



**HAL**  
open science

# ANALYSE STATISTIQUE DE DONNÉES BIOGRAPHIQUES : MÉTHODES, PREMIERS RÉSULTATS ET PERSPECTIVES POUR L'ÉTUDE DES MIGRATIONS

Olivier Barbary

► **To cite this version:**

Olivier Barbary. ANALYSE STATISTIQUE DE DONNÉES BIOGRAPHIQUES : MÉTHODES, PREMIERS RÉSULTATS ET PERSPECTIVES POUR L'ÉTUDE DES MIGRATIONS. atelier méthodologique COLEF-ORSTOM sur la mesure de la migration internationale - Tijuana, Mexique 2-4 mai 1994, May 1994, TIJUANA, Mexique. hal-02481334

**HAL Id: hal-02481334**

**<https://hal.science/hal-02481334>**

Submitted on 17 Feb 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Olivier Barbary  
Statisticien ORSTOM,  
Departamento de matemáticas y estadística  
Universidad Nacional de Colombia - Santafé de Bogotá D.C.  
Colombia

Atelier méthodologique COLEF-ORSTOM sur la mesure de la migration internationale - Tijuana, Mexique  
2-4 mai 1994

## **ANALYSE STATISTIQUE DE DONNÉES BIOGRAPHIQUES : MÉTHODES, PREMIERS RÉSULTATS ET PERSPECTIVES POUR L'ÉTUDE DES MIGRATIONS**

### ***Introduction : les enjeux de la collecte de données biographiques***

La collecte de données biographiques se généralise en sciences sociales. Dans l'étude des formes de la mobilité humaine, elle tente de répondre à des questionnements nouveaux, dans des contextes thématiques, disciplinaires et méthodologiques variés. Qu'il s'agisse en effet de la mobilité spatiale, sociale ou économique, de celle des individus ou des groupes sociaux, d'approches démographiques, géographiques, économiques ou socio-anthropologiques, de méthodes quantitatives ou qualitatives, l'objet de la recherche est à chaque fois un phénomène continu. Et ce terme est bien à prendre au sens que lui donnent les mathématiques puisque les vies humaines sont composées d'une infinité d'instantanés, de lieux, d'événements, etc. Dès lors, même si aucun système d'observation ne peut prétendre rendre compte de cette continuité "théorique", l'appareil des concepts et des méthodes de collecte doit viser l'observation la plus exhaustive possible du temps et des espaces dans lesquels se déroule la mobilité des hommes. Le corpus de données biographiques est à l'évidence un mode de recueil de l'information approprié à cet objectif.

La réflexion qui précède n'en est plus aujourd'hui au stade des débats théoriques, du moins à ce niveau de généralité. En effet, les systèmes d'observation statistique de la mobilité humaine - entendons par là le domaine qui s'étend de la définition des populations et des concepts de base de l'observation, à la réalisation d'enquêtes sur des échantillons représentatifs - ont indiscutablement connu de grands progrès récemment. En décembre 1992 à Bogota, l'atelier méthodologique CEDE/ORSTOM sur les nouvelles formes de mobilité des populations urbaines d'Amérique Latine s'est fait l'écho de ces avancées et de leur application à la recherche et à la planification urbaine dans le continent. Un des constats fait par M. Villa en conclusion des débats, qui dépasse d'ailleurs le contexte sud-américain, est que la mise en oeuvre de nouveaux concepts et le développement de nouvelles catégories d'analyse de la mobilité passent par la collecte de données qui permettent une reconstitution longitudinale des trajectoires individuelles. Deux méthodologies, plus complémentaires que concurrentes, peuvent s'appliquer : les questionnaires biographiques rétrospectifs et les enquêtes à passages répétés <sup>1</sup>.

Dans cet atelier consacré à la mesure de la migration internationale, le COLEF et l'ORSTOM nous proposent, comme point de départ à la réflexion, un système d'information déjà opérationnel, conçu autour de l'articulation entre l'enquête sur les migrations à travers la frontière nord du Mexique (observatoire des flux migratoires) et un système d'information géographique et statistique sur les régions d'origine de la migration (SIGEF). Pour le propos de cette communication, c'est l'enquête d'observation des flux qui nous intéresse au premier chef. Elle est décrite par D. Delaunay et J. Santibañez.

---

<sup>1</sup> : Voir les conclusions présentées par M. Villa lors de la session de clôture de l'atelier in "Memorias del taller sobre nuevas formas de movilidad urbanas en America Latina" ed. sc. F. Dureau et C. E. Flores, CEDE/ORSTOM, Bogota, 1992.

*"Une enquête par sondage appliquée à la population mobile des migrants dans une vingtaine de villes frontalières. Le mouvement est saisi tant vers les Etats-Unis que lors du retour des mexicains expatriés. La biographie migratoire de chacun est reconstituée, de même que le parcours exact de ses déplacements et les emplois exercés."*

Il s'agit donc d'une enquête basée sur un questionnaire biographique rétrospectif.

Mais si l'on sait de plus en plus précisément pourquoi et comment collecter ces corpus de données biographiques, de nombreuses difficultés subsistent dans les phases d'analyse de ces données, particulièrement dans l'approche quantitative. On doit alors employer des méthodes d'analyse statistique respectant la richesse et la continuité des données. De l'avis général, il existe à l'heure actuelle un décalage important entre l'évolution des concepts et des techniques de collectes, qui ont permis une avancée significative dans la saisie quantitative des réalités complexes de la mobilité des populations, et le caractère encore très frustré des outils statistiques généralement employés lors de l'exploitation. On s'en tient souvent à des analyses purement transversales et, lorsque ce n'est pas le cas, à l'examen d'indicateurs longitudinaux monovariés, nécessairement réducteurs de l'information originale. Globalement, cet état de fait a pour conséquence une sous-exploitation des enquêtes; la pratique de l'analyse multivariée des données biographiques est encore loin d'être généralisée.

Il faut dire à la décharge des équipes de chercheurs confrontés à ce problème, que les résultats applicables de la recherche statistique sur ces questions sont eux-mêmes assez récents, à cause de la structure complexe des données. Généralement formalisée mathématiquement à l'aide de la théorie des processus stochastiques, l'analyse statistique des données biographiques sur gros échantillons est un domaine resté longtemps assez théorique ou d'application très spécifique : ce fut le cas par exemple des travaux de Cox sur l'analyse des données de survie.

Aujourd'hui cependant, les acquis méthodologiques en termes de techniques d'analyse des données biographiques se sont diversifiés et ils deviennent très prometteurs. C'est ce que nous tenterons de montrer dans cet exposé, en insistant particulièrement sur leur application à l'étude de la mobilité spatiale des populations.

La première partie dresse, en des termes schématiques et généraux, qui ne font aucun usage de formalisme mathématique, le panorama des débats et des résultats qui caractérisent la recherche statistique actuelle dans ce domaine. Dans la seconde partie, nous illustrerons l'application de certaines techniques simples d'analyse multivariées à l'étude de la mobilité spatiale. Nous présenterons des exemples de méthodes et de résultats types dans l'analyse historique et géographique des migrations vues depuis les villes d'accueil. Enfin, la troisième partie aborde les perspectives qu'ouvrent l'emploi de méthodes plus sophistiquées, prenant l'exemple de l'analyse harmonique qualitative. Elle permet d'élargir la discussion aux problèmes de typologie d'itinéraires biographiques (en particulier migratoires) et d'identification des réseaux.

On l'aura compris, l'objet de recherche présenté ici est essentiellement méthodologique, donc par nature transverse à un champ d'application assez vaste. Le thème des enquêtes à partir desquelles nous testons ces nouvelles méthodes n'est pas spécifiquement celui des migrations internationales. Il se peut par conséquent que les méthodes et les exemples que nous allons présenter, ne soient pas directement transposables à ce thème, tel que l'étudient l'ORSTOM et le COLEF dans leur programme sur la frontière nord du Mexique. La raison principale en serait que notre point de vue sur les phénomènes migratoires s'est toujours situé depuis les villes d'accueil, ce qui n'est pas le cas dans l'observatoire des flux conçu par le programme COLEF/ORSTOM. Cependant d'un point de vue ou de l'autre, ce que l'on observe sont des itinéraires.

## 1. Méthodologies d'analyse statistique des données biographiques

### 1.1 Le problème posé : respecter au mieux la continuité des temps et des espaces biographiques

Sans entrer dans l'exposé des problématiques de sciences sociales dans l'étude de la mobilité humaine, il faut dire rapidement comment se pose au statisticien le problème de l'exploitation des données biographiques dans ce cadre. On comprendra mieux pourquoi les questionnements thématiques des chercheurs en sciences sociales génèrent un objet de recherche proprement statistique.

Si l'on ne fait usage d'aucun formalisme mathématique, le problème est simple. Pour identifier et qualifier les processus de la mobilité des individus (mobilité spatiale, professionnelle, sociale etc.), il faut conserver, tout au long de l'analyse des données biographiques, l'ordre de la séquence d'événements qu'ils connaissent et les durées passées dans chacun des états (états résidentiels, professionnels, sociaux etc.). La caractérisation globale des itinéraires que nous cherchons ne peut pas découler de l'analyse transversale séparée de chaque type d'événement : premier changement de résidence, premier emploi, résidence ou emploi à tel âge ou telle date etc. En effet, on aboutirait ainsi à caractériser les individus indépendamment pour chaque événement considéré. Le croisement de ces typologies, outre qu'il risque de générer des classes d'effectifs trop faibles, regroupera des individus ayant certes connu un même ensemble d'événements, mais éventuellement dans des ordres, à des dates et avec des durées de séjour dans chaque état totalement différents. On ne pourra de ce fait rien conclure sur l'homogénéité des itinéraires des individus du groupe, pas plus que sur le type de dépendance statistique existant éventuellement entre ces événements. Comme le disent D. COURGEAU et E. LELIEVRE [1989] : "On voit dès lors que l'unité d'analyse ne doit plus être l'événement mais la biographie individuelle, considérée comme un processus complexe".

### 1.2 Approche théorique : décrire avant de modéliser

Posée en ces termes, l'analyse statistique des biographies est, comme nous l'avons dit, un domaine de recherche relativement récent. Dans son développement actuel, on peut distinguer deux approches théoriques. Certains les jugent plus complémentaires que concurrentes. Elles visent cependant des buts radicalement différents et font l'objet d'un débat d'école, parfois plus sectaire que scientifique, mais dont l'enjeu est, pensons-nous, important. Présentons d'abord rapidement les deux approches.

L'approche probabiliste est principalement développée par les statisticiens démographes, ou travaillant à des applications cliniques, à partir de modèles non-paramétriques, paramétriques ou semi-paramétriques (par exemple COX [1975], COURGEAU & LELIEVRE [1989] ...). Leur point de départ est la formalisation des biographies individuelles comme des processus aléatoires : la biographie d'un individu est représentée par une suite de variables aléatoires  $T_1, T_2, \dots, T_n$ , qui sont les durées de séjour dans les divers états qui la composent, pris dans l'ordre chronologique. La démarche consiste à estimer, à partir des données de l'enquête, un modèle de distribution (paramétrique ou non) de chacune de ces variables, pour tenter ensuite de modéliser la distribution plus complexe de l'ensemble de la trajectoire (distribution conjointe). Les méthodes statistiques et les programmes informatiques que produit cette approche visent à réaliser une analyse fine des dépendances probabilistes entre événements démographiques et à apprécier les effets des caractéristiques individuelles (sociales, économiques, culturelles etc.) et des événements "extérieurs" (contexte économique, socio-politique, naturel etc.) sur les durées de séjour dans des états donnés. Mais cela n'est possible qu'à la condition qu'on s'intéresse à un système d'événements et d'états en nombre limité et précisément défini a priori. Cette approche ne fournit pas d'outils exploratoires, descriptifs ou typologiques, des trajectoires individuelles.

L'approche par l'analyse des données utilise les techniques d'analyse exploratoire des grands tableaux (analyses factorielles, classifications automatiques), qui ont souvent été appliquées aux séries chronologiques synchrones (principalement en économie). Mais le problème est en fait totalement différent ici. Contrairement aux séries chronologiques habituelles, les données des biographies individuelles sont d'une part asynchrones : il ne s'agit pas d'une série de "mesures" effectuées à des dates fixe pour tous, mais au contraire d'un ensemble de séquences d'événements datés ayant, pour chaque individu, son "horloge" propre. D'autre part les nombres d'événements et l'ordre dans lequel ils surviennent, sont a priori différents d'un individu à l'autre. Du fait de ces caractéristiques particulières des données, c'est relativement tard, il y a environ vingt ans, que les théoriciens et les praticiens de l'analyse des données se sont attaqués à l'élaboration, au test et à l'application d'outils spécifiques pour

traiter le temps passé dans des situations successives. Leurs résultats et l'utilisation qu'on peut en faire dans les études sur la mobilité qui nous intéressent, ne visent pas la modélisation des trajectoires biographiques mais plus modestement leur description statistique, caractérisation et typologie. Je donnerais plus loin quelques principes de certaines de ces méthodes.

En étant prétentieux, on pourrait dire que l'alternative théorique se situe entre une approche structuraliste et une approche déterministe, dans la mesure où l'on peut assimiler grossièrement à ces deux courants de pensée les démarches de la statistique descriptive (structuraliste) et probabiliste (déterministe). En d'autres termes, doit-on d'abord décrire les réalités que permet d'appréhender l'enquête et tenter d'y déceler des relations structurantes, ou bien modéliser les phénomènes que l'on croit sous-jacents aux données, ces dernières ne servant alors qu'à l'inférence de lois déterminantes ?

J.P. Benzecri, dans des écrits parfois polémiques, mais dont on ne peut contester la profondeur de l'argumentation - du moins si l'on se réfère aux ouvrages fondamentaux qu'il a consacré à l'analyse des données, apporte des pièces maîtresses à ce débat. Il n'est pas question ici de rendre compte de l'ensemble des principes qui, selon lui, doivent présider à la pratique moderne de l'analyse statistique, et encore moins de leurs justifications mathématique ou épistémologiques. J'en retiens un seul qui me paraît s'appliquer particulièrement au traitement de l'information biographique.

*"Le modèle doit suivre les données, non l'inverse. Autre trait fâcheux des mathématiques appliquées aux sciences humaines : l'abondance de modèles, forgés a priori puis confrontés aux données par ce qu'on appelle des "tests". Et tantôt le "test" sert à justifier un modèle ou il y a plus de paramètres à ajuster que l'on a déterminé de données. Tantôt au contraire il sert à rejeter comme invalides les plus judicieuses remarques de l'expérimentateur. Mais ce dont nous avons besoin, c'est d'une méthode rigoureuse qui extraie des structures à partir des données."*<sup>2</sup>

C'est dans la même logique que M. Vole fait des techniques descriptives et typologiques de l'analyse des données un préalable aux approches statistiques probabilistes :

*"On remarquera que la référence au calcul des probabilités est pratiquement absente de l'analyse des données : celle-ci se situe dans une étape du raisonnement logiquement antérieure à l'induction probabiliste (...). Pour un praticien de l'analyse des données, ce n'est qu'après avoir identifié les structures qui sous-tendent les données, après les avoir classées et découpées, que l'on peut s'engager dans leur approche probabiliste."*<sup>3</sup>

En fin de compte, ce sont ces mêmes principes qui guident le programme de recherche que j'ai entrepris. La recherche bibliographique initiale portait effectivement sur les deux approches théoriques. Mais, outre qu'elle offre un champ de recherche beaucoup moins exploré que l'approche modélisatrice, l'approche par l'analyse des données semble plus adaptée à une phase de la recherche sur les formes de la mobilité humaine où la priorité est de décrire les phénomènes et d'en identifier les déterminants possibles. Il s'agit en effet d'un contexte où les successions d'événements observés sont très complexes et où l'univers des variables potentiellement explicatives est foisonnant. Dans ce domaine de recherche, je reprendrais volontier l'argument de J.P. Benzecri : on est pas encore en mesure de formuler des hypothèses précises, pour les valider (ou les invalider) à l'aide de modèles et de tests statistiques. Comment, de surcroît, dans le cadre des sciences sociales et de "plans d'expériences" si peu contrôlés (parce que incontrôlables !), peut-on être raisonnablement sûr que les hypothèses probabilistes qui fondent l'inférence sur le modèle ne soit pas partiellement ou totalement contredites par les données ? Doit-on même employer le terme de plan d'expérience s'agissant de ces domaines de recherche, lorsqu'on songe aux protocoles d'échantillonnage et de mesure utilisés dans les sciences physiques ou biologiques, où s'est fait le développement et le succès des théories statistiques de l'estimation et du test d'hypothèse. Pour toutes ces raisons, j'ai choisi de donner la priorité à une démarche exploratoire, à l'aide d'outils statistiques typologiques, qui visera entre autre à définir plus précisément des systèmes d'événements et des sous-populations pouvant faire l'objet d'une approche modélisatrice<sup>4</sup>.

<sup>2</sup> : J.P. Benzecri & collaborateurs, L'analyse des données -tome 2 : l'analyse des correspondances, p 6, Dunod, Paris, 1982.

<sup>3</sup> : M. VOLLE, Analyse des données, p 21, Economica, Paris, 1981.

<sup>4</sup> : Pour toutes ces raisons aussi, je n'ai pas voulu ici détailler, même partiellement, la bibliographie sur l'approche modélisatrice, très riche puisqu'elle couvre en fait tous les développements théoriques fait sur les processus stochastiques appliqués aux données de survie ("failure time data" ou "failure time series" dans la littérature anglo-saxonne).

### 1.3 En statistique descriptive, des pistes méthodologiques variées

Je vais me limiter maintenant à un panorama très rapide et non exhaustif de quelques acquis récents dans le domaine qu'on peut appeler l'analyse exploratoire de données chronologiques individuelles. Théories ou techniques d'analyse factorielle ou de classification automatique, il me semble que chacune d'elles, directement ou indirectement (certains articles ne traitent pas uniquement ou spécifiquement de calendriers biographiques), peuvent apporter une contribution au problème de l'analyse typologique des biographies.

J. PICARD [1987] introduit ainsi la méthode de classification de profils individuels évolutifs qu'il a développée :

*"L'analyse de l'évolution multivariée d'individus se ramène à l'étude des structures d'un tableau tridimensionnel dont les trois indices correspondent respectivement aux individus, aux variables et au temps. Les méthodes utilisées pour cette étude reposent essentiellement sur deux types d'approches factorielles :*

*- la première approche a été proposée par LEBART [1969] sous le nom d'analyse factorielle locale (AFL) qu'il a appliquée à des contiguïtés spatiales. Indépendamment, LE FOLL [1982] introduit l'analyse factorielle des évolutions (AFE) puis en fait une généralisation sous le nom d'analyse factorielle pondérée (AFP). Dans un article récent, CARLIER [1985] reprend l'AFE et montre que c'est une ACP sur un tableau d'observations associé à un graphe descriptif des contiguïtés, comme l'a fait LEBART pour l'AFL."*

Les données biographiques migratoires sont un exemple de données que l'on peut considérer comme se réalisant sur des graphes de contiguïté (graphe des contiguïtés temporelles ou spatiales). Cette première classe de méthodes d'analyses factorielles est construite autour de l'idée que, bien souvent, l'évolution temporelle ou spatiale des données observées n'est pas indépendante ces structures de graphe. Leur pratique est fondée sur la comparaison des résultats de l'analyse en composantes principales (ACP) de la matrice de variance-covariance des données, avec ceux des ACP des matrices de contiguïté à différents niveaux. L'application de ces méthodes au cas des biographies migratoires devrait permettre la mise en évidence de l'échelle et de la localisation des liaisons spatiales et temporelles entre les données.

PICARD poursuit :

*"- la deuxième approche est celle de la méthode STATIS développée par L'HERMIER DES PLANTES et ESCOUFIER [1976, 1978, 1980] dont un prolongement très intéressant de FOUCAULT [1983] aboutit à l'analyse factorielle des opérateurs."*

L'outil statistique principal qui utilise les méthodes multi-tableaux STATIS est l'analyse en composante principale de différents tableaux de distances construit à partir des différentes dimensions des tableaux de données de départ : distances inter-individus, inter-variables, inter-dates. Il s'agit en fait d'une "boîte à outils" complète d'analyse typologique, qui peut également déboucher sur des procédures de classification automatique des trajectoires individuelles.

Mais il faut ajouter deux autres approches factorielles de ces problèmes à l'inventaire de PICARD.

La première est l'analyse harmonique qualitative, une application particulière de l'analyse des correspondances multiples, développée et justifiée par DEVILLE et SAPORTA dans deux articles théoriques [1980, 1982]. Cette méthode nous paraît une des plus prometteuses dans le champ de l'analyse typologique des itinéraires biographiques. Nous y reviendrons à la fin de cet exposé.

La seconde est certainement la plus simple de toutes; elle ne doit pas pour autant être oubliée. Elle consiste à appliquer des techniques d'analyse multivariées classiques à certains tableaux extraits de l'information biographique "complète", qui se prêtent alors à une technique d'analyse spécifique : ACP, Analyse Factorielle des Correspondances simple ou multiple (AFC ou AFCM), Classification Ascendante Hiérarchique (CAH) etc. Avec cette stratégie on peut par exemple aborder l'analyse longitudinale d'événements biographiques particuliers comme l'arrivée des migrants dans leur première

résidence dans la ville d'accueil, l'entrée dans la vie active, la première union etc. En revenant à l'objectif d'une typologie générale des données complètes, on peut également soumettre à l'analyse des correspondances multiples classique un tableaux où l'on a résumé les trajectoires individuelles à étudier à l'aide d'un certain nombre de variables quantitatives ou qualitatives qui les décrirons synthétiquement. On perd alors l'ordre exact et les dates précises des événements et tout repose sur un choix judicieux des résumés. Un excellent exemple de résultats intéressants obtenus grâce à cette méthode se trouve dans MARPSAT [1984]. Comme illustration de ces démarches méthodologiques simples, nous passerons en revue quelques exemples.

## **2. Exemples d'applications et résultats types**

Depuis le démarrage, en 1991, de ce programme de recherche méthodologique, nous nous sommes appuyé, pour les applications, sur deux corpus de données issus d'enquêtes démographiques par sondages comprenant différents modules biographiques :

- l'enquête "migrations" réalisée en 1987 à Quito par l'ORSTOM sous la direction de F. Dureau.
- l'enquête "insertion urbaine" réalisée en 1989 à Dakar par L'IFAN et l'ORSTOM sous la direction de Ph. Antoine.

Les exemples qui suivent sont tirés de l'exploitation de ces deux enquêtes.

### 2.1 Exemple d'analyse longitudinale d'un événement particulier : la première installation des migrants en ville

Dans les problématiques d'étude de l'insertion urbaine, à Quito comme à Dakar, le moment décisif que constitue dans la vie des migrants l'étape d'installation en ville apparaît particulièrement important puisqu'il détermine le point de départ de leurs trajectoires dans la ville. Nous avons donc choisi de consacrer à cet événement une étude statistique spécifique avec, dans chaque ville, un objectif d'analyse historique : décrire l'évolution de la géographie des zones d'accueil des migrants.

Les données et méthodes de cette analyse ont des structures très simples<sup>5</sup>. on part du tableau qui croise la localisation des premières résidences des migrants dans la ville avec leurs dates d'arrivée. Cette information provient du recueil, dans le questionnaire biographique des enquêtes, du quartier de résidence pour toutes les étapes résidentielles en ville connues par les individus de l'échantillon. On ne retient, pour cette analyse, que la première étape. A Quito par exemple, la population analysée comprends 1587 individus migrants, chefs de ménage au moment de l'enquête, pour lesquels sont connus la date de première arrivée à Quito et le quartier de cette première résidence dans la capitale. Cet échantillon est représentatif de la population des chefs de ménages migrants résidents à Quito en 1987. Pour constituer, à partir de cette population, un tableau de contingence où les effectifs soient suffisants dans chaque case, les quartiers de résidence ont été regroupé en 20 zones et les dates d'arrivée en 11 périodes quinquennales. C'est ce tableau 20x11 qui est soumis à l'AFC.

La démarche d'interprétation des résultats est classique. Elle repose d'abord sur l'examen des plans factoriels des AFC, et peut ensuite donner lieux à plusieurs types d'illustration graphique des résultats.

A Dakar, le premier plan factoriel de l'analyse montre très bien que la géographie des zones d'accueil de la migration dans l'agglomération s'est radicalement modifiée depuis 1945 (graphique 1). L'allure parabolique du nuage, classique en AFC (effet GUTMAN), signale une forme presque diagonale du tableau de données, autrement dit un transfert régulier des zones d'installation des migrants au cours du temps, depuis les quartiers d'accueil "anciens" jusqu'aux "modernes" en passant par des quartiers dont la fonction d'accueil a connu un maximum à un moment donné de la période d'observation. La structure particulièrement claire des données, permet de fonder sur les deux premiers facteurs une typologie très satisfaisante des 14 quartiers, selon l'évolution de leur rôle dans l'accueil de la migration. Cette typologie

---

<sup>5</sup> : En revanche, lors de l'application, de multiples précautions s'imposent, que je ne peux pas détailler ici. Elles sont liées d'une part aux différents types de biais et troncatures qui affectent les données d'enquêtes rétrospectives, d'autre part aux caractéristiques du sous-échantillon biographique. Pour plus de détails, voir par exemple : Barbary (1993) pp 2-9 et pp 31-35.

en 5 classes est illustrée par le graphique 2, qui synthétise l'évolution spatiale et temporelle de la fonction d'accueil de la migration dans Dakar.

Suivant l'extension spatiale de la ville, avec toutefois un certain décalage temporel, l'accueil de la migration se décentralise tout d'abord du Plateau et de la Médina vers les banlieues "intégrées" que sont grand Dakar, HLM et Pikine ancien (de 1950 à 1969). A partir de la fin des années soixante, le déplacement se poursuit vers les extensions "régulières" et "irrégulières" de ces banlieues : Patte d'Oie, Cambérène, Médina Gounass, Thiaroye, Guédiawaye, Yeumbeul et Malika.

On peut également choisir une représentation cartographique du phénomène (graphique 3), indépendamment d'ailleurs de la démarche d'analyse factorielle, puisque ce sont directement les données du tableau de contingence qui sont cartographiées (une carte par date). L'expérience montre cependant que l'analyse typologique apporte une aide précieuse, voire indispensable, lors du choix des unités spatiales et des dates que l'on doit représenter. D'autre part, la description statistique du phénomène de transfert spatio-temporel illustré par la carte permet aussi une meilleure structuration du commentaire. Enfin le recours à l'analyse factorielle des données devient incontournable si l'une ou l'autre des dimension du tableau (nombre de dates ou nombre d'unités spatiales) augmente beaucoup.

**Graphique 1** : évolution des zones d'accueil des migrants depuis 1945, le premier plan factoriel de l'AFC

**Graphique 2** : évolution des zones d'accueil des migrants depuis 1945, typologie en cinq groupes de quartiers

(1x2)

**Graphique 3** : évolution des zones d'accueil des migrants depuis 1945, cartes de répartition

## 2.2 La même méthodologie factorielle appliquée à l'analyse de "transitions migratoires" <sup>6</sup>

La démarche d'analyse factorielle que je viens de présenter s'applique identiquement au tableau croisant les lieux d'origine des migrants avec leurs dates d'arrivées. Les résultats conduisent d'abord à la mise en évidence de transitions migratoire à l'échelle "macro" (nationale et internationale).

**Graphique 4 :** Répartition des dates d'arrivées à Quito des populations migrantes originaires de trois sous ensembles géographiques, par tranche quinquennale depuis 1935

Ainsi à Quito (graphique 4), la composition géographique de la migration s'est radicalement modifiée en vingt ans, entre 1960 et 1980. Durant cette période l'espace qui alimente la majorité de la migration vers la ville s'est transféré de l'arrière-pays rural (la sierra rurale centre et nord-équatorienne fournissait 55% des migrants en 1960, moins de 30% à partir de 1980), vers un espace plus lointain, intégrant la côte, les campagnes de la sierra sud et l'étranger (20% de la migration en 1960, plus de 40% à partir de 1980). La part relative de l'espace urbain de la sierra, qu'on pourrait appeler intermédiaire en terme de distance géoculturelle à la capitale, est demeurée stable durant cette période (25% à 30% des migrants).

A Dakar (graphique 5), l'évolution est moins rapide et s'étale sur toute la période d'observation. Avant 1960, la migration en provenance de l'arrière-pays rural et du nord du pays domine largement (autour de 80% du total). L'espace nord-sénégalais joue alors un rôle très important : dépassant 50% du flux migratoire total vers Dakar entre 1935 et 1939 et de 1945 à 1949, sa part relative est encore de 42% entre 1955 et 1959. Quant au troisième espace migratoire (Sud Sénégal et Etranger), il n'a avant 1950 qu'un rôle relativement mineur : moins de 20% du flux migratoire total. C'est précisément la croissance presque continue depuis 1945 de la migration originaire du Sud du Sénégal et de l'étranger qui constitue l'évolution majeure dans l'histoire récente de la migration vers Dakar. La part de ces migrations a crû très fortement de 1945 à 1964, passant de 14% à 32%. Elle rejoint à ce moment les migrations originaires des autres espaces, puis les dépasse pour atteindre 43% de la migration totale en 1989. Cette croissance s'est accomplie presque exclusivement (sauf entre 1970 et 1974) au détriment de la migration en provenance du nord du pays, qui ne représente plus que 23% de la migration en 1989. En revanche, la part de la migration en provenance du bassin arachidier et du pays Sérère est restée globalement stable depuis 1950 : elle est de 34% en 1989.

---

<sup>6</sup> : Nous donnons ici à ce terme le sens très général d'une modification durable (à l'échelle de la période historique que permet d'observer la donnée rétrospective), de la répartition des régions d'origine des flux migratoires vers la ville étudiée. L'usage, essentiellement pratique, de cette expression ne renvoie pas à une analogie stricte avec la définition plus précise que lui donnent certains auteurs dans l'analyse historique de la transition démographique en Europe par exemple; il s'agit alors du renversement structurel de tendance, au cours du vingtième siècle, qui a fait passer les pays européens de l'émigration nette à l'immigration nette (voir par exemple J.C. CHESNAIS -1986- Chap. 6).

*Graphique 5 : Répartition des dates d'arrivées à Dakar des populations migrantes originaires de trois sous ensembles géographiques, par tranche quinquennale depuis 1935*

L'unicité de la structure des données et de la méthode d'analyse facilite évidemment les études comparatistes. Similarités et différences entre les deux contextes, équatorien et sénégalais, apparaissent dès les premiers résultats de l'analyse. Dans les deux cas, un transfert bien marqué des lieux d'origine de la migration au profit de régions plus lointaines. Cependant les modalités temporelles et spatiales de cette transition diffèrent d'un pays à l'autre, par exemple quant au rôle relatif et à l'évolution du flux migratoire en provenance de l'hinterland rural des deux villes. On est alors renvoyé à des questions de structuration spatiale et économiques des deux espaces nationaux : histoire du développement urbain dans les deux pays (monocéphalie au Sénégal, bicéphalie en Equateur), histoire des crises de l'économie rurale etc.

Mais l'identification du schéma global d'évolution des flux migratoires vers la ville étudiée n'est pas le seul intérêt de cette méthodologie. Elle permet également, si l'on inverse le point de vue et que l'on s'intéresse à l'évolution du flux en provenance d'une région d'origine donnée, d'aborder la question des déterminants de la migration dans les zones de départ. On doit bien sûr, pour que ce point de vue aie un intérêt, ne pas s'en tenir aux grands agrégats très hétérogènes que nous considérons précédemment. C'est l'un des mérites de la méthode que de fournir plusieurs typologies d'évolution des flux dont les plus fines sont constituées d'espaces relativement homogènes, du point de vue de la dynamique du flux migratoire comme du point de vue géographique et socio-économique <sup>7</sup>. Enfin, à ce niveau d'analyse, il est également intéressant d'ajouter à la description de la dynamique du flux, la caractérisation de sa composition démographique. Pour montrer à quel type de commentaire une telle démarche peut aboutir, je prendrais l'exemple des villes de l'arrière-pays dakarois où l'on observe la stabilisation et la féminisation de l'émigration vers Dakar.

Les villes des régions de Thiès et Diourbel (Thiès, Mbour, Tivaouane, Bambey, Diourbel, Touba, Mbaké etc.) constituent un espace assez particulier et homogène. Son intégration économique et sa grande proximité géographique et culturelle avec Dakar en font "l'hinterland urbain" de la capitale. L'expression est particulièrement justifiée jusqu'en 1950 si l'on considère (graphique 6a) l'importance qu'avaient ces villes dans le flux migratoire total vers Dakar (plus de 20% sauf entre 1940 et 1945). Depuis 1935, on assiste à une évolution en deux phases : une forte baisse de la contribution relative à l'immigration totale de 1935 à 1960 (33% à 13%), suivie d'une stagnation jusqu'à nos jours (10,2% entre 1985 et 1989). En effectifs, ce mouvement correspond à une légère chute entre 1935 et 1955, puis à une croissance lente depuis 1960 (1100 migrants/an en 1970, 2000 en 1985).

---

<sup>7</sup> : Le degré de finesse auquel permet d'accéder la technique est bien sûr fonction d'une part de la précision avec laquelle on a collecté l'information sur les lieux d'origine des migrants, d'autre part de la taille de l'échantillon.

**Graphiques 6a et 6b :** *Part relative dans la migration totale vers Dakar (a) et structures par sexe et âge à la dernière installation à Dakar (b) des populations migrantes originaires des villes de l'arrière-pays de Dakar*

Graphique 8a

Graphique 8b

Dans le contexte de l'évolution du réseau urbain sénégalais, ces villes de l'arrière-pays dakarois semblent s'être engagées dans une transition de leur régime de migration. Espace d'émigration importante vers Dakar jusqu'en 1950, elles ont connu depuis lors un développement démographique et économique régulier : Thiès par exemple, avec 175 000 habitants en 1976, est aujourd'hui la deuxième ville du Pays, la population de Mbour a cru entre 1976 et 1988 au rythme de 6,3% par an (ANTOINE et all.,1992). Parallèlement, il est probable que le taux d'émigration vers Dakar a baissé, puis s'est stabilisé. Considéré globalement, cet espace conserve maintenant une part importante de son accroissement de population et certaines de ces villes jouent même sûrement un rôle important dans l'accueil de la migration en provenance de leurs arrière-pays ruraux : Thiès et Mbour par exemple, ont des soldes migratoires positifs entre les recensements de 1976 et 1988.

La composition par sexe et âge de cette population migrante (graphique 8b), est caractéristique d'un regroupement familial généralisé : 58% de femmes (c'est la sous-population de migrants la plus féminine de toutes) et 32% d'enfants de moins de dix ans (contre 26% dans l'ensemble des migrants). Des structures par sexe très spécifiques, chez les migrants célibataires comme dans la migration familiale,

aboutissent à cette prédominance globale des femmes. D'une part, chez les célibataires originaires de ces villes la structure par sexe est parfaitement équilibrée, alors qu'on trouve 57% d'hommes dans l'ensemble des migrants célibataires. D'autre part, pour l'ensemble des autres statuts matrimoniaux, la proportion de femmes atteint 62,4%. Ici la polygamie joue un rôle important puisque 51% des femmes mariées de cette origine ont des époux polygames, alors que cette proportion n'est que de 37% pour les autres migrantes urbaines. Bizarrement, le phénomène n'est pas symétrique pour les hommes, qui sont même un peu moins polygames que la moyenne des migrants (29,4% des hommes mariés contre 30,8%), ce qui suggère une certaine exogamie chez les femmes de cette origine résidant à Dakar. Le niveau d'activité est, par ailleurs, nettement moindre chez ces femmes que pour l'ensemble des migrantes : entre 15 et 60 ans, 32% travaillent ou sont au chômage contre 40% pour l'ensemble des migrantes. Ainsi se dessine, pour une part importante des femmes originaires de l'arrière-pays urbain, un schéma migratoire différent de celui qu'on observe dans les campagnes du bassin arachidier et du pays Sérère. La migration féminine s'inscrit majoritairement ici dans une stratégie de mobilité familiale où l'activité professionnelle des femmes n'a qu'une place secondaire; avec des modalités différentes, les deux phénomènes migratoires concourent en tout cas à la féminisation globale de la migration vers Dakar.

### ***3. Perspectives dans le domaine de l'analyse statistique des itinéraires individuels : l'exemple de l'analyse harmonique qualitative***

Les exemples précédents ont montré comment des méthodologies désormais classiques d'analyse multivariée permettent de conduire certaines études historiques des flux migratoires à partir de données biographiques rétrospectives. Cependant, comme on l'a vu en introduction, un des enjeux de la collecte de données biographiques est la reconstitution et l'analyse longitudinale d'itinéraires individuels. Décrire et comprendre la mobilité des hommes suppose en effet de savoir resituer des événements particuliers dans l'ensemble du cycle de vie de l'individu (temps et espace biographique), comme dans le contexte historique et géographique (temps et espace historique). Identifier les déterminants fondamentaux de cette mobilité que sont les réseaux migratoires, implique également une description typologique et une classification des trajectoires spatiales et sociales des individus. Il est clair que ces objectifs ne peuvent être atteints à l'aide des méthodes simples présentées jusqu'ici. Nous allons donc revenir maintenant sur une voie méthodologique récemment ouverte qui nous semble une des plus pertinentes : l'analyse harmonique qualitative.

#### **3.1 Le principe de la méthode**

L'analyse harmonique qualitative a été présentée une première fois dans un court article théorique par DEVILLE et SAPORTA en 1980, puis à nouveau plus longuement par DEVILLE, dans un article de 1982 où après avoir donné les fondements théoriques de la méthode, il l'applique à des données biographiques concernant la vie matrimoniale d'un échantillon de femmes mariées trois fois ou plus. DEVILLE le dit lui-même, plus que l'intérêt même du sujet, c'est la nature très particulière des données et leur caractère à la fois accessible et pédagogique qui l'a orienté dans ce choix. En effet la nature des données considérées par DEVILLE et SAPORTA est typique des corpus biographiques. Le problème et l'objectif sont clairement annoncés :

*"De nombreuses données d'enquêtes permettent de retracer l'histoire d'individus pendant une certaine période de temps : évolution de l'activité professionnelle, de la situation matrimoniale, de la résidence. Le but de cet article est de montrer que l'on peut analyser de telles données de façon analogue à ce que l'on ferait pour un processus scalaire (analyse harmonique) et pour un ensemble fini de variables qualitatives non temporelles (analyses canonique généralisée ou analyse des correspondances multiples)."*

La réalisation de cet objectif entraîne DEVILLE et SAPORTA dans des développements théoriques assez difficiles. Je ne ferais ici que résumer leur démarche.

Le point de départ est de considérer ce type de données biographiques comme une réalisation d'un processus qualitatif en temps continu  $X_t(i)$  à  $m$  modalités qui sont les  $m$  états dans lesquels peuvent se trouver les individus ( $X_t(i)$  est donc l'état, parmi les  $m$  états possibles, dans lequel se trouve l'individu  $i$  à l'instant  $t$ ). Puisque l'on a affaire à un processus qualitatif que l'on ne sait pas analyser, on va rechercher

un codage scalaire de ce processus :  $Y_t = f_t(X_t)$  où  $f_t$  est une fonction qui transforme un élément de l'espace produit temps x états en un nombre réel (en d'autres termes on associe à toutes les modalités possibles du croisement des états avec les instants d'observation, un codage réel). On peut ensuite faire l'analyse harmonique du processus  $Y_t$ , c'est à dire le décomposer en une suite de variables aléatoires indépendantes du temps et une suite de fonctions du temps (dites harmoniques du processus).

*"L'interprétation pratique peut se faire comme celle des composantes et des harmoniques en analyse harmonique : graphiques des codages de chaque état en fonction du temps, corrélation des génératrices avec certaines variables connues (cercle des corrélations)."*

Mais DEVILLE et SAPORTA font ensuite une remarque de grande importance pratique :

*"Cela dit, l'analyse harmonique qualitative est une généralisation de méthodes familières d'analyse des données et s'y ramène d'ailleurs sur le plan numérique. Toutes les techniques d'interprétations habituelles dans ces méthodes sont donc encore valables dans notre cas."*

Suit dans l'article la démonstration de ces propriétés de l'analyse harmonique qualitative, puis l'exposé d'une méthode d'analyse en temps discret opérationnelle. C'est cette méthode qui nous intéresse pratiquement. Le principe en est simple.

On commence par discrétiser l'intervalle de temps d'observation du processus : les dates  $t_0$  (début de l'observation),  $t_1 \dots t_n$  (fin de l'observation) sont les bornes des intervalles de temps durant lesquels le processus reste stable, c'est à dire durant lesquels aucun individu ne change d'état. La décomposition de l'analyse harmonique en temps continu devient, en temps discret, la décomposition en facteurs obtenue par l'analyse des correspondances multiples d'un tableau disjonctif particulier à  $n$  lignes et  $mp$  colonnes (où  $n$  est le nombre d'individus,  $m$  le nombre d'états possibles et  $p$  le nombre de périodes durant lesquelles le processus est stable). La case courante  $k(i, j, k)$  du tableau vaut 1 si l'individu  $i$  est dans l'état  $j$  durant la période  $k$ , 0 sinon. Mais un problème pratique apparaît immédiatement : dans les cas concrets d'application les nombres  $n$ ,  $m$  et  $p$  génèrent un tableau de dimension pharamineuse. DEVILLE et SAPORTA proposent alors plusieurs solutions d'approximation du problème complet par recodage du tableau de départ. La plus simple et efficace est celle de la division de l'intervalle d'observation en un nombre raisonnable de périodes (de durée constante ou non) sans tenir compte des changements d'états individuels. Le tableau est ensuite construit en calculant le temps passé par chaque individu dans chacun des états possibles au cours de chaque période (on pondérera donc chaque colonne par la durée de la période). Il suffit ensuite de soumettre ce tableau à l'analyse des correspondances multiples et d'utiliser les techniques d'interprétations habituelles.

### 3.2 Les perspectives d'applications

Aux plans théorique et pratique, si l'on envisage son application aux données biographiques complètes qui nous intéressent, l'analyse harmonique qualitative satisfait deux conditions importantes.

Tout d'abord la discrétisation et synchronisation des temps biographiques individuels est modulable et permet diverses stratégies de recodification des données afin de conserver au mieux l'information asynchrone collectée dans l'enquête. On peut en effet adopter plusieurs logiques de découpage de la période d'analyse. Par exemple, celle du temps historique calendaire où les bornes des intervalles de recodification sont choisies en fonction de dates qu'on estime déterminantes dans le contexte conjoncturel, social et politique etc. de la période. Celle d'un temps biographique individuel liée à l'âge de l'individu, ou encore celle d'un temps biographique plus collectif découpé en fonction d'événements clefs du cycle de vie. La question de l'optimalité du découpage correspondant à chacune de ces logiques peut d'ailleurs être posée en terme statistique : optimisation, au sens d'un certain critère numérique, du nombre et des limites des intervalles de codification du processus.

En second lieu, l'usage de l'analyse des correspondances multiples permet tous les croisements possibles entre événements biographiques pour aboutir à la variable d'état qu'on va analyser. En d'autres termes, si l'on collecte de plus en plus dans une même enquête, plusieurs types de biographies (résidentielles, professionnelles, familiales etc.), c'est évidemment que l'on présume des relations étroites entre elles; on attend alors de la méthode d'analyse qu'elle puisse les mettre en évidence et les décrire

précisément. L'analyse harmonique qualitative, appliquée à la variable d'état complexe qui résulte du croisement des différentes situations résidentielles, professionnelles, familiales etc., est potentiellement capable d'atteindre cet objectif. Il ne faut pourtant pas croire qu'on fera un bon usage de la méthode en croisant à tout va toutes les informations biographiques de l'enquête, laissant au calcul algébrique le soin d'exhiber une typologie "complète" qui rende compte de toutes les relations structurant les données. Il y a des limites de plusieurs ordres à ce genre d'approche.

Un préalable "épistémologique" d'abord : même si la vocation des méthodes d'analyse des données est, dit-on, exploratoire, leur emploi ne dispense jamais d'avoir une problématique bien définie lorsqu'on conçoit et analyse une enquête. D'autre part, du point de vue de la fiabilité statistique, la taille de l'échantillon ne permet pas de complexifier, au delà d'un certain nombre de modalités, l'espace des états biographiques (cette remarque s'applique d'ailleurs aussi au découpage de la période de temps de l'analyse). Enfin, s'agissant d'une méthodologie récente, dans un domaine où l'accumulation d'expérience est fondamentale, on manque encore beaucoup de pratique dans l'interprétation des résultats de ce type d'analyse <sup>8</sup>. Et bien sûr, plus l'espace des états est complexe, plus l'interprétation des résultats sera difficile.

Une méthode prometteuse donc, dans le champs de l'analyse typologique des corpus de données biographiques, et particulièrement adaptée à l'étude des formes de la mobilité humaine. Mais au plan de l'interprétation des résultats et pour assurer la pertinence de la démarche, il reste beaucoup à faire. Il faut accumuler les expériences, ne pas céder à la tentation de "l'arme absolue" et savoir, au moins un peu, ce qu'on cherche...

## ***Conclusion***

On ne peut pas reprocher aux chercheurs en sciences sociales et moins encore aux gestionnaires et décideurs politiques, de placer souvent dans les théories et les techniques statistiques des espoirs d'applications concrètes et immédiates : identification et mesure de l'effet des déterminants d'un phénomène social, et plus encore, ajustement de ces phénomènes par des modèles aux vertus descriptives, mais surtout explicatives voire prédictives. Ces résultats sont bien sûr très attractifs lorsqu'on a pour perspective la compréhension d'une réalité sociale, a fortiori quand il s'agit de prévoir et d'agir. Le thème des migrations internationales, s'agissant d'un domaine où les enjeux politiques et économiques sont importants et le débat souvent vicié par des contextes xénophobes, suscitera par excellence ce genre d'attente.

En donnant la priorité à une approche statistique descriptive des données d'enquêtes biographiques, nous ne pensons pas négliger dans la recherche ces objectifs d'application, mais bien plutôt tenter de réunir les conditions préalables pour que les approches probabilistes soient pertinentes. Ce n'est pas qu'il soit urgent d'attendre, mais nous pensons qu'en la matière, il est indispensable de décrire avant de modéliser si l'on veut éviter que les modèles, face à la demande sociale pressante, ne fonctionnent comme des leures ou des outils de désinformation.

---

<sup>8</sup> : Il n'existe à l'heure actuelle aucun logiciel commercial proposant l'analyse harmonique qualitative et l'un des produit de notre recherche est précisément une procédure paramétrée qui permet cette analyse sous le logiciel SAS.

**Notes bibliographiques :**

- AMSELLE J-L. et al. (1976) "Aspects et significations du phénomène migratoire en Afrique" in *Les migrations africaines*, sous la direction de J.L. AMSELLE, ed. MASPERO, Paris, pp 9-39.
- ANTOINE Ph. et al. (1991), "L'insertion urbaine : le cas de Dakar", compte rendu d'une recherche financée par le ministère de la recherche et de la technologie, Dakar, 230 p multigr. + annexes.
- BARBARY O. (1993) avec la collaboration de E.H.A. DIOP, "L'insertion urbaine : le cas de Dakar, rapport d'exploitation d'enquête", Dakar, 213 p multigr.
- BARBARY O. (1994) "Dakar et la Ségambie, évolution d'un espace migratoire transnational", à paraître dans la revue Espace Temps Société, Dakar, 24 p.
- BERET P. [1988] : Analyse de données chronologiques relatives à l'insertion professionnelle, "Les Cahiers de l'Analyse des Données" vol. XIII, n° 2, pp 159-174.
- CEDE y ORSTOM (1992), Memorias del taller metodológico sobre las nuevas formas de movilidad urbanas de las poblaciones de America Latina, 7-11 de diciembre de 1992, CEDE, Bogota xxx p.
- CHESNAIS J.C. (1986), "La transition démographique, étapes, formes, implications économiques", Col. Travaux et Documents n° 113, INED, PUF, Paris 580 p.
- COURGEAU D., LELIEVRE E. [1989] : Analyse démographique des biographies, éditions de l'INED, Paris, 268 p.
- COX D.B. [1972] : Regression models and life tables (with discussion), "Journal of Royal Statistical Society", B 34, pp 187-220.
- DEVILLE J.C., SAPORTA G. [1980] : Analyse harmonique qualitative, "Data Analysis and Informatics", E. DIDAY et al. éditeurs, North Holland Publishing Compagny, pp 375-389.
- DEVILLE J.C. [1982] : Analyse des données chronologiques qualitatives, comment analyser les calendriers ? "Annales de l'INSEE", n° 45, pp 45-104.
- DIOP A.B. (1965) : "Société Toucouleur et migration, enquête sur l'immigration Toucouleur à Dakar", Col. Initiations Africaines, Vol. XVIII, IFAN, Dakar, 224 p.
- DUREAU F. (1989) : Quito, Estadísticas de Población y Vivienda 1987, I.M.Q./ORSTOM Quito, 182 p.
- DUREAU F. [1990] : Recueil et analyse de biographies migratoires et professionnelles à Quito (Equateur), "Travail et Travailleurs", ORSTOM, 10 p.
- DUREAU F. BARBARY O. (1993) : "Citadinos en movimiento : aproximación conceptuel y metodológica a las prácticas residenciales", in Desarrollo y Societdad n° 31, CEDE, Universidad de los Andes, Bogota pp 27-48.

- ESCOUFIER Y., L'HERMIER DES PLANTES [1978] : A propos de la comparaison graphique des matrices de variance, "Biometrical Journal", Vol 20, 5, pp 491-497.
- FARELL G. (1988) "Migración campesina y mercado de trabajo urbano" in Pachano (ed., 1988), pp 287-304.
- FARELL G., PACHANO S., CARRASCO H., (1988) "Caminantes y Retornos", Quito, I.E.E., 167 p.
- FOUCART T. [1983] : Une nouvelle approche de la méthode STATIS, "Revue de Statistique Appliquée", Vol. 31, 2, pp 61-95.
- GREGORY J. (1988), "Migrations et urbanisation", in *Population et Société au sud du Sahara*, sous la direction de D. TABUTIN, L'HARMATTAN, Paris, pp. 369-400.
- HAMER A. (1980), "Les femmes Diolas et la migration : Etude de cas", in *Les migrants de l'arachide*, sous la direction de L. GALLISTEL COLVIN, Dakar, pp. 204-228.
- LEBART L. [1969] : Analyse statistique de la contiguïté, Publications de l'Institut de Statistique de l'Université de Paris, 1969, Vol. XVIII pp 81-112.
- LE FOLL Y.[1982] : Pondération des distances en analyse factorielle, in "Statistique et analyse des données", 1982, vol 1, Paris, pp 13-31.
- MARPSAT M. [1984] : Chômage et profession dans les années trente, in "Economie et Statistique" n° 170, Paris, pp 53-69.
- PACHANO S. (1986) "Se fue a volver" in *Se Fue a Volver*, Seminario sobre las migraciones temporales en America Latina, PISPAL/CUIDAD/CENEP, México, pp 19-40.
- PACHANO S. ed. (1988) *Población, Migración y Empleo en el Ecuador*, Quito, ILDIS, 351 p.
- PICARD J. [1987] : Classification des profils évolutifs incomplets et asynchrones, "Revue de Statistique Appliquée", vol. XXXV (2) pp 27-38.
- SECK A. (1970), "Dakar, métropole ouest-africaine", Col. Mémoires de l'IFAN, IFAN, Dakar, 508 p.
- UBERTALLI B., PERNIN M.O. [1990] : Utilisation d'une méthode multitableaux en sciences sociales. Une étude longitudinale de carrières : les 12 premières promotions de l'école d'infirmières de Roanne. "Population", n°6, pp 1092-1100.