

2 AVRIL 1985.

- 46 O.R.S.T.O.M. Fonds Documentaire
N° : 17.240 ex 1
Cote : B

UNE UTILISATION DE LA MÉTHODE DES NUÉES
DYNAMIQUES LORS D'UNE ENQUÊTE SOCIO-DÉMOGRAPHIQUE

PAR PHILIPPE ANTOINE
ET AGNÈS GUILLAUME

La tentation est toujours grande, lors d'une enquête statistique, de multiplier les questions afin de saisir le mieux possible la réalité sociale. Or il est souhaitable qu'une enquête portant sur un échantillon conséquent d'individus ne comporte qu'un nombre restreint de questions (contraintes de temps, de coût, de fiabilité des réponses, ...). Afin de mieux comprendre et expliquer les données recueillies, il semble préférable de se livrer à des investigations plus approfondies auprès de certains enquêtés. Divers critères peuvent être envisagés : tirer au hasard un individu sur 100 ou 1 000, retenir les personnes les plus communicatives, procéder à un choix raisonné, etc. Ces différents modes de tirage d'un sous-échantillon ne sont guère satisfaisants car aucun moyen ne permet de connaître le degré de représentativité des individus sélectionnés.

Certaines méthodes d'analyse de données peuvent aider dans ce choix et la méthode des nuées dynamiques utilisée dans cet exemple présente de nombreux avantages qui justifieraient son utilisation plus fréquente lors de l'exploitation et de l'analyse des enquêtes démographiques et statistiques portant sur des effectifs importants.

1 - OBJECTIF INITIAL : PASSER DU QUANTITATIF AU QUALITATIF

A partir des données recueillies à Abidjan lors de l'enquête ivoirienne à passages répétés (EPR) (1) nous comptons prolonger l'analyse des relations

(1) Cette enquête a été réalisée et financée par la Direction de la Statistique de Côte d'Ivoire (01 BP V 55 - ABIDJAN 01). Les premiers résultats sont disponibles auprès de ce service. Pour une présentation de l'enquête d'Abidjan, voir aussi ANTOINE Ph. et HERRY Cl. (1983).

Ph. ANTOINE et A. GUILLAUME sont des démographes de l'ORSTOM.

- 47 -

familiales en essayant de mieux cerner l'espace de vie des chefs de ménage tant dans les liens entretenus avec le milieu d'origine que dans le champ des relations familiales en ville. Une démarche plus anthropologique s'imposait, il restait donc à choisir les individus susceptibles d'être enquêtés. L'hypothèse initiale émise est la suivante : il existe des comportements différenciés selon les types d'habitat et il convient donc de distinguer les chefs de ménage selon cette catégorie et également selon le type de famille auquel ils appartiennent. L'habitat nous semble être une variable assez synthétique traduisant bien la diversité du fait abidjanais. Quatre types peuvent être distingués :

- l'habitat de haut et moyen standing (ou résidentiel) qui regroupe les catégories sociales les plus élevées ;
- l'habitat économique moderne, où résident surtout des salariés qualifiés du secteur moderne ;
- l'habitat évolutif, ou cour, constitue l'habitat le plus répandu et le plus ancien de l'agglomération ; chaque cour comprend plusieurs logements ;
- l'habitat sommaire comprend des constructions en bois de palme et surtout des baraques. Il est malgré sa précarité relativement bien structuré et doit être considéré comme un habitat de substitution à l'habitat évolutif.

Les individus retenus doivent donc être représentatifs des différentes modalités d'une typologie combinant catégories d'habitat, formes de ménage (familiale nucléaire, famille élargie, groupement de célibataires et ménage d'une personne) et incluant d'autres variables pouvant influencer la typologie (CSP, âges, ...). Notre but est de sélectionner des individus, a priori représentatifs des différentes classes, que nous devons ensuite retrouver sur le terrain et interviewer pour mieux connaître à la fois l'évolution de la composition de leur famille, leur réseau familial en ville et les modalités de leur installation à Abidjan. Par ailleurs, cette nouvelle investigation amène à appréhender le devenir de la population cinq ans plus tard (1978 - 1983).

2 - LE CHOIX DE LA MÉTHODE

Les méthodes de classification automatique effectuent une partition d'un ensemble de données mais peu d'entre elles permettent de dégager les individus

in: Statéco, Bull. de liaison, n° 38, juin 1984

17.240 ex 1

représentatifs de cette typologie. Différentes approches sont possibles comme rechercher une partition qui minimise un certain critère, selon un certain seuil ; ou bien retenir des techniques basées sur l'agrégation de groupes par itérations successives.

Nous avons retenu les méthodes de partitionnement par agrégation à partir de noyaux (1). En effet, elles permettent de traiter de grands ensembles de données (plusieurs milliers d'individus).

Deux méthodes de "partitionnement" ont été prises en considération :

- la méthode des boules optimisées ;
- la méthode des nuées dynamiques.

Ces deux méthodes sont des algorithmes d'agrégation autour de noyaux variables. La différence essentielle entre elles étant le mode de choix des noyaux à partir duquel on agrège.

Dans le cas des boules optimisées, l'agrégation se fait en fonction d'une distance inférieure à un rayon fixé a priori (mais dont le choix peut être guidé). Si la méthode dégage facilement des typologies, elle ne permet pas d'extraire automatiquement les individus les plus représentatifs. Le nombre de classes de la typologie est très sensible à une variation même minime du rayon.

La méthode nuées dynamiques mise au point par E. DIDAY (1971) est la mieux adaptée à nos objectifs. Elle permet de "générer une partition à partir d'un corps de données sur lequel on ne demande pas de faire d'hypothèses a priori. Cette partition doit réaliser les deux propriétés suivantes :

(1) Les méthodes de classification hiérarchique (méthodes ascendantes) procèdent par agrégations successives (selon un critère de distance) d'éléments isolés au départ, à la différence d'autres méthodes de classification qui agrègent selon un noyau central fixé a priori. Les méthodes hiérarchiques n'ont pu être utilisées pour un fichier aussi important que le nôtre (7 800 individus x 60 modalités).

- les individus de chaque partie se ressemblent le plus possible ;
- les individus de deux parties différentes se ressemblent le moins possible" (DIDAY, p. 20).

Le processus est itératif et à chaque étape la qualité de la partition s'améliore. Le nombre de classes souhaité est déterminé a priori ainsi que le nombre d'éléments centraux désiré, c'est-à-dire le nombre d'éléments au centre du noyau qui seront énumérés. Au départ, un ensemble de points ou noyaux d'une classe peut être tiré au hasard. Autour de ces points se regroupent les éléments les plus proches pour former une partition. La distance calculée par rapport au centre de classe est la distance du Chi-Deux. A partir de cette partition créée, une autre famille de noyaux est définie, elle regroupe les points les plus proches formant une nouvelle classe et ainsi de suite jusqu'à obtention d'un nombre fini de classes. Si, après un certain nombre d'itérations, les classes formées sont stables, les données sont dites "classifiables" et constituent des "formes fortes". Les individus qui changent de classes selon les tirages sont les "individus charnières".

3 - L'APPLICATION

L'application présentée ici a été réalisée à l'aide des logiciels mis à la disposition des utilisateurs par l'ADDAD (1).

Un fichier "chef de ménage" a été établi à partir des données de l'EPR (strate Abidjan). Il comprend environ 7 800 chefs de ménage africains résidents au 1er passage de l'enquête et seulement une dizaine de variables démographiques et socio-économiques parmi les plus discriminantes (âge, situation matrimoniale, instruction, profession, habitat, type de famille, ...) sont retenues pour l'analyse typologique. Les modalités de ces variables sont tout d'abord regroupées et

(1) ADDAD : Association pour le Développement et la Diffusion de l'Analyse des Données.

Laboratoire du Professeur BENZECRI.
Siège social : Laboratoire de Statistique
4 rue André Chenier
92130 ISSY-les-MOULINEAUX.

transformées en variables disjonctives (code 0 ou 1), c'est-à-dire qu'une variable ayant n modalités a donné naissance à n variables disjonctives. Cette transformation s'opère très facilement avec le programme RECODE du logiciel utilisé. A l'issue de cette opération le fichier comprenait 60 variables disjonctives.

Lors de l'utilisation du programme intitulé NUEDYN, l'utilisateur doit seulement opérer le choix de quelques paramètres comme le nombre de classes désiré, le mode de tirage des individus étalons (tirage au hasard ou numéros introduits par l'utilisateur).

Nous avons, dans notre cas, procédé en deux temps. Dans une première phase, très empirique, les éléments étalons (au nombre de 10 pour chacune des classes) sont tirés au hasard ; différentes partitions sont ainsi obtenues permettant de dégager diverses catégories de chefs de ménage. Les individus sélectionnés au cours de ces différents essais servent de base au choix des éléments étalons utilisés lors de la phase ultérieure. Nous avons, dans un second temps, constitué 9 classes, de chacune 10 éléments étalons. A priori les définitions de chacune de ces phases sont les suivantes :

- Type 1 : Chef de ménage famille élargie de l'habitat résidentiel
- Type 2 : Chef de ménage masculin de plus de 65 ans résidant dans les cours
- Type 3 : Chef de ménage famille élargie résidant dans l'habitat moderne
- Type 4 : Chef de ménage famille élargie résidant dans les cours
- Type 5 : Chef de ménage féminin
- Type 6 : Chef de ménage famille nucléaire résidant dans l'habitat de cour
- Type 7 : Groupement de célibataire cour et sommaire
- Type 8 : Chef de ménage famille nucléaire résidant dans l'habitat sommaire
- Type 9 : Ménage de une personne résidant dans l'habitat de cour ou le sommaire.

Cette typologie est loin d'être fine mais elle couvre les principales catégories perceptibles à Abidjan. Les résultats obtenus se présentent sous forme de tableaux (Cf page suivante). Le premier donne la liste des individus étalons des noyaux. Le second tableau, le plus complet, donne pour chaque classe son effectif et la liste exhaustive des numéros des individus appartenant à chacune

TABLEAU 1

LECTURE DES ETALONS NUMERO 1

1.1	10.1	5049	5965	5953	5755	5186	3409	5750	5188	6638	5212
2.1	10.1	653	1566	7894	493	6079	1554	1671	9405	6662	2698
3.1	10.1	3610	2142	3594	7221	3617	2361	7202	3613	2130	2464
4.1	10.1	9282	8606	8260	7890	8012	9179	654	21	179	8507
5.1	10.1	8872	8785	5499	605	8194	184	2917	1550	1488	8532
6.1	10.1	2738	4456	3116	2991	2988	2755	2982	3098	1130	9321
7.1	10.1	3464	2168	4470	3067	3065	1641	1416	1510	4456	4474
8.1	10.1	4165	4030	4020	3899	3847	3831	4022	3809	4173	3966
9.1	10.1	3481	3171	1542	3062	1511	2951	1676	2039	8951	2986

ITERATION NUMERO 1

INDIVIDUS DES CLASSES	En entrée numéros des éléments étalons										En sortie																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																									
	a) Individus par classe à chaque itération																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																			
1	266	303	399	491	581	671	761	851	941	1031	1121	1211	1301	1391	1481	1571	1661	1751	1841	1931	2021	2111	2201	2291	2381	2471	2561	2651	2741	2831	2921	3011	3101	3191	3281	3371	3461	3551	3641	3731	3821	3911	4001	4091	4181	4271	4361	4451	4541	4631	4721	4811	4901	4991	5081	5171	5261	5351	5441	5531	5621	5711	5801	5891	5981	6071	6161	6251	6341	6431	6521	6611	6701	6791	6881	6971	7061	7151	7241	7331	7421	7511	7601	7691	7781	7871	7961	8051	8141	8231	8321	8411	8501	8591	8681	8771	8861	8951	9041	9131	9221	9311	9401	9491	9581	9671	9761	9851	9941	10031	10121	10211	10301	10391	10481	10571	10661	10751	10841	10931	11021	11111	11201	11291	11381	11471	11561	11651	11741	11831	11921	12011	12101	12191	12281	12371	12461	12551	12641	12731	12821	12911	13001	13091	13181	13271	13361	13451	13541	13631	13721	13811	13901	13991	14081	14171	14261	14351	14441	14531	14621	14711	14801	14891	14981	15071	15161	15251	15341	15431	15521	15611	15701	15791	15881	15971	16061	16151	16241	16331	16421	16511	16601	16691	16781	16871	16961	17051	17141	17231	17321	17411	17501	17591	17681	17771	17861	17951	18041	18131	18221	18311	18401	18491	18581	18671	18761	18851	18941	19031	19121	19211	19301	19391	19481	19571	19661	19751	19841	19931	20021	20111	20201	20291	20381	20471	20561	20651	20741	20831	20921	21011	21101	21191	21281	21371	21461	21551	21641	21731	21821	21911	22001	22091	22181	22271	22361	22451	22541	22631	22721	22811	22901	22991	23081	23171	23261	23351	23441	23531	23621	23711	23801	23891	23981	24071	24161	24251	24341	24431	24521	24611	24701	24791	24881	24971	25061	25151	25241	25331	25421	25511	25601	25691	25781	25871	25961	26051	26141	26231	26321	26411	26501	26591	26681	26771	26861	26951	27041	27131	27221	27311	27401	27491	27581	27671	27761	27851	27941	28031	28121	28211	28301	28391	28481	28571	28661	28751	28841	28931	29021	29111	29201	29291	29381	29471	29561	29651	29741	29831	29921	30011	30101	30191	30281	30371	30461	30551	30641	30731	30821	30911	31001	31091	31181	31271	31361	31451	31541	31631	31721	31811	31901	31991	32081	32171	32261	32351	32441	32531	32621	32711	32801	32891	32981	33071	33161	33251	33341	33431	33521	33611	33701	33791	33881	33971	34061	34151	34241	34331	34421	34511	34601	34691	34781	34871	34961	35051	35141	35231	35321	35411	35501	35591	35681	35771	35861	35951	36041	36131	36221	36311	36401	36491	36581	36671	36761	36851	36941	37031	37121	37211	37301	37391	37481	37571	37661	37751	37841	37931	38021	38111	38201	38291	38381	38471	38561	38651	38741	38831	38921	39011	39101	39191	39281	39371	39461	39551	39641	39731	39821	39911	40001	40091	40181	40271	40361	40451	40541	40631	40721	40811	40901	40991	41081	41171	41261	41351	41441	41531	41621	41711	41801	41891	41981	42071	42161	42251	42341	42431	42521	42611	42701	42791	42881	42971	43061	43151	43241	43331	43421	43511	43601	43691	43781	43871	43961	44051	44141	44231	44321	44411	44501	44591	44681	44771	44861	44951	45041	45131	45221	45311	45401	45491	45581	45671	45761	45851	45941	46031	46121	46211	46301	46391	46481	46571	46661	46751	46841	46931	47021	47111	47201	47291	47381	47471	47561	47651	47741	47831	47921	48011	48101	48191	48281	48371	48461	48551	48641	48731	48821	48911	49001	49091	49181	49271	49361	49451	49541	49631	49721	49811	49901	50000	50090	50180	50270	50360	50450	50540	50630	50720	50810	50900	50990	51080	51170	51260	51350	51440	51530	51620	51710	51800	51890	51980	52070	52160	52250	52340	52430	52520	52610	52700	52790	52880	52970	53060	53150	53240	53330	53420	53510	53600	53690	53780	53870	53960	54050	54140	54230	54320	54410	54500	54590	54680	54770	54860	54950	55040	55130	55220	55310	55400	55490	55580	55670	55760	55850	55940	56030	56120	56210	56300	56390	56480	56570	56660	56750	56840	56930	57020	57110	57200	57290	57380	57470	57560	57650	57740	57830	57920	58010	58100	58190	58280	58370	58460	58550	58640	58730	58820	58910	59000	59090	59180	59270	59360	59450	59540	59630	59720	59810	59900	60000	60090	60180	60270	60360	60450	60540	60630	60720	60810	60900	60990	61080	61170	61260	61350	61440	61530	61620	61710	61800	61890	61980	62070	62160	62250	62340	62430	62520	62610	62700	62790	62880	62970	63060	63150	63240	63330	63420	63510	63600	63690	63780	63870	63960	64050	64140	64230	64320	64410	64500	64590	64680	64770	64860	64950	65040	65130	65220	65310	65400	65490	65580	65670	65760	65850	65940	66030	66120	66210	66300	66390	66480	66570	66660	66750	66840	66930	67020	67110	67200	67290	67380	67470	67560	67650	67740	67830	67920	68010	68100	68190	68280	68370	68460	68550	68640	68730	68820	68910	69000	69090	69180	69270	69360	69450	69540	69630	69720	69810	69900	70000	70090	70180	70270	70360	70450	70540	70630	70720	70810	70900	70990	71080	71170	71260	71350	71440	71530	71620	71710	71800	71890	71980	72070	72160	72250	72340	72430	72520	72610	72700	72790	72880	72970	73060	73150	73240	73330	73420	73510	73600	73690	73780	73870	73960	74050	74140	74230	74320	74410	74500	74590	74680	74770	74860	74950	75040	75130	75220	75310	75400	75490	75580	75670	75760	75850	75940	76030	76120	76210	76300	76390	76480	76570	76660	76750	76840	76930	77020	77110	77200	77290	77380	77470	77560	77650	77740	77830	77920	78010	78100	78190	78280	78370	78460	78550	78640	78730	78820	78910	79000	79090	79180	79270	79360	79450	79540	79630	79720	79810	79900	80000	80090	80180	80270	80360	80450	80540	80630	80720	80810	80900	80990	81080	81170	81260	81350	81440	81530	81620	81710	81800	81890	81980	82070	82160	82250	82340	82430	82520	82610	82700	82790	82880	82970	83060	83150	83240	83330	83420	83510	83600	83690	83780	83870	83960	84050	84140	84230	84320	84410	84500	84590	84680	84770	84860	84950	85040	85130	85220	85310	85400	85490	85580	85670	85760	85850	85940	86030	86120	86210	86300	86390	86480	86570	86660	86750	86840	86930	87020	87110	87200	87290	87380	87470	87560	87650	87740	87830	87920	88010	88100	88190	88280	88370	88460	88550	88640	88730	88820	88910	89000	89090	89180	89270	89360	89450	89540	89630	89720	89810	89900	90000	90090	90180	90270	90360	90450	90540	90630	90720	90810	90900	91000	91090	91180	91270	91360	91450	91540	91630	91720	91810	91900	92000	92090	92180	92270	92360	92450	92540	92630	92720	92810	92900	93000	93090	93180	93270	93360	93450	93540	93630	93720	93810	93900	94000	94090	94180

des classes. Le troisième tableau présente la liste des individus centraux de chacune des classes, correspondant à la sélection désirée (10 individus par classe dans notre cas). Deux tableaux d'aide à l'interprétation sont fournis également (non présentés ici) : une mesure d'homogénéité des classes (variance) et une mesure de la valeur du critère pour chaque partition obtenue. Enfin, quand la convergence est atteinte, un tableau des individus des formes fortes figure en fin de listing.

Il faut bien distinguer entre les éléments centraux obtenus en sortie et l'ensemble des individus composant la classe. Un simple retour au fichier permet de connaître les caractéristiques des individus tirés, qui ont pour une même classe des caractéristiques voisines. Par contre l'ensemble des individus composant une même classe peut être plus hétérogène.

Le tableau 2 donne la répartition des 9 classes selon l'habitat et selon la catégorie socio-professionnelle. Ainsi le type 1 regroupe essentiellement des cadres, indépendamment de l'habitat où ils résident (une minorité logeant dans les cours plus traditionnelles). Dans ce cas, le statut professionnel prime sur le type de logement de résidence. De même, pour la classe 7 qui regroupe exclusivement des chefs de ménage apprentis ou aides familiaux. Il s'agit le plus souvent dans ce cas de jeunes qui logent dans la boutique de leur patron. A l'inverse, la classe 8 rassemble essentiellement des chefs de ménage résidant dans l'habitat sommaire quel que soit le type de famille auquel ils appartiennent. Après analyse des différents tableaux, il s'avère que certains types doivent comporter des intitulés différents. C'est le cas pour les types :

Type 1 : Chef de ménage cadres ou employeur

Type 3 : Chef de ménage famille élargie salarié des secteurs commerce et service moderne

Type 5 : Chef de ménage sans emploi (ce type comprend aussi des hommes, alors que les individus étalons et les individus représentatifs sont des femmes)

Type 6 : Chef de ménage famille nucléaire ou groupement de célibataire résidant dans l'habitat de cour

Type 7 : Chef de ménage apprenti ou aide familial

Type 8 : Chef de ménage résidant dans l'habitat sommaire.

REPARTITION DE LA TYPOLOGIE SELON L'HABITAT ET LES CATEGORIES SOCIO-PROFESSIONNELLES

TABLEAU 2

HABITAT	1	2	3	4	5	6	7	8	9	TOTAL
Résidentiel	163	5	154	27	13	23	2	1	17	405
Economique Moderne	76	5	971	215	42	86	3	7	44	1 449
Cour	25	106	390	1 905	230	1 618	90	0	623	4 987
Sommaire	1	8	14	14	22	0	19	731	144	953

CATEGORIES SOCIO-PROFESSIONNELLES

Sans emploi	4	61	9	4	280	2	0	6	27	393
Cadres - Employeur	261	2	14	41	0	24	0	3	7	352
Exploitant indépendant	0	24	68	579	27	563	0	206	296	1 763
Salarié qualifié commerce - service	0	3	1 157	161	0	105	0	35	93	1 554
Salarié non qualifié commerce - service	0	5	15	139	0	255	0	133	83	630
Salarié qualifié industrie	0	0	250	383	0	194	0	68	121	1 016
Salarié non qualifié industrie	0	6	5	802	0	537	0	240	186	1 776
Aide familial - Apprenti	0	0	0	0	0	0	114	0	0	114
Travailleur de l'agriculture	0	23	11	52	0	47	0	48	15	196
TOTAL	265	124	1 529	2 161	307	1 727	114	739	828	7 794

Ainsi pour certains types la définition doit être revue et l'hypothèse avancée doit donc être modifiée. Il n'en demeure pas moins, dans notre cas, que les types retenus se stabilisent rapidement, 91 % des individus se positionnent dès la première itération et dès la troisième itération chaque individu a rejoint sa partition. A titre d'exemple nous donnons les effectifs de chacune des combinaisons possibles (tableau 3). Le code correspond au type affecté à chacune des trois itérations : le code 463 par exemple correspond à un individu qui rejoint le type 4 à la première itération, le type 6 à la seconde et le type 3 à la dernière itération. Ce tableau permet d'évaluer le nombre des individus charnières, ainsi par exemple 31 individus passent de la classe 9 (ménage 1 personne) à la classe 7 (apprenti ou aide familial). Il n'y a aucune difficulté à lister ces individus afin de les identifier, et éventuellement les retrouver ensuite sur le terrain.

Notre objectif premier était d'obtenir une liste d'individus appartenant à chacun des types. La méthode de sélection a répondu à notre attente, toutefois dans notre cas le temps écoulé entre l'enquête initiale et l'investigation sociologique est trop long : cinq ans, mais malgré ce handicap, nous avons pu retrouver de nombreux chefs de ménage. Reste à s'interroger sur la notion d'individu central d'une classe, d'individu moyen. Effectivement, pour une classe donnée, les principales caractéristiques retenues sont semblables (par exemple pour les vieux : l'âge, l'habitat, la date d'installation à Abidjan, ...), mais derrière ces indications ponctuelles se cachent des vies bien singulières. Il est évident que la méthode ne permet de sélectionner les individus qu'en fonction des variables introduites et si les personnes sont interrogées sur un autre sujet (les conditions de leur insertion à Abidjan par exemple) les réponses divergent. Mais même si le passé diffère, des traits saillants communs apparaissent comme l'importance du tuteur au moment de l'adolescence, le rôle joué par l'épouse, etc.

4 - INTERET ET LIMITES DE LA METHODE

Cette méthode permet d'obtenir une représentation synthétique de grands ensembles de données en utilisant un temps de calcul relativement limité. Elle fournit de façon plus lisible le classement de chaque individu au sein de la typologie, ce qui constitue, pour cet aspect, un avantage par rapport aux méthodes d'analyse factorielle des correspondances. En effet, la projection des individus

INDIVIDUS STABLES ET INDIVIDUS CHARNIERES

TABLEAU 3

CODE	FREQ	EFF. DEF.	CODE	FREQ	EFF. DEF.	CODE	FREQ	EFF. DEF.
111.	262	265	564.	1		855.	10	
133.	1		566.	2		866.	10	
155.	3		599.	2		877.	1	
222.	124	124	611.	1		886.	1	
333.	1 386	1 529	633.	33		888.	718	739
336.	4		636.	2		911.	2	
344.	30		644.	43		933.	11	
355.	7		646.	2		939.	1	
363.	2		655.	16		944.	4	
366.	11		663.	2		949.	1	
399.	1		664.	74		955.	4	
433.	82		665.	2		959.	1	
444.	1 979	2 161	666.	1 528	1 727	966.	7	
446.	15		686.	1		969.	2	
455.	36		688.	3		977.	31	
463.	1		699.	1		988.	6	
464.	26		766.	1		993.	3	
466.	142		777.	82	114	995.	1	
488.	7		833.	8		998.	5	
544.	1		844.	3		999.	819	828
555.	228	307	846.	1				

sur le plan factoriel est bien souvent d'une lecture difficile lorsque de grands tableaux de données sont pris en considération, la superposition des points constituant une perte d'information.

La partition peut être réalisée, on l'a vu, de deux façons :

- soit à partir d'éléments tirés au hasard, ce qui est utile lorsque aucune information n'est disponible a priori sur le corps des données ; cette procédure semblant un bon préalable à un dépouillement classique ;

- soit à partir d'éléments étalons : la typologie obtenue vérifie ou non les hypothèses préalablement établies. Cette deuxième solution est préférable car, en principe, tout utilisateur a un corps d'hypothèses qu'il désire tester. La typologie peut être réalisée en introduisant des individus fictifs possédant des caractéristiques bien déterminées de façon à obtenir ainsi les individus réels qui s'agrègent à chacun de ces modèles.

Un autre intérêt de la méthode réside dans la possibilité d'extraire la classe affectée à chacun des individus et l'adjoindre au fichier initial. Cette typologie devenant elle-même une nouvelle variable, plus synthétique, permettant d'autres analyses.

A notre sens, la possibilité d'extraire nominalement les individus centraux de chacun des types constitue l'avantage principal de la méthode. Ces individus centraux possèdent en principe les caractéristiques moyennes de la classe. Le nombre de ces éléments est fixé a priori par l'utilisateur. Une identification rapide de ces individus à l'issue de l'enquête permet de réaliser conjointement une investigation plus approfondie sur un sous-échantillon représentatif de chacune des classes obtenues. A l'inverse, les cas marginaux, révélés par leur plus faible capacité à se stabiliser rapidement dans le modèle, méritent aussi d'être étudiés : leur marginalité statistique traduit peut-être une marginalité sociale.

Comme toute méthode d'analyse des données, celle des nuées dynamiques présente quelques limites.

Les modalités comprenant des effectifs peu nombreux tendent à constituer des classes parfois non pertinentes ; mais cette limite est aussi un avantage dans la mesure où certains cas aberrants peuvent être décelés de cette façon, surtout si un fichier de données brutes est utilisé.

Une modalité composant majoritairement une variable occulte les autres possibilités : ainsi nous avons dû constituer, lors du travail préparatoire, un second fichier ne comprenant pas l'habitat de cour, afin de mieux dégager les types intégrant les populations résidant dans les trois autres formes d'habitat.

Des variables trop redondantes risquent de masquer l'influence des autres caractéristiques.

Le choix des étalons laisse une grande part d'arbitraire dans la mesure où certaines variables peuvent être privilégiées à tort. Mais l'utilisation dans un premier temps d'un tirage au hasard des étalons, puis l'élaboration de la typologie à partir des premiers individus tirés limite les risques de manipulation inconsidérée des données.

Cette méthode comporte une part d'arbitraire mais il faut surtout concevoir son utilisation comme le moyen de tester un corps d'hypothèses. Si le choix des étalons est judicieux, ces derniers se retrouvent en grand nombre comme éléments centraux ou pour le moins dans la classe qu'ils ont aidé à déterminer. Si l'hypothèse est fautive, ces individus étalons se dispersent dans diverses classes à l'issue de la partition.

--- ---

BIBLIOGRAPHIE

- ANTOINE Ph., HERRY Cl. - 1983
Quelques éléments pour l'évaluation des résultats de l'enquête à passages répétés menée à Abidjan
STATECO n° 34, juin 1983, INSEE, pp. 15-38.
- DIDAY E. - 1971
Une nouvelle méthode en classification automatique et reconnaissance des formes. La méthode des nuées dynamiques
Revue de Statistiques Appliquées, vol. XX, n° 2, pp. 19-33.
- FLAMENBAUM G., THIERY J., BENZECRI J.P. - 1979
Agrégation en boules de rayon fixe et centre optimisés (boules optimisées)
Les Cahiers de l'Analyse des Données, n° 3, pp. 357-364.
- LEBART L., MORINEAU A., TABARD N. - 1977
Techniques de la description statistique
Dunod, Paris, 351 pages.
- VOLLE M. - 1981
Analyse des données
Economica, Paris, 317 pages.