

APPLICATION DU LOGICIEL LISA A UN EXEMPLE DE TRAITEMENT D'ENQUETE : Analyse démographique des communes de la Haute Loire

Georges FRANCILLON*

RESUME

En analysant la démographie des communes d'un département français, la Haute-Loire, l'article se propose de montrer un certain nombre de facettes du Logiciel LISA (Logiciel Intégré pour les Systèmes Agraires), développé et diffusé par le D.S.A. depuis le premier Août 1988.

Après une présentation sommaire du Logiciel, l'étude du département est abordée à différents niveaux :

- Gestion des données.*
 - Analyse statistique élémentaire, tracé de cartes thématiques.*
 - Recherche d'une typologie : on enchaîne analyse factorielle classification automatique et étude statistique des classes obtenues...*
 - Pour terminer, on édite une carte synthétique des classes de communes ainsi qu'une carte agrégeant les zones homogènes.*
- LISA fera l'objet ultérieurement d'applications aux systèmes agraires.*

MOTS-CLES

Méthode - Analyse des données - Informatique - Logiciel - Démographie - Commune - France.

INTRODUCTION

Le logiciel LISA a été développé au DSA pour répondre aux besoins des chercheurs en matière de traitement d'enquête et d'analyse de données. L'hétérogénéité des logiciels existants (structures des fichiers, syntaxe des commandes...), leurs limites concernant le nombre, le type de variables, l'absence de gestion des données manquantes, à vérifier, leur évolution impossible à contrôler et leur coût à l'achat, nous ont conduit à créer nous-mêmes un logiciel « maison » construit autour d'un gestionnaire de fichiers performant, « le cœur du système », possédant aussi un module d'analyse statistique élémentaire, un module d'analyse multidimensionnelle très poussé et un module permettant de cartographier les résultats, faisant le lien indispensable entre l'analyse statistique et l'espace.

Le gestionnaire de fichiers permet une création, correction et mise en forme de l'information à analyser : c'est sans doute l'étape la plus longue et la plus pénible de l'analyse. Les modules statistiques et cartographiques permettent, grâce à leur inter-activité d'enchaîner simplement toute une série d'analyses conduisant par exemple à l'élaboration d'une typologie et à une représentation cartographique des classes obtenues.

Cette représentation, essentielle pour les systèmes agraires, sera l'objet de développements importants dans les versions ultérieures de LISA.

A travers l'étude démographique des communes du département de la Haute-Loire, nous voudrions montrer un certain nombre de facettes de ce logiciel. Ultérieurement, des études plus appliquées aux systèmes agraires seront présentées. Cependant, à travers l'exemple, le lecteur pourra se rendre compte des potentialités du logiciel et des domaines d'application envisageables...

* DSA - CIRAD

Présentation sommaire de LISA

LISA (Logiciel Intégré pour les Systèmes Agraires) est un outil informatique d'analyse, composé de 3 modules de base : gestion de fichiers, statistiques et infographie, sur lequel on peut greffer des «modules à la carte» : programmation linéaire, optimisation, simulation de systèmes agraires, outils de diagnostic (systèmes experts...) etc.

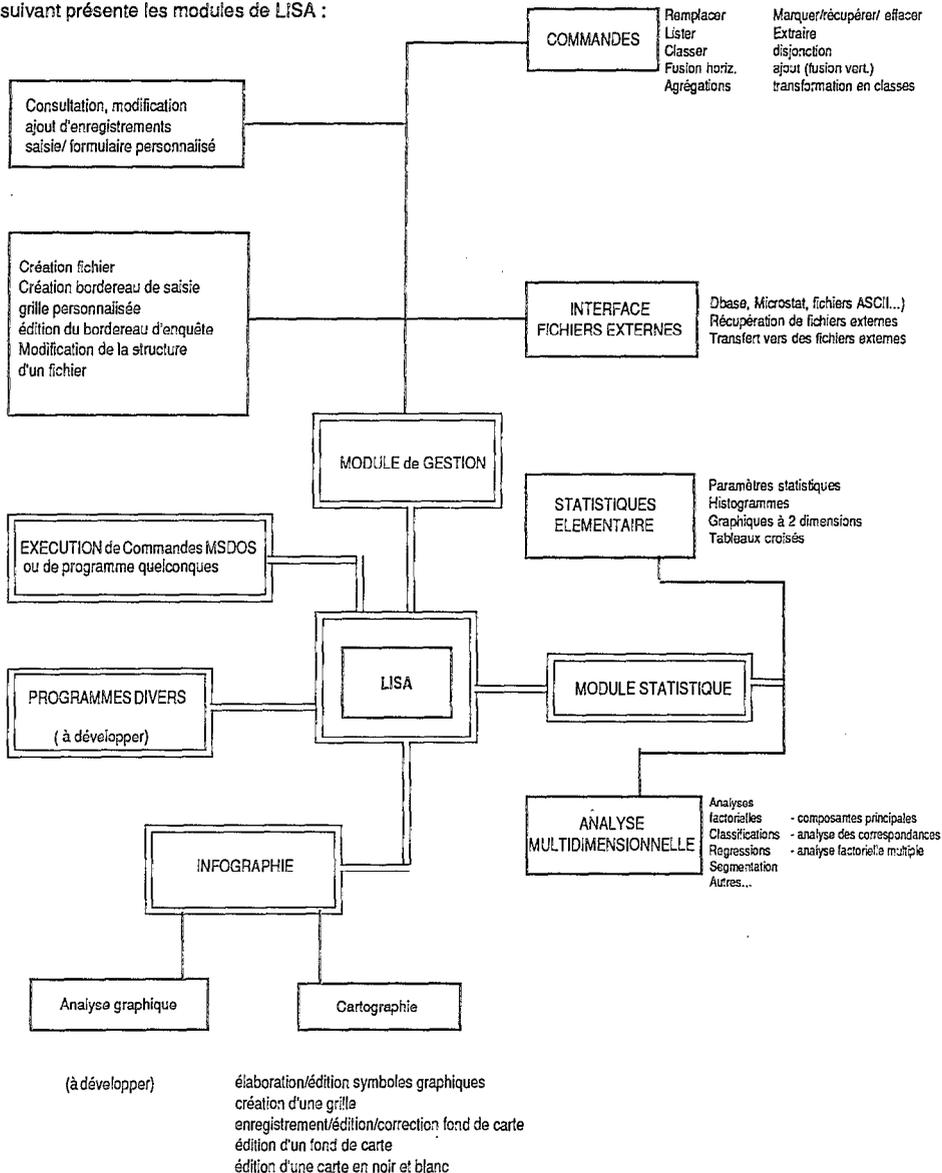
LISA s'appuie sur les travaux de l'ADDAD pour les statistiques, de l'université de Franche-Comté pour la partie carto, de l'école Bertin (EHESS de Paris) pour le traitement graphique de l'information. LISA est diffusé par le DSA dans le cadre et selon les règles du réseau ID animé par No Chang Bang (DSA), réseau qui regroupe les universités Paul Valéry, de Franche-comté, de Haute-Normandie, l'EHESS, la Maison de la Géographie, le DSA.

Portable, évolutif, adapté aux micros de base, interfacé aux structures de fichiers les plus communes (DBASE...) LISA est caractérisé par :

- la rapidité d'exécution (langages C et FORTRAN),
- son encombrement : environ 3 méga-octets sur disque dur ou 10 disquettes... (versions disque dur et disquettes), mémoire centrale nécessaire ≥ 256000 octets..., fonctionne sur MSDOS, adaptable facilement à d'autres systèmes (UNIX...)
- une gestion dynamique de la mémoire : adaptation à la mémoire disponible de l'ordinateur (pas de recompilation !), aucune limitation sur le nombre de variables et d'individus (sauf ordinateur lui-même),
- une gestion de variables caractères, qualitatives (avec leurs modalités) et quantitatives
- une gestion des données manquantes, à vérifier, à effacer (contrôle des limites)

la mise au point d'un interpréteur de formule arithmétique et logique faisant intervenir un ensemble évolutif d'opérateurs et de fonctions.

L'organigramme suivant présente les modules de LISA :



(On se reportera à la bibliographie pour une description plus précise de ce logiciel.)

I — PRESENTATION DE L'ETUDE

Ce département, dont la préfecture est «Le Puy», est constitué de 266 communes, en majorité rurales ou groupées autour de gros bourgs comme Yssingaux, Brioude...

On dispose d'un recensement effectué en 1982 permettant de dessiner une image du département ainsi que de données d'évolution entre 1975 et 1982.

Les variables relevées sont les suivantes :

variables du recensement

2 - P 019 ANS - % de la classe d'âge 0 à 19 ans
3 - P 2039 ANS - % de la classe d'âge de 20 à 39 ans
4 - P 4059 ANS - % de la classe d'âge de 40 à 59 ans
6 - SEXRATIO - (rapport nombre de femmes/nombre d'hommes) pour la classe 20-39 ans
7 - POURCETLAN - % d'étrangers par rapport à la population totale
8 - PVARTOTALE - % de variation de la population totale entre 75 et 82
9 - DENSPOP82 - Densité de population (en nombre d'habitants au km ²)
10 - VARRESPRIN - Variation du nombre de résidences principales entre 75 et 82
11 - VARPOPAC - % de la population active entre 75 et 82
12 - POURHORCOM - % de la population active travaillant hors de la commune de résidence
13 - POURACTTOT - % des actifs agricoles par rapport à la population active totale
14 - POURCHOMAC - % de chômeurs par rapport à la population active totale

Remarque : on ne dispose pas de la classe d'âge supérieure à 59 ans : c'est pourtant une donnée indispensable pour cerner le vieillissement du département. On calculera cette variable avec LISA.

1. La problématique

pour un objectif

La liste des paramètres relevés ci-dessus éclaire suffisamment sur les questions que se posent les responsables de l'enquête !

Nous pouvons aborder l'étude de plusieurs façons :

Soit variable par variable : on compare la distribution des communes et on essaie de les cartographier. Puis on effectue une analyse thématique : par exemple, quelles sont les communes dont la population a diminué, est restée stable ou a augmenté ? Où se trouvent-elles géographiquement ? Peut-on, par une carte, mettre en évidence les zones homogènes ?

L'âge de la population (classe > 59 ans) et le sexe-ratio sont essentiels pour mesurer le devenir des communes : quelles sont les zones qui paraissent condamnées ?

On pourrait de la même façon aborder l'étude de toutes les variables mesurées...

Soit à partir de l'ensemble des variables en se posant trois types de questions : comment ces variables sont-elles liées ? Indépendance, synergie ou antagonisme ?

Peut-on synthétiser ces relations pour décrire les communes et en effectuer une typologie ?

Peut-on élaborer une carte de synthèse permettant de représenter dans l'espace différents types de communes retenus ? Mettra-t-on ainsi en évidence des zones bien délimitées d'avenir plus ou moins incertain ?

2. La méthode suivie

quatre étapes

Sans vouloir être exhaustif, on abordera l'étude avec LISA en quatre étapes :

- Gestion des données : création, enregistrement, modification des données, élaboration d'un fond cartographique pour représenter les communes.

- Analyse statistique élémentaire des variables : première prise de contact avec les données

- Analyse thématique : distribution et cartographie de quelques variables «typiques»

- Analyse de synthèse à partir de l'ensemble de l'information : étude des relations entre variables par une analyse en composantes principales ; recherche d'une typologie par une méthode de classification automatique ; mise en classes et élaboration d'une carte de synthèse permettant de caractériser les communes.

Dans tout ce qui suit, on ne donnera que les résultats. Le manuel d'utilisation de LISA explique en détail la façon d'utiliser le logiciel...

II — PREMIERE ETAPE : MISE EN FORME DE L'INFORMATION

création de la structure du fichier

Les données obtenues au cours des enquêtes doivent être classées dans des fichiers selon les paramètres précis en utilisant le module «Gestion de fichier».

Il faut donc commencer par créer ici le fichier DEMO43.GST dans lequel on a rajouté deux variables : un indicateur ou numéro de commune de 1 à 266 et la variable PSUP59ANS à calculer.

enregistrement et calcul des variables manquants

Puis on enregistre l'information en la corrigeant éventuellement...

(5 communes n'ont aucune information, on remplit les enregistrements correspondants par le code «donnée manquante»)

et enfin on calcule la variable PSUP59ANS :

Ce qui donne par exemple :

```
Fichier...:
demo43.GST

Etendue.....((Enr (- Enr(, ))) : (Sélection par N° enr.)

Condition..... : (Sélection conditionnée. F1 : aide)

Champ à remplacer (F1) :
PSUP59ANS

... par : (F1 aide)
100-P019ANS-P2039ANS-P4059ANS
```

Ch.	Nom	Nom_e	Type	long	Déc.	Modalités	Limites de signification
1	COMMUNE	COMM	c	4			
2	P019ANS	P019	F	7	2	0	[: :]
3	P2039ANS	P230	F	7	2	0	[: :]
4	P4059ANS	P450	F	7	2	0	[: :]
5	PSUP59ANS	P859	F	7	2	0	[: :]
6	SEXRATIO	SEXR	F	7	2	0	[: :]
7	POURCETRAN	ETRA	F	7	2	0	[: :]
8	PVARTOTALE	VTOT	F	7	2	0	[: :]
9	DENSPOP82	DENS	F	7	2	0	[: :]
10	VARRESPRIN	VRES	F	7	2	0	[: :]
11	VARPOFACTI	VACT	F	7	2	0	[: :]
12	POURHORCOM	HORC	F	7	2	0	[: :]
13	POURACTTOT	AAAT	F	7	2	0	[: :]
14	POURCHOMAC	CHOM	F	7	2	0	[: :]

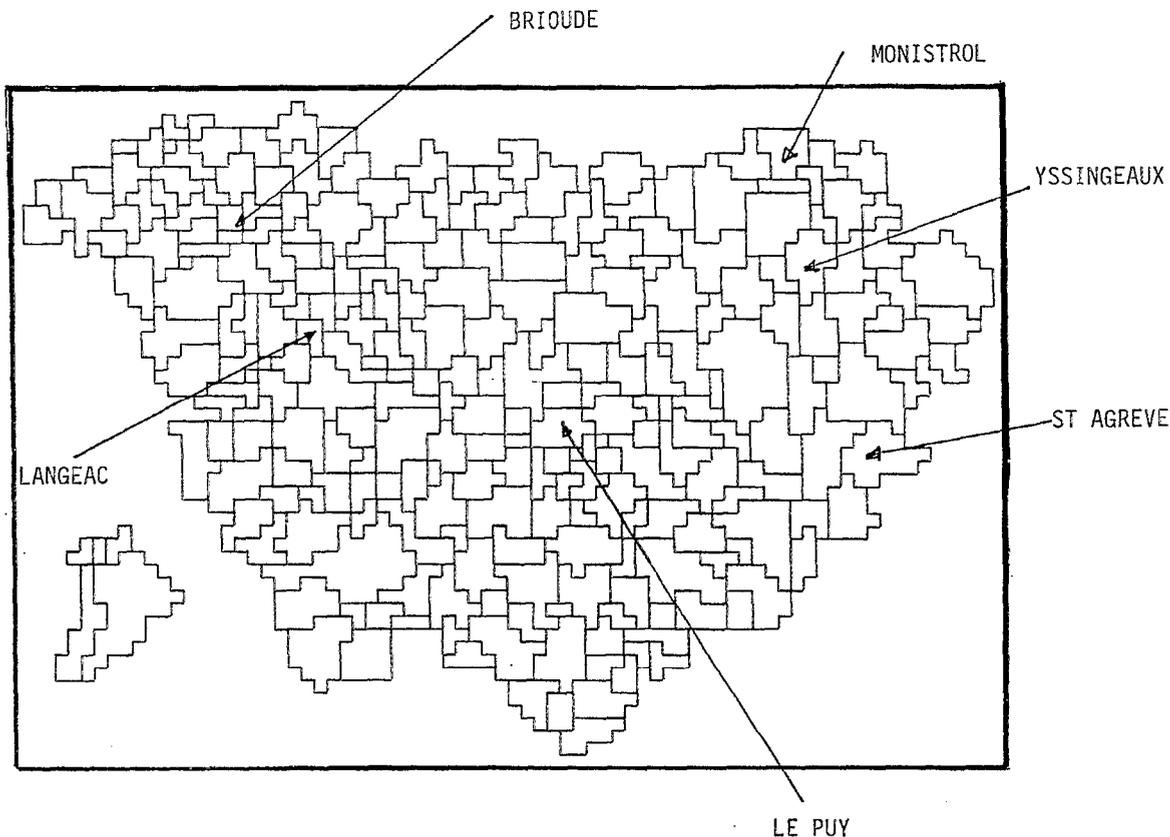
création du fichier "fond de carte"

Le fichier est alors prêt pour l'analyse...

Il faut ensuite créer le fichier «fond de carte» HL. GRL avec le module infographie de LISA : en s'aidant d'une grille dessinée par LISA, on dessine la carte des communes de la Haute-Loire, le fond de carte est ensuite enregistré dans un fichier HL. GRL, et on édite enfin sur l'imprimante (cf. fig 1).

Figure 1 - Carte des communes de Haute-Loire

* Les flèches correspondent aux agglomérations les plus importantes. Les 5 communes manquantes sont placées en dehors de la carte



III — DEUXIEME ETAPE : ANALYSE STATISTIQUE ELEMENTAIRE

Le module «statistiques élémentaires» de LISA permet de calculer les paramètres statistiques de base donnant ainsi un premier aperçu sur les données (fig 2).

Figure 2 - Paramètres statistiques de base de la démographie des communes de Haute-Loire

VARIABLE	EFFECTIF	MOYENNE	MINIMUM	MAXIMUM	ECART-TYPE	COEFF VARIA	LIMITES DE CONFIANCE 95 %	
P019	261	24.1881	3.70000	39.6000	5.57721	23.0577	23.4977	24.8786
P230	261	23.9031	12.5000	36.0000	4.33439	18.1332	23.3665	24.4397
P450	261	24.2203	14.9000	36.0000	3.55279	14.6686	23.7805	24.6601
P559	261	27.6885	7.70000	53.3000	7.31461	26.4175	26.7830	28.5940
SEXR	261	.802184	.120000	2.11000	.209817	26.1558	.776209	.828158
ETRA	261	.868582	.000000	11.4000	1.74486	200.886	.652574	1.08459
VTOT	261	-2.86107	-30.9000	69.8000	15.8314	553.337	-4.82095	-.901198
DENS	261	61.3870	4.00000	1433.00	167.918	273.540	40.5992	82.1747
VRES	261	3.04866	-25.7000	78.7000	15.4297	506.114	1.13851	4.95881
VACT	261	1.06077	.570000	2.23000	.237013	22.3435	1.03142	1.09011
HORC	261	29.2192	.000000	81.0000	16.6914	57.1248	27.1528	31.2855
AAAT	261	41.1912	.600000	86.4000	22.8103	55.3767	38.3673	44.0150

montre
les variabilités
des paramètres

La grande variabilité de certains paramètres laisse prévoir de grosses différences selon les communes... : par exemple, les proportions des différentes classes d'âge sont très variables, le sex-ratio varie dans des limites étonnantes et le pourcentage d'actifs agricoles également.

Remarquez notamment que certaines communes subissent une perte de population et de résidences principales totales alors que d'autres en gagnent beaucoup ! etc.

L'analyse des corrélations apparentes (non éditée) montrerait également des liaisons importantes entre certaines variables : elles seront étudiées en détail lors de l'analyse en composantes principales...

IV — TROISIEME ETAPE : ANALYSE THEMATIQUE

L'analyse thématique permet d'étudier les unes après les autres les variables choisies.

A titre d'exemple, nous étudierons deux variables importantes :

- Le % de population âgée de 60 ans ou plus,
- La variation de population totale de 1975 à 1982

Le module «histogrammes» de LISA permet de préciser la distribution de ces variables en grandes classes (fig 3).

distribuer
les variables
en grandes classes

Figure 3 - Editions partielles de la distribution de variables

CLA	LIMITE SUP	EFFECT	POURCENT % CUMULE	
1	15.0	9	3.45	3.45
2	20.0	29	11.11	14.56
3	25.0	63	24.14	38.70
4	30.0	62	23.75	62.45
5	35.0	56	21.46	83.91
6	40.0	33	12.64	96.55
7	53.0	9	3.45	100.00

CLA	LIMITE SUP	EFFECT	POURCENT % CUMULE	
1	-25.0	10	3.83	3.83
2	-15.0	44	16.86	20.69
3	-10.0	41	15.71	36.40
4	-5.00	40	15.33	51.72
5	.000	45	17.24	68.97
6	5.00	24	9.20	78.16
7	10.0	15	5.75	83.91
8	20.0	21	8.05	91.95
9	30.0	8	3.07	95.02
10	.40.0	4	1.53	96.55
11	69.8	9	3.45	100.00

Ces deux variables mettent en évidence les grandes différences entre les communes du département. Il existe tous les intermédiaires entre les communes «âgées» et les communes «jeunes» et celles où la population diminue et celles où la population augmente.

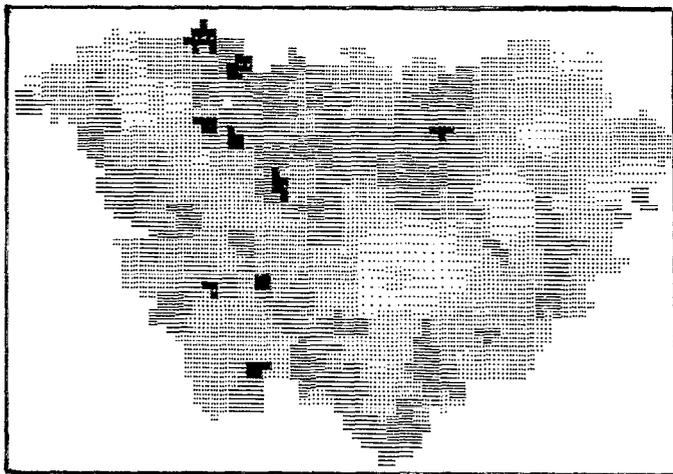
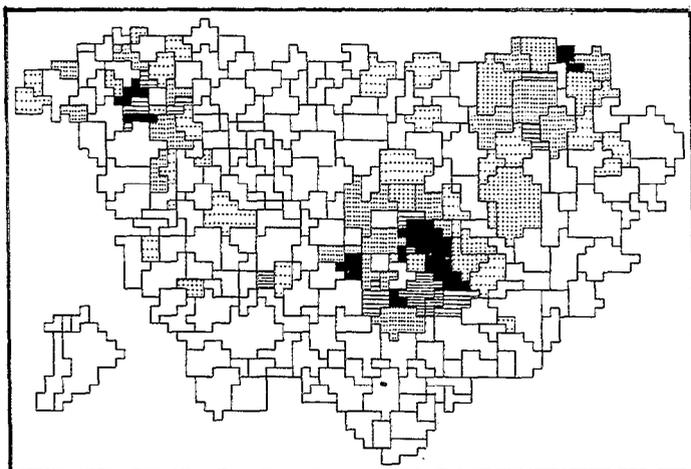


Figure 4 - Population > 60 ans



Figure 5 - Variation de la population

Augmentation — Population Totale



Diminution — Population Totale

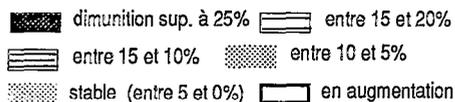
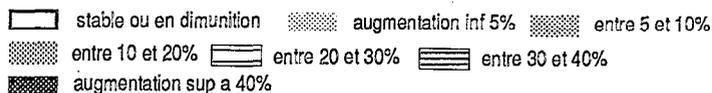
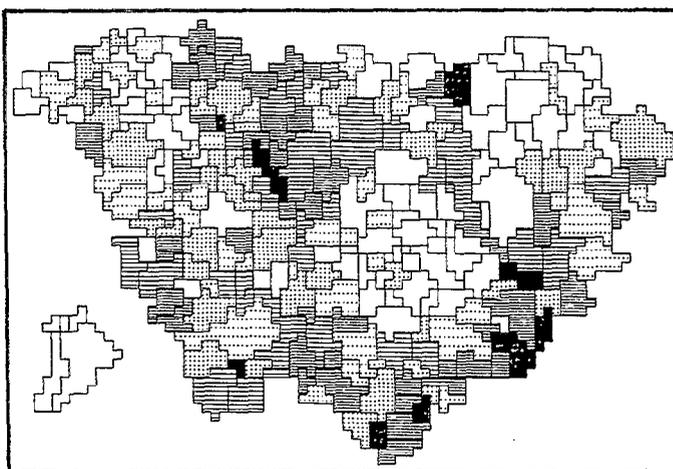


Figure 6 - Matrice des corrélations
(Tous les coefficients sont multipliés par 1000)

	P019	P230	P450	PS59	SEXR	ETRA	VTOT	DENS	VRES	VACT	HORC	AAAI	CHOM
P019	1000												
P230	526	1000											
P450	-435	-558	1000										
PS59	-863	-722	177	1000									
SEXR	333	283	-195	-327	1000								
ETRA	234	333	-182	-288	215	1000							
VTOT	580	542	-372	-583	340	161	1000						
DENS	253	344	-125	-336	256	446	220	1000					
VRES	508	525	-271	-567	288	135	905	260	1000				
VACT	507	504	-246	-565	230	144	764	188	710	1000			
HORC	427	536	-249	-522	299	67	645	322	597	155	1000		
AAAI	-459	-600	398	513	-390	-450	-574	-452	-514	-478	-575	1000	
CHOM	47	89	-107	-37	-41	41	42	82	54	110	-34	-184	1000

NUM	ITER	VAL. PROPRE	POURCENT	CUMUL	P*	HISTOGRAMME DES VALEURS PROPRES DE LA MATRICE
1	0	5.79406	44.570	44.570	!*
2	0	1.44642	11.126	55.696	!*
3	3	1.08764	8.366	64.062	!*
4	3	.95974	7.383	71.445	!*
5	2	.83661	6.435	77.881	!*
6	3	.76082	5.852	83.733	!*
7	2	.63219	4.863	88.596	!*
8	3	.47711	3.670	92.266	!*
9	2	.37059	2.851	95.117	!*
10	5	.30539	2.349	97.466	!*
11	2	.25649	1.973	99.439	!*
12	2	.07294	.561	100.000	!*
13	2	-.00001	.000	100.000	!*

Figure 7 - Les valeurs propres
val (1) = 5.79406

les cartographier

Pour éditer une carte thématique, il faut transformer les variables du fichier en classes à l'aide du **module de gestionnaire** de LISA .

Deux cartes ont donc été éditées :

- La carte de la classe d'âge de 60 ans et plus (fig 4)
- Les cartes des variations de population totale (fig 5) et zones où la population diminue et zones où la population augmente.

et les analyser

On remarquera que les zones où la population est la plus âgée correspondent aux communes rurales, plus particulièrement de moyenne altitude.

La carte de la classe > 59 ans et celle des diminutions de population sont en concordance totale : les zones «vieillissantes» ont subi l'exode de plein fouet au profit des zones urbaines, péri-urbaines ou situées à proximité d'un bourg important (cf. carte des augmentations).

On constate aussi que la population totale a augmenté essentiellement dans 3 zones : la limagne de Brioude, la zone du Puy, l'axe Monistrol-Yssingeaux-le Puy.

V — QUATRIEME ETAPE : ANALYSE SYNTHETIQUE

Permet, à partir de l'ensemble des variables d'avoir une image «visuelle» des ressemblances ou des différences.

Cette analyse classique enchaîne : analyse factorielle, classification automatique, étude des partitions obtenues, mise en classe dans un fichier et éventuellement cartographie des résultats...

1. Relations entre les variables. Analyse des composantes principales (ACP)

Partant d'un espace de représentation à 13 variables, donc impossible à visualiser, on ne peut pas juger les ressemblances entre les communes, analyser les inter-relations entre les variables et expliquer celles qui discriminent les communes.

L'ACP s'efforce de répondre aux questions suivantes :

- Peut-on réduire l'espace de représentation ? Autrement dit, peut-on passer avec le moins de perte possible d'informations d'une représentation à 13 dimensions à une ou deux représentations planes beaucoup plus simples à analyser ?

- Quelles sont les variables les plus importantes et les inter-relations entr'elles ?

Y-a-t'il indépendance, antagonisme ou synergie ?

- Comment se situent les individus «communes» les uns par rapport aux autres ?

Peut-on mettre en évidence des groupes, des gradients, des communes qui se ressemblent ou qui sont différentes ?

- Quelles sont les variables qui discriminent le plus les communes ?

On effectue alors une ACP avec LISA dont voici quelques extraits.

- L'examen de la matrice des corrélations (fig. 6) apporte déjà un certain nombre d'indications :

étude de la matrice de corrélation

De façon plus ou moins précise, on note une opposition entre les classes d'âge jeunes et âgées. La variation de la population totale, des résidences principales, de la population active, la densité de population, le % de gens travaillant hors commune augmentent avec les classes d'âge jeunes. Il en est de même, mais beaucoup moins nettement, pour le sex-ratio (sans doute en raison de quelques institutions religieuses) et du % d'étrangers.

A l'opposé, le % d'actifs agricoles augmente avec les classes d'âge âgées.

Notez que le % de chômeurs est à peu près indépendant des autres variables...

- On va préciser les inter-relations entre ces variables par l'ACP.

Par un changement de variable, on essaie d'obtenir des «facteurs» expliquant au mieux le phénomène. Les variances (valeurs propres) de ces facteurs sont dans l'ordre (fig 7).

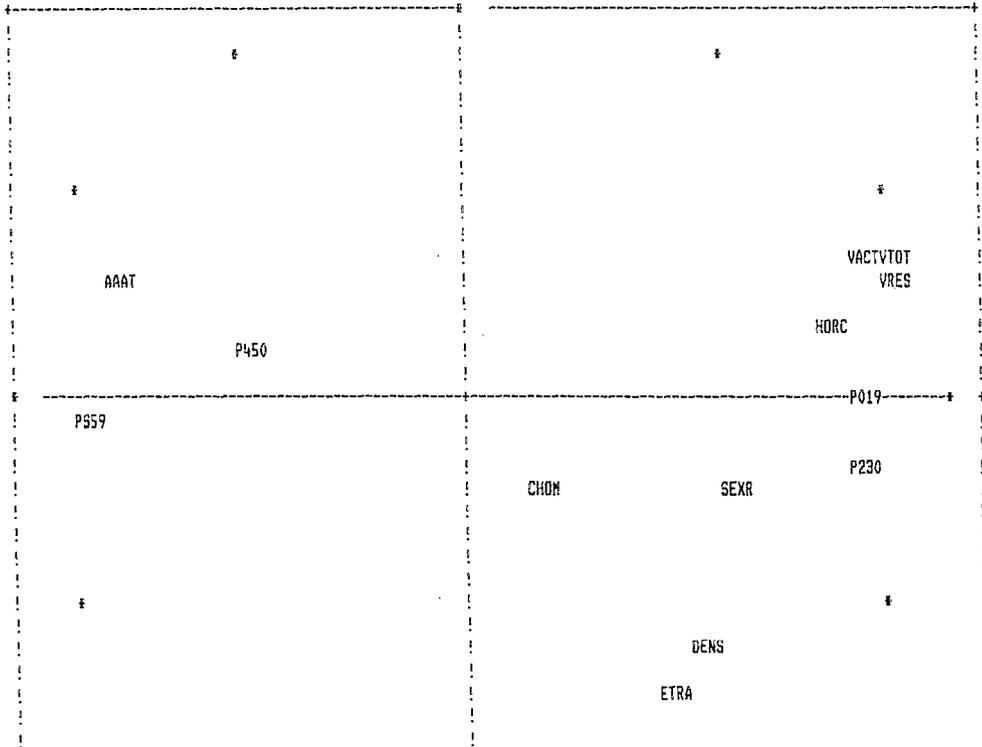
**des inter-relations
entre les variables**

On constate que le premier facteur explique près de 50 % (44,57 %) du phénomène. La variabilité des troisième et quatrième facteurs ne dépasse pas celle des variables de départ. On verra donc l'essentiel sur le plan des 2 premiers facteurs. Au-delà, on aura à faire à des influences secondaires. Le programme calcule les coordonnées des communes dans l'espace des facteurs ainsi que les corrélations (entr'autres) entre les facteurs et les variables de départ. On édite seulement ce dernier résultat (fig 8). Il représente les variables et les #F les corrélations (*1000) entre les variables et les facteurs 1,2... on édite aussi le graphe des variables et leur cercle de corrélation sur le plan des facteurs 1 et 2.

Figure 8 - Corrélation entre les facteurs et les variables de départ - tableau et graphique

		axe 1	axe 2	axe 3	axe 4	axe 5	axe 6
1	P019	763	40	-10	301	429	198
2	P230	796	-131	134	251	57	-164
3	P450	-493	139	-420	-621	330	136
4	P559	-814	-20	133	-76	-521	-120
5	SEXR	470	-195	-342	98	-357	683
6	ETRA	382	-718	-111	-66	99	-192
7	VTOT	852	347	9	-110	-142	-44
8	DENS	464	-575	-192	-341	-22	-145
9	VRES	804	350	-11	-210	-100	-75
10	VACT	757	345	72	-207	29	-60
11	HORC	732	219	-144	-151	-217	-127
12	AAAT	-774	286	-78	119	213	39
13	CHOM	105	-180	823	-387	88	329

NOMBRE DE POINTS : 25



L'analyse du tableau et du graphe précédents conduit aux résultats suivants :
 Le premier facteur (axe 1) met très nettement en évidence l'opposition entre les classes jeunes (partie positive de l'axe) et les classes âgées (surtout > 60 ans, partie négative de l'axe). Les classes âgées sont caractérisées par les actifs agricoles. La variation de population totale, active, des résidences secondaires et le % de gens travaillant hors commune varient dans le même sens que les classes d'âges jeunes.
 Le second facteur (axe 2) caractérise les étrangers et à un degré moindre la densité de population. La présence d'étrangers est liée aux communes de forte densité (donc urbaines), densité bien expliquée par les deux premiers facteurs.
 L'axe 3 caractérise le chômage, pratiquement indépendant des autres variables. Sex-ratio et classe d'âge 40-59 ans sont moins bien expliqués. Il faut aller à l'axe 4 pour cette dernière et à l'axe 6 pour le sex-ratio.

Cela signifie que la classe d'âge 40-59 ans est beaucoup moins discriminante pour typer les communes et que le sex-ratio doit être interprété avec nuance compte tenu des nombreuses congrégations religieuses du département...

leur représentation graphique permet ici de classer les communes

et d'étendre les facteurs qui les discriminent

afin de réaliser une typologie

Si l'on représente les communes dans le plan des deux premiers facteurs, on entrevoit (cf. Fig 9) comment vont se classer les communes et ce qui les discrimine.

Horizontalement, de la gauche vers la droite, on passe progressivement des communes les plus rurales, à faible densité, à population active fortement agricole, dominées par les personnes âgées, où l'exode est fortement marqué, aux communes où les classes jeunes sont dominantes, la population plus dense, l'augmentation de population nette (au détriment des précédentes), où la main-d'œuvre va souvent travailler hors-commune mais réside sur place...

Dans le cadran IV, on trouvera les communes les plus urbaines, avec une main-d'œuvre étrangère plus importante qu'ailleurs...

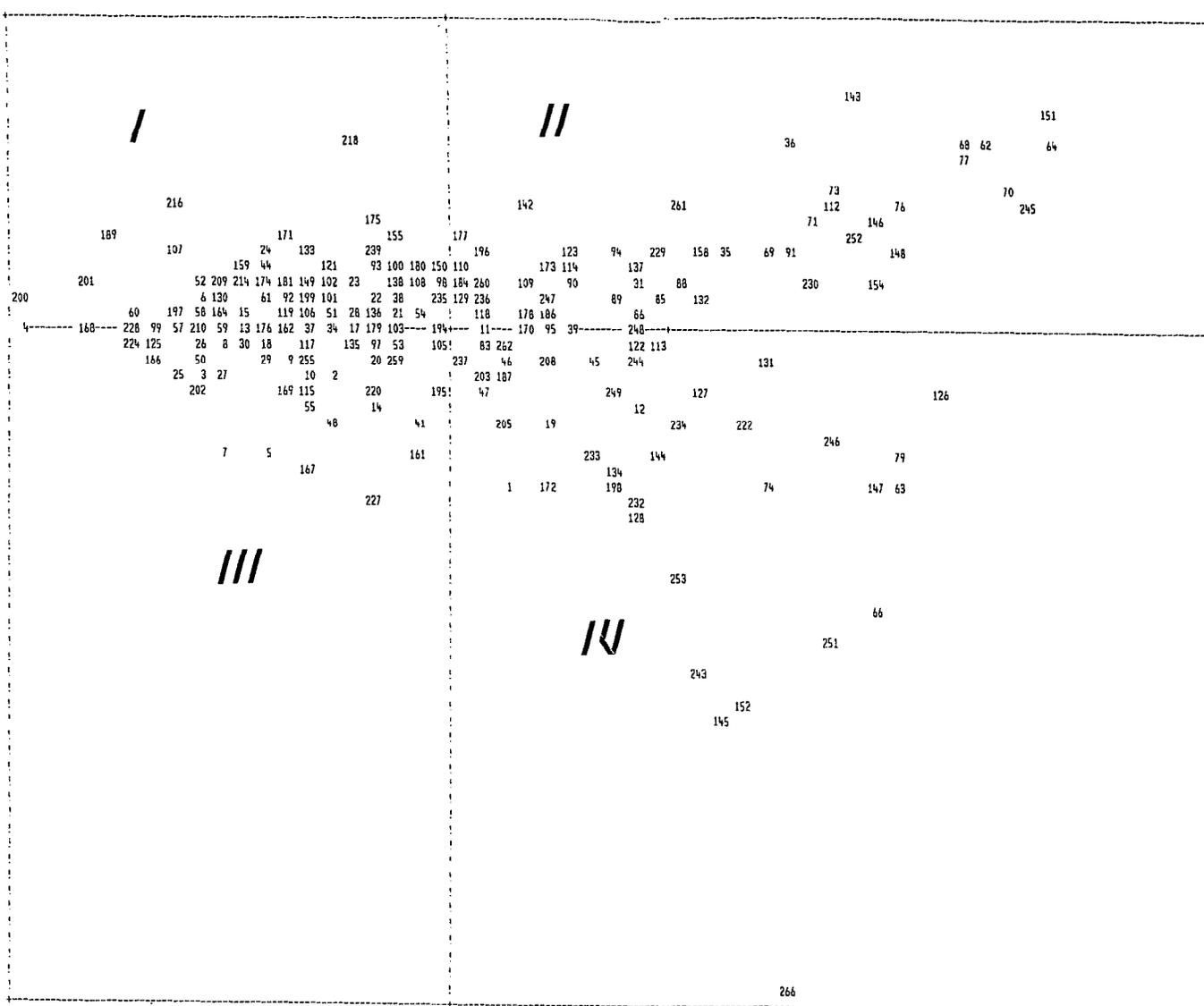
Dans le cadran II, on trouvera les communes où les variations de populations et de résidences ont été les plus importantes et où on travaille le plus hors commune.

Les synergies et antagonismes entre ces diverses variables sont suffisamment nettes pour que l'on puisse tenter par une classification automatique d'établir une typologie synthétique des communes.

Il y a deux façons de faire :

- Partir des variables de départ
 - Partir des facteurs mis en évidence par l'ACP. Cette dernière technique est préférable car elle hiérarchise l'importance des variables, élimine le bruit de fond lié à des variations secondaires...
- Nous retiendrons les 6 premiers facteurs, le sixième prenant encore en compte le sex-ratio...

Figure 9 - Représentation des communes selon les axes 1 et 2

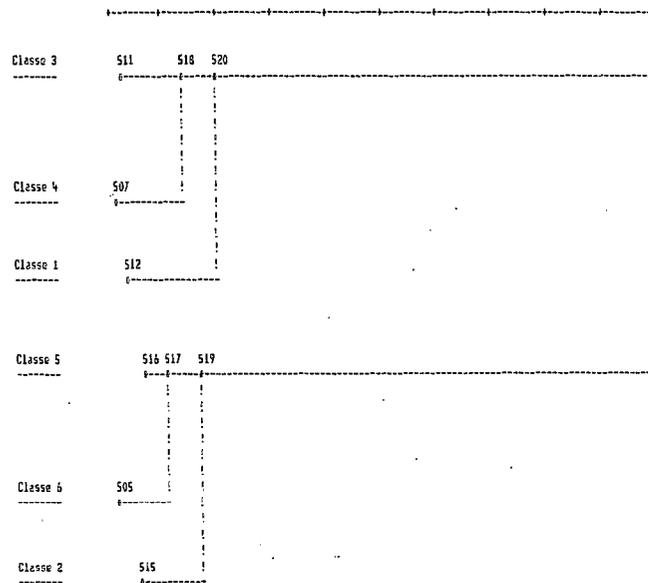


2. Classification des communes

automatique selon 4 possibilités

LISA propose 4 types de classifications automatiques : deux hiérarchiques et deux directes. Voici à titre d'exemple une classification **ascendante hiérarchique** : cette méthode fournit une hiérarchie de classe emboîtées de plus en plus fines, allant de la classe réunissant l'ensemble des 261 communes aux classes terminales réduites à une seule commune. Il est par conséquent possible d'étudier les différents niveaux de la hiérarchie et de proposer une classification de plus en plus précise. Dans le cas présent, on s'est limité à 6 classes de communes, les classes plus fines correspondant à des différences minimales (fig 10)... On ne représente que l'arbre hiérarchique et les numéros des classes retenues... La règle horizontale, graduée en dixièmes représente «la distance» entre la branche de droite et gauche d'une classe. Cela permet de juger de la proximité et de l'homogénéité des classes.

Figure 10 - Représentation de la classification hiérarchique : extraits



La population des communes se sépare en deux classes relativement homogènes 520 et 519 très différentes l'une de l'autre.

La classe 520 se sépare elle-même en 2 classes 518 et 512, la classe 512 étant très homogène.

La classe 518 éclate en 2 classes 511 et 507 très homogènes.

Les classes 511, 507 et 512 sont proches les unes des autres.

A l'opposé, la classe 519 donne naissance aux classes 517 et 515, cette dernière étant très homogène.

La classe 517 se décompose en 2 classes homogènes 516 et 505.

Les classes 516, 505 et 515 sont proches les unes des autres.

caractérise les classes

Comment ces classes se caractérisent-elles ?

L'ACP nous a fourni des indications sur les variables qui discriminent ces classes.

LISA permet d'analyser, niveau par niveau, les classes obtenues en calculant pour chacune les statistiques élémentaires.

Ainsi, au niveau le plus élevé, la séparation se fait sur une opposition entre les communes à caractère rural et les communes à caractère urbain :

- Les communes de la classe 520 sont de caractère urbain, les communes de la classe 519 sont de caractère rural,

- La classe 520, par rapport à la classe 519, est caractérisé par : une population plus jeune, un sex-ratio plus élevé donc plus «de filles à marier», un % d'étrangers supérieur, une population totale et

un nombre de résidences principales qui augmentent tandis qu'elles diminuent en 519, une densité de population très nettement supérieure, un % d'actifs agricoles bien plus faible, un % d'actifs travaillant hors commune nettement supérieur.

En continuant l'analyse, on montrerait que :

- La classe 520 donne naissance à deux classes 518 et 512, la dernière étant de caractère urbain plus fortement accentué
- La classe 519 donne naissance aux classes 517 et 515, la classe 517 étant de caractère rural plus marqué.

En définitive, on arrête l'analyse à 6 classes (classes 512, 515, 511, 507, 516 et 505) : ces classes sont renumérotées dans l'ordre par le programme de 1 à 6. (cf. arbre).

3. Analyse des 6 classes obtenues

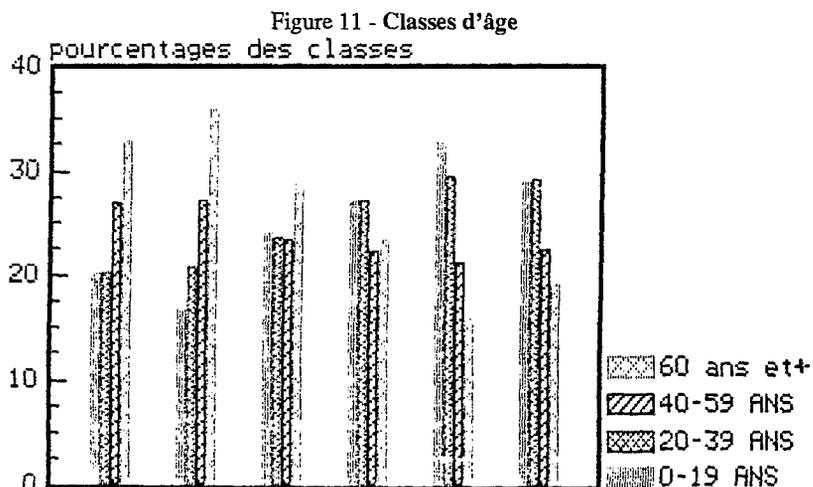
les analyses

Voici le tableau des moyennes des classes ordonnées de haut en bas dans le sens rural vers ville, c'est à dire dans l'ordre des numéros de classe 5, 6, 2, 4, 3 et 1.

CLASSE	P019ANS	P2039ANS	P4059ANS	PSUP59ANS	SEXRATIO	%ETRAN	%VARTOT	DENS POP 82	VARRES PRIN	VARPOP ACTI	%HORCOM	%AATOT	%CHOMEURS
5	20.25	20.23	27.00	32.72	0.71	0.32	-14.13	11.90	-7.22	0.89	20.03	61.00	6.03
6	16.61	20.59	27.08	35.73	0.55	0.44	-10.89	13.00	-0.81	1.04	20.05	45.79	15.26
2	24.13	23.70	23.43	28.73	0.81	0.49	-6.88	23.51	-1.27	1.03	23.70	45.65	7.38
4	27.22	27.10	22.39	23.29	0.90	0.93	7.24	59.17	11.58	1.18	41.13	21.83	7.33
3	33.24	29.72	21.26	15.78	0.92	0.63	34.06	66.90	37.48	1.51	58.14	16.32	7.32

à l'aide de graphiques

En utilisant le logiciel résidant en mémoire «graph in the box», on peut illustrer ce tableau par trois graphiques :



celui des classes d'âges (fig 11)

celui des variables sex-ratio à l'agriculteurs/actifs rapportées à l'indice 100 (fig 12)

celui des variables étrangers à chômeurs rapportées à l'indice 100 (fig 13)

Interprétation

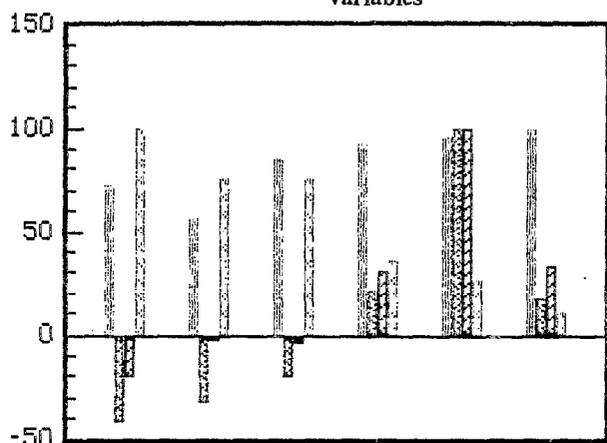
La classification confirme les conclusions de l'ACP.

La plupart des variables s'ordonnent dans le sens rural → ville à quelques nuances près.

Ainsi :

- l'importance des classes jeune augmente
- les classes âgées diminuent ainsi que le % d'actifs agricoles
- le sex-ratio et le % d'étrangers augmentent
- la densité de population augmente
- la population totale, le nombre de résidences principales diminuent dans les zones rurales et augmente dans les zones urbaines.

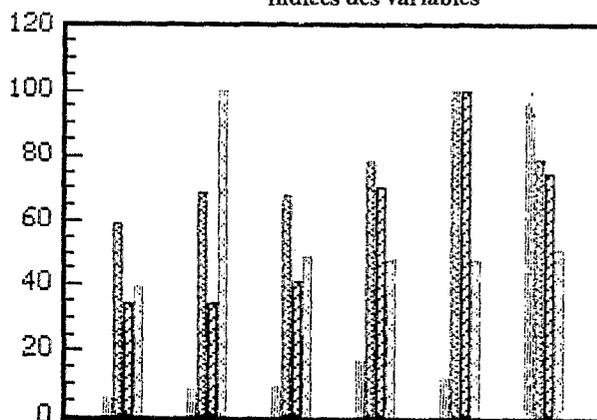
Figure 12 - Classes de communes-indices des variables



Classes dans l'ordre 5,6,2,4,3 et 1

	AGRI/ACTIF
	VARESPRIN
	%VARPOPTOT
	SEXRATIO

Figure 13 - Classes de communes - indices des variables



Classes dans l'ordre 5,6,2,4,3 et 1

	%CHOMEURS		VARPOPACTI
%HORS COMM pattern"/>	%HORS COMM		%ETRANGERS

Le % des personnes travaillant hors commune augmente dans les zones à caractère rural alors qu'il est très faible dans les zones rurales.

Il faut mettre à part la population active qui reste stable et le % de chômeurs qui très curieusement n'est élevé que dans la classe rural 6.

Les classes 5, 6 et 2 à caractère rural sont donc très différentes des classes 3, 4 et 1 à caractère urbain.

Si l'on compare les classes rurales 5, 6 et 2 on constate que :

- Les classes 5 et 6 ont un caractère rural très marqué et sont les plus menacées : population âgée et sex-ratio très faible donc pas d'épouses pour les jeunes..., fort exode et % d'actifs agricole élevé, chômage important pour la classe 6

- La classe 2 est intermédiaire entre les classes précédentes et les classes à caractère urbain. Il s'agit des communes groupées autour de villages d'une certaine importance.

Concernant les classes à caractère urbain, on passe de la classe 4 constituée de communes regroupées autour de bourgs importants ou à proximité de villes aux communes péri-urbaines de la classe 3 et enfin aux communes urbaines (Le Puy...) de la classe 1.

Les communes péri-urbaines de la classe 3 sont caractérisées par la plus forte augmentation de population totale, des résidences principales, et que cette classe a le plus fort % d'actifs travaillant hors commune.

4. Cartographie des classes des communes

LISA permet de récupérer dans un fichier les classifications obtenues aux différents niveaux de la hiérarchie.

On se limitera à la représentation cartographiques des 6 classes précédentes (fig 14) qui permet de noter les transitions entre les zones des classes 5 et 6 (essentiellement de montagne), la classe 2 dans les plateaux et les vallées et les auréoles péri-urbaines à urbaines des classes 4, 3 et 1 et de remarquer les principaux centres d'attraction urbains que sont : au nord, la limagne de Brioude et Langeac plus au sud-est, la ville du Puy au centre, l'axe Monistrol-Yssingeaux au nord-est, la ville de saint Agrève à l'est.

permettant de visualiser une typologie

5. Agrégation des communes des 6 classes

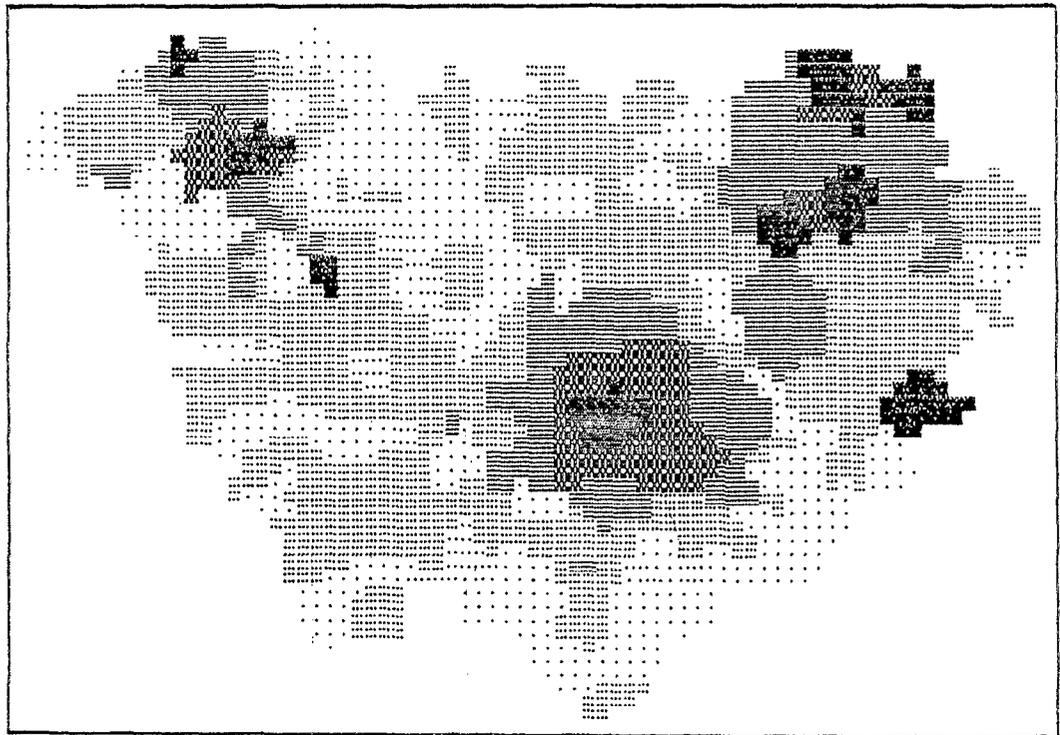
Pour terminer, nous voudrions montrer l'une des fonctionnalités les plus originales de LISA.

L'agrégation permet de passer d'un niveau à un autre niveau d'étude plus global, par exemple des communes aux cantons, des cantons aux départements...de la parcelle à l'exploitation (temps de travaux...), de l'exploitation au village etc...

permet de changer de niveau d'étude

LISA permet aussi d'agréger les données sur un nombre quelconque de critères.

Fig 14 — Typologie des communes



LEGENDE

-
-

a partir du nombre de critères

L'agrégation est réalisée à deux niveaux :

- Au niveau fichier de données : on crée un nouveau fichier contenant autant d'articles que d'unités d'agrégation. Les variables reçoivent les sommes, valeurs moyennes ou %.
- Au niveau cartographique : on crée un nouveau fond cartographique où les zones d'agrégation regroupent les zones élémentaires.

Par exemple : supposons que l'on veuille redécouper le département en zones homogènes correspondant aux 6 classes de communes. On crée par agrégation un fichier de 6 classes de zones contiguës ou non.

Le fichier contient l'information agrégée suivante (on a choisi les valeurs moyennes) :

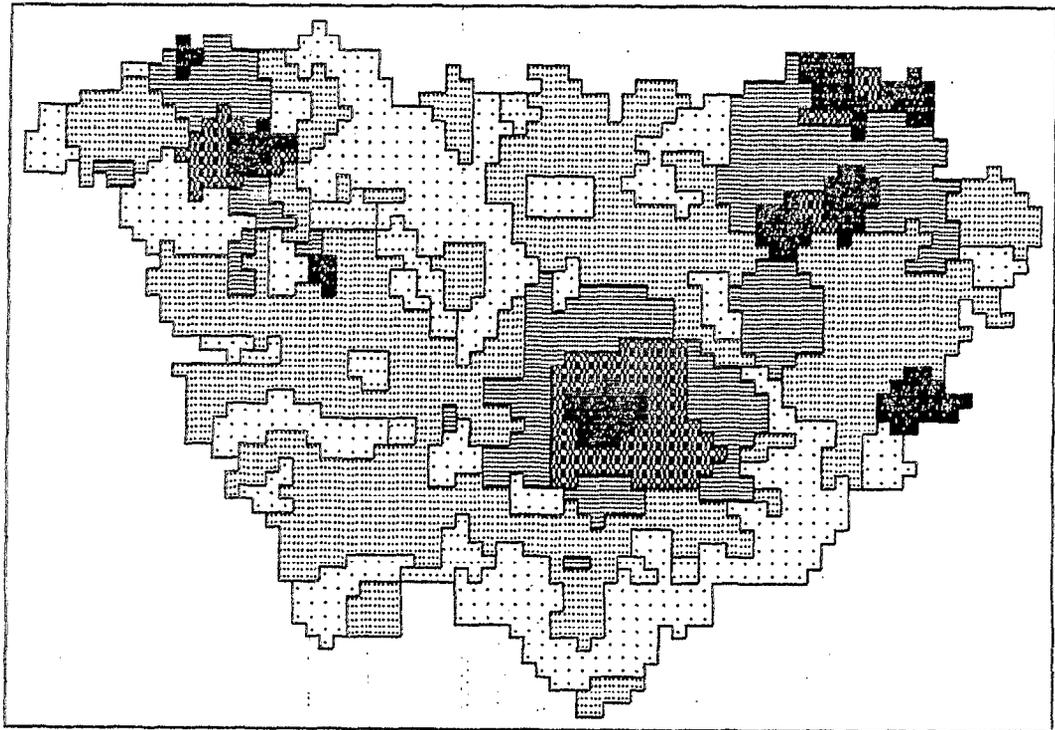
On crée également un fond de carte regroupant les communes en 6 classes.

CLASSE P019ANS P2039ANS P4059ANS PSUP59ANS SEXRATIO %ETRAN %VARTOTALE DENSPOP8 VARRESPRIN VARPOFACTI %HORCOM %AATOT %CHOMEURS

CLASSE	P019ANS	P2039ANS	P4059ANS	PSUP59ANS	SEXRATIO	%ETRAN	%VARTOTALE	DENSPOP8	VARRESPRIN	VARPOFACTI	%HORCOM	%AATOT	%CHOMEURS
1	29.18	29.29	22.52	19.01	0.97	5.52	5.85	497.89	12.46	1.19	43.28	6.33	7.84
2	24.13	23.70	23.43	28.73	0.81	0.49	-6.88	23.51	-1.27	1.03	23.70	45.65	7.38
3	3.24	29.72	21.26	15.78	0.92	0.63	34.06	66.90	37.48	1.51	58.14	16.32	7.32
4	27.22	27.10	22.39	23.29	0.90	0.93	7.24	59.17	11.58	1.18	41.13	21.83	7.33
5	20.25	20.03	27.00	32.72	0.71	0.32	-14.13	11.90	-7.22	0.89	20.03	61.00	6.03
6	16.61	20.59	27.08	35.73	0.55	0.44	-10.89	13.00	-0.81	1.04	20.05	45.79	15.26

En représentant les 6 zones homogènes avec les mêmes symboles que pour la typologie, on trace la carte des 6 zones (fig 15). Comparez avec la typologie : les zones agrègent bien les communes semblables.

Figure 15 - Les 6 zones homogènes



LEGENDE

 CLASSE1	 CLASSE4
 CLASSE2	 CLASSE5
 CLASSE3	 CLASSE6

CONCLUSION

A travers l'exemple traité, on a voulu montrer quelques fonctionnalités du logiciel LISA. La méthode présentée ne prétend pas être universelle. Bien des typologies peuvent être établies sans utiliser la batterie lourde des analyses multidimensionnelles ! LISA permet, grâce à ses procédures de tri et extraction conditionnelles notamment, d'effectuer des typologies pas à pas, raisonnées, sans faire appel aux techniques d'analyse multivariable... Le lecteur fera facilement le lien avec ses propres préoccupations. Au cours de l'année 88-89, le logiciel LISA évoluera en intégrant

- de nouveaux modules statistiques
- de nouveaux modules cartographiques
- un module de traitement graphique de l'information
- un module de traitement des données de télédétection.

BIBLIOGRAPHIE

- DE LAGARDE, 1983. — Initiation à l'analyse de données. — Paris : Dunod.
 FENELON LEFONEN J.P., 1981. — Qu'est ce que l'analyse des données.
 FOUART, 1982. — Analyse factorielle . — Paris : Masson.
 FRANCILLON G., 1988. — Présentation de LISA. — in : Gazette des Systèmes DSA.
 FRANCILLON G., 1988. — Manuel d'utilisation de LISA . — Montpellier : DSA service informatique.
 ROUX M., 1985 . — Algorithmes de classification. — Paris : Masson.

Application of the program LISA to an example of survey processing - G. FRANCILLON

The article aims at demonstrating a number of aspects of the computer program LISA (Logiciel Intégré pour les Systèmes Agraires - Integrated program for farming systems) developed and distributed by D.S.A. since 1st August 1988 by analysing the demography of the communes in a french department (Haute Loire). A short presentation of the program is followed by a study of the department at several levels :

- data management,
- basic statistical analysis, plotting of theme maps,
- typology : factorial analysis, automatic classification and statistical study of the classes obtained were carried out one after the other,
- finally, a synthetic map of the classes of communes and a map showing homogeneous zones were produced. LISA is subsequently to be applied to farming systems.

KEY WORDS Method - Data analysis - Software - Farming system - Demography - Data processing - Commune - France

Aplicación del programa LISA para un ejemplo de tratamiento de encuesta : el análisis demográfico de las comunas de Haute-Loire - G. FRANCILLON

Analizando la demografía de las comunas de una provincia francesa (Haute-Loire) el presente artículo se propone revelar ciertas particularidades del programa LISA (Logiciel Integrado para los Sistemas Agrarios), programa desarrollo y divulgado por el DSA desde el primero de agosto de 1988. Tras un presentación esquemática del programa, se inicia el estudio de la provincia a diferentes niveles :

- gestión de datos,
 - análisis estadístico elemental, realización de mapas temáticos
 - búsqueda de una tipología : se realizan uno tras otro el análisis factorial, la clasificación automática y el estudio estadístico de las clases obtenidas...
 - para terminar, se realiza un mapa sintético de las clases de comunas así como un mapa en el que se vean asociadas las zonas homogéneas.
- Ulteriormente, LISA será aplicado a los sistemas agrarios.

Palabras claves : Método - Análisis de datos - Informática - Sistemas Agrarios - Programa de informática - Demografía - Comuna - Francia