



HAL
open science

Visualisation graphique agrégée des trajectoires individuelles : revue de l'existant et application en géographie

Hadrien Commenges, Pierre Pistre

► To cite this version:

Hadrien Commenges, Pierre Pistre. Visualisation graphique agrégée des trajectoires individuelles : revue de l'existant et application en géographie. *M@ppemonde*, 2014, 115 (3). halshs-01238371

HAL Id: halshs-01238371

<https://shs.hal.science/halshs-01238371v1>

Submitted on 4 Dec 2015

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.

Visualisation graphique agrégée des trajectoires individuelles : revue de l'existant et application en géographie

Hadrien Commenges, Pierre Pistre

Résumé.— Les données longitudinales forment un pan important de la statistique en sciences sociales. La démographie a développé des représentations graphiques spécifiques pour ces données. Celles-ci restent peu utilisées pour l'analyse des dynamiques spatiales. Après une revue de ces modes de visualisation à partir d'un jeu de données fictives, l'article propose un mode original de représentation « en coulées », utile pour analyser les trajectoires d'individus sociologiques comme d'individus géographiques. Son utilisation est illustrée à partir de trois exemples : mobilités résidentielles, trajectoires modales de transport, dynamiques de l'emploi et de la résidence.

Données longitudinales • Individu • Logiciel R • Visualisation graphique

Abstract.— Longitudinal data are an important part of statistics in the social sciences. Demography has developed specific graphic visualization for these data but their use remains residual for the analysis of spatial dynamics. After a review of these graphic displays using a toy dataset, the paper proposes an original mode of visualization called « slide plot » conceived to study trajectories of individuals or spatial units. Its use is illustrated with three examples : residential mobility, changes of modal choice in transportation, dynamics of spatial mismatch.

Longitudinal data • Individual • R software • Graphic visualization

Resumen.— Los datos longitudinales constituyen una importante parte de la producción estadística en las ciencias sociales. En el campo de la demografía se han desarrollado representaciones gráficas específicas para este tipo de datos. Estos siguen siendo poco utilizados para el análisis de las dinámicas espaciales. Después de una revisión de estos modos de visualización a partir de datos ficticios, el artículo propone un modo original de representación « en ríos », útil para el análisis de las trayectorias de individuos o unidades espaciales. Su uso se ilustra con tres ejemplos : movilidad residencial, cambio de modos de transporte, repartición de los lugares de trabajo y de residencia.

Datos longitudinales • Individuo • Programa R • Visualización gráfica

Introduction : de l'intérêt des analyses longitudinales en géographie

La spatialisation des phénomènes humains et l'étude de leur évolution au cours du temps constituent l'approche principale d'analyse de nombreux géographes, qui l'appliquent à des objets variés : les dynamiques de peuplement, les structures productives, les migrations internationales ou encore la mobilité quotidienne. Lorsque leur travail empirique s'appuie sur une information quantitative, la démarche dominante est agrégée (Durand-Dastès, 2001) : elle mobilise avant tout des agrégats d'individus regroupés selon leurs caractéristiques sociodémographiques, leur lieu de résidence ou leurs pratiques spatiales. Dans une approche synchronique, le dispositif transversal renvoie une photographie du phénomène étudié à un temps t (ex. : recensements de la

population, registres d'entreprises, enquêtes sur les flux d'individus ou de marchandises). Dans une approche diachronique, l'évolution est alors saisie par la comparaison des photographies du phénomène renvoyées par le dispositif transversal à plusieurs temps successifs. Approche agrégée et dispositif transversal de quantification vont donc souvent de pair, puisque l'approche agrégée ne nécessite pas de suivre les mêmes individus au cours du temps. Et ceci explique, pour une bonne part, que les données longitudinales restent peu exploitées dans l'analyse des dynamiques spatiales.

Dans une approche transversale, les géographes mobilisent massivement les méthodes statistiques descriptives et modélisatrices, ainsi que les modes de visualisation graphique et cartographique. Ils sont en revanche peu nombreux à utiliser des modes de visualisation et des méthodes statistiques dans une démarche véritablement longitudinale. Cet article a un double objectif, théorique et méthodologique : il amène une réflexion sur l'utilisation des données longitudinales et l'étude des trajectoires en géographie, et plus généralement dans les sciences sociales. Il propose parallèlement un mode de visualisation agrégée des trajectoires individuelles au cours du temps, utile pour l'étude d'individus sociologiques comme d'individus géographiques sous forme d'unités spatiales. Ce mode est dit de visualisation *en coulées* : il a été pour la première fois utilisé dans une étude empirique des mobilités résidentielles à destination des communes rurales françaises (Pistre, 2011, 2012). Le présent article systématise l'usage de ce mode de représentation des données longitudinales, par sa mise en application dans plusieurs champs d'analyse. De plus, il est associé à un deuxième texte qui met à disposition l'outil de visualisation *en coulées* sous deux formes : une fonction à utiliser avec le logiciel R et une plateforme web d'exploration graphique des données longitudinales (Commenges *et al.*, 2014 ; <http://slider.parisgeo.cnrs.fr>).

Un cadrage méthodologique et historique sur les données et les dispositifs longitudinaux est tout d'abord présenté. En se concentrant sur la représentation graphique de ce type de données, nous proposons ensuite une revue des principaux modes existants (*en chronogramme* et *en tapis*) ainsi qu'une présentation appliquée du mode de visualisation *en coulées*. La troisième partie est dédiée à l'utilisation de cette forme de représentation : trois exemples sont présentés pour illustrer plusieurs possibilités d'application. Le premier exemple traite des mobilités spatiales d'individus sociologiques, à travers l'étude des trajectoires résidentielles de personnes qui ont débuté leur vie dans une grande ville française entre 1940 et 1960. Le deuxième exemple concerne l'évolution du mode dominant de transport utilisé par les résidents de l'agglomération de Seattle pour aller travailler. Le dernier exemple porte sur l'évolution de la répartition des lieux de résidence et des lieux d'emploi dans les communes de Paris et de sa petite couronne depuis les années 1960.

1. Cadrage sur les données et les dispositifs longitudinaux

Deux grands types de dispositifs d'enquête sont à distinguer dans la production de données statistiques : les *dispositifs transversaux* (« cross sectional design ») et les *dispositifs longitudinaux* (« longitudinal design »). Les dispositifs transversaux produisent la photographie d'un phénomène à un instant donné : le recensement de la population et la plupart des enquêtes INSEE (enquêtes Logement, enquêtes Patrimoine, enquêtes Emploi du temps, etc.) relèvent de ce type de dispositif. Les

dispositifs longitudinaux procèdent de leur côté à des mesures répétées sur les mêmes individus, et trois grands types sont à distinguer (Taris, 2000) :

- les dispositifs d'intervention qui mesurent la réaction ou la réponse des individus à un traitement et qui procèdent, en général, à deux vagues de mesures, avant et après traitement. Ces dispositifs concernent essentiellement des enquêtes et des protocoles mis en œuvre dans le domaine de la santé ;

- les dispositifs panels qui suivent un groupe d'individus et qui procèdent, la plupart du temps, à des vagues de mesures plus nombreuses. Citons par exemple, dans le cas français, l'Échantillon Démographique Permanent (EDP), mis en place par l'Insee à partir du recensement de 1968 et qui comptait près de 900 000 individus, résidant ou ayant résidé sur le territoire métropolitain, au milieu des années 2000 (Couet, 2006) ;

- les dispositifs rétrospectifs qui comportent une seule vague d'enquête et qui recueillent des informations sur un laps de temps passé. L'enquête Biographies et entourage conçue par des chercheurs de l'INED en est un exemple (Lelièvre, Vivier, 2001).

En outre, il convient de distinguer les dispositifs longitudinaux des données longitudinales. Les dispositifs sont dits longitudinaux lorsqu'ils procèdent à des mesures répétées sur les mêmes individus, alors que les données sont dites longitudinales lorsqu'elles caractérisent les mêmes individus à différents instants ou périodes. Ainsi, un dispositif rétrospectif n'est pas à proprement parler longitudinal (une seule vague), mais il produit des données longitudinales.

La naissance des dispositifs longitudinaux est liée au développement de la médecine moderne. L'histoire des dispositifs d'intervention débute en 1753, avec la publication du *Treatise of the scurvy* (« Traité du scorbut ») par James Lind, médecin dans la marine britannique. Considéré comme le premier essai clinique contrôlé, il y réalise une expérience sur des marins atteints du scorbut, en différenciant un groupe de contrôle qui ne reçoit aucun traitement d'un autre groupe d'individus traités au jus de citron (Martini, 2005).

La mise en place de méthodologies systématiques dans les études cliniques correspond ensuite au développement de la statistique, entre la fin du XIX^e et le début du XX^e siècle (Desrosières, 1993). Les dispositifs d'intervention se diffusent dans tous les domaines de la médecine et de la psychologie (Campbell, Stanley, 1966). Dans le champ restreint des études cliniques, des enquêtes rétrospectives et celles dites prospectives (du type des enquêtes panels) sont également mises en œuvre à partir des années 1930 (Doll, 2001a, 2001b). Ces dispositifs de mesure et les méthodes statistiques associées se diffusent par la suite dans d'autres domaines scientifiques et, en premier lieu, en démographie. À la fin des années 1960, plusieurs panels démographiques notamment sont mis en place : le *Panel Study of Income Dynamics* (PSID) aux États-Unis ou l'EDP en France (Duncan *et al.*, 1987 ; Couet, 2006). Enfin, les dispositifs rétrospectifs sont développés dans les études démographiques : on peut citer les travaux réalisés à l'INED, dans le cadre de l'analyse des biographies et la mise en place d'enquêtes *ad hoc* (Triple Biographie, Peuplement et dépeuplement de Paris, Biographies et entourage) (Courgeau, Lelièvre, 1989 ; Bonvalet, 2009 ; Lelièvre, Vivier, 2001).

Historiquement, les dispositifs longitudinaux occupent aussi une grande place dans les études économiques sur la consommation. L'origine du terme panel est d'ailleurs attribuée à P. F. Lazarsfeld et M. Fiske (1938) dans une analyse des effets de la publicité à la radio sur les comportements d'achat. Ils concluaient sur l'intérêt du panel

comme instrument de mesure de l'opinion, le considérant même comme le dispositif le plus pertinent pour saisir les effets de la publicité sur la consommation des individus. Le panel occupe depuis une place de choix dans la boîte à outils des études de marché et des pratiques de consommation.

2. Visualisation de données longitudinales : revue de l'existant et proposition

Différents rappels historiques et précisions méthodologiques ayant été faits sur la production et l'utilisation des données longitudinales, cette deuxième partie a pour objectif de présenter plusieurs modes de visualisation graphique de ce type de données, puis de proposer une forme originale de représentation. Complémentaire des modes existants, cette dernière a pour intérêt de mieux rendre compte de la capacité des données longitudinales à décrire les changements d'état au cours d'une période donnée, qu'il s'agisse d'un individu isolé ou surtout d'un groupe d'individus présentant chacun des trajectoires d'états spécifiques.

Cette revue de l'existant ne prétend pas à l'exhaustivité. Elle présente les principaux modes de visualisation graphique de données longitudinales, pour dégager leurs intérêts et limites respectifs. La conception proprement dite des figures a été réalisée avec le logiciel R, et notamment la bibliothèque *TraMineR* développée par Gabadinho *et al.* (2011a, 2011b), qui fait aujourd'hui référence dans l'analyse des données longitudinales. Cependant, ces modes de visualisation de données longitudinales existent dans la littérature indépendamment de l'outil informatique utilisé ici.

L'état de l'art est mené sur un jeu de données unique afin de rendre plus aisée la comparaison des différentes méthodes de visualisation. Dans un objectif pédagogique, et pour introduire plus efficacement ces méthodes, un jeu de données fictives a été créé décrivant un monde dans lequel les individus, les espaces et les relations individus-espaces sont très simplifiés. Dans ce monde, il existe trois catégories d'espaces clairement identifiables : des espaces ruraux, des espaces urbains et des espaces intermédiaires. Deux types d'individus y vivent et circulent dans ces catégories d'espace au cours de leur vie, et cette circulation dessine de façon stylisée des dynamiques de peuplement rural-urbain. Les données utilisées renseignent sur les caractéristiques des individus au premier pas de temps de l'analyse (soit 2020) et retracent leurs parcours résidentiels sur une période de 25 ans (soit de 2020 à 2045).

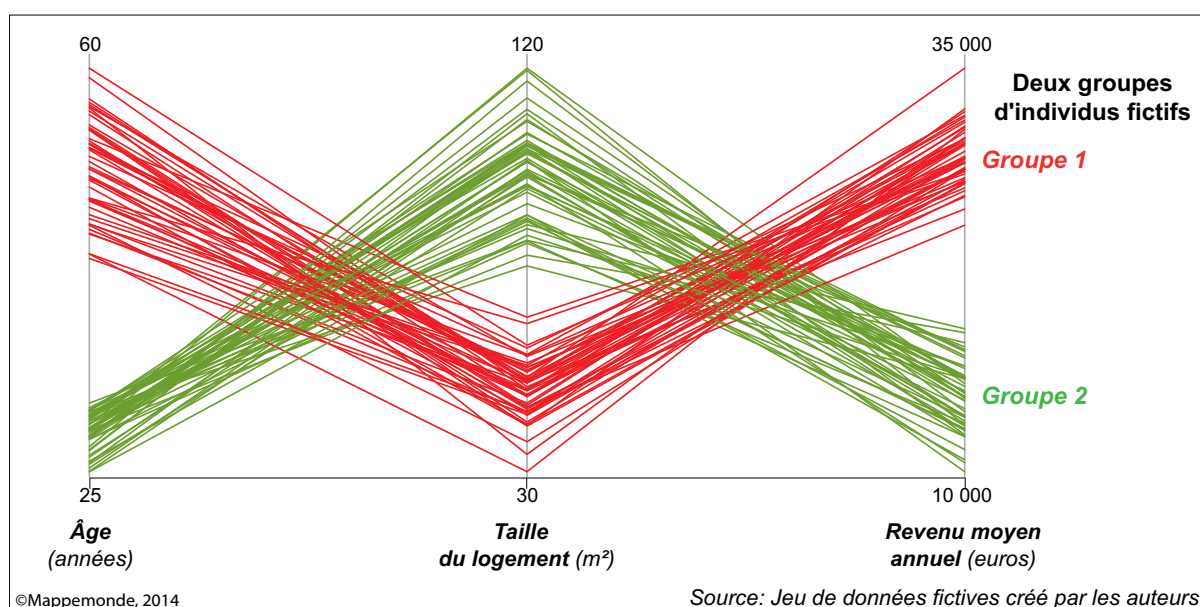
Le premier type est constitué d'individus vivant majoritairement dans des espaces urbains : ils sont âgés, ont un revenu élevé et résident dans des logements de petite taille. Le second type est constitué d'individus vivant majoritairement dans des espaces ruraux : ils sont jeunes, ont un faible revenu et résident dans des logements de grande taille. Le parcours résidentiel de ces deux groupes d'individus dessine un chassé-croisé entre l'urbain et le rural : les individus âgés se sont enrichis dans l'urbain, où ils ont passé la plus grande partie de leur vie active, puis ils partent progressivement dans le rural autour du passage à la retraite. Les individus jeunes, quant à eux, partent progressivement dans l'urbain en début de vie active. Dans les deux cas, le passage de l'urbain au rural ou du rural à l'urbain se fait en passant par des espaces intermédiaires.

2.1. Trajectoires graphiques : exemple de représentation en coordonnées parallèles (*parallel coordinates*)

En préambule à la présentation de plusieurs exemples de visualisation graphique des données longitudinales, cette revue débute par l'évocation du mode de représentation *en coordonnées parallèles* (« parallel coordinates plot », cf. Inselberg, 2008). Ce type de représentation ne s'applique pas à des données longitudinales, mais il est mentionné pour amener le mode de visualisation *en coulées* présenté en fin de partie. D'une manière générale, le mode de représentation *en coordonnées parallèles* consiste à relier, au moyen d'un figuré graphique linéaire, les valeurs enregistrées par un individu pour différentes variables quantitatives qui le caractérisent. Des trajectoires graphiques individuelles sont ainsi constituées et la formation d'agrégats visuels rend compte de profils différenciés (figure 1).

Sur notre jeu de données, les deux types d'individus se distinguent très clairement : en moyenne, les premiers (Groupe 1) sont âgés de 50 ans, ils vivent dans des logements de 50 m² et ils ont un revenu annuel de 30 000 euros ; en moyenne, les seconds (Groupe 2) sont âgés de 25 ans, ils vivent dans des logements de 100 m² et ils ont un revenu annuel de 15 000 euros.

Pourquoi amorcer notre revue par ce type de représentation graphique alors même qu'il n'utilise pas de données longitudinales ? Les modes de visualisation présentés ci-après ont plusieurs intérêts pratiques et analytiques. Cependant, la très grande majorité néglige la représentation des données longitudinales sous forme de trajectoires, alors même qu'il s'agit de l'intérêt majeur de ce type de données. La proposition de visualisation présentée ensuite met précisément l'accent sur la figuration des trajectoires individuelles et collectives, à l'instar du mode de représentation *en coordonnées parallèles* qui le propose pour des données non longitudinales.



1. Exemple de visualisation en coordonnées parallèles.

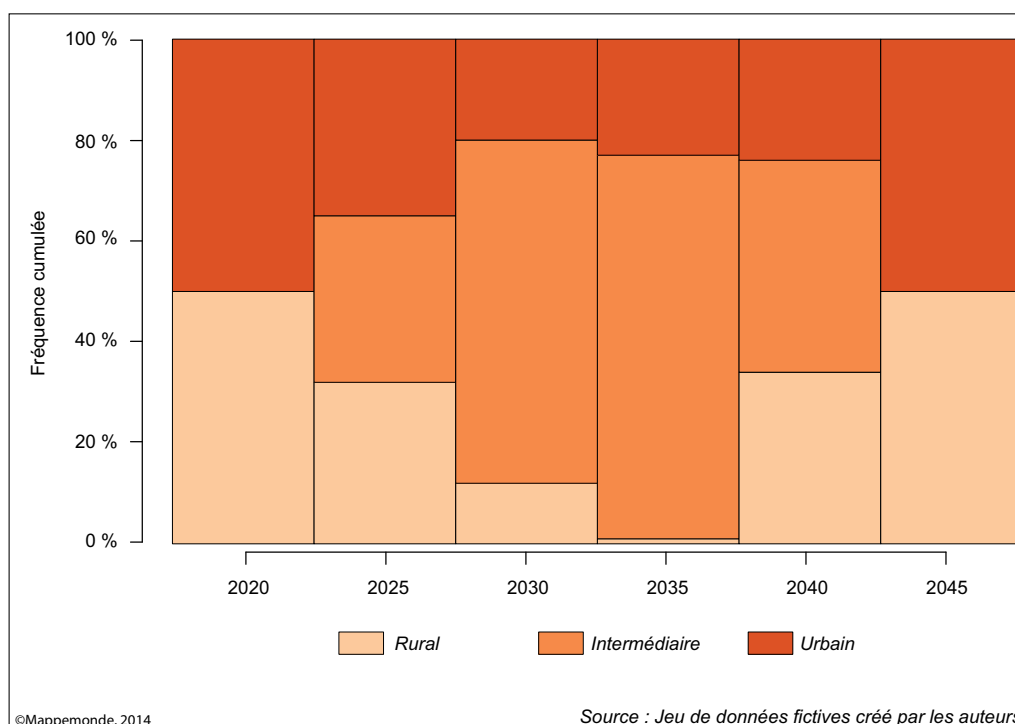
2.2. Modes de visualisation graphique des données longitudinales

Deux formes principales de représentation graphique des données longitudinales existent dans la littérature : *en chronogramme* (« state distribution plot ») et *en tapis* (« sequence index plot »). Ces deux modes de visualisation sont notamment utilisés pour présenter des types de trajectoires individuelles obtenus après classification statistique, comme dans le cas de parcours professionnels (Grelet, 2002), de trajectoires de vie (Robette, 2011) ou de programmes d'activités (Thévenin *et al.*, 2007).

Le mode *en chronogramme* représente la distribution des individus, répartis dans plusieurs modalités, à chaque pas de temps de la période d'étude. Il montre ainsi la fréquence des différentes modalités enregistrées dans les tranches temporelles successives. À la différence du précédent, la représentation *en tapis* conserve l'aspect individuel des trajectoires et elle est utilisée pour visualiser tout ou partie d'un échantillon de trajectoires individuelles. Chacune d'elles y est représentée de manière autonome, et des segments de couleurs, de grisés, voire de figurés, permettent de différencier les états successifs de chaque individu. La plus ancienne représentation en tapis que nous ayons recensée a été proposée par Bonnafous *et al.* (1981). Il s'agit d'un ensemble de fiches cartonnées dont la tranche est colorée en fonction des déplacements et des activités effectués par un échantillon d'individus au cours de la journée.

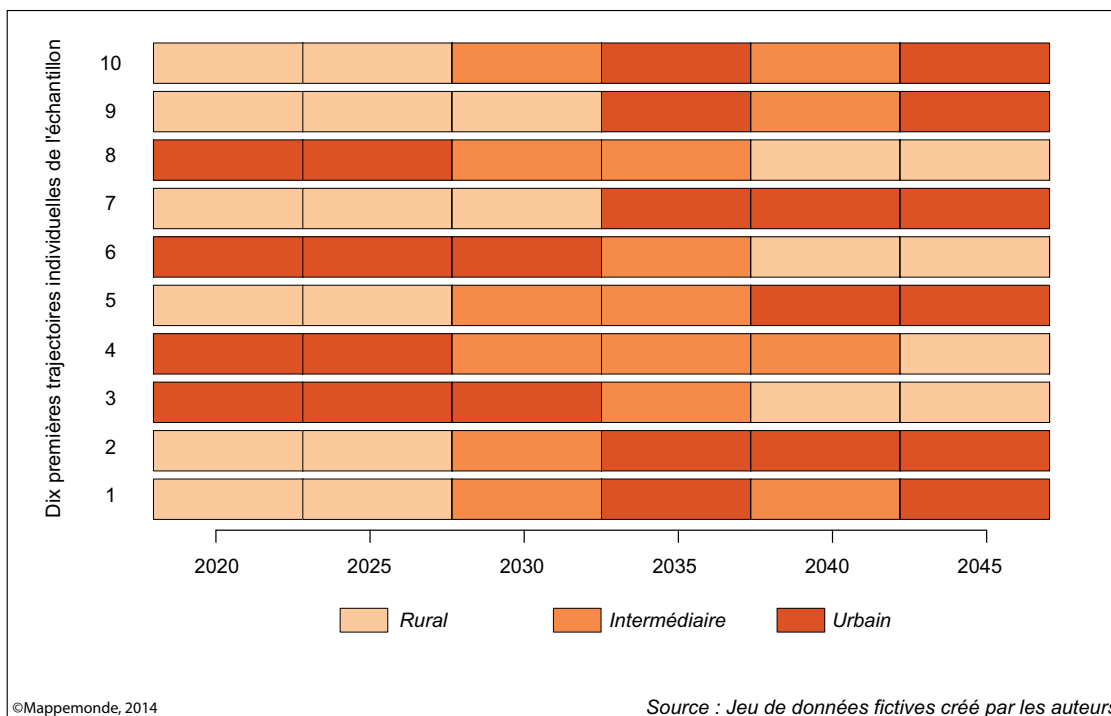
Les représentations graphiques qui suivent (figures 2, 3, 4 et 5) illustrent les principaux modes de visualisation des données longitudinales. Quatre exemples sont développés à partir du même jeu de données fictives.

Le premier exemple correspond à une représentation typique *en chronogramme* des données longitudinales (figure 2) : les fréquences cumulées des trois catégories d'espaces sont associées pour les six pas de temps considérés. La représentation permet de repérer les modalités récurrentes à chaque pas de temps et l'évolution de leur poids respectif au cours de la période d'étude : en début de période, la moitié des individus réside dans des espaces ruraux et l'autre moitié réside dans des espaces

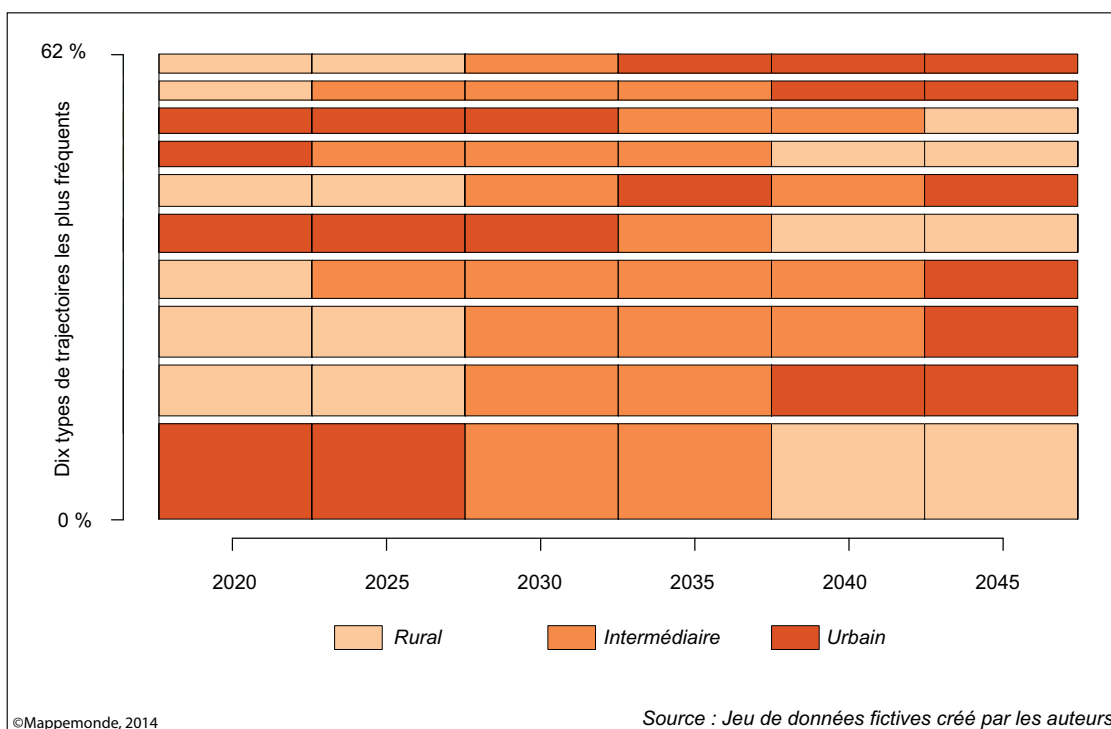


2. Exemple de visualisation en chronogramme.

urbains. En milieu de période, les espaces intermédiaires sont les plus peuplés et, en fin de période, on retrouve la même équirépartition entre espaces ruraux et urbains. Cependant, ce mode de représentation ne permet pas de connaître les étapes successives franchies par tel individu ou groupe d'individus pour passer, par exemple, d'une catégorie initiale « Rural » à une catégorie finale « Urbain ». En outre, il ne requiert pas forcément des données longitudinales : l'utilisation d'un échantillon représentatif d'une population générale à chaque pas de temps, c'est-à-dire une succession de mesures transversales, permet de réaliser le même type de visualisation graphique.



3. Exemple de visualisation en tapis des dix premiers individus de l'échantillon.



4. Exemple de visualisation en tapis des dix trajectoires individuelles les plus fréquentes.

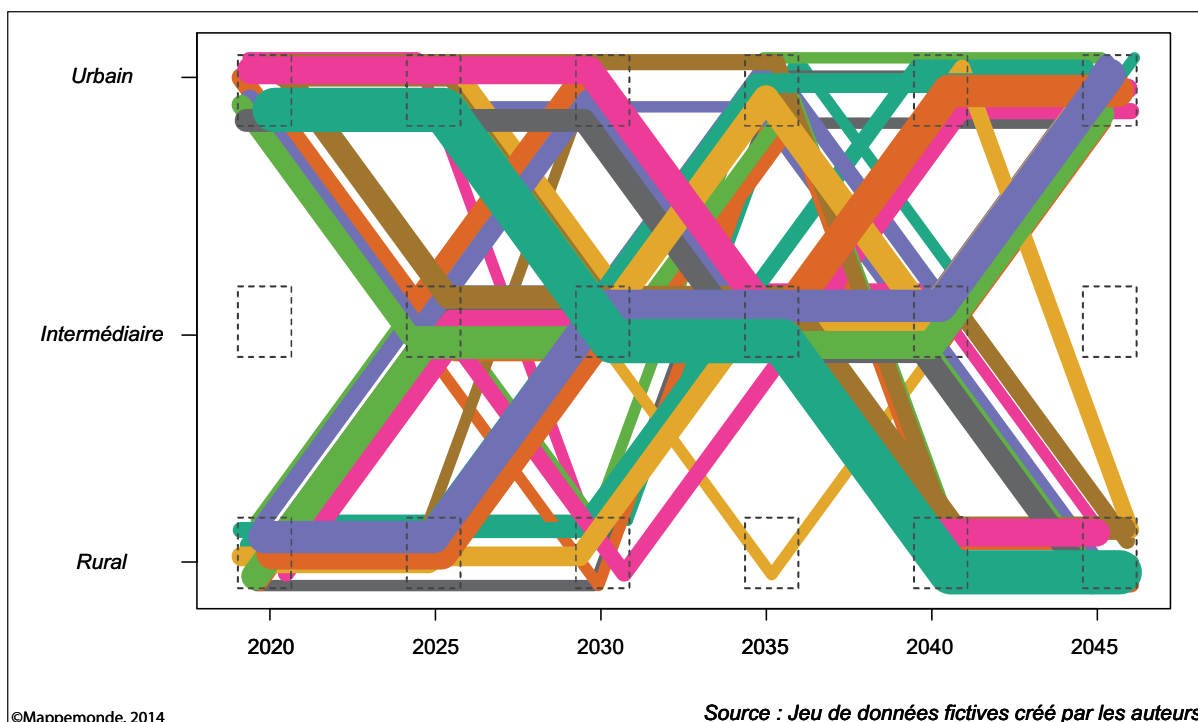
La deuxième représentation est un exemple de visualisation *en tapis* des données longitudinales (figure 3). Seules les trajectoires des dix premiers individus de l'échantillon sont ici représentées et chacune est composée de plusieurs états successifs. La confrontation des dix trajectoires individuelles renvoie une image de la structure des données et permet de dégager quelques récurrences. Les parcours résidentiels ont l'air de suivre une logique unidirectionnelle : les individus se dirigent soit du rural vers l'urbain, soit de l'urbain vers le rural, mais il n'y a pas de va-et-vient entre ces deux catégories au sein des mêmes parcours. Principale limite de cette visualisation *en tapis* : les dix premières trajectoires individuelles ne sont en rien représentatives de l'ensemble de l'échantillon.

Le troisième mode de représentation relève également d'une visualisation *en tapis* des données longitudinales, mais les trajectoires des dix premiers individus sont remplacées par les dix trajectoires individuelles les plus fréquentes dans l'ensemble de l'échantillon (figure 4). Cette visualisation constitue une forme de synthèse graphique des deux représentations précédentes, en conservant à la fois l'intérêt d'une analyse agrégée de l'échantillon général et d'une véritable représentation des trajectoires individuelles. Plusieurs observations complémentaires peuvent être faites sur notre exemple fictif. D'abord, les dix types de trajectoires individuelles les plus fréquents correspondent à 62% de la population d'étude. Ensuite, les parcours sont effectivement unidirectionnels et les trajectoires les plus fréquentes sont composées d'une transition d'un type d'espace à un autre en passant par la catégorie d'espace intermédiaire en milieu de parcours.

Le dernier exemple de représentation des données longitudinales a été développé récemment par les concepteurs de la bibliothèque *TraMineR* (figure 5). Il s'inspire du type de visualisation *en coordonnées parallèles* présenté plus haut dans la section 2.1 et il est baptisé « parallel coordinate plot for sequence data » (Bürgin, Ritschard, 2014). L'objectif est de mettre en évidence des configurations récurrentes en agrégeant les mêmes trajectoires ou parties de trajectoires, tout en conservant leur aspect individuel. Dans notre exemple fictif, la représentation révèle des concentrations individuelles plus ou moins fortes selon les modalités et les moments de l'étude (ex : rural et urbain en début de période, intermédiaire en milieu de période). Elle met aussi en évidence les trajectoires les plus fréquentes, mais celles-ci sont difficiles à déterminer et à hiérarchiser avec précision.

Comme pour tout mode de visualisation, il y a un équilibre à trouver entre la conservation de l'information et la lisibilité graphique. La conservation de l'information individuelle des trajectoires se fait ici au prix d'une perte de lisibilité générale, et notamment des tendances agrégées. Ce mode de visualisation s'avère ainsi peu adapté à la représentation de gros échantillons de données longitudinales, en particulier lorsque les trajectoires individuelles sont complexes (nombreuses modalités d'états ou d'évènements, et forte hétérogénéité des comportements individuels). D'un point de vue sémiologique, ce graphique ne peut pas être satisfaisant ; en effet, il n'est pas possible de trouver une palette de couleurs correcte permettant de distinguer plus de trente modalités distinctes.

En définitive, la représentation *en chronogramme* (figure 2) donne une vue transversale de la distribution des états à chaque pas de temps, mais elle perd la dimension longitudinale des données. La représentation classique *en tapis* (figure 3) ne peut, en restant vraiment lisible, s'appliquer qu'à une partie de l'échantillon d'étude. La représentation *en tapis*, qui met en avant les trajectoires les plus



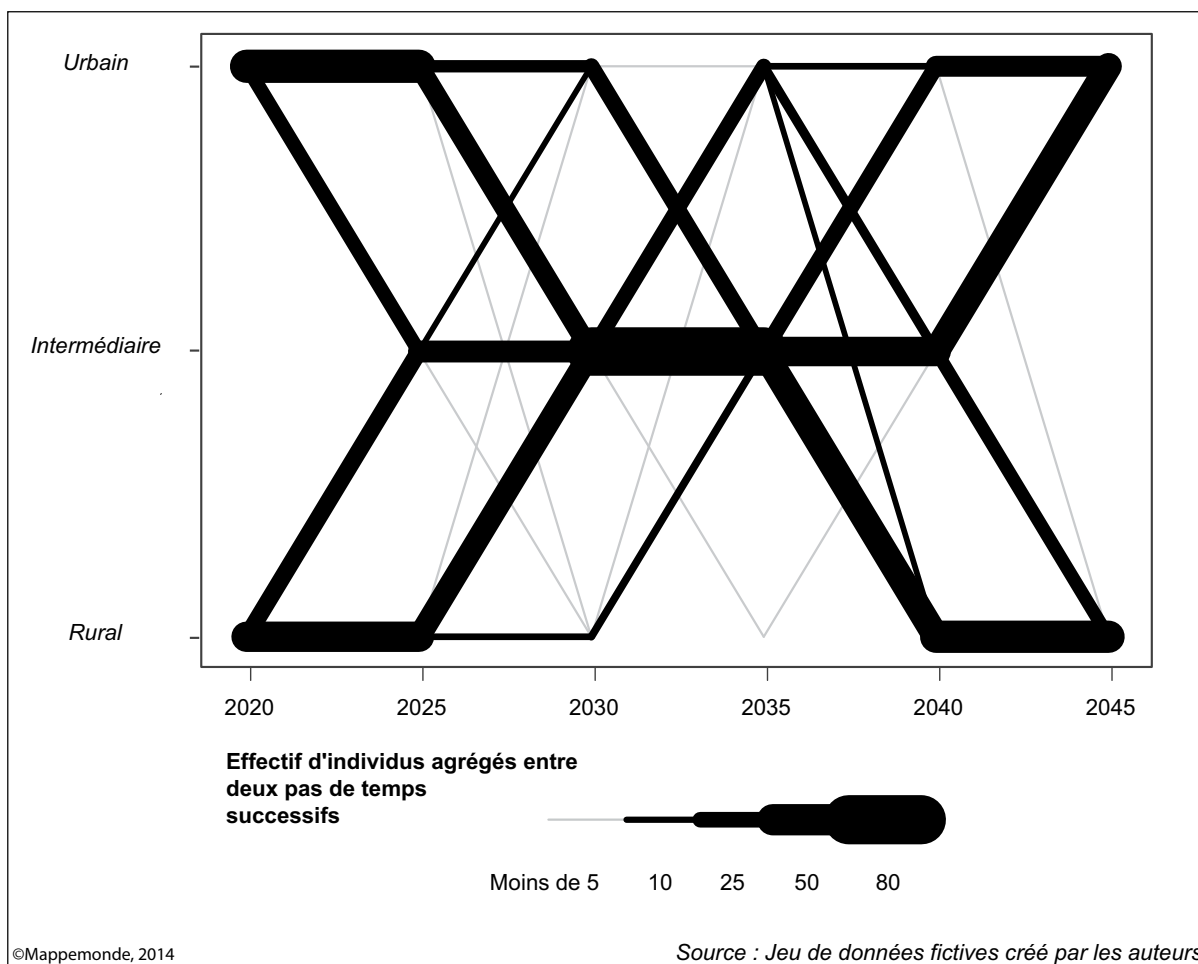
5. Exemple de visualisation en coordonnées parallèles adaptées aux données longitudinales.

fréquentes (figure 4), ne permet pas de repérer les états les plus fréquents à un pas de temps donné, ou les changements d'état les plus fréquents entre deux pas de temps successifs. Enfin, la représentation *en coordonnées parallèles* adaptées aux données longitudinales (figure 5) élimine plusieurs de ces limites. Mais la conservation de l'aspect individuel des trajectoires se fait alors aux dépens de la lisibilité des tendances lourdes, qui émanent de toutes les trajectoires individuelles. Dans la lignée du dernier exemple, le mode de visualisation *en coulées*, ci-dessous proposé, comble certains de ces manques.

2.3. Proposition du mode de visualisation graphique *en coulées*

La visualisation *en coulées* repose sur la représentation agrégée des parties communes de trajectoires individuelles et des changements ou continuités d'état entre deux pas de temps successifs (figure 6). Elle ne permet pas nécessairement de dégager les trajectoires les plus fréquentes (figure 4, figure 5), mais elle permet d'identifier aisément les tendances lourdes et d'évaluer le poids des changements ou des continuités d'états au cours de la période d'étude. Deux principales lectures sont possibles. La première consiste à dégager une ou plusieurs tendances pour l'ensemble de l'échantillon, lesquelles correspondent à la combinaison de différents types de trajectoires individuelles intégrales. La seconde consiste à détecter les relations privilégiées entre deux ou plusieurs états, à des moments spécifiques ou sur toute la période d'étude.

Ce mode de visualisation graphique repose sur le croisement de deux variables qualitatives. En abscisse, figure un axe temporel discret constitué de pas de temps successifs. En ordonnée, figure une variable qualitative nominale ou ordinale, qui correspond aux modalités de la variable étudiée. Le cœur du graphique est composé de parties de trajectoires individuelles agrégées au moyen d'un figuré linéaire dont l'épaisseur est proportionnelle à l'effectif d'individus concernés par tel changement ou telle continuité d'états entre deux pas de temps successifs.



6. Exemple de visualisation en coulées sur l'ensemble de l'échantillon.

Dans l'exemple des parcours résidentiels fictifs, la visualisation graphique *en coulées* est d'une aide précieuse pour compléter l'analyse. Les catégories « Rural » et « Urbain » sont bien les plus fréquentes en début et en fin de période, recueillant chacune la moitié de la population. La catégorie « Intermédiaire » se peuple progressivement en milieu de période et se dépeuple ensuite. Les décalages de temporalité dans les parcours résidentiels sont bien visibles et la hiérarchie des espaces l'est également : très peu de parcours résidentiels court-circuitent la hiérarchie urbaine. En définitive, la visualisation *en coulées* permet de dégager des tendances lourdes à partir d'un entrelacs de trajectoires individuelles hétérogènes. Dans notre exemple fictif, l'hétérogénéité est assez faible, mais sur des données réelles le mode de visualisation proposé prendra tout son intérêt.

La confrontation des observations aux principaux modes de représentation (*en chronogramme*, *en tapis*, *en coordonnées parallèles*), et à celui *en coulées*, souligne l'intérêt de ce dernier et sa complémentarité avec les modes existants. En outre, la diversité de ses champs d'application et des types d'individus étudiés, esquissée dans la troisième partie, renforce son intérêt pour la représentation de données longitudinales variées.

La section suivante présente un éventail d'utilisations (figure 7) : elle commence par la mobilité géographique d'individus sociologiques qui circulent entre des types d'espaces (migrations résidentielles). Puis, elle s'intéresse à des changements de catégorie opérés par des individus sociologiques (changement du mode de transport

	Mobilité géographique	Changement de catégories
Individus sociologiques	Section 3.1	Section 3.2
Individus géographiques	–	Section 3.3

7. Triptyque d'exemples d'application du graphique en coulées en géographie.

principal pour se rendre au lieu de travail). Enfin, elle montre l'utilité du mode de visualisation *en coulées* pour analyser les changements de catégorie observés sur des individus géographiques (caractérisation des communes au regard des flux de navettes domicile-travail).

3. Exemples de visualisation en coulées de données longitudinales

Cette dernière partie présente plusieurs exemples de visualisation graphique *en coulées* sur divers objets de recherche habituels des géographes : les mobilités résidentielles (section 3.1), les modes de transport (section 3.2), les dynamiques communales d'emploi et de résidence (section 3.3). Ces exemples s'appuient sur des sources statistiques variées : l'enquête Histoire de vie INSEE/INED de 2003, la *Puget Sound Transportation Panel Survey* de la région de Seattle de 1989 à 1999, et le Recensement français de la population depuis 1968.

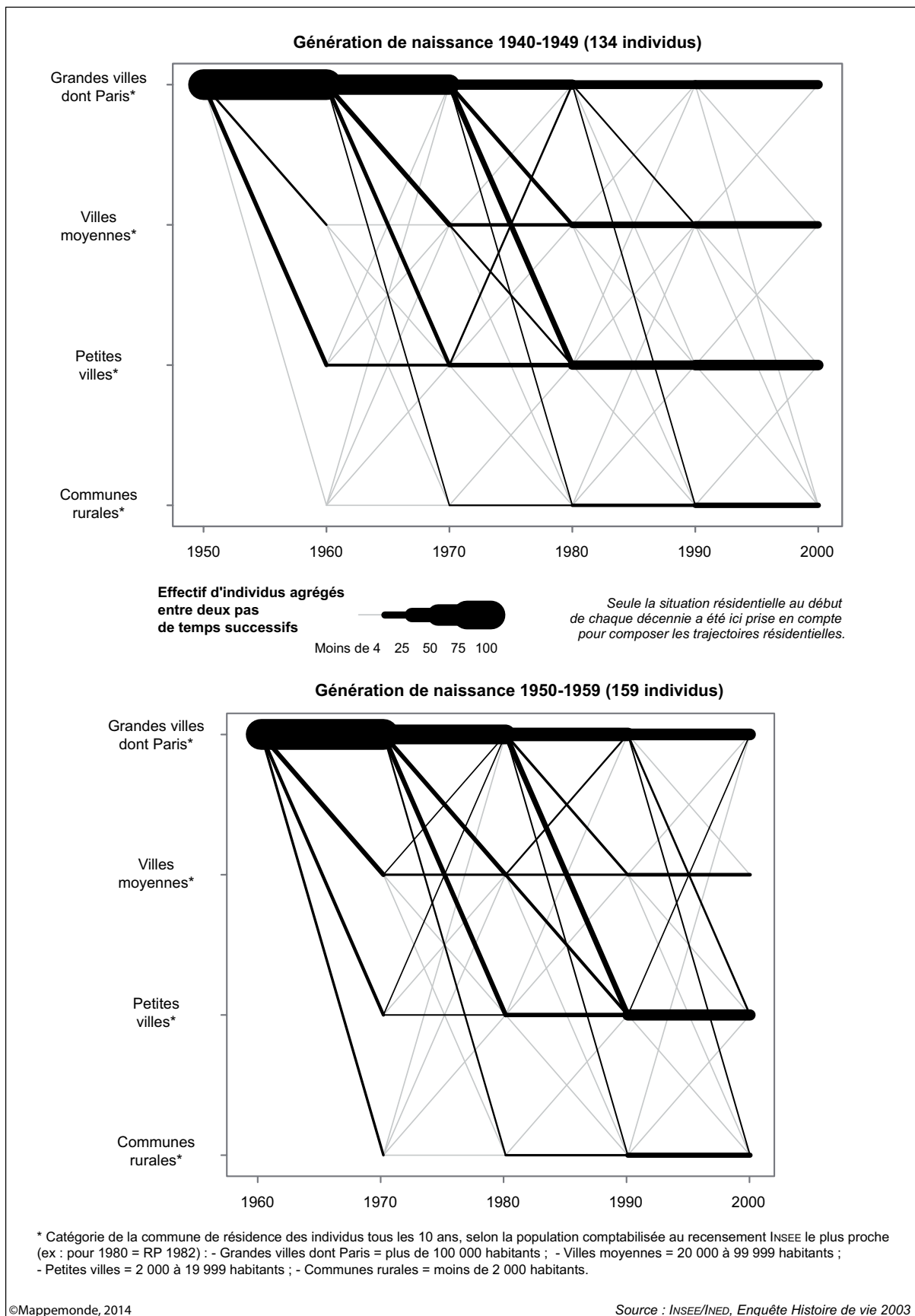
Avec ces trois exemples il s'agit de montrer les spécificités et les intérêts du mode de visualisation proposé, ainsi que la variété de ses champs d'application : de l'étude des trajectoires spatiales effectuées par des individus sociologiques à l'analyse des changements d'états d'individus géographiques, en passant par l'examen des changements d'états d'individus sociologiques. Chaque exemple est traité en associant des considérations méthodologiques et un bref commentaire thématique.

3.1. Migrations et trajectoires résidentielles au départ des grandes villes françaises

Le premier exemple traite des mobilités résidentielles d'individus sociologiques entre quatre catégories spatiales (figure 8). Cette utilisation du mode de visualisation *en coulées* est l'une des plus intuitives : la trajectoire y est conçue comme le suivi dans le temps de la mobilité géographique d'individus sociologiques. Il s'agit ainsi d'analyser les étapes successives de résidence des individus d'un échantillon donné, en visualisant à la fois les changements de types de lieux de résidence et les phases de continuité résidentielle. L'exemple s'intéresse plus précisément aux comportements d'individus qui ont en commun d'avoir débuté leur trajectoire résidentielle dans une grande ville française.

L'échantillon utilisé émane de l'enquête Histoire de vie, réalisée en 2003 en collaboration principale entre l'INSEE et l'INED. Cette enquête s'intéresse à la construction des identités individuelles à l'articulation de dimensions familiales, professionnelles, sociales ou encore spatiales (Crenner *et al.*, 2006 ; Guérin-Pace *et al.*, 2009). Une grille biographique a été intégrée dans le questionnaire d'enquête et a permis de recueillir, rétrospectivement, un grand nombre de trajectoires résidentielles (échantillon de 8 403 individus de 18 ans et plus).

Deux générations de naissance sont ici étudiées : les individus nés dans les années 1940-1949 (134 individus) et les individus nés dans les années 1950-1959



8. Trajectoires résidentielles des individus de deux générations de naissance au départ des grandes villes françaises.

(159 individus), et seuls ceux qui ont débuté leur trajectoire résidentielle dans une grande ville ont été sélectionnés. En outre, quatre grands types de lieux de résidence ont été distingués : les grandes villes, dont Paris (100 000 habitants et plus), les villes moyennes (de 20 000 à 99 999 habitants), les petites villes (de 2 000 à 19 999 habitants) et les communes rurales (moins de 2 000 habitants). Enfin, les pas de temps pris en compte donnent une photographie de la situation résidentielle des individus à chaque début de décennie, de 1950 ou 1960 à 2000.

Les visualisations *en coulées* des deux générations d'individus présentent plusieurs ressemblances. Les tendances lourdes sont assez semblables : le nombre d'individus qui résidaient encore dans une grande ville au début des années 2000 est minoritaire (respectivement 24,6% et 34,5%). De plus, le basculement vers des tranches urbaines de taille inférieure (villes moyennes et surtout petites villes) s'est effectué principalement en début de période d'activité professionnelle et de mise en ménage (autour de 20-30 ans). Les retours vers les grandes villes ont été peu nombreux durant la décennie 1990, à l'inverse des départs vers des communes rurales qui ont été assez réguliers au cours de toute la période d'étude : en 2000, 20,9% des individus de la génération 1940-1949 résidaient dans une commune rurale, et 19,5% des individus de la génération 1950-1959. En définitive, la visualisation *en coulées* est ici d'une aide précieuse pour déterminer les principales tendances résidentielles et pour évaluer les périodes du cycle de vie où ont eu lieu les basculements entre types d'espaces.

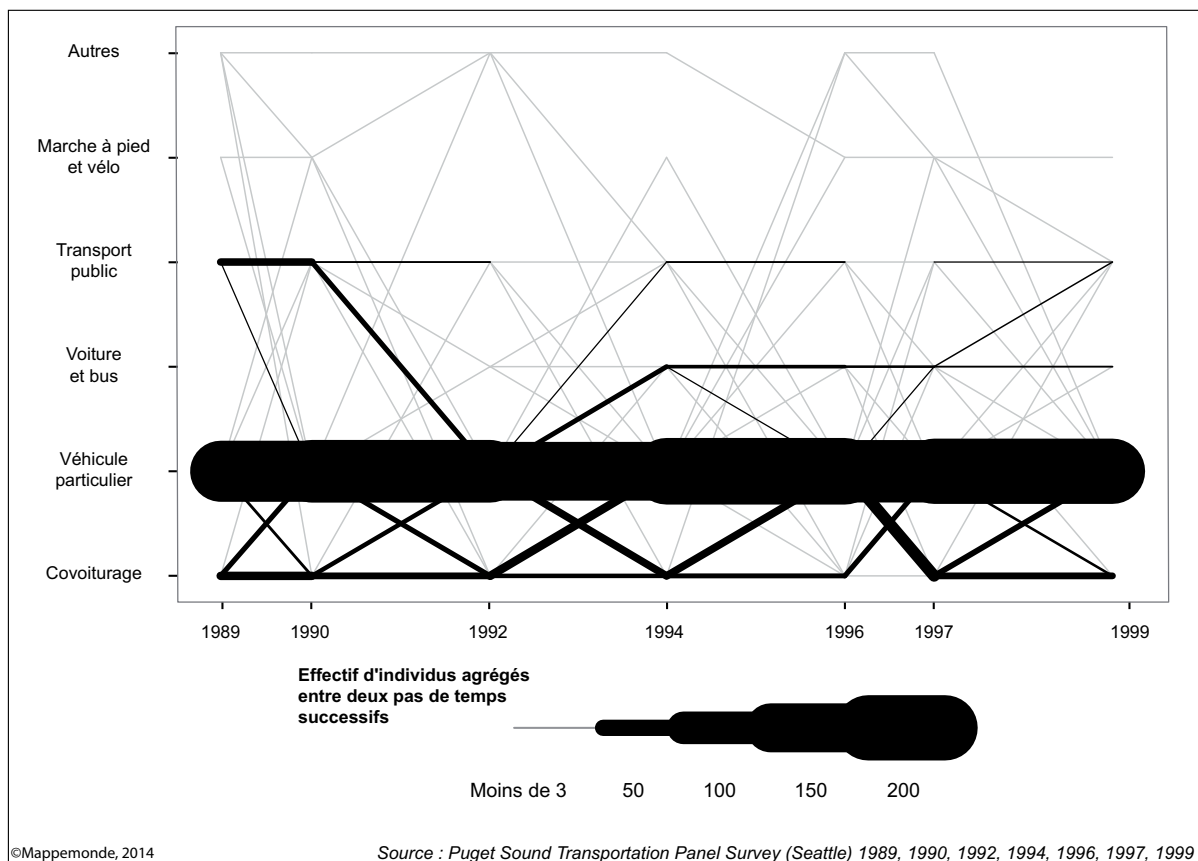
3.2. Transport et trajectoires modales dans la région de Seattle

Le deuxième exemple concerne aussi les trajectoires d'individus sociologiques et des continuités ou changements d'états parmi les modalités d'une même variable (figure 9). Cependant, il ne s'agit pas de mobilités effectives entre des types d'espaces, mais de modifications des pratiques de transport enregistrées par une variable qualitative.

Cet exemple de visualisation *en coulées* utilise les données de l'enquête de mobilité de la région de Seattle (PSTPS – *Puget Sound Transportation Panel Survey*), qui est un modèle d'enquête panel dans le domaine de la mobilité et des transports. Celle-ci a suivi un échantillon d'environ 1 800 ménages de 1989 à 2000, avec une nouvelle vague d'enquête en moyenne tous les deux ans (Murakami *et al.*, 1992). La variable sélectionnée pour cet exemple renseigne sur le mode de transport principal utilisé par un individu pour aller au travail. Elle est construite sur un échantillon de l'enquête PSTPS (286 individus), qui n'est constitué que des individus actifs occupés et renseignés tout au long de la période (c'est-à-dire non affectés par l'érosion du panel), et qui ont utilisé leur véhicule individuel ou le covoiturage à au moins une des dates d'enquête (1989, 1990, 1992, 1994, 1996, 1997, 1999).

L'objectif est de suivre l'évolution des pratiques de mobilité quotidienne en considérant le mode de transport principal utilisé pour se rendre au travail. Deux questions peuvent être posées à ce sujet. La première concerne les individus eux-mêmes : sont-ils turbulents, changeant sans cesse de mode de transport, ou bien restent-ils fidèles à un même mode de transport ? La seconde concerne les relations entre les modalités de la variable étudiée : les transferts sont-ils nombreux et plutôt uni- ou bidirectionnels ?

Au cours de la période 1989-1999, la proportion d'individus utilisant le covoiturage diminue (de 13,4 à 10,9%), alors que leur proportion augmente pour l'usage du véhicule individuel (de 70,8 à 76%). Ce résultat agrégé confirme une tendance lourde



9. Trajectoires modales des individus actifs occupés dans la région de Seattle.

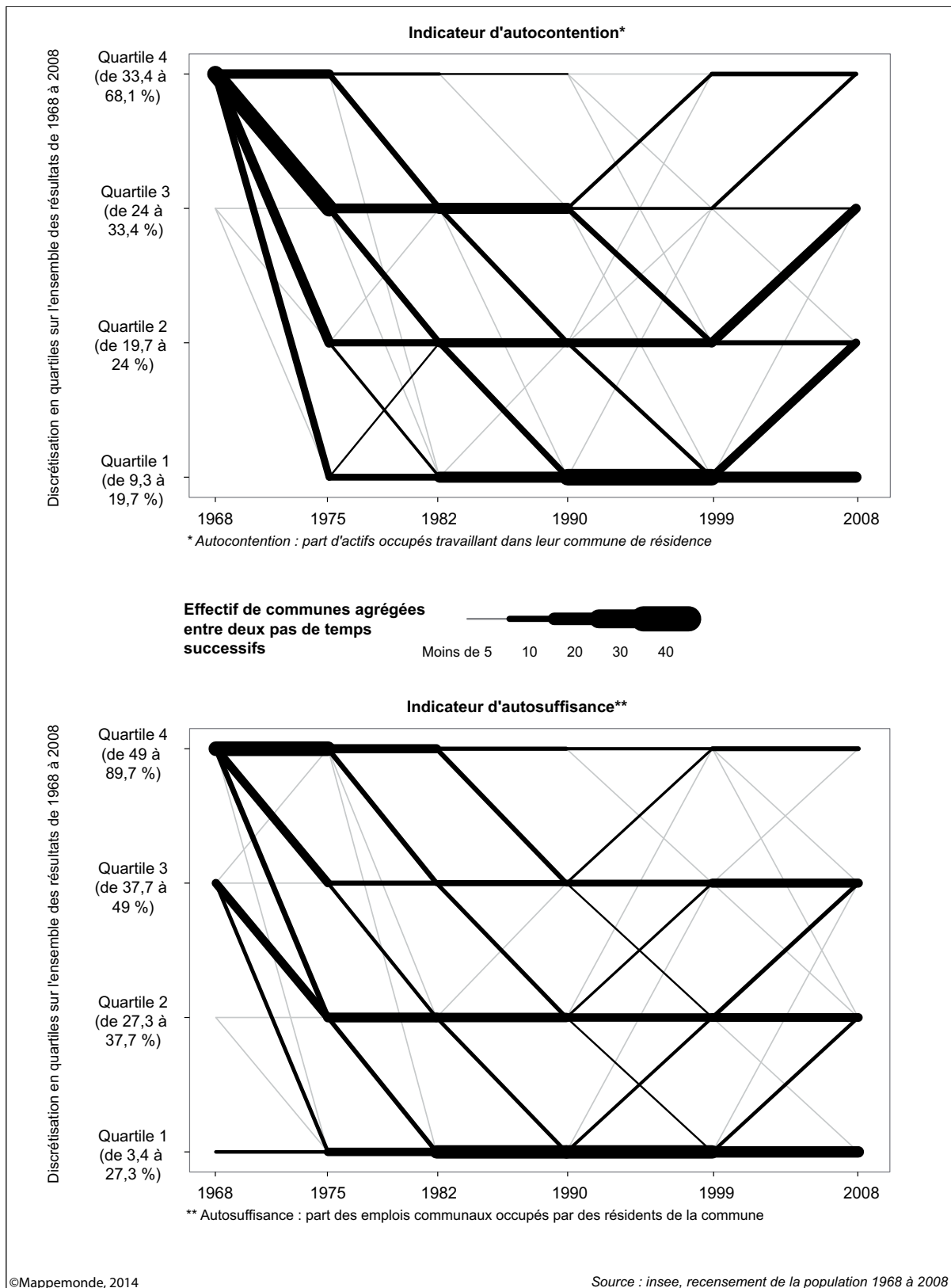
qui peut aussi bien être observée par comparaison de deux enquêtes transversales. Cependant, l'association d'un dispositif d'enquête panel et de la visualisation *en coulées* présente au moins deux intérêts supplémentaires. Le premier concerne la mise en évidence de tendances générales, qui émanent de la conjonction des séquences composant les trajectoires individuelles. Dans le cas présent, la visualisation dessine un patron spécifique : le passage entre le véhicule individuel et le covoiturage est clairement bi-directionnel, l'intensité de changement est presque similaire. Pour le second intérêt, la porosité entre les catégories covoiturage et véhicule individuel contraste avec l'étanchéité entre elles des catégories covoiturage et autres modalités de transport. Il n'y a par exemple aucun transfert important entre covoiturage et transport public : dans la région de Seattle, la niche de report modal vers le covoiturage se situe avant tout chez les utilisateurs d'un véhicule individuel.

3.3. Emploi, résidence et trajectoires communales à Paris et en petite couronne

L'exemple présenté dans cette dernière section ne concerne plus des individus sociologiques, mais des individus géographiques. Considérant des unités spatiales comme des individus géographiques, une même démarche d'analyse longitudinale peut être appliquée à l'évolution de tous types d'unités spatiales. En ce sens, les données produites par des dispositifs transversaux deviennent longitudinales lorsque l'on raisonne au niveau agrégé d'une unité spatiale, à condition que les limites de ces unités restent stables sur la période d'étude.

Les communes de Paris et de la petite couronne (départements 75, 92, 93, 94), soit 143 communes, constituent les individus géographiques étudiés (figure 10). Les données sont tirées de six Recensements de la Population (RP 1968 à 2008). Si le

recensement n'est pas à proprement parler un dispositif de recueil de données longitudinales, il le devient indirectement lorsque l'individu considéré est une unité spatiale, elle-même agrégat d'individus sociologiques.



10. Trajectoires communales d'autocontention et d'autosuffisance à Paris et en petite couronne.

Les lieux de travail et les lieux de résidence, enregistrés dans le recensement de la population, sont plus précisément utilisés pour cet exemple francilien. Deux indicateurs ont été calculés selon les définitions suivantes :

- l'autocontention correspond au pourcentage d'actifs occupés travaillant dans leur commune de résidence ;
- l'autosuffisance correspond à la part des emplois dans la commune occupés par des résidents de la commune.

Ces deux indicateurs sont à la fois liés et distincts, d'où l'intérêt de confronter leurs résultats. Certaines communes enregistrent une forte autocontention et une forte autosuffisance : c'est par exemple le cas de plusieurs arrondissements périphériques parisiens, en particulier du Nord-Est (18^e, 19^e, 20^e), avec des valeurs avoisinant les 50% au RP 2008 pour les deux indicateurs. D'autres communes franciliennes présentent une forte autocontention et une faible autosuffisance, l'exemple le plus marquant étant celui du 8^e arrondissement de Paris. Enfin, en petite couronne parisienne, une majorité de communes présentent de faibles valeurs, autour de 20% pour les deux indicateurs.

Dernier point méthodologique : après calcul des indicateurs d'autocontention et d'autosuffisance pour chaque commune, aux six dates de recensement, les résultats ont été discrétisés en quartiles sur l'ensemble de la série temporelle. En effet, la visualisation *en coulées* n'est valide que si les états et les modalités statistiques ont été définis de manière stable au cours du temps. Une discrétisation indépendante à chaque pas de temps ne serait ainsi pas acceptable.

La représentation des trajectoires d'évolution communale des deux indicateurs dessine un patron semblable. En 1968, la majorité des communes se trouvent dans le quartile supérieur, soit plus de 33,4% pour le premier indicateur et plus de 49% pour le deuxième. De 1968 à 1990, les communes enregistrent une baisse relative de l'autocontention et de l'autosuffisance. Mais, à partir de 1990, la dynamique tend à se stabiliser et à s'inverser pour la plupart des communes, et pour les deux indicateurs étudiés. Comment interpréter ces tendances ? Le levier explicatif principal renvoie au décalage temporel entre desserrement de l'habitat et desserrement de l'emploi observé en région parisienne. Dans un premier temps, les communes de la petite couronne ont accueilli une périurbanisation résidentielle, de courte et moyenne portée, une majorité d'emplois continuant d'être concentrés dans les arrondissements parisiens. Dans un second temps, le desserrement des emplois a entraîné des implantations de plus en plus nombreuses dans les communes de la petite couronne et, de fait, une moindre dépendance vis-à-vis de Paris. Ce phénomène, qui mériterait un examen bien plus approfondi, n'est d'ailleurs pas propre à la région parisienne : il a en effet été observé dans de nombreuses aires urbaines françaises (Chalonge, Beaucire, 2007).

Perspectives géographiques d'analyse et de visualisation de données longitudinales

Les trois exemples présentés dans la dernière partie de l'article correspondent à trois démarches d'utilisation et de mise en forme graphique de données longitudinales. Les deux premiers exemples traitent les pratiques spatiales d'individus sociologiques (types successifs d'espaces de résidence, évolution du mode de transport dominant) et font partie des utilisations les plus classiques d'un dispositif d'analyse longitudinal.

Le troisième exemple s'intéresse aux trajectoires d'évolution d'individus géographiques (dynamiques communales de l'emploi et de la résidence). Ce dernier, plus original, ouvre un grand nombre de perspectives de recherche en géographie pour les raisons non exhaustives suivantes :

- l'analyse des trajectoires d'évolution d'individus géographiques ne requiert pas nécessairement le recours à des données longitudinales. Elle peut notamment être menée par l'exploitation des résultats agrégés du recensement de la population à une échelle géographique donnée ;

- l'évolution d'une grande variété d'unités spatiales et d'indicateurs statistiques, dans de nombreux champs thématiques, gagnerait à être analysée sous le prisme d'une approche longitudinale, afin de mieux prendre en compte les trajectoires individuelles ;

- l'analyse de données longitudinales ne se limite pas aux modes de visualisation graphique : plusieurs méthodes statistiques d'analyse exploratoire et de modélisation (analyse séquentielle, modèle de survie ou de durée) paraissent également pertinentes pour l'étude des trajectoires d'évolution d'individus géographiques.

En outre, le mode de visualisation *en coulées* proposé dans cet article, qui relève de l'analyse exploratoire des données longitudinales en géographie et dans les sciences humaines, vient compléter l'offre existante de représentation graphique (*en chronogramme, en tapis, en coordonnées parallèles*). D'autres formes de visualisation notamment cartographiques, ainsi que le développement d'indicateurs statistiques de synthèse, peuvent encore améliorer l'utilisation de ces données particulièrement riches pour l'analyse de phénomènes spatio-temporels.

Bibliographie

- BÜRGIN R., RITSCHARD G. (2014). «A Decorated Parallel Coordinate Plot for Categorical Longitudinal Data», *The American Statistician*, 68:2, 98-103, DOI:10.1080/00031305.2014.887591
- BONNAFOUS A., PATIER-MARQUE D., PLASSARD F. (1981). *Mobilité et vie quotidienne*. Lyon : Presses universitaires de Lyon, 172 p. ISBN : 2-7297-0103-6
- BONVALET C. (2009). « L'enquête "Peuplement et Dépeuplement de Paris" », dans Groupe de réflexion sur l'approche biographique, *Biographies d'enquêtes - Bilan de 14 collectes biographiques*. Paris : INED, coll. « Méthodes et savoirs », p. 105-127. ISBN : 978-2-7332-6007-4
- CAMPBELL D.T, STANLEY J.C. (1966). *Experimental and quasi-experimental designs for research*. Chicago : Rand Mc Nally, 84 p. ISBN : 0-5286-1400-2
- CHALONGE L., BEAUCIRE F. (2007). « Le desserrement des emplois au sein des aires urbaines ». *Les Annales de la Recherche Urbaine*, n°102, p. 97-102.
- COMMENGES H., PISTRE P., CURA R. (2014). « SLIDER : Software for Longitudinal Data Exploration with R ». *Cybergeog : European Journal of Geography*. [En ligne], GeOpenMod, document 693, mis en ligne le 07 novembre 2014. URL : <http://cybergeog.revues.org/26530> ; DOI : 10.4000/cybergeog.26530
- COUET C. (2006). « L'échantillon démographique permanent de l'INSEE ». *Courrier des Statistiques*, n°117-119, p. 5-14.
- COURGEAU D., LELIÈVRE É. (1989). *Analyse démographique des biographies*. Paris : INED, 268 p., ISBN : 2-7332-2010-1

- CRENNER E., DONNAT O., GUÉRIN-PACE F., HOUSEAUX F., VILLE I. (2006). « L'élaboration d'une enquête quantitative sur la construction des identités ». *Économie et Statistique*, n°393-394, p. 7-18.
- DESROSIÈRES A. (1993). *La Politique des grands nombres : histoire de la raison statistique*. Paris : La Découverte, 440 p. ISBN : 2-7071-2253-X
- DOLL R. (2001a). « Cohort studies. History of the method I. Prospective cohort studies ». *Sozial- und Präventivmedizin*, vol. 46, n°2, p. 75-86.
- DOLL R. (2001b). « Cohort studies. History of the method II. Retrospective cohort studies ». *Sozial- und Präventivmedizin*, vol. 46, n°2, p. 152-160.
- DUNCAN G.J., JUSTER T., MORGAN J.N. (1987). « The role of panel studies in research on economic behavior ». *Transportation Research Part A*, vol. 21, n°4-5, p. 249-263.
- DURAND-DASTÈS F. (2001). « Les concepts de la modélisation en analyse spatiale ». In Sanders L. (dir.), *Modèles en analyse spatiale*. Paris : Lavoisier, p. 31-59. ISBN : 2-7462-0320-0
- GABADINHO A., RITSCHARD G., MÜLLER N.S., STUDER M. (2011a). « Analyzing and visualizing state sequences in R with TraMineR ». *Journal of Statistical Software*, vol. 4, n°4, p. 1-37.
- GABADINHO A., RITSCHARD G., STUDER M., MÜLLER N. S. (2011b). *Mining sequence data in R with the TraMineR package : À user's guide*. Genève : Université de Genève, 128 p. [en ligne : <http://mephisto.unige.ch/traminer/index.shtml>]
- GRELET Y. (2002). « Des typologies de parcours : méthodes et usages ». *Notes de travail Génération 92*, n°20, 47 p.
- GUÉRIN-PACE F., SAMUEL O., VILLE I. (dir.) (2009). *En quête d'appartenance. L'enquête Histoire de vie sur la construction des identités*. Paris : INED, coll. « Grandes enquêtes », 224 p. ISBN : 978-2-7332-8001-0
- INSELBERG A. (2008). « Parallel Coordinates : Visualization, Exploration and Classification of High-Dimensional Data ». In CHEN C.-H., HÄRDLE W., UNWIN A. (eds), *Handbook of Data Visualization*. Cleveland : Springer, p. 643-680. ISBN : 978-1-2811-8034-6
- LAZARSELD P.F., FISKE M. (1938). « The panel as a new tool for measuring opinion ». *Public Opinion Quarterly*, vol. 2, n°4, p. 596-612.
- LELIÈVRE E., VIVIER G. (2001). « Évaluation d'une collecte à la croisée du quantitatif et du qualitatif. L'enquête Biographies et entourage ». *Population*, vol. 56, n°6, p. 1043-1073.
- MARTINI E. (2005), « Comment Lind n'a pas découvert le scorbut ». *Histoire des Sciences Médicales*, vol. 39, n°1, p. 79-92.
- MURAKAMI E., WATTERSON W.T. (1992). « The puget sound transportation panel after two waves ». *Transportation*, vol. 19, n°2, p. 141-158.
- PISTRE P. (2011). « Life course and biographical analysis of residential migrations to French rural areas ». Actes de la conférence *European Colloquium on Quantitative and Theoretical Geography*, Athènes, 2-6 septembre 2011.
- PISTRE P. (2012). *Renouveaux des campagnes françaises : évolutions démographiques, dynamiques spatiales et recompositions sociales*. Paris : Université Paris Diderot (Paris 7), Thèse de doctorat de géographie, 407 p. [en ligne : <http://halshs.archives-ouvertes.fr/tel-00764869/>]
- ROBETTE N. (2011). *Explorer et décrire les parcours de vie : les typologies de trajectoires*. Paris : CEPEP, 86 p. ISBN : 978-2-8776-2184-7

- TARIS T.W. (2000). *A primer in longitudinal data analysis*. Londres : SAGE Publications, 163 p., ISBN : 978-1-4129-3358-2
- THÉVENIN T., CHARDONNEL S., COCHEY E. (2007). « Explorer les temporalités urbaines de l'agglomération de Dijon : une approche activité-centrée de l'Enquête Ménage Déplacement ». *Espace Populations Sociétés*, vol. 2-3, p. 179-190.

Adresses des auteurs

Hadrien Commenges, Université Paris 1 Panthéon-Sorbonne, UMR 8504 Géographie-cités. Courriel : hcommenges@parisgeo.cnrs.fr

Pierre Pistre, Institut français des sciences et technologies des transports, de l'aménagement et des réseaux (IFSTTAR), Aménagement, mobilités et environnement (AME), Dynamiques Économiques et Sociales des Transports (DEST), Associé à l'UMR 8504 - Géographie-cités. Courriel : pierre.pistre@parisgeo.cnrs.fr