

Rapport final du projet Harmonie-cités (ANR Corpus et Outils de la Recherche en SHS édition 2007)

Anne Bretagnolle, Marianne Guérois, Hélène Mathian, Fabien Paulus, Céline Vacchiani-Marcuzzo, François Delisle, Liliane Lizzi, Thomas Louail, Sébastien Martin, Elfie Swerts

▶ **To cite this version:**

Anne Bretagnolle, Marianne Guérois, Hélène Mathian, Fabien Paulus, Céline Vacchiani-Marcuzzo, et al.. Rapport final du projet Harmonie-cités (ANR Corpus et Outils de la Recherche en SHS édition 2007) : Bases de données harmonisées sur la dynamique et les compétences des villes en réseau selon les régions du monde. 2011. halshs-00688491

HAL Id: halshs-00688491
https://shs.hal.science/halshs-00688491

Submitted on 17 Apr 2012

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.

Rapport final du projet Harmonie-cités

A. Identification

Programme – année	Corpus et outils de la recherche en sciences humaines et sociales Edition 2007
Projet (acronyme)	Harmonie-cités
Titre complet du projet	Bases de données harmonisées sur la dynamique et les compétences des villes en réseau selon les régions du monde
Coordinateur du projet Nom, laboratoire, organisme de rattachement	Bretagnolle Anne, UMR Géographie-cités (8504), Université Paris 1 Panthéon-Sorbonne
Période du projet (date début – date fin)	28 décembre 2007-27 juillet 2011

Rédacteur de ce rapport

Civilité, prénom, nom	Mme Anne Bretagnolle*
Téléphone	01 40 46 40 01
Adresse électronique	anne.bretagnolle@parisgeo.cnrs.fr
Date de rédaction	Septembre-Octobre 2011

*Co-rédacteurs de ce rapport (par ordre alphabétique) :
François Delisle, Marianne Guérois, Liliane Lizzi, Thomas Louail, Hélène Mathian, Sébastien Martin, Fabien Paulus, Céline Vacchiani-Marcuzzo, Elfie Swerts.
*Contributeurs : Sylvie Dubuc, Gabriel Vatin, Sébastien Oliveau, Denise Pumain, Céline Rozenblat, Lena Sanders, Adrien Van Hamme

B. Rapport scientifique

B.1 - Résumé du travail accompli

Le projet Harmonie-cités vise à construire des bases de données cohérentes sur les villes, pour la comparaison dynamique et internationale. Les réflexions sur l'harmonisation des bases partent du constat que l'agencement des définitions de l'urbain, des maillages géométriques, des sources statistiques, ou l'architecture des bases de données sont autant de paramètres qui conditionnent fortement les résultats donnés par les indicateurs et les modèles utilisés pour rendre compte des évolutions des systèmes urbains dans le monde. Les bases construites ou expertisées portent sur des populations, des activités et des réseaux de transport inter-urbains, dans plusieurs régions du monde et dans des durées de temps longues. Ces régions sont très différenciées et représentatives, en quelque sorte, de la diversité des systèmes de villes dans le monde : le cas de la France et de l'Europe renvoient à des systèmes de villes anciens et relativement stables, celui de l'Afrique du Sud et des Etats-Unis à des systèmes récents, tout en étant très différents du point de vue du niveau de développement. L'Inde et la Chine représentent enfin des systèmes anciens mais perturbés par un choc colonial.

Un premier chantier a consisté à réfléchir sur la conception et la formalisation de modèles de données génériques, permettant de suivre de manière harmonisée les objets villes dans le temps long et dans une variété de contextes géographiques. Des modèles de données permettant une harmonisation à la fois longitudinale et transversale ont été proposés, à partir d'un travail d'expertise de bases déjà existantes et à partir d'un approfondissement des méthodes d'harmonisation en fonction du budget-temps d'une heure, qui avaient été élaborées pour les villes des Etats-Unis en 2006 et 2007 et qui ont été automatisées dans le cadre de l'ANR. Concernant les bases de données sur les délimitations et les populations des villes, plusieurs résultats importants ont été obtenus. Pour les Etats-Unis (1790-2000), la base a été élargie, incluant désormais, en plus des grandes villes, les petites comprises entre 2500 et 10 000 habitants. Un travail de validation a été mené sur la base Urban Morphological Zones (produite de manière quasiment automatique à partir de la base CORINE Land cover) et permet son utilisation pour suivre l'évolution des villes d'Europe depuis 1990 et pour les années futures. Pour la France, la construction d'agglomérations historiques réalisée manuellement en 1982 dans le Fichier Historique de l'INED (1830-1954) est en cours d'automatisation. La base sur l'Afrique du Sud a été complétée pour la fin du 19ème siècle et les bases sur l'Inde et la Chine sont en cours de réalisation, en collaboration avec d'autres projets de recherche. Les bases de données sur les activités ont été construites de manière harmonisée pour les Etats-Unis, la France et l'Afrique du Sud de 1960 à nos jours. Les premières exploitations, transversales et longitudinales, montrent de fortes proximités entre ces trois pays (structuration hiérarchique des systèmes, trajectoires fonctionnelles des villes). L'articulation des bases de données sur les réseaux de transport inter-urbains, d'une part, et sur les villes, d'autre part, a nécessité un travail important de réflexion sur la mise en cohérence de deux types de bases de données (appariement et généralisation cartographique). La numérisation du réseau ferroviaire sud-africain a été réalisée entre 1860 et 1960, celle du réseau ferroviaire des Etats-Unis (1830-1960) est en cours et pour la France une collaboration a été réalisée avec d'autres projets de recherche. Enfin, une interface permettant d'interroger de manière dynamique et interactive les bases

de données constituées dans le projet a été construite de manière générique et implémentée pour la France et les Etats-Unis.
Les retombées à court terme de l'ANR Harmonie-cités sont importantes, comme en témoignent le nombre de publications, communications dans des colloques internationaux ou rapports qui ont déjà été réalisés (17 références). L'investissement mené par l'équipe dans la construction de modèles conceptuels génériques permet d'envisager des retombées à plus long terme, et pas seulement pour les projets menés au sein du laboratoire. En outre, les perspectives pour les 5 ans à venir sont riches, puisque les projets qui n'ont pas été achevés dans le cadre de l'ANR sont actuellement poursuivis à travers trois projets de recherche internationaux (ESPON Data Base, l'ERC Geodivercity et l'ANR Transmondyn). En conclusion, cette ANR a marqué un tournant fort pour un laboratoire spécialisé depuis longtemps dans la construction et le suivi de bases de données urbaines : l'accent mis sur la conception de modèles de données génériques, sur l'utilisation de formalismes eux aussi génériques et sur des méthodes automatisées permet de construire des bases de données caractérisées par une traçabilité complète.

The project Harmonie-cités aims to build coherent databases for cities, for dynamic and international comparison. It started from the statement that the choice of definitions, geometrical meshes, statistical sources or database architectures are basic parameters that largely influence the different results given by indicators or by urban system modelling. The databases that have been constructed or evaluated focus on populations, activities and inter-urban transportation networks, in several regions of the world and in the long term period. These regions are very different and representative, in a sense, of the diversity of urban systems in the world: the cases of France and Europe are related to old and quite stable urban systems, the cases of South Africa and United States of America illustrate recent urban systems (with a very different level of economic development). India and China represent old urban systems that have been modified by an external choc (trading post then colonisation).
A first task consisted to work on the conception and formalisation of generic data models, that could allow following in a harmonized way the urban entities on the long term period and in a variety of geographical contexts. Different data models allowing longitudinal and transversal comparability have been proposed, starting from an expertise of a bunch of current urban databases (constructed by researchers and census bureaux) and from a deepening of harmonization methods based on the one-hour time-budget, elaborated for the U.S. data base built by the team research in 2006-2007 and that have been automated in the frame work of the project Harmonie-cités.
A second task was dedicated to delineations and populations of cities. Several results have been obtained. For United States (1790-2000), the data base has been enlarged and includes now, in addition to cities larger than 10 000 inhabitants, the smaller containing more than 2500 inhabitants. For Europe, a validation protocol has been applied to the database UMZ (Urban Morphological Zones, constructed quite automatically from Corine Land cover by the European Environment Agency). It now allows the use of UMZ for working on European urban settlements since 1990 to 2006, with the perspective of future updates of the database by EEA. Concerning France, the construction of historical agglomerations (1830-1954) that had been done manually in 1982 by INED (Institut National d'Etudes Démographiques) has been revised using automatic methods. The database for South Africa has been

completed for the end of the 19th century, and databases concerning India and China are currently in process, based on collaboration with other research projects.
A third task concerns databases on urban activities, constructed in a harmonized way. The first explorations of the results, transversal (France, U.S., South Africa) and longitudinal (1960-2000), reveal very strong proximities between the three countries (hierarchical structure of the urban system, co-evolution of functional trajectories of the cities). The fourth task, dedicated to inter-urban exchanges, gave rise to a deep reflection on the alignment of the two type of databases that are involved (matching and cartographic generalisation). The digitalization of the south-african railroad network has been conducted for one century (1860-1960), the one concerning U.S. is in progress and for France a collaboration with another research project has been engaged. Lastly, a web interface allowing a dynamic and interactive interrogation of the different databases constructed in Harmonie-cités has been realized. This interface is generic and has been first implemented for France and United States.
The short-term impacts of the ANR are important, as reflected by the number of publications, communications in international colloquiums and reports already realized (17 references). The investment of the team in the construction of conceptual generic models should also produce long-term effects, to be expected not only for projects concerning the laboratory itself. Furthermore, the future perspectives for the next 5 years are rich, as the projects which are still in course are under the scope of three international projects (ESPON Data Base, the ERC Geodivercity and the ANR Transmondyn). In order to conclude, this ANR marked a turning point in a research laboratory that has been specialized for a long time in the construction and the following of urban databases: the focus on the conception of generic database models, on the use of generic formalisms and automated methods allows constructing databases characterized by a full traceability.

B.2 - Rapport scientifique proprement dit

1. Enjeux, situation du sujet, objectifs et problématique

Les enjeux soulevés par l'élaboration de bases de données cohérentes pour la comparaison dynamique et internationale sont importants. Les statistiques proposées par l'ONU, qui constituent la référence pour la plupart des chercheurs, aménageurs ou hommes politiques, ne sont pas harmonisées et résultent d'un simple assemblage de définitions nationales des villes. Comment interpréter alors le « franchissement du seuil de 50% d'urbains dans le monde » (récent, selon l'O.N.U), les chiffres alarmistes de l'accroissement de certaines mégapoles du Sud (8 à 10% par an en Afrique Noire, toujours selon l'O.N.U), l'attention extrême portée au rythme de la croissance des grandes villes dans les pays anciennement industrialisés (avec des expressions qui reviennent régulièrement depuis une trentaine d'années, telles que « contre-urbanisation », « renouveau des petites villes », et plus récemment « renouveau rural ») ? La sensibilité des mesures à la nature et à la qualité des bases de données utilisées pour les comparaisons mondiales ou longitudinales appelle une réflexion sur la notion d'harmonisation : l'agencement des définitions de l'urbain, des maillages géométriques, des sources statistiques, l'architecture des bases de données sont autant de paramètres qui conditionnent fortement les résultats donnés par les indicateurs et les modèles utilisés pour rendre compte des évolutions des systèmes urbains dans le monde.

L'harmonisation est néanmoins une opération complexe, qui ne repose pas uniquement sur la qualité et le soin apportés à la construction de bases de données. Elle dépend aussi largement des résultats d'une expertise, menée par les spécialistes du champ urbain dans des contextes géographiques et historiques variés. D'une région du monde à une autre, des décalages s'opèrent depuis deux siècles dans les transitions démographiques et urbaines ainsi que dans les phases de développement économique. En raison de la diversité des formes de peuplement dans le monde (types d'habitat, densités, organisation des transports, nature des activités et des institutions de gouvernance), l'analyse dynamique du poids des villes et de leur rayonnement ne peut s'affranchir d'une prise en compte, dans la définition même de ce qu'est une ville, de la variabilité des contextes économiques, historiques et culturels.

Menée dans un laboratoire de recherche associant depuis plusieurs décennies des spécialistes de la géomatique et des spécialistes du champ urbain dans différentes régions du monde, l'ANR Harmonie-cités s'est donnée pour objectif de construire, améliorer ou développer des bases de données harmonisées sur les populations, les activités et les réseaux de transport inter-urbains, dans plusieurs régions du monde et dans des durées de temps longues (un à deux siècles pour les populations et les réseaux, 3 à 4 décennies pour les activités économiques). Ces régions sont très différenciées et représentatives, en quelque sorte, de la diversité des systèmes de villes dans le monde : le cas de la France et de l'Europe renvoient à des systèmes de villes anciens et relativement stables, celui de l'Afrique du Sud et des Etats-Unis à des systèmes récents, tout en étant très différents du point de vue du niveau de développement. L'Inde et la Chine représentent enfin des systèmes anciens mais perturbés par un choc colonial.

L'ANR a été structurée en quatre volets.

Le premier porte sur les délimitations et les populations des villes. Selon les époques et les régions du monde, les enveloppes correspondant aux contours des villes peuvent être construites à partir d'une pluralité d'approches conceptuelles (la ville comme entité juridique, comme entité morphologique, comme aire fonctionnelle définie par le trajet des navetteurs…). Une manière de choisir et d'harmoniser les définitions dans le temps et dans l'espace consiste à se fonder sur un invariant anthropologique, celui du budget temps d'une heure de transport consacré en moyenne par les habitants des villes aux trajets domicile-travail, avec des variations selon le niveau hiérarchique de la ville (budget temps plus important pour les grandes villes que pour les petites). Les bases de données ainsi construites doivent reposer sur des modèles de données multi-niveaux (trois niveaux sont requis : l'entité élémentaire composant la ville, la ville elle-même et le système de villes) et dynamiques (c'est-à-dire permettant de suivre dans le temps long l'évolution des maillages administratif, les changements de niveaux de maillage dans le suivi de l'objet ville, les changements dans les définitions mêmes des objets villes). Ces bases doivent ensuite être testées, validées ou corrigées par des analyses de croissance et des modélisations des hiérarchies urbaines.

Le deuxième volet concerne l'analyse de la diversité fonctionnelle des systèmes de villes. Cette diversité est le produit de l'inégale répartition des activités économiques présentes dans les villes, donnant à chacune ou à un sous-ensemble une place spécifique dans la production de richesse. L'enjeu est de comprendre dans quelle mesure les différenciations de ces profils économiques urbains résultent en partie de la diffusion hiérarchique, selon la taille des villes, de grands cycles

d'innovation (comme la première révolution industrielle, la deuxième centrée sur l'automobile et l'électricité ou plus récemment les Nouvelles Technologies de l'Information et de la Communication). Ce second volet découle du premier et utilise des délimitations urbaines précises et cohérentes, et pour des systèmes de villes différents (Europe, France, Afrique du Sud, Etats-Unis). L'objectif est de rassembler au niveau des villes l'information éparse sur leurs activités et les professions, ou d'en proposer des approximations, en s'appuyant si possible sur le travail d'harmonisation des nomenclatures issues des différents pays. Les travaux portent sur la collecte systématique des descripteurs généraux des fonctions urbaines sur le long terme. A l'aide des d'analyses factorielles multi-tableaux, les notions de générations de villes et les cycles de croissance associés aux innovations doivent ensuite être utilisées et testés de manière récursive dans cette démarche. Les paramètres des lois d'échelle (scaling) doivent permettre d'affiner les comparaisons entre systèmes de villes.

Le troisième volet met l'accent sur les réseaux d'échange. Les travaux récents sur les villes confirment qu'il n'est pas possible de comprendre le devenir d'une ville à partir de ses seuls constituants, sa population, ses activités ou sa gouvernance. Les villes ne sont pas des entités isolées, elles ont besoin pour fonctionner et pour évoluer d'entretenir des relations avec les autres villes. La connaissance sur les réseaux d'échanges qui ont participé à la structuration des systèmes de villes est cependant faible, en raison de la difficulté à rassembler les sources et construire des bases de données harmonisées des transports, dans le temps et dans l'espace. Parmi les réseaux d'échanges structurants, les transports ferroviaires qui se sont mis en place tout au long du 19ème siècle constituent une innovation majeure, dont les effets ont davantage fait l'objet de suppositions que de mesures. Pourtant, à partir de la numérisation des réseaux ferroviaire, il est possible d'évaluer la position relative de chaque ville dans le système, de manière locale (indices de nodalité) ou globale (indices d'accessibilité de Shimbel, par exemple). L'étude des liens entre l'accessibilité des villes et leur dynamique démographique ouvre un large champ d'analyses et de modélisations (notamment par des systèmes multi-agents) sur la nature des co-évolutions entre réseaux de transport et systèmes de villes.

Enfin, le quatrième et dernier volet consiste dans la construction d'une interface, permettant d'interroger de manière dynamique et interactive les bases de données ainsi constituées. L'originalité du projet est de choisir comme niveau de référence celui du système de villes. En effet, la plupart des interfaces existant à ce jour montrent des évolutions de villes au cours du temps sans s'intéresser aux liens d'interdépendance avec des ensembles plus vastes (systèmes de villes) ni aux complémentarités qui peuvent exister entre des dimensions différentes des villes (population et emprise spatiale, caractéristiques socio-économiques, position relative dans des réseaux d'échanges). L'objectif de l'interface Harmonie-cités est de mettre l'accent sur une approche multi-scalaire et multi-dimensionnelle de la ville, tout en gardant un niveau de généricité important pour permettre son utilisation pour différentes époques et différentes régions du monde.

2. Matériels (terrains, sources, données…), méthodes et travaux réalisés

2.1 Conceptualisation des modèles de données, automatisation des processus, réflexions sur la notion d'harmonisation

Un premier chantier consistait à réfléchir sur la conception et la formalisation de modèles de données génériques, permettant de suivre de manière harmonisée les objets villes dans le temps long et dans une variété de contextes géographiques. Ces bases de données spatio-temporelles sont particulièrement complexes car elles intègrent généralement plusieurs niveaux d'échelle (l'objet ville étant composé d'une agrégation de d'entités élémentaires, par exemple des unités de collecte des recensements) et portent sur des objets évolutifs. Suivre une ville dans une durée relativement longue suppose en effet que l'on fasse évoluer au cours du temps la définition même d'un objet qui a été profondément modifié par les révolutions dans les techniques de transport (passant de la commune éponyme que l'on parcourait à pied jusqu'au début du 19ème siècle à l'aire urbaine fonctionnelle actuelle sillonnée par des réseaux de transport rapides sur plusieurs dizaines de kilomètres). En outre, les géométries, voire la nature même des entités élémentaires a parfois considérablement varié aussi au cours du temps (changement de type de mailles, changements individuels de type fusion ou scission etc.).

Deux approches ont été suivies. D'une part, nous avons mené une réflexion sur la notion d'harmonisation, à la fois longitudinale (comment rendre comparables dans le temps long des villes d'un même pays ou d'une même région) et transversale (comparabilité de villes provenant de pays ou de régions différentes). Ces réflexions ont été alimentée par l'étude d'un certain nombre de bases de données dites « harmonisées », utilisées aujourd'hui par des chercheurs dans différentes régions du monde. D'autre part, nous sommes partis des méthodes d'harmonisation en fonction du budget-temps d'une heure, élaborées sur la base de données des villes des Etats-Unis en 2006 et 2007 (Bretagnolle e al. 2008), afin de les tester sur d'autres régions du monde et d'automatiser les procédés de construction d'agglomérations lorsque cela était envisageable (stage de Master 2 Pro Carthageo de Gabriel Vatin, 2010).

2.2 Bases de données sur les Populations et les Espaces urbains

Initialement conçu pour s'étendre sur la première année de l'ANR, le volet Population et Espaces Urbains s'est progressivement étoffé et a finalement porté sur les trois années du projet, soit en raison de l'arrivée de forces supplémentaires (Elfie Swerts, doctorante, démarrant en 2010 une thèse sur l'Inde et la Chine), soit parce que l'articulation des bases de données sur les villes et sur les réseaux de transport, analysée dans la troisième année du projet, a révélé les insuffisances de certaines bases qui avaient été construites avant le démarrage de l'ANR (c'est le cas de la France et des Etats-Unis, mais aussi, de manière plus ponctuelle, de l'Afrique du Sud).

Concernant **l'Europe** (pour laquelle nous rappelons qu'il n'existe, à notre connaissance, aucune base de données permettant de suivre dans le temps long les populations des quelques 5000 agglomérations urbaines de plus de 10 000 habitants en 2000), le projet envisageait d'articuler trois bases de données couvrant des segments temporels différents :

- 1200-1850 : base de Paul Bairoch et ses collaborateurs (Bairoch et al. 1987)

- 1950-1990 : base Géopolis de François Moriconi-Ebrard (Moriconi-Ebrard, 1994)
- 1990-2006 : base *Urban Morphological Zones*, de l'Agence Européenne de l'Environnement (EEA, 2002, Milego 2007)

Cette articulation reposait notamment sur deux préalables : d'une part recueillir des données sur les agglomérations européennes entre 1850 et 1950, à partir des recensements de population dans les différents pays concernés ; d'autre part, s'assurer de la pertinence de la base UMZ pour qualifier le fait urbain en Europe à la fin du 20ème siècle, cette base n'ayant encore jamais été utilisée en ce sens par des experts.

L'expertise concernant la période 1850-1950 a été menée à partir de l'étude des recensements de population des différents pays européens, conservés à la Bibliothèque Nationale. A partir du choix d'une date intermédiaire, autour de 1900, ces sources ont systématiquement été dépouillées (lorsque la langue ou la disponibilité du document le permettait) par Jean-Baptiste Arrault. Les recherches ont porté sur les types de définition utilisés par chaque pays (juridique, morphologique, statistique…), les briques de base composant la ville lorsque celle-ci procèdait d'une agrégation d'entités, le nombre de « villes » (agglomérations, municipalités) de plus de 10 000 habitants et la présentation des tableaux de données.

Concernant la base *Urban Morphological Zones*, la démarche d'expertise a consisté à s'assurer de la pertinence d'objets villes reconstitués à partir d'une approche hautement automatisée. Les *UMZ* reposent en effet sur l'application systématique d'un même algorithme aux espaces bâtis extraits de la base CORINE Land cover, sans opération de validation *a posteriori*. Par comparaison, les bases nationales ou européennes qui émanent de l'application de standards internationaux (comme par exemple le critère de discontinuité de bâti inférieure à 200m) intègrent en aval une phase d'expertise pouvant conduire à des réajustements ponctuels, en particulier dans le cas délicat des conurbations. Il importait donc de confronter les *UMZ* aux agglomérations définies par les offices statistiques de quelques Etats européens, suivant des critères proches (Danemark, France, Suède), afin de saisir et d'interpréter les écarts observés entre les *UMZ* et des agglomérations nationales par ailleurs bien documentées. C'est dans cette perspective qu'un protocole de comparaison entre les deux types de bases a été mis en œuvre. Il procède d'une part de la mise en correspondance des objets de la base *UMZ* et de ceux des bases nationales, d'autre part de la confrontation d'indicateurs fondamentaux du niveau d'urbanisation (nombre, population, surface des villes), qui permet de préciser le degré de convergence des deux bases ainsi que de détecter les formes d'urbanisation à l'origine des principaux écarts.

La construction d'une base de données harmonisée pour la **France** n'était pas prévue initialement. Néanmoins, un volet a été ajouté au printemps 2011 pour permettre l'articulation de la base de données sur les transports ferroviaires en France (collaboration avec Thomas Thévenin, voir section C2) et le fichier de l'urbanisation historique de l'INED (Pumain et Riandey 1986), principale source utilisée jusque là par les chercheurs travaillant sur l'évolution de la croissance urbaine dans le temps long (voir les travaux de France Guérin Pace, Denise Pumain, Fabien Paulus etc.). Entre 1831 et 1954, des communes sont regroupées dans une agglomération si elles dépassent une certaine population (2500 habitants

agglomérés au chef-lieu), si elles sont contiguës à la commune-centre ou à une commune déjà intégrée et si elles font partie d'une même unité urbaine au recensement de 1962. Construit de manière manuelle dans les années 1980, ce fichier comporte un certain nombre d'incertitudes pour les petites villes, le critère d'agglomération morphologique n'ayant pas toujours été respecté lors de sa création. Grâce à l'utilisation de la base numérique historique 1801-2001 des communes de France (Motte, Séguy, There 2003), nous avons pu estimer à environ 10% la part des communes figurant par erreur dans le fichier. La collaboration menée avec Claude Motte (voir section C2) nous a permis d'avoir une extraction de la base historique 1801-2001 des communes de France pour les 5954 communes qui composent les unités urbaines de la France métropolitaine en 1999 et les 345 communes urbaines avant 1999 apparaissant dans le fichier de l'INED. Un premier travail d'appariement des bases historiques de l'INED et de celle de Claude Motte a été effectué, car la codification des communes qui ne sont plus urbaines de nos jours différait entre les deux bases. Dans le même temps, une campagne de saisie aux archives de l'INSEE a été lancée au printemps 2011 (complétée, pour les dates les plus anciennes, par des relevés à la Bibliothèque Nationale et à l'INED, et par des relevés à Strasbourg, pour les départements de l'Est et du Nord-Est de la France, annexés à l'Allemagne fin 19ème).

Comme pour la France, la construction d'une base de données harmonisée n'était pas prévue pour les **Etats-Unis**, ce travail ayant été réalisé les deux années précédant le début de l'ANR (Bretagnolle, Giraud, Mathian 2008). Néanmoins, un volet a été là aussi ajouté en 2010 lorsque la campagne de numérisation du réseau ferroviaire a été programmée. En effet, le seuil minimal qui avait été utilisé lors de la construction de la base pour le 19ème siècle (8000 habitants en 1860 puis 10 000 habitants à partir de 1870) est trop élevé pour une analyse de la dynamique des petites villes en lien avec l'établissement du réseau. Une deuxième campagne utilisant le seuil minimal de 2500 habitants tout au long du 19ème (soit le même seuil que pour la France) et 5000 habitants dans la première moitié du 20ème siècle a été programmée en 2010. Elle repose, en outre, sur l'utilisation des méthodes de construction automatiques d'agglomérations mises en place dans le stage de Gabriel Vatin (voir section précédente).

La base de données de la population des agglomérations urbaines en **Afrique du Sud** est antérieure au projet ANR pour la période 1911-2001 (Vacchiani-Marcuzzo Céline, 2005) et a fait l'objet de plusieurs publications dont un Atlas aux éditions IRD (Giraut, Vacchiani-Marcuzzo, 2008). Le cadre de l'ANR a cependant permis de lancer une campagne de collecte des données de populations dans des recensements plus anciens, échelonnés entre 1865 et 1904 selon les provinces.

Pour **l'Inde**, la construction d'une base historique d'évolution de la population indienne des villes de plus de 10 000 habitants (entre 800 et 2001) a été envisagée, en collaboration avec le projet e-geopolis[1], dans le cadre de la thèse d'Elfie Swerts, portant sur la comparaison de l'évolution des systèmes de villes indiennes et chinoises[2]. Cette collaboration visait à éviter les redondances et homogénéiser les résultats.

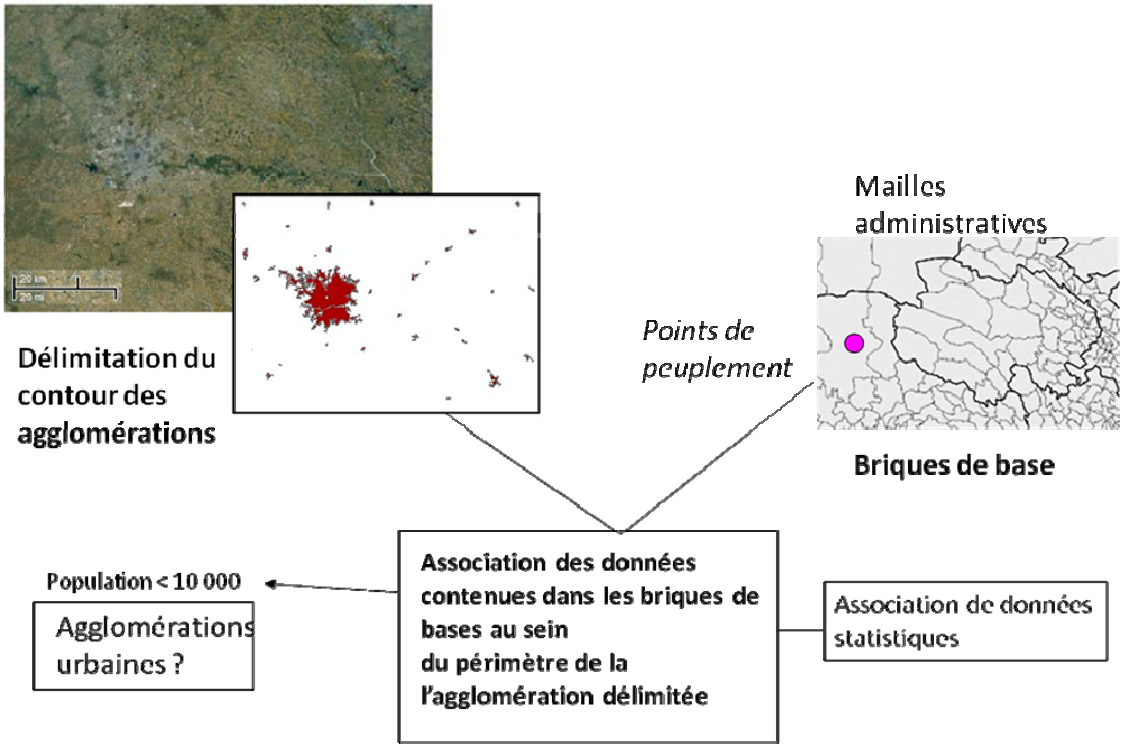
[1] Url e-géopolis : http://e-geopolis.eu/

[2] Elfie Swerts, *Evolution des systèmes de villes en Inde et en Chine : Observation et modélisation statistique et géographique, comparaison*, sous la direction de Denise Pumain et Eric Denis.

Entre 800 et 1750, les données sont disponibles tous les 50 ans, et tous les dix ans entre 1750 et 2001. La base devrait être complétée pour l'année 2010 une fois que l'ensemble des données seront parues.

De 800 à 1891, les données proviennent des publications du *census of India* et ont été complétées par les bases construites par Chandler et Fox (1974) et par Elisée Reclus (1876-1894). Afin que la base soit cohérente et robuste malgré la diversité des sources, divers tests, comme la mesure du taux de croissance des villes entre différents pas de temps ou de leur trajectoire tout au long de cette période, ont été effectués sur les données. De 1901 à 1951, les données proviennent d'une base construite par François Moriconni Ebrard puis successivement complétée par Sylvie Dubuc (1995) et Joël Quercy (2010). Enfin, de 1961 à 2001, les données démographiques ont été produites par l'équipe de l'Institut Français de Pondichéry dans le cadre du projet Indiapolis et comprennent les chiffres de population de 6 000 villes de plus de 10 000 habitants. Ces données ont été harmonisées en éliminant les fluctuations dues aux définitions administratives du pays. Pour ce faire, pour chaque ville, la population issue des données des recensements officiels ont été associée au sein d'un système d'information géographique à un contour délimité sur des critères morphologiques à l'aide d'images satellites (Figure 1). Plusieurs séjours à l'Institut Français de Pondichéry (IFP) ont été effectués pour coordonner ces travaux et mettre au point un protocole comparable pour réaliser la base de données sur les villes chinoises.

Figure 1 : Méthodologie de construction de la base de données de population des villes indiennes de 1961 à 2001 (Swerts, 2011)



2.3 Bases de données sur les activités et sociétés urbaines

La construction des bases de données statistiques sur les fonctions urbaines, prévue initialement en 2ème année du projet, s'est poursuivie aussi en 3ème année. Elle a été menée en se focalisant sur les activités économiques afin d'identifier, sur une période de temps suffisamment longue d'une cinquantaine d'années, l'émergence

d'activités nouvelles et leurs localisations spécifiques dans les systèmes de villes. Nous avons également choisi, lorsque cela était possible, de collecter des informations sur les catégories sociales et professionnelles. En effet, saisir les lieux de l'innovation, c'est aussi considérer les populations les plus innovantes, notamment en termes de qualification. Les tâches à effectuer étaient les suivantes : 1) Construire les bases de données sur les fonctions urbaines (dans 3 pays : Afrique du Sud, Etats-Unis, France) et sur une cinquantaine d'années ; 2) Identifier des grands types de fonctions et qualifier les positions relatives des villes dans les systèmes respectifs ; 3) Analyser les dynamiques du changement urbain, en nous appuyant notamment sur les comparaisons entre systèmes. A cette fin, un travail d'harmonisation des nomenclatures d'activités, à la fois dans le temps et entre les différents pays, a été produit.

Un préalable à la réalisation de ces tâches était de disposer des résultats du premier volet sur les délimitations des villes. C'est pourquoi, pour l'Europe, le travail s'est focalisé essentiellement sur la France.

2.3.A – Sources et collecte des données

La première étape a consisté à collecter les données les plus détaillées possibles. Plusieurs sources ont permis cette collecte, souvent complexe selon les pays. Les tableaux ci-dessous résument les données collectées, les sources et les années par pays.

Les recensements décennaux de la population aux **Etats-Unis** dénombrent la population active suivant les nomenclatures d'activités économiques et des professions en vigueur aux dates des recensements et à une échelle fine (*census tracts, places, counties*). Les informations de l'avant-dernier recensement, celui de 2000, sont accessibles en ligne sur le site du U.S. Census Bureau. Notons que la diffusion des données du recensement 2010 doit s'achever en 2012. Les données des recensements plus anciens sont également en libre accès sur ce site, mais sous une forme non immédiatement exploitable (images des livrets publiés pour chacun des recensements). Nos investigations ont permis d'identifier des équipes de chercheurs américains qui ont amorcé un travail de compilation des données des recensements et de diffusion dans des formats informatiques directement exploitables. Ainsi, l'ICPSR (*Interuniversity Consortium for Political and Social Research*), projet d'archivage des données en sciences sociales de *l'Institute for Social Research* de l'Université du Michigan, a construit une base de données historique où un premier travail d'harmonisation des informations, notamment géographiques, a été mené. Cependant, cette base de données a le défaut de proposer des données économiques et sociales très agrégées (une dizaine de postes seulement).

Une autre source s'est alors révélée essentielle pour le projet : le NHGIS (*National Historical Géographic Information System*, mis en place par l'Université du Minnesota). Il s'agit d'un projet de mise à disposition de l'ensemble des données des recensements de la population américaine de 1790 à 2000 dans des cadres spatiaux du plus fin (census tract lorsque disponible, comtés dans tous les cas) jusqu'aux Etats mais aussi de la mise à disposition des fonds cartographiques associés. Ces données sont accessibles sous divers formats depuis mars 2007.

La possibilité d'exploiter tous les recensements et donc de construire une base de données des fonctions urbaines sur deux siècles, s'est heurtée au niveau de détail des nomenclatures. Ainsi, les nomenclatures utilisées jusque dans les années 1950

ne comptent que quelques postes. Elles ne sont très détaillées depuis 1960. Le tableau résume ces informations.

Tableau 1 : Données disponibles pour les Etats-Unis à l'échelle du county (et parfois subcounty) pour les Etats-Unis et niveaux de détail des nomenclatures (Paulus, Vacchiani-Marcuzzo 2011)

	Année	Source	Niveau de détail de la nomenclature
Industry	1960	NHGIS	25
Occupations	1960	NHGIS	12 ou 231
Industry	1970	NHGIS	41
Occupations	1970	NHGIS	42
Industry	1980	NHGIS	49
Occupations	1980	NHGIS	70
Industry	1990	NHGIS	72
Occupations	1990	NHGIS	94
Industry	2000	NHGIS	80
Occupations	2000	Census	46

Nota-Bene : Secteurs économiques : *industry* ; Catégories sociales : *occupation*

Pour **l'Afrique du sud**, la collecte a nécessité un travail de terrain afin de récupérer et de photocopier les recensements (format papier de 1960 à 1980) au département des publications gouvernementales de la bibliothèque de l'Université de Cape Town et les données numérisées au bureau de *Stats SA* pour les années 1980 (missions Baffi, 2010 et Vacchiani-Marcuzzo, 2011).

Tableau 2 : Sources pour la collecte des données d'activités et de catégories sociales des villes sud-africaines (Solène Baffi 2010-2011)

	Année	Source	Maillage	Niveau de détail de la nomenclature
Occupation	1960	Census	Province/ magisterial district	16
			Metropolitan area (14)	11 ou 70
Industry	1960	Census	Province/ magisterial district	12
			Metropolitan area (14)	11 ou 59
Occupation	1970	Census	Province/ magisterial district	9
			Metropolitan area (13)	9
Industry	1970	Census	Province/ magisterial district	10
			Metropolitan area (13)	11
Industry	1980	Census	Metropolitan area/ Statistical region/ suburb	11
Occupation	1980	Census	Metropolitan area/ Statistical region/ suburb	12
Industry	1996	Census	Transitional Local Councils (TLC) et	11
Occupation	1996	Census	Transitional Metropolitan Councils (TMC)	20
Industry	2001	Census	Main Place/ Sub place	11
Occupation	2001	Census	Main Place/ Sub place	29

Pour la **France**, une base de données sur les activités économiques des aires urbaines a été construite précédemment par Fabien Paulus, pour la période 1962-1999 (Paulus, 2004). De plus, les informations du recensement de 2006, diffusées

en 2009, ont été introduites. Un travail de conception d'une base de données temporelle sur les catégories socioprofessionnelles a été mené.

2.3.B – Harmonisation des nomenclatures

La deuxième étape concerne l'harmonisation des différentes nomenclatures au cours du temps et entre les différents pays en cherchant la mise en concordance en termes de niveau de détail et de signification des postes. Ce travail a été fait sur les activités économiques et non sur les catégories sociales, trop différentes à la fois dans le temps et dans l'espace (fortes diversités nationales des spécialisations sociales relatives des villes). En revanche, une analyse croisée des catégories sociales urbaines a été menée pour l'année 2000.

Un travail important a été mené pour harmoniser les nomenclatures des **Etats-Unis** (pays sur lequel les analyses ont demandé le plus d'investigation). Les emplois ont été dénombrés suivant deux nomenclatures, la SIC (*Standard Industrial Classification*) et la NAICS (*North American Industry Classification*). La SIC a été développée dans les années 1930 et utilisée pour le recensement de 1940. Cette nomenclature a été actualisée pour chaque recensement décennal jusqu'à l'entrée en vigueur en 1997 de la NAICS (utilisée également au Canada et au Mexique et construite comme une déclinaison de l'ISIC, *International Standard Industrial Classification*), laquelle est utilisée pour le recensement de 2000. Si le changement de nomenclature correspond à une nécessaire prise en compte des mutations économiques, et en particulier le déclin de l'emploi industriel au profit des activités de services, le *Census Bureau* propose une table de correspondances entre ancienne et nouvelle nomenclatures qui permet, à un niveau d'agrégation intermédiaire (une vingtaine de secteurs), de reconstruire des séries relativement cohérentes dans le temps.

Quant aux professions, la nomenclature utilisée est la SOC (*Standard Occupational Classification*) du *Bureau of Labor Statistics*. Mise à jour régulièrement, elle distingue en 2000 23 catégories professionnelles dans son niveau le plus agrégé.

Armés de ces outils, nous avons pu construire une base de données pour les Etats-Unis pour environ 900 villes (Metropolitan et micropolitan areas) et pour 5 dates, de 1960 à 2000. Le niveau de détail possible dans la construction d'une nomenclature harmonisée des activités économiques nous a conduit à distinguer deux bases : l'une en 25 postes pour la période 1960-2000 et l'autre de 40 postes pour 1970-2000. L'harmonisation des nomenclatures d'activités avait été menée pour la base de données France avant ce programme ANR. Celui-ci a été l'occasion d'approfondir ce travail d'harmonisation dans la perspective des comparaisons internationales. Ainsi, par exemple, des regroupements de postes élémentaires ont été proposés afin de mieux saisir les services avancés aux entreprises ou les activités de communication, comme c'est le cas dans la nomenclature nord-américaine. Enfin, pour l'Afrique du Sud, le travail s'est révélé le plus simple. En effet, les informations disponibles, est ventilée dans une dizaine de postes qui restent globalement cohérents à toutes les dates.

2.4 Bases de données sur les réseaux de transport inter-urbains

Avant même d'envisager la numérisation de réseaux de transport inter-urbains dans les différents pays concernés, un premier chantier concernait la question de la mise en relation de la dynamique des villes avec celles des réseaux. La mise en cohérence de deux bases de données peut s'apparenter à une question

d'appariement de bases du point de vue des objets « villes » et de généralisation cartographique du point de vue du réseau. Les réflexions ont été nourries à la fois par des tests effectués sur les sources sud-africaines et sur les sources françaises et ont donné lieu à un ensemble de propositions génériques, reprises dans chacun de ces deux projets.

2.4.A – Afrique du Sud

A ce jour, il n'existe pas de base numérique de l'évolution du chemin de fer en Afrique du Sud, mais uniquement quelques cartes dessinées par Ron J. Davies (Davies, 1972) et reprise par Céline Vacchiani-Marcuzzo dans sa thèse (Vacchiani-Marcuzzo, 2005).

Un premier travail a donc consisté à rechercher les sources permettant de numériser l'évolution du réseau. Au cours de la mission de Solène Baffi (janvier 2011), les archives détenues par l'Université de Cape Town (UCT) ont été analysées et photocopiées lorsqu'elles présentaient un intérêt pour la numérisation des chemins de fer. Deux types d'archives ont été rapportés.

Tout d'abord, le *Report of the South African Railways and Harbours* (SAR&H), qui donne, pour chaque année à partir de 1863, la liste des ouvertures de lignes et lignes en construction. Ces documents ont été photocopiés et ramenés au laboratoire, puis ont fait l'objet d'une saisie informatique jusqu'en 1960 (Tableau 3). D'autres rapports existent jusqu'en 1980 mais n'ont pour le moment pas fait l'objet d'une saisie (le réseau ferroviaire est peu dynamique dans la deuxième moitié du 20ème siècle, exception faite des lignes créées pour desservir les zones d'urbanisation déplacée, les bantoustans).

Tableau 3 : Extrait d'une page de rapport du SAR&H, photocopiée puis numérisée

OPENED		SECTION		GOVERNMENT LINES				
Year	Month	From	To	Cape of Good Hope	Natal	Transvaal	Orange Free State	South West Africa
1875	22nd Sept	Swartkops	Uitenhage	13 _ 39				
1875	3rd Nov	Tulbagh Road	Wolseley	9 _ 45				
1876	1st Jan	Cape Town	Wynberg	8 _ 06				
1876	1st March	Addo	Coerney	7 _ 48				
1876	1st April	Coerney	Sand Flats	14 _ 68				
1876	16th June	Wolseley	Worcester	24 _ 30				
1876	14th Sept	Bellville	Muldersvlei	14 _ 14				
1876	18th Dec	East London	Blaney	24 _ 79				

Un deuxième type d'archives consiste dans un corpus de cartes anciennes, appelées *Standard Railway Map of South Africa* et photographiées par Solène Baffi puis transformées en format image (Figure 2). Onze cartes sont ainsi exploitables, et s'échelonnent[3] entre 1892 et 1961. S'il n'est pas possible d'utiliser directement ces cartes pour le géo-référencement des stations, elles constituent néanmoins une aide précieuse pour révéler les interconnexions entre les tronçons du SAR&H et permettre ainsi la numérisation du réseau.

[3] Il s'agit des années 1892, 1912, 1916, 1921, 1925, 1931, 1935, 1938, 1945, 1955 et 1961.

Figure 2 : Extrait de la carte du réseau ferroviaire sud-africain de 1912



Source : Standard Railway Map of South Africa, 1912.

2.4.B – France

Dans le projet initial, un travail de numérisation du réseau avant été proposé, les années 1830 à 1860 ayant déjà été saisies (Bretagnolle 2009). Cependant, plutôt que poursuivre ce travail de numérisation, une collaboration avec Thomas Thévenin et son équipe, à l'université de Bourgogne, a été envisagé. Il s'agit d'un système d'information geo-historique, achevé récemment et qui couvre les années 1821 à nos jours. Le travail au sein de l'ANR Harmonie-cité a donc consisté à réfléchir sur l'articulation entre la base villes et la base transports.

Une fois construits les différents réseaux ferroviaires, plusieurs types de **modélisation** ont été prévu pour explorer les liens entre accessibilité urbaine et dynamique des villes. Des indices d'accessibilité locale, fondés sur l'évaluation de la nodalité des villes, peuvent être calculés très facilement à partir de requêtes de type SIG. L'évaluation d'une accessibilité globale, reposant sur des calculs de plus courts chemins entre une ville et l'ensemble des autres villes du système, nécessitent un outillage méthodologique un peu plus élaboré, tel que celui mis en place pour le réseau des routes postales en France aux 17ème et 18ème siècles (Bretagnolle, Giraud, Verdier 2010). Enfin, l'utilisation du modèle multi-agents Simpop2 permet d'envisager une amélioration des simulations obtenues grâce à la prise en compte de l'accessibilité des villes. La modélisation et la simulation informatique offrent, en effet, un corpus de méthodes et d'outils pour étudier des phénomènes spatiaux complexes. Elles peuvent notamment être utilisées afin de tester des raisonnements hypothético-déductifs difficiles à dérouler mentalement ou mathématiquement, du fait de l'imbrication de nombreuses interactions entre les entités composant le système étudié. La simulation informatique à base d'agents est l'une des approches privilégiées par le laboratoire pour ces études calculatoires. L'origine de ces travaux remonte à une quinzaine d'années et au modèle Simpop, application pionnière des systèmes multi-agents à la géographie (Bura et al., 1996). Le modèle Simpop2[4] a été conçu et développé au milieu des années 2000. Il complète et raffine Simpop1

4 www.simpop.parisgeo.cnrs.fr

et a été permis par l'augmentation de la puissance de calcul, le gain en expressivité des langages et librairies de programmation, et l'expérience accumulée de l'équipe. Il a été appliqué à un panel étendu de systèmes de villes du monde (Europe, Etats-Unis, Afrique du Sud notamment, Pumain et al. 2009, Bretagnolle et Pumain, 2010). Cependant une des limites de ce modèle est qu'il n'intègre pas explicitement les réseaux de transport entre villes, et ne permet donc pas de tester d'hypothèses sur les liens pouvant exister sur le long-terme entre l'accessibilité (évolutive) d'une ville et sa trajectoire au sein du système des villes. Jusqu'ici, les distances entre les villes sont calculées à vol d'oiseau dans le modèle, et ce sont les villes en position centrale qui sont avantagées dans le fonctionnement des échanges. Ces hypothèses sont cependant très rudimentaires, comme l'atteste le cas de Londres, située à l'écart des autres villes d'Europe si l'on ne tient compte que de la distance euclidienne, alors qu'elle est en réalité au cœur des réseaux d'échanges maritimes puis ferroviaires à partir du 18ème siècle. La comparaison de résultats de simulation obtenues avec des distances à vol d'oiseau et obtenues avec des distances-réseau devrait permettre de mieux préciser, dans la dynamique des villes, la part des effets structurants dus à l'établissement de nouveaux réseaux de transport (19ème – 1ère moitié du 20ème siècle).

2.5 Interface Harmonie-cités

L'élaboration de l'interface du projet Harmonie-cités s'est articulée autour du stage de Master 2 Pro Carthagéo réalisé par Adrien Van Hamme en 2010 (encadré par Hélène Mathian et Anne Bretagnolle) et s'est prolongée en 2011. Un travail important de réflexion conceptuelle et méthodologique a précédé les premières réalisations techniques de l'interface. En effet, le projet était ambitieux et original, par rapport à l'ensemble des sites de cartographie dynamique et interactive des villes qui avaient été repéré sur Internet.

Tout d'abord, le niveau de référence n'est pas la ville mais le système de villes. Les implications méthodologiques sont fortes, car cela suppose que l'interface soit suffisamment souple pour que l'utilisateur puisse sélectionner une variété de contextes géographiques. La notion de systèmes de villes est, en effet, tout aussi complexe que celle de ville. Selon les époques et selon le niveau hiérarchique de la ville ou sa spécialisation économique, les réseaux d'interdépendances tissées avec d'autres villes peuvent être mondiaux (cas non pris en compte dans l'interface actuelle), macro-régionaux, nationaux, régionaux ou non contigus (par exemple les villes les plus grandes à une certaine époque, impliquées dans des relations à longue distance et constituant une sorte d'archipel). Différents contextes systémiques doivent donc être prévus sur la plate-forme afin que l'utilisateur puisse choisir, en fonction de l'époque et du pays considéré, les ensembles de villes qui co-évoluent et qui sont susceptibles d'avoir un impact sur la trajectoire d'une ville au sein de ce système. Cela suppose aussi de guider le choix des utilisateurs peu familiers avec cette notion de système de villes, par des contextes proposés par les experts. Par exemple, dans l'application dédiée aux Etats-Unis, les villes ne forment pas encore un système national au 19ème siècle mais ressortent de plusieurs sous-systèmes (la région des Grands Lac et ses prolongements au nord du littoral atlantique, le sud-est des Etats-Unis, la côte pacifique, les villes pétrolières du Texas etc.).

Ensuite, l'approche des villes n'est pas mono-dimensionnelle (par exemple la ville dans sa composante démographique) mais multi-dimensionnelle. La pluralité des bases de données rassemblées dans le projet Harmonie-cités permet d'envisager de

montrer plusieurs facettes des villes, en termes de populations, d'emprise spatiale, d'activités économiques et d'accessibilité dans les réseaux de transport. L'objectif est alors de permettre à l'utilisateur d'explorer différentes facettes des relations d'interdépendances entre les villes, par exemple des liens entre l'accessibilité des villes et leur dynamique économique (les « hubs » de Chicago ou de Paris par exemple), entre des ressources régionales et des cycles d'innovation économique (Rust-belt, Sun-belt…), entre les tailles des villes et leur profil social (par exemple la concentration des « classes créatives » dans les grandes métropoles). Cela suppose là aussi des contraintes fortes dans la création de l'interface, qui doit pouvoir intégrer les applications différentes : Population et emprises spatiales, Activités économiques, Accessibilité dans les réseaux de transport.

3. Résultats scientifiques obtenus, présentation de leur caractère original, mise en perspective avec les attentes initiales et l'état de l'art

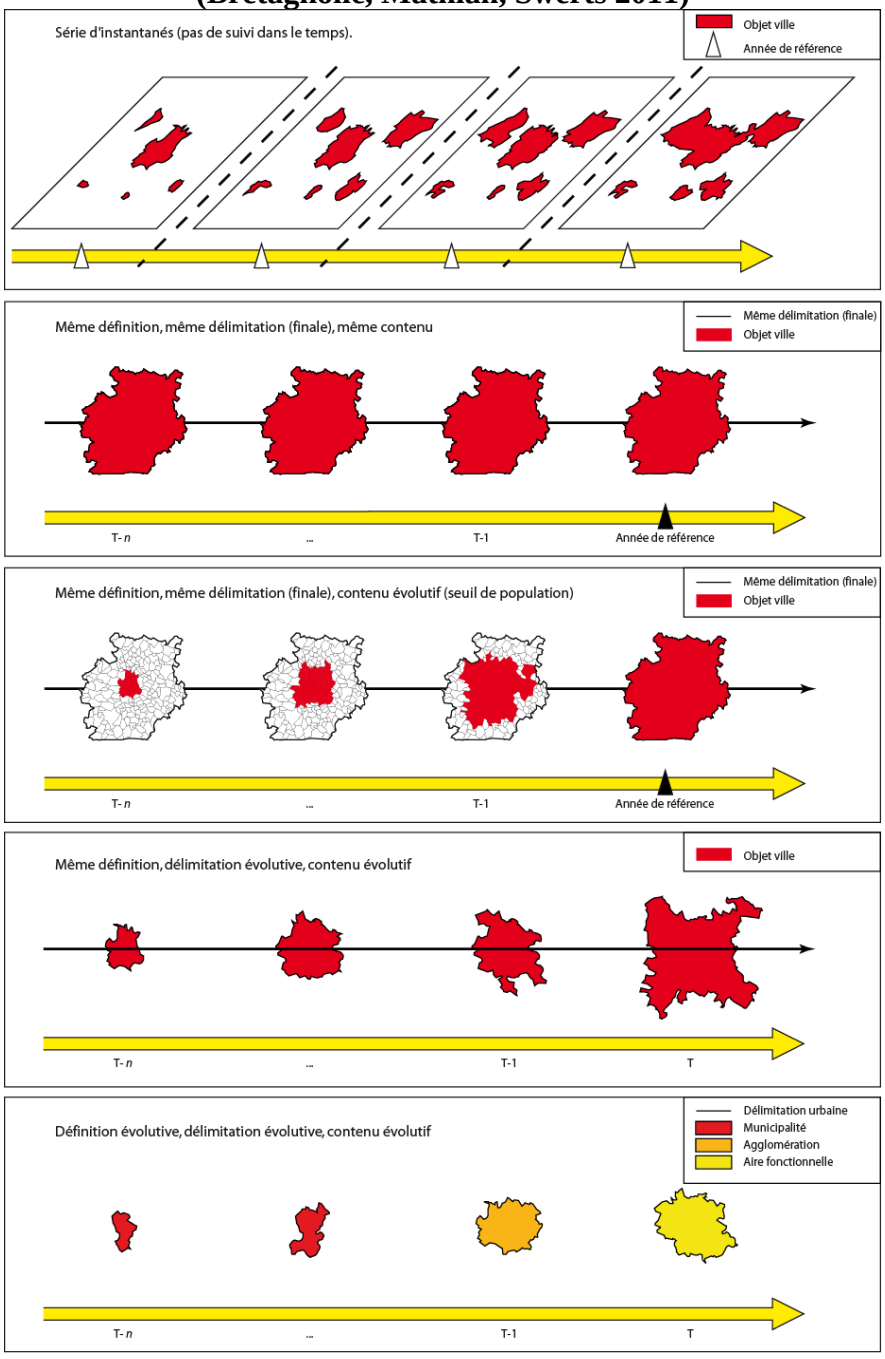
3.1 Conceptualisation des modèles de données, automatisation des processus, réflexions sur la notion d'harmonisation

Un premier chantier consistait à réfléchir sur la notion d'harmonisation longitudinale et transversale, à partir d'une étude comparée des bases de données permettant de suivre des ensembles de villes au cours du temps, en Europe, aux Etats-Unis, en Chine ou en Inde. Les résultats ont été présentés aux Journées de Rochebrune en janvier 2011 et sont en cours de publication (Bretagnolle, Mathian, Swerts 2011). Ils consistent dans la formalisation des différents types de modèles conceptuels utilisés pour suivre des villes au cours du temps (Figure 3). Le premier consiste dans la succession d'une série d'instantanés des villes, sans chercher à les suivre au cours du temps. C'est l'approche utilisée, par exemple, par le Bureau du Recensement des Etats-Unis depuis le milieu du 20ème siècle : les identifiants des villes varient d'un recensement à un autre, notamment lorsque les critères de définition sont modifiés, ce qui rend extrêmement compliquée toute tentative de suivi temporel des objets urbains. La deuxième approche est régulièrement utilisée par les Offices statistiques, car c'est la plus simple à mettre en place pour suivre les villes dans le temps. Il s'agit de garder la même délimitation tout au long de la période considérée, celle-ci correspondant généralement à la plus récente. C'est le choix, par exemple, de l'Audit Urbain (organe émanant d'Eurostat pour la collecte et l'harmonisation de données urbaines) pour la base européenne des *Larger Urban Zones* (Eurostat 2004). Cette approche est aussi suivie dans la base de données des aires fonctionnelles des Etats-Unis de Duncan Black et Vernon Henderson (2003). Ils partent de la délimitation de 1990 et recalculent les populations à l'intérieur de ce périmètre[5] pour les recensements plus anciens, compris entre 1950 et 1980. Pour les dates antérieures (1900-1940), ils reprennent un travail semblable effectué par D. Bogue (1953), qui rétropole les SMAs dans leur limite de 1950. Cette approche entraine une surestimation de la population des plus grandes villes aux dates les reculées. Une troisième approche consiste à utiliser la délimitation finale de l'objet, mais à ajouter un critère de population minimale (par exemple 10 000 habitants, pour la base Géopolis de François Moriconi-Ebrard, 1994). Seules les entités élémentaires composant la ville à la date finale et dépassant ce seuil (en contigüité) sont retenues à chaque date, ce qui donne un contenu évolutif à l'objet urbain. Notons que c'est la même qui a été utilisée dans le Fichier Urbanisation Historique de l'INED, pour les dates comprises en 1831 et 1954. Un risque de surestimation des superficies et donc des populations des grandes villes est envisageable lorsque la période considérée est large. Une quatrième approche, utilisée par exemple dans la base Urban Morphological Zones, de l'Agence Européenne de l'Environnement, ou dans les bases des unités urbaines de l'INSEE, du Danemark ou de la Suède, consiste à utiliser même définition mais une délimitation évolutive des villes. Ainsi, le seuil d'espacement minimal de 200 mètres entre les bâtiments conduit à

[5] Avec quelques variations : ainsi, pour les 18 CMSAs correspondant aux plus grandes aires métropolitaines, les auteurs utilisent les délimitations de leurs composantes, appelées PMSAs. Par exemple, l'aire de New York ne compte dans leur base que 8 millions d'habitants et s'étend sur un périmètre restreint, le même qu'en 1980. En outre, seule la population urbaine des *counties* composant l'aire est considérée à chaque date, ce qui minimise les risques de sur-évaluation des populations totales dans les décennies les plus reculées.

réévaluer périodiquement l'extension de la zone bâtie en continuité. Cette approche est très satisfaisante pour des périodes de temps relativement courtes (de l'ordre d'une à trois décennies). Néanmoins, elle se révèle insuffisante lorsqu'on se place sur des durées de temps plus longues, marquées notamment par des changements importants dans les technologies de transport. C'est pour cette raison que nous proposons une cinquième approche, celle de ville harmonisée, dans laquelle ce sont les définitions même des villes qui changent au cours du temps (passage de la municipalité à l'agglomération puis à l'aire fonctionnelle).

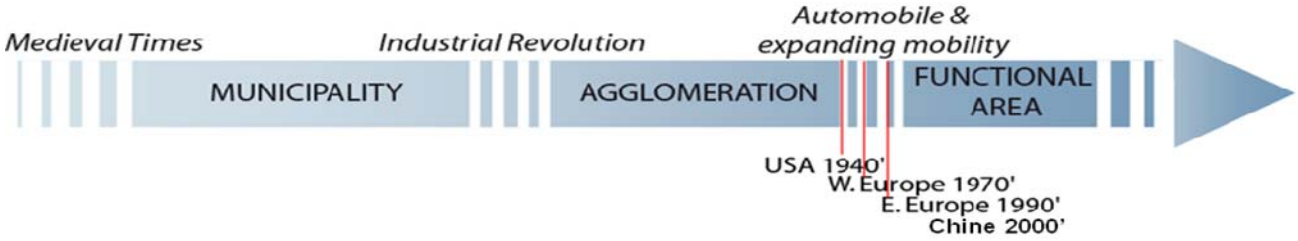
Figure 3 : Cinq approches pour l'intégration du temps dans les bases de données urbaines (Bretagnolle, Mathian, Swerts 2011)



Les règles de passage d'une définition à une autre dans les bases de données harmonisées doivent cependant faire l'objet d'un travail d'expertise car elles varient

selon la position hiérarchique de la ville (pour les grandes villes, le budget-temps est légèrement supérieur à une heure et pour les petites légèrement inférieur) et selon le stade technologique atteint par le pays au moment des deux grands basculements que sont la fin du 19ème siècle (passage de la municipalité à l'agglomération, pour les grandes villes) et le milieu du 20ème siècle (passage de l'agglomération à l'aire fonctionnelle pour les grandes villes). Par exemple, la diffusion de l'automobile pour des trajets domicile-travail s'échelonne sur près de soixante ans, entre les années 1940 (Etats-Unis) et les années 2000 (Chine) (Figure 4).

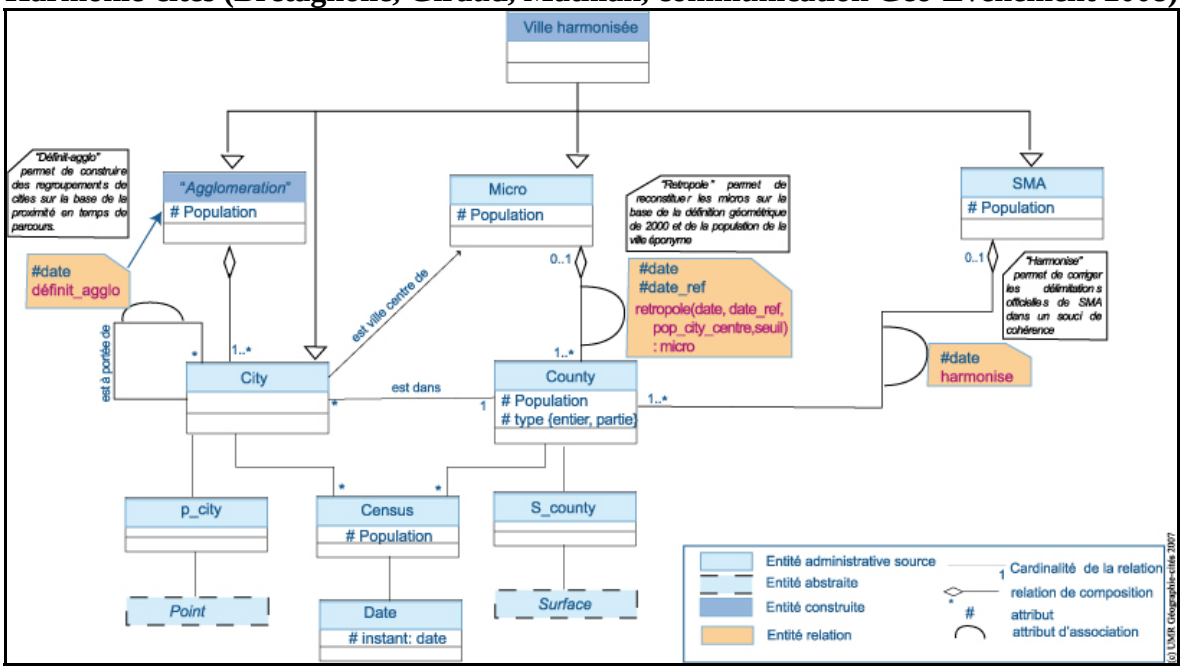
Figure 4 : Harmonisation des définitions des villes et prise en compte des contextes technologiques, l'exemple de la diffusion de l'automobile

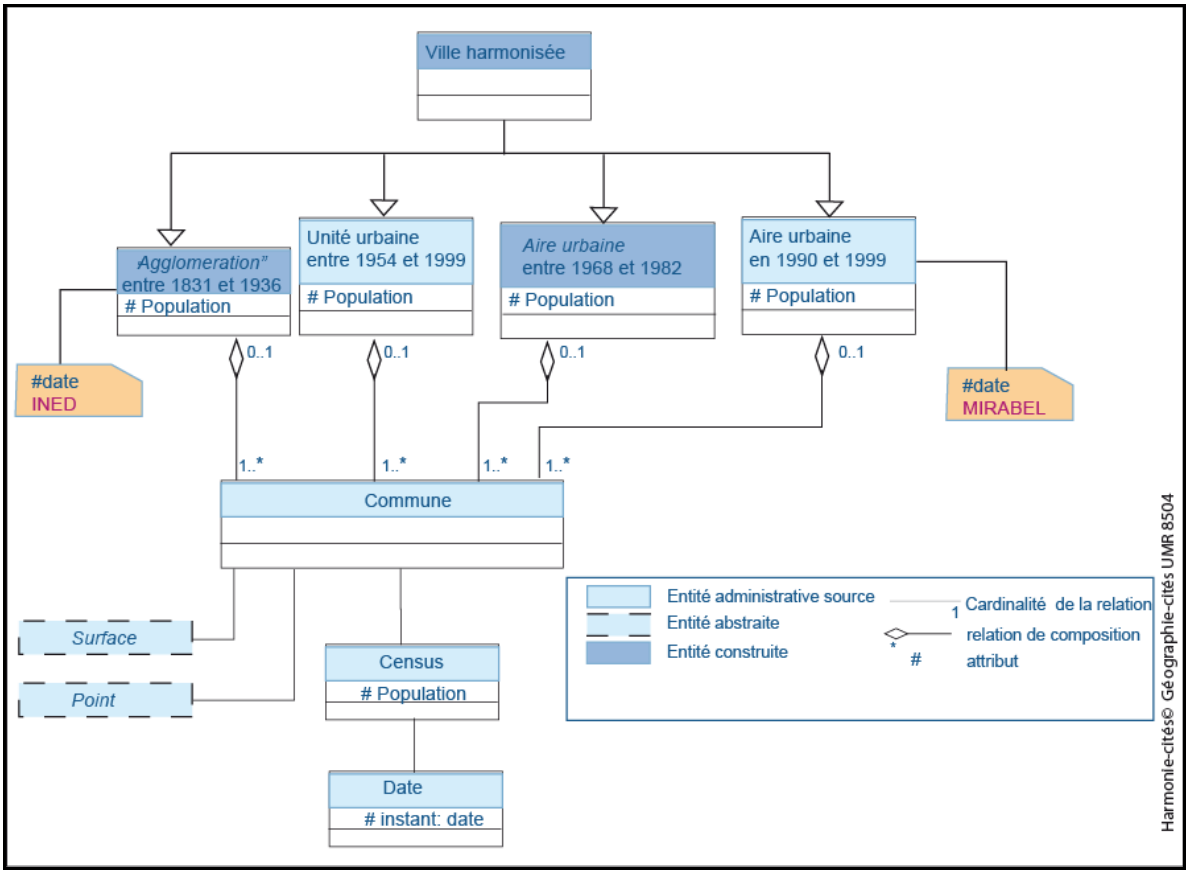


Un deuxième chantier concernait la création de modèles de données adaptés à des bases harmonisées dans le temps long, dans les différentes régions du monde concernées par le projet Harmonie-cités.

Un premier résultat a consisté dans la création d'un modèle de données permettant de suivre les villes françaises de manière harmonisée sur deux siècles. La création de l'interface Harmonie-cités pour la France rendait, en effet, nécessaire ce travail d'harmonisation des définitions. Sur la base du modèle multi-niveau et dynamique qui avait été conçu pour les villes des Etats-Unis en 2006 et 2007 (Bretagnolle, Giraud, Mathian 2008), un modèle de données a été élaboré pour la France (Figure 5). Entre 1831 et 1954, la ville harmonisée correspond à une agglomération morphologique, définie selon la méthodologie de la base INED. Entre 1954 et 1968, ce sont les unités urbaines de l'INSEE qui ont été considérées. Enfin, à partir de 1968, comme pour la base Etats-Unis, les agglomérations peuvent être remplacées par les aires urbaines lorsque celles-ci existent. Les aires urbaines françaises ont été définies en 1996 mais ont été rétropolées par Philippe Julien et un certain nombre de chercheurs de l'INSEE (Julien 2003) aux recensements de 1968, 1975, 1982 et 1990, sur la base de sources existant à chaque époque (fichiers des navettes Mirabelle). Les bases de données obtenues ont été géocodées, mises en forme et exploitées par Fabien Paulus dans le cadre de sa thèse (Paulus 2004).

Figure 5 : Le modèle de données construit pour les Etats-Unis (Bretagnolle, Giraud, Mathian 2008), en regard du modèle de données construit pour la France dans l'ANR Harmonie-cités (Bretagnolle, Giraud, Mathian, communication Géo-Evenement 2008)





Un deuxième résultat a consisté dans la conception et la création d'un algorithme permettant de construire, de manière automatique, des aires fonctionnelles ou des agglomérations fonctionnelles sur la base du budget-temps d'une heure.

Comme il est spécifié dans la Figure 6, la conception d'une base de données spatio-temporelle cohérente a nécessité l'automatisation de processus. Nous présentons ici celui qui correspond à l'agrégation automatique des agglomérations en « Villes fonctionnelles » (definit_agglo). Le modèle sous jacent à cette automatisation est qu'une aire urbaine fonctionnelle est définie autour d'un centre par l'ensemble des agglomérations qui sont à portée d'une heure de trajet du centre.

L'application développée permet de trouver automatiquement les « grappes » d'agglomérations, selon cette définition, à une année donnée. La procédure est itérative et est composée de plusieurs étapes. Elle est fondée sur la notion de potentiel d'interaction, qui permet de hiérarchiser les grappes ainsi détectées :

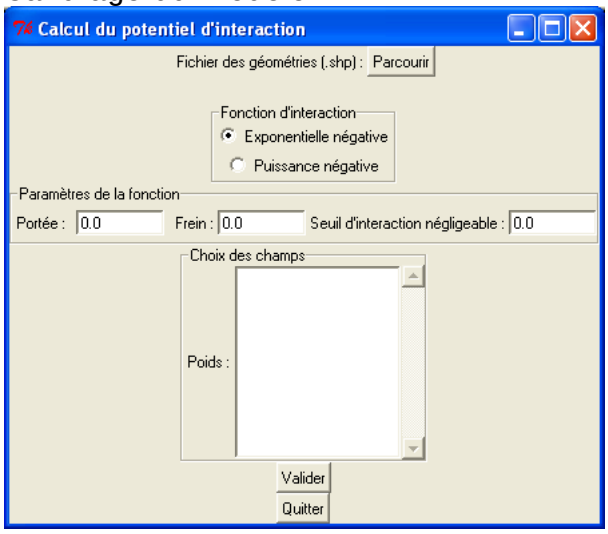
$$I_{ij} = k \frac{P_j}{f(d_{ij})}$$

Une entité j exerce sur les entités voisines i une « attraction » Iij, proportionnelle à la population de j et inversement proportionnelle à la distance qui les sépare selon une fonction f à définir. La procédure développée doit être calibrée en fonction de la distance géographique associée au référentiel d'une heure de trajet, cette distance étant variable au cours des années. Si Iij > Iji alors l'entité j « polarise » l'entité i, sinon c'est l'inverse.

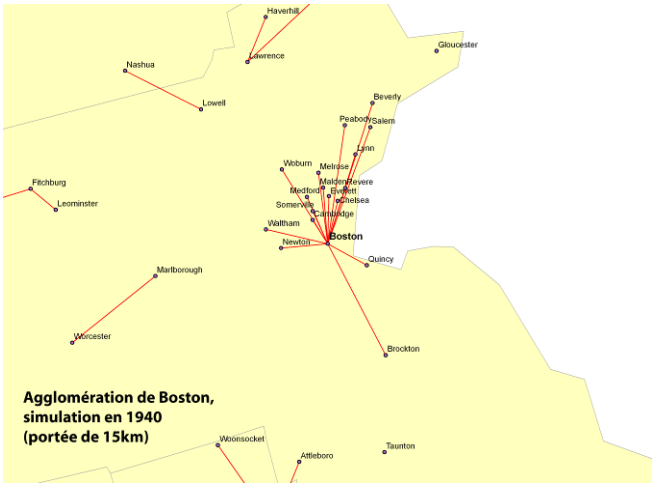
A la fin de la procédure, on obtient un graphe reliant les entités distantes de moins d'une heure. Pour chaque composante connexe de plus d'une entité, on identifie un pôle central et ses périphéries. A chaque composante connexe est donc associée une aire fonctionnelle.

Figure 6 : Modélisation des agglomérations fonctionnelles des Etats-Unis à partir des municipalités (cities et towns), en 1940 : la procédure « grappe »

Calibrage du modèle



Résultat de la procédure



L'application a été développée en langage python, et elle est entièrement compatible avec ArcGIS. Elle se décompose en 4 modules :

M1 - Calcul du potentiel d'interaction sur l'ensemble des points en entrée.

Entrée : un fichier géométrique des points à étudier, n lignes de type [ID, géométrie,

masse] (SHP)

Sortie : une table de type [i, j, I(i, j)] (DBF).

M2 – Recherche des Pôles / Périphéries (graphe de dominance)

Entrée : une table du potentiel d'interaction (DBF)

Sortie : une nouvelle table de potentiel, ne gardant que le maximum pour chaque entité

(DBF)

M3 - Simplification de la hiérarchie du graphe (chaînage)

Entrée : une table de potentiel maximum (DBF)

Sortie : une table représentant le graphe final chaîné (DBF)

M4- Représentation graphique

Entrée : une table représentant un graphe orienté (DBF)

Sortie : un fichier géométrique de lignes représentant les grappes d'interaction (SHP)

L'application a été testée sur des semis déjà agrégés manuellement. Un utilitaire de comparaison de graphes a été développé pour permettre une analyse de sensibilité des résultats aux paramètres et donc un calibrage plus raisonné de l'application.

Les limites de la méthode restent l'évaluation des distances à vol d'oiseaux. Le modèle serait bien sur amélioré en introduisant des distances calculées sur le réseau.

3.2 Bases de données sur les Populations et espaces urbains

Concernant **l'Europe**, un premier chantier consistait à examiner les possibilités de jointure entre la base de Paul Bairoch et ses collaborateurs (1988) et celle de François Moriconi-Ebrard (Géopolis).

Un premier travail a consisté à faire un état précis de la base Bairoch, notamment des délimitations utilisées par les auteurs et des métadonnées. Cette base rassemble les populations des 2200 villes d'Europe (Russie comprise) qui ont compté, une fois au moins, 5000 habitants entre l'an 800 et 1800. Les informations utilisées sont issues de recensements, de chiffres publiés par des chercheurs ou de bases de données plus anciennes, celle de Jan De Vries (1984) ou celle de T. I. Chandler et G. Fox (1976). Les villes sont définies selon deux types de critères, lorsque les informations sont disponibles. Tout d'abord, des critères d'agglomération urbaine : « La ville inclut la population non seulement de la ville proprement dite, mais également des 'faubourgs, des 'banlieues', 'communes', 'hameaux', 'quartiers' etc. qui lui sont directement adjacents » (p. 289). Les métadonnées sont de faible qualité et la traçabilité de la base assez médiocre : par exemple, les auteurs ne fournissent pas d'information sur les entités composant les grandes villes, ou sur les

sources permettant de réaliser ces agrégations6. Ensuite, des critères socio-économiques : dans certains pays, comme la Hongrie ou l'Italie du sud, de véritables villages de plus de 10 000 habitants existent et ont été écartés de la base. Pour des raisons de fiabilité des sources, les auteurs suggèrent d'utiliser le seuil minimal de 10 000 habitants dans l'exploitation de la base de données. Le nombre total de villes est alors de 618 en 1800 et 967 en 1850. L'utilisation de la base pour l'année 1850 introduit cependant un biais non négligeable, souligné par les auteurs eux-mêmes. Seules sont recensées pour cette année les villes qui ont dépassé 5000 habitants au cours des dates précédentes (Bairoch et alii 1988, p. ix). Comme le reconnaissent les auteurs eux-mêmes, cela conduit à sous-estimer largement le nombre des villes dans un pays comme la Grande-Bretagne, déjà engagé dans la révolution industrielle et caractérisé par l'apparition de nombreuses villes au milieu du 19ème siècle. Lors du dépouillement de deux volumes du recensement anglais de 1901, nous avons pu constater que nombre de villes n'étaient pas mentionnées dans la base Bairoch.

L'expertise réalisée sur la disponibilité de définitions des villes européennes en tant qu'agglomération autour de 1900 a porté sur 18 pays. Les recensements relatifs à la Grèce (1913), la Hongrie (1891), l'Autriche (1890), les Pays-Bas (1899) et la Pologne n'ont pas été dépouillés, principalement pour des raisons linguistiques. Les résultats de cette expertise ont été valorisés dans le premier chapitre du Mémoire d'Habilitation à Diriger des Recherches (Bretagnolle 2009), en ligne sur Hal-SHS. Ils révèlent une grande diversité dans les approches de la ville en Europe, avec 4 pays seulement sur les 18 analysés proposant des définitions d'agglomérations (la Suisse, la Belgique, la France et l'Allemagne) (Tableau 4). Cela signifie qu'il n'est pas possible de s'appuyer sur des définitions nationales des agglomérations et qu'il faut construire nous-mêmes les agrégats d'entités élémentaires, en nous appuyant sur le budget-temps d'une heure. Cette entreprise est extrêmement complexe, notamment en raison de l'évolution des maillages nationaux au cours du temps. En effet, comme nous avons pu le constater en examinant les archives autour de 1900, il est parfois impossible de localiser les communes citées dans les recensements en utilisant le fond actuel des NUTS 5. Certains maillages ont beaucoup évolué, par exemple à l'occasion de réformes administratives (la Belgique en 1977, l'Angleterre en 1974, la Grèce en 1997, mais aussi le Portugal, l'ex Yougoslavie, le Danemark ou d'autres, voir Moriconi-Ebrard et Hubert 1999) ; en outre, les mailles ont souvent été modifiées localement, par des annexions, des fusions ou des créations, ce qui implique la mise en place de méthodologies particulières pour assurer leur suivi. Une autre difficulté concerne la construction d'agglomérations morphologiques. Même en utilisant le seuil relativement élevé de 10 000 habitants en 1900, nous avons estimé à plus de 2000 le nombre d'agglomérations pour l'ensemble des pays d'Europe, et ce nombre est d'environ 5000 en 1950, d'après la base Géopolis. Si l'on arrive à régler le problème du géo-référencement des toponymes, il reste celui des sources sur les réseaux de transport ferroviaires suburbains. Celles-ci doivent être collectées pour toutes les grandes villes d'Europe entre 1870 et 1950 pour pouvoir transposer la méthode de construction de grappes mise au point pour les Etats-

[6] Comme le fichier Urbanisation historique de l'INED, cette base a été conçue dans les années 80, et Paul Bairoch défend l'idée d'un travail *a minima*, pour en permettre la faisabilité : « La méthode de travail utilisée durant ces années a été très artisanale. En fait, je me contentais (…) de remplacer un chiffre chaque fois qu'une source plus récente en proposait un autre, mais sans relever systématiquement la référence (sauf lorsque celle-ci fournissait plus de huit renseignements différents) » (Bairoch et *alii*, 1988, p. xii).

Unis. Un tel chantier nécessiterait la mise en place d'une collaboration internationale entre des historiens, démographes, géographes et géomaticiens.

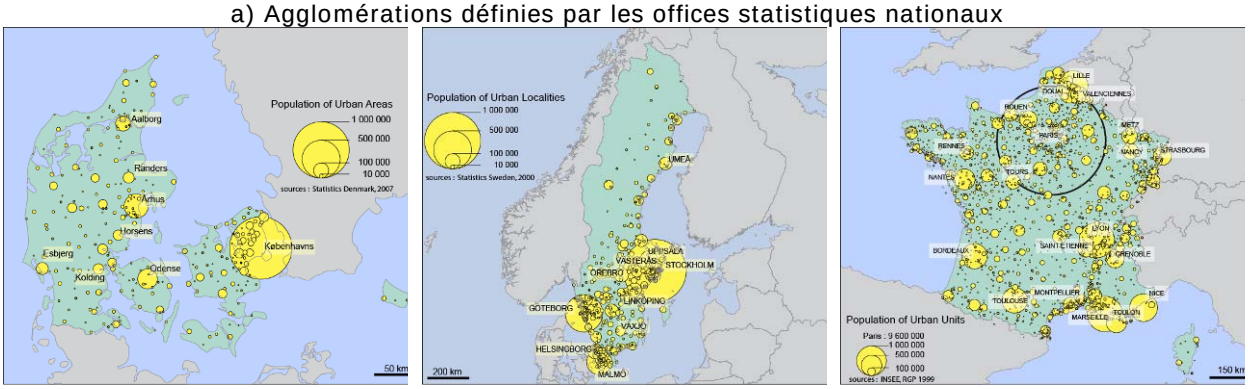
Tableau 4 : Les définitions de la ville en Europe au tournant des 19ème- 20ème siècles
(d'après l'expertise de J.-B. Arrault)

Pays	Recensement ou autre source	Briques élémentaires (et nombre lorsque mentionné)	Entité « ville » (et nombre lorsque mentionné)
Albanie (nord)	1916-1918, Akademie … (BN)	*Gemeinde*	*Stadtgemeinde*
Allemagne	1900, Statistik… (BN)	*Gemeinde* (3397)	*Agglomeration (27)* *Stadtgemeinde*
Angleterre	1901, Parl. Papers (BN), 2 des 7 volumes dépouillés	Paroisse *(parish),* District *(ward)*	*County boroughs, Municipal borough, Urban district*
Belgique	1890 et 1910, Min. Int. (INED)	Commune (2596)	Agglomération (4)
Bosnie-Herzégovine	1895, Statistischen Dep. (INED)	*Gemeinde*	*Stadtgemeinde*
Danemark	1901, Danmarks Stat. (INED)	Paroisse (*beforlknong*)	*By* (paroisse urbaine) + « agglomération » Copenhague
Espagne	1900, Ministerio de Instruct. (en ligne)	Municipalité (*ayuntamientos*)	*Capitales de provincias*
France	1911 (Dénomb., INED)	Commune	Commune urbaine
Irlande	1902, Jordan 1998	Paroisse (*parish*) *Town*	*Town, City* *Urban district*
Islande	1901, Guðmundur et al. 1997	Paroisse (*parish*) *Nucleus*	*Town* *Urban nucleus*
Italie	1901, Ministerio di agric. (BN)	*Communi*, *mandamento*, *frazioni*	*Centri* *Citta*
Luxembourg	1900, Min. de l'Eco. (BN)	Communes (118)	« canton de Luxembourg-ville »
Norvège	Sundbärg (1906)	Non mentionné	Oslo « avec ses faubourgs »
Portugal	1900, Min. dos Negocios (INED)	Paroisses (*freguesias*)	*Cidades*
Roumanie (Royaume + Transylvanie)	1899, Ministerul agriculturei et Rotariu (1999), INED et BN	*Comune* *Municipii* (transyl.)	Comune urbane Orasele (villes principales)
Serbie	1900, Direction de la Stat. d'Etat, (INED)	Commune	Commune urbaine
Suède	Sundbärg (1906)	« communes » (ouvrage en français)	6 agglomérations (ville + « faubourg »)
Suisse	1900, Stat. de la Suisse (INED)	Commune	Agglomération

L'exploration de la **base *UMZ*** a conduit à deux séries de résultats qui concernent tout aussi bien la validation de cette base que son exploitation thématique. La confrontation entre la base *UMZ* et trois bases nationales d'agglomérations a tout d'abord permis de vérifier la pertinence des *UMZ* en tant qu'objets urbains, même si la rareté des bases d'agglomérations nationales nous a amené à limiter l'échantillon étudié: il ressort d'une part que la délimitation des agglomérations dans la base *UMZ* n'est pas systématiquement biaisée par la résolution de la base CORINE Land cover, pourtant moins précise que les images du bâti utilisées par les offices statistiques nationaux. D'autre part, la base *UMZ* et les bases nationales renvoient des images fortement convergentes du fait urbain, avec par exemple des écarts moyens de population urbaine de 5 à 8%, selon les Etats. Enfin, les plus forts écarts observés localement entre UMZ et agglomérations nationales concernent des conurbations industrielles ou littorales, qui sont associées à plusieurs agglomérations distinctes dans la base *UMZ*. Ces résultats, qui peuvent préparer à la constitution de bases d'agglomération dans les pays ne disposant pas de telles données (soit la majorité des Etats d'Europe centrale et orientale), ont été valorisés lors de deux conférences internationales (*European Colloquium of Theoretical and Quantitative Geography*, Bretagnolle et al., 2009 ; *Association of American Geographers*, Bretagnolle et al., 2010) et dans un article de la revue *Environment and Planning B* (Guérois, Bretagnolle, Giraud, Mathian, 2011, sous presse).

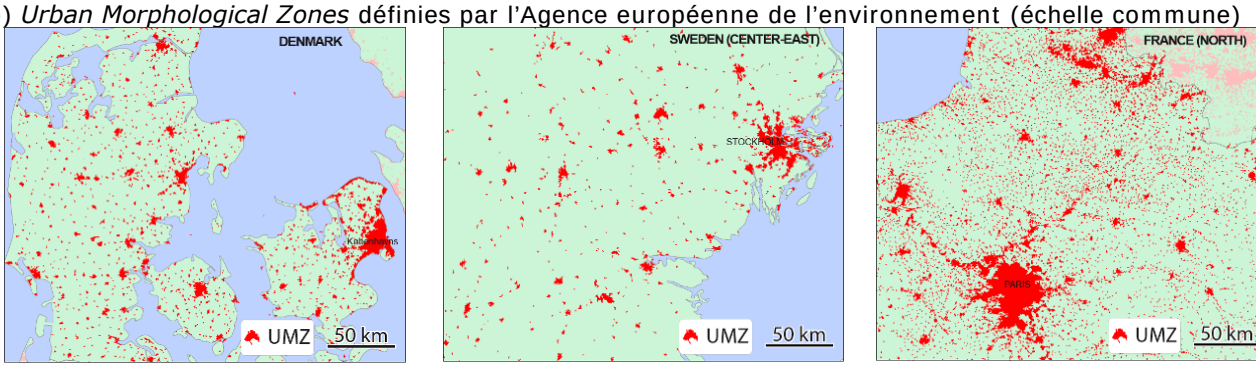
Figure 7 : Confrontation des UMZ à trois bases nationales d'agglomération (Danemark, France, Suède) (Guérois, Bretagnolle, Giraud, Mathian 2011)

a) Agglomérations définies par les offices statistiques nationaux



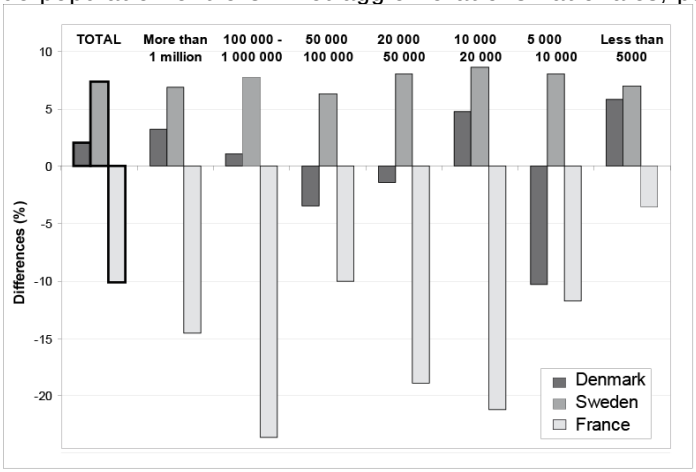
Sources : Danish Statistics agency (DST), Swedish statistics agency (SCB), INSEE).

b) *Urban Morphological Zones* définies par l'Agence européenne de l'environnement (échelle commune)



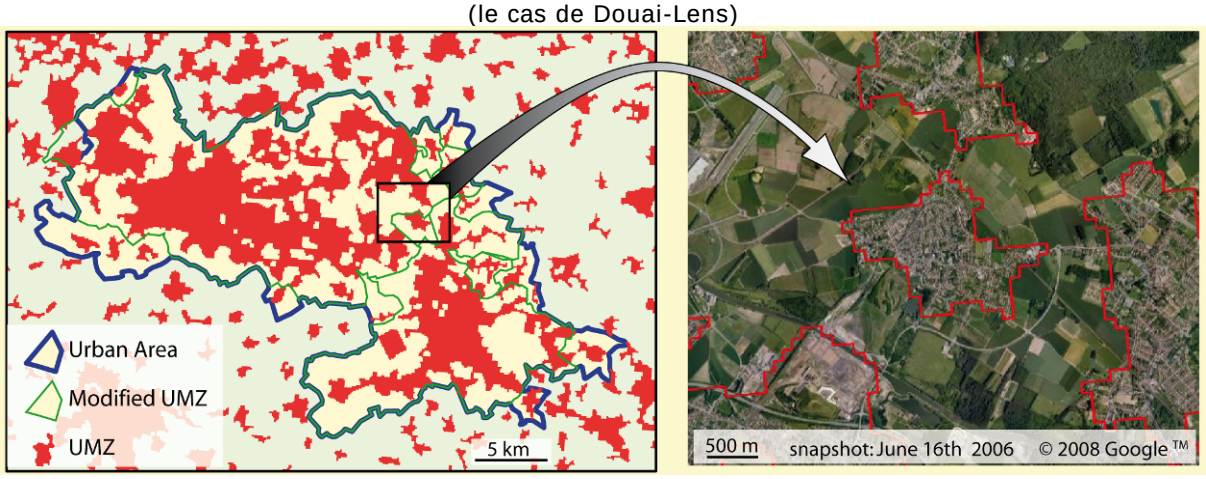
Source : European Environment Agency (EEA)

c) Ecarts moyens de population entre UMZ et agglomérations nationales, par classes de taille



Sources des délimitations : European Environment Agency (EEA), Danish Statistics agency (DST), Swedish statistics agency (SCB), INSEE. Source des données de population : population density grid, Joint Research Centre (JRC

d) Ecarts locaux : UMZ et agglomérations nationales confrontées à l'image Google Earth (le cas de Douai-Lens)

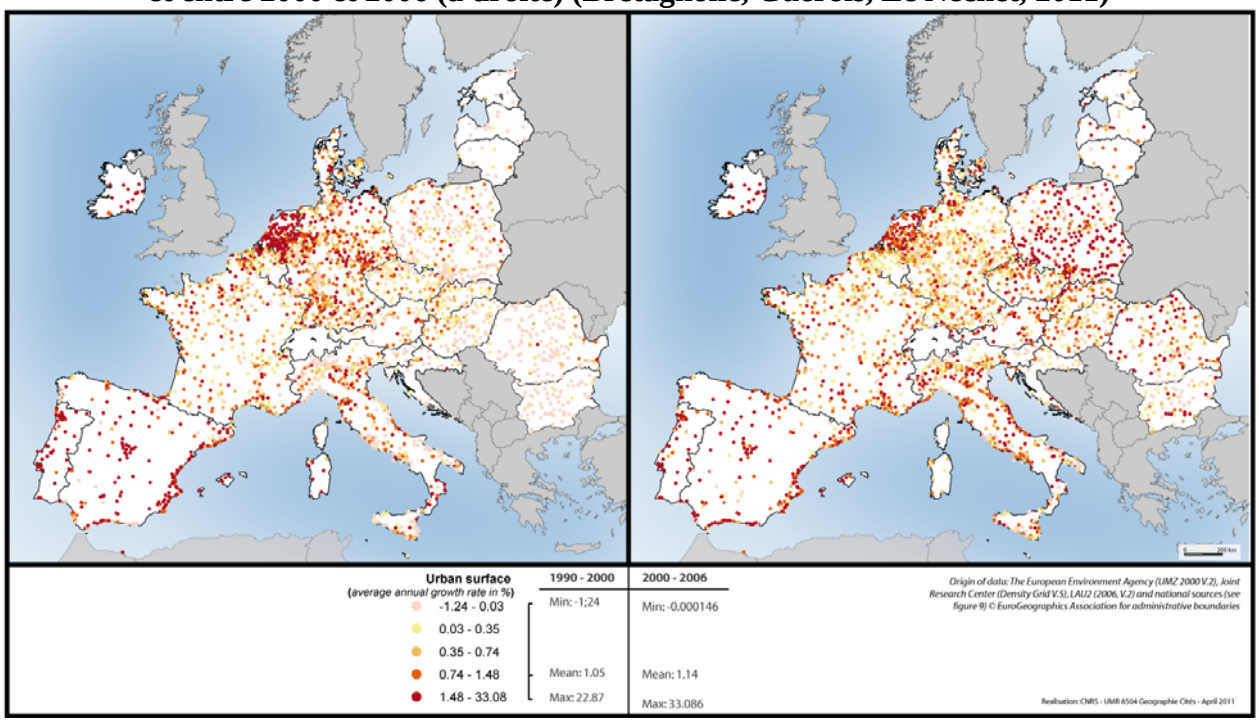


Sources: European Environment Agency (EEA), INSEE, Google Earth

Dans le prolongement de cette principale étape de validation, nous avons posé les bases d'une exploration multi-échelle du fait urbain en Europe, en collaboration avec Florent le Néchet (LVMT, Université Paris-Est). Le suivi de l'extension spatiale des *UMZ* entre 1990 et 2006 confirme l'exploitation de bases plus anciennes ou plus partielles, et précise notamment la situation de l'Europe centrale et orientale, dont les agglomérations connaissent l'expansion la plus rapide depuis 2000. D'autre part, la grande diversité de l'emprise spatiale des villes, étudiée au moyen d'un indicateur de densité nette, reflète l'influence partielle de facteurs géographiques (gradients nord/sud et est/ouest), le poids plus important de spécificités nationales, le lien plus

insignifiant avec la taille de ces villes. Enfin, l'appréciation des formes urbaines observées a été complétée, pour les 20 plus grandes agglomérations, par l'exploration d'un indicateur synthétique de compacité (distance moyenne relative entre habitants), qui décrit la répartition des densités à l'intérieur des *UMZ* et ressort comme une dimension indépendante des niveaux de densité nette. Les premiers résultats de ce travail ont fait l'objet d'une communication lors du dernier colloque de l'Association des géographes américains (Bretagnolle, Guérois, Le Néchet, 2011a) et ont été soumis en vue d'une communication au prochain colloque du Collège International des Sciences du Territoire (Bretagnolle, Guérois, Le Néchet, 2011b).

Figure 8 : Le basculement de la croissance de la surface urbanisée vers l'Europe Centrale. Taux de croissance moyen annuel de la surface urbanisée, entre 1990 et 2000 (à gauche) et entre 2000 et 2006 (à droite) (Bretagnolle, Guérois, Le Néchet, 2011)



Sources: European Environment Agency (EEA), Population density grid, Joint Research Centre (JRC)

Pour la **France**, les sources principales pour la correction et la validation du fichier de l'urbanisation historique de l'INED sont les volumes "quinquennaux" donnant les résultats du dénombrement général de la population depuis 1801. Pour faciliter la saisie et diminuer le risque d'erreurs, les fichiers numériques de travail ont été structurés afin qu'ils correspondent à l'organisation des sources papier, par exemple recherche par département puis par arrondissements dans lesquels les communes sont listées par ordre alphabétique. Il ne s'agit pas de saisir de nouvelles populations mais de valider le statut urbain ou non de la commune dans la base historique de l'INED en regardant la valeur de sa population agglomérée (rappelons que nous utilisons le seuil de 2500 habitants agglomérés au chef-lieu). La base historique de Claude Motte pourra servir dans ce cadre à lever des incertitudes liées aux modifications des noms de communes ou aux évènements (annexions, fusions, scissions etc.) qui ont pu les concerner.

Pour les **Etats-Unis**, la campagne de saisie débutée en 2010 a apporté plusieurs résultats importants. Tout d'abord, elle a permis d'approfondir l'expertise sur les

définitions de la ville aux Etats-Unis, notamment sur les notions de « city » et de « town », qui ne revêtent pas le même sens selon les périodes de temps et selon les Etats de ce pays. Lors de la première campagne de saisie en 2006-2007, le seuil minimal de population qui avait été choisi (plus de 10 000 habitants) était suffisamment élevé pour que le degré d'incertitude sur le statut, urbain ou rural, de certaines entités administratives soit faible. En revanche, le choix d'un seuil minimal de 2500 habitants, pour la deuxième campagne de saisie, a nécessité un travail approfondi d'analyse des sources accompagnant les recensements historiques. Rappelons que dans cette toute jeune démocratie élective que constituent les Etats-Unis du 19ème siècle, le passage de la simple localité (*place*) à une municipalité urbaine (*city* et/ou *town*, selon les Etats) s'effectue par un acte d' « incorporation », qui s'établit par un vote et qui entraîne la création d'un gouvernement local, s'inscrivant dans un cadre urbain (*cities*, *towns*), rural (*villages*), ou dans celui d'un district (*boroughs*). La difficulté provient de l'absence d'adéquation qui peut exister parfois entre le statut choisi par les habitants et le caractère urbain de la localité (au sens morphologique du terme, à savoir une concentration d'habitats rapprochés dans l'espace et dépassant un certain seuil de population). Des critères statistiques sont donc préconisés par le bureau du recensement dès 1870, mais qui fluctuent au cours du temps (8000 habitants en 1874, puis il s'abaisse à 4000 habitants en 1880 et 2500 à partir de 1910), ce qui traduit un certain désarroi face à la difficulté de regrouper sous un ensemble de critères uniques la diversité des situations rencontrées selon les Etats. Dans ceux de la Nouvelle Angleterre notamment (Connecticut, Maine, Massachusetts, New Hampshire, Rhodes Island et Vermont), le maillage administratif date de la période coloniale et présente des particularités fortes. Ainsi, les *city* et/ou *town* (il existe des variations selon les Etats) correspondent à des subdivisions territoriales reconnues sur le plan législatif, pouvant contenir d'autres entités bénéficiant ou non d'une reconnaissance législative, s'appelant *village*, *borough*, etc. D'un recensement à un autre, le *Census Office* propose de traiter à part la Nouvelle Angleterre en variant les solutions préconisées d'une date à une autre, jusqu'en 1930, date où les premiers critères morphologiques (densité et population minimale) apparaissent. Les travaux en cours visent à réunir un ensemble de critères pour décider, au cas par cas, de l'intégration ou non de chaque city et town dans la base : statut de l'entité au recensement de 1930 (en fonction de sa densité de population), trajectoire de population sur l'ensemble de la période, statut de l'entité au recensement de 1940 (intégration ou non dans une aire urbaine fonctionnelle, appelée *Standard Metropolitan Area*), population moyenne sur l'ensemble de la période etc.

Un deuxième résultat consiste dans la construction d'un dictionnaire des changements, permettant de suivre les trajectoires des entités élémentaires composant les villes dans la base. Ce dictionnaire n'avait pas été intégré dans la version de la base construite en 2006-2007. Une première phase d'expertise a recensé systématiquement les types de changements survenus d'un recensement à un autre (séparation, fusion, scission, annexion, changement de noms, changement de statut, etc.). Une typologie des changements possibles a ensuite été réalisée, de manière à établir la filiation des entités lorsqu'elles se scindent, fusionnent ou disparaissent (Tableau 5). Les cas identifiés dans cette typologie sont : l'annexion, l'absorption, le fractionnement par scission, la sécession, l'apparition. De même,

une entité peut être contenue dans une autre, contenir une autre entité ou recouvrir les limites de cette entité[7].

Figure 9 : Exemple de recouvrement dans la Nouvelle Angleterre : Ansonia town et Ansonia city en 1900 (Connecticut) (François Delisle, 2011)

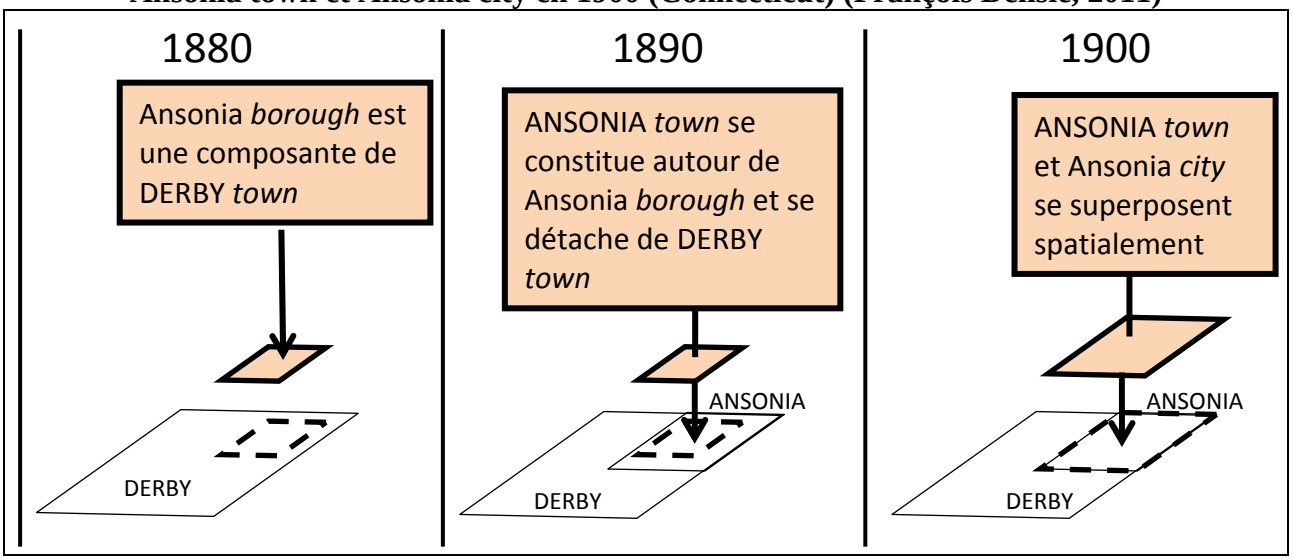
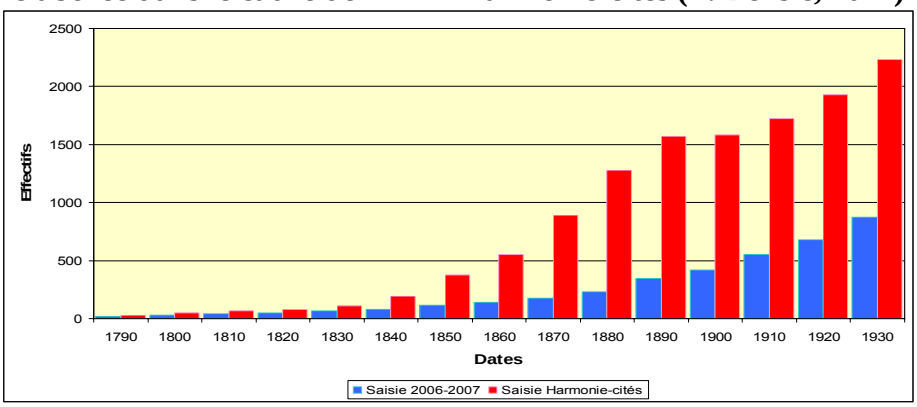


Tableau 5 : Nombre de changements enregistrés, selon la typologie retenue (F. Delisle, L. Lizzi 2011)

Type de changement	Annexion/ Absorption	Scission/ Sécession	Inclus/ est inclus	Recouvrement
Nombre total de cas, sur l'ensemble de la période	111	48	46	16

Un troisième résultat consiste dans l'importance numérique des changements apportés. Le nombre total de briques de bases (entités élémentaires composant les villes) est passé de 874 à 2334, sur l'ensemble de la période concernée (1790 à 1930) (Figure 10).

Figure 10 : Nombre d'entités élémentaires composant les villes des Etats-Unis dans la base 2006-2007 (Bretagnolle et al. 2008) et dans la nouvelle version élaborée dans le cadre de l'ANR Harmonie-cités (F. Delisle, 2011)



[7] Cette typologie se fonde sur des relations de réciprocité entre deux unités, par exemple entre annexion et absorption, entre fractionnement par scission et sécession (par exemple, sur la Figure 9, lorsque Derby perd une partie de son territoire qui sert à l'apparition d'Ansonia). Dans le cas de la Nouvelle Angleterre, lorsqu'une entité est contenue dans une autre, cela signifie que cette-ci contient la première. Enfin, lorsqu'un *borough*, *city* ou *village* étend ses limites jusqu'à coïncider celles de la *town* le/la contenant, cela est signalé pour les deux entités concernées.

Les différentes missions menées sur le terrain sud-africain ont permis de collecter des données présentes dans des recensements antérieurs à la formation de l'Union sud-africaine en 1911. On a ainsi pu exploiter des données dès 1865 (pour la seule Province « mère » du Cap) puis d'autres années, en fonction des dates de recensements variables d'une province à l'autre. Ces archives nous ont permis d'intégrer de nouvelles localités urbaines dans la base de données sur les populations et de proposer une plus grande temporalité dans l'analyse de la mise en place et de l'émergence des villes, en fonction des différentes avancées des fronts de colonisation et des découvertes minières, à l'origine d'un puissant mouvant d'urbanisation du territoire (Tableau 6).

Tableau 6 : Nombre de localités par province en Afrique du Sud 1865-1904

	1865	1875	1880	1889-90-91	1904
Orange			14	31	41
Transvaal				20	
Cape	2	19		19	
Natal				22	50

Pour **l'Inde**, l'analyse des trajectoires des villes indiennes a été effectuée des années 800 aux années 2000. Pour la période 800-1891, le nombre de villes pour lesquelles les données de population ont pu être associé est faible (moins d'une cinquantaine, tout au moins jusqu'en 1871). Néanmoins ces données ont permis de suivre l'évolution de villes telles que Delhi, Agra ou Vârânasî, qui ont perduré tout au long de l'histoire. A partir de 1871, le nombre de ville est suffisamment conséquent pour réaliser des études statistiques.

Pour la période 1961-2001, l'ensemble des agglomérations morphologiques (bâti urbain éloigné de moins de 200 m) de plus de 1000 habitants a été couvert, ce qui permet d'analyser l'évolution de l'urbanisation sur l'ensemble des villes indiennes sur une période de près d'un demi-siècle. En particulier, il est possible de suivre l'évolution de « gros villages » et de petites villes qui ne figurent pas dans les recensements officiels. Cette base historique sera complétée dans des travaux futurs en analysant les données de recensement disponibles en Inde et non encore exploitées.

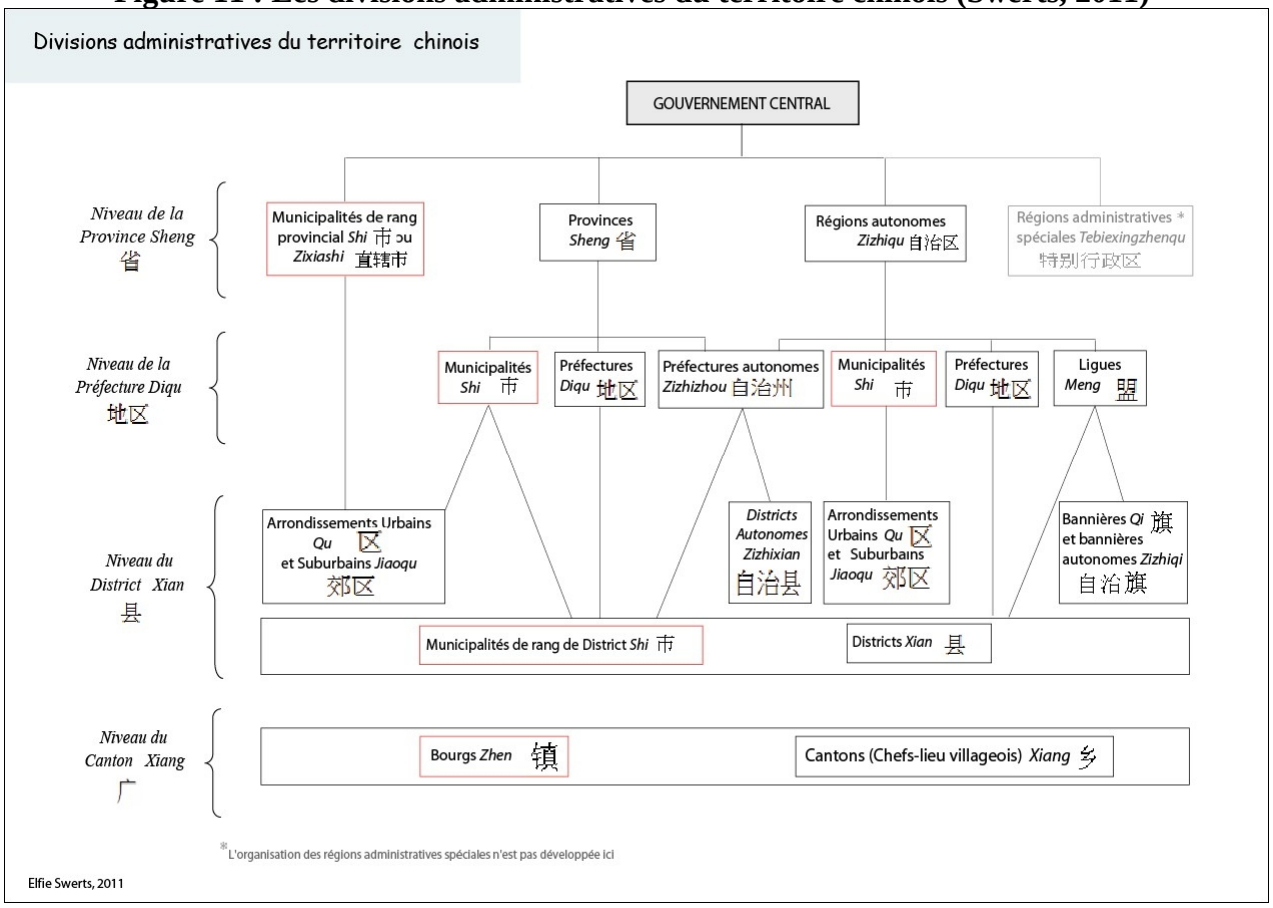
Tableau 7 : Nombre de villes dont les données de population ont été associées pour reconstituer les trajectoires des villes indiennes (Swerts, 2011)

Année	Nombre de villes	Année	Nombre de villes	Année	Nombre de villes	Année	Nombre de villes
800	9	**1600**	27	**1881**	250	**1941**	1398
1000	6	**1700**	22	**1891**	205	**1951**	1641
1200	4	**1750**	32	**1901**	1152	**1961**	6202
1300	7	**1800**	23	**1911**	900	**1981**	6202
1400	6	**1850**	36	**1921**	1253	**1991**	6202
1500	16	**1871**	124	**1931**	1312	**2001**	6202

Un deuxième chantier, qui n'avait pas été prévu dans la réponse à l'appel d'offre de l'ANR, concerne la Chine. Dans le cadre de la thèse d'Elfie Swerts, des travaux ont pu être menés sur ce pays en 2010 et 2011. Rappelons l'importance de l'enjeu consistant à établir des bases de données pour suivre la croissance des villes dans ce pays : l'urbanisation est d'une ampleur sans précédent et pour une large part

non maîtrisée. Enrichir la connaissance du développement urbain en Chine est ainsi un enjeu à la fois humain, économique et écologique. C'est également un enjeu théorique car la Chine présente des singularités politiques et historiques qui pourraient remettre en cause les modèles déjà établis pour les systèmes de villes d'autres régions du monde dans leur capacité à expliquer la coévolution des villes au sein d'un territoire. Identifier les principaux caractères et les éventuelles spécificités de l'urbanisation chinoise nécessite cependant une expertise préalable du poids politique et du rôle des villes dans le développement économique de ce pays hors normes. Les données du BNSC (Bureau National des Statistique Chinois) fournie par le Data China Center (Université du Michigan), acquises dans le cadre de cette ANR a permis de mettre en évidence l'importance croissante de la place des villes dans le système administratif et territorial de la Chine (Figure 11). L'apparition progressive de la ville dans toutes les strates de la hiérarchie administrative du pays rend complexe les analyses portant sur la mesure de l'évolution de la croissance des villes au cours du temps. Dans ce contexte, l'analyse des données fournies par le Data China Center a permis de poser les jalons de la construction d'une base de données harmonisées selon la même méthode que pour la base Indiapolis, citée plus haut (association au sein d'un SIG d'agglomération morphologique et de points de population). Elle permettra également de poursuivre la réflexion engagée sur la mesure de l'objet « ville », et en particulier de repenser le concept d'aire fonctionnelle, dans un contexte culturel, politique et historique différent de celui de l'Europe et des Etats-Unis.

Figure 11 : Les divisions administratives du territoire chinois (Swerts, 2011)



Ces résultats ont été présentés aux Rencontres interdisciplinaires sur les systèmes complexes naturels et artificiels de Rochebrune (16-22 janvier 2011), « Echelles et modélisations multi-niveaux », et font l'objet d'une publication (Bretagnolle, Mathian et Swerts, à paraître).

3.3 Bases de données sur les activités et les sociétés urbaines

Le deuxième axe d'analyse de l'ANR concerne les profils d'activités économiques et les spécialisations des villes. Les résultats présentés ici concernent 3 pays : les Etats-Unis, l'Afrique du Sud et la France, à la fois suivant une approche spécifique à chaque système de villes dans son contexte historique et à la fois dans une approche comparative dans l'objectif de faire émerger des faits stylisés et intégrables dans une théorie évolutive des villes. Les modalités du changement urbain, déjà identifiées dans le cas de la France, sont-elles aussi pertinentes et valables dans d'autres systèmes de villes et notamment ceux des pays neufs comme les USA et l'Afrique du Sud ?

Les premières investigations ont consisté à s'assurer de la cohérence des données collectées sur l'emploi et les catégories sociales en les confrontant avec d'une part, les résultats issus des bases sur les populations des villes (partie 1 du projet) et d'autre part avec des indicateurs globaux produits par d'autres sources, comme le taux d'emploi urbain ou l'évolution du nombre d'emplois dans certains secteurs d'activités. Ces premiers résultats ont confirmé la robustesse des bases de données constituées et ouvrent la voie aux analyses plus approfondies.

3.3.A – Innovations et dynamiques urbaines

La théorie urbaine évolutive (Pumain 2010) postule que l'innovation est le moteur de la dynamique des systèmes de villes. Nous entendons par innovation une invention qui est devenue socialement acceptée. Plus précisément, nous considérons un cycle d'innovation comme un ensemble de nouveaux produits, de nouvelles activités économiques, de nouvelles professions et les nouvelles pratiques sociales qui en découlent et qui émergent plus ou moins simultanément sur une courte période de temps. Ces cycles d'innovation ne sont pas ubiquistes mais sélectionnent dans un premier temps quelques villes, avant de se diffuser dans le reste du système.

La diffusion de ce cycle suit le plus souvent un canal hiérarchique : c'est en effet dans les plus grandes villes qu'il émerge parce que ces dernières présentent, « par accumulation antérieure une diversité d'activités et une complexité sociale qui (les) favorisent » (Pumain, 2010) ou augmente leur capacité d'adaptation.

Ces lieux pionniers de l'innovation profitent de ces nouveautés en produisant et en captant plus de richesses et d'emplois. Selon ce schéma, les villes les plus grandes sont d'autant plus assurées d'une croissance sur le long terme grâce à leur plus grande capacité à capter tous les cycles d'innovation. Les villes de taille intermédiaire et les plus petites peuvent être pionnières à un moment donné pour un cycle précis mais leur spécialisation peut aussi devenir un handicap dans la captation de cycles ultérieurs. Cette approche théorique permet d'expliquer l'adaptation générale des systèmes de villes au changement socio-économique qui maintient les structures fondamentales de ces systèmes (hiérarchie urbaine et variété des fonctions).

Afin de tester cette approche théorique dans des systèmes de villes à la genèse et à l'histoire différentes, les analyses s'organisent dans un premier temps selon une

approche transversale, à une date donnée (2000), puis dans un deuxième temps, privilégient l'approche longitudinale (1960-2000).

3.3.B – Analyse transversale menée pour les trois pays (France, USA, Afrique du Sud) en 2000

3.3.B.a – Lois d'échelles

Les structures des systèmes de villes, systèmes complexes, peuvent être exprimées sous la forme de lois d'échelles (*Scaling laws*) qui ont été reconnues comme spécifiques de leur dynamique. Une loi d'échelle s'exprime de la manière suivante :

$$x = y^{\beta}$$

(où *x* est le nombre d'emplois dans un secteur d'activité et *y* est la taille de la ville)

A un moment donné, on fait l'hypothèse que les activités les plus innovantes se retrouvent dans les plus grandes villes ; les activités banales sont plus ubiquistes tandis que les activités matures voire déclinantes sont concentrées dans quelques villes petites et moyennes. Ces trois stades correspondent à trois paramètres de lois d'échelles :

- *Innovant, moteur (β > 1)*
- *Banal (β ≈ 1)*
- *Mature, déclinant (β <1)*

Les résultats obtenus sur les trois pays confirment cette hypothèse (Figure 12, Tableau 8). Ainsi, les services de finance et d'assurance, moteurs du développement économique des villes dans un contexte mondialisé, apparaissent avec un *β> 1,* quel que soit le pays.

Figure 12 : Paramètres de lois d'échelle supérieurs à 1 pour les secteurs innovants : l'exemple de la finance et de l'assurance (Paulus, Vacchiani-Marcuzzo 2010)

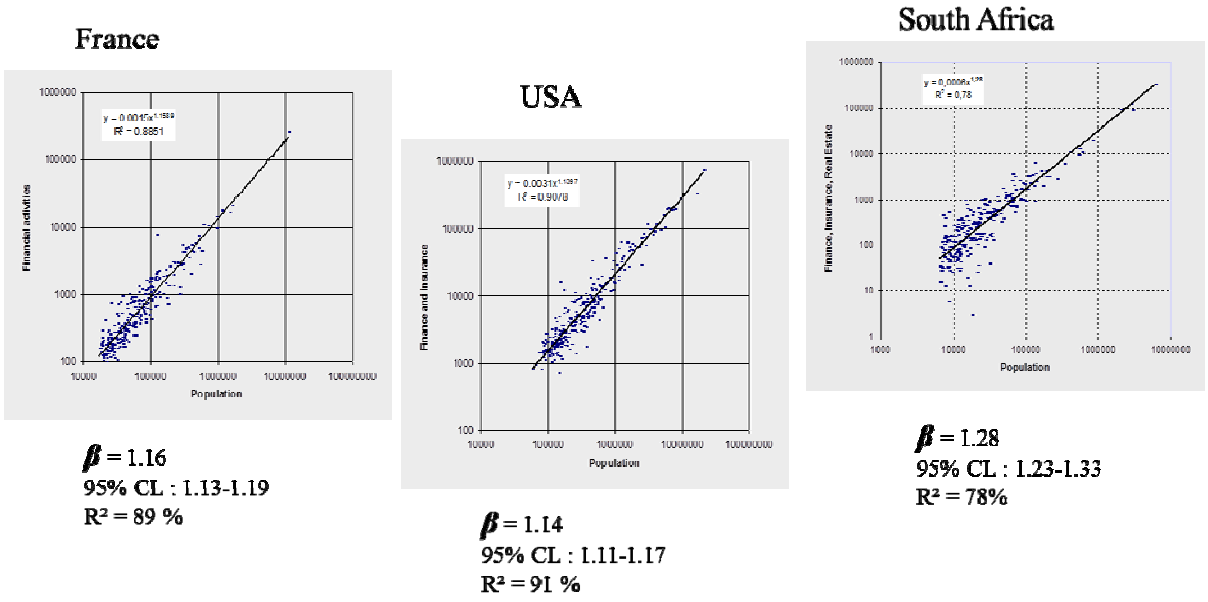


Tableau 8 : Paramètres de lois d'échelles par secteur, pour les trois pays (Paulus, Vacchiani-Marcuzzo 2010)

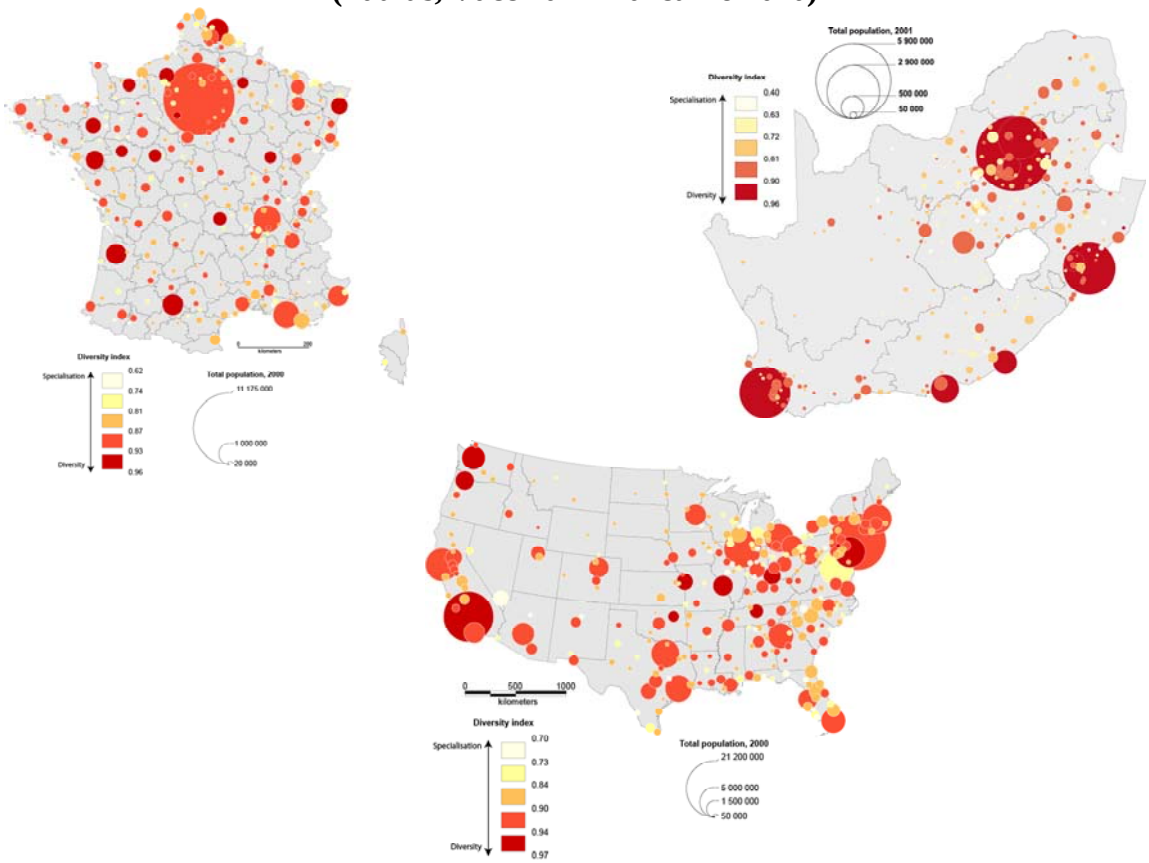
Stages in innovation cycle	France	USA	South Africa
Innovative sectors ß > 1	- Financial activities, Insurance, Real Estate		
	- Research and development - Business services, Consultancy		- Manufacturing
Common sectors ß ˜ 1	- Hotels and Restaurants - Community, social, personal services		- Retail Trade - Community, social and personal services - Utilities
Mature sectors ß <1	- Manufacturing - Retail Trade	- Retail Trade - Utilities	- Private Households

Le Tableau 8 montre que la répartition de l'emploi par secteur économique dans les villes est fortement dépendante de leur taille et en lien avec leur position dans leur cycle de vie. L'Afrique du Sud marque ainsi son décalage (et par là, son rattrapage en cours) par rapport aux USA et à la France puisque les industries font encore partie des activités innovantes, se concentrant encore dans le haut de la hiérarchie urbaine (.*β> 1).*

3.3.B.b – Diversité fonctionnelle des villes

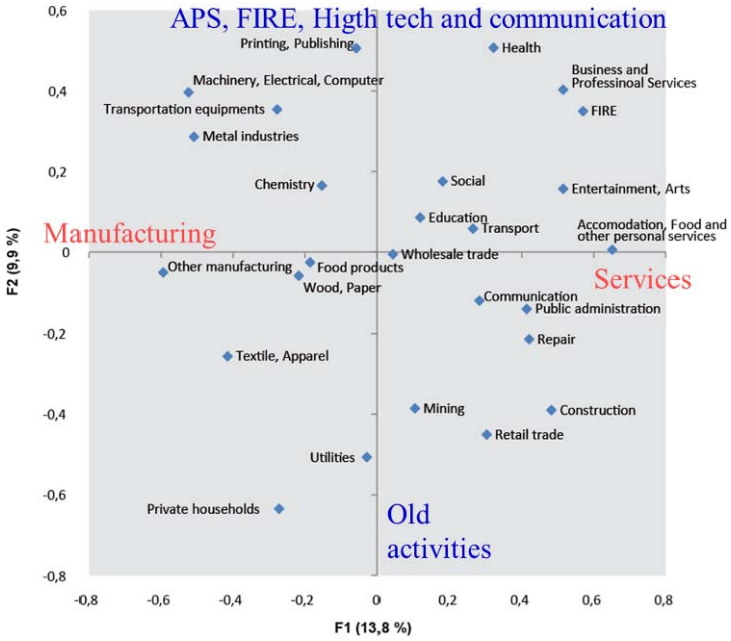
Les grandes villes étant ainsi par excellence des matrices de l'innovation, leur base fonctionnelle présente une diversité d'autant plus forte. Cette insertion a été vérifiée en calculant des indices de diversité fonctionnelle sur les trois cas d'étude (Figure 13).

Figure 13 : Indices de diversité économique des villes en 2000 (Paulus, Vacchiani-Marcuzzo 2010)



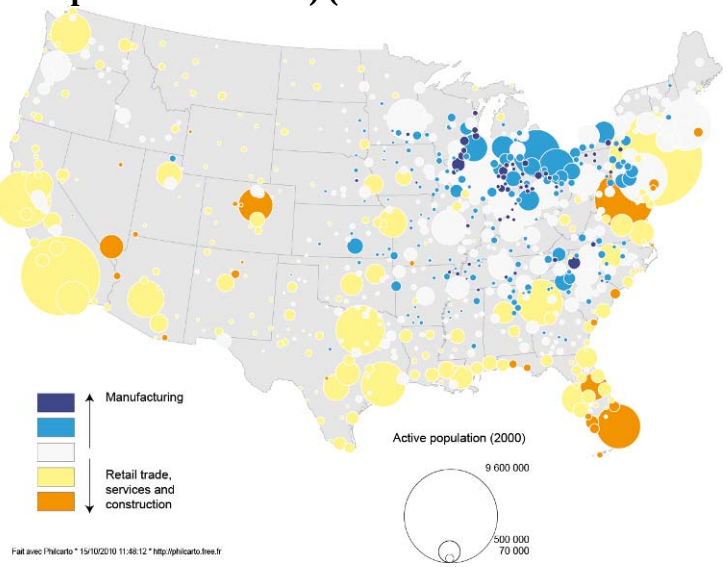
Pour approfondir la nature de ces spécialisations, nous avons mené des analyses factorielles sur les activités économiques présentes dans les villes. L'investigation a porté sur les trois pays, montrant des similitudes dans la structuration de l'espace économique des villes. Les résultats présentés ici portent sur les Etats-Unis. La Figure 14 montre les résultats d'une Analyse en Composantes Principales (corrélation des secteurs d'activité économique avec les deux premières composantes).

Figure 14 : Principales dimensions de la différenciation économique des villes aux Etats-Unis en 2000 (Paulus, Vacchiani-Marcuzzo 2010)



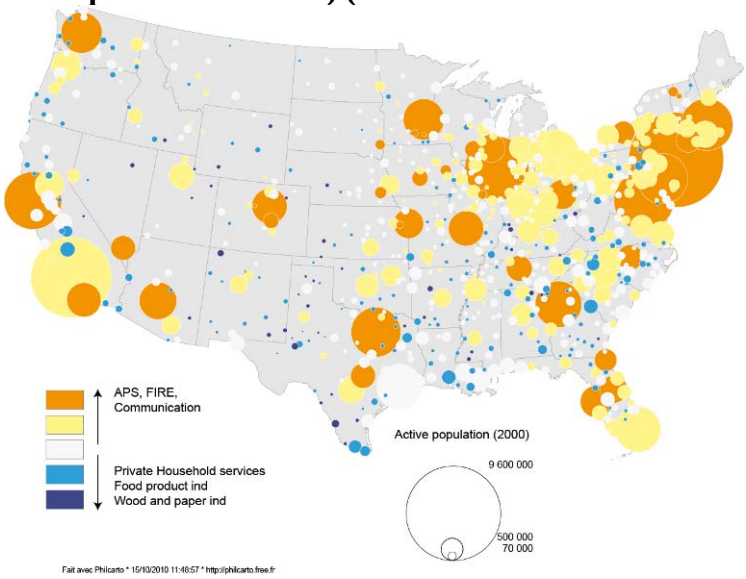
La cartographie des coordonnées des villes sur la première composante révèle l'opposition régionale ancienne entre villes industrielles et villes tertiaires reste la principale différenciation du système des villes en 2000 (Figure 15).

Figure 15 : L'opposition entre villes industrielles et tertiaires (première composante de l'ACP) (Paulus et Vacchiani-Marcuzzo, 2010)



La cartographie des coordonnées des villes sur la deuxième composante révèle, quant à elle, une opposition entre des villes qui concentrent les activités innovantes et d'autres spécialisées dans des activités plus banales et anciennes et suit, globalement, la hiérarchie urbaine (Figure 16).

Figure 16 : L'opposition entre villes spécialisées dans des activités innovantes et celles caractérisées par des activités plus banales et anciennes (deuxième composante de l'ACP) (Paulus et Vacchiani-Marcuzzo, 2010)



Ces différenciations viennent renforcer les premiers résultats globaux issus de la méthode des lois d'échelle avec une structuration nettement hiérarchique des activités du cycle contemporain. Pour autant, la principale différenciation reste l'héritage de cycles plus anciens (structuration plus régionale). Ces interprétations liant innovation, taille et dynamique urbaine, ont été davantage étayées par le prisme d'analyses longitudinales.

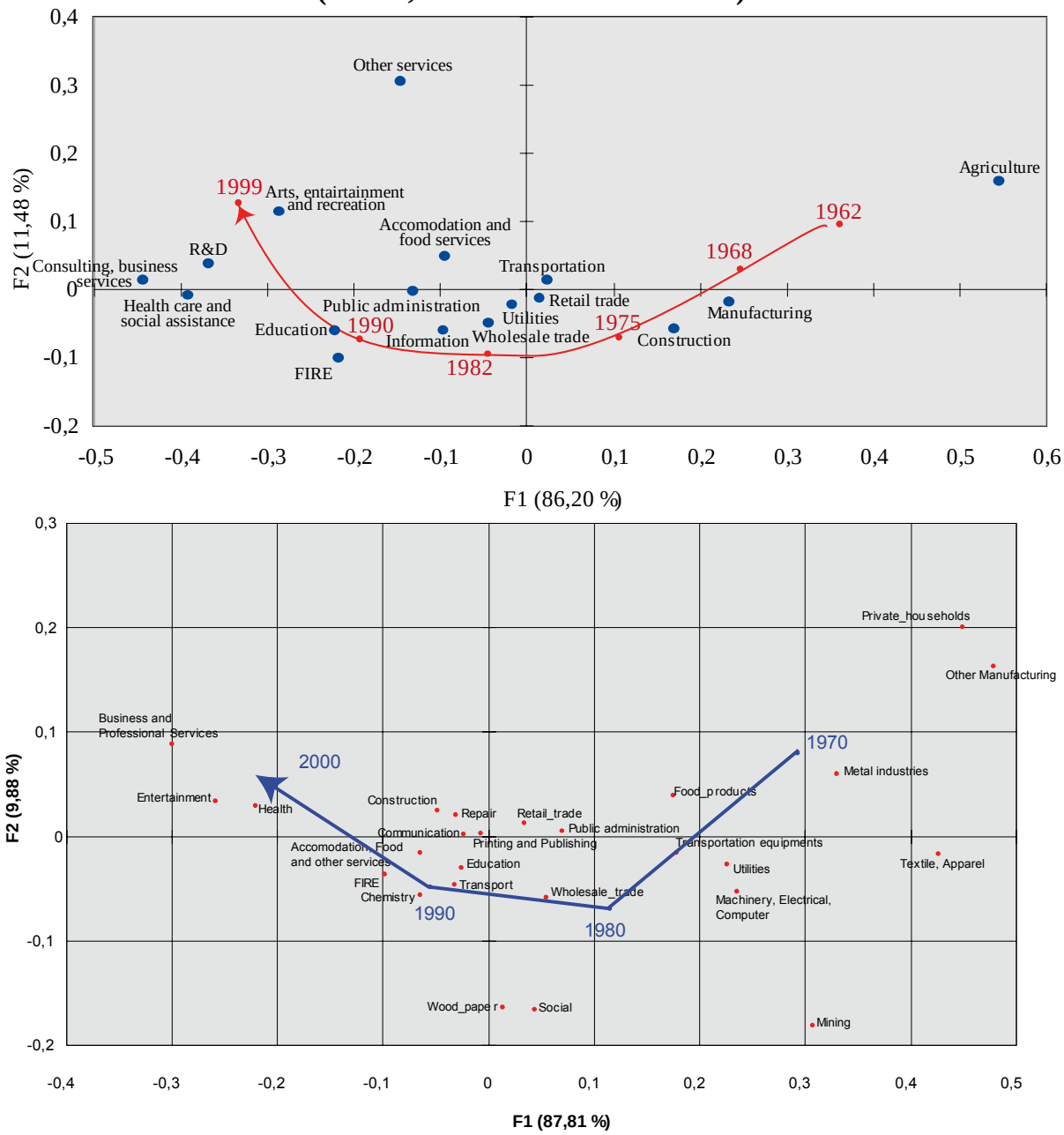
3.3.C – Analyse longitudinale menée pour deux pays (France, USA) 1960-2000

Dans un premier temps, les analyses longitudinales ont été menées dans les cas états-unien et français et dans des périmètres urbains fixes dans le temps, ceux de 1999 pour la France et 2000 pour les Etats-Unis. La constitution des bases de données spatio-temporelles dans le cas sud-africain ont nécessité des investigations plus lourdes et plus longues (mission sur le terrain, collecte de donnés sur papier et saisie informatique) et n'ont pu être analysées dans la temporalité du projet ANR.

Les traitements des bases de données inédites sur les Etats-Unis et la France dans le temps long permettent d'approcher les correspondances entre cycles de croissance urbaine, cycles d'innovations économiques et trajectoires de villes sur la deuxième moitié du XX^{ème} siècle.

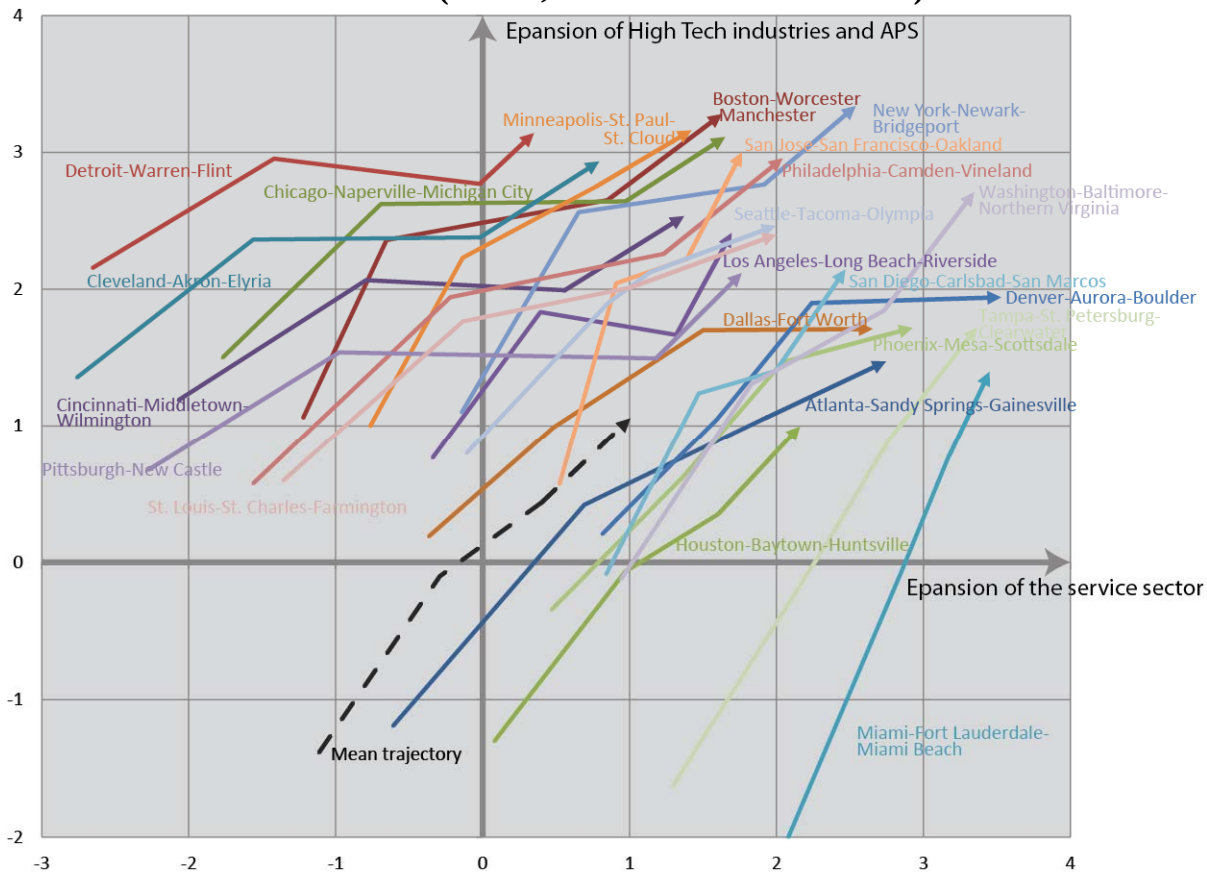
Les graphiques (Figure 17) mettent en avant la trajectoire globale de la transformation des profils économiques des villes de 1962 à 1999 en France et de 1970 à 2000 aux Etats-Unis. Dans les deux cas, ces trajectoires sont similaires et montrent le déclin relatif de l'industrie au profit du développement des services avancés et destinés aux entreprises, de la recherche et du développement, de la santé et des activités culturelles et de loisirs.

Figure 17 : Activités économiques et tailles des villes au cours du temps (En haut : France 1968-1999, En bas : Etats-Unis 1970-2000) (Paulus, Vacchiani-Marcuzzo 2010)



Une des dernières investigations menées dans le cadre du projet ANR a été de s'interroger sur la manière dont chaque ville, prise individuellement, s'est adaptée à ce changement global. L'analyse factorielle multi-tableaux permet de dessiner les trajectoires des villes dans l'espace de leurs différenciations économiques au cours du temps (Figure 18).

Figure 18 : Trajectoires économiques des grandes aires métropolitaines aux Etats-Unis, 1970-2000 (Paulus, Vacchiani-Marcuzzo 2010)



La similitude de ces trajectoires traduit la cohérence des évolutions. Ce résultat fort montre à quel point la dynamique d'une ville n'est pas isolée mais au contraire à l'intersection de multiples interactions qui la lient aux autres villes, ce que nous appelons *la coévolution des villes en système***.** Néanmoins, des décalages ou décrochages peuvent s'observer. Deux cas de figure ressortent : l'un concernant les villes plus petites et moyennes en raison de facteurs plus spécifiques et plus locaux, l'autre relevant de la très forte spécialisation de grandes villes, davantage freinées dans le processus de captation de nouvelles activités et d'adaptation. Dans le premier cas, ce frein est aussi lié à un profil social moins enclin à l'innovation.

L'analyse de ces trajectoires a déjà été faite pour la France et l'intérêt s'est donc porté sur les Etats-Unis. Le résultat majeur de ces analyses est la confirmation de processus communs dans l'évolution fonctionnelle des systèmes de villes et la temporalité de l'évolution urbaine. Celle-ci distingue i. une dynamique rapide d'ajustement (les fluctuations des trajectoires sur le cours terme 1-10 ans), ii. une dynamique continue banale liée à la diffusion des innovations et iii. une dynamique lente des transformations relatives des villes avec l'émergence d'un nouveau cycle d'innovation.

Ce résultat conduit à une investigation des catégories sociales car nous faisons l'hypothèse que les villes les plus innovantes sont celles qui concentrent les populations les plus qualifiés. En effet, ce qui semble relever de l'évidence dans notre société actuelle de la connaissance et de l'information est sans doute un élément majeur et ancien dans l'explication de l'avantage des villes à innover ou à capter l'innovation. Une première approche de cette hypothèse a été menée par les

analyses transversales en 2000 en France et aux USA (Figure 19, France, et Figure 20, Etats-Unis). Dans les deux cas, la relation entre taille des villes et qualification professionnelle est forte et montre la transition des plus petites villes aux plus grandes. Les plus petites accueillent davantage des populations peu qualifiées (ouvriers non qualifiés), les villes intermédiaires des employés et ouvriers plus qualifiés, tandis que les plus grandes villes concentrent les cadres et techniciens supérieurs. L'analyse de l'évolution du lien entre profils sociaux et taille des villes est en voie d'achèvement pour les Etats-Unis.

Figure 19 : Une autre représentation de loi d'échelle entre taille des villes et catégories socioprofessionnelle, France, 1999 (Paulus et Vacchiani-Marcuzzo, 2010

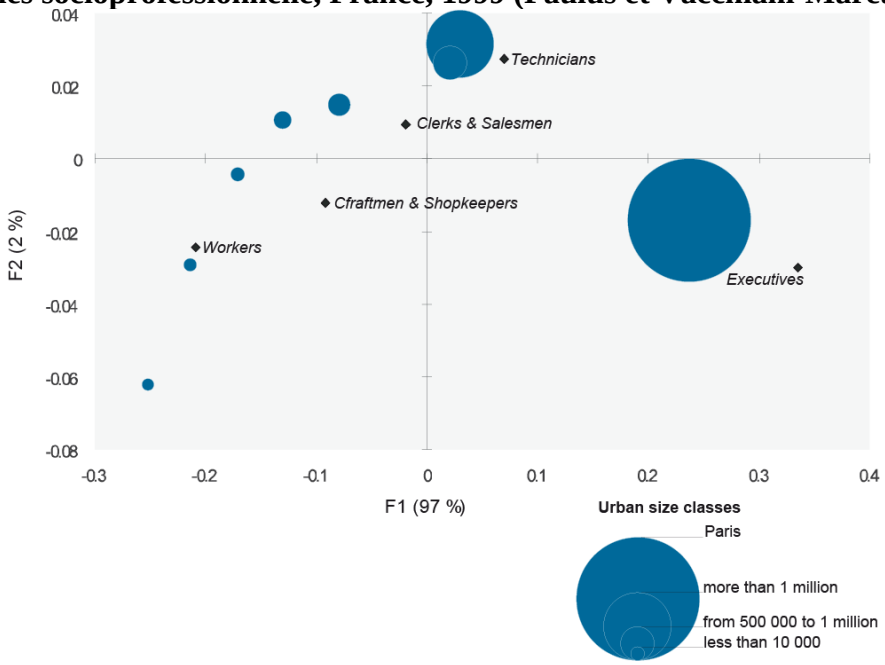
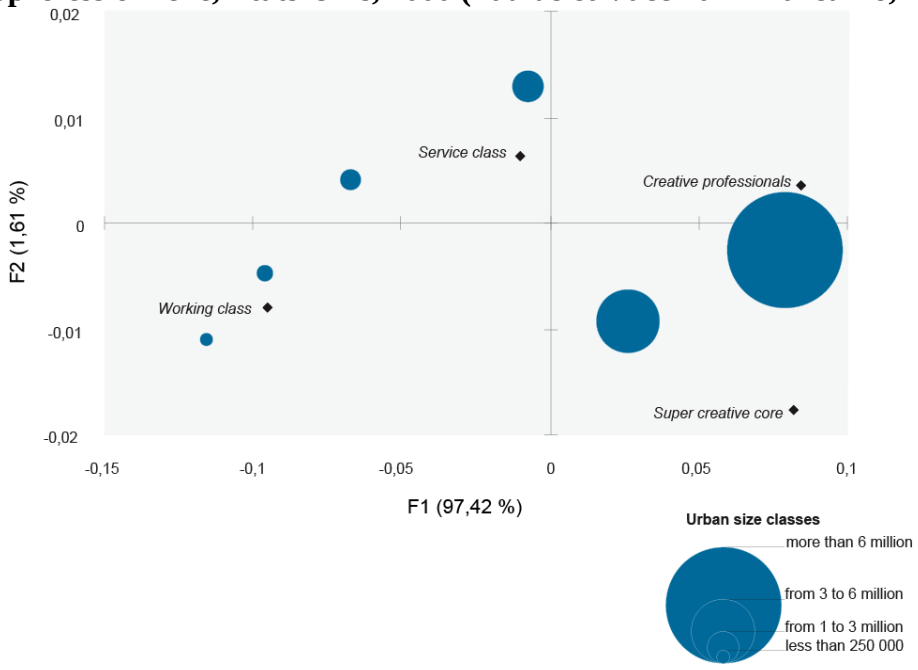


Figure 20 : Une autre représentation de loi d'échelle entre taille des villes et catégories socioprofessionnelle, Etats-Unis, 2000 (Paulus et Vacchiani-Marcuzzo, 2010)



La construction des bases de données cohérentes sur les fonctions urbaines en France, aux Etats-Unis et en Afrique du Sud a permis de progresser dans la description et la compréhension des transformations urbaines sur le temps long. L'exploitation fine, par des méthodes similaires, de ces données conduit ainsi à confirmer la théorie de l'évolution des villes qui lie innovation, taille et croissance.

3.4 Bases de données sur les réseaux de transport inter-urbains

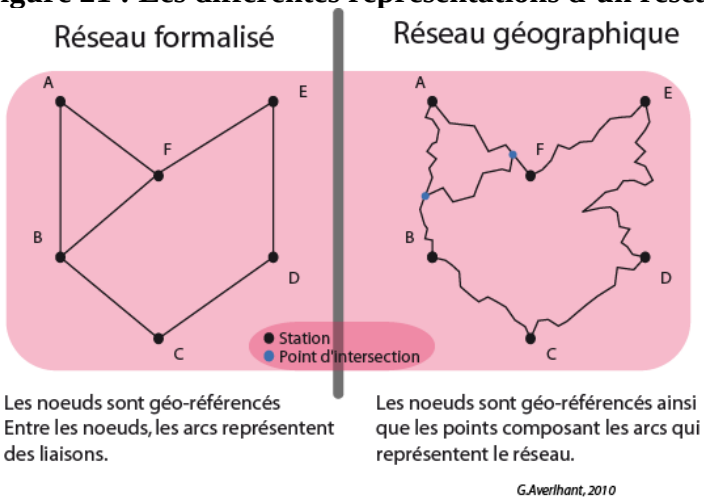
Les travaux sur les **méthodes génériques d'appariement** des bases de données réseaux et villes et sur les étapes de la généralisation cartographique des réseaux ont permis d'avancer plusieurs résultats.

L'enregistrement et la construction de base de données réseau suivent généralement une logique indépendante de celle qui a sous-tendu la modélisation des données pour l'analyse des dynamiques urbaines. Le réseau se construit sur la base d'arcs et de nœuds. Les nœuds peuvent être associés à un lieu ayant une sémantique ou non (gares de triage, carrefour d'aiguillage…) pour la question posée ou la résolution à laquelle le réseau est numérisé. La Figure 21 présente en exemple deux formalismes associés au réseau :

1- une représentation topologique du réseau par un graphe. Les nœuds seuls sont géoréférencés et peuvent être associés à des objets « villes ». Les arcs symbolisent des liaisons et n'ont pas d'inscription spatiale.

2- une représentation géographique du réseau. L'ensemble des arcs et des nœuds sont géoréférencés. Certains nœuds correspondent à des lieux identifiés, du type gare ou ville (points noirs), tandis que d'autres sont seulement des carrefours sans identité spécifique (points bleus).

Figure 21 : Les différentes représentations d'un réseau



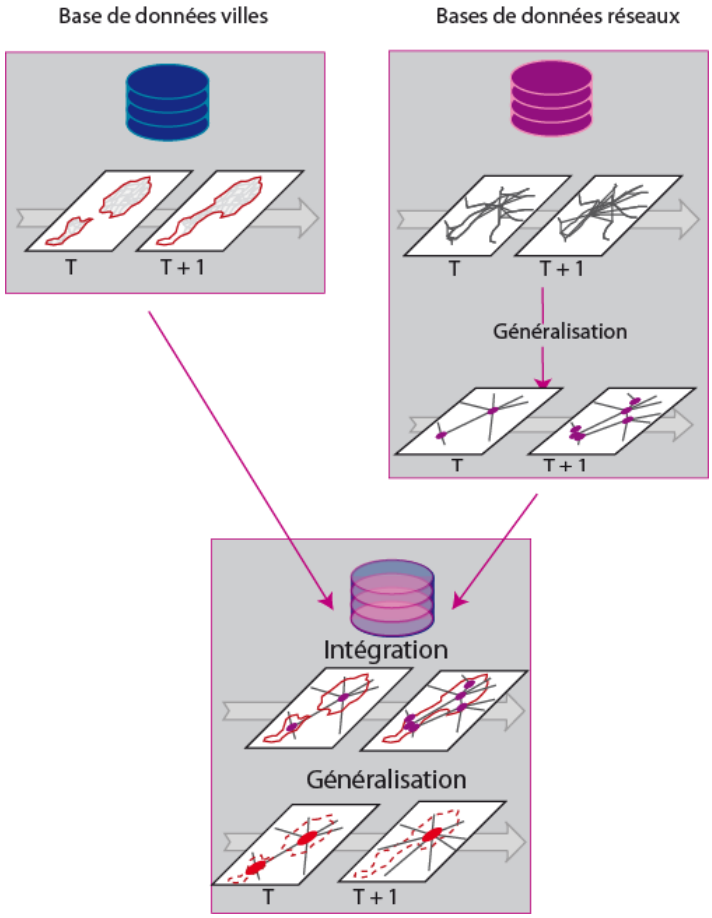
La Figure 22 illustre les étapes associées à la mise en relation de la base de données des villes et celle des réseaux. En considérant que l'on s'abstrait des questions de cohérence temporelle entre les différentes sources, la mise en relation nécessite :

1- D'apparier les nœuds « station » de la base réseau avec les objets « villes » de la base ville.

2- De généraliser le réseau sur la base des appartenances des proximités entre nœuds et villes.

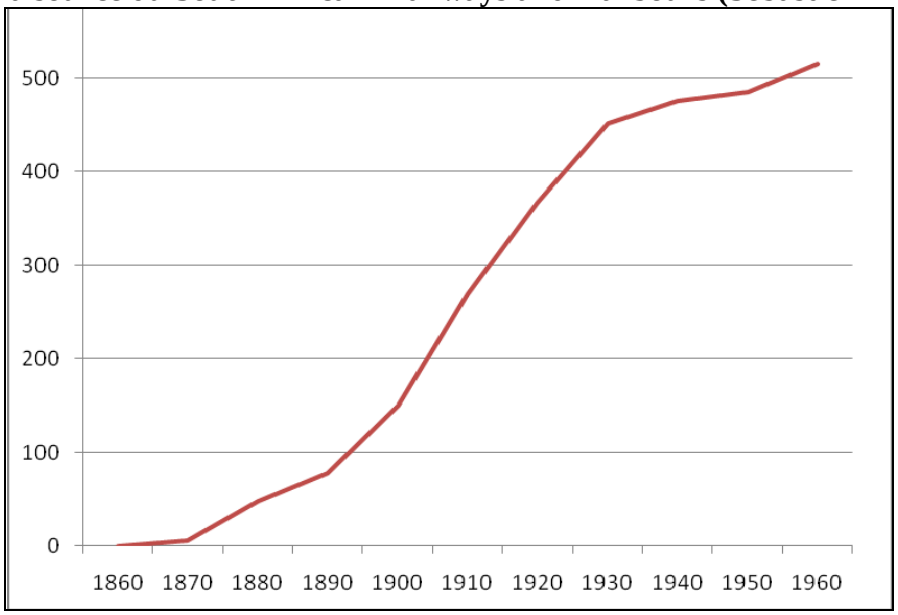
Cette illustration est générique. Selon les pays, le processus peut être différent, en fonction des sources existantes et du format des données.

Figure 22 : Etapes de l'intégration et harmonisation de la base ville et de la base réseau



Pour **l'Afrique du Sud**, l'ensemble de la période 1860-1960 a été couvet. Lle nombre de tronçons saisis à partir des sources du SAR&H est égal à 515 et le nombre total de stations ferroviaires est de 576 (Figure 23). Comme pour d'autres réseaux de transport innovants, on aboutit à une courbe de croissance logistique.

Figure 23 : Evolution du nombre de tronçons ferroviaires en Afrique du Sud, entre 1860 et 1960, selon la source du South Africain Railways and Harbours (Sébastien Martin, 2011)



L'étape du géo-référencement des stations ferroviaires a été menée au cours de l'année 2011. Ce géo-référencement a été effectué à l'aide de plusieurs bases de données : OpenStreet Map (production de données basées sur des enregistrements de données GPS) fournit des informations sur le réseau ferroviaire sud africain ainsi que sur les gares desservies, en format Shapefile. Un script effectuant des requêtes sur les données OpenStreet Map puis sur GoogleMaps (en cas d'échec avec la base OSM) a permis de géoréférencer automatiquement 79% des stations répertoriées dans le SAR&H. Les stations qui n'ont pu être localisées présentent des problèmes de correspondance avec les données OSM ou Google Maps; ou des problèmes de toponymie (entre l'afrikaneer et l'anglais).

L'étape en cours de réalisation consiste à recréer la géométrie du réseau, date après date. Outre l'utilisation des cartes anciennes, une piste possible est de repartir du tracé actuel et remonter dans le temps. Deux bases ont été testées, celle d'Open Street Map, relativement récente (2010) et en libre accès, et celle de la NGIA (National Geospatial Intelligence Agency), diffusant des données Vmap de niveau 0, libres de droits. La première base est exhaustive et précise (cette précision est relative au GPS utilisé lors de l'acquisition de la donnée). Les données Vmap ont été numérisées à l'échelle du 1:1 000 000 (système géodésique WGS84) et sont donc moins précises que les données OSM (écarts de l'ordre quelques kilomètres en moyenne). Mais elles ont l'avantage d'être plus généralisées. L'utilisation de ces tracés récents reste néanmoins en discussion, car les tracés ont pu être modifiés au cours du temps. Les cartes historiques sont donc incontournables pour intégrer à la base qui existe actuellement (une ligne par tronçon, avec un point de départ et un point d'arrivée) des informations sur les interconnexions entre les tronçons (Figure 24).

Figure 24 : Reconstruire la géométrie du réseau ferroviaire sud-africain à partir du croisement des sources (Sébastien Martin, 2011)

Year	Month	From	To	Reference	Cape of Good Hope	Natal	Transvaal	Orange Free State	South West Africa
1884	May	Pietermaritzburg	Merrival	Carte 1912		15_19			
1884	16th Sept	Sterkstroom	Molteno	Carte 1912	20_73				
1884	3rd Nov	De Aar	Orange River	Carte 1912	70_15				
1884	1st Dec	Grahamstone	Port Alfred	Carte 1912					
1885	19th March	Molteno	Burgersdorp	Carte 1912	32_76				
1885	2nd Sept	Burgersdorp	Aliwal North	Carte 1912	37_05				
1886	21st June	Estcourt	Ladysmith	Carte 1912		45_52			



Données SAR&H

+

Données Standard Railway
Map of South Africa

Year	Month	From	To	Interconnexion	Commentaire	Reference	Cape of Good Hope	Natal
1884	May	Pietermaritzburg	Merrival	Cedara;Hilton Road;Boshoff's Road;Winter's Kloof;Sweetwaters;Blackridge	Boshoff's Road = Boshoffweg;	Carte 1912		15_19
1884	16th Sept	Sterkstroom	Molteno	Lower Incline;Carrickmore;Cyphergat	Cyphergat = Syfergat	Carte 1912	20_73	
1884	3rd Nov	De Aar	Orange River	Paardavlei;Behrshoek;Houtkraal;Baartman;Potfontein;Kalkbult;Paauwpan;Krankuil;Wickham	Paardavlei = Perdevlei;Paauwpan=Poupan	Carte 1912	70_15	
1884	1st Dec	Grahamstone	Port Alfred	Manley Flats;Martindale;Trapps Valley;Clumber		Carte 1912		
1885	19th March	Molteno	Burgersdorp	Stormberg		Carte 1912	32_76	
1885	2nd Sept	Burgersdorp	Aliwal North	Druenberd	(Actuellemnt Dreunberg)	Carte 1912	37_05	
1886	21st June	Estcourt	Ladysmith	Ennersdale Junction;Heavitree;Frere;Chieveley;Colenso,Harts Hill		Carte 1912		45_52

Enfin, deux méthodes sont envisagées pour l'intégration spatiale de la base des tronçons : utiliser les tracés actuels (OSM, Vmap0…), ou écrire un script permettant de générer de manière linéaire chaque tronçon.

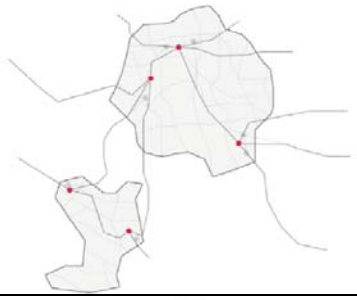
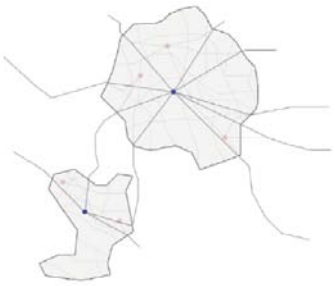
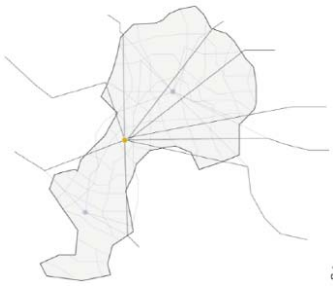
Pour la **France**, une analyse de faisabilité pour le croisement avec la base de données réseau (système d'information géo-historique du système ferroviaire et démographique développé par Thomas Thévenin et son équipe à l'université de Bourgogne) a été entamée. Le système d'information représente 11350 gares localisées et 28000 km de lignes.

Des recommandations des auteurs de cette base ont été faites pour la première étape de généralisation : il s'agit de ne pas considérer les réseaux locaux, et de ne garder que les lignes principales. Ce système d'information ayant été développé pour une analyse démographique, le réseau présente l'avantage d'avoir déjà été « raccordé » aux communes. L'opération d'intégration ne posera donc aucun problème, les villes étant dans le cas de la France définies à partir d'une agrégation de communes.

En revanche l'étape de généralisation nécessite des tests qui sont en cours. Il s'agit de dénombrer les différents cas de figures et de proposer les méthodes d'agrégation permettant de passer d'un réseau à l'échelon communal à un réseau à l'échelon des villes.

A titre d'exemple la Figure 25 présente le cas où plusieurs communes d'une même « ville » ont une gare. La question qui doit être réglée est le niveau de généralisation du réseau qui va être associé à l'intégration « gares-ville ».

Figure 26 : Les 2 étapes d'appariement des deux bases de données

	Etape d'intégration : réseau + ville	Etape de généralisation
T1 : 2 villes		
T2 : agrégation des 2 villes		

Pour la modélisation des interactions entre réseaux d'échanges et dynamiques des villes, un travail de préparation du **modèle multi-agents Simpop2** a été mené. En effet, une condition importante pour la poursuite des recherches en modélisation dynamique était le passage à une nouvelle génération de modèles (SimpopLocal et SimpopNet, thèses en cours de Clara Schmitt et Sébastien Rey) inspirés de Simpop2. Pour ce faire, une phase préalable et nécessaire a consisté à modulariser l'application Simpop2 et à porter la partie factorisable sur une plate-forme permettant son passage à l'échelle[8] : couplage au SIG, déploiement sur grille pour effectuer du calcul intensif, intégration dans des systèmes de management de workflow scientifique. Ces évolutions auraient été coûteuses à réaliser en gardant l'application Simpop2 en l'état, développée en Objective-C et couplée à la librairie de programmation Swarm, en voie d'obsolescence. Un redéveloppement, intégrant un travail de modularisation du code, de définitions d'interfaces et une documentation approfondie s'avérait nécessaire. Ce travail a été initié au cours d'une vacation d'un mois effectuée par Thomas Louail fin 2010. Il a été poursuivi et pérennisé lors d'un stage de 3 mois d'école d'ingénieurs à l'été 2011. Le cœur de l'application Simpop2 (correspondant à la partie générique de Simpop2) a été nettoyé et ré-écrit en Scala, et interfacé avec OpenMole, une application de création et de déploiement sur grille de workflows d'expérimentation numérique, ainsi qu'avec le SIG libre QGIS. Ce travail a servi à préparer l'intégration future de réseaux de transports inter-urbains dans les modèles dynamiques inspirés par Simpop et l'étude par simulation des liens entre accessibilité et trajectoire urbaine. Le code résultant de ce travail est sous licence libre et disponible au téléchargement sur la forge logicielle[9] de l'institut des systèmes complexes de Paris-Ile-de-France.

[8] Programme qui a été adapté de façon à ce qu'il puisse être exécuté par de nombreux utilisateurs successivement ou parallèlement pour accélérer le temps de calcul.

[9] Système de gestion collaboratif de logiciel. Il s'agit d'un outil en ligne qui offre des fonctionnalités telles qu'héberger et gérer l'historique des versions des programmes de la forge (codes sources et documentation), gérer

3.5 Interface Harmonie-cités

Les travaux réalisés sur l'interface en 2010 et 2011 ont permis d'achever l'application portant sur les Populations et espaces urbains, pour les Etats-Unis et pour la France. Plusieurs résultats principaux ont pu être obtenus.

Tout d'abord, conformément aux objectifs du projet, l'accent a été mis sur le caractère générique de l'interface, tant dans la formalisation des bases de données cartographiques que dans le choix et la présentation des outils d'interactivité, des formats de légende etc. Grâce à cette généricité, l'application Afrique du Sud devrait compléter prochainement la liste des pays disponibles, et les applications dédiées aux activités économiques des villes sont en cours de conceptualisation.

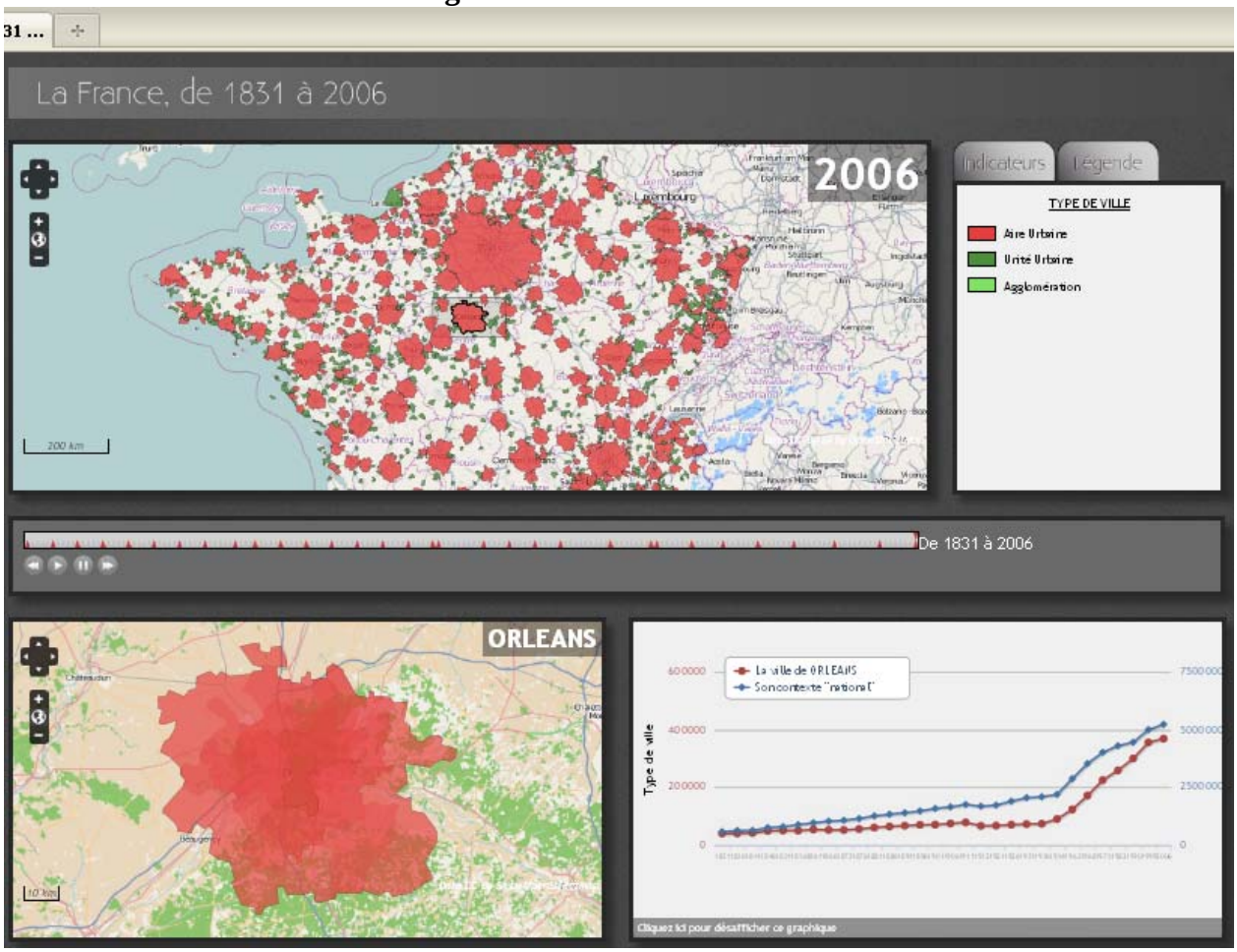
Ensuite, l'accent a été mis sur la dynamique, qui est au cœur de l'interface. La fenêtre est organisée selon trois sous-fenêtres, reliées dynamiquement par la flèche du temps, en position centrale, qui représente l'ensemble de la période (Figure 27). L'unité de temps est l'année. Les outils d'animation sont classiques (play,stop, pause…) et un curseur permet de suivre la date représentée au fur et à mesure de l'animation. L'animation automatique (touche Play) donne l'impression d'une continuité sur l'ensemble des deux siècles, au-delà des dates des évènements (années d'enregistrement des données dans la base), marquées par une encoche rouge. Le temps de l'animation est proportionnel au temps réel : le curseur se déplace sur la ligne de temps et la représentation cartographique change seulement lorsqu'il pointe sur une encoche. Nous avons choisi de ne pas faire d'interpolation des géométries entre deux années d'enregistrement afin de rester au plus près des bases de données. Par ailleurs, une telle interpolation n'aurait pas été possible dans le cas des Etats-Unis avant 1940, période pour laquelle les villes sont identifiées par les coordonnées de leur centroïde mais pas par une enveloppe spatiale.

L'utilisateur peut aussi choisir de déterminer lui-même les dates des représentations et leur succession. Dans ce cas, c'est lui qui conduit l'exploration du système des villes : le temps devient celui de l'utilisateur.

---

des listes de discussion entre développeurs, recenser et hiérarchiser les bugs sur lesquels il faut travailler, gérer les traductions des documentations, etc.

Figure 27 : L'interface France



Un deuxième ensemble de résultats concerne le caractère multi-scalaire de l'interface. Conformément aux objectifs initiaux, trois niveaux sont représentés : celui du système de villes (qui est le niveau de référence), celui de la ville, et celui des entités élémentaires qui composent la ville. La fenêtre supérieure permet de choisir le contexte systémique parmi différentes possibilités. Quelques outils sont installés (Déplacement et Zoom), d'autres devraient être prochainement rajoutés pour que le système choisi par l'utilisateur puisse être :

- un contexte administratif (Pays, Etat des Etats-Unis, Région administrative…)

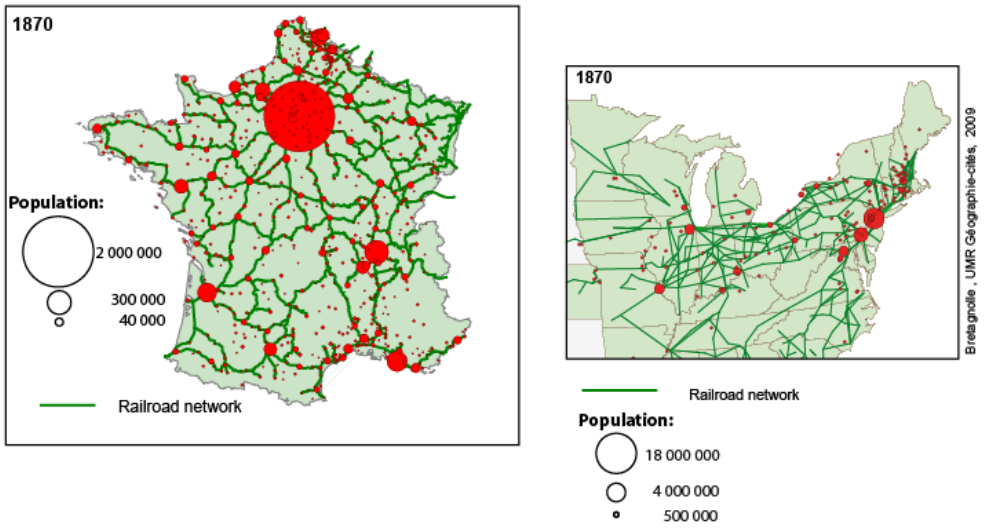
- un contexte sémantique (top 10 ou top 20, de l'année en cours, de l'année initiale ou de l'année finale)

- un contexte « expert » (proposé à l'utilisateur en fonction des connaissances sur les villes du pays et les réseaux de transport à telle ou telle époque, voir l'exemple de l'année 1870 pour la France et les Etats-Unis dans la Figure 28)

- un contexte « utilisateur » (rectangle de sélection).

Figure 28 : Deux contextes systémiques définis par l'expertise, en 1870, en fonction notamment des connexions ferroviaires

Gauche: un contexte national (France, 1870),
Droite: un contexte régional (Région des Grandes Lacs, 1870)

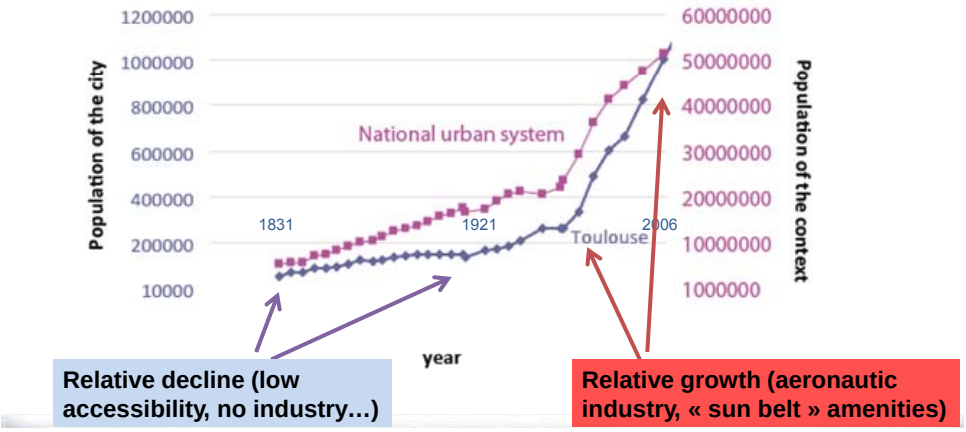


Sources: Railroad network digitalized from Caron 1997 Tome 1 Planche 11; Harmonized data base on French cities, Bretagnolle 1999.

Sources : US harmonized data base (Bretagnolle, Giraud 2008) and railroad network (Paffoni 2008, Thomas W., Railroads and the Roots of Modern America /Aurora Project, University of Nebraska)

Le choix du contexte est surtout important pour la mise en regard de la trajectoire de la ville et de celle du contexte systémique (fenêtre en bas à droite). Cette trajectoire est établie à partir d'un attribut, par exemple la population, la surface urbaine, la densité de population, le nombre d'actifs dans telle ou telle activité économique. Le fait de porter sur le même graphique l'évolution individuelle et celle du système permet d'apporter des éclairages tout à fait essentiels pour enrichir la connaissance de la ville. Par exemple pour Toulouse (Figure 29), on voit nettement apparaître deux périodes différentes dans l'évolution de la ville aux 19ème et au 20ème siècle, la première marqué par un déclin relatif (vraisemblablement lié à une faible accessibilité et à la faiblesse des activités industrielles), la seconde par une croissance relative (s'expliquant notamment par l'implantation de l'industrie aéronautique et par des aménités de type « sun-belt »).

Figure 29 : Une vue relative de la trajectoire d'une ville (Toulouse, France)



La troisième fenêtre (en bas à gauche) permet de mettre en regard l'évolution de la trajectoire de la ville et celle de son emprise spatiale. Pour Toulouse, par exemple,

la concordance est frappante entre la période de déclin relatif, où la ville reste composée d'une seule commune, et la période de croissance, marquée par un fort étalement spatial.

Les interfaces France et Etats-Unis ont été présentées au Salon Multimedia du Festival de Géographie de Saint-Dié (Van Hamme et al. 2010) et au Colloque International de Cartographie (Bretagnolle, Lizzi, Mathian, Van Hamme 2011).

Références

Bairoch P., Batou J., Chèvre P., 1988: *La population des villes européennes. Banque de données et analyse sommaire des résultats: 800-1850*. Centre d'Histoire Economique International de l'Université de Genève, 339 p.

Black Duncan, Henderson Vernon (2003), "Urban evolution in the USA". *Journal of Economic Geography 3*, 343-372.

Bretagnolle A. (2009), *Villes et réseaux de transport : des interactions dans la longue durée (France, Europe, Etats-Unis)*. Habilitation à diriger des recherches, Université Paris 1, 17 Juin 2009, http://tel.archives-ouvertes.fr/tel-00459720_v1/.

Bretagnolle A., Giraud Timothée, Mathian Hélène (2008), « La mesure de l'urbanisation des Etats-Unis, des premiers comptoirs coloniaux aux Metropolitan Statistical Areas (1790-2000), in Cybergeo [revue en ligne], 427, 40 p.

Bretagnolle A., Giraud T., Verdier N. (2010), « Modéliser l'efficacité d'un réseau : le cas des routes de poste en France (1632, 1833). *L'Espace Géographique*, n°2, 17 pages.

Bretagnolle A., Mathian H., Van Hamme A. (2011), "Animating the cities…Dynamic exploration of harmonized urban databases (United-States, France 1800-2000)", Proceedings of the 25th International Cartographic Conference (Paris, 3-8 july 2011), http://halshs.archives-ouvertes.fr/halshs-00624473/fr/ (**joint au dossier**)

Bretagnolle A., Pumain D. (2010), « Simulating urban networks through multiscalar space-time dynamics (Europe and United States, 17th -20th centuries) », *Urban Studies* 47, 13 (2010) 2819-2839.

Bretagnolle A., Mathian H., Swertz E. (2011, sous presse), « La question de l'échelle dans la construction de bases de données urbaines harmonisées », in Pumain (dir), *Echelles et modélisations multi-niveaux*, Editions Chemins de Traverse, Collection Systèmes complexes (**joint au dossier**).

Bura S., Guérin-Pace F., Mathian H., Pumain D., Sanders L. (1996), « Multi-agent systems and the dynamics of a settlement system », Geographical Analysis, 2, 161-178.

Chandler T. et Fox G, (1974), *3000 years of Urban Growth*, New York, London: Academic Press, - Studies in population.

Dubuc S. (1995), *L'urbanisation de l'Inde au XXe siècle*. Mémoire de Master 1 de Géographie, sous la direction de Denise Pumain, Université Paris1 Panthéon Sorbonne.

European Environment Agency, 2002, CORINE land cover update, Technical guidelines, http://www.eea.europa.eu/publications/technical_report_2002_89.

Eurostat (2004), *Urban Audit Reference Guide,* Eurostat Methodologies and Working Paper, 274 p.

Giraut F., Vacchiani-Marcuzzo C. (2009), *Territories and Urbanization, Atlas and geo-historical information system* (DYSTURB), Paris, IRD Editions, Collection CD-Rom et DVD. http://www.cartographie.ird.fr/dysturb.html

Guérois M., Bretagnolle A., Giraud T., Mathian H. (2011, sous presse), « A new database for the cities of Europe? Urban Morphological Zones (CLC2000) confronted to three national databases of urban agglomerations (Denmark, France Sweden) », *Environment & Planning B* (**joint au dossier**)

Julien P. (2003), « L'évolution des périmètres des aires urbaines 1968-1999 », in Pumain D., Mattéi F. (eds.), *Données urbaines*, tome 4. Paris, Anthropos, Economica, pp. 11-20.

Milego R., 2007, *Urban Morphological Zones, Definition and procedural steps,* Report, Copenhagen: European Environment Agency, European Topic Centre Terrestrial Environment, 9 p.

Motte C., Séguy I., Théré C. (2003), *Communes d'hier, communes d'aujourd'hui. Les communes de la France métropolitaine, 1801-2001. Dictionnaire d'histoire administrative*, Paris, Institut National d'Études Démographiques, 2003, 407 p. + 1 CD-Rom.

Moriconi-Ebrard F. (1994), *Géopolis, Pour comparer les villes du monde*, Paris, Anthropos, Economica, Collection Villes, 246 p.

Moriconi-Ebrard François, Hubert Jean-Paul (1999), « Terrae Statisticae : The Database on European Local Territorial Units », in *Sistema Terra*, vol. 8, n°1-3, pp. 120-125.

Paulus Fabien (2004), *Coévolution dans les systèmes de villes : croissance et spécialisation des aires urbaines françaises de 1950 à 2000*. Thèse de Doctorat de Géographie, Université de Paris 1, 407 pages.

Paulus F., Vachiani-Marcuzzo C, 2010, « Innovation and city-size, a cross-country analysis », Ohio State University, Columbus, novembre 2010 (conférence invitée).

Pumain Denise, Riandey Benoît (1986), « Le fichier de l'INED : Urbanisation de la France », in *Espace Populations Sociétés*, vol. 2, pp. 269-278.

Pumain D., Paulus F., Vacchiani-Marcuzzo C. (2009), « Innovation Cycles and Urban Dynamics" in D. Lane, D. Pumain, S. Van der Leeuw, G. West (ed.), *Complexity perspectives on innovation and social change*, chapter 8, ISCOM, Springer, Methodos Series, p. 237-260.

Pumain D., Sanders L., Bretagnolle A., Glisse B., Mathian H. (2009), The future of urban systems : exploratory models, in Lane, Pumain, Van der Leeuw, West (eds), *Complexity perspective in innovation and social change*. Springer, Methodos Series, p. 331-361.

Pumain D. (2010), Une théorie géographique des villes, in *Bulletin de la Société Géographique de Liège*, vol. 55, pages 5-15.

Querci J. (2010), *Le système urbain indien au XX^{ème} siècle*. Mémoire de Master 1 de Géographie, sous la direction de Sébastien Oliveau, Université de Provence (Aix-Marseille I).

Reclus E. (1876-1894), *Nouvelle géographie universelle : la terre et les hommes*, Tome VIII. L'Inde et l'Indo-Chine, Paris, Hachette.

Vacchiani-Marcuzzo C. (2005), « Mondialisation et système de villes : les entreprises étrangères et l'évolution des agglomérations sud-africaines, Thèse de doctorat, Université de Paris 1, 361p.

Van Hamme Adrien (2010), *Conception et développement d'un module générique de cartographie dynamique et interactive*. Mémoire de stage du Master 2 Carthageo, 50 pages. (**joint au dossier**)

Van Hamme A., Bretagnolle A., Mathian H. (2010), « Animer les villes. Exploration dynamique de bases de données urbaines harmonisées dans le temps long », in

Mappemonde, n°99, http://mappemonde.mgm.fr/num27/fig10/fig10301.html (**joint au dossier)**

Vatin Gabriel (2011), *Développement d'algorithmes de construction de grappes d'interaction*. Mémoire de stage du Master 2 Carthagéo, 32 pages. (**joint au dossier**)

4. Perspectives ouvertes par les résultats

L'ANR Harmonie-cités a permis d'ouvrir un véritable champ de réflexion, riche et nourri par la présence de nombreux chercheurs et ingénieurs (une vingtaine impliqués dans le projet) autour de la notion d'harmonisation. Outre les bases de données améliorées ou produites dans le cadre du projet, qui se diffuseront et s'enrichiront, un travail fondamental de conceptualisation et de réflexion théorique a été mené et permettra au projet de rayonner au-delà du seul laboratoire concerné : citons par exemple les algorithmes de construction automatique d'agglomération, les modèles de données permettant d'intégrer une évolution de plusieurs décennies, la construction d'une interface générique pour l'interrogation des bases, l'expertise rassemblant des bases de données urbaines européennes et américaines, un travail de fond sur l'harmonisation des nomenclatures d'activité et sur les méthodes permettant d'exploiter ces bases harmonisées, des méthodes génériques d'appariement de bases de données sur les villes et sur les réseaux de transport etc.

5. Appréciation de l'impact (scientifique, social, culturel, économique…)

Les communications et publications réalisées par les chercheurs et ingénieurs du projet sont multiples (17 références à ce jour) et permettent d'apprécier l'impact scientifique du projet. En particulier, les propositions envoyées dans des colloques avec comité de sélection ont toutes été acceptées et témoignent de l'intérêt fort porté par la communauté des chercheurs autour de ces questions d'harmonisation. L'interface générique Harmonie-cités, permettant d'interroger les bases construites, a été présentée au Concours de géo-visualisation du Festival International de Saint-Dié en 2010 et a reçu le premier prix. Les travaux sur la construction de modèles de données génériques et dynamiques doivent être réutilisés dans trois programmes de recherche internationaux.

C. Publications, diffusion et valorisation

C.1 - Liste des publications et communications

1) Liste détaillée

Articles acceptés dans des revues à comité de lecture

Guérois M., Bretagnolle A., Giraud T., Mathian H. (2012, accepté), « A new database for the cities of Europe? Urban Morphological Zones (CLC2000) confronted to three national databases of urban agglomerations (Denmark, France Sweden) », Environment & Planning B

Van Hamme A., Bretagnolle A., Mathian H. (2010), « Animer les villes. Exploration dynamique de bases de données urbaines harmonisées dans le temps long », in Mappemonde, Palmarès du Concours de Géovisualisation du F.I.G. Saint-Dié, 4 pages, http://mappemonde.mgm.fr/num27/fig10/fig10301.html

Ouvrages ou chapitres d'ouvrages

Bretagnolle A., Mathian H., Swertz E. (2012, accepté), « La question de l'échelle dans la construction de bases de données urbaines harmonisées », in Pumain (dir), Echelles et modélisations multi-niveaux, Editions Chemins de Traverse, Collection Systèmes complexes.

Communications

Bretagnolle A., Giraud T., Guérois M., Mathian H. (2010), "Building European urban areas from Corine Land cover and UMZ database?", The Association of American Geographers Meeting, Washington DC (USA), April 14-18.

Bretagnolle A., Mathian H., Van Hamme A. (2011), "Animating the cities…Dynamic exploration of harmonized urban databases (United-States, France 1800-2000)", Proceedings of the 25th International Cartographic Conference (Paris, 3-8 july 2011), http://halshs.archives-ouvertes.fr/halshs-00624473/fr/

Bretagnolle A., Swertz E. (2011), « La question de l'échelle dans la construction de bases de données urbaines harmonisées », Rencontres interdisciplinaires sur les systèmes complexes naturels et artificiels, Rochebrune, 16-22 janvier 2011.

Guérois M., Bretagnolle A., Giraud T., Mathian H. (2009), « From Urban Morphological Zones (UMZ) to harmonised urban objects in Europe », 16th European Colloquium on Quantitative and Theoretical Geography, Maynooth (Ireland) 4-8 sept. 2009.

Bretagnolle A., Le Nechet F. (2011), "European Urban Patterns and Trends: New Evidences from Urban Morphological Zones data base (CLC1990, 2000, 2006)", The Association of American Geographers Meeting, Seattle (USA), April 12-16.

Paulus F, Vacchiani-Marcuzzo C. 2010, "Innovative activities and city size, 1960-2000 (United States-France), AAG Annual Meeting, Washington

Paulus F., Vachiani-Marcuzzo C, 2010, « Innovation and city-size, a cross-country analysis », Ohio State University, Columbus, novembre 2010 (conférence invitée)

Vacchiani-Marcuzzo C. Paulus F., 2008 "Scaling laws and economic specialization: an experiment United States-France-South Africa", AAG Annual Meeting, Boston.

Publications soumises ou en préparation

Paulus F., Vachiani-Marcuzzo C (en cours de rédaction), "Exploring Functional Diversity in Urban Systems: France, USA, South Africa".

Bretagnolle A., Guérois M., Le Nechet Florent (2011), Exploration multi-échelles d'une base de données harmonisée sur les villes d'Europe (UMZ). Proposition acceptée pour le Colloque « Fonder les Sciences du Territoire », CIST, Paris 23-25 novembre 2011