



**HAL**  
open science

# EL ANÁLISIS ARMÓNICO CUALITATIVO : TEORÍA Y APLICACIÓN A LA TIPOLOGÍA DE TRAYECTORIAS INDIVIDUALES

Olivier Barbary, Luz Mary Pinzon Sarmiento

► **To cite this version:**

Olivier Barbary, Luz Mary Pinzon Sarmiento. EL ANÁLISIS ARMÓNICO CUALITATIVO :  
TEORÍA Y APLICACIÓN A LA TIPOLOGÍA DE TRAYECTORIAS INDIVIDUALES. Revista  
colombiana de Estadística, 1999. hal-02478142

**HAL Id: hal-02478142**

**<https://hal.science/hal-02478142>**

Submitted on 13 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.

# EL ANÁLISIS ARMÓNICO CUALITATIVO : TEORÍA Y APLICACIÓN A LA TIPOLOGÍA DE TRAYECTORIAS INDIVIDUALES

Olivier BARBARY \*, Luz Mary PINZON SARMIENTO \*\*

## Resumen

**Este artículo presenta el análisis armónico cualitativo como herramienta para la estadística descriptiva de los procesos aleatorios. En la primera parte, nos interesamos el análisis armónico de un proceso cualitativo y su aproximación por el análisis de correspondencia de la tabla de las duraciones de presencia de los individuos en los estados. A continuación, su aplicación a una encuesta sobre las trayectorias residenciales, profesionales y familiares de residentes en el área metropolitana de Bogotá, muestra cómo el método permite conseguir una tipología de la movilidad intra-urbana y relacionarla con los demás elementos de la biografía.**

## INTRODUCCIÓN

La reflexión teórica en ciencias sociales concede cada vez más importancia a los enfoques dinámicos y a la noción de estrategia de los individuos y de los grupos sociales. En este contexto se implementan nuevos métodos de recolección de la información y las encuestas retrospectivas sobre biografías individuales, basadas en muestras representativas, toman importancia para alcanzar una observación más precisa del tiempo y de los espacios en los cuales transcurren las vidas humanas. Pero aunque sabemos con más precisión por qué y cómo recolectar datos biográficos, subsisten enormes dificultades para su análisis, particularmente en el campo de la estadística descriptiva. En efecto por lo general la prioridad es conseguir una tipología de las trayectorias individuales; sin embargo el análisis se queda en muchos casos a un nivel transversal o en indicadores longitudinales univariados, necesariamente reductores de la información original.

El análisis armónico cualitativo (AAC), presentado por primera vez por Deville y Saporta en 1980, es una herramienta estadística exploratoria multivariada que permite respetar la riqueza y la continuidad de este tipo de datos. Con este método, el análisis factorial de correspondencias (Benzecri [1973]) encuentra un nuevo campo de aplicación : la estadística descriptiva y tipológica de los datos longitudinales. La justificación teórica del AAC es objeto de la primera parte de este artículo. La segunda parte presenta rápidamente la encuesta sobre la movilidad de las poblaciones del área metropolitana de Bogotá y describe los pasos y los problemas de la aplicación del método a los datos biográficos. En la tercera parte intentamos ilustrar, con el comentario de algunos ejemplos de los resultados, las ventajas, limitaciones y perspectivas de este enfoque.

---

\* : Estadístico - ORSTOM, Profesor invitado de la Universidad Nacional de Colombia  
Email : olibarba@mafalda.univalle.edu.co

\*\* : Postgrado de estadística, Universidad Nacional de Colombia  
Email : lpinzon@ciencias.ciencias.unal.edu.co

# 1. BREVE TEORÍA DEL ANÁLISIS ARMÓNICO CUALITATIVO

A continuación se introduce el método de análisis armónico cualitativo (AAC), cuyos fundamentos teóricos se describen sintéticamente hasta llegar a una propiedad fundamental en la práctica: la equivalencia con un análisis de correspondencia particular. En seguida se revisan las técnicas de aproximación numéricas empleadas para su ejecución. Este resumen de la teoría sólo presenta a grandes rasgos los principales pasos de la justificación del método sin dar todas las demostraciones. El lector interesado en una presentación matemática más completa debe remitirse a los trabajos de J.C. Deville y G. Saporta [1980], G. Saporta [1981] y J.C. Deville [1982]), mientras en Saporta [1996] se encuentra otra presentación sintética del método orientada hacia el análisis de los datos biográficos.

## 1.1 Nomenclatura, definiciones y principio

Como marco para la exposición, vamos a considerar los siguientes elementos :

- un intervalo de tiempo  $T=[0,T]$ ,  $\beta$  el conjunto de los borelianos de  $T$ , la medida de Lebesgue sobre  $(T, \beta)$ , de forma que  $(T, \beta)$  es el espacio de tiempo medible.
- un conjunto finito de estados, de cardinal  $m$ ,  $\mathcal{X} = \{1, \dots, m\}$ .
- un espacio de probabilidad  $(\mathcal{X}, A, P)$  y el conjunto infinito de los procesos aleatorios cualitativos  $X$  sobre  $\mathcal{X}T$  :

$$X : \mathcal{X}T \rightarrow \mathcal{X} \\ (t) \rightarrow X_t = x$$

Si  $1_t^x$  es la función indicadora del evento  $X_t = x$ , y  $E(\cdot)$  la esperanza matemática bajo  $P$ , notaremos :

$$P^x(t), \text{ la probabilidad } P(X_t = x) = E(1_t^x) \text{ y}$$

$$P^{x,y}(t,s), \text{ la probabilidad } P(X_t = x \mid X_s = y) = E(1_t^x 1_s^y)$$

La idea directriz del análisis armónico cualitativo del proceso  $X$  es seguir los pasos del análisis factorial en el marco del álgebra de operadores : se empieza por buscar una codificación real del proceso con el fin de obtener un proceso aleatorio con valores reales y luego se hace la descomposición espectral del operador de varianza asociado a dicho proceso (descomposición llamada también análisis armónico, de donde proviene el nombre del método).

Una codificación real del proceso  $X$ , es una función  $f$  tal que :

$$f : \mathcal{X} \times T \rightarrow \mathfrak{R} \\ (x, t) \rightarrow y = f_t(x)$$

Obviamente, dentro del espacio de todas las codificaciones reales posibles, unas son más naturales y prácticas que otras. Se consideran las codificaciones de cuadrado integrable, es decir las que verifican :

$$\int_T f_t^2(x) P^x(t) dt < \infty \quad \forall x \in \mathcal{X}$$

## 1.2 Análisis espectral de un proceso cualitativo

Sea  $H$  el espacio de Hilbert de los procesos aleatorios reales sobre  $T$  de cuadrado integrable es decir ,  $H = L^2(\times T) = \{Y : \int_T \int_{\Omega} Y^2(\omega, t) dp(\omega) dt < \infty\}$ . Sea  $L^2(X)$  el

subespacio de  $H$  conformado por los procesos constantes en el tiempo y  $L^2(X_t)$  el subespacio engendrado por las codificaciones reales del proceso  $X$ . Para cada  $t$ ,  $L^2(X_t)$  es el conjunto de las variables  $\beta_t$  medibles, donde  $\beta_t$  es el conjunto de los borelianos engendrado por  $X_t$ , que se compone de variables numéricas (dependientes del tiempo) de la forma :

$$f_t(X_t) = \xi_t = \sum_{x \in \mathcal{X}} a_t^x I_t^x, \quad (a_t^x \in \mathfrak{R}) \quad (1)$$

Ahora  $E_t$  es el operador de esperanza condicional a  $X_t$ ,  $E_t(\xi) = E(\xi / X_t)$ , es decir la proyección ortogonal de  $L^2(\times T)$  sobre el subespacio  $L^2(X_t)$ . Las propiedades de  $E_t$  permiten demostrar que cada variable  $X_t$  es determinada por el operador  $E_t$  de esperanza condicional. Entónces queda claro que las dependencias estadísticas entre dos instantes  $t$  y  $s$  del proceso son resumidas por el producto  $K(t,s)$  de los operadores asociados a las v.a.  $X_t$  y  $X_s$ ;  $K$  es el operador sobre  $L^2(\times T)$  definido por :

$$K: L^2(\times T) \rightarrow L^2(\times T)$$

$$\xi \xrightarrow{K} \eta$$

con :  $\eta: (\omega, t) \rightarrow \eta_t(\omega)$

donde  $\eta_t(\omega) = \int_T E_t E_s(\xi_s) ds$  para todo  $t$  en  $T$ .

### Procesos propios y generatrices

Saporta [1981, pp 101-103] y Deville [1982, pp 61-62]) demuestran que  $K$  es compacto y de traza finita igual a  $mT$ . Como tal,  $K$  admite una descomposición espectral en vectores propios  $\xi^i$ , llamados también “procesos propios” :

$$K = \sum_{i=1}^{\infty} \lambda^i \xi^i \otimes \xi^i$$

Los  $\xi^i$ , asociados a los valores propios positivos  $\lambda^i$  ( $\sum_{i=1}^{\infty} \lambda^i = mT$ ), forman un conjunto ortonormal en  $L^2(\times T)$ , y los  $\lambda^i$  satisfacen la ecuación :

$$\lambda^i \xi_t^i = \int_T K(t,s) \xi_s^i ds \quad \forall i \quad (2)$$

Dado que los procesos  $\xi_t^i$  son medibles, es legítimo hacer intervenir en (2) las integrales y esperanzas condicionales, así :

$$\lambda^i \xi_t^i = \int_T E_t E_s(\xi_s^i) ds$$

$$\lambda^i \xi_t^i = E_t \int_T E_s(\xi_s^i) ds$$

Así, dado que  $E_s(\xi_s) = \xi_s$ , tenemos :

$$\lambda^i \xi_t^i = E_t \int_T \xi_s^i ds$$

Ahora se define la variable aleatoria  $z^i$ , llamada generatriz del proceso propio  $\xi^i$  por :

$$z^i = \int_T \xi_s^i ds,$$

de esta manera queda claro que la generatriz  $z^i$  no depende del tiempo y por consiguiente pertenece al subespacio  $L^2(\cdot)$  de  $L^2(\times T)$ , de los procesos constantes en el tiempo. Tenemos entonces  $\lambda^i \xi_t^i = E_t(z^i)$ , y los procesos propios  $\xi^i$  se obtienen, para cada  $t$ , como proyecciones ortogonales de los  $z^i$  sobre  $L^2(X_t)$ .

Utilizando la nueva variable e integrando la ecuación precedente, se obtiene :

$$\begin{aligned} \lambda^i \int_T \xi_t^i ds &= \int_T E_t(z^i) dt \\ \lambda^i z^i &= \int_T E_t(z^i) dt \end{aligned} \quad (3)$$

Si llamamos  $Q$  el operador  $\int_T E_t(\cdot) dt$ , la ecuación (3) se convierte en :

$$\lambda^i z^i = Qz^i \quad (4)$$

El análisis espectral de  $K$  (procesos propios  $\xi^i$ ) en  $L^2(\times T)$  puede deducirse del análisis espectral más simple de  $Q$  (generatrices  $z^i$ ) en  $L^2(\cdot)$ .

### Códigos propios

Otra forma de la ecuación de valores propios se obtiene si volvemos a la expresión (1) para  $\xi_t$ . Dejando de lado el índice  $i$  para simplificar la escritura, (2) se transforma en :

$$\begin{aligned} \lambda \xi_t &= E_t \left[ \int_T \xi_s ds \right] = \int_T E_t \xi_s ds \\ \lambda \sum_{x \in \mathcal{X}} a_t^x 1_t^x &= \int_T E_t \left( \sum_{y \in \mathcal{X}} a_t^y I_t^y \right) ds \\ &= \sum_{y \in \mathcal{X}} \left( \int_T a_s^y E_t(1_s^y) ds \right) \end{aligned}$$

Ahora,

$$E_t(1_s^y) = \sum_{x \in \mathcal{X}} \left( 1_s^y \frac{E(1_t^x 1_s^y)}{E(1_t^x)} \right) = \sum_{x \in \mathcal{X}} \left( 1_s^y \frac{P_{(t,s)}^{x,y}}{P_{(t)}^x} \right),$$

por consiguiente,

$$\lambda \sum_{x \in \mathcal{X}} a_t^x 1_t^x = \sum_{x \in \mathcal{X}} 1_t^x \sum_{y \in \mathcal{X}} \left( \int_T a_s^y \left[ \frac{P_{(t,s)}^{x,y}}{P_{(t)}^x} \right] ds \right),$$

y finalmente el sistema de ecuaciones :

$$\lambda a_t^x = \sum_{y \in \mathcal{X}} \left( \int_T a_s^y \left[ \frac{P_{(t,s)}^{x,y}}{P_{(t)}^x} \right] ds \right) \quad x = 1, 2, \dots, m \quad (5)$$

En resumen, las ecuaciones (4) y (5) muestran que los  $z^i$  y los  $a_t^x$  forman dos descomposiciones espectrales del proceso. En la primera, la serie de variables aleatorias ( $z^i$ ,  $i=1, \dots, mT$ ) son independientes del tiempo y en la segunda, los códigos reales no aleatorios ( $a_t^x$ ,  $x=1, \dots, m$ ) dependen del tiempo.

El análisis armónico cualitativo, al igual que el análisis canónico generalizado, se basa en el principio de identificar dos vectores aleatorios que tengan máxima correlación. El análisis armónico del proceso cualitativo  $X_t$  se reduce entonces a la búsqueda del elemento  $z^i$  de  $L^2()$  y del proceso codificado  $\xi_t = \sum_{x \in \mathcal{X}} a_t^x 1_t^x$  de  $L^2(X_t)$  que tengan máxima

correlación dentro de  $L^2(\times T)$ , e iterar este proceso bajo condición de ortonormalidad. Se trata pues del análisis canónico de los subespacios  $L^2()$  y  $L^2(X)$  en el espacio  $L^2(\times T)$ .

### 1.3 Equivalencia con el análisis de correspondencia y aproximación numérica

Ahora suponemos que existen  $p+1$  instantes ( $p$  finito)  $0=t_0 < t_1 \dots < t_{p-1} < t_p=T$  tales que el proceso sea estable sobre los  $p$  intervalos  $[t_{j-1}, t_j]$ , es decir que ningún individuo cambie de estado dentro de cada intervalo. Si  $\xi_j$  es el valor del código propio sobre  $[t_{j-1}, t_j]$ ,  $E_j$  la esperanza condicional de  $\xi_j$  y  $l_j$  la longitud del intervalo  $j$  ( $l_j = t_j - t_{j-1}$ ), entonces las anteriores ecuaciones se convierten en :

$$\begin{aligned} \xi_k^i &= \sum_{j=1}^p l_j E_k E_j \xi_j, \text{ y ya que } z = \sum_{j=1}^p l_j \xi_j, \\ \xi_k^i &= E_k(z_i) \end{aligned} \quad (2')$$

Dado que los códigos propios asociados a las variables  $\xi_k$  son funciones constantes sobre cada uno de los intervalos de la partición, el sistema (5) se puede escribir así :

$$\lambda a_k^x = \sum_{j=1}^p \sum_{y=1}^m \left( l_j \frac{P_{(k,j)}^{x,y}}{P_{(k)}^x} a_j^y \right) \quad x = 1, 2, \dots, m \quad (5')$$

La ecuación (5') corresponde a la ecuación fundamental del análisis canónico generalizado. Cuando es finito, compuesto por  $n$  individuos, el análisis se reduce al análisis de correspondencia (ver G. Saporta [1981], pgs 110-111 y J.C. Deville [1982], pgs 68-74).

En efecto en este caso las ecuaciones (2') y (5') son las ecuaciones del análisis de correspondencia del conjunto de las variables indicadoras de los eventos  $\{X_j=y, j \in T, y \in \chi\}$  multiplicadas por la medida del intervalo  $[t_{j-1}, t_j[$ . El código real  $a_k^x$  es entonces el vector propio asociado al valor propio en el análisis de correspondencia de una tabla disyuntiva particular con  $n$  filas y  $mp$  columnas (recordamos que  $n$  es el número de individuos,  $m$  el número de estados posibles y  $p$  el número de períodos en los que el proceso se mantiene estable). La casilla elemental de la tabla vale 1 si el individuo está en el estado considerado durante el período y 0 si no lo está. Esta tabla puede ser sometida al análisis de correspondencia pero inmediatamente surge un problema : en la mayor parte de los casos los números  $n$ ,  $m$  y  $p$  generan un cuadro de tamaño gigantesco, lleno de ceros, y el análisis no proporciona resultados interesantes. La solución práctica consiste en dividir el intervalo de observación en un número razonable de períodos (de duración constante o no) sin tener en cuenta los cambios individuales de estado. Posteriormente se construye el cuadro calculando la proporción de tiempo que ha permanecido cada individuo en cada uno de los estados a lo largo de cada período (densidad individual de presencia en los estados). En seguida se aplica el análisis de correspondencias simple y las técnicas de interpretación habituales. Cabe resaltar que esta codificación conserva íntegramente la información relativa a las duraciones individuales en los estados, pero no el orden cronológico de los cambios de estado cuando ocurren en un mismo intervalo de codificación.

Existen entonces distintas posibilidades para la definición de los intervalos de codificación y el cálculo de las frecuencias. Consideramos ahora el intervalo  $[0, T]$  dividido en  $p$  intervalos :  $[t_0, t_1[, \dots, [t_k, t_{k+1}[, \dots, [t_{p-1}, t_p]$ , de longitud igual o no. Los  $t_k$  son escogidos sin tomar en cuenta los cambios de estados individuales, según el conocimiento previo del proceso (argumentos propios de la(s) disciplina(s) involucrada(s) en la temática) o con base en la distribución de frecuencia de los cambios de estado en el tiempo (argumentos estadísticos). Por otra parte, abandonando el marco probabilístico definido en la parte teórica, podemos llamar  $I$  el conjunto finito de individuos ( $card(I)=n$ ) y  $J$  el conjunto finito de estados del proceso ( $card(J)=m$ ).

La codificación real  $Y(i, j, k)$  ( $i=1, n; j=1, m; k=1, p$ ) del proceso cualitativo es una función real definida sobre el producto cartesiano  $I \times J \times [0, T]$ . Ahora si  $\tau_{ijk}$  es el tiempo pasado por el individuo  $i$  en el estado  $j$  durante el intervalo de tiempo  $[t_{k-1}, t_k[$ , Deville y Saporta [1980] definen el valor de  $Y$  de la siguiente manera :

$Y(i, j, k) = \tau_{ijk} / T$ , proporción del tiempo total de observación del Proceso ( $T$ ) que el individuo  $i$  ha pasado en el estado  $j$  durante el intervalo  $[t_{k-1}, t_k[$ . En este caso, cualquiera sean los cortes del intervalo  $[0, T]$ , la métrica sobre el tiempo es uniforme y la suma de cada fila de la tabla de frecuencia (dimensión individuos) vale 1.

Sin embargo, para el cálculo algebraico efectuado en el análisis de correspondencia, no es para nada obligatorio que los distintos períodos de codificación sean de duración constante o que sea uniforme la métrica sobre el tiempo. Al contrario, son varios los argumentos a favor de otra alternativa. Por una parte, desde el punto de vista estadístico, tenemos interés en detallar la codificación durante los períodos en donde ocurren muchos cambios de estados, fijándose en el histógrama de la distribución temporal de los eventos (ver J.C. Deville [1982] y A. Florette [1988]). Por otra parte parece natural, en una problemática de análisis de las trayectorias residenciales de los individuos, hacer más énfasis en los cambios durante la edad adulta (es decir cambiar la métrica de acuerdo a cada periodo). En efecto, en esta etapa de la vida las decisiones pertenecen generalmente al mismo individuo, mientras durante la niñez o la vejez dependen más de otros individuos.

Así en la práctica se adopta a menudo la solución que consiste en calcular la densidad de presencia de los individuos en los estados en relación a la duración de cada intervalo de codificación :

$Y(i,j,k) = \tau_{ijk} / (t_k - t_{k-1})$ , proporción de la duración del intervalo  $[t_{k-1}, t_k]$  que el individuo  $i$  ha pasado en el estado  $j$  durante el intervalo  $[t_{k-1}, t_k]$ . En este caso, si los intervalos  $[t_{k-1}, t_k]$  no son de longitud constante, la métrica sobre el tiempo no es uniforme y la suma de cada fila de la tabla de frecuencia vale  $p$ , el número de períodos definidos en  $[0, T]$ .

Esta es la metodología que vamos a ilustrar ahora con un ejemplo.

## 2. LA ENCUESTA Y LA APLICACIÓN DEL A.A.C.

### 2.1 El operativo y el material de encuesta

#### *Contexto y problemática*

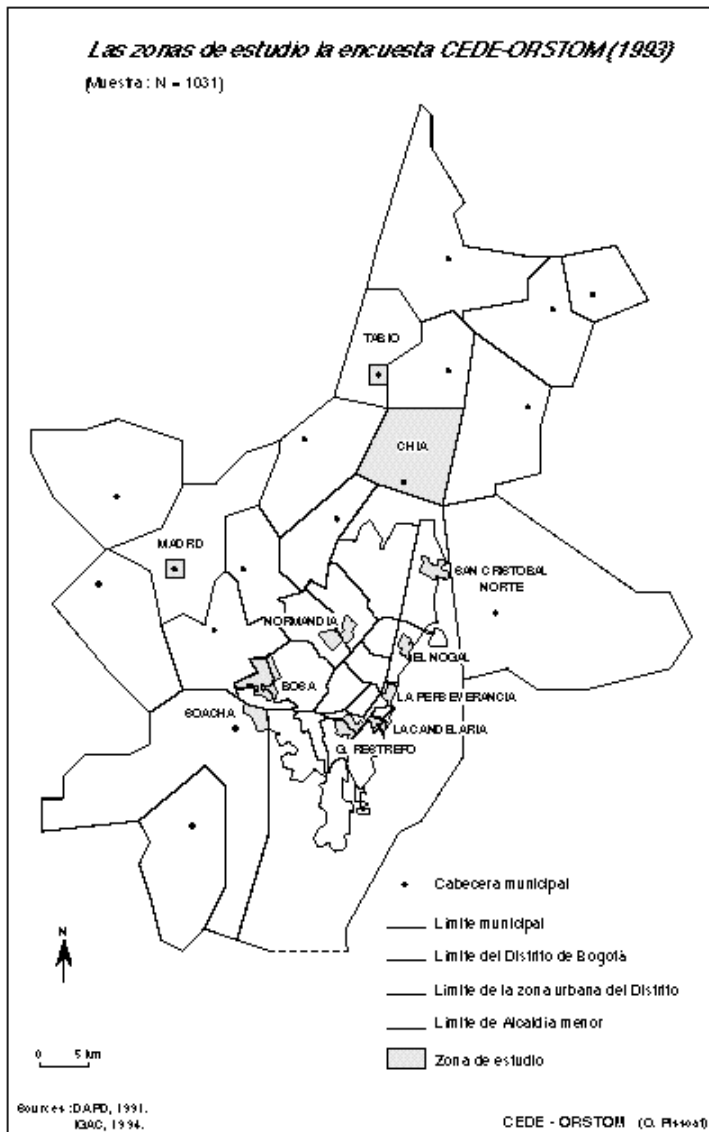
La encuesta y el trabajo estadístico presentados aquí se enmarcan en una investigación emprendida en agosto de 1992 por un equipo de investigadores del Centro de Estudios sobre Desarrollo Económico (C.E. Florez) y de ORSTOM (F. Dureau y M.C. Hoyos), sobre la movilidad de las poblaciones de Bogotá y su impacto en la dinámica del área metropolitana. Paralelamente a este programa la Universidad Nacional de Colombia y ORSTOM realizaron un programa de cooperación con el propósito de investigar los métodos de análisis estadístico de datos longitudinales.

Entre las capitales latinoamericanas, Bogotá fue la que experimentó el mayor crecimiento demográfico durante las décadas de los años cincuenta y sesenta (más de 6% anual). Entre 1951 y 1964 duplicó su población y alcanzó 2.5 millones de habitantes en 1970. Desde hace unos veinte años, el ritmo de crecimiento de la capital colombiana, al igual que las demás capitales del sub-continente, ha venido desacelerándose para llegar aproximadamente a 2.5% anual en el censo de 1985. Al momento de la encuesta CEDE/ORSTOM (final de 1993), Bogotá tenía una población cercana a 5.5 millones de habitantes y crecía a un ritmo ligeramente superior al 2% anual. La disminución del ritmo de crecimiento se explica por los efectos combinados de tres factores : la caída rápida de la fecundidad, el descenso en la intensidad de los flujos migratorios hacia la capital y la transformación del patrón de distribución geográfica de la población en beneficio de los municipios periféricos del área metropolitana. Dicho proceso de cambio demográfico se acompaña de nuevas estrategias de localización residencial, las cuales producen a su vez importantes modificaciones en el reparto de la población y en las modalidades de segregación socio-espacial en la aglomeración. Esta recomposición, que ha sido poco estudiada hasta el presente, constituye el centro de la problemática de la encuesta.



## Diseño de muestra

**Mapa 1:** Distribución de las zonas de estudio



Por razones de costo, el diseño muestral no buscó garantizar la observación del conjunto del área metropolitana de Bogotá, sino solamente ser representativo de 11 zonas de estudio (4 comunas o partes de comunas del área metropolitana y 7 barrios de Bogotá) que tienen un valor heurístico para la problemática del programa (mapa 1). En cada una de las 11 zonas, el diseño muestral consiste en un muestreo de áreas estratificado trietápico con probabilidades de selección desiguales de las unidades primarias. En la primera etapa, las áreas seleccionadas (U. P) son manzanas, definidas como unidades espaciales enteramente delimitada por vías o elementos naturales (ríos, quebradas). Se asegura la buena repartición geográfica de la muestra al seleccionar las manzanas por medio de una cuadrícula de puntos con tamaño definido en función de la tasa en cada estrato (muestreo espacial sistemático). Debido a este proceso, la probabilidad de selección de cada U.P. es proporcional a su superficie. En la segunda etapa se seleccionan en cada U.P. cinco U.S viviendas en

forma equiprobable a partir de una lista establecida en cada manzana de la muestra. En la última etapa, las unidades de observación son todos los hogares de las viviendas seleccionadas.

### *Población encuestada*

La población de las unidades de observación hogares (un grupo de personas que ocupan parte o totalidad de una vivienda y comparten las comidas) incluye los residentes habituales (personas que viven la mayoría del año en el hogar aunque al momento de la encuesta se hayan ausentado por corto tiempo - menos de tres meses) y los no residentes habituales (personas que viven la mayoría del año fuera del hogar pero que hayan habitado por lo menos 30 días -consecutivos o no- en el hogar encuestado, sean o no presentes en el momento de la visita del encuestador). Así forman parte del hogar, siempre y cuando cuplen con los requisitos de permanencia descritos anteriormente : las personas que prestan servicio militar y regresan regularmente a sus propios hogares, los alumnos internos o los

trabajadores que ejercen su actividad fuera del hogar o de Bogotá y regresan periódicamente a vivir en el hogar, las personas presas u hospitalizadas por cortos tiempo, los empleados del servicio doméstico u otros trabajadores cuando duermen en la vivienda, los huéspedes y pensionista. En cambio, las personas que tienen en arriendo uno o varios cuartos de la vivienda y cosinan por separado (inquilinos), forman hogares diferentes.

### *Cuestionario*

El cuestionario de la encuesta comprende, luego de un módulo clásico de información socio-demográfica sobre el conjunto de los individuos de los hogares (1031 hogares en la muestra), una serie de calendarios cuyo propósito es registrar la biografía residencial, profesional y familiar de una submuestra. Para esta ‘muestra biográfica’, se selecciona un individuo de más de 18 años en cada hogar (1031 personas), controlando su estructura con cuotas de sexo, edad, relación de parentesco con el jefe de hogar y estatus migratorio. Desde luego, las conclusiones del análisis no son válidas sino en el interior de este universo, para esa población particular y no para el conjunto de la población de Bogotá y de su área metropolitana <sup>1</sup>.

### *Variables para el análisis*

La variable longitudinal activa del análisis se basa en el conjunto de las etapas residenciales (con duración de un año o más) que han declarado los encuestados <sup>2</sup>. El nivel de agregación geográfica de la variable debe suministrar la máxima precisión en el análisis de la movilidad espacial intraurbana, siempre y cuando conservemos un número suficiente de casos en cada modalidad. La nomenclatura que le conviene a este objetivo es aquella de las Alcaldías menores de Bogotá (19 unidades, ver mapa 1). A estas 19 modalidades se agrega otra para las residencias dentro de la ciudad no especificadas, cuatro para las zonas de estudio de la periferia metropolitana (Chía, Tabio, Madrid y Soacha), una para los otros municipios del área metropolitana (A.M.) y, por último, una modalidad para las residencias situadas fuera del A.M. En total pues, contamos con 26 modalidades de estado.

Con el fin de obtener una visión de conjunto de las relaciones existentes entre la trayectoria espacial y otros componentes de la biografía individual, introducimos 7 variables longitudinales ilustrativas como resumen de cuatro capítulos de la biografía que podemos suponer determinados o determinantes de la trayectoria residencial : etapas del ciclo de vida (relación de parentesco con el jefe del hogar, estado matrimonial, coresidencia con los hijos, composición del hogar), acceso a la vivienda (condición de ocupación de la vivienda), carrera educativa (nivel educativo cursado) y movilidad socioprofesional (categoría socio-ocupacional). Cada una de estas variables posee su propio ‘calendario’, independiente del número de etapas residenciales.

Finalmente, una caracterización socioeconómica de los individuos y de los hogares correspondiente a cada patrón de movilidad se obtiene con una serie de variables ilustrativas transversales. En primer lugar se consideran las características sociodemográficas de los individuos : sexo, edad, status migratorio, nivel de instrucción, categoría socio-ocupacional, etc. Posteriormente los hogares se caracterizan por su tamaño y composición, las características de la vivienda (número de cuartos, condición de

---

<sup>1</sup> : Sobre el control de cuota, ver también el primer párrafo de la sección 3, pero no ha lugar, en este artículo, una presentación detallada de los temas ligados a la representatividad de los datos. El lector interesado debe remitirse a DUREAU F., FLOREZ C.E., BARBARY O., GARCIA L., HOYOS M.C. [1994].

<sup>2</sup> : Una presentación de la metodología de recolección y de la estructura informática de los datos biográficos se encuentra en F. Dureau y C.E. Florez [1996] y A. Morales [1996].

ocupación, hacinamiento) y del jefe del hogar (sexo, edad, status migratorio, edad promedio de los hijos).

## 2.2 Los pasos de la ejecución del A.A.C. hasta la clasificación de las trayectorias

### Manejo del tiempo

El análisis se realizó de acuerdo con el ‘tiempo biográfico’, es decir, siguiendo los individuos desde su nacimiento hasta la edad alcanzada en la fecha de la encuesta. Por supuesto, otras opciones son posibles, como por ejemplo el tiempo histórico (seguir a los individuos entre dos fechas) o un seguimiento a partir de un evento determinado (primera unión, primer nacimiento, etc.). No obstante, el tiempo biográfico es el más conveniente para nuestra problemática ya que permite relacionar, a nivel individual como estadístico, los diferentes itinerarios residenciales, familiares, educativos y profesionales.

**Figura 1 : Distribución de los cambios de residencia (RESIALC) según la edad**

RESIALC	Freq	Perct	Cum. Perct	
0	8	0.28	0.28	
1	53	1.86	2.14	
2	44	1.54	3.68	1 : 0-5 años
3	39	1.37	5.05	
4	33	1.16	6.21	
5	50	1.75	7.96	
6	44	1.54	9.50	
7	49	1.72	11.22	2 : 6-9 años
8	53	1.86	13.08	
9	58	2.03	15.11	
10	69	2.42	17.53	
11	42	1.47	19.00	3 : 10-12 años
12	68	2.38	21.39	
13	62	2.17	23.56	4 : 13-14 años
14	77	2.70	26.26	
15	102	3.58	29.84	5 : 15-16 años
16	96	3.37	33.20	
17	117	4.10	37.31	6 : 17-18 años
18	120	4.21	41.51	
19	112	3.93	45.44	7 : 19-20 años
20	117	4.10	49.54	
21	118	4.14	53.68	8 : 21 años
22	111	3.89	57.57	9 : 22-23 años
23	88	3.09	60.66	
24	93	3.26	63.92	10 : 24-25 años
25	83	2.91	66.83	
26	86	3.02	69.85	
27	76	2.66	72.51	11 : 26-28 años
28	52	1.82	74.33	
29	54	1.89	76.23	
30	79	2.77	79.00	12 : 29-31 años
31	42	1.47	80.47	
32	47	1.65	82.12	
33	44	1.54	83.66	13 : 32-35 años
34	45	1.58	85.24	
35	52	1.82	87.06	
36	39	1.37	88.43	
37	27	0.95	89.38	
38	30	1.05	90.43	
39	25	0.88	91.30	14 : 36-42 años
40	32	1.12	92.43	
41	18	0.63	93.06	
42	16	0.56	93.62	
43	18	0.63	94.25	
44	22	0.77	95.02	15 : 43-65 años
45	15	0.53	95.55	
etc...				

Una vez adoptada la opción del tiempo biográfico, conviene escoger un período de tiempo común a todos los individuos. Si se desea conservar la totalidad de las etapas, el individuo más viejo de la muestra (92 años) es quien fija el rango del período. Con el fin de evitar que el final de la tabla esté casi vacío, decidimos interrumpir el análisis a los 65 años de edad para la variable activa y a los 70 años para las variables ilustrativas, con excepción de la carrera educativa que se cortó desde los 45 años.

Dado que no han alcanzado estas edades al momento de la encuesta, la mayor parte de los individuos salen de observación antes del final del período de análisis (‘censura a la derecha’ en la terminología del análisis longitudinal). Hay que anotar que el tipo de censura depende del manejo del tiempo : con un tiempo histórico, por ejemplo, tendríamos censuras a la izquierda para los individuos que todavía no han nacido al principio del período. En este análisis las censuras se trataron, para cada una de las variables, con una modalidad

adicional en la que el individuo entra a partir de su edad a la fecha de la encuesta. Esta opción, que permite conservar todos los individuos durante la totalidad del período de análisis, puede criticarse ya que la censura incide en el resultado tipológico. Aunque se pueda limitar este efecto al tratar las modalidades de censura como ilustrativas en el AFC, o realizando análisis separados por cohortes de edades, decidimos no hacerlo puesto que buscamos una tipología global de la movilidad (para el conjunto de la muestra) y durante toda la duración de permanencia en el área metropolitana; en estas condiciones, un individuo de 20 años de edad no puede pertenecer al mismo tipo que un individuo de 50 años.

Como se dijo anteriormente, la puesta en práctica del AAC reposa sobre el corte del período de análisis en un número ‘razonable’ de intervalos de recodificación. El examen de la frecuencia de los cambios de estado según la edad (figura 1) permite definir 15 intervalos de amplitud variable siguiendo los cuantiles de la distribución; la precisión es buena entre 15 y 25 años, menor antes y después. La misma estrategia se aplica a las variables longitudinales ilustrativas. Para todas las variables, las frecuencias se calculan en proporción de la duración de cada intervalo de recodificación (la suma de cada fila vale 15). Así la tabla final comprende 1031 filas, 398 columnas activas ( $15 \times 27$  menos 7 columnas vacías) y 625 columnas ilustrativas.

#### *Ponderación de la muestra*

Habida cuenta del hecho que el diseño muestral no es autoponderado y que la selección de la muestra biográfica no es equiprobable (un solo individuo por hogar) ni tampoco probabilista (método de cuotas), surge el problema de saber si el análisis factorial debe hacerse sobre los datos observados o sobre los datos ponderados por los factores de expansión de los hogares. Los dos AFC dan resultados prácticamente equivalentes; en cambio para la clasificación en el espacio de los factores, la ponderación tiende a volver muy ‘volátiles’ los individuos encuestados en Chía, Madrid y Tabio (dominios con poblaciones pequeñas), mientras que la inercia de los individuos de zonas más pobladas es muy fuerte. Para la identificación de los diferentes tipos de itinerario presentes en la muestra, es preferible, entonces, proceder con AFC y clasificación no ponderadas. Al contrario, nos pareció importante tener en cuenta los pesos demográficos, aunque aproximados, para describir y caracterizar las clases. Desde luego, todas las indicaciones cifradas en la discusión de la tipología (sección 3.) son valores obtenidos con ponderación por el inverso de la probabilidad final de inclusión de los hogares.

#### *Análisis factorial*

El análisis de correspondencias suministra un histograma de valores propios muy liso (figura 2), lo que no debe sorprender dada la estructura de la tabla : gran número de columnas respecto al número de filas y abundancia de celdas nulas. Hay que recordar que para el resultado del AFC, cualquier permutación de las columnas es indiferente. En otras palabras, dos individuos con los mismos itinerarios, simplemente rezagados en uno o dos años, pueden tener coordenadas factoriales muy diferentes. La tabla analizada es entonces muy ‘ruidosa’ con respecto a la estructura de proximidades que nos interesa y el histograma de valores propios refleja este ruido.

**Figura 2 :** *Histograma de los diez primeros valores propios del AFC (Tabla 1031×398)*

```

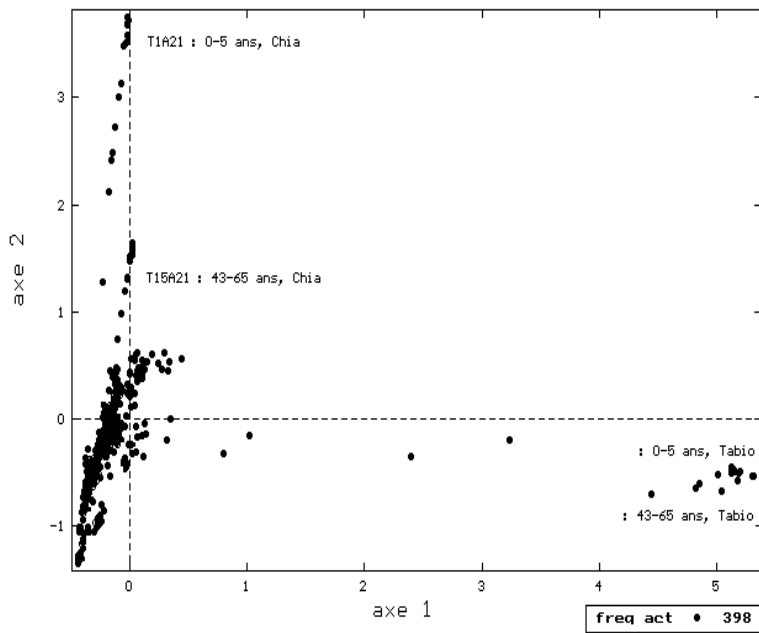
-----
SPAD.N Sistema Portable para el Analisis de Datos
Copyright (C) CISIA, 1987, 1991 - Version 2.52
-----

```

NO	VALOR PROPI	PCENT	PCENT ACUMU.
1	.7099	3.17	3.17
2	.6291	2.81	5.98
3	.5973	2.67	8.65
4	.5800	2.59	11.24
5	.5562	2.48	13.72
6	.5291	2.36	16.09
7	.5046	2.25	18.34
8	.4848	2.17	20.51
9	.4834	2.16	22.67
10	.4602	2.06	24.72

etc....

**Figura 3 : Variables activas sobre el plano factorial 1x2**



A pesar de todo, la interpretación de los primeros factores es clara : si bien es cierto que con la codificación hemos perdido la cronología, esta meta-información estructura tan fuertemente los datos que resurge sobre todos los ejes factoriales útiles. El plano 1x2 (figura 3) muestra el tipo de estructura evidenciada por los primeros ejes : las variables que representan un mismo lugar son agrupadas y ordenadas cronológicamente a lo largo de los ejes. Esto se debe a unos grupos de individuos minoritarios pero muy fuertemente caracterizados por cada lugar. En el caso de

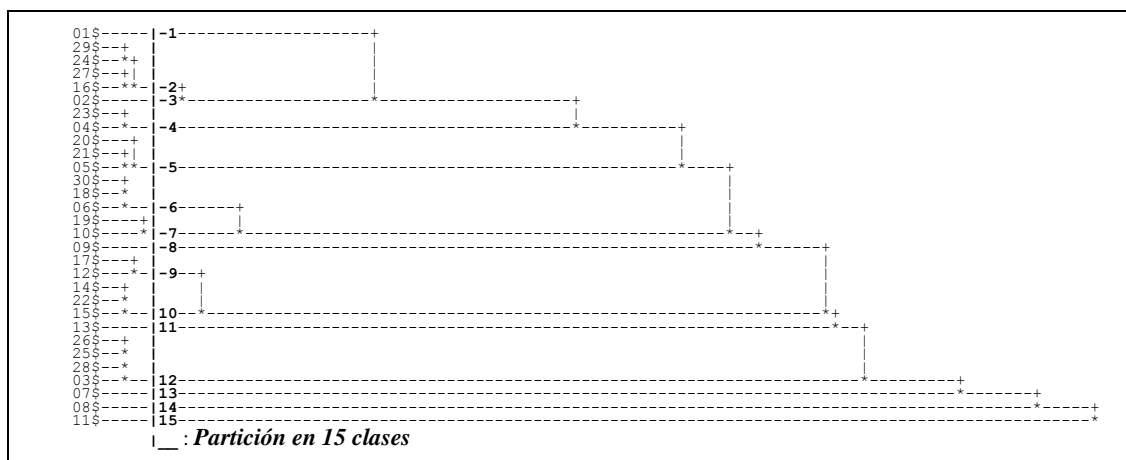
los dos primeros factores, se trata de los subgrupos de la muestra estables durante toda la observación en Tabio, Chía y Madrid. A partir de los primeros factores del análisis, se logra sin dificultad la caracterización de los grupos estables de la muestra. Estos individuos representan aproximadamente el 22% del conjunto de la muestra y alrededor del 52% de la inercia explicada por los diez primeros factores. Los patrones de movilidad que caracterizan el resto de los individuos (el 78% de la muestra) se identifican menos fácilmente en los factores. Un cierto número de ejes muestran asociaciones entre modalidades que significan transiciones frecuentes a una cierta edad, entre diferentes lugares. No evidencian grupos de individuos con trayectorias globalmente idénticas, sino grupos que tienen en común una cierta transición a una edad dada. Este tipo de resultado, clásico en la aplicación del análisis armónico cualitativo (J.C. Deville [1982], P. Béret [1988] [1995]), es insuficiente con respecto a nuestro objetivo tipológico.

#### *Clasificación de la nube de individuos*

Para conseguir una tipología completa de las trayectorias, hemos clasificado los individuos en el espacio vectorial de los primeros factores del A.F.C. Veremos que se pueden considerar los diez primeros ejes como los factores útiles, totalizando así el 25% de la inercia total. Despreciar el 75% de la información original puede sorprender, pero

recordemos que con 26 modalidades activas tenemos una tabla bastante dispersa donde una gran parte de la variabilidad (los pequeños desfases entre trayectorias equivalentes), carece de interés para nuestro propósito. En este espacio la nube de individuos puede someterse a algoritmos de clasificación ascendente jerárquica como CLUSTER - criterio de Ward - en SAS o SEMIS en SPADN. Ambos dan resultados muy próximos y conservamos los de SPADN, ya que tienen la ventaja de optimizar la partición una vez escogido el número de clases. Haciendo variar el número de factores constatamos que hasta 10 factores las clases ganan homogeneidad, pero a partir del décimo primer factor, el tamaño de la clase más importante aumenta. Esta base empírica conduce a la clasificación final (figura 4).

**Figura 4 :** *Árbol de clasificación en el espacio de los diez primeros factores del AFC*



La partición en 15 clases constituye la tipología que vamos a describir. Una vez maximizada la inercia interclase (PROC PARTI de SPADN), dicha clasificación conserva el 82% de la inercia total de la nube en el espacio de los diez primeros factores. Podemos entonces concluir que la metodología permite captar la mayor parte de la información significativa (es decir excluyendo el ruido) que proporciona la encuesta acerca de la heterogeneidad de las trayectorias residenciales.

#### *Caracterización de las clases y otros resultados*

Debido al carácter longitudinal de los datos, el trabajo de caracterización de las clases se vuelve un tanto particular. En efecto, aunque es posible dar una interpretación ‘indirecta’ de las clases a partir de los factores, siempre es más directo y enriquecedor volver a la tabla de frecuencias individuales. Los perfiles de movilidad de las clases se obtienen calculando las frecuencias promedio de permanencia de los individuos de cada clase en los estados durante los intervalos de codificación, y se traducen en gráficas (figuras 8 a 22 del anexo). Asimismo calculamos los perfiles correspondientes a las variables longitudinales ilustrativas (figuras 5 a 7). Para que estos perfiles sean comparables entre ellos y al perfil del conjunto de la muestra, basta con eliminar el efecto de la estructura por edad diferente, que no es otra cosa que la frecuencia de censura en las diferentes edades. Se calculan entonces las frecuencias, normalizadas a 100% para cada edad, entre los individuos de la clase que alcanzaron esta edad (excluyendo los individuos censurados). Por último, se produce la serie de tablas cruzadas de la tipología con las características transversales de los individuos y de los hogares. El grado de ‘asociación’ de las clases con cada característica se mide con la prueba de  $\chi^2$  asociada a la tabla, correlaciones de rangos o regresiones logísticas.

Este material sirve de base para la descripción del patrón de comportamiento residencial de cada clase (trayectoria espacial de los encuestados). A partir de las variables

longitudinales ilustrativas, podemos relacionar esta trayectoria con la sucesión de los eventos del ciclo de vida. Apuntamos también las características demográficas y socioeconómicas de los individuos y de los hogares, completando así la ‘cédula de identidad’ de la clase. Como lo vamos a ver a continuación en los ejemplos de resultados, se desprenden entonces algunas hipótesis acerca de las interacciones que estructuran la biografía de los individuos, que podemos someter a verificación.

### 3. EJEMPLOS DE RESULTADOS

Antes de examinar los perfiles de las variables longitudinales, hay que destacar que el control de cuotas permite que la estructura de la submuestra biográfica según el sexo, la edad y el parentesco con el jefe del hogar no se desvíe mucho de aquella de la muestra compuesta de todos los miembros del hogar con 18 años de edad o más. Las únicas diferencias notables se deben a la condición del individuo dentro del hogar: sobre-representación de los jefes de hogar (42% vs. 36% en el conjunto de la población de los hogares) y de los cónyuges (33% vs. 27%) en detrimento de los hijos del jefe del hogar (17% vs. 25%). Esta distorsión repercute lógicamente sobre el estado conyugal (solteros menos representados: 25% vs. 32%) y sobre la condición laboral (amas de casa más representadas: 28% vs. 23%). Son diferencias de amplitud razonables y podemos abordar sin temor el análisis de los datos biográficos en el conjunto de la muestra. Sin embargo, cuando se considera la estructura de la muestra biográfica dentro de cada zona de estudio, las distorsiones pueden acentuarse. Por lo tanto habrá que estar alerta en la interpretación de las clases que agrupan individuos procedentes de ciertas zonas de estudio. Por ejemplo, las muestras biográficas en Bosa y Soacha contabilizan demasiadas mujeres, respectivamente 65% vs. 51% y 68% vs. 53%, mientras que aquellas de Tabio y Madrid son demasiado masculinas, respectivamente 62% de hombres vs. 47% y 58% vs. 50%. Por último, cuando el número de integrantes de una clase es pequeño (es el caso en las clases que describen grupos de poblaciones estables), aparecen distorsiones más fuertes que complican la interpretación de los resultados.

#### 3.1 Un esquema global de la movilidad espacial de los individuos de la muestra

Vista en su conjunto, la tipología en 15 clases refleja dos fenómenos principales: la movilidad intra-urbana entre alcaldías menores y la migración procedente del exterior del área metropolitana - más precisamente, la edad y el lugar de llegada de los migrantes al área metropolitana. Si se consideran los dos indicadores sintéticos de la tabla 1 (Imobalc - col. 8- y Pctmig -col. 5-), la correlación con el rango de las clases es evidente y podemos entonces distinguir 4 grupos :

*1. Muy estables* (8% de la muestra, figuras 8 a 10 del anexo)

Este grupo conformado por las clases con la movilidad intra-urbana más baja se ubica en su totalidad fuera de Bogotá: estables en Chía (clase 14), Madrid (clase 13) y Tabio (clase 15). Lógicamente, encontramos en las tres clases una gran mayoría de nativos del área metropolitana, aunque los migrantes no están totalmente ausentes (35% en Chía).

*2. Estables* (18% de la muestra, figuras 11 a 15)

Un poco más móviles, las clases 9, 10, 11 y 12 están constituidas por núcleos de población nativa estable en diversos barrios de Bogotá : Usaquéen (clase 10), Rafael Uribe y Antonio Nariño (clase 11), La Candelaria y Santa Fé (clase 12), a los cuales se agregan migrantes jóvenes estabilizados en Madrid (clase 2) y en Usaquéen (clase 9).

*3. Movilidad intra-urbana media* (44% de la muestra, figuras 16 a 20)

El grupo más importante de la tipología está formado por individuos que han tenido alrededor de dos cambios de alcaldía desde su entrada al área metropolitana y se caracterizan por estadías largas en ciertos barrios, sin que necesariamente las residencias en la fecha de la encuesta estén muy concentradas : clase 8 (estadías largas en otros municipios del área metropolitana), clase 7 (estadías largas en la Candelaria y Teusaquillo), clase 5 (estadías largas en Rafael Uribe), clase 4 : estadías largas en la periferia Oeste), clase 1 (viejos migrantes en todos los barrios).

*4. Movilidad intra-urbana fuerte (30% de la muestra, figuras 21 y 22)*

El grupo más móvil comprende la clase 6, compuesta de viejos migrantes llegados jóvenes y nativos que han tenido una fuerte movilidad en los barrios del peri-centro norte y de la periferia Norte y Oeste; y la clase 3 que reagrupa a los jóvenes migrantes llegados entre los 10 y 25 años, cuyos itinerarios se ubican frecuentemente en la periferia Oeste y Sur. En la clase 3 las residencias en la fecha de la encuesta están bastante concentradas en Bosa, Soacha y Gustavo Restrepo.

**Tabla 1** : Indicadores de movilidad promedia para las quince clases de la tipología

CLASES (1)	NINDIV (2)	PESO (3)	PCTPESO (4)	PCTMIG (5)	DURMED (6)	IMOBRES (7)	IMOBALC (8)
1	217	36156	22,32	95,9	21,3	12,7	7,86
2	25	2326	1,44	69,0	21,8	9,32	4,02
3	196	36557	22,57	83,3	15,8	18,5	14,23
4	72	13890	8,58	39,9	27,6	9,57	7,08
5	65	14706	9,08	30,1	30,0	9,33	6,67
6	89	11889	7,34	54,7	38,6	14,0	11,05
7	32	3039	1,88	12,2	47,4	8,69	6,13
8	34	4244	2,62	24,8	40,8	9,16	6,00
9	37	6515	4,02	40,3	23,6	10,2	5,56
10	25	4294	2,65	11,3	34,8	5,91	2,83
11	26	6157	3,80	25,8	35,6	4,98	3,15
12	90	9486	5,86	19,2	39,3	7,23	3,60
13	39	3529	2,18	13,6	29,7	5,20	1,56
14	48	7956	4,91	35,2	29,2	7,44	1,16
15	36	1234	0,76	6,6	43,2	5,41	2,17
Total	1031	161978	100	58,4	26,4	10,9	7,23

- (1) : Rango de la clase en la tipología
- (2) : Número de individuos
- (3) : Población expandida de la clase
- (4) : % de la población total expandida
- (5) : % de individuos nacidos fuera del a.m.
- (6) : Duración promedia de residencia en el a.m.

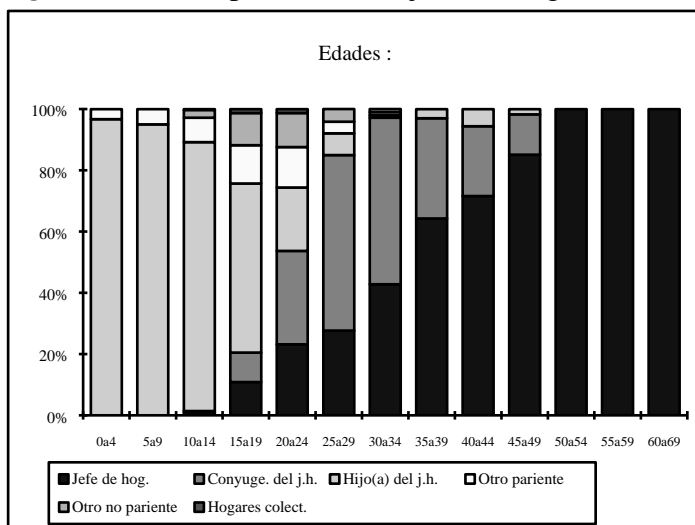
- (7) : Frecuencia de cambios de viv. en el a.m. (100 x número de cambios de viv. / núm. de años de resid. en el a.m.)
- (8) : Frecuencia de los cambios de alcald. en el a.m. (100 x núm. de cambios de alc. / núm. de años de resid. en el a.m.)



### 3.2 Algunos elementos sobre la inserción de los jóvenes migrantes en Bogotá

#### Modalidades de inserción residencial al llegar a la ciudad

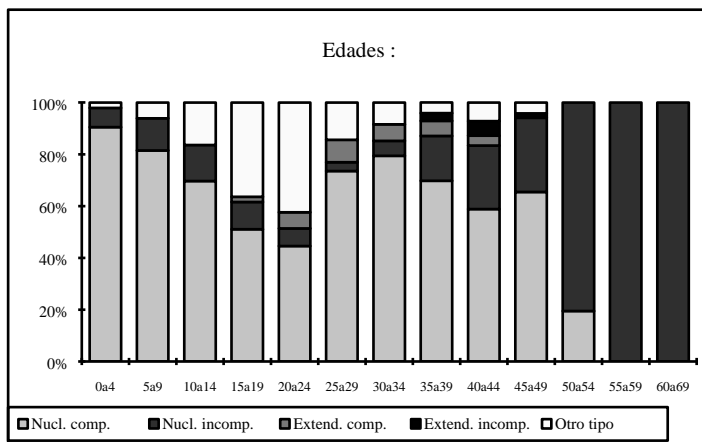
**Figura 5 :** rel. de parent. con el jefe del hogar (clase 3)



En la clase 3, el hecho destacable en el perfil longitudinal de la relación de parentesco con el jefe del hogar (figura 5) es la salida muy rápida del hogar de los padres, pero más que todo, que estas salidas no siempre corresponden a entradas directas a la condición de jefe del hogar o de cónyuge: entre los 15 y 24 años, las frecuencias de “otros parientes” y “otros no parientes” del jefe del hogar alcanzan el doble de los valores observados en toda la muestra. En estas situaciones de transición,

correspondientes a distintas formas de hospedaje gratuito o no, la ‘circulación’ es muy rápida: así, el 54% de los individuos de la clase experimentó un episodio de este tipo en algún momento (vs el 27% en el resto de la muestra), siendo el hecho mucho más frecuente entre las mujeres (65%) que entre los hombres (37%). Con tal nivel de frecuencia, estas formas de alojamiento aparecen como una de las modalidades principales de inserción residencial de los jóvenes migrantes a su llegada a Bogotá.

**Figura 6 :** composición del hogar (clase 3)



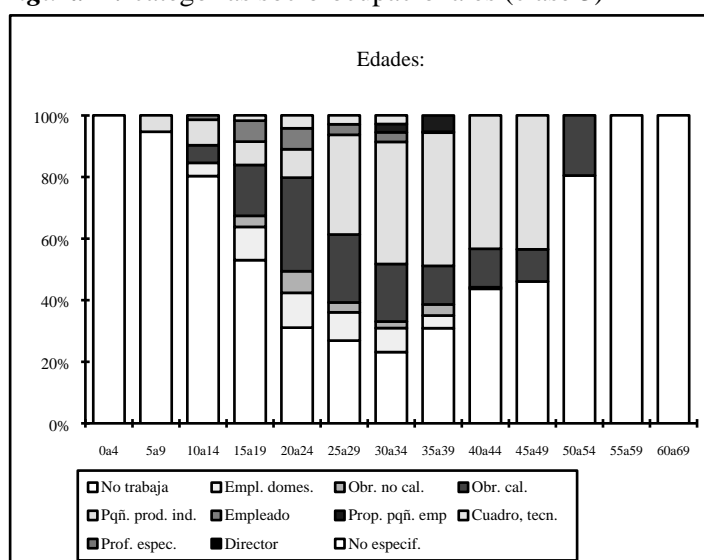
El perfil de la composición de los hogares (figura 6) también muestra particularidades al momento de la llegada al área metropolitana: entre los 10 y los 25 años es impactante la frecuencia de “otros tipos”, modalidad que reagrupa los hogares unipersonales y los hogares no familiares. Allí de nuevo, a causa de la fuerte ‘circulación’, la frecuencia real de estos episodios es superior a lo que parece en la gráfica: abarcan

el 72% de los individuos de la clase (vs 41% en el resto de la muestra). En esta ocasión el echo es más frecuente para los hombres (84%, contra 64% de las mujeres). Volviendo a los datos individuales, aparece que para las mujeres el 93% de los episodios de vida en hogares no familiares se dan como otro parientes o otro no parientes del jefe de hogar (hospedajes), mientras que para los hombres una fracción importante (30.4%) corresponde a una condición de jefe de hogar, donde comparten la vivienda con amigos o parientes colaterales.

Así, la inserción residencial de los jóvenes migrantes entre los 10 y 25 años de edad se caracteriza por etapas de hospedaje o de vida en hogares no familiares. Por una parte, este patrón es bastante distinto de aquel que se da, a las mismas edades, para los otros individuos de la muestra. Por otra parte, existe un fuerte diferencial según género, con hospedajes en hogares de familiares muy frecuentes para las mujeres, mientras los hombres experimentan casi sistemáticamente episodios en hogares unipersonales o compartiendo vivienda con amigos o colaterales. Sin embargo, para las mujeres como para los hombres, estas situaciones son eminentemente transitorias, ya que a partir de los 25 años se observa un acceso rápido a las condiciones de jefes de hogares y cónyugues (figura 16) así como la formación de hogares familiares (nucleares completos o incompletos, figura 17).

### Una movilidad profesional fuerte y particular

**Figura 7 :** categorías socio-ocupacionales (clase 3)



Los jóvenes migrantes se distinguen también por sus formas de inserción laboral (figura 7), en particular a causa de una movilidad específica entre empleos asalariados e independientes. La distribución de los empleos no se aleja mucho del patrón global hasta los 24 años, aunque se nota una proporción un poco superior de empleados domésticos (11% entre 15 y 24 años vs. 8% global). Después de los 25 años, en cambio, el perfil se destaca por la proporción muy superior al promedio de pequeños productores independientes (sin

equivalentes en las otras clases de la tipología) : 44% de los empleos ocupados entre 25 y 29 años (vs. 19% global) y hasta 77% entre 40 y 44 años (vs. 33%). Dado que se trata de una clase que contiene cerca del 23% de la muestra, estamos en presencia de un hecho de importancia mayor. Volviendo a los datos individuales constatamos que su explicación reside en una movilidad entre el empleo asalariado y el independiente particularmente fuerte y concentrada en el tiempo, ameritando que se detalle su mecanismo.

En el conjunto de los individuos que han ejercido un empleo en la fecha de la encuesta (84% de la clase), el 35% ocupó por lo menos un empleo independiente, pero sus itinerarios profesionales son de dos tipos distintos. En efecto, para el 22%, la actividad independiente ha sido el primer empleo y, en este caso, todos la iniciaron antes de los 28 años y 90% antes de los 25 años. Los otros (13%) acceden a un empleo independiente después de otros empleos y esta transición se da para el 21% de ellos antes de los 22 años, para el 74% entre 22 y 25 años y solamente para el 5% después de los 25 años, lo que causa la modificación marcada del perfil global de actividad a esta edad. Por otro lado, es importante anotar que la transición inversa (pequeños productores hacia otros empleos) no es menos frecuente : 15% de los individuos que han tenido un empleo en la fecha de la encuesta la han experimentado, pero todos antes de los 22 años. Esto contribuye también a limitar las densidades de presencia en los empleos independientes antes de los 24 años. Así, lo que explica la recomposición del perfil no es el saldo de la movilidad entre los empleos

de pequeños productores y los otros (en la clase este es más bien favorable a los segundos), sino el desfase en el tiempo de los dos tipos de movilidad. Las salidas de los empleos de pequeños productores hacia otros ocurren entre los 15 y los 22 años mientras que, en la mayoría de los casos, las entradas en estos empleos se dan entre los 22 y los 25 años. Queda por saber si el tipo de actividad y el ingreso procurado son equivalentes en ambos casos...

## CONCLUSIONES

Cuando se plantea su aplicación a datos biográficos complejos, el análisis armónico cualitativo presenta una serie de ventajas importantes a nivel teórico y práctico.

En primer lugar, el manejo del tiempo es flexible y permite varias estrategias con el fin de conservar mejor la información continua y asincrónica original, y ajustar su codificación a la problemática del análisis. Se puede adoptar lógicas de división del tiempo diferentes a las empleadas en este artículo. Por ejemplo, cuando los intervalos de recodificación son escogidos con fechas que se consideran determinantes en el contexto conyuntural, político o social del período, el análisis toma un giro más histórico, lo que puede proporcionar otro punto de vista interesante sobre los datos (véase por ejemplo O. Barbary [1993], pp 30-64). Alternativamente se puede también adoptar un tiempo biográfico más colectivo, donde la sincronización no se hace según la edad de los individuos, sino según el tiempo transcurrido después (o antes) de un acontecimiento clave del ciclo de vida : primera llegada de los migrantes al lugar de la encuesta, primera salida del hogar de los padres, primera unión, primer acceso a la propiedad de la vivienda etc...

Otros aspectos a considerar son los relacionados con la división óptima del período de análisis (número y límites de los intervalos de codificación del proceso), y la elección de la métrica sobre el tiempo. Estas preguntas se pueden plantear en un sentido teórico, con optimización de un criterio estadístico, lo que abre espacios para investigaciones futuras. Limitándonos al terreno empírico, hemos visto que no hay obligación de que los intervalos de codificación sean de duración constante o que la métrica sea uniforme en el tiempo; estos parámetros son también importantes para el ajuste del método.

Para realizar el análisis multivariado de diferentes tipos de eventos biográficos, el A.A.C. ofrece dos posibilidades. La primera, que no hemos adoptado aquí, consiste en analizar la variable compleja resultante del cruce de las diferentes situaciones residenciales, profesionales, familiares etc. Esta alternativa parece atractiva ya que agrupa los distintos temas en la variable activa y, desde luego, les da la misma importancia en el resultado tipológico. Sin embargo, no creemos que un buen uso del método es el que consiste en mezclar todos los capítulos de la biográfica y esperar que el cálculo nos muestra el conjunto de relaciones que los estructura. Existen limitantes de varios órdenes a esta práctica. En primer lugar una condición epistemológica : aún cuando el enfoque del análisis de datos tiene vocación exploratoria, su empleo nunca dispensa de una problemática bien definida de la cual se desprende entonces una jerarquización de la variables. Por otra parte, desde el punto de vista de la fiabilidad estadística, el tamaño de la muestra no permite generalmente aumentar mucho el número de modalidades de estados biográficos.

La segunda alternativa, la técnica de las variables longitudinales ilustrativas, nos parece la mejor. En primer lugar, la prioridad que se le da a la variable activa es afirmación del propósito experimental asignado al análisis para contestar a un interrogante preciso. En este marco, la interpretación de los resultados encuentra un principio director que lleva al

estadístico, si se esfuerza un poco, a formular hipótesis o conclusiones interesantes para sus interlocutores de ciencias sociales. En segundo lugar, permite mantener normas de representatividad aceptables con muestras que no pueden ser muy grandes (la recolección de biografías resulta compleja y costosa!).

Para finalizar, cabe resaltar que, además de los resultados descriptivos que logra evidenciar, el método presenta también un interés evidente desde una perspectiva de modelage. Por una parte, el resultado tipológico permite, cuando se busca modelar el efecto de variables independientes, ajustar modelos sobre subpoblaciones más homogéneas que el conjunto de la muestra. Por otra parte, cuando el número de casos en la clase es suficiente (en general 30 a 50 individuos), los perfiles de movilidad sugieren a menudo las formas y los parámetros de las distribuciones de los tiempos de permanencia en los estados biográficos. De esta manera las hipótesis pueden ser confirmadas, eliminando el efecto de los datos censurados, con estimaciones no-paramétricas (Kaplan Meier). Este material empírico, cuya ausencia es una crítica frecuentemente dirigida a los enfoques de estadística inferencial en ciencias sociales, es de gran utilidad en la construcción de modelos paramétricos o semiparamétricos.

#### **AGRADECIMIENTOS**

Queremos agradecer la profesora Jeanne FINE y el profesor Leonardo BAUTISTA para la lectura cuidadosa que hicieron de la primera versión de este artículo, sus comentarios y sugerencias fueron de gran utilidad.

## **NOTAS BIBLIOGRÁFICAS :**

BARBARY O. [1996] "Análisi tipológico de datos biográficos en Bogota, Universidad Nacional de Colombia, Col. Textos n° 24, Bogota, 254 p.

BARBARY O. Ed. Cient., [1996] : "Recolección y Análisis de Datos Longitudinales", Memorias del seminario de capacitación e investigación, Bogotá, 9-13 de diciembre de 1996. Universidad Nacional de Colombia, PRESTA, ORSTOM, Bogotá, 425 p.

BARBARY O. [1996] : Una aplicación del análisis armónico cualitativo : la tipología de trayectorias individuales, in Memorias del seminario de capacitación e investigación "Recolección y análisis de datos longitudinales", Bogotá 9-13 dec. 1996, Universidad Nacional de Colombia - ORSTOM - PRESTA, Bogotá pp. 121-144.

BENZECRI J.P. et al. [1973], "L'analyse des données. Tome 2 : L'analyse des correspondances" DUNOD, Paris, 632 p.

BERET P. [1988] : Analyse de données chronologiques relatives à l'insertion professionnelle, in "Les Cahiers de l'Analyse des Données" vol. XIII, n° 2, pp. 159-174.

DEGENNE A., LEBEAUX M.O., MOUNIER L. [1995] : Construction d'une typologie de trajectoires à partir de l'enquête de suivi des jeunes des niveaux V, Vbis et VI. Communication aux Journées CEREQ-LASMAS-IDL sur les données longitudinales dans l'analyse du marché du travail, Caen, 28-29 juin 1995, 14 p.

DEVILLE J.C., SAPORTA G. [1980] : Analyse harmonique qualitative, in "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..

DUREAU F., FLOREZ C.E., BARBARY O., GARCIA L., HOYOS M.C. [1994] : La movilidad de las poblaciones y su impacto sobre la dinámica del área metropolitana de Bogotá", metodología de la encuesta cuantitativa. Documento de trabajo n°2 : CEDE/ORSTOM, Bogotá, 393 p.

DUREAU F., BARBARY O., FLOREZ C.E. [1994] : "La movilidad de las poblaciones y su impacto sobre la dinámica del área metropolitana de Bogotá", resultados preliminares de la encuesta cuantitativa. Documento de trabajo N° 3, CEDE/ORSTOM, Bogotá, 309 p.

FLORETTE A. [1988] : Approximation et choix du découpage dans le cadre de l'analyse harmonique qualitative, Mémoire de DEA, ENSAE, Paris.

FLOREZ C.E, MENDEZ R. [1994] : Boletín de estadísticas sobre hogar y familia en Colombia, 1972-1992. Informe final, Universidad de los Andes, CEDE, Bogotá, 43p.

MORALES A. [1996] : Estructuración, captura y control de bases de datos longitudinales, Memorias del seminario de capacitación e investigación "Recolección y análisis de datos longitudinales", Bogotá 9-13 dec. 1996, Universidad Nacional de Colombia - ORSTOM - PRESTA, Bogotá pp. 59-70.

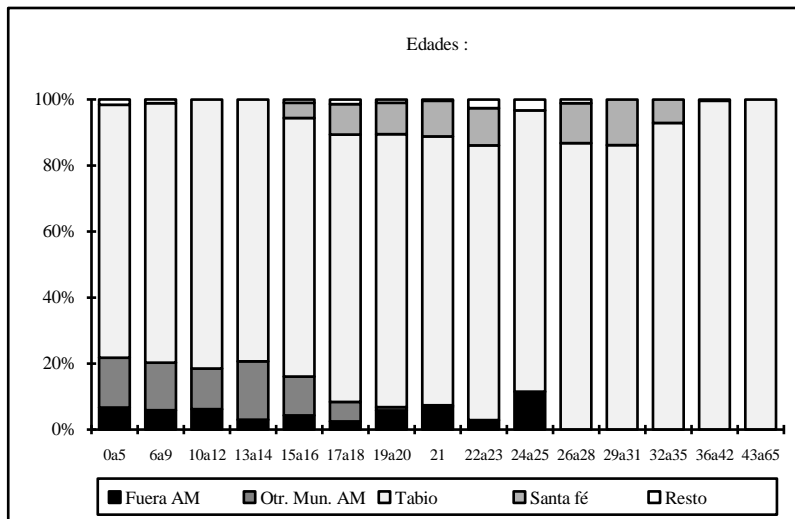
SAPORTA G. [1981] : Méthodes exploratoires d'analyse de données temporelles, in "Cahiers du bureau universitaire de recherche opérationnelle", n° 37-38, Université Pierre et Marie Curie, Paris, 194 p.

SAPORTA G. [1996] : L'analyse harmonique qualitative, une synthèse de la théorie, Memorias del seminario de capacitación e investigación "Recolección y análisis de datos longitudinales", Bogotá 9-13 dec. 1996, Universidad Nacional de Colombia - ORSTOM - PRESTA, Bogotá pp. 111-120.

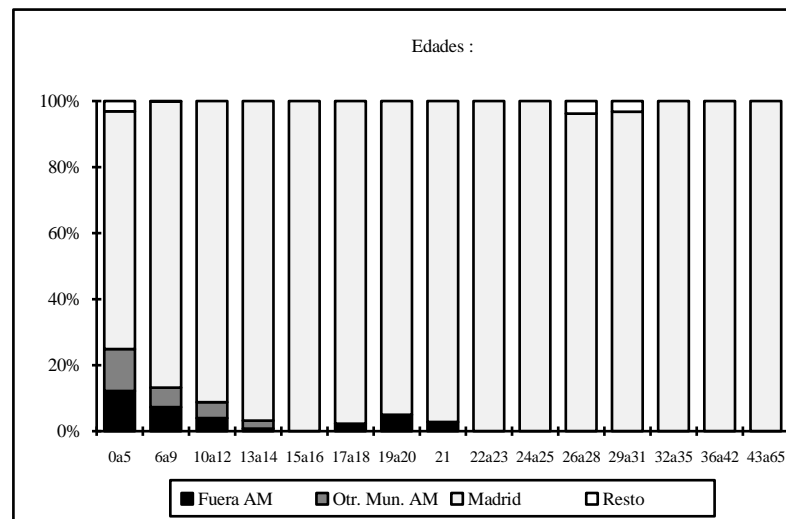
VAN DER HEIJDEN P. G. M. [1987] : Correspondence analysis of longitudinal categorical data, DSWO PRESS, Leiden.

**Anexo : Los grupos estables y estabilizados en los municipios periféricos (Tabio, Chía, Madrid)**

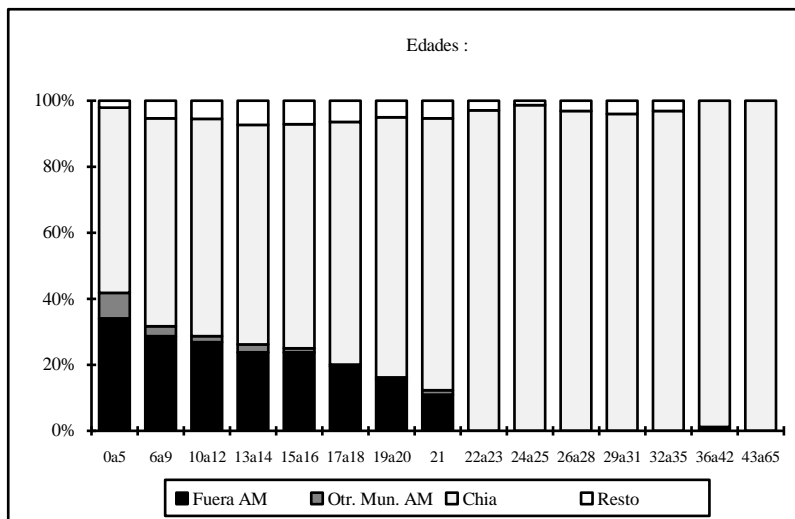
**Figura 8 : Clase 15, estable en Tabio**



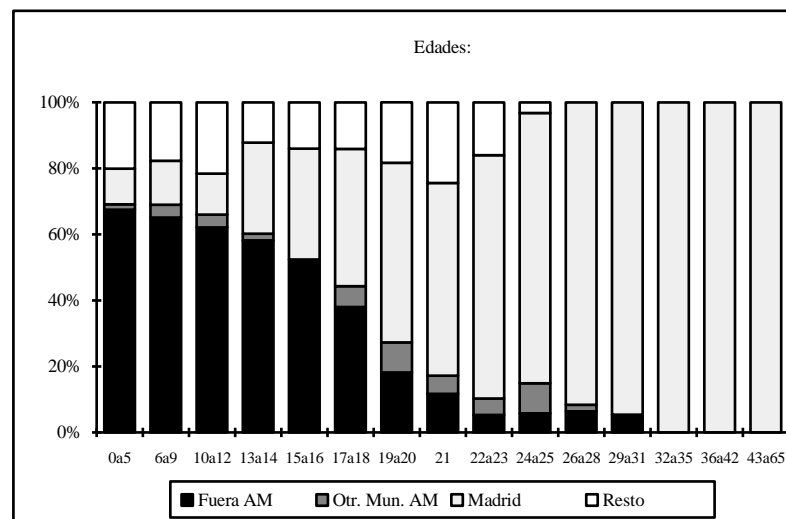
**Figura 10 : Clase 13, estable en Madrid**



**Figura 9 : Clase 14, estable en Chía**

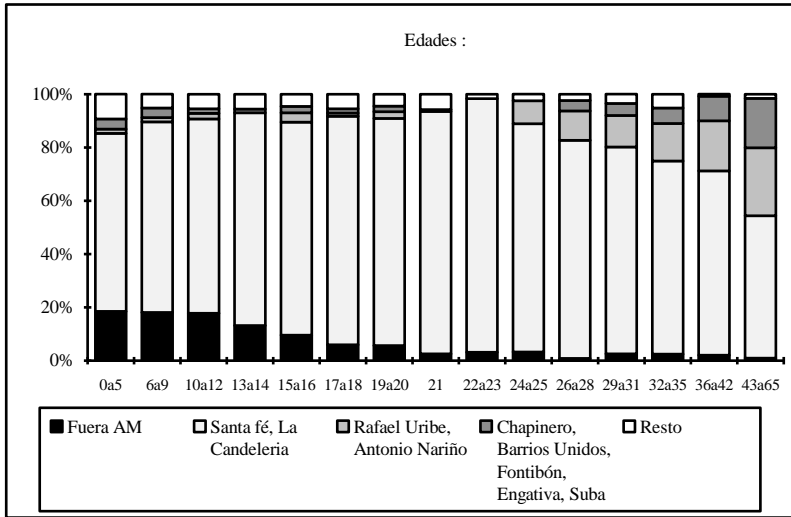


**Figura 11 : Clase 2, estabilizada en Madrid**

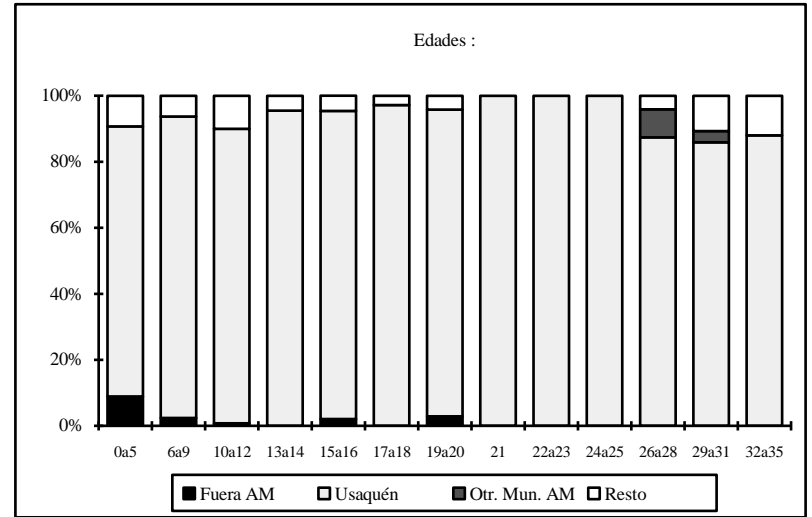


**Anexo : Los grupos estables y estabilizados en Bogotá**

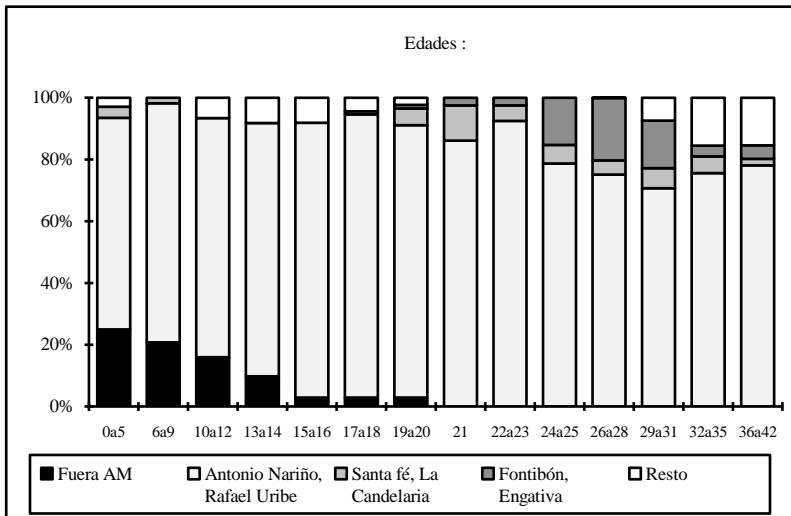
**Figura 12 : Clase 12, estable en Santa Fé y La Candelaria**



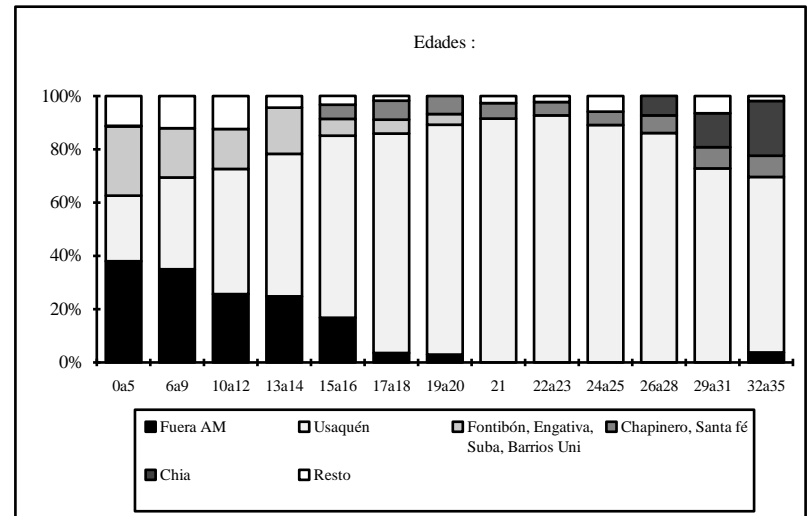
**Figura 14 : Clase 10, estable en Usaquéen**



**Figura 13 : Clase 11, estable en Antonio Nariño y Rafael Uribe**

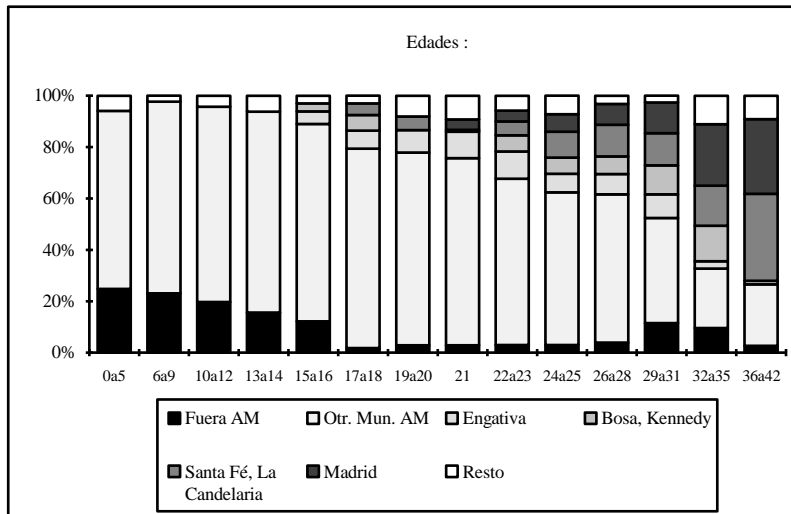


**Figura 15 : Clase 9, estabilizada en Usaquéen**

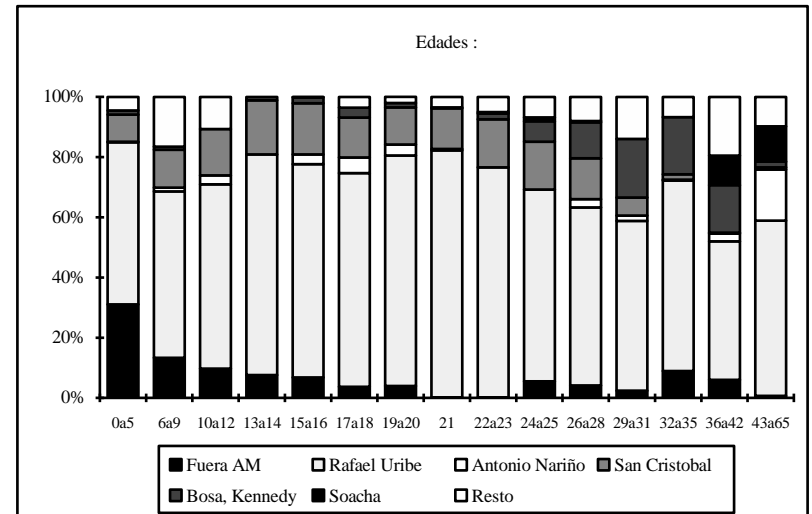


**Anexo : Los grupos con movilidad intra urbana media**

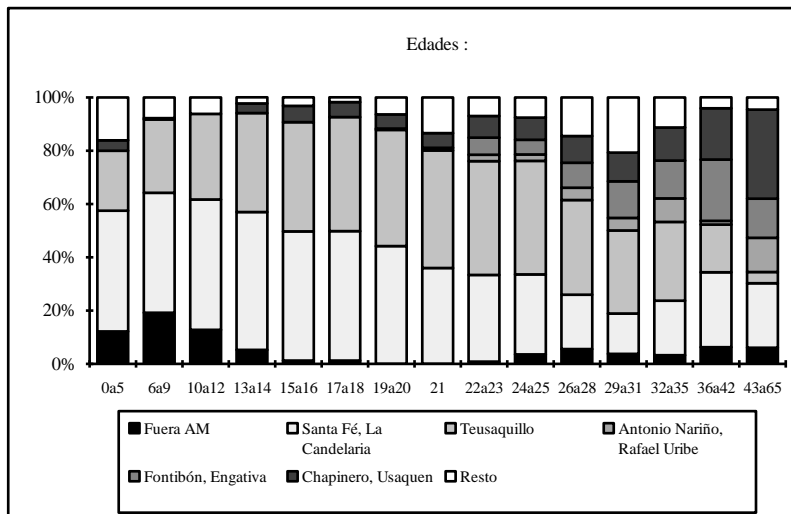
**Figura 16 : Clase 8, estadías en otros municipios del a.m.**



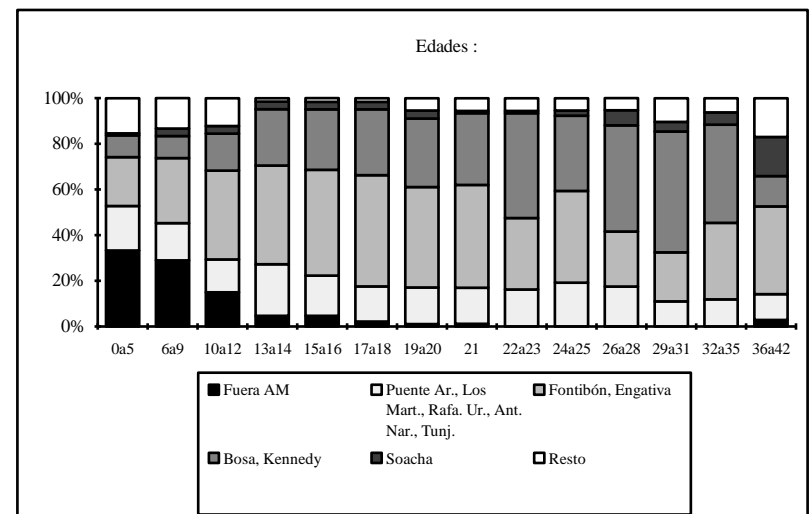
**Figura 18 : Clase 5, estadías en Rafael Uribe y San Cristóbal**



**Figura 17 : Clase 7, estadías en La Candelaria y Teusaquillo**



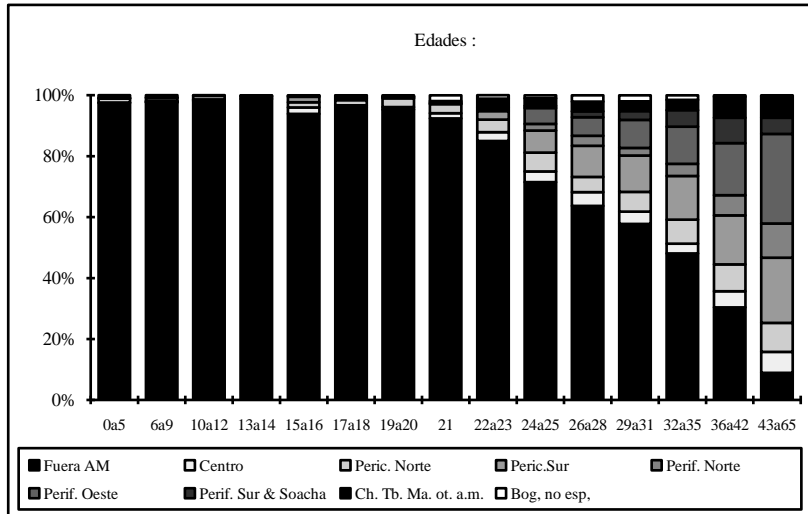
**Figura 19 : Clase 4, estadías en la periferia Oeste**



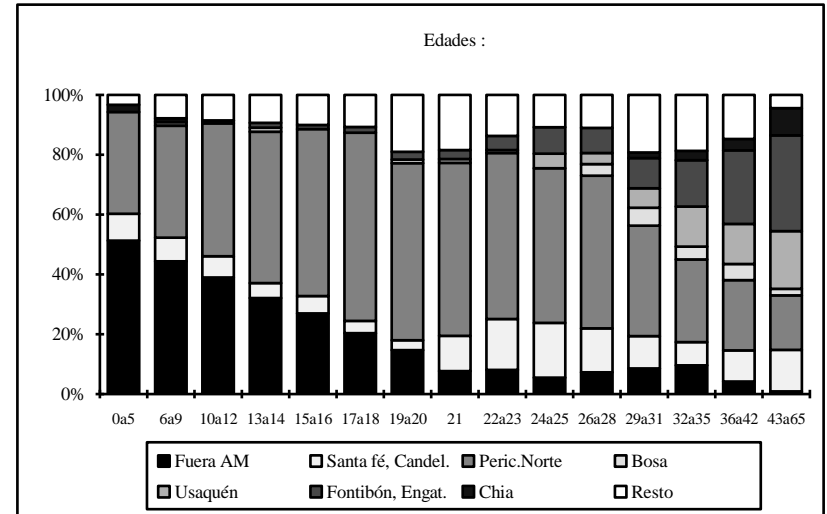


**Anexo : Los migrantes y los grupos con movilidad intra-urbana fuerte**

**Figura 20 : Clase 1, migrantes mayores llegados entre los 20 y 40 años de edad**



**Figura 21 : Clase 6, migrantes mayores llegados antes de los 20 años de edad**



**Figura 22 : Clase 3, migrantes jóvenes llegados entre los 10 y 25 años de edad**

