



**HAL**  
open science

## Curso introductivo al análisis de datos biográficos en ciencias sociales

Olivier Barbary, Estadístico Orstom

► **To cite this version:**

Olivier Barbary, Estadístico Orstom. Curso introductivo al análisis de datos biográficos en ciencias sociales. lectureType\_12. Colombia. 1994. hal-02483080

**HAL Id: hal-02483080**

**<https://hal.science/hal-02483080v1>**

Submitted on 18 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  
Estadístico ORSTOM,  
Departamento de matemáticas y estadística  
Universidad Nacional de Colombia - Santafé de Bogotá D.C.  
Colombia

## EL ANALISIS ESTADÍSTICO DE DATOS BIOGRÁFICOS EN CIENCIAS SOCIALES

Curso introductorio al uso de nuevas metodologías de análisis multivariado en el estudio de las trayectorias individuales

### Índice

---

	páginas
<b>INTRODUCCIÓN</b>	7
<b>PRIMERA PARTE : Métodos, ejemplos y perspectivas en el estudio de itinerarios migratorios</b>	9
1. Introducción : los envites de la recolección y del análisis de datos biográficos	9
1.1 Observación biográfica y movilidad humana	9
1.2 El problema estadístico : respetar la continuidad de los tiempos y de los espacios biográficos	10
1.3 Enfoque teórico : describir antes de proponer modelos	11
2. Metodología de análisis estadístico de datos biográficos, revisión bibliográfica y discusión	13
2.1 El Análisis Estadístico de la Contigüidad (o AFL) de Lebart	13
2.2 El Método STATIS de L'Hermier des Plantes y Escoufier	15
2.3 El Análisis Armónico Cualitativo de Deville y Saporta	16
2.4 Especificidad y Complementariedad de los métodos exploratorios para el análisis de datos biográficos	17
3. Ejemplos de aplicaciones y resultados tipos con métodos simples	19
3.1 El análisis longitudinal de un evento particular : la primera instalación de los migrantes en la ciudad	19
3.2 La misma metodología factorial aplicada al análisis de "transiciones migratorias"	23
Conclusion	26

	páginas
<b>SEGUNDA PARTE : Los datos biográficos de encuestas retrospectivas, estructuras, manejo, sesgos y datos censurados</b>	27
Introducción	29
1. La recolección de datos biográficos : cuestionarios, diseños muestrales	30
1.1 Un ejemplo de problemática : analizar la dinámica de Bogotá a través de la movilidad espacial de la población	31
1.2 Información sobre movilidad y cuestionarios biográficos	31
1.3 Muestreo, unidades de recolección, unidades de análisis	33
2. Estructura y manejo de las bases de datos biográficos	35
2.1 Estructura de los archivos	35
2.2 Manejo de las bases de datos	36
3. Sesgos característicos de los datos retrospectivos	37
3.1 identificación de los sesgos : el efecto de la mortalidad y de la emigración	37
3.2 Corrección del sesgo debido a la mortalidad	38
3.3 Consecuencias de los sesgos sobre las estimaciones retrospectivas	40
4. Crítica "externa" de los datos : comparación de las estimaciones retrospectivas con otras fuentes estadísticas	42
5. Datos censurados "a la derecha"	44
5.1 exposición del problema	44
5.2 El caso de los datos de permanencia en la primera vivienda de los migrantes	45
Conclusión	48
<b>TERCERA PARTE : Aplicaciones de los métodos clásicos multivariados a los datos retrospectivos, ejemplos de análisis históricos</b>	49
Introducción	51
1. Aplicación del análisis de correspondencia al estudio de la evolución del origen geográfico de la migración	52
1.1 Datos y métodos	52
1.2 Los primeros planos factoriales proporcionan esquemas generales de evolución	54
1.3 Análisis histórico con niveles geográficos más finos	61

	páginas
2. La misma metodología aplicada al estudio de la inserción residencial de los migrantes	75
2.1 Datos, métodos y sesgos específicos de las submuestras biográficas	76
2.2 El análisis factorial de correspondencia evidencia la transición geográfica de las zonas de recepción de los migrantes	80
Conclusión	92
<b>CUARTA PARTE : El análisis armonico cualitativo, principios y programación, perspectivas de aplicación</b>	<b>101</b>
Introducción	103
1. Principios metodológicos	104
1.1 El enfoque de Deville y Daporta	104
1.2 El algoritmo de recodificación	105
2. puesta en práctica con SAS	106
2.1 Estructura del programa	106
2.2 Desarrollo y parametrización del análisis	120
Conclusión : perspectivas de aplicación	121
Bibliografía	123

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

## EL ANALISIS ESTADÍSTICO DE DATOS BIOGRÁFICOS EN CIENCIAS SOCIALES

Curso introductorio al uso de nuevas metodologías de análisis multivariado en el estudio de las trayectorias individuales

---

### INTRODUCCIÓN

El presente curso se origina en un programa de investigación que llevan a cabo conjuntamente el Departamento de Matemática y Estadística de la Universidad Nacional de Colombia y el Instituto Francés de Investigación Científica para el Desarrollo en Cooperación (ORSTOM). Dicho programa se intitula "metodología y aplicaciones del análisis de datos al estudio estadístico de los itinerarios biográficos", y tiene como principal campo de aplicación el análisis de la formas y evoluciones de la movilidad espacial y social de la población del gran Bogotá. Desde los primeros pasos de su concepción, los investigadores y profesores participantes han querido dar al proyecto tres objetivos :

- desarrollar nuevas metodologías de análisis de datos o adaptarlas a las especificidades de los datos biográficos
- multiplicar las aplicaciones en varios cuerpos de datos, con la máxima variedad de contextos geográficos, metodológicos o temáticos, con fin de acumular experiencia en la interpretación de los resultados y poder concluir sobre la pertinencia y robustesa de cada método
- aprovechar cada ocasión de producir resultados explotables en la problemática social colombiana y en particular buscar relaciones de intercambios con los investigadores de ciencias sociales involucrados en el estudio de la movilidad de la población del área metropolitana de Bogotá <sup>1</sup>.

Este programa lleva actualmente cuatro meses de existencia en Colombia, pero vino a continuación de un trabajo metodológico en los mismos temas, que empecé en Dakar hace dos años y medio. En este curso no podré, por lo tanto, hacer referencia a datos y resultados obtenidos en Colombia. En cambio, trataré de dar cuenta de las dos primeras dimensiones del programa : la investigación metodológica y su aplicación en varios contextos.

Siendo desprendido de un trabajo de investigación en proceso, esta presentación no puede tener pretensión exhaustiva sobre el tema. Se tratará más bien de un conjunto de consideraciones "a priori", a veces teóricas, de hipótesis y experiencias metodológicas y de algunos resultados parciales. Para tratar de dar apariencia de orden a esta colección un poco disparatada, escogimos el plan siguiente.

En la primera sesión, presentaremos primero el contexto de la producción de los datos biográficos en ciencias sociales y la discusión teórica relacionada al análisis estadístico de estos datos. Luego daremos una rápida revisión a la bibliografía del tema y presentaremos los enfoques y resultados típicos de lo que se puede hacer con metodología clásica de análisis multivariado. Concluiremos esta primera sesión con las perspectivas que abren metodologías más sofisticadas.

La segunda sesión estará consagrada a las características de los datos biográficos que proporcionan los cuestionarios retrospectivos; problemas ligados a la técnica de recolección, estructura y manejo de los datos después de la captura, caracterización y manejo de los diferentes tipos de sesgo y datos censurados que afectan la información.

---

<sup>1</sup> : Tenemos en particular muy estrecha colaboración con el programa desarrollado por el CEDE y ORSTOM sobre la movilidad de las poblaciones de Bogotá y su impacto en la dinámica de la ciudad, programa presentado en el presente simposio por F. Dureau.

En la tercera sesión volveremos más en detalles sobre el método clásico de análisis de correspondencia y su aplicación a los datos biográficos, con una serie de ejemplos procedentes de encuestas en Quito y Dakar.

Finalmente, en la cuarta sesión abordaremos un método recientemente desarrollado en Francia y que nos parece muy prometedor : el análisis armónico cualitativo. Una exposición esquemática de sus principios, estadísticos y algorítmicos, nos permitirá luego mirar hacia las perspectivas que podría abrir este método para el análisis tipológico de los itinerarios individuales.

Dada la falta total de experiencia que tengo en docencia, es posible que este programa no se pueda cumplir en su totalidad o, a lo contrario, que nos sobra tiempo. Pero en todo caso me esforzaré de conservar en cada sesión una buena parte del tiempo para aclarar preguntas y permitir el debate.

---

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

## EL ANALISIS ESTADÍSTICO DE DATOS BIOGRÁFICOS EN CIENCIAS SOCIALES

Curso introductorio al uso de nuevas metodologías de análisis multivariado en el estudio de las trayectorias individuales

---

### PRIMERA PARTE : MÉTODOS, EJEMPLOS Y PERSPECTIVAS EN EL ESTUDIO DE INTINERARIOS MIGRATORIOS

#### *1. Introducción : los envites de la recolección y del análisis de datos biográficos*

##### *1.1 Observación biográfica y movilidad humana*

La observación biográfica es cada vez más frecuente en las ciencias sociales. Para el estudio de las formas de movilidad humana, este método corresponde a interrogantes nuevos que surgen en contextos temáticos, disciplinarios y metodológicos muy variados. En efecto, cualquiera que sea el propósito del análisis, la movilidad espacial, social o económica, la movilidad individual o de grupos sociales, cualquiera que sea también el enfoque disciplinario (demográfico, geográfico o socio-antropológico) o metodológico (métodos cuantitativo o cualitativo), el objeto observado es cada vez, un fenómeno continuo. Aquí, el término debe entenderse con el mismo sentido matemático, pues las vidas humanas se componen de infinitudes de instantes, lugares, eventos etc. Es desafortunadamente cierto que ningún sistema de observación puede pretender la restitución de esta continuidad "teórica". No obstante, los conceptos y métodos de la recolección deben alcanzar una descripción la más exhaustiva posible del tiempo y de los espacios en los cuales se desarrolla la movilidad humana.

Hoy en día, esta reflexión ha dejado de ser solamente teórica. Desde la definición de las poblaciones y de los conceptos básicos de observación, hasta la puesta en práctica en encuestas con muestras representativas, los sistemas de observación estadística de la movilidad humana han experimentado recientemente grandes progresos. En el mes de diciembre de 1992 en Santafé de Bogotá, el taller metodológico CEDE/ORSTOM sobre las nuevas formas de movilidad de las poblaciones de América Latina fue la ocasión de hacer el balance de estos alcances y de sus aplicaciones en la investigación y la planificación urbana a través del continente. Entre otras conclusiones, se constató que la aplicación de nuevos conceptos y el desarrollo de nuevas categorías de análisis de la movilidad tienen como punto de partida una recolección de datos que permita la reconstrucción longitudinal de las trayectorias individuales. Existen entonces dos metodologías que aparecen en realidad complementarias : el cuestionario retrospectivo y las encuestas por visitas repetidas.<sup>2</sup>

En un taller sobre la medición de la migración internacional al cual asistí hace poco en México (Tijuana 5-7 de mayo de 1994), el COLEF y ORSTOM nos proponían, como punto de partida a la

---

<sup>2</sup> : Véase las conclusiones presentadas por M. VILLA en "Memorias del taller sobre nuevas formas de movilidad de las poblaciones urbanas en América Latina", Ed. Cie. F. Dureau y C. E. Florez, CEDE/ORSTOM, Santafé de Bogotá, 1992.

reflección, un sistema de observación basado en la articulación de una encuesta sobre migración en la frontera norte de México y un Sistema de Información Geográfica y Estadística de la misma región (SIGEF). Para el propósito nuestro, es la encuesta de observación de los flujos que nos interesa en primer lugar. Se trata de una metodología de encuesta biográfica retrospectiva que J. Santibañez y D. Delaunay describen así :

*" Una encuesta por muestreo probabilístico de poblaciones móviles, que se aplica en más de veinte ciudades fronterizas. El movimiento migratorio es capturado tanto hacia Estados Unidos, como de regreso a México. La biografía migratoria de los entrevistados es reconstituida, así como la ruta de sus desplazamientos y las ocupaciones ejercidas."*

### *1.2 El problema estadístico : respetar la continuidad de los tiempos y de los espacios biográficos*

Así, en la investigación sobre la movilidad humana, el marco problemático, conceptual y operativo de la recolección de datos ha evolucionado y se ha precisado : se sabe hoy en día de manera mucho más precisa porque y cómo definir sistemas de informaciones basados en datos biográficos. Sin embargo, subsisten muchas dificultades en los diferentes pasos de análisis de esta información, particularmente en el enfoque cuantitativo. Como lo dice M. Villa, es necesario "perfeccionar los mecanismos de procesamiento de datos aludiendo a un amplio empleo de procedimientos estadísticos que permitan resaltar las expresiones estocásticas de los eventos relacionados con la movilidad de la población". En otros términos, los métodos de análisis estadístico deben respetar la riqueza y la continuidad de los datos biográficos. Aparece entonces, un importante desfase entre la evolución de los conceptos y técnicas de recolección y el carácter relativamente básico de las herramientas estadísticas generalmente empleadas durante la explotación. Todavía existe frecuentemente una limitación a un enfoque de análisis puramente transversal o, cuando nos es el caso, el manejo de indicadores longitudinales univariados, los cuales son necesariamente reductores de la información original. Esas estrategias tienen generalmente como consecuencias una subexplotación de la información; una verdadera práctica del análisis multivariado de los datos biográficos debe generalizarse.

Debemos reconocer un hecho que explica parcialmente esta situación : debido a la estructura compleja de los mismos datos, los resultados aplicables de la investigación estadística son relativamente nuevos. A causa de su formalización matemática dentro de la teoría de los procesos estocásticos, el campo del análisis estadístico de datos biográficos ha sido durante mucho tiempo bastante teórico t con aplicaciones muy específicas. Ha sido el caso, por ejemplo, de los trabajos de Cox sobre el análisis de los datos de sobrevivencia.

Pero recientemente, nos parece haber cambiado el panorama. Existen desde más o menos veinte años, nuevas propuestas metodológicas en el ámbito del análisis de datos, lo cual constituye ahora un legado para el análisis tipológico de datos biográficos. Es lo que intentaremos mostrar en esta sesión, concentrándonos en la aplicación de tales metodologías al estudio de la movilidad espacial de la población.

Sin entrar en detalles en las problemáticas del estudio de la movilidad humanas por la ciencias sociales, hay que decir brevemente cómo se plantea, para el estadístico, el problema de la explotación de los datos biográficos en este marco. De ello viene que los interrogantes temáticos de las ciencias sociales generan nuevos objetos de investigación para la estadística.

Mientras no se recurra a algún formalismo matemático, el problema es sumamente simple. Para identificar y cualificar los procesos de movilidad de los individuos (movilidad espacial, profesional, social, etc.), es preciso conservar, a lo largo de los diferentes pasos del análisis, el orden de la secuencia de eventos que han conocido y la duración pasada en cada uno de los estados (estados residenciales, profesionales, sociales etc.). La caracterización global de los itinerarios que estamos buscando, no puede desprenderse de varios análisis transversales aplicados separadamente a cada tipo de evento (primer cambio de residencia, primera actividad, residencia o actividad en tal edad o fecha, etc), cada uno con su respectivo resultado tipológico. Al cruzar estas tipologías, además de que surge probablemente un problema de confiabilidad debido al volumen reducido de ciertas clases, uno obtiene grupos de individuos que han conocido el mismo conjunto de eventos, pero posiblemente en orden, con fechas y con duración de transcurso en cada estado totalmente distintas. En tal caso, nada se puede inferir en cuanto a la



homogeneidad de los itinerarios individuales, como tampoco sobre el tipo de dependencia probabilista entre eventos, que evidencian generalmente los datos. Como lo dicen D. Courgeau y E. Lelievre [1989, p 2] : "Se ve desde entonces que la unidad de análisis no puede seguir siendo el evento (fallecimiento, matrimonio, nacimiento, migración, etc.), si no la biografía del individuo, considerada como un proceso complejo".

### 1.3 Enfoque teórico : describir antes de proponer modelos

Planteado así, el análisis de las biografías es, como se ha dicho, un campo de investigación relativamente nuevo. En su extensión actual, se pueden distinguir dos enfoques teóricos : el enfoque descriptivo (o de análisis de datos) y el enfoque inferencial (o probabilista). Según algunos criterios, serían más complementarios que concurrentes. No obstante, siendo dirigidos hacia objetivos radicalmente distintos, estos dos enfoques provocan un debate científico, cuyos argumentos principales trataremos de resumir. Pero antes de ello, presentaremos una rápida síntesis de los principios de estos dos acercamientos estadísticos.

El enfoque de modelización probabilista, desarrollado por los estadísticos y demógrafos a partir de modelos no paramétricos, paramétricos o semiparamétricos (Cox, 1975. Courgeau y Lelievre, 1989. Florez, 1990) tiene como punto de partida una formalización matemática que plantea a las biografías individuales como procesos aleatorios: la biografía de un individuo está representada por una serie de variables aleatorias que son el tiempo de permanencia en los diferentes estados que la componen, tomados en orden cronológico. El trabajo consiste entonces en estimar a partir de los datos de la encuesta, un modelo de distribución, sea paramétrico o no, de cada una de estas variables, para intentar luego la modelización de la distribución más compleja del conjunto de la trayectoria (distribución conjunta). Los métodos estadísticos y los programas informáticos generados por este enfoque, permiten un sofisticado análisis de interacciones entre fenómenos demográficos y apreciar los efectos de las características individuales (sociales, económicas, culturales, etc.) y de los eventos "exteriores", (contexto económico, sociopolítico, catástrofes naturales, etc.) sobre el tiempo de permanencia en diferentes estados.

Una limitación importante de este método tiene que ver con el carácter necesariamente limitado del sistema de eventos o estados que pueda considerarse; además de que el sistema debe definirse *a priori*. Finalmente el enfoque de la modelización no brinda ninguna herramienta descriptiva o tipológica de las trayectorias individuales.

Las técnicas de análisis exploratorio de grandes tablas de datos constituyen, por su parte, herramientas competitivas de descripción y clasificación: análisis factoriales, clasificaciones automáticas. Se ha clasificado este conjunto de métodos, particularmente en el caso de la escuela francesa, bajo el nombre de enfoque descriptivo. Cómo pueden aplicarse en nuestro caso?

Los métodos de análisis de datos han sido aplicados a menudo, especialmente en economía, a series cronológicas sincrónicas, con un fin descriptivo o incluso de previsión. Pero el problema aquí es totalmente diferente: opuestos a las series cronológicas habituales, los datos cronológicos de las biografías individuales son por un lado asincrónicos -no se trata de una serie de "medidas" efectuadas en fechas fijas para todos los individuos sino por el contrario, de un conjunto de secuencias de eventos fechados, cada uno con "reloj" propio-, y por otro lado, el número de eventos y el orden en el que ocurren son diferentes *a priori* de un individuo a otro.

Debido estas características particulares de los datos, sólo recientemente, hace más o menos unos diez años, teóricos y usuarios del análisis exploratorio de datos se han interesado en la elaboración, la evaluación y el uso de instrumentos específicos para analizar el tiempo pasado en situaciones sucesivas. Una búsqueda bibliográfica, aunque incompleta, muestra que existe desde entonces un legado en materia de análisis tipológico de los datos biográficos: los desarrollos y aplicaciones expuestos en varios artículos consagrados al análisis exploratorio de datos cronológicos individuales abren perspectivas interesantes para el

procesamiento de las biografías. Estas herramientas deberían permitir en efecto, la descripción, caracterización y tipología de las trayectorias biográficas. Daremos en adelante algunos principios de algunos de estos métodos.

Siendo poco humilde, se podría decir que la alternativa teórica está entre el estructuralismo y el determinismo, en la medida que, explícitamente o no, se pueden referir a estas dos tendencias filosóficas el enfoque descriptivo (estructuralismo) y el enfoque inferencial (determinismo). En otros términos, los enfoques descriptivos dan la prioridad a la caracterización estadística de la información de la encuesta, con el propósito de permitir su síntesis la más completa y pertinente posible, y de buscar las relaciones que estructuran dicha información. Del otro lado, los enfoques probabilistas proponen modelos con la hipótesis que los datos se desprenden de procesos (o leyes de tipo determinante) que se pueden ajustar o aproximar. Los datos son entonces solamente útiles a la inferencia del modelo.

J.P. Benzecri, en los importantes tratados que consagró al análisis de datos, a ofrecido contribuciones decisivas a este debate. No se puede aquí, dar cuenta del conjunto de los principios que conforman la filosofía de lo que él llama "una práctica moderna del análisis estadístico". Mucho menos todavía presentar su justificación matemática o epistemológica. Daremos un solo ejemplo de sus argumentos, lo cual nos parece aplicarse particularmente al caso del procesamiento de la información biográfica.

*"El modelo debe seguir los datos, no lo contrario. Otra característica deplorable de las matemáticas aplicadas a las ciencias humanas : la abundancia de modelos, forjados a priori y luego confrontados a los datos por medio de lo que llamamos "pruebas". Y unas veces la "prueba" sirve para justificar un modelo donde los parámetros que deben ajustarse son más numerosos que los datos recolectados. Otras veces, al contrario ella sirve para rechazar como inválidas las más pertinentes observaciones del experimentador. Pero lo que nosotros necesitamos, es un método riguroso que nos permita extraer estructuras a partir de los datos."*<sup>3</sup>

En esta misma lógica, M. Volle piensa que las técnicas descriptivas y tipológicas del análisis de datos constituyen una condición previa a todo enfoque probabilista :

*"Uno nota que la referencia al calculo de probabilidades es practicamente ausente del análisis de datos : el análisis de datos se ubica en una etapa del razonamiento logicamente anterior a la inducción probabilista (...). Para el que practica el análisis de datos, es solo despues de haber identificado las estructuras que subyacen en los datos, despues de haberlos clasificado y recortado, que es posible abordarlos con un enfoque probabilista."*<sup>4</sup>

Finalmente son estos principios los que guían el programa de investigación que hemos emprendido. En efecto, además de ofrecer un campo de investigación mucho menos explorado que al del enfoque probabilista, el enfoque del análisis de datos nos parece adaptarse mejor a un paso de la investigación sobre la movilidad de las poblaciones cuyo objetivo prioritario es la descripción de los fenómenos y la identificación de sus posibles determinantes. Cuando se trata de un contexto donde las sucesiones de los eventos observados son muy complejas y el universo de variables potencialmente explicativas muy extenso, es tentador volver al argumento de J.P Benzecri : no estamos en posición de formular hipótesis suficientemente precisas par validarlas (o invalidarlas) por medio de modelos o pruebas estadísticas. En otros términos, como podemos estar seguros, en el marco de las ciencias sociales, con "diseños experimentales" muy poco controlados (porque son en realidad incontrolables!), de que las hipótesis probabilísticas en las cuales basamos el modelo no sean parcialmente o totalmente en contradicción con los mismos datos. Cabe recordar por ejemplo que son muy escasos los datos procedentes de encuestas de ciencias sociales, que se ajustan a una hipótesis de distribución normal. De acuerdo con todos estos argumentos, nos pareció preferible dar la prioridad a un enfoque exploratorio, utilizando instrumentos estadísticos tipológicos.

## **2. Metodología de análisis estadístico de datos biográficos, revisión bibliográfica y discusión**

<sup>3</sup> : J.P. Benzecri & colaboradores, L'analyse des données -tome 2 : l'analyse des correspondance, p 6, Dunod, Paris, 1982.

<sup>4</sup> : M. Volle, Analyse des données, p 21, Economica, Paris, 1981.

Haremos solamente ahora una revisión muy rápida y sin pretención exhaustiva de algunos legados recientes que ha proporcionado el campo de investigación que podríamos llamar "análisis exploratorio de datos cronológicos individuales". Se trata pues de teorías y técnicas de análisis factorial o de clasificación automática (cluster análisis), y cada una de ellas nos parece poder contribuir de alguna manera al análisis tipológico de las biografías. Luego aprovecharemos de esta presentación sintética para evaluar, a priori y de manera teórica, la especificidad y/o complementariedad de estos diferentes métodos.

En su artículo consagrado a un método de clasificación de perfiles individuales evolutivos, J. Picard (1987) introduce su exposición con las siguientes referencias:

*"El análisis de la evolución multivariada de individuos conlleva el estudio de las estructuras de un cuadro tridimensional, en el que los tres índices corresponden respectivamente a los individuos, las variables y el tiempo. Los métodos utilizados para este estudio se basan esencialmente en dos tipos de enfoques factoriales:*

*El primer enfoque ha sido propuesto por Lebart (1966-1969) bajo el nombre de análisis factorial local (AFL), que aplicó en contiguidades espaciales.*

*El segundo enfoque es el del método STATIS desarrollado por L'Hermier des Plantes y Escoufier (1978)."*

De estas dos familias de métodos de análisis factorial, presentaremos aquí las ideas principales del artículo de Lebart (1969) y una aplicación de los métodos STATIS.

Es necesario igualmente, agregar otras dos aproximaciones factoriales de estos problemas al inventario de Picard :

La primera es el análisis armónico cualitativo, una aplicación particular del análisis de las correspondencias múltiples desarrollado y justificado por Deville y Saporta, en dos artículos teóricos (1980-1982). Este método nos parece de los más prometedores en el campo del análisis tipológico de los itinerarios biográficos. Por eso volveremos en ello más prolongamente en la cuarta sesión, hoy día solamente resumiremos su principio.

La segunda, por ser la más simple de todas, no puede por ello omitirse. Consiste en aplicar técnicas clásicas de análisis multivariado a ciertos cuadros extractos de la información biográfica "completa". Se puede así, mediante métodos específicos (ACP, AFC o AFCM, CAH etc.), abordar el análisis longitudinal de ciertos eventos biográficos particulares como son la etapa de llegada de los migrantes a sus lugares de destino, la entrada en la vida activa, la primera unión, etc. Como ilustración de estos enfoques metodológicos simples, daremos más luego algunos ejemplos. Para volver al objetivo más general de una tipología de las biografías completas, el análisis de las correspondencias múltiples clásico también se puede aplicar a un cuadro en el que se han resumido las trayectorias individuales a estudiar, en un determinado número de variables cuantitativas o cualitativas que las describirán en forma resumida. Todo depende entonces del escogimiento de estos resúmenes. Dada su carácter clásico, no hace falta presentar esta metodología aquí. Un excelente ejemplo de resultados interesantes obtenidos mediante estos métodos se encuentra en Marpsat (1984).

## 2.1 El Análisis Estadístico de la Contiguidad (o AFL) de Lebart <sup>5</sup>

*"Los estadísticos tienen que ver a menudo con conjuntos de medidas u observaciones que no pueden ser consideradas como realizaciones independientes de variables o de vectores aleatorios. En economía, en Ciencias Humanas o Biológicas, las repeticiones de pruebas idénticas son muy escasas.*

*Sin embargo, el campo de observaciones sugiere a menudo la forma de los enlaces entre ellas. Es el caso de las series cronológicas y los modelos con errores, ligadas en el tiempo, de los econométricos. Estudiaremos aquí el caso más genérico, en el que las observaciones se realizan sobre un grafo (por ejemplo el caso en que se privilegia un conjunto de parejas de observaciones: parejas de*

---

<sup>5</sup> : Vease Lebart (1969).

*observación sucesivas para las series temporales, parejas de observaciones contiguas para las variables regionales, etc.)."*

Es así como Lebart plantea el problema al que aporta elementos de solución en su artículo publicado en 1969.

Los datos biográficos migratorios son un ejemplo de datos relacionados con grafos de contiguidad (grafos de contiguidad espacial o temporal). Esta primera clase de métodos se desprende de la idea que en muchos casos, la evolución temporal y/o espacial de los datos no es independiente de la estructura de contiguidad entre fechas o lugares, representada por dichos grafos. Como se trata precisamente de la hipótesis que se puede formular a propósito de los datos biográficos individuales, los desarrollos de Lebart pueden abastecernos certeramente de pistas metodológicas interesantes.

En este artículo el autor tiene dos objetivos principales:

- Probar la validez (con ayuda de un test estadístico) de la hipótesis de dependencia de los datos frente al grafo y describirla.
- Fundar sobre esta idea una nueva técnica de análisis factorial que permita evidenciar la escala y la localización de los enlaces espaciales o temporales entre los datos.

Al iniciar la empresa, Lebart retoma la noción de coeficiente de contiguidad definido por Geary en 1933. Este coeficiente puede comprenderse como la razón obtenida con dos estimadores de la varianza de la variable: en el numerador, el estimador toma en cuenta la contiguidad o no de las observaciones sobre el grafo, mientras que en el denominador encontramos la varianza empírica clásica. Si las observaciones varían independientemente de la estructura del grafo, los estimadores poseen la misma esperanza, y el coeficiente de Geary se acerca a uno. El cálculo de los momentos de la distribución de los coeficientes de contiguidad permite construir un test de la hipótesis de dependencia con relación a el grafo.

Habiendo alcanzado su primer objetivo Lebart generaliza el coeficiente de Geary a la noción de coeficiente de contiguidad en el nivel  $\beta$ , tomando las parejas de vértices distantes de  $\beta$  sobre el grafo. Esta generalización le permite probar, no solamente la dependencia con respecto a la estricta vecindad en el grafo, sino además las dependencias con respecto a vecindades de nivel  $\beta$ . Así se logra identificar la escala espacial o temporal en la que se ejerce la autocorrelación de los datos.

Lebart pasa luego a la aplicación de estas nociones en el análisis factorial de un conjunto de observaciones contiguas. Considera algunas variables concurrentes en un mismo grafo (grafo de regiones o de fechas de observaciones cronológicas por ejemplo) e introduce la matriz de contiguidad de las observaciones: matriz de covarianza de variables, calculada sobre las observaciones que poseen determinada distancia sobre el grafo. La idea es que si hay independencia de las evoluciones espaciales o temporales en relación con la estructura de el grafo, el análisis en Componentes Principales (ACP) de la matriz de contiguidad, no variará significativamente frente a la ACP de la matriz de covarianza clásica de las variables. Inversamente, si la hipótesis de independencia es falsa, podremos interpretar las diferencias entre los planos factoriales de la ACP de la matriz de contiguidad y aquellos de la matriz de covarianza. El ejemplo de aplicación dado es el de una dependencia espacial de datos socioeconómicos en el conjunto de los departamentos franceses para los que se prueban diferentes niveles de contiguidad.

La aplicación del análisis de contiguidad a los datos biográficos, debería permitir identificar las escalas temporales o espaciales en las que se ejerce la autocorrelación de los eventos migratorios y profesionales, en los grandes tipos estadísticos de itinerarios biográficos. En cambio, el método no proporciona directamente instrumentos tipológicos pero veremos en lo sucesivo que todos los pasos de análisis tipológico completos, requieren una reducción más o menos considerable de la información recolectada inicialmente. Interesaría entonces, utilizar los resultados del análisis de las contiguidades. Los concernientes a la autocorrelación temporal guiarán el escogimiento del período de referencia considerado para clasificar los itinerarios que precedan o sigan un evento clave: la llegada de los migrantes a Quito, por ejemplo. La calificación de la autocorrelación que opera en los espacios geográficos o económicos, determinará por su lado, las nomenclaturas de la recodificación eventual de las variables que describen los estados residenciales y profesionales.

Pero la aplicación de este método puede resultar compleja: cómo definir los criterios de contigüidad en los espacios temporal, geográfico y económico? Qué niveles de análisis de contigüidad escoger *a priori* para construir las matrices objeto de análisis? (Ya aparece aquí el problema de la discretización del tiempo continuo de la observación). El método basado en la ACP, requerirá en todo caso la construcción de variables métricas que resuman las trayectorias residenciales y profesionales durante los períodos elementales.

## 2.2 El Método STATIS de L'Hermier des Plantes y Escoufier <sup>6</sup>

El principal instrumento estadístico que utilizan los métodos multicuadros STATIS es el ACP de los diferentes cuadros de distancias, construidos a partir de las dimensiones de las tablas de datos originales: distancias interindividuos, intervariables, interfechas. Cuando se quieren analizar calendarios de eventos, el uso de estos métodos implica transformaciones de los datos originales. Daremos una ojeada esquemática a esta aproximación, tomando como ejemplo la aplicación presentada en el artículo de Ubertalli y Pernin (1990).

Ubertalli y Pernin, observan las trayectorias profesionales de las primeras promociones de la escuela de enfermería de Roane. Tratándose de ACP de cuadros de distancias, es necesario en primer lugar lograr que las tres dimensiones del problema (individuos, variables, fechas), puedan ser representadas en espacios euclidianos. Para situarse en este marco, los autores deben discretizar y sincronizar el tiempo de observación de los individuos: a partir de la fecha de obtención del diploma, se escogen "instantes clave" de observación de individuos (diploma más un año, diploma más tres años, etc.). Luego se resume la carrera individual para cada una de estas fechas por medio de un número determinado de variables métricas: tiempo de ejercicio como enfermero, tiempo de estudios complementarios, número de períodos de inactividad, etc.

Una vez realizada esta "mediatización" de las informaciones recolectadas sobre las carreras individuales, la aplicación de los métodos STATIS se desarrolla en cuatro etapas principales:

### a. Análisis de la "interestructura"

Esta etapa tiene como objetivo la descripción de las relaciones globales entre las diferentes fechas. Para cada fecha se calcula la matriz de distancias euclidianas entre individuos, en el espacio de las variables. Esta matriz es la de un operador lineal que traduce la "estructura individuos" en cada fecha (de ahí el nombre de "interestructura". El producto escalar de Hilbert-Smith entre operadores lineales, permite construir una matriz de distancias interfechas que es sometida al ACP. La interpretación de los planos factoriales proporciona una tipología de las estructuras globales individuos por variables en las diferentes fechas; ésta proporciona una idea de la evolución temporal promedio de los individuos.

### b. Análisis de las "distancias-compromiso".

A partir de las matrices de las distancias interindividuales en cada fecha, se construye ahora una matriz de "distancias-compromiso", es decir la suma de las distancias en cada fecha ponderada por las componentes del primer vector propio a la "interestructura" (es en síntesis, una distancia interindividual promedio a lo largo del conjunto de fechas). La ACP de este cuadro de distancias describe entonces la estructura individuos, común al conjunto de fechas, una especie de resumen temporal de las trayectorias individuales.

### c. Análisis de la "intraestructura" de las variables.

Consideremos ahora el gran cuadro rectangular obtenido "fusionando" los diferentes cuadros individuos por variables en cada fecha, nuevamente ponderados por las raíces cuadradas de las componentes del primer vector propio de la "interestructura", para permitir la superposición de planos factoriales con aquellos obtenidos en la ACP de las "distancias-compromisos".

El análisis factorial de este cuadro y la proyección de la nube de "variables x fechas", permite entonces observar las correlaciones entre estas "variables x fechas" y las componentes principales de "compromiso". De esta manera se interpretan los ejes del "compromiso" a partir de las variables

<sup>6</sup>: Véase Escoufier y L'Hermier des Plantes (1978) y Ubertalli y Pernin (1990).

descriptivas de las trayectorias individuales y, se podrá juzgar la importancia relativa de cada variable en la discriminación de las diferentes formas de trayectorias.

d. Análisis de la "intraestructura" de los individuos.

El análisis de la "intraestructura" de los individuos se hará por el medio de la adición, como elementos suplementarios, de los cuadros individuos x variables en cada fecha en la ACP del cuadro de "distancias-compromisos". Sobre los planos factoriales del "compromiso", existe un punto por individuo y por fecha que se puede relacionar en orden cronológico para obtener trayectorias individuales interpretables en función de su posición con respecto a los ejes del "compromiso".

Aplicada en su conjunto a los datos biográficos de la encuesta Migraciones de Quito, el método STATIS debe permitir un análisis tipológico completo de los itinerarios residenciales y profesionales :

- Evidenciar a partir de la "interestructura," las fechas claves en la diferenciación de los itinerarios;
- Explicar con la ayuda de la "intraestructura" de las variables, el papel de los eventos residenciales y profesionales en esta diferenciación;
- Clasificar las trayectorias individuales, ("intraestructura" de los individuos) en el espacio factorial del "compromiso", interpretar la tipología e ilustrarla por medio de trayectorias-tipos.

Su ejecución propiciará las mismas preguntas generales planteadas por el análisis de las contigüidades: discretización y sincronización de los tiempos de observación individuales, escogimiento de los resúmenes de la información residencial y profesional por medio de variables métricas. Sin embargo, tal como lo veremos luego, el riesgo de reducción en la precisión de la información sobre la cronología de los eventos individuales, aumenta significativamente con este método.

### 2.3 El Análisis Armónico Cualitativo de Deville y Saporta <sup>7</sup>

El análisis armónico cualitativo fue presentado por primera vez en un corto artículo teórico, por Deville y Saporta en 1980, luego más extensamente por Deville, en un artículo de 1982.

La estructura de los datos considerados por Deville y Saporta y los objetivos que se proponen de alcanzar son realmente típicos de los cuerpos de datos biográficos. Podemos juzgarlo de acuerdo a la presentación que hacen los autores, del problema planteado :

*"Diversos datos de encuestas permiten trazar la historia de los individuos durante un periodo de tiempo determinado: la evolución de la actividad profesional, de la situación matrimonial, de la residencia. El objetivo de este artículo es mostrar que datos de este tipo pueden analizarse de manera análoga a lo que haríamos con un proceso escalar (análisis armónico) y para un conjunto finito de variables cualitativas no temporales (análisis canónico generalizado o análisis de correspondencias múltiples)".*

El alcance de este objetivo arrastra a Deville y Saporta hacia desarrollos teóricos complejos que aquí tan sólo resumiremos.

El punto de partida es considerar este tipo de datos biográficos como la realización de un proceso cualitativo en un tiempo continuo  $X_t(i)$  de  $m$  modalidades ( $X_t(i)$  es el estado, entre los  $m$  estados posibles, en el que se encuentra el individuo  $i$  en un instante  $t$ ). Puesto que se trata de un proceso cualitativo que no sabemos analizar, buscaremos un código escalar del mismo:  $Y_t = f_t(X_t)$ , (en otros términos, asociamos a todas las modalidades posibles de cruces de los estados con los instantes de observación, un código real). Podemos luego hacer el análisis armónico del proceso  $Y_t$ , es decir descomponerlo en una serie de variables aleatorias independientes del tiempo y una serie de funciones del tiempo (llamadas armónicas de procesos). Pero no es la interpretación de este análisis armónico la que nos interesa, porque Deville y Saporta hacen luego una observación de gran importancia práctica :

*"El análisis armónico cualitativo es una generalización de métodos familiares de análisis de datos y es reductible a estos métodos, al plano numérico. Todas las técnicas de interpretaciones habituales en estos métodos tienen validez en nuestro caso".*

Sigue en el artículo la demostración de estas propiedades de análisis armónico cualitativo, luego la exposición de un método de análisis en tiempo discreto operacional que es una aplicación particular del análisis de correspondencias. El principio es sencillo.

Se comienza por discretizar el período de observación del proceso, considerando todos los intervalos de tiempo en los que el proceso se mantiene estable, sin que individuo alguno cambie de estado. Es necesario apuntar aquí, que a pesar de ello, no se pierde información sobre la cronología de los eventos recolectados. El análisis armónico en tiempo continuo se convierte, en tiempo discreto, en la descomposición en factores resultante del análisis de correspondencias múltiples de un cuadro disyuntivo particular con  $n$  líneas y  $mp$  columnas (en el que  $n$  es el número de individuos,  $m$  es el número de estados posibles y  $p$  el número de períodos en los que el proceso se mantiene estable). La casilla elemental del cuadro vale uno si el individuo está en el estado considerado durante el período y cero si no lo está. Pero inmediatamente surge un problema práctico: en los casos concretos de aplicación, los números  $n$ ,  $m$  y  $p$  generan un cuadro de tamaño asombroso. Deville y Saporta proponen entonces una solución de orden aproximativo al problema global, recodificando el cuadro original. Se dividirá el intervalo de observación en un número razonable de períodos (de duración constante o no) sin tener en cuenta los cambios de estado individuales. Se construye luego dicho cuadro calculando el tiempo de permanencia de cada individuo en cada uno de los estados posibles a lo largo de cada período; en seguida se aplica el análisis de correspondencias y las técnicas de interpretación habituales.

### 2.4 Especificidad y Complementariedad de los métodos exploratorios para el análisis de datos biográficos

<sup>7</sup>: Vease Deville y Saporta (1980) y Deville (1982).

Evidentemente no podemos prejuzgar ahora los resultados, que en el plano temático, arroje el procesamiento estadístico de las biografías recolectadas. Desde el punto de vista de la metodología estadística, en cambio, quisiéramos intentar una síntesis inicial, examinando globalmente la adecuación de los métodos de análisis exploratorio que hemos revisado frente al problema que encabeza el procesamiento de los datos biográficos de las encuestas retrospectivas.

Como base de una respuesta a este interrogante encontramos el problema de la descripción estadística y del análisis tipológico de los itinerarios residenciales y profesionales recogidos. Ya sea con el objeto de identificar las prácticas y las lógicas de ocupación del espacio geográfico y económico por parte de los ciudadanos, ya sea para medir el impacto que estas prácticas tienen sobre la dinámica de la ciudad, es necesario antes que nada, describir y clasificar con el fin de medir la importancia relativa de cada tipo, las diferentes trayectorias individuales. Este trabajo absolutamente necesario debe tomar en cuenta tanto las variables constitutivas de los itinerarios individuales -el tiempo, los espacios geográficos y económicos en los que se desarrollan-, como las variables que describen el contexto histórico, geográfico y económico de estos eventos individuales.

Por una parte tal problemática general justifica, como lo hemos visto, escogimientos metodológicos generales para la recolección y para el análisis:

- El enfoque estadístico a partir de una muestra representativa del conjunto de residentes urbanos que permite la conformación de cohortes de efectivos significativos, homogéneas desde el punto de vista de los contextos geográfico y cultural de ciertos eventos claves como el comienzo de la migración o la llegada en la ciudad;

- El empleo de técnicas de análisis exploratorio de los cuadros de datos de tamaño importante, capaces tan sólo de dar respuesta al primer objetivo de descripción y análisis tipológico de las trayectorias.

Por otra parte, la rápida revista metodológica que nos precede, ha mostrado que la adecuación de métodos de análisis tipológico al cuestionamiento temático, debe medirse con uno de los dos criterios principales relacionados con los problemas de recodificación de los datos:

- El grado en que se han tenido en cuenta, en el análisis, la duración y del orden de los eventos constitutivos de las biografías individuales (la recodificación discreta y sincrónica de los tiempos de observación individuales conserva acaso la información continua y asincrónica de la variable original ?);

- La conservación más o menos adecuada, mediante la recodificación, de la riqueza de la información acerca de las trayectorias individuales en los espacios geográficos y económicos.

Arriesgándonos en un primer intento de síntesis, pero desde un punto de vista estrictamente teórico y sin ninguna pretensión definitiva, por la falta de trabajo práctico con los datos, podemos proponer una "clasificación" inicial de los cuatro métodos arriba considerados, según estos dos criterios, desvelando así la especificidad de cada una y sus complementareidades, que pensamos aprovechar en el análisis tipológico de los datos biográficos recolectados en Quito, en el marco del interrogante delimitado.

El análisis de las correspondencias múltiples sobre el cuadro que resume las trayectorias individuales permite desembocar en una tipología de las trayectorias, probablemente muy reduccionista desde el punto de vista de los dos criterios que nos interesan: la discretización/sincronización de los tiempos es total; la selección y codificación de las variables para describir sintéticamente el itinerario residencial y profesional determina en gran medida la pertinencia de los resultados, de donde nace la necesidad de conservar en tal estado las nomenclaturas más refinadas. Su ejecución es simple y rápida; bien aplicada, puede prestar apreciables servicios en una primera fase de descripción general de los datos biográficos.

El análisis de la contigüidad, aunque no sea propiamente un método completo de análisis tipológico, puede proveer herramientas valiosas para la preparación de datos para el análisis. Permite la identificación



de escalas y la localización de las autocorrelaciones temporales, espaciales o económicas que se ejercen en el seno de los itinerarios biográficos. La discretización/sincronización de los tiempos practicada en este análisis es maleable, pero, basado sobre el análisis en componentes principales, implica la constitución, para cada intervalo de tiempo, de variables métricas que resuman los eventos residenciales y profesionales acaecidos.

El método STATIS, constituye una muy completa "caja de herramientas" para el análisis tipológico aplicable a datos biográficos individuales. El tiempo se encuentra discretizado y sincronizado en el seno de un número necesariamente reducido de intervalos, ya que cada fecha introduce un nuevo cuadro de análisis; reducir la información se convierte en un imperativo. Además, también basada en el ACP, presentará la misma debilidad que el método precedente, desde el punto de vista del segundo criterio. Sin embargo, en un marco de discusión bien definido, la riqueza de la interpretación puede convertirle en un precioso método.

Finalmente, el análisis armónico cualitativo, evaluado en el terreno teórico de acuerdo a los dos criterios que nos interesan, parece el más prometedor: conserva óptimamente el orden y la cronología de los eventos residenciales y profesionales, y el empleo del análisis de correspondencias múltiples permite la mezcla de variables cualitativas y cuantitativas en la descripción de los estados. De la mano con las técnicas clasificatorias automáticas, desemboca en tipologías de los itinerarios individuales susceptibles de interpretación con el auxilio de reglas sencillas.

Subrayemos también la óptima complementareidad potencial de los cuatro métodos. El primer método, utilizado en la etapa de descripción general de los itinerarios biográficos, permitirá la definición de tipos globales en los que luego operará el análisis tipológico detallado. El análisis de contiguidades, aportando conclusiones sobre las escalas en que interviene la autocorrelación temporal, geográfica y económica de los datos biográficos, contribuirá a definir intervalos de tiempo y nomenclaturas para recodificar variables de estadios residenciales y profesionales. Podremos luego alternar o llevar conjuntamente los análisis tipológicos refinados con la ayuda de métodos STATIS y con el análisis armónico cualitativo, mediante la selección que mejor se adapte al problema, o sacando provecho, en un campo común de aplicación, de las dos iluminaciones complementarias que permiten técnicas de cálculo y de interpretación diferentes.

Pero antes de abordar, con una o otra de las metodologías precedentes, el análisis de biografías individuales completas, nos parece importante mostrar cuáles son los alcances y limitaciones de los métodos más simples y tradicionales de análisis multivariados.

### ***3. Ejemplos de aplicaciones y resultados tipos con métodos simples***

Desde su inicio en 1991, nuestra investigación se apoyó, para las aplicaciones, sobre dos cuerpos de datos que proceden de encuestas demográficas incluyendo varios módulos biográficos :

- la encuesta "migraciones", realizada en 1987 en Quito por ORSTOM, dirigida por F. Dureau.
- la encuesta "inserción urbana" realizada en 1989 en Dakar por IFAN y ORSTOM, dirigida por Ph. Antoine.

Seleccionemos aquí algunos ejemplos típicos procedentes de la explotación de estas dos encuestas.

#### ***3.1 El análisis longitudinal de un evento particular : la primera instalación de los migrantes en la ciudad***

En una primera etapa de la explotación de las encuestas de Quito y Dakar, hemos querido estudiar particularmente un momento decisivo de la vida de los migrantes, el de la primera llegada a la ciudad. Este acontecimiento constituye obviamente una etapa clave para las problemáticas de estudios urbanos : determina al mismo tiempo el punto de partida de las trayectorias urbanas de los migrantes y las zonas de impacto de la migración en la ciudad. Desde cualquiera de estas dos perspectivas interpretativas, una dimensión histórica en el análisis es imprescindible. Tenemos entonces un objetivo muy específico para este

primer acercamiento estadístico a los datos biográficos : describir la evolución geográfica y temporal de las zonas de acogimiento de la migración.

Las estructuras de los datos y métodos de este análisis son bastante simples <sup>8</sup>. Construimos primero el cuadro de contingencia que cruce las localizaciones de las primeras residencias de los migrantes en la ciudad con sus fechas de llegada. Esta información, proporcionada por la parte biográfica de los cuestionarios, proviene de la recolección de los barrios y fechas de residencias de cada etapa residencial en la ciudad de los individuos de la muestra. En el caso de Quito por ejemplo, la población analizada incluye 1587 individuos migrantes, jefes de hogares en el momento de la encuesta. Conocemos, para cada uno de ellos, la fecha de primera llegada a Quito y el barrio de esta primera residencia en la capital. Esta muestra es representativa de la población de los jefes de hogares residentes en Quito en 1987 (llegados desde 1935). Para formar, a partir de esta información, un cuadro de correspondencias con efectivos suficientes en cada modalidad, clasifiquemos los barrios de residencia en 20 zonas y las fechas de llegada en 11 periodo quinquenales. Fue este cuadro 20 x 11 el sometido al análisis factorial

Los pasos de interpretación de los resultados son los clásicos del análisis factorial, basados primero en el examen de los planos factoriales y de los indicadores estadísticos, luego en diferentes tipos de ilustraciones gráficas. Los resultados obtenidos en Dakar son talvez los más pedagógicos.

El primer plano factorial evidencia muy claramente que la geografía de las zonas de llegada de los migrantes en la urbe, se ha modificado radicalmente desde 1945 (gráfico 1). La nube de puntos (fechas y barrios) presenta un aspecto parabólico muy frecuente en AFC (efecto Guttman), lo cual señala una forma cuasi diagonal del cuadro de datos. Dicho de otra manera, los datos demuestran una transferencia regular de las zonas de instalación de los migrantes en el transcurso del tiempo. Así hay barrios de recepción de la migración que se pueden calificar de "antiguos", otros de "modernos", otros finalmente, cuya función de recepción ha pasado por un máximo durante un momento dado del periodo de observación. La estructura particularmente clara de los datos nos permite deducir de estos dos primeros factores, una tipología muy satisfactoria de los 14 barrios, según la evolución de sus papeles en la recepción de la migración. Esta tipología se ilustra con el gráfico 2 que sintetiza la evolución espacial y temporal de la función de recepción de la migración en Dakar.

En fin, podemos dar de esta evolución una imagen cartográfica (gráfico 3); en realidad de manera independiente del análisis factorial, porque son los mismos datos de la tabla de contingencia los que representamos (un mapa por fecha). No obstante, el análisis tipológico puede brindar una valiosa ayuda para la elección de las unidades espaciales y de las fechas que resaltarán mejor el fenómeno. Por otra parte, la caracterización estadística de la transferencia espaciotemporal ilustrada por el mapa, permite también una mejor estructuración del comentario. Y finalmente el análisis factorial de los datos se vuelve indispensable cuando una o otra de las dimensiones del cuadro (número de fechas o de unidades espaciales) aumenta significativamente.

---

<sup>8</sup> : En cambio, es necesario tomar muchas precauciones en la aplicación y interpretación de los resultados, debidas por una parte a los diferentes tipos de sesgos y datos censurados que afectan a la información retrospectiva, por otra parte a las características eventualmente particulares de la submuestra biográfica. Para más detalles, vease por ejemplo : Barbary (1993) pp 2-9 y pp 31-35.

**Gráfico 1 :** *evolución de la zonas de recepción de los migrantes en Dakar desde 1945, primer plano factorial del AFC (1x2)*

**Gráfico 2 :** *evolución de la zonas de recepción de los migrantes en Dakar desde 1945, tipología en cinco grupos de barrios*

De 1950 a 1969 , siguiendo con cierto desfase temporal la extensión espacial de la ciudad, la recepción de la migración se ha transferido desde los barrios céntricos (Plateau y Médina) hacia la primera cintura de los suburbios más antiguos (Grand Dakar, HLM y Pikine). A partir del final de los años sesenta, el desplazamiento sigue hacia las extensiones, legales o ilegales, más excentradas de la ciudad : Patte d'Oie,

Cambérène, Médina Gounass, Thiaroye, Guédiawaye, Yeumbeul y Malika. Más que la transferencia de la función de recepción de los migrants por ella misma, es la evolución de su ritmo, apreciado en relación al propio crecimiento de la ciudad, que constituye el fenómeno importante. Mientras antes de 1960, un desfase relativamente importante existe entre la emergencia y consolidación de nuevas zonas de residencia, y la llegada de una población inmigrada, el periodo moderno se caracteriza por un desarrollo mucho más rápido de la función de recepción de los migrantes en los barrios nuevos.

**Gráfico 3** : *evolución de la zonas de recepción de los migrantes en Dakar desde 1945, mapas de repartición en porcentaje del flujo total*

### 3.2 La misma metodología factorial aplicada al análisis de "transiciones migratorias"<sup>9</sup>

El método de AFC precedente puede aplicarse de la misma manera al cuadro cruzado obtenido con los lugares de origen de los migrantes y sus fechas de llegada. Los resultados evidencian en primer lugar transiciones migratorias a la escala nacional e internacional (escala "macro").

**Gráfico 4 :** *evolución de la regiones de origen de los migrantes hacia Quito desde 1935, tipología en tres espacios geográficos*

Así en Quito (gráfico 4), la composición del origen geográfico de la migración ha cambiado bastante brutalmente entre 1960 y 1980. Durante este periodo, el espacio de origen de la mayoría de los migrantes hacia la ciudad se trasladó desde el hinterland rural (así la sierra rural centro y norecuatoriana proporcionaba 55% de la migración en 1960 y menos de 30% a partir de 1980), hacia un espacio más lejano, integrando a la costa, los campos de la sierra sur y el extranjero (20% de la migración en 1960, más de 40% a partir de 1980). En cuanto al espacio urbano de la sierra, su contribución al flujo migratorio global se mantuvo casi estable durante todo el periodo (25% a 30% de los migrantes).

En Dakar (gráfico 5), la evolución fue menos rápida y se extiende a todo el periodo de observación. Antes de 1960, la migración procedente del hinterland rural y del norte del país domina ampliamente (más o menos 80 % del total). La parte norte de Senegal juega en este momento un papel muy importante : 50% del flujo entre 1935 y 1939 y entre 1945 y 1949, todavía 42% entre 1955 y 1959. Antes de 1950, el tercer espacio de la tipología (parte sur de Senegal y extranjero) no tiene sino un papel menor con menos de 20% del flujo migratorio total. Es precisamente el crecimiento casi continuo desde 1945 de la migración originaria del sur del país y del extranjero, lo que constituye el fenómeno sobresaliente de la historia moderna de la migración hacia Dakar. En 1945, su contribución a la migración total era de 14%, 32% en 1965 y 43% en 1989. Este crecimiento ha sido casi siempre desventajoso para la migración procedente del norte del país, que no representa más de 23% de la migración en 1989. En cambio, y al contrario de lo que se observó en Quito, la contribución del hinterland rural a la migración total hacia Dakar no ha variado mucho (34% en 1989).

**Gráfico 5 :** *evolución de la regiones de origen de los migrantes hacia Dakar desde 1935, tipología en tres espacios geográficos*

---

<sup>9</sup> : Daremos aquí a este término un sentido muy general : la modificación durable (a escala del periodo histórico de la observación retrospectiva), del reparto de las zonas de origen de los flujos migratorios hacia un lugar dado. El uso de esta expresión, esencialmente práctico, no implica equivalencia con la definición más precisa que le dan muchos autores en el análisis histórico de la transición demográfica en Europa : se trata entonces de una inversión estructural de tendencia, durante el siglo veinte, que hizo pasar muchos países europeos de la emigración neta a la inmigración neta. (vease por ejemplo J.C. Chesnais -1986- Cap. 6).

La unicidad de la estructura de los datos y del método de análisis facilita obviamente los estudios comparativos. Entre los dos contextos, senegalés y ecuatoriano, surgen inmediatamente de los resultados, similitudes y diferencias. En ambos casos un traspaso bien marcado de los lugares de origen de los migrantes hacia regiones más lejanas. Sin embargo las modalidades temporales y espaciales de esta transición se oponen de un país al otro, por ejemplo en cuanto al ritmo del fenómeno, o a la evolución de la contribución del hinterland rural a la migración total hacia las dos ciudades. Para ir más allá de estas simples constataciones, deberíamos por supuesto entrar en el examen de la estructuración espacial y económica de los dos espacios nacionales y considerar la historia del desarrollo urbano y de las crisis de la economía rural en ambos países...

Pero la identificación de un esquema global de evolución de los flujos migratorios hacia la ciudad estudiada no es el único alcance de esta metodología. Permite igualmente, cambiando el punto de vista para mirar la evolución del flujo procedente de un lugar de origen dado, abordar la cuestión de los factores determinantes de la migración en las zonas de expulsión <sup>10</sup>. Pero para que la perspectiva tenga interés, no se pueden considerar los dominios geográficos muy grandes y heterogéneos de las tipologías anteriores. Es una de las bondades del método, poder proporcionar diferentes tipologías de evolución de los flujos, a diferentes escalas. La más finas pueden agrupar espacios relativamente homogéneos, sea del punto de vista de la dinámica del flujo migratorio, o sea del punto de vista geográfico o socioeconómico <sup>11</sup>. En fin, es igualmente interesante a este nivel de análisis, añadir a la descripción de la dinámica del flujo, la caracterización de su composición demográfica. Para mostrar a cual tipo de comentario este enfoque puede llegar, tomaremos el ejemplo de las ciudades de los alrededores de Dakar (regiones de Thies y Diourbel), en donde observamos una estabilización y feminización de la migración hacia la capital.

Las ciudades de las regiones de Thies y Diourbel constituyen un conjunto con cierta homogeneidad y particularidad. Respecto a su integración económica y proximidad geográfica y cultural con Dakar, se lo podría caracterizar como el "hinterland urbano" de la capital. La expresión es en todo caso válida hasta los años cincuenta si se considera (gráfico 6a) la importancia que tenían entonces estas ciudades en el flujo migratorio total hacia Dakar. Desde 1935, la dinámica del flujo tuvo dos fases : un fuerte decrecimiento de la contribución relativa a la inmigración total de 1935 a 1960 (33% a 13%), seguido por un período de estabilidad hasta hoy en día (10,2% en 1989).

***Gráficos 6a et 6b : Contribución relativa a la migración total hacia Dakar (a) y estructura por sexo y edad a la fecha de última instalación en Dakar (b) de la población migrante originaria de las ciudades de las regiones de Thies y Diourbel***

---

<sup>10</sup> : Para este propósito, es por supuesto necesario tener informaciones adicionales (procedentes de censos por ejemplo) acerca de la dinámica migratoria global de estas zonas.

<sup>11</sup> : El grado de homogeneidad que permite alcanzar la técnica es por supuesto función, por una parte de la precisión geográfica con la cual se recolectó el lugar de origen de los migrantes, por otra parte del tamaño de la muestra.

## Gráfico 6a

## Gráfico 6b

En el contexto de la evolución de la red urbana senegalesa, y si se relaciona el constato echo anteriormente desde Dakar con las informaciones más generales que proporcionan los censos de 1976 y 1988, estas ciudades parecen involucradas en un proceso de transición de sus regímenes migratorios. Caracterizadas hasta 1950 por una importante emigración hacia Dakar, han conocido desde entonces un desarrollo demográfico y económico constante : Thiès, por ejemplo, con 175,000 habitantes en 1976, es hoy en día la segunda ciudad del país. En paralelo con este desarrollo, es muy probable que la tasa de emigración hacia Dakar aya bajado, para luego estabilizarse. En su globalidad, este espacio conserva ahora una parte muy importante de su crecimiento poblacional y algunas de estas ciudades juegan un papel importante en la retención de la migración de sus propios alrededores rurales. : Thiès y Mbour por ejemplo tienen saldos migratorios positivos entre los censos de 1976 y 1988.

La estructura por sexo y edad de esta población migrante es característica de una fuerte componente de migración familiar : 58% de mujeres (es la población de migrantes la más femenina de todas)y 32% de menores de diez años (solo 26% en el conjunto de los migrantes). Son estructuras muy femenina, en los migrantes solteros como en los de otros estados civiles, que determinan este promedio tan elevado : en la población de migrantes casados, viudos o separados por ejemplo, la proporción de mujeres alcanza 62,4%. Aquí se debe subrayar el papel importante de la poligamia, pues el 51% de las mujeres casadas de este origen tienen maridos polígamos cuando esta proporción es solo de 37% para las otras migrantes urbanas. De otra parte, el nivel de actividad fuera del hogar es significativamente menor para estas mujeres que en el



conjunto de las migrantes : entre 15 y 60 años, 32% trabajan o buscan empleo cuando la tasa es de 40% para el conjunto de las migrantes. Así aparece, para un grupo importante de mujeres originarias de estas ciudades, un esquema migratorio diferentes de los que se han identificados en otras regiones de alta migración femenina como son la Casamance o el país Serer, y en las cuales la movilidad de las mujeres es intimamente ligada a la actividad económica en Dakar. En el caso de las ciudades de las regiones de Thies y Diourbel, la migración femenina (en su mayoría Wolof) se relaciona principalmente con una estrategia de movilidad familiar en la cual la actividad económica de la mujer no tiene el papel principal.

## ***Conclusion***

No se puede hacer el reproche a los investigadores de ciencias sociales, y mucho menos a los planificadores o decisores políticos, el confiar en las teorías y técnicas estadísticas. Tampoco de tener esperanzas de aplicaciones concretas e inmediatas : identificación y medición del efecto de los determinantes de tal fenómeno social, pero más que todo, ajustes de dichos fenómenos por modelos, de alcances explicativos y preferiblemente prospectivos. Estos resultados son por supuesto muy atractivos cuando la perspectiva es de comprender una realidad social, y todavía más cuando se trata de preveer y actuar. El tema de las migraciones internacionales, siendo un campo de debates políticos y económicos importantes, con frecuencia lamentablemente viciados en contextos de xenofobia, no puede escapar a la regla y por lo contrario suscita por excelencia este tipo de esperanza.

Cuando escogimos dar la prioridad a un enfoque estadístico descriptivo de los datos biográficos migratorios, no pensamos despreciar en la investigación estos objetivos de aplicación, sino más bien tratar de reunir las condiciones previas para que los enfoques probabilistas sean más confiables y pertinentes. Así, mismo frente a una demanda social legítima y urgente, nos parece indispensable tomar el tiempo de describir muy cuidadosamente la realidad antes de pretender encerrarla en un modelo. Al no tomar esta precaución, los modelos se vuelven, demasiado a menudo, herramientas de desinformación.

En realidad, quizás el principal debate no es de oponer los enfoques estadístico descriptivo o probabilista de los datos de ciencias sociales. Cuando se aborda un fenómeno social a partir de su medición cuantitativa, la cuestión fundamental no sería a caso, más que de elegir uno o otro de estos enfoques metodológicos, de saber cual es la extensión y la naturaleza del espacio de los factores determinantes de dicho fenómeno. En tal caso es cierto que el estadístico no puede, solo, contestar a la pregunta.

### ***Bibliografía :***

- 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 Sénégambie, 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 conceptual y metodológica a las prácticas residenciales", in *Desarrollo y Sociedad* 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.