, Université de Montpellier), avec la collaboration de V. CABOS et T. PANOUILLERES (Atlas Régional du Languedoc-Roussillon), F. AURIAC, M.C. BERNARD, R. FERRAS, M. VIGOUROUX (Groupe DUPONT), M. BRUYNNOGHE.*

NOTE 2 : *Des essais d'analyses en composantes principales appliquées aux cas de Grenoble et Montpellier ont été réalisés par les géographes du groupe DUPONT pour tenter de définir une ou les distances synthétiques à la ville (cf. La distance à la ville; Groupe DUPONT, L'espace géographique, 1975. Les aires d'étude circonscrites autour de Grenoble (176 communes) et de Montpellier (150 communes) comprennent des communes très diversifiées allant de la banlieue immédiate jusqu'à la campagne exclusivement agricole et constituent deux milieux a priori bien différents offrant une grande diversité de formes d'activité et d'occupation de l'espace: "haute et moyenne montagne touristique, vallées industrielles et résidentielles, espaces ruraux ou caractérisés par une activité industrielle diffuse de l'aire grenobloise et basse plaine viticole, garrigue plus ou moins occupée, littoral touristique, moyenne vallée de l'Hérault à forte activité agricole, banlieues résidentielles discontinues de l'aire montpelliéraine".*