

L'inscription dans l'espace des biographies individuelles¹

Daniel Delaunay²

**Communication présentée à la XXIV^{ième} Conférence Générale de l'IUSSP
Salvador (Brésil), 18-24 août 2001**

¹ Texte provisoire, prière de ne pas citer sans l'autorisation de l'auteur. Je tiens à remercier Françoise Dureau qui m'a communiqué les données de l'enquête biographique de Bogota, assorties de sa connaissance des mobilités urbaines.

² Institut de Recherche pour le Développement, 213 rue La Fayette, Paris, Cedex 75480 ; mel : Daniel.Delaunay@bondy.ird.fr

Résumé :

L'analyse biographique ne tire pratiquement aucun parti des informations sur la localisation spatiale des événements, pourtant disponibles dans un grand nombre d'enquêtes. Elle contient pourtant la promesse d'une modélisation spatio-temporelle des comportements au niveau individuel. Cette absence est d'autant plus étonnante que les outils géomatiques et statistiques existent pour en tirer un parti analytique. Ce travail explore trois voies : a) la cartographie des itinéraires biographique à l'aide de la gestion du temps dans les systèmes d'information géographique ; b) le suivi dynamique des recompositions territoriales et leur modélisation multiniveau ; c) l'élaboration d'une analyse contextuelle biographique, selon trois niveaux d'agrégation (individu, quartier et contexte historique). Ces trois approches sont appliquées à la mobilité résidentielle des habitants de Bogotá.

Summary

Actually, the analysis of biographies ignore the information on the spatial localization of events, however available in a number of surveys. It nevertheless contains the promise of a space-time modeling of demographic behavior at the individual level. This absence is all the more astonishing since the GIS and statistical techniques exist to draw an analytical benefit from it. This work explores three approaches: a) the cartographic observation of life courses based on integration of time in geographical information systems; b) the dynamic follow-up of territorial compositions and their multilevel modeling; c) the development of a contextual analysis of hazard models, according to three levels of aggregation (individual, district and historical context). These three approaches are applied to the residential mobility in Bogotá.

Mots-clés : Démographie. Géographie. Modèles de durée. Analyse biographique. Analyse multiniveau. Systèmes d'information géographique. Modélisation spatio-temporelle. Recomposition spatiale. Mobilité résidentielle. Colombie. Bogota

Keywords: Demography. Geography. Event history models. Multilevel analysis. Geographical Information Systems. Space-time modeling. Space composition. Residential mobility. Colombia. Bogotá

L'analyse des pratiques spatiales dans le temps est une préoccupation assez tardive de la géographie. Elle débute véritablement dans les années soixante avec l'école de Lund³, fondée sur des concepts qui, aujourd'hui, résonnent étonnement avec la modélisation individuelle du temps biographique employée en démographie (Courgeau et Lelièvre 1989). Il est dès lors surprenant que l'emploi, désormais général, des modèles de durée en sciences humaines ne fasse pratiquement pas usage de l'information géo-référencée contenue dans bon nombre d'enquêtes biographiques. Une question a inspiré le présent travail : s'il existe une utilité à attendre de l'inscription dans l'espace des biographies, pourquoi cette dimension est-elle absente tant des méthodes que des problématiques des études longitudinales ? Une explication à cet oubli est peut-être à rechercher dans les réelles difficultés techniques rencontrées lors de la construction des modèles spatialisés de durée. Cette difficulté justifie de présenter quelques arguments qui relèvent de la méthode statistique ; d'autant qu'aujourd'hui de nouvelles méthodes élargissent le champ de la modélisation spatio-temporelle des comportements individuels.

A ce stade exploratoire, l'étude commencera par un rapide argumentaire en faveur de la prise en compte de la localisation dans l'analyse biographique. Elle envisagera ensuite des pistes de recherche, trois au total, choisies pour illustrer ces promesses comme les obstacles rencontrés. L'accent ne sera pas mis sur l'analyse du cas choisi — la mobilité résidentielle à Bogota⁴ — mais plutôt sur des solutions méthodologiques empruntées aux Systèmes d'Information Géographique et à l'analyse multiniveau. Bien évidemment, ces propositions ne couvrent pas l'ensemble des traitements possibles, ou simplement évoqués.

Propositions pour l'analyse spatiale des biographies

Pourquoi s'intéresser à l'espace biographique ? Trois ensembles d'arguments peuvent être avancés selon l'ambition du projet, pragmatique, méthodologique ou théorique. Si les applications proposées illustrent quelques-unes des questions théoriques soulevées, leur réalisation n'ira pas bien au-delà des deux premières aspirations.

La présente recherche se justifierait par le seul dessein de tirer parti d'une référence spatiale tout simplement disponible dans bien des enquêtes biographiques, pour peu qu'elles enregistrent les trajectoires résidentielles. Dès lors, les états de l'individu sont localisables à la fois dans l'espace et dans le temps, dessinant l'espace des vies matrimoniales, professionnelles, éventuellement sociales. Songeons également que des durées plus courtes, des espaces plus restreints sont susceptibles d'être décrits sur cette double échelle : on songe aux mobilités quotidiennes de loisir et de travail, les trajets urbains de transports, les fréquentations familiales ou sociales. D'emblée, sans effort statistique particulier, des espaces parcourus seront observables dans la durée, leur comparaison pour des sous-populations choisies deviendra possible. La référence géographique aidera l'analyse en lui fournissant des mesures des distances couvertes, des indicateurs de la configuration des réseaux tissés, une information complémentaire sur les territoires traversés. La première approche proposée utilise cer-

³ Proceeding of the IGU Symposium in urban geography, Lund, 1960.

⁴ Encuesta Movilidad espacial en el área metropolitana de Bogotá - Octubre 1993 - CEDE-ORSTOM

taines fonctionnalités des Systèmes d'Information Géographique (SIG) pour servir la cartographie des itinéraires et leur observation dynamique (Fotheringham, Wegener et al. 2000), (CASSINI 1999).

Ensuite, l'avantage méthodologique à retirer de l'inscription spatiale des événements biographiques se situe dans la continuité de l'effort de désagrégation d'une analyse démographique qui s'est longtemps arrêtée à la personne indivisible. Cette barrière a été levée par les modèles de durée, pour lesquels l'atome de l'analyse n'est plus l'individu mais le temps passé dans un état⁵. Avec la prise en compte du lieu, une nouvelle désagrégation est obtenue : le temps passé dans une même entité spatiale et un état inchangé. Le gain en finesse de l'observation (plusieurs états peuvent être combinés) ouvre, pour le moins, deux voies analytiques nouvelles. Tout d'abord, la modélisation spatio-temporelle des comportements individuels, et non plus seulement pour les unités spatiales qui sont des agrégats d'individus. Une application immédiate, à être illustrée dans le présent travail, sera d'associer à l'analyse biographique des informations contextuelles relatives aux lieux fréquentés, ou encore des mesures de distance, de proximité. Cela se justifie, par exemple, si l'on suppose que le même espace partagé contribue à une certaine corrélation entre les durées de séjour dans un état particulier. Une autre perspective méthodologique est de regrouper les événements individuels par unité de lieu et donc de suivre, en temps réel, la recomposition de l'espace démographique. Sans s'inquiéter pour l'instant du détail statistique exigé, voyons-y le moyen de comprendre comment les calendriers vitaux remodelent les lieux ; en quoi ils changent les distributions spatiales de populations singulières : célibataires, chômeurs, immigrés... Une telle application est particulièrement intéressante pour les recompositions spatiales rapides, invisibles aux recensements décennaux, tels les espaces urbains modifiés par les mobilités résidentielles ou quotidiennes.

Les solutions méthodologiques au traitement de l'information spatio-temporelle individuelle — que les registres de population détaillent avec une grande précision —, ouvrent la voie à l'exploration statistique de théories familières, comme de questionnements renouvelés. Le démographe qui inscrit l'analyse biographique dans l'espace se rapproche des options théoriques de l'école suédoise de Lund (Hägerstrand, Carlestam et al. 1991), il y trouvera une mine de questionnements et d'outils théoriques (les budgets espace-temps, par exemple). Ces problématiques sont toujours d'actualité avec le développement d'une analyse contextuelle, multiniveau des données longitudinales, mais aussi pour suivre en temps réel les dynamiques spatiales.

Trois pistes de recherche seront proposées. La première est une représentation cartographique des itinéraires résidentiels parcourus par des habitants de Bogota. Il s'agit d'une cartographie dynamique, réalisée grâce à l'intégration dans un SIG de géo-événements, soit des événements datés et localisés. Chaque géo-événement, ici le déplacement résidentiel, est assorti des attributs enregistrés par l'enquête ; lesquels sont employés pour sélectionner les parcours et construire des cartes thématiques de cette mobilité. La base de données géo-référencées ainsi constituée facilite la cons-

⁵ L'individu reste l'atome de l'analyse quand on ne retient qu'un seul événement par personne, ce qui est malheureusement une situation encore fréquente en démographie mais qui ne correspond pas à la réalité des événements renouvelables observés dans trajectoires matrimoniales, reproductives, migratoires, professionnelles, soit la grande majorité des événements

truction d'indicateurs spatiaux qui serviront l'analyse statistique et testeront les hypothèses visuelles de l'observation. La seconde approche explorée est une reconstruction de l'espace vécu à partir des trajectoires biographiques. L'objectif est d'évaluer, en temps réel, le peuplement des territoires par des personnes en mouvement. C'est un peuplement « qualifié » qui est reconstitué puisque l'on connaît les caractéristiques changeantes des occupants grâce au suivi biographique. Une méthode multiniveau sera proposée pour modéliser les évolutions spatio-temporelles territoriales et, conjointement, pour consolider l'imprécision qui résulte généralement d'échantillons insuffisants pour un découpage spatial fin. Le troisième essai de cette recherche est une analyse contextuelle biographique, laquelle fait également appel à des modèles multiniveaux, mais cette fois-ci de durée ou biographiques. Les enjeux analytiques sont considérables, pas seulement pour la démographie, car ils concernent un éventail de contextes qui ne se limite pas à l'environnement spatial. La méthode proposée s'appuie sur la mesure de l'hétérogénéité inobservée, signe statistique d'effets contextuels, mais aussi outil indispensable de l'analyse biographique en sciences humaine- quand la détermination des modèles ne peut atteindre la complétude familière aux sciences expérimentales.

Ces méthodes seront rapidement décrites pour introduire les exemples, mais auparavant quelques précautions sémantiques sont utiles pour éviter de faciles confusions. Le terme de *trajectoire* sera conservé pour désigner le calendrier des changements d'un état particulier. Une trajectoire se composera fréquemment, mais pas toujours, de plusieurs événements ce qui complique un peu les modèles de durée⁶ (Cleves 1999). On emploie quelquefois le terme d'*itinéraire* pour désigner les trajectoires biographiques. Il sera ici réservé aux déplacements dans l'espace, mais en rapport avec un état biographique particulier. Un itinéraire professionnel, par exemple, suivra les lieux des changements de ce statut particulier ; un itinéraire quotidien sera établi selon l'activité ludique ou professionnelle de la personne. L'agrégation ou la répétition des itinéraires conforme un *réseau*, dont la nature est éminemment spatiale — ils peuvent être décrits et paramétrés par les outils de la théorie des graphes — mais se distinguent des territoires par leur morphologie et les formes de leur organisation. Les diverses échelles traitées, ainsi que le projet d'entreprendre une analyse contextuelle, nous amènera à traiter une information hiérarchique, qui réunit plusieurs niveaux d'agrégation. Biologiquement l'individu est indivisible et, implicitement, l'analyse démographique le conçoit comme l'unité statistique élémentaire. Du point de vue de l'analyse statistique entreprise, cependant, l'individu est divisé dès que l'on décompose son calendrier et ses déplacements dans l'espace. Dans la hiérarchie multiniveau alors constituée, l'unité d'observation élémentaire est une fraction d'espace et de temps ; soit la durée dans un état biographique en un lieu enregistré. Cela a une conséquence analytique et heuristique importante, c'est de faire de l'individu un contexte de ses calendriers comme de ses déplacements. La non-prise en compte de cette éventuelle cohérence biographique d'une personne risque, lors de l'interprétation, de la reporter sur des contextes de niveau supérieur. Contrairement aux modèles linéaires, cette même négligence produira des estimations à la fois biaisées et inconsistantes pour les modèles de durée (Trussel et Rodríguez 1990). L'analyse multiniveau

humains qu'observe le démographe.

⁶ Plusieurs événements de même nature, parce qu'ils sont le fait d'un même individu, risquent de présenter une certaine similitude : les caractéristiques stables de la personne contribuant à leur ressemblance. Des options doivent être prises pour la mesure du temps, des précautions sont nécessaires pour produire des estimateurs robustes des prédictions.

incitera à distinguer la mobilité de la migration, si l'on admet que la notion de mobilité est plus focalisée sur la capacité individuelle aux déplacements et leur cohérence biographique ; alors que la mesure de la migration se place en un point de l'espace, lieux d'arrivée, de départ ou de passage. La terminologie de l'analyse multiniveau mériterait d'être uniformisée pour dissiper la confusion qui résulte de la multiplication des formulations et modèles élaborés par des disciplines différentes. Pour ne s'en tenir qu'aux modèles de durée ou biographiques, rappelons que le terme d'hétérogénéité inobservée des unités d'observation est le signe d'une corrélation entre les durées et donc en fait d'une similitude — une homogénéité interne — qui signale des facteurs contextuels communs. Cette hétérogénéité inobservée est désignée par le terme de *frailty* (la fragilité sous-jacente de l'individu) dans les modèles de durée bio-démographiques (Aalen 1988; Hougaard 1995; Keiding, Andersen et al. 1997).

Cartographie et topologie des itinéraires biographiques

La localisation spatiale des événements biographiques répétés compose autant d'itinéraires que d'états recensés : qu'il soit matrimonial, professionnel, familial... La curiosité qu'ils éveillent tient à ce que leur géographie change dans le temps, appelant une cartographie dynamique, c'est-à-dire comportant une échelle temporelle. Malheureusement, parce qu'elle s'applique à chaque individu, la cartographie exploratoire des itinéraires devient d'une désarmante complexité qui risque d'en brouiller la perception. Aux deux échelles spatiales et temporelles s'ajoutent les attributs, ou caractéristiques, de chaque géo-événement, également candidats à la représentation cartographique. Un outil géomatique flexible est donc requis, que propose la présente section.

Il existe un autre argument en faveur de l'intégration des itinéraires biographiques dans une base de données géo-référencées qui est d'accéder, par le biais de la localisation, à toute l'information disponible sur le lieu, si elle est contenue dans la base (de Jong et Hoheus 1997). Conjointement, la description topologique des itinéraires, facilitée par des outils géomatiques, apporte une dimension analytique supplémentaire à l'examen de leur calendrier ; des indicateurs de distance, de configuration spatiale devenant disponibles pour la modélisation des temps de séjour. Faute de place, et de questionnement approprié sur la mobilité résidentielle à Bogota⁷, cet aspect ne sera qu'allusivement abordé dans le présent article. Deux aspects de cette intégration méritent une remarque. Une première dimension peut être qualifiée de réticulaire puisque les déplacements composent des réseaux orientés, avec des arcs, des nœuds, des connections. On dispose des concepts et instruments forgés par la théorie des graphes pour les décrire mais l'intérêt de cet examen est plus manifeste pour les parcours répétés, ou les cycles de déplacements : navettes de travail ou de loisirs, migrations tournantes... Il est néanmoins concevable d'appliquer cet outil, ainsi que d'autres traitements statistiques multivariés, à des regroupements d'itinéraires ou d'individus. Une autre dimension est territoriale quand les séjours en un lieu et un état durent plus que le déplacement ; l'espace séjourné se distinguant de l'espace parcouru. L'analyse sera amenée à privilégier cette dimension

⁷ Le calcul des indicateurs spatiaux répondra à la problématique du thème étudié : on s'intéressera, par exemple, à l'endogamie territoriale des unions, la distance des mobilités spatiales professionnelles, ou l'étendue et la vitesse des déplacements quotidiens.

pour un itinéraire unique, le parcours résidentiel ou professionnel de l'individu par exemple, quand la durée retient l'individu en un lieu qu'il transforme et dont il subit les influences. Un lieu de séjour appartient à un territoire délimité — un quartier résidentiel, une aire matrimoniale, un bassin d'emploi — auquel on peut prêter une certaine homogénéité, un mode d'organisation ou des perceptions singulières de la part des occupants, pour le moins des traits distinctifs. L'analyse contextuelle biographique (Cf. infra) utilisera avec profit ce type d'information issue des SIG ; elle documentera éventuellement la simulation multi-agent du choix par les individus de leurs itinéraires.

Revenons à ce qui, vraisemblablement, constituera le premier pas d'une observation des itinéraires démographiques : leur représentation cartographique. Les fonctionnalités de certains SIG s'appliquent à des géo-événements, soit des objets qui, au même titre que les autres entités spatiales sont décrits par des attributs (ESRI 1998). La localisation de ces événements — ici les changements résidentiels des habitants de Bogota — de même que leurs caractéristiques — âge au déplacement, sexe et état civil de l'individu — produisent autant de cartes que de dates ou de périodes d'observation, les deux pouvant être combinées. Les deux cartes qui suivent retiennent tous les déplacements résidentiels des générations nées entre 1930 et 1934. Il ne s'agit pas d'une cartographie dynamique puisque toutes les dates sont confondues, seulement distinguées par des symboles. L'étape de la migration est également signalée. Une carte englobe la totalité de la Colombie pour faire apparaître l'origine des nouveaux citoyens, une autre se concentre sur l'aire métropolitaine où, en 1993, se retrouvent la cinquantaine de personnes appartenant à cette cohorte. Ces deux cartes identiques, mais à la couverture différente, mettent en évidence des informations difficiles à discerner dans les chiffres bruts : les origines provinciales des migrants, les choix de localisation dans la ville (l'axe central est privilégié, de même que la proximité de la région d'origine), le calendrier tardif de l'arrivée dans la capitale, l'étendue des déplacements intra-urbains, etc.

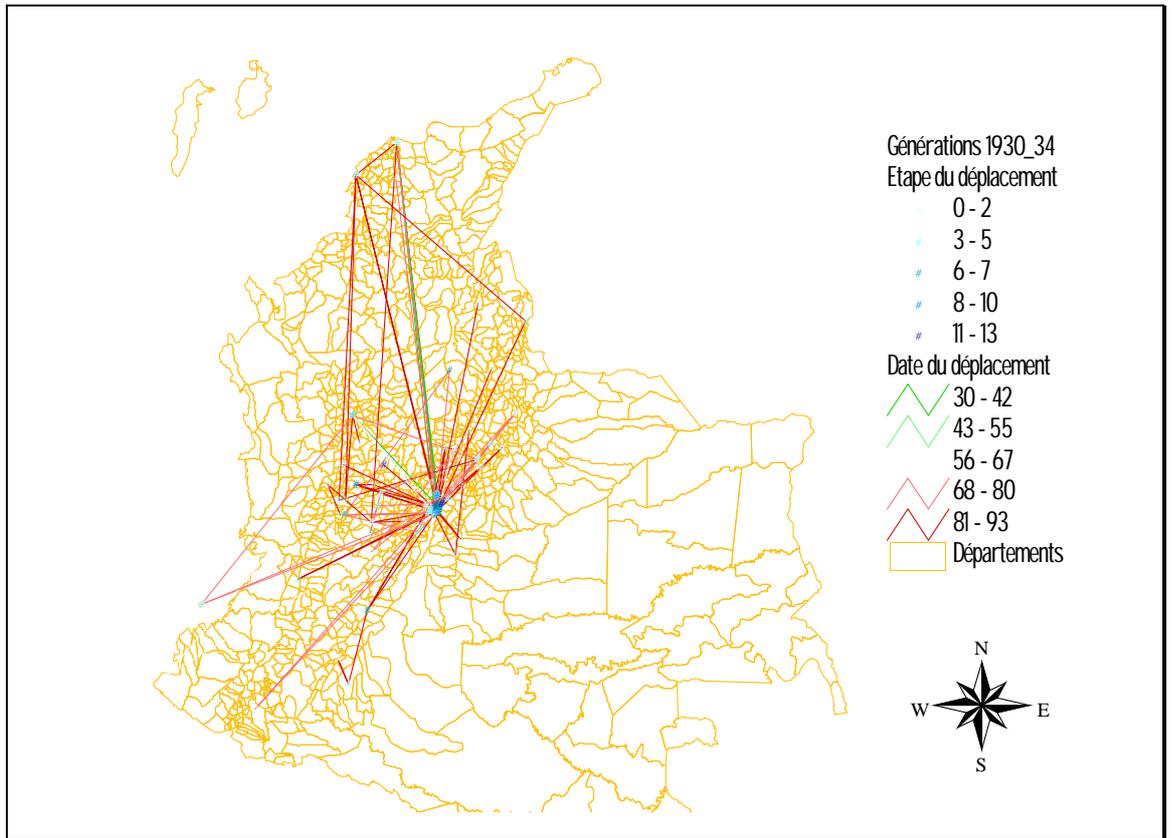


Figure 1 Itinéraires des personnes nées entre 1930 et 1934

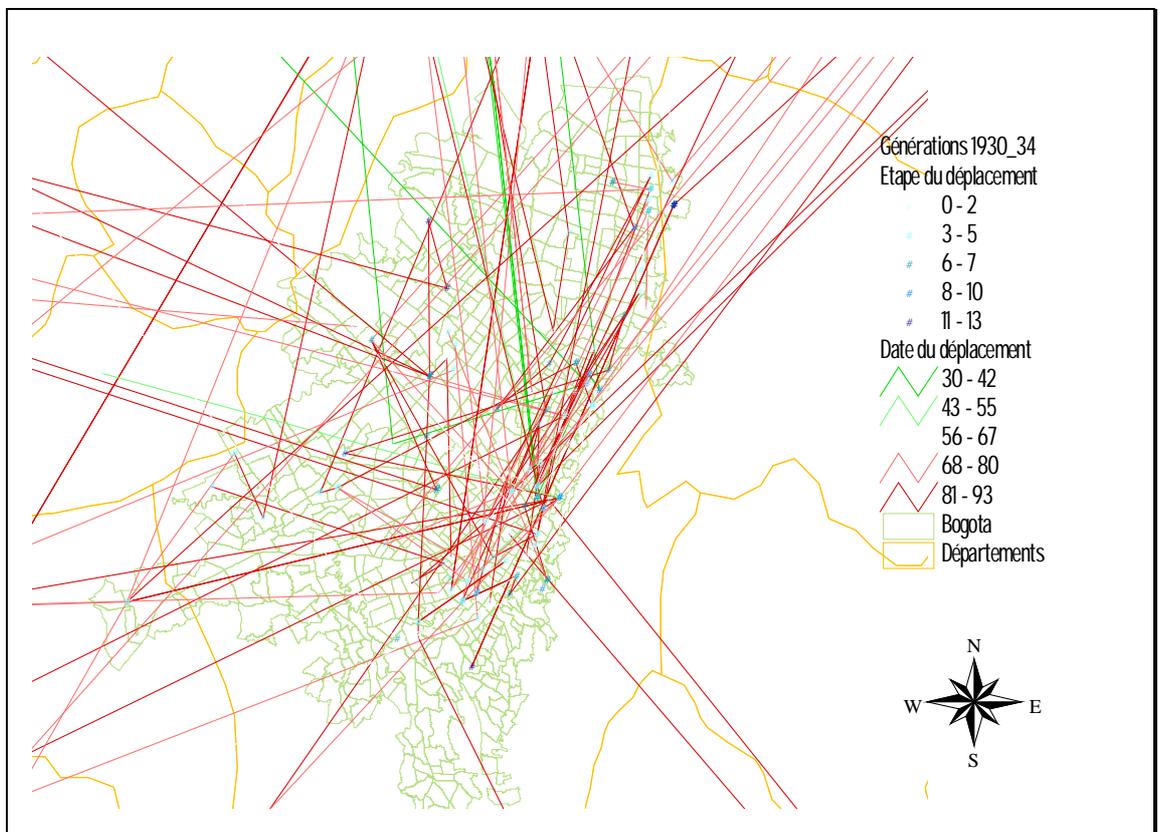


Figure 2 Les mêmes itinéraires pour l'aire métropolitaine de Bogota

Potentiellement, l'intérêt de cette cartographie automatisée réside dans la comparaison des itinéraires pour des groupes bien identifiés ; on est en effet restreint par la capacité de l'œil à n'interpréter qu'un nombre limité de graphes représentés. Confronter les espaces de vie de plusieurs générations, de divers groupes ethniques ou vagues d'immigrants (pour relever les ségrégations spatiales vécues, par exemple), de différentes catégories socioprofessionnelles demande seulement un effort d'imagination puis d'analyse, tant les combinaisons et les choix sont étendus. Mais, sans conteste, la valeur ajoutée de la cartographie biographique est la décomposition du temps. Et cela, conformément à deux paramètres : la périodicité des dates d'observation et l'étendue de la période couverte. La séquence historique représentée dans la Figure 3 donne la succession des mobilités résidentielles selon le sexe et le niveau d'éducation pour des périodes décennales : ce sont tous les mouvements de la décennie qui sont accumulés à des moments réguliers. Un trait orienté rejoint deux événements dans l'intervalle de temps pris en compte. La forte différenciation spatio-temporelle des mobilités n'est pas interprétable dans les mêmes termes qu'une analyse statistique qui serait orientée vers les variables. Telle qu'elle est présentée, il s'agit bel et bien d'une lecture monographique d'un regroupement de cas singuliers. Cette lecture spatio-temporelle globale peut guider l'analyse statistique, la compléter.

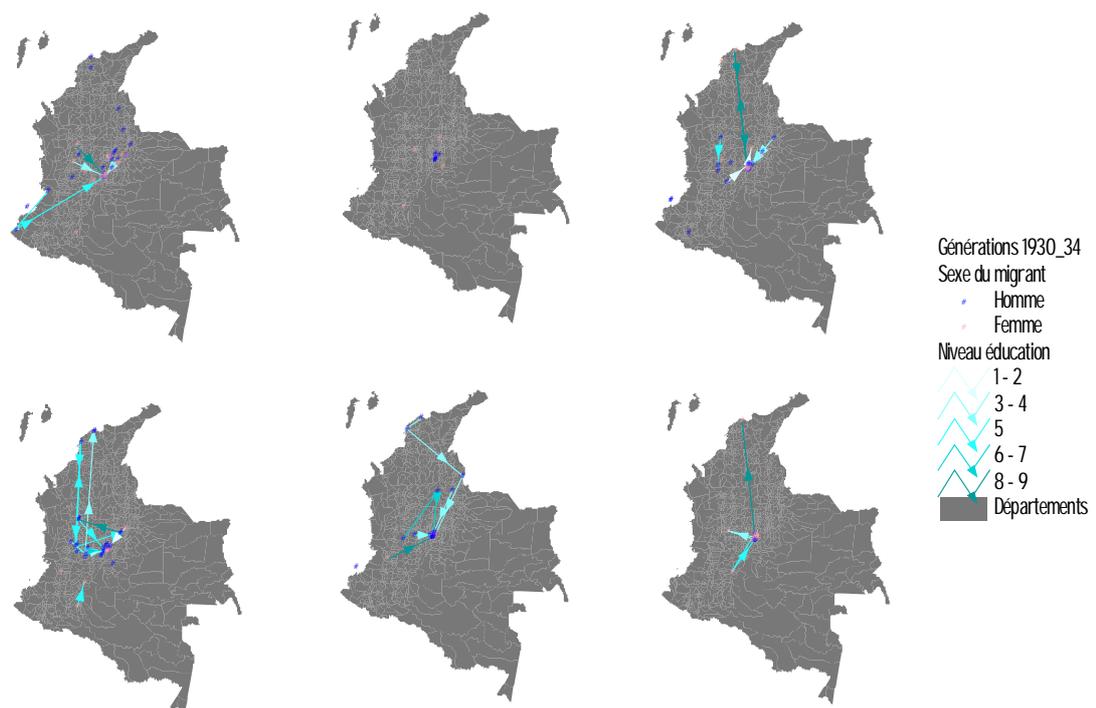


Figure 3 Evolution des mobilités résidentielles par tranches de dix ans (génération 1930-34)

Mais qu'importe l'approche envisagée, qualitative (étude de cas) ou quantitative (sur les variables), puisque chaque géo-événement conserve, dans la base de données, toutes ses caractéristiques d'origine. Il en acquiert d'autres, géographiques cette fois, par l'intégration dans le SIG : appartenance aux unités des couches d'information superposées (image satellitaire, inventaires naturels ou

centitaires, réseaux de transport...), appartenance fondée sur des distances fonctionnelles... Toutes sont accessibles par le biais d'un langage de programmation propre au SIG (ESRI 1996) et donc disponibles pour des traitements statistiques ultérieurs. Notons que le géocodage de chaque événement doit rester cohérent avec la structure hiérarchique de l'information localisée du SIG, afin de se ménager toutes les possibilités de l'analyse multiniveau (Delaunay 1997).

L'analyse dynamique des compositions territoriales

Un second bénéfice, plus proprement géographique, à retirer de l'analyse biographique localisée est le suivi, en temps réel, de la composition des territoires. Connaissant pour chaque individu ses états, à un moment et en un lieu, le décompte périodique des populations locales est immédiat. Par cet aménagement de l'information biographique, on passe d'une statistique individuelle de flux — l'enregistrement des passages d'un état/lieu à l'autre — à une mesure spatio-temporelle de stocks rapportés à des unités spatiales. Cela documente une connaissance dynamique du peuplement territorial, un peuplement « qualifié » puisque l'on en connaît, à tout instant, les caractéristiques. C'est la précision dans le temps qui est ici précieuse puisque les recompositions démographiques du territoire sont connues à l'événement individuel près. On suivra, année par année, la diffusion spatiale de l'état de santé d'une population ou d'une génération, la dispersion géographique d'une vague d'immigrants, la configuration spatiale de l'emploi, des unions... La plus grande limitation de cet exercice est que l'information biographique n'est généralement connue qu'au moyen d'enquêtes rétrospectives portant sur quelques milliers d'observations ; elles ne procurent pas la finesse spatiale des recensements, beaucoup s'en faut. Ceux-ci, en effet, compensent leur grossière périodicité par un détail spatial exhaustif⁸. Les qualités respectives des deux sources deviennent complémentaires : au reste de la population, on pourra comparer la contribution territoriale du groupe dont on suit les trajectoires, donnant les éléments d'une étude des discriminations induites par la mobilité résidentielle, par exemple. Naturellement, l'utilité géographique des biographies sera plutôt à attendre dans les domaines délaissés par les recensements, notamment pour les mobilités courtes, résidentielles ou quotidiennes. Le « peuplement » urbain change au cours d'une journée de déplacements professionnels ou de loisirs ; les mobilités résidentielles risquent d'infléchir rapidement la demande de logement de certains quartiers.

L'exemple de la mobilité résidentielle de Bogota

L'illustration proposée va dans le sens de cette remarque : il s'agit d'une description du peuplement de Bogota à partir de la même enquête biographique, dont on ne retiendra que les changements de résidence. L'objectif est d'abord descriptif : montrer comment s'est dispersée telle génération, ou la population de tel quartier, au cours de ses parcours résidentiels. Tous les déménagements sont examinés, incluant le premier à la naissance de l'individu. Pour chaque quartier et chaque année, on retient la présence des individus selon le statut du logement qu'ils occupent. Par convention, le changement de résidence une année donnée compte pour une présence cette année-là.

Pour illustrer cette reconstitution du peuplement à l'aide des biographies, suivons de 1950

à 1993, l'évolution d'un indice : le rapport entre les proportions de propriétaires du quartier et la même mesure pour l'ensemble de la ville, selon l'enquête. La valeur « 1 » sur l'axe des ordonnées traduit l'égalité des proportions (le quartier ne se distingue pas de tout Bogota), les valeurs inférieures à un une moindre présence de logements en propriété, le contraire pour les valeurs supérieures à l'unité. Cet indice a été retenu pour tenir compte de l'évolution de l'accès à la propriété au cours de la vie de toutes les personnes interrogées. Cette simple mesure souffre d'une restriction due à la nature de l'échantillon qui ne prétend pas représenter l'ensemble de l'aire métropolitaine (Dureau, Florez et al. 1994). Pour cette raison, l'interprétation doit préalablement bien identifier les populations et les phénomènes que l'on observe. Quatre arrondissements⁹ ont été retenus à titre d'exemple. Les graphiques de la Figure 4 doivent être interprétés de la manière suivante : à la date de 1960, par exemple, les informateurs présents dans l'arrondissement de Usaquen ont un accès à la propriété identique (indice = 1.02) à l'ensemble de l'échantillon dispersé dans le reste de la ville (59% contre 58%). Délaissions l'analyse de ces cas particuliers pour relever deux remarques générales. D'abord, les tendances longues sont clairement discernables : la stabilité dans Usaquen à partir de 1960, dans Teusaquillon dès les années soixante-dix, la hausse depuis 1950 à Fontibón, à la baisse constante à Tunjuelito. Deuxièmement, des variations cycliques ne remettent pas radicalement en cause ces évolutions générales : pour l'ensemble des arrondissements, elles tendent à s'atténuer avec l'arrivée des sujets appartenant aux générations récentes. Ce qui suggère que la faiblesse des effectifs dans le passé en est la cause.

Reconnaissons que cet exemple n'échappe pas aux dangers auxquels nous expose l'information biographique. D'une part, elle perd en qualité avec le recul, introduisant une incertitude sur les calendriers et l'identification des lieux. Pour le géographe, elle présente un défaut plus grave encore : au fur et à mesure que l'on s'éloigne dans le passé, l'univers se réduit aux générations les plus anciennes interrogées au moment de l'enquête, soit une double sélection par la mort et la migration.

⁸ Sous réserve de restrictions conçues pour protéger la confidentialité de l'information censitaire.

⁹ L'échantillon en compte vingt-et-un au total, incluant le territoire colombien hors de l'aire métropolitaine, pour les migrations externes à la ville.

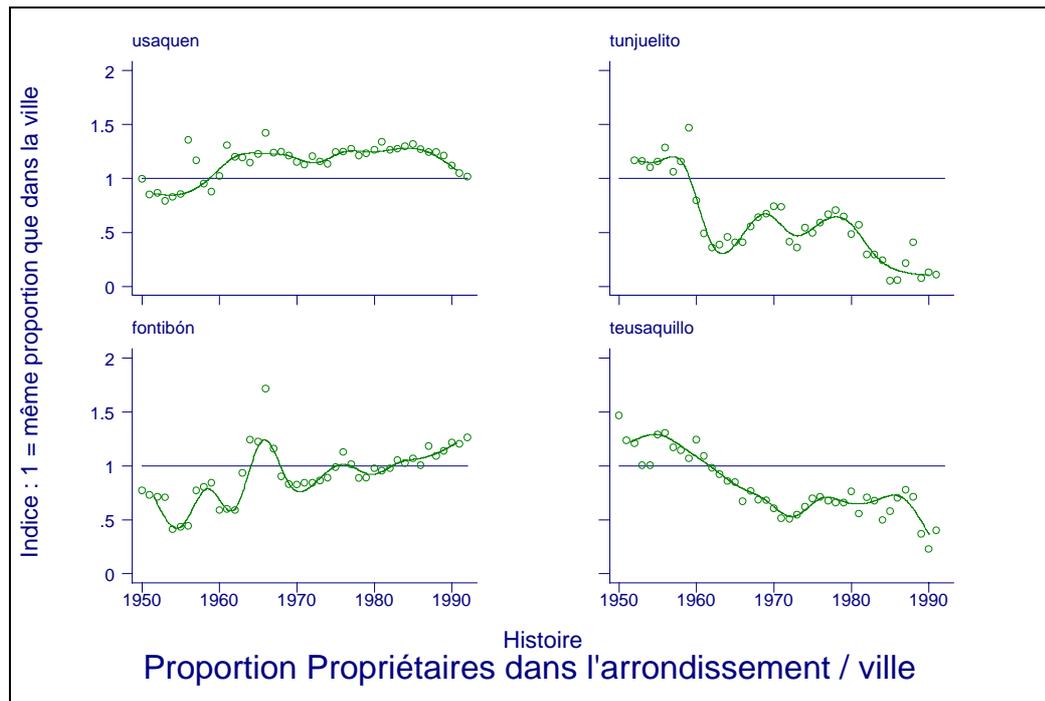


Figure 4 Evolution des compositions territoriales dans quatre arrondissements

Evolutions spatio-temporelles de la location de logements par les habitants du quartier de Bosa et Soacha en 1993

Au cours de leurs trajectoires résidentielles, les habitants des quartiers de Bosa et Soacha enquêtés en 1993 séjournèrent dans différents arrondissements de Bogota, et hors de la ville, dont ils composent une partie de la population. Supposons que l'on s'intéresse à leur situation immobilière, soit la proportion de logement en location pour l'exemple illustré par la Figure 5. Sans rentrer dans le détail d'une analyse qui n'a pas sa place ici, le suivi des compositions territoriales met conjointement en évidence des niveaux et des calendriers contrastés de la tenure du logement. En dehors de la ville, le pourcentage de locations est faible et constant dans le temps ; ces migrants sacrifient un statut de propriétaire. D'autres quartiers, de création récente (Usme est occupé à la fin des années soixante-dix, Tunjuelito une décennie plus tôt) ont, au contraire, été peuplés selon un mode locatif, avec une tendance à l'augmentation de ce type d'occupation dans le cas de Tunjuelito. L'intérêt du suivi temporel précis de ce statut met en évidence deux vagues dans le cas de Bosa : une première dès le début des années soixante où la propriété semble dominer (si l'on en croit la faible proportion de locataires), une seconde qui naît autour de 1980 quand la location prend subitement de l'importance (toujours pour ces habitants particuliers) pour décroître ensuite régulièrement.

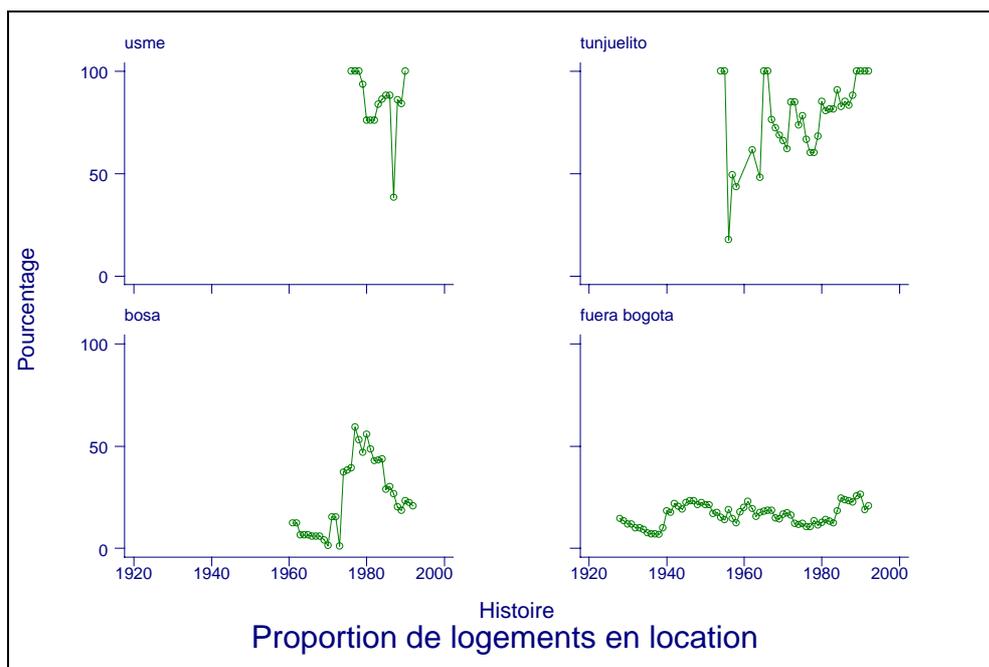


Figure 5 Proportion de propriétaires parmi les habitants de Bosa et Soacha en 1993

Des analyses similaires sont envisageables pour d'autres statuts (matrimoniaux, professionnels, familiaux), pour autant que leur configuration spatio-temporelle ait un sens. Si l'on se penche sur des populations très caractérisées, soit du point de vue ethnique ou migratoire, il sera aisé de suivre leur situation contextuelle dans le temps en les comparant aux autres occupants du lieu. En l'absence de recensements ou de registres, à condition de concevoir un échantillonnage adapté, l'analyse biographique spatialisée permet de retrouver l'histoire du peuplement.

Modéliser les recompositions territoriales : la méthode multiniveau

Quand diminuent la taille des unités spatiales et le nombre d'événements biographiques leur correspondant, les variations erratiques des mesures territoriales rendent difficile ou suspecte l'interprétation des évolutions. Des ajustements de tendance deviennent nécessaires, mais vite fastidieux pour un découpage fin de l'espace ; si l'on passe par exemple de la vingtaine d'arrondissements de la ville de Bogota aux 622 secteurs urbains. L'emploi de modèles mixtes, ou multiniveaux, apporte une solution commode et élégante à cette double difficulté, pour autant que l'on dispose de quelques observations par unités. Il suffit d'introduire deux composantes aléatoires simultanées : une pour les variations dans le temps au sein de chaque entité territoriale, une autre pour la différenciation entre les unités spatiales. Les paramètres de chaque ajustement temporel sont introduits pour modéliser, au niveau supérieur, les variations spatiales, éventuellement grâce à l'introduction de caractéristiques du quartier ou de la période. Ce faisant, on lève une difficulté très présente dans les enquêtes biographiques aux échantillons réduits, puisque l'estimation des paramètres est réalisée sur l'ensemble des effectifs et non plus pour chaque unité territoriale. L'avantage pratique est qu'un grand nombre de tendances temporelles sont obtenues par une seule équation. Dans le cas de Bogota, la proportion de propriétaires est connue à chaque date (i) pour chaque quartier (j). La variation totale de cette proportion se décompose en trois parties : e_{ij} les écarts temporels

autour de la moyenne du lieu (j), u_{0j} les variations de l'intersection de chaque district à l'origine du temps mesuré et u_{1j} la variation de la pente des ajustements linéaires.

$$\begin{aligned}
 F^{\circ} \text{ Propriétaires}_{ij} &\sim N(\lambda E, \Omega) \\
 F^{\circ} \text{ Propriétaires}_{ij} &= \beta_{0j} \text{cons} + \beta_{1j} \text{Date}_{ij} \\
 \beta_{0j} &= \beta_0 + u_{0j} + e_{0j} \\
 \beta_{1j} &= \beta_1 + u_{1j}
 \end{aligned}$$

$$\begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} \sim N(0, \Omega_u) : \Omega_u = \begin{bmatrix} \sigma_{u0}^2 & \\ & \sigma_{u1}^2 \end{bmatrix}$$

$$\begin{bmatrix} e_{0j} \end{bmatrix} \sim N(0, \Omega_e) : \Omega_e = \begin{bmatrix} \sigma_{e0}^2 \end{bmatrix}$$

$-2 * \log \text{likelihood(TOLS)} = 16788.920$ (1988 of 1988 cases in use)

Pour illustrer la procédure, voici l'ajustement des proportions de propriétaires pour cinq arrondissements de la ville : dans ce cas, les mesures reconstituées sont assez nombreuses pour s'assurer d'une tendance linéaire¹⁰ dégagée du bruit statistique, due rappelons-le aux faibles effectifs de l'enquête. Cette variation parasite n'a pas d'intérêt analytique et seule cette pente mérite d'être « expliquée » par un jeu de caractéristiques spatiales du quartier.

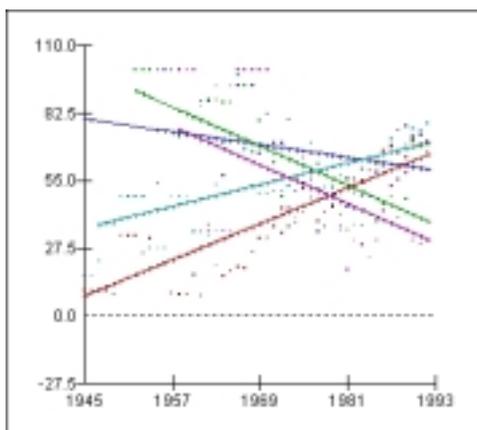


Figure 6 Ajustements linéaires multiniveau pour six arrondissements

On s'en doute, il en sera tout autrement pour les quelques six cents secteurs urbains de la métropole, mal décrits par les biographies d'un millier de personnes (d'autant qu'un peu moins de la moitié des déplacements se situent hors de l'aire métropolitaine). Ici encore l'ajustement est immédiat pour les secteurs qui disposent d'au moins deux mesures temporelles depuis 1950 ; certains hélas portant sur des périodes très courtes. Il suffit de porter sur la carte des secteurs les variations de la pente (Longford 1993) pour construire la configuration spatiale des changements de statut. La méthode, très prometteuse quand une information adaptée est disponible, peut s'appliquer à toutes modélisations spatio-temporelles, surtout si d'autres recompositions territoriales sont calculées conjointement : évolution du statut matrimonial, de l'emploi, de la structure familiale...

¹⁰ Un ajustement polynomial est également possible

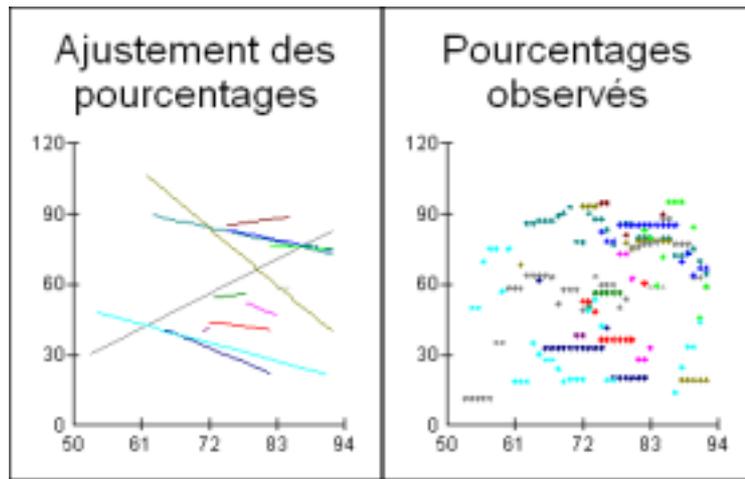


Figure 7 Ajustement selon les secteurs urbains (une sélection)

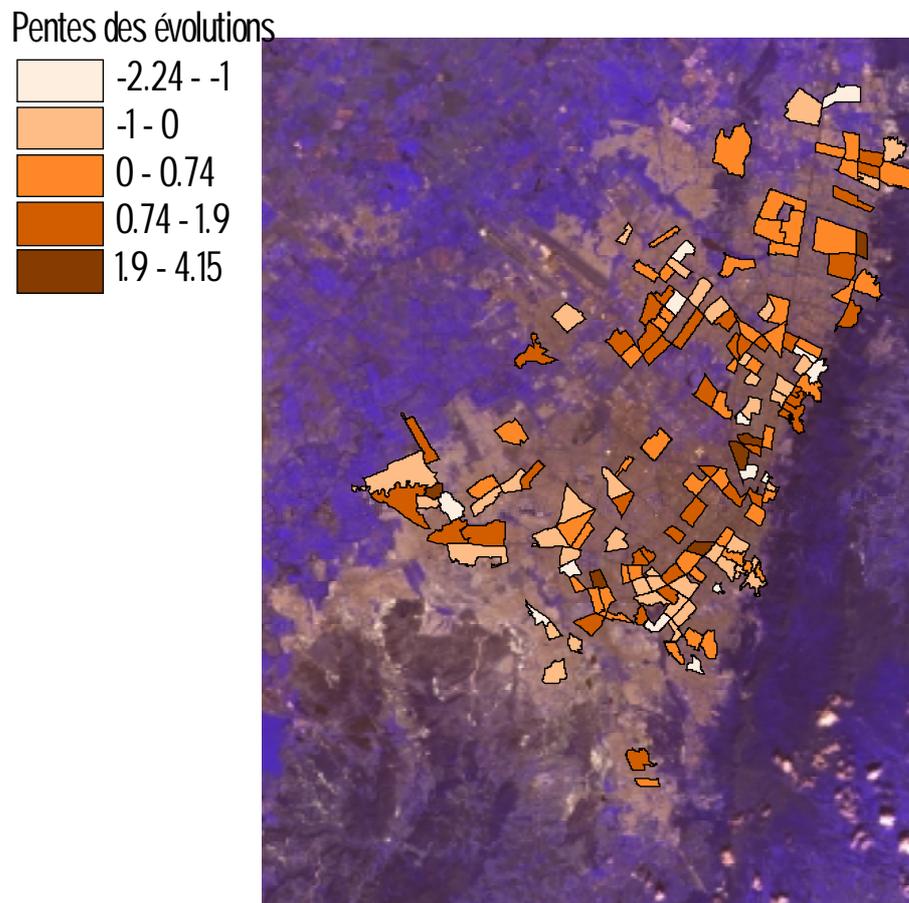


Figure 8 Variations annuelles de la proportion de propriétaires selon l'enquête biographique

Analyser l'influence des contextes spatiaux

Les restrictions de lieu et de temps obligent un individu à choisir une localisation à une date donnée pour chaque acte de la vie. Ce choix, inspiré par les ressources inégalement réparties dans l'espace (Hägerstrand 1973), se prête à des simulations (Duley et Rees 1991), que la présente étude n'abordera pas. Par ailleurs, connaissant les lieux qui accueillent les événements marquants de la vie d'un individu, il est donné d'observer les conditions locales de leur survenue et donc d'envisager une analyse contextuelle (Boyd et Iversen 1979; Jones et Duncan 1995). Une fois établi, l'individu va progressivement interpréter et modifier l'environnement qui l'accueille (il exerce un effet de composition) ; un environnement qui va à son tour infléchir le calendrier des événements démographiques (effet contextuel). Aucun événement vital n'y échappe¹¹, mais ce n'est qu'assez récemment que cette influence est envisagée dans les modèles de durée (Guo 1993; Brewster 1994; Sastry 1996; Courgeau 2000). Cela se réalise par l'introduction de variables stochastiques dans les fonctions de risque (Little, Schnabel et al. 2000) et l'identification de possibles corrélations entre les durées de séjour au sein d'une même communauté, fratrie ou région. (Xue et Ding 1999).

La première difficulté rencontrée, est qu'à ce niveau fin de désagrégation de l'espace-temps individuel, l'environnement spatial n'est qu'un contexte parmi d'autres. Il convient non seulement de l'isoler, mais également d'évaluer les autres pour éviter d'attribuer à l'un ce qui est redevable à l'autre. Pour les événements répétés (la plupart de ceux que la démographie observe), l'individu compose le premier contexte de sa biographie, qui associe diverses trajectoires¹². Depuis plus d'une décennie, les bio-statisticiens modélisent à l'aide d'une composante aléatoire — *frailty* — la corrélation suspectée des risques vitaux imputables à un patrimoine génétique ou un milieu partagé (Aalen 1988; Vaupel 1988). Enfin, l'instant vécu par tous, qui s'ajoute au cycle de vie individuel, crée d'autres conditions, historiques cette fois, susceptibles de précipiter ou ajourner la survenue de l'événement¹³. L'étude de cas proposée doit donc envisager ces trois contextes : l'individu, le quartier et l'histoire de la ville.

La conception même de l'influence contextuelle, notamment territoriale, tout autant que sa mesure, ont inspiré diverses approches et quelques réserves. Il sera probablement trop optimiste ou réducteur de supposer qu'un lieu exerce un effet uniformément réparti à l'intérieur des limites retenues¹⁴ ; il peut se partager aussi entre deux lieux éloignés¹⁵. Dans le temps, une relative inertie des influences contextuelles est prévisible, la mémoire comme la lenteur d'adaptation au lieu risquent d'atténuer l'effet du nouvel environnement, prolonger celui de l'ancien. Envisager une seule influence territoriale est incompatible avec la fréquentation individuelle de plusieurs espaces, dont

¹¹ Le nombre de mariables en présence dans l'aire matrimoniale peu modifier l'âge à la première union, l'incidence de l'offre de logements du quartier va accélérer ou ralentir la mobilité résidentielle, le marché du travail local va infléchir le calendrier de la carrière professionnelle ou de la migration, l'environnement médical et naturel a un impact sur la mortalité...

¹² Les caractères susceptibles d'expliquer cette influence contextuelle sont ceux en particulier qui ne changent pas au cours de sa vie, et qui sont donc des mesures répétées pour tous les événements qu'il connaîtra.

¹³ Ce sera une union (du fait d'une disposition légale, par exemple), un emploi (soumis à la conjoncture économique), un changement de résidence répondant au lotissement d'un espace urbain.

¹⁴ Un réseau, par exemple, a un impact anisotropique sur l'espace, pour le moins discontinu.

¹⁵ Un changement de résidence, par exemple, sera décidé autant en fonction du lieu de départ que de celui d'arrivée.

certaines réticulaires, en tout cas pas nécessairement hiérarchiques. L'emploi de pondérations différentielles, généralement employées dans ces cas, alourdit une analyse déjà malaisée et introduit une part d'arbitraire (Goldstein, Rasbash et al. 2000). Dans la meilleure des situations, une évaluation statistique d'un effet contextuel reposera sur la connaissance précise de son contour et la disposition de variables décrivant cet environnement ; autorisant d'en tester l'influence sur les durées de séjours ou les fonctions de risque. Il convient seulement de prendre soin de calculer des estimateurs robustes de leur paramètres¹⁶. Mais si l'étendue spatiale (ou sociale) et la caractérisation de ces effets sont méconnue, la présomption d'une influence contextuelle sera déduite de la présence d'une hétérogénéité inobservée, entre des unités spatiales données a priori, des circonscriptions administratives, par exemple. La mise en évidence d'une hétérogénéité résiduelle des entités territoriales entre elles est le signe d'une homogénéité, ou corrélation, des durées à l'intérieur de ces unités ; et par conséquent de l'influence de facteurs communs inobservés. Celle-ci sera évaluée moins par la confirmation statistique de coefficients de régression significatifs que par la réduction de cette variation résiduelle, à la suite de leur introduction. Mais l'analyse multiniveau — l'association d'effets fixes et aléatoires — des modèles de durée complexifie les conditions particulières de leur application. Dans les modèles à risques proportionnels par exemple, l'effet des composantes fixes comme aléatoires doit être constant en chaque point de la fonction de risque ; on doit avancer l'hypothèse également que les courbes des quotients instantanés sous-jacents sont parallèles d'un contexte à l'autre (Reardon, Brennan et al. 2001).

Ces réserves cependant ne doivent pas faire oublier que ce qui risque d'apparaître comme un raffinement méthodologique requis pour la mise en œuvre d'une analyse biographique contextuelle est en vérité une exigence fondamentale de la modélisation des durées de séjour en démographie. La raison en est que, contrairement aux sciences expérimentales, les modèles en sciences sociales sont, trop souvent, largement indéterminés. Or il est prouvé que dans les modèles non linéaires, et en particulier les modèles biographiques, omettre cette hétérogénéité produit des résultats inconsistants et biaisés (Trussel et Rodríguez 1990).

Un modèle multiniveau de la mobilité à Bogota

À ces exigences d'ordre général, qui sont autant d'appels à la prudence, s'en ajoutent de singulières à chaque cas traité. Le cas de la mobilité à Bogota sera examiné pour en illustrer quelques-unes, avec les solutions ou constats d'échec éventuels qu'elles inspirent. Ce sera l'occasion de présenter un chemin parmi toutes les solutions techniques possibles (et non décrites ici). En effet, leur diversité est parfois désarmante, selon que l'on retient un modèle totalement ou partiellement paramétrique (Goldstein 1995), une formulation en termes de risque proportionnel ou de durée accélérée (Pan 2001), une mesure continue, discrète ou regroupée des temps de séjour (Allison 1982; Yue et Chan 1997; Ross et Moore 1999; Goldstein 2000), les hypothèses sur la distribution de la composante aléatoire (Shih et Louis 1995; Sahu, Dey et al. 1997; Yue et Chan 1997; Biller 2000)...

¹⁶ Dans les données traitées, une même mesure contextuelle est répétée autant de fois que l'individu ou l'événement s'y reproduisent : la répétition crée une illusion de précision statistique qui induit des prévisions trop optimistes sur les intervalles de confiance.

L'enquête biographique utilisée porte sur un individu de chaque ménage, dont le parcours résidentiel est enregistré depuis sa naissance (Dureau et Lorez 1999). Cette disposition nous dispense d'envisager un effet contextuel du ménage : l'individu en change plusieurs fois au cours de sa vie et on ignore les trajectoires des membres les composant. Une autre conséquence est l'atypie des premières migrations qui, pour la plupart, correspondent à la sortie tardive du ménage des parents. Pour celles-ci, l'hypothèse de proportionnalité des risques n'est pas remplie ; de surcroît, elle déforme la répartition des risques (Figure 9), invalidant les modélisations paramétriques. Pour atténuer cet inconvénient, le traitement n'a retenu que les déplacements de personnes adultes (de plus de dix-huit ans), ce qui revient à écarter les migrations de premier rang (Figure 10).

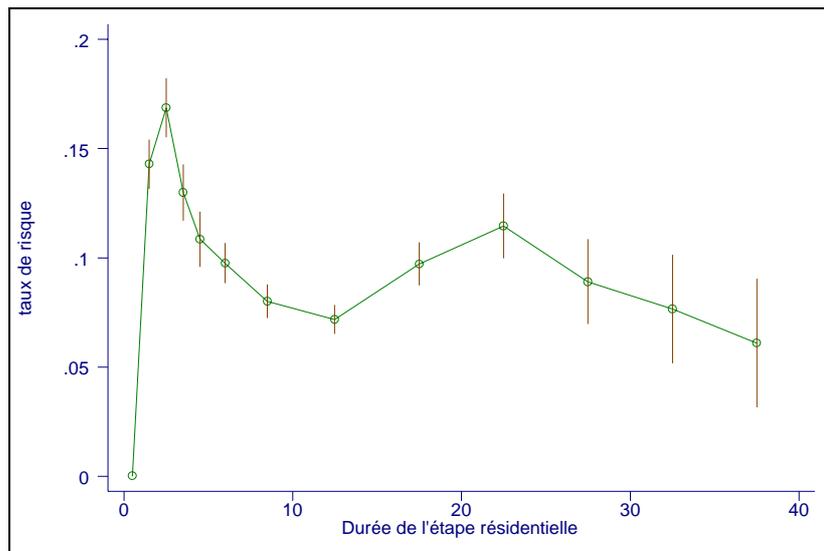


Figure 9 Répartition des risques, toutes migrations confondues

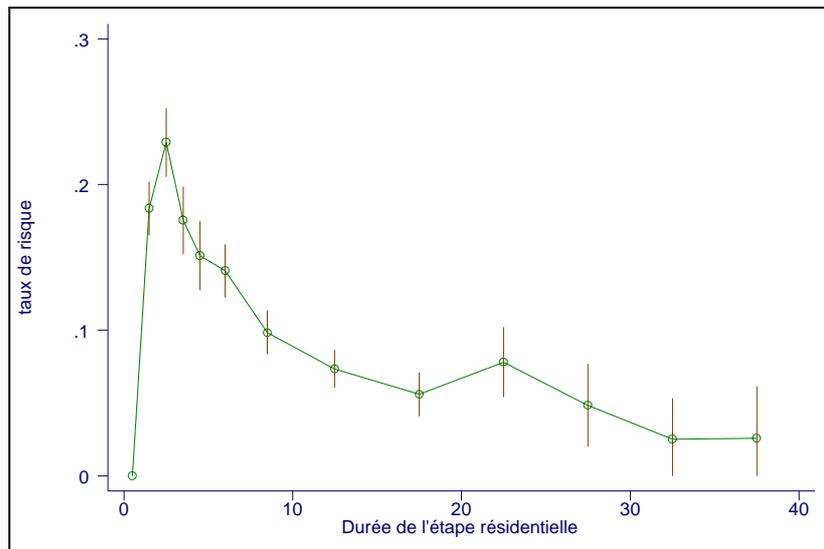


Figure 10 Répartition des risques pour les adultes de plus de 18 ans

De cette comparaison, il ressort néanmoins que la recrudescence des risques entre vingt et vingt-cinq ans n'est pas totalement gommée par la sélection. Cela peut être propre à l'échantillon (une histoire particulière de certaines générations enquêtées), ou révéler une pratique résidentielle

inhabituelle (le déménagement de certains parents une fois leurs enfants partis) ; qu'importe les raisons, le résultat est que cette fonction de risque n'est toujours pas paramétrable. Comme il s'agit d'une situation fréquemment rencontrée, une solution flexible était préférable. Elle est offerte par un découpage de la fonction de risque en autant de tronçons à l'intérieur desquels le logarithme du risque est une fonction linéaire de la durée (*piecewise Gompertz transformation*)¹⁷ : il suffit alors d'identifier les points de rupture pour s'adapter à toutes les formes rencontrées dans la pratique (Rafiq 1987; Panis 1994). Un autre avantage de cette transformation est de pouvoir introduire dans le modèle, et de manière additive, plusieurs fonctions temporelles du risque (Lillard et Panis 2000) ; une possibilité qui sera exploitée pour modéliser le contexte historique. Pour la mobilité résidentielle à Bogota, les points de rupture des ajustements linéaires se situent à 2, 20, 25, 30 années de séjour, conformément à la configuration de la Figure 10¹⁸.

La structure hiérarchique est la suivante : chaque séjour résidentiel compose les observations élémentaires. Parmi le jeu réduit de variables testées, seul l'âge au moment du déplacement est, à ce niveau, significatif. Vient ensuite l'individu, d'abord parce que certaines caractéristiques individuelles stables (sexe, génération, éducation...) lui correspondent, et sont répétées pour tous ses déplacements, mais aussi pour juger de l'hétérogénéité inobservée entre les individus. Afin de tester un modèle simple, seule l'appartenance à des groupes quinquennaux de générations a été introduite dans le modèle, dans un ordre croissant avec l'ancienneté. L'unité supérieure de regroupement est une unité spatio-temporelle, définie afin de conserver la structure hiérarchique de l'information : soit un arrondissement de la ville de Bogota¹⁹ par période historique de politiques urbaines. L'observation commence en 1947 (pour éliminer les trop rares occasions observées auparavant) jusqu'en 1966 — la période de planification spatiale —, de 1967 à 1978 pour la phase de planification économique, puis de 1979 à 1989 pour la période qualifiée de « normativisation » par T. Lulle qui propose ce découpage historique, pour enfin distinguer les années 1990 « du chaos au réveil » (Lulle 2000). Pour chacune de ces unités espace-temps de regroupement, la proportion de logement en propriété a été calculée afin de tester l'impact de cet effet contextuel sur les durées résidentielles. Pour faciliter la convergence des modèles multiniveaux, ont été écartées de l'analyse les unités espace-temps qui ne rassemblaient pas pour le moins dix déplacements²⁰.

La modélisation ici proposée se fixe trois objectifs : a) tester une batterie de variables sélectionnées à chacun des trois niveaux d'observation de la mobilité résidentielle²¹ ; b) introduire deux

¹⁷ Sur les modèles paramétriques, cette solution offre l'avantage de ne pas avoir à se préoccuper de la conformité des risques observés à la fonction théorique. Sur le modèle semi-paramétrique de Cox, elle conserve l'avantage d'une estimation par la vraisemblance complète et d'une mesure simple des risques sous-jacents. Dans le cadre de l'analyse multiniveau, le modèle de Cox revient à une discrétisation du temps et oblige à estimer chaque taux sous-jacent (éventuellement par une fonction polynomiale) Goldstein, H. (1995). *Multilevel statistical models*. London, New York, Edward Arnold ;Halsted Press.

¹⁸ Cette figure rapporte les taux observés, mais l'ajustement linéaire se fait sur les logarithmes de ces taux (Gompertz)

¹⁹ Dont une seule zone pour le reste de la Colombie.

²⁰ Ce fort déséquilibre résulte du mode de tirage des ménages, à partir de zones d'enquêtes choisies de manière raisonnée pour représenter le type d'habitat, pas la totalité de la population de l'aire métropolitaine. Cette extraction des unités peu documentées se justifie par la préoccupation méthodologique de ce papier.

²¹ Que nous rappelons : l'âge au moment du déplacement dont l'origine est déplacé à 18 ans, l'appartenance de l'individu à un groupe quinquennal de générations de plus en plus anciennes, la proportion de propriétaires dans l'unité spatio-temporelle.

mesures d'hétérogénéité inobservée, une entre les individus et l'autre entre les unités spatio-temporelles ; c) modéliser les variations historiques de la mobilité conjointement avec les calendriers individuels. Le traitement utilise le logiciel aML pour sa flexibilité, sa simplicité dans la construction de modèles mixtes de durée (Lillard et Panis 2000). Chaque tableau de résultats correspond à une étape de ce cheminement, qui sera plus méthodologique que thématique.

Tableau 1 Modélisation de la mobilité, sans composante aléatoire

	Modèle A1	Modèle A2	Modèle A3	Modèle A4
<i>Piecewise Gompertz linear baseline : la pente du log des risques en fonction de la durée</i>				
de 0 à 2 années de séjour	0.7301 ***	0.7515 ***	0.7931 ***	0.8129 ***
de 2 à 20 années de séjour	-0.1150 ***	-0.0938 ***	-0.0926 ***	-0.0845 ***
de 20 à 25 années de séjour	0.2137 ***	0.2106 ***	0.1933 ***	0.2076 **
plus de 25 années de séjour	-0.1501 **	-0.1493	-0.1456	-0.1419 *
Intersection	-2.9411 ***	-1.9675 ***	-1.4245 ***	-1.1795 ***
<i>Caractéristiques de la mobilité (trois niveaux d'observation)</i>				
Age au début du séjour (origine = 18 ans)		-0.0161 ***	-0.0060	-0.0001
Rang d'ancienneté des générations		-0.0758 ***	-0.1272 ***	-0.1428 ***
Proportion de logements en propriété		-0.0036 ***	-0.0042 ***	-0.0117 ***
Période 1967-78 (versus 1947-66)			-0.2794 ***	-0.2404 **
Période 1979-89 (versus 1947-66)			-0.3607 **	-0.2698
Période après 1990 (versus 1947-66)			-0.1602	0.0038
Déplacement hors aire métropolitaine				0.8202 ***
<i>Mesures de validité du modèle</i>				
Log de vraisemblance	-4647.39	-4570.71	-4565.59	-4503.93

Signification : * = 10%; ** = 5%; *** = 1%.

Le premier tableau donne une évaluation des seuls effets fixes, servant de référence pour les propositions multiniveaux à suivre. Son originalité est d'utiliser une fonction des risques scindée en quatre régressions linéaires (*Piecewise Gompertz linéaire*, Cf. supra). Le modèle A1 est inconditionnel, soit avec une évaluation des risques sans variable explicative. Elle confirme la forte montée durant les deux premières années, puis une décroissance régulière, seulement rompue par la recrudescence entre les vingtièmes et vingt-cinquièmes années dont la cause n'est pas claire, mais dont il faut tenir compte pour le réalisme du modèle. Le modèle suivant (A2) inclut les effets retenus à chaque niveau d'observation. L'âge au début de la migration (avec l'origine replacée à dix-huit ans) abaisse sensiblement tous les risques d'un changement de résidence, de même l'appartenance à des cohortes plus anciennes. L'accès à la propriété contribue également à réduire la mobilité résiden-

tielle²². Le modèle A3 introduit des variables de contrôle pour la périodisation des politiques de l'aménagement urbain, retenant la plus ancienne (1947-66) pour référence. Elles témoignent d'une réduction de la mobilité depuis la fin des années soixante, qui ne retrouverait son niveau original que très récemment, durant les années quatre-vingt-dix. Ce résultat, un peu étonnant, est peut-être l'indication d'un exode rural autrefois nourri par la transition démographique. Pour s'en assurer, le dernier modèle (A4) distingue les déplacements réalisés en dehors de l'aire métropolitaine. Manifestement, Bogota stabilise ses habitants, plus que la province en tout cas. Mais subsiste la trace statistique d'une plus forte mobilité des vingt premières années, même au sein de la ville où se prolongent les turbulences de l'exode rural. Notons que l'effet de l'âge est effacé par la prise en compte de ce calendrier urbain, il était donc le produit de ces particularités historiques. L'effet du rang d'ancienneté des générations se maintient d'un modèle à l'autre.

La seconde série de modèles (Tableau 2) reprend le cheminement précédent mais avec deux composantes aléatoires qui mesurent l'hétérogénéité résiduelle ou inobservée entre les individus d'une même unité spatio-temporelle, puis entre celles-ci. L'objectif est d'abord de tester la présence d'un effet contextuel, confirmé par les chiffres. Conjointement, il affine le modèle en corrigeant les paramètres par la prise en compte de son indétermination, due à des caractéristiques individuelles ou régionales méconnues. Examinons ces deux composantes aléatoires supposées normalement distribuées et donc paramétrées par leur écart-type : elles sont significativement différentes de zéro. Pour le modèle indéterminé (B1), elles décomposent la variation totale des durées ; dont on doit attendre la réduction au fur et à mesure de l'introduction des covariants. Ce que confirme la tendance de l'hétérogénéité individuelle pour les quatre formulations du modèle ; mais ce que contredit la composante régionale. En effet, l'hétérogénéité entre les unités spatio-temporelles augmente dès que l'on ajoute les périodes historiques. Est-ce le signe d'une mauvaise conception du modèle multiniveau ? Probablement car ces unités dont on évalue l'hétérogénéité sont construites à la fois sur la division territoriale et la périodisation du calendrier, leur hétérogénéité estimée absorbe donc l'effet de l'histoire de la ville. Et de fait, sa signification statistique s'estompe. La troisième série de modèle cherchera une solution à cette difficulté. Il est une autre imperfection : selon cette formulation, l'accession à la propriété perd son impact sur les durées de séjour. Or cet indicateur fut délibérément construit à partir des mesures biographiques individuelles (voir page 12). Il convenait d'introduire d'abord cette information au niveau individuel auquel elle appartient, avant de juger son impact contextuel via la composition résultante du quartier habité.

Cet essai montre que l'évaluation du contexte par les variables de contrôle est insatisfaisante, il ne l'est pas seulement du fait d'une définition inadéquate des environnements spatio-temporels. Doit être critiqué le choix de périodes définies a priori, de manière exogène par les politiques urbaines, sans être assuré du fondement empirique de ce choix. De plus, on oblige la part de mobilité afférente au calendrier à varier par pallier, elle est fixée à l'intérieur des périodes testées, changeant brutalement de l'une à l'autre, alors que les transitions sont inévitablement progressives.

²² Dans une mesure dont il faut remarquer qu'elle correspond à un point du pourcentage de logements en propriété dans l'arrondissement à la période observée

Tableau 2 Modélisation de la mobilité, avec deux mesures d'hétérogénéité inobservée

	Modèle B1	Modèle B2	Modèle B3	Modèle B4
<i>Piecewise Gompertz linear baseline : la pente du log des risques en fonction de la durée</i>				
de 0 à 2 années de séjour	0.9356 ***	0.9240 ***	0.9117 ***	0.9183 ***
de 2 à 20 années de séjour	-0.0689 ***	-0.0616 ***	-0.0645 ***	-0.0584 ***
de 20 à 25 années de séjour	0.1751 **	0.1723 **	0.1759 **	0.1857 **
plus de 25 années de séjour	-0.1343 *	-0.1317 *	-0.1303	-0.1328 *
Intersection	-3.1601 ***	-1.9985 ***	-1.5831 ***	-1.6304 ***
<i>Caractéristiques de la mobilité (trois niveaux d'observation)</i>				
Age au début du séjour (origine = 18 ans)		-0.0089 *	-0.0072	0.0009
Rang d'ancienneté des générations		-0.1265 ***	-0.1308 ***	-0.1546 ***
Proportion de logements en propriété		-0.0020	-0.0077	-0.0037
Période 1967-78 (vs 1947-66)			-0.2965	-0.3345 *
Période 1979-89 (vs 1947-66)			0.0041	-0.2839
Période après 1990 (vs 1947-66)			0.2317	-0.4583
Déplacement hors aire métropolitaine				1.0560 ***
<i>Composante aléatoires (exprimées en écart-type)</i>				
Hétérogénéité inobservée entre les individus	0.5693 ***	0.4891 ***	0.4452 ***	0.3999 ***
Hétérogénéité inobservée entre les unités spatio-temporelles	0.4235 ***	0.4097 ***	0.4345 ***	0.5833 ***
<i>Mesures de validité du modèle</i>				
Log de vraisemblance	-4582.75	-4516.14	-4513.09	-4450.94

Signification : * = 10%; ** = 5%; *** = 1%.

Une alternative est de concevoir une fonction de risque propre au calendrier historique, superposée à celle des séjours. La formulation dans les deux cas est identique, soit une périodisation par portions de régressions linéaires. L'ensemble de l'histoire de la ville est couvert (depuis 1947) mais l'addition des deux tracés commence à la date variable de chaque changement de résidence. L'avantage de la méthode est que l'on définit les points de rupture en fonction des évolutions effectivement observées. Pour les retrouver à partir des biographies individuelles, on découpe le calendrier historique en tranches régulières de temps assez nombreuses (quinquennales par exemple) pour suivre l'évolution de leurs pentes ; puis un regroupement séquentiel est retenu selon leurs ressemblances.

Tableau 3 Modélisation de la périodisation historique par une fonction additive du risque

	Modèle C1	Modèle C2
<i>Piecewise Gompertz linear baseline : la pente du log des risques en fonction de la durée du séjour</i>		
de 0 à 2 années de séjour	0.8829 ***	0.9409 ***
de 2 à 20 années de séjour	-0.0784 ***	-0.0689 ***
de 20 à 25 années de séjour	0.1976 **	0.1813 *
plus de 25 années de séjour	-0.1403	-0.1429
<i>Fonction de risque selon une périodisation historique (pente du log des risques sous-jacents)</i>		
Intersection	0.6293	-0.7138
Avant 1955	-0.2020	-0.1828
De 1955 à 1964	0.0314	0.0224
De 1965 à 1987	-0.0142	-0.0074
Après 1987	0.1212 ***	0.1232 ***
<i>Caractéristiques de la mobilité (trois niveaux d'observation)</i>		
Intersection	-2.8109	-0.3179
Age au début du séjour (origine = 18 ans)		-0.0150
Rang d'ancienneté des générations		-0.0818
Proportion de logements en propriété		-0.0038
Déplacement hors aire métropolitaine		1.4128 ***
<i>Composante aléatoires (exprimées en écart-type)</i>		
Hétérogénéité inobservée entre les individus	0.5367 ***	0.4587 ***
Hétérogénéité inobservée entre les unités spatio-temporelles	0.4651 ***	0.3593 ***
<i>Mesures de validité du modèle</i>		
Log de vraisemblance	-4564.98	-4434.33

Signification : * = 10%; ** = 5%; *** = 1%.

Il en résulte quatre périodes à retenir pour l'estimation finale du modèle : de 1947 à 1955 quand la mobilité, apparemment élevée, tend à décroître, suivie d'une décennie de sensible reprise avant une stabilisation d'une vingtaine d'année, très légèrement négative. Mais ces mouvements ne sont pas statistiquement prouvés ; à l'exception de la nette recrudescence à la fin des années quatre-vingt. L'introduction des quelques variables retenue ne change pratiquement pas ces évolutions. En revanche, l'effet de trois d'entre elles sur quatre est absorbé par la tendance lourde urbaine, voire en partie colombienne car cette mobilité résidentielle saisie à Bogota passe en partie par la province. La

plus grande brièveté des séjours hors de la ville, pour cet échantillon de citoyens, est confirmée selon un coefficient multiplicatif plus élevé que dans les estimations précédentes. Si ce modèle reste partiellement indéterminé, on constate, avec satisfaction, que l'hétérogénéité inobservée est grandement réduite par l'introduction des variables testées.

Conclusion

Il faudra juger des promesses de ces trois applications par des études de cas plus approfondies, documentées par des statistiques mieux adaptées à l'analyse spatiale des biographies individuelles. On songe aux registres de population ou le relevé de mobilités courtes, moins exposées aux défaillances de la mémoire des informateurs. Reste, bien sûr, à inventer les bonnes questions à poser à l'espace biographique, reprenant les acquis théoriques de la géographie du temps (*time-geography*). D'autres représentations cartographiques méritent être explorées — la graphique des réseaux, tout juste évoquée dans ce travail, par exemple —, de même que d'autres formulations des modèles de durées, notamment en termes de temps discrets, peut-être plus faciles à mettre en œuvre. Ce qui ne fait pas de doute, c'est l'utilité des nouveaux outils géomatiques dont les SIG et surtout le caractère indispensable de la modélisation multiniveau. Ses méthodes sont désormais suffisamment robustes pour traiter avec rigueur et commodité les modélisations spatio-temporelles, dont il est finalement question ici, à plusieurs niveaux d'observation. Les trois applications développées démontrent que le traitement de la localisation, conjointement avec la durée biographique, est possible et riche de possibilités analytiques. Ses applications sont aussi variées que nécessaires : configuration spatiale de la mortalité, mobilités urbaines et recompositions territoriales... Et bien sûr l'étude de la migration dont on peut se demander comment elle a pu faire l'économie des informations géo-référencées contenues dans les enquêtes biographiques.

Référence des ouvrages cités

- Aalen, O. O. (1988). "Heterogeneity in survival analysis." Stat Med **7**(11): 1121-37.
- Allison, P. D. (1982). "Discrete-time methods for the analysis of event histories." Sociological Methodology **1982**: 61-98.
- Billier, C. (2000). "Discrete duration models combining dynamic and random effects." Lifetime Data Analysis **6**(4): 375-90.
- Boyd, L. H. et G. R. Iversen (1979). Contextual analysis : concepts and statistical techniques. Belmont, Wadsworth Publishing Company.
- Brewster, K. L. (1994). "Race differences in sexual activity among adolescent women; the role of neighborhood characteristics." American Sociological Review **59**: 408-424.
- CASSINI (1999). "Représentation de l'espace et du temps dans les SIG." Revue Internationale de Géomatique **9**(1).
- Cleves, M. (1999). "Analysis of multiple failure-time survival data." Stata Technical Bulletin **13**: 14.
- Courgeau, D. (2000). Vers une analyse biographique multiniveau. Journées de méthodologie statistique de l'INSEE, Paris.
- Courgeau, D. et E. Lelièvre (1989). Analyse démographique des biographies. Paris, Ined.
- de Jong, T. et H. Hoheus (1997). GIS functionality for multilevel research. Place and People: multilevel modelling in geographical research. W. G. R. and R. N. Verhoeff. Utrecht, Nederlandse Geografische Studies: 227.
- Delaunay, D. (1997). L'analyse multiniveau de l'information géoréférencée, applications à la démographie. XXXIII^{ème} Conférence générale de population, Pékin.
- Duley, C. et P. Rees (1991). Incorporating migration into simulation models. Migration models - macro and micro approaches-. J. Stillwell and P. Congdon. London, Great Britain, Belhaven Press: 228-261.
- Dureau, F., C. E. Florez, et al. (1994). La movilidad de las poblaciones y su impacto sobre la dinámica del área metropolitana de Bogotá -documento de trabajo no 2, metodología de la encuesta cuantitativa, anexos: cuestionario y manuales. Bogota, Colombia, ORSTOM-Universidad de los Andes.
- Dureau, F. et C. E. Lorez (1999). Enquêtes mobilité spatiale à Bogota et dans trois villes du Casanare (Colombie). Renseignements concernant le déroulement et l'évaluation de la collecte. Biographies d'enquête. Paris, INED: 241-278.
- ESRI (1996). Avenue: customization and application development for ArcView, ESRI, Inc.
- ESRI (1998). ArcView Tracking Analyst: real-time data display, interactive playback and analysis, TACS, Inc.
- Fotheringham, A. S., M. Wegener, et al. (2000). Spatial models and GIS : new potential and new models. London ; Philadelphia, Taylor & Francis.
- Goldstein, H. (1995). Multilevel statistical models. London, New York, Edward Arnold ;Halsted Press.
- Goldstein, H. (2000). Discrete time event history models. Working Paper: 6.
- Goldstein, H., J. Rasbash, et al. (2000). "Multilevel Models in the study of dynamic household structure." European Journal of Population **16**: 373-387.
- Guo, G. (1993). "Use of the sibling data to estimate family mortality effects in Guatemala." Demography **30**: 15-32.
- Hägerstrand, T. (1973). The domain of human geography. Human Geography. C. R. London, Methuen.
- Hägerstrand, T., G. Carlestam, et al. (1991). Om tidens vidd och tingens ordning. [Stockholm] : Solna :, Byggeforskningsradet ; Distribution, Svensk byggtjänst.
- Hougaard, P. (1995). "Frailty models for survival data." Lifetime Data Anal **1**(3): 255-73.
- Jones, K. et C. Duncan (1995). "Individuals and their ecologies; analysing the geography of chronic illness within a multilevel framework." Health and place **1**(1): 27-40.
- Keiding, N., P. K. Andersen, et al. (1997). "The role of Frailty models and accelerated failure time models in describing heterogeneity due to omitted covariates." Stat Med **16**(1-3): 215-24.
- Lillard, L. A. et C. W. A. Panis (2000). aML user's guide and reference manual. Release 1.0. Los Angeles, California, EconWare.

- Little, T. D., K. U. Schnabel, et al. (2000). Modeling longitudinal multilevel data : practical issues, applied approaches, and specific examples. Mahwah, N.J., Lawrence Erlbaum Assoc.
- Longford, N. T. (1993). Random coefficient models. Oxford, Charenton Press.
- Lulle, T. (2000). Bogotá: les coûts du laissez-faire. Métropoles en mouvement. Une comparaison internationale. F. Dureau, V. Dupont, E. Lelièvre, J.-P. Lévy and T. Lulle. Paris, Anthropos - Institut de Recherche pour le Développement: 343-350.
- Pan, W. (2001). "Using frailties in the accelerated failure time model." Lifetime Data Anal 7(1): 55-64.
- Panis, C. W. A. (1994). "The piecewise linear spline transformation with an application to age at first marriage." Stata Technical Bulletin 18: 27-29.
- Rafiq, S. (1987). The use of a piecewise constant proportional hazards model in issues related to standardisation : a review and some results. Interuniversity Programme in Demography. Centrum voor Sociologie, Vrije Universiteit Brussel: 36.
- Reardon, S. F., R. Brennan, et al. (2001). Estimating multilevel discrete-time hazard models using cross-sectional data: neighborhood effects on the onset of adolescent cigarette use. Working Paper: 43.
- Ross, E. A. et D. Moore (1999). "Modeling clustered, discrete, or grouped time survival data with covariates." Biometrics 55(3): 813-9.
- Sahu, S. K., D. K. Dey, et al. (1997). "A Weibull regression model with gamma frailties for multivariate survival data." Lifetime Data Anal 3(2): 123-37.
- Sastry, N. (1996). "Community characteristics, individual and household attributes, and child survival in Brazil." Demography 23(2): 211-229.
- Shih, J. H. et T. A. Louis (1995). "Assessing gamma Frailty models for clustered failure time data." Lifetime Data Anal 1(2): 205-20.
- Trussel, J. et G. Rodríguez (1990). Heterogeneity in demographic research. Convergent issues in genetics and Demography. J. L. Adams, David A.; Hermalin Albert I. ; Smouse, Peter E. London, Oxford University Press: 111-132.
- Vaupel, J. W. (1988). "Inherited Frailty and longevity." Demography 25(2): 277-87.
- Xue, X. et Y. Ding (1999). "Assessing heterogeneity and correlation of paired failure times with the bivariate Frailty model." Stat Med 18(8): 907-18.
- Yue, H. et K. S. Chan (1997). "A dynamic Frailty model for multivariate survival data." Biometrics 53(3): 785-93.