

**Projet de thèse :**

**Exploration et analyse de données spatio-temporelles :  
Application à la modélisation des transformations des  
systèmes de peuplement sur le temps long.**

**Table des matières**

1 Contexte de la proposition : un projet interdisciplinaire d'identification de transitions dans les systèmes de peuplement.....	2
2 Transitions des systèmes de peuplement : un objet de recherche.....	3
2.1 Abstraire et formaliser le corpus existant.....	3
2.2 Modéliser les transitions.....	3
2.3 Comparer les modèles.....	4
2.4 Vers la conception d'un environnement adapté.....	4
3 État de l'existant.....	5
3.1 L'exploration de données temporelles.....	5
3.2 Outils d'exploration de données spatiales .....	6
3.3 Outils intégrés d'analyses spatio-temporelles.....	6
4 Proposition et programme de recherche.....	7
4.1 Mise en place d'une démarche d'accompagnement des thématiciens dans la modélisation	7
4.2 Développement d'une plate-forme d'exploration de données spatio-temporelles : application à des transitions dans le système de peuplement.....	8
4.3 Application : analyse des transitions du système de peuplement modélisées.....	9
5 Faisabilité et temporalités.....	10
5.1 Étapes clés du projet.....	10
5.2 Contexte de travail et collaborations possibles.....	11
6 Bibliographie.....	12

# **1 Contexte de la proposition : un projet interdisciplinaire d'identification de transitions dans les systèmes de peuplement.**

Ce projet de thèse s'inscrit dans le cadre conceptuel du projet ANR TransMonDyn, et dans la continuation des activités que j'y mène en tant qu'ingénieur d'études depuis Septembre 2012. Il s'agit d'un projet interdisciplinaire, porté par des géographes et des archéologues, visant à modéliser et comparer un ensemble de transitions, sur le temps long, ayant abouti à des transformations du système de peuplement.

Dans ce projet, les thématiciens<sup>1</sup> (géographes, archéologues et linguistes) ont isolé 12 périodes spatio-temporelles dans lesquelles ils ont pu repérer des modifications qui les ont amené à identifier des *transitions* dans le système de peuplement. Ces transitions sont d'une diversité importante, aussi bien en ce qui concerne leurs localisation (de la Gaule méridionale à l'Afrique australe<sup>2</sup>...), leurs échelles d'application (depuis la région du Mesa Verde jusqu'à l'échelle mondiale de la littoralisation<sup>3</sup>), leurs dates de commencement (-70 000 pour la sortie d'Afrique, 1960 pour la métropolisation<sup>4</sup>), ou encore leurs étendues temporelles (300 ans pour la romanisation de la Gaule méridionale, 10 000 ans pour l'émergence des systèmes de villes<sup>5</sup>).

Cette diversité sur tous les plans amène la question de la comparaison de ces transitions, et avec elle, de la comparabilité des faits observés. Afin de pouvoir mener une étude comparative, il a été choisi par des géographes et archéologues<sup>6</sup>, de modéliser ces transitions en identifiant des régularités, « faits stylisés »<sup>7</sup>, qui permettraient l'analyse systématique des influences et répercussions de phénomènes récurrents et génériques sur la trame des structures spatiales constituées par l'Homme.

En tant qu'ingénieur d'étude j'ai eu à travailler sur plusieurs des axes thématiques de TransMonDyn. En premier lieu, dans le domaine de la modélisation orientée-agents, j'ai effectué une veille technologique et un recensement des plate-formes de simulation. En prenant part à un groupe de travail, rassemblant thématiciens et modélisateurs, visant à modéliser l'une des transitions identifiées, j'ai aussi pu m'immerger dans les logiques et contraintes et l'interdisciplinarité, en particulier dans la communication entre archéologues et géographes.

De manière plus transverse, j'ai participé à la création d'une grille d'analyse systémique des transitions du projet TransMonDyn, premier pas vers l'identification de processus communs et de similarités et sous-groupes dans les transitions.

Sur une base plus régulière de travail, j'avais pour objectif de concevoir une plate-forme d'exploration des données résultant des simulations à venir. Cet outil, dont un premier prototype est en cours de réalisation, doit permettre aux thématiciens de mener différentes analyses sur les données dont ils disposent, en y associant des modes de visualisation interactifs, afin de leur permettre de discerner des structures spatiales et temporelles dans leur corpus de données.

Avec la participation à l'organisation des séminaires, j'ai ainsi pu acquérir une vision globale de l'intérêt et même de la nécessité de l'interdisciplinarité pour répondre à des problématiques plus globales que l'étude d'un unique système spatialisé.

---

1 Membres du projet qui apportent les connaissances et expertises thématiques, et sont donc spécialistes des périodes et étendues spatiales étudiées. Dans le projet, on les distingue des modélisateurs, qui n'ont pas pour rôle d'apporter une connaissance directe sur les phénomènes étudiés, mais sur la manière de les envisager dans une approche systémique.

2 Transitions : « Concentration de l'habitat de l'Âge du Fer » (P. Garmy, J-L. Fiches, L. Nuninger) et « Antiquité tardive » (F. Favory) pour la Gaule, et « Urbanisation de l'Afrique du Sud », C. Vacchiani- Marcuzzo

3 « Village formation in Pueblo societies », T. Kohler, « Littoralisation des systèmes de peuplement », C. Ducruet

4 « Out of Africa », J-M. Hombert et C. Coupé, « Émergence de métropoles polycentriques "Mega-City Regions" », F. Le Néchet

5 « Romanisation », M-J. Ouriachi et F. Bertoncello, et « Émergence des villes », D. Pumain

6 Dans la continuité des projets Archaeomedes I et II qui les avaient fait collaborer (Archaeomedes, 1998)

7 Un "fait stylisé" est une présentation simplifiée (ie. taux, ratio ou écart) d'une régularité empirique sur l'observation de laquelle il y a un large accord. Le terme a été popularisé en économie par Kaldor (1961). (D. Phan, à paraître)

## 2 Transitions des systèmes de peuplement : un objet de recherche

Cette expérience m'a convaincu du besoin de composer un « méta-modèle » de l'évolution des systèmes de peuplement. L'approche d'une telle modélisation est progressive et itérative, basée sur une succession de modèles de niveaux conceptuels différents, et débute par la mise au point d'un modèle descriptif commun et adapté à tous les systèmes étudiés dans le projet, ce qui permet dès lors une formalisation systémique des transitions. Ce méta-modèle allierait modélisation géographique (de la conception à l'implémentation) et analyse formelle (via la définition d'un cadre théorique et applicatif) des dynamiques évolutives d'une structure spatiale.

Le projet de recherche doit permettre de mener une analyse originale des dynamiques et mécanismes qui interagissent et provoquent une transition du système de peuplement étudié. Dans ce projet de thèse, je m'appuierai sur un *corpus*, composé des transitions de TransMonDyn, qui nourrira la démarche méthodologique en faisant office de cas d'application et de source thématique.

### 2.1 Abstraire et formaliser le corpus existant

La première des activités, consiste à analyser et à synthétiser le corpus des connaissances existantes (Observation et Réflexions/Analogies dans la Figure 1), amenées par les thématiciens, sur chacune des transitions. Ce corpus, constitué des cas d'étude des thématiciens, de leurs interprétations et connaissances, doit être collecté et formalisé pour pouvoir être envisagé comme une source de modélisation systémique (voir par exemple Costa, 2012). Cette tâche, qui revient à un travail d'extraction de connaissances spécialisées, a été facilitée par la mise en place d'un schéma commun de description des transitions, amené par J-L. Fiches<sup>8</sup> et par la volonté partagée des « porteurs » de transition de s'approprier ce mode de synthèse. Dans ce schéma descriptif, les *transitions* surviennent entre deux *régimes*, qui constituent des *états* de systèmes dynamiques (Figure2) caractérisé par une structure propre : ce sont des « clichés » instantanés d'un système.

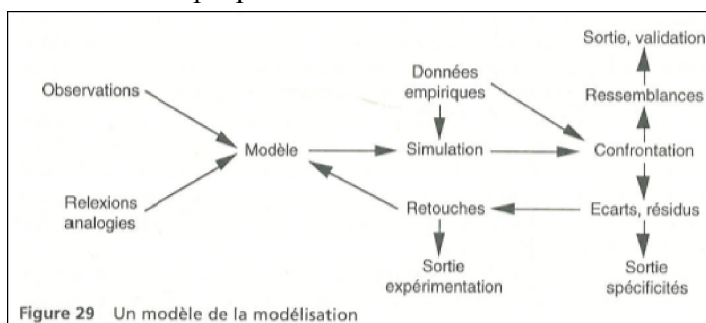


Figure 29 Un modèle de la modélisation

Figure 1 : Un modèle de modèle en géographie - F. Durand Dastès, « Les modèles en géographie », in Encyclopédie de Géographie (dir. Bailly, Ferras, Pumain).

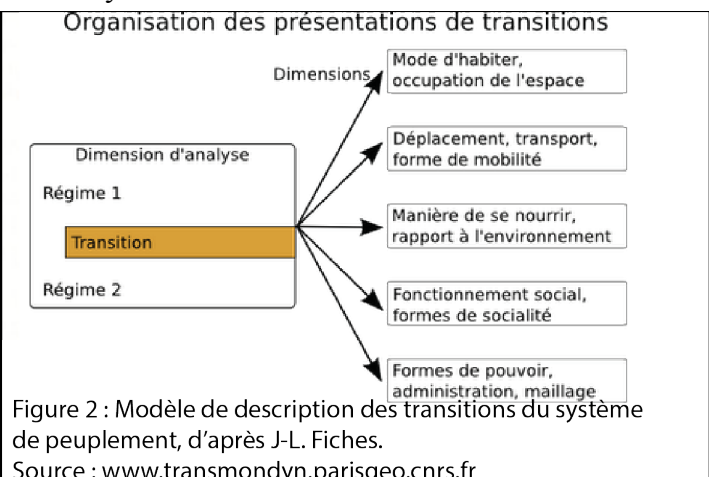


Figure 2 : Modèle de description des transitions du système de peuplement, d'après J-L. Fiches.  
Source : [www.transmondyn.parisgeo.cnrs.fr](http://www.transmondyn.parisgeo.cnrs.fr)

Au delà des connaissances thématiques de chaque spécialiste d'un domaine, on dispose aussi de sources de données qui permettent d'appuyer chacune des interprétations effectuées pour les transitions. Ces données (Données empiriques sur la Fig. 1), qui sont très inégales, tant dans leur quantité que leur fiabilité, ou encore dans les formats à disposition, devraient pouvoir être analysées en y appliquant des méthodes et traitements communs, afin de préparer le travail de comparaison des transitions.

La mise en place d'une structure de données commune à l'ensemble des transitions constituerait donc un préalable nécessaire à une réflexion sur la structure des données que devront générer les modèles (Sorties d'expérimentation dans la Fig. 1), et qui devront pouvoir être absolument comparables entre les transitions afin de permettre l'identification de processus et tendances similaires.

### 2.2 Modéliser les transitions

Afin d'aboutir à la création du méta-modèle poursuivi, il faudra ensuite modéliser ces transitions en s'inscrivant dans un cadre de modélisation commun. Le niveau d'abstraction de la description des transitions étudiées est de nature très disparate, conséquence de la forte interdisciplinarité. La

démarche d'ensemble est cependant identique dans tous les cas : les thématiciens procèdent en premier lieu à une modélisation ontologique des *régimes* encadrant la *transition*, en repérant et distinguant les entités, ainsi que leurs attributs et leurs relations, qui participent d'un système. Puis, en analysant les changements entre ces régimes, on exclut les propriétés des régimes qui ne participent pas (directement ou indirectement) aux transformations du système de peuplement, et on isole les processus d'interaction et de rétroaction.

L'étape suivante, comme indiqué dans l'introduction, consistera à modéliser de manière systémique, par exemple à travers l'utilisation de diagrammes sagittaux, ces modèles conceptuels. La modélisation systémique peut prendre plusieurs formes, et on devra ici viser à avoir un modèle centré sur les interactions entre entités sociales et spatiales, préalable à la conception des modèles de simulation. Celle-ci se fera en suivant les concepts de la modélisation orientée-agents, qui permettrait d'insérer des faits stylisés, représentant les hypothèses des transformations.

Pour que le contenu thématique généré par les modèles soit ré-utilisable, il convient de mener des étapes de « validation », tout au long du développement des modèles. Ces phases de vérifications se divisent en de multiples étapes de « validation interne », permettant de s'assurer que les dynamiques introduites dans un modèle sont bien celles que décrit et observe le thématicien.

La validation passe aussi par un retour cyclique aux données et expertises, au travers d'une démarche constante d'allers-retours aux données empiriques, formalisé dans le concept de « validation externe » (Amblard et al, 2006). On devra donc confronter les résultats obtenus aux connaissances des thématiciens, mais aussi, sur le plan micro, comparer les données simulées et les données empiriques du corpus.

L'enjeu de cette étape consiste aussi à la mise en place d'une démarche facilitant le dialogue et la compréhension interdisciplinaire, ainsi que l'harmonisation de ces modèles.

### **2.3 Comparer les modèles**

Une fois les modèles constitués et validés par les thématiciens, il sera alors possible de travailler sur la comparaison des modèles, lesquels peuvent ainsi être vus comme des proxys permettant de mener la comparaison des systèmes modélisés, trop complexes pour pouvoir être eux-même comparés dans leur intégralité.

Cette comparaison se basera notamment sur la coexistence de faits stylisés similaires au sein de différentes transitions, qui devraient permettre d'identifier de manière formelle les dynamiques qui mènent à des modifications spatiales variées. Si chaque transition présente une majorité de caractéristiques uniques, les thématiciens sont néanmoins convaincus de la généricité de certains phénomènes, qui, selon le contexte, peuvent provoquer des effets inverses. Il conviendra donc de regrouper les transitions dans des méta-modèles d'évolution spatiale, ce qui permettra alors des comparaisons plus poussées.

Celles-ci, par exemple, en comparant terme à terme les paramétrages de faits stylisés similaires, pourraient faire émerger des règles génériques de transformation des systèmes de peuplement face à un stimulus, endogène ou exogène, et la comparaison porterait alors sur les spécificités de chacune des transitions face à ces stimuli commun, permettant dès lors d'approfondir la connaissance des systèmes modélisés ainsi que leurs résiliences et modes de réaction.

### **2.4 Vers la conception d'un environnement adapté**

En sus du besoin de modélisation, il apparaît donc nécessaire de disposer d'un environnement permettant la mise à disposition de méthodes d'analyses des données empiriques, ainsi que la comparaison de celles-ci avec les données résultant de la simulation. Ces méthodes devront être accessibles à des non spécialistes en informatique, et de fait, leur environnement d'exécution devra être interactif et d'une prise en main aisée.

L'ensemble des données comportant une dimension spatiale, les méthodes de base doivent être tournées autour de ce type de données, où l'on ne peut se contenter de considérer les informations spatiales comme de simples variables quantitatives de géolocalisation.

Les structures étudiées dans chacune des transitions revêtant un caractère profondément complexe, elles seront modélisées avec une approche agent, mais cette même approche impose des méthodes

d'analyses particulières, basées sur les agrégats, les analyses de sensibilité, et d'une manière générale, l'étude du poids de la stochasticité dans les phénomènes.

L'objet d'étude résulte d'interactions spatiales qui sont de plus présentes à des échelles emboîtées (les interactions entre agents sociaux de différents niveaux amenant nécessairement de multiples échelles spatiales). Les modes d'analyse doivent de fait être adaptés et spécifiques à cette approche multi-scalaire, et permettre d'en dissocier les échelles et les boucles de rétro-action qui opèrent entre elles.

Tout le projet ayant pour objectif d'étudier des changements spatiaux sur le temps long, la dimension temps doit aussi revêtir une importance particulière, et ne peut être traitée comme une dimension générique ne servant qu'à ordonner des événements. Le corpus pris comme exemple d'application présentant un large éventail d'étendues temporelles, l'environnement d'exploration des données devra qui plus est être adaptés aux contraintes de cette gestion non standardisée du temps.

### **3 État de l'existant**

#### **3.1 L'exploration de données temporelles**

Le cadre d'analyse le plus fréquent, en géographie, de données spatio-temporelles, est théorisé et appliqué au sein de la « Time-Geography » (Hägerstrand, 1970, via Chardonnel, 2001). Les études qui s'inscrivent dans ce paradigme visent bien à l'analyse de données et de phénomènes spatio-temporels, et pourraient constituer une norme d'étude adaptée aux problématiques de notre corpus si celui-ci n'était pas aussi varié.

La Time-Geography vise en effet majoritairement à suivre des mouvements, des déplacements<sup>9</sup>, ce qui constitue une large source d'inspiration, toutefois insuffisante dans le cas des problématiques de restructuration du territoire que l'on rencontre dans le projet. Au niveau même des données disponibles pour les transitions présentant des déplacements, on peut certes interpoler ces mouvements, mais les sources à disposition sont discrètes dans le temps et dans l'espace, ne présentant pas de moyens de suivis fiables. Qui plus est, le corpus d'étude est multi-scalaire et son échelle micro ne concerne que rarement l'individu.

Si le mode d'appréhension du temps est bien exploitable, les techniques mobilisées pour l'étude des transitions ne peuvent donc s'appuyer entièrement sur le cadre théorique de la Time-Geography, et encore moins sur les modes de représentation et d'exploration qui lui sont associés, qui s'ils font l'objet de recherches, ne présentent pas encore de mode de visualisation simple (Kwan, 2004).

On peut discerner deux principales manières de représenter et d'analyser le temps dans les outils et méthodes existant<sup>10</sup>: soit en présentant des instantanés successifs correspondant aux différents pas de temps, soit en présentant de manière intégrée les variations entre ces pas de temps.

La première méthode est le plus souvent représentée par le biais d'animations, qui peuvent, dans les cas les plus avancés, être interactives. La dimension temps y est donc présentée comme une variable discrète, que l'on fait évoluer sans se préoccuper de l'état antérieur, et qui ne permet la comparaison entre deux états que par une analyse personnelle visuelle des différences.

La seconde méthode est surtout utilisée pour représenter les évolutions entre deux pas de temps successifs, par la représentation statique d'une évolution (par exemple un taux de variation), donc résultat d'une analyse finie et non extensible. Si l'approche est comparative, elle est difficilement applicable à des périodes non consécutives, et nécessite surtout que la morphologie des structures spatiales n'évolue pas entre les périodes analysées.

Pour ce dernier cas, des méthodes existent, à l'intérieur du champ de la géographie autant que dans d'autres disciplines (Andrienko, 2006), souvent issues du domaine du « data-mining ». Les méthodes de suivi temporel, qui œuvrent à la reconnaissance d'ensembles (dans de gros volumes de données) présentant les mêmes caractéristiques au cours du temps, mais dont les formes varient, sont en effet très utilisées, par exemple dans le champs de l'analyse des mobilités individuelles, où l'on peut chercher à agréger des individus sur la seule base de la forme de leurs déplacements<sup>11</sup>.

9 «la time-geography met, cependant, davantage l'accent sur l'emprise spatiale des trajectoires [individuelles]», Chardonnel, 2001

10 ESPON Cartographic Language, Interim Report, à sortir Mai 2013

11 Spielman, S.E. Bit City: Data, Technology and Understanding the City, Association of American Geographers Annual Meeting 2012

Les dimensions temporelles sont bien prises en compte dans ces démarches, mais le spatial n'est vu que dans son aspect de localisation. Ne serait-ce que pour le cas le plus simple, à savoir des informations spatiales ponctuelles, on n'y tient aucunement compte des notions de contexte, de cohérence, et dans l'ensemble de structure spatiale.

### **3.2 Outils d'exploration de données spatiales**

Dans le domaine de l'exploration de données spatiales, avec l'informatisation et la multiplication des sources de données disponibles, de nombreux outils sont apparus depuis les années 1980. Le plus souvent, ces outils visent à mettre en avant une méthode originale, qu'ils viennent donc mettre à disposition d'un public de géographes avertis. Ils sont donc support d'un traitement particulier, ou, pour les plus avancés, essaient de concentrer dans un outil unique l'ensemble des avancées d'un laboratoire ou d'un groupement de chercheurs.

Les Systèmes d'Information Géographiques ne peuvent être ici vus comme des outils d'exploration à part entière, en ce qu'ils sont fondamentalement pensés comme des outils de description et d'analyse d'un territoire donné, visant à caractériser un ensemble spatial par le croisement de couches d'informations, plus que par l'analyse de ces couches elles-mêmes. Ces outils sont plus dans la synthèse et la communication que dans l'exploration elle-même, quand bien même ils peuvent intégrer des fonctions d'analyse spatiale avancées.

On prêterait un intérêt particulier aux outils de géovisualisation (Kraak, 1998), en ce qu'ils permettent une exploration visuelle des données, en offrant des modes de représentation innovants pilotés par une interactivité qui constitue le cœur de conception de ces outils (Theus, 2005).

Bien que très performants et ergonomiques pour de l'exploration spatiale, ces outils ne prêtent toutefois que rarement une place aux données temporelles. Quand c'est le cas, le temps n'est vu que comme une succession de valeurs instantanées, et n'est donc manipulable que via des « timelines » (VIS-STAMP, Guo et al, 2006) et autres dispositifs d'animation de cartes. Certaines méthodes ont été créées qui permettent une interaction plus forte avec cette dimension, mais ils se cantonnent à l'étude de séries temporelles, basées sur des mesures régulières et exhaustives. Et à nouveau, la tâche de comparaison est laissée à l'utilisateur, qui doit noter les éventuelles différences via la visualisation, sans disposer d'indicateur réellement pensé pour lui faire remarquer les changements spatiaux.

En dehors de ces aspects qui pourraient se voir résolus par l'ajout de méthodes de comparaison spatio-temporelles dans ces outils existants, on peut noter le manque d'outils de manipulation des données. Les logiciels disponibles permettent ainsi la représentation de différents indicateurs (GeoDA, Anselin 2006, et GeoVista Studio, Takatsuka 2002), et des sélections communes à chacun, qui permettent d'observer les données sur plusieurs plans à la fois. Mais en dehors de cette sélection partagée, il n'y a pas de communication possible entre les indicateurs, entre les différents traitements possibles sur les données.

La modélisation nécessite des étapes exploratoires, et ne peut être validée qu'au moyen d'allers-retours sur le modèle et les données produites, ce qui implique de pouvoir visualiser les résultats et de pouvoir, face à la masse de données qui peut être générée, les synthétiser au moyen de chaînes de traitement de l'information. Comme cela se fait en statistique, il faut donc un environnement permettant d'enchaîner des opérations simples, constituant ainsi une chaîne de traitement complexe, dont chaque membre peut être modifié à la volée, ce qui permettrait alors d'explorer les interactions entre indicateurs.

Si l'on peut utiliser ces plate-formes d'exploration de données statistiques, et les adapter en partie aux besoins de l'analyse spatiale, il réside un besoin fondamental d'accéder à des plate-formes de même ambition, mais dédiées aux statistiques spatiales. Les opérations d'agrégations spatiales, de classifications sont ainsi parfois possibles, mais elles sont alors vues comme une finalité, alors qu'elles devraient aussi pouvoir être des étapes, des pré-traitements préalables à l'exploration d'autres dimensions des données.

### **3.3 Outils intégrés d'analyses spatio-temporelles**

On peut cependant noter l'existence de nombreux outils alliant une exploration interactive de données spatiales à un travail sur la prise en compte du temps et de l'évolution. Ces outils ne sont pour autant pas génériques, faisant le plus souvent office de présentoirs à données. Ce sont des outils de

valorisation de données connues, qui permettent à un public souvent large d'appréhender les notions entremêlées de temps et d'espace.

Ces outils ne permettent pas l'import de données externes, et surtout, ne font qu'appliquer une unique méthodologie d'exploration des données embarquées. Il s'agit de donner à voir un phénomène, par le biais d'un type d'indicateurs, sur un espace donné.

L'origine de ces outils explique cette spécialisation : ils sont majoritairement produits par des institutions qui cherchent à communiquer sur un jeu de données en particulier (par exemple, Bretagnolle et al, 2011, ou Pison et al, 2011). Ce sont donc des applications métiers, pensées par et pour un jeu de données spécifique, qui mettent en avant une méthodologie de représentation permettant une bonne analyse visuelle et interactive de ce jeu.

Malgré leur spécialisation, ces outils, par la nécessité qu'ils ont d'innover pour séduire, permettent toutefois la mise en avant de nouveaux modes de représentation et de manipulation de l'information spatio-temporelle.

Les utilisation de la 3D par exemple, qui est souvent le parent pauvre de la visualisation se contentant d'apporter une information redondante au fond de couleur utilisé, peut aussi être vecteur d'une information supplémentaire (L'Hostis, 2003), pour peu que l'on puisse « naviguer » dans l'environnement et qu'on ne se contente donc pas d'une pseudo perspective.

Les outils d'exploration, pour des structures de données fixes et originales, peuvent eux aussi présenter des innovations, comme dans le cadre de séries temporelles, où l'on peut ainsi interroger les données temporelles non seulement dans leur succession, mais aussi dans leurs répétitions et cycles (Timewheel, Edsall & Peuquet, 1997). On peut par exemple observer les valeurs d'indicateurs, selon des intervalles de temps définis, mais aussi s'intéresser aux données de certains horaires sur des temporalités bien plus longues (Time Coil, Edsall & Peuquet 1997).

Afin de répondre aux problématiques de ce projet de recherche, il faut donc un environnement adapté, tirant avantage des innovations dans la représentation apportées par la géovisualisation et l'exploration interactive, mais aussi suffisamment générique pour pouvoir travailler sur les données très hétérogènes qui ont été mentionnées auparavant.

Le besoin de reproductibilité et de traçabilité des méthodes employées rend nécessaire la création d'un environnement dédié, dont le cadre théorique devra s'appuyer sur des règles et méthodes fortement formalisées. On devrait donc aboutir à une grammaire, qui fait actuellement l'objet de recherche, permettant l'expression de modèles conceptuels d'intégration du temporel et du spatial.

## 4 Proposition et programme de recherche

### 4.1 Mise en place d'une démarche d'accompagnement des thématiciens dans la modélisation

La modélisation d'un phénomène en sciences sociales implique de concilier des connaissances propres à la discipline et des savoirs-faire de modélisation. Dans le projet TransMonDyn auquel j'ai participé, et dont certaines productions constituent le corpus sur lequel je m'appuierai, le contenu thématique est apporté par des géographes, archéologues et linguistes, et les autres membres du projet (mathématiciens, informaticiens et philosophes) ont pour rôle d'aider à l'extraction des connaissances des thématiciens dans un but de modélisation. Cette modélisation, qui passe par des étapes de formalisation systémique des transitions étudiées, peut revêtir plusieurs formes, d'un modèle conceptuel, basé sur une ontologie, à une implémentation informatique sous forme de système multi-agents.

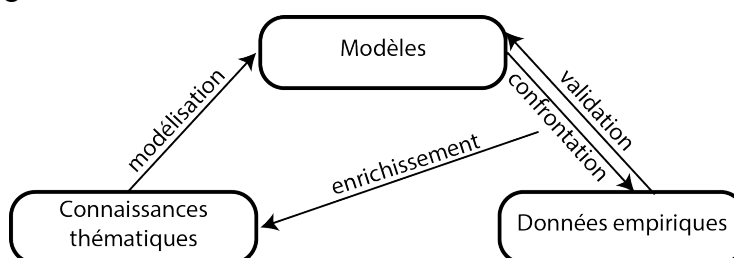


Figure 3 : Articulation de la démarche d'analyse et d'apprentissage par la modélisation.

Dans un tel contexte interdisciplinaire, une démarche cyclique, privilégiant la réinjection des savoirs acquis par la thématique dans les modèles, puis par les modèles dans la thématique, s'avère judicieuse (voir Figure 3). La réalisation des objectifs du projet de thèse devra dès lors reposer sur l'accomplissement des

étapes suivantes, présentées dans leur succession chronologique principale.

En premier lieu, et c'est sans doute l'étape qui demandera le plus de travail en interdisciplinarité, il conviendra d'extraire les connaissances des thématiciens, et de les « décrypter » afin d'en tirer des processus de causalité, de dépendance et d'inter-relations compréhensibles par le modélisateur géomaticien que je suis.

Cette phase, qui n'est possible que par un dialogue approfondi et par une compréhension mutuelle des enjeux, permettra l'isolement de « faits stylisés », c'est-à-dire des règles d'interactions entre les entités d'un modèle. Chaque thématicien ayant son propre niveau de langage, d'abstraction et de vulgarisation vis-à-vis de son objet d'étude, il est nécessaire, à chaque étape de la description du système et de l'identification des lois qui régissent son évolution, que les faits stylisés soient suffisamment précis et explicites pour que le thématicien en cerne les implications directes.

Ces règles peuvent être formalisées conjointement par le modélisateur et le thématicien à l'aide de plusieurs méthodes de modélisation, qu'elles soient systémiques, sagittales, ou plus proches de l'implémentation, comme l'UML.

Une fois les logiques de fonctionnement du modèle décidées, il faudra en créer les implémentations informatiques en tant que systèmes multi-agents. Selon les niveaux d'abstraction employés par les thématiciens, les modalités d'implémentation peuvent changer, par exemple selon que l'on souhaite que le modèle anime des faits stylisés très génériques, auquel cas on se tournera vers des plate-formes de développement de modèles généralistes comme NetLogo (Tisue & Wilenski, 2004), que le modèle fasse appel à des données géométriques, ce qui pousserait à l'utilisation de GAMA (Amouroux et al, 2009), ou encore, si le modèle présente de forts besoins de calcul, auquel cas on se tournera vers une implémentation désolidarisée d'une plate-forme, directement dans un langage de programmation (Reuillon et al, 2013).

Après la mise en place du cœur des modèles, il sera nécessaire, par des allers-retours avec les thématiciens, de calibrer ceux-ci, afin qu'ils collent autant que possible aux représentations que les thématiciens ont des résultats attendus. Cette étape de validation externe reposera, quand les données seront présentes, sur des comparaisons entre données simulées et données empiriques.

En l'absence de ces dernières, ou si leur précision est trop faible, on pourra toutefois se baser sur l'analyse des dynamiques émergeant du modèle, comparer ces dynamiques celles que les thématiciens ont observées et/ou supposées.

Quand plusieurs modèles auront été développés, on pourra passer à une phase de comparaison des modèles implémentés. Cette comparaison pourra s'opérer sur plusieurs plans de manière simultanée, c'est-à-dire qu'elle portera autant sur la comparaison systémique des mécanismes implémentés, par leur niveau de généralité et de similarité, que, lors de l'identification de faits stylisés récurrents, sur la comparaison des différentes modalités d'implémentation.

Cette étape devrait permettre de trouver des similarités et points de divisions dans les phénomènes conduisant à des transformations des systèmes de peuplement. Ce sera aussi l'occasion de situer ces phénomènes dans leurs contextes de systèmes complexes, et ainsi d'analyser leurs différents types de « réponses » dans des environnements différents.

Afin de simplifier une démarche de comparaison systématique des modèles de transitions, cette étape de comparaison des modèles passera aussi par la définition et l'application d'un standard dans les sorties numériques des modèles, standard qui devra autant que possible rester générique afin de pouvoir formaliser les données empiriques du corpus d'étude.

## **4.2 Développement d'une plate-forme d'exploration de données spatio-temporelles : application à des transitions dans le système de peuplement.**

Parallèlement à la conception et au développement des modèles, et afin de permettre la validation interne et externe de ceux-ci, l'objectif sera de créer une plate-forme exploratoire et ouverte aux données des thématiciens. Celle-ci doit permettre à des utilisateurs parfois novices en analyse de données spatiales de synthétiser leurs données en y faisant émerger des structures spatiales et temporelles. Pour que cette plate-forme puisse véritablement s'inscrire dans une démarche exploratoire, elle doit s'inscrire dans le cadre conceptuel de l'environnement d'exploration de données



spatio-temporelles dont le besoin a été identifié dans ce projet de thèse. On développera ici les principales fonctions que cette plate-forme doit être en mesure d'offrir.

S'agissant d'un outil d'exploration dédié à un large public de scientifiques, de degré de spécialisation très divers, la prise en main doit être très aisée. Pour cela, et pour faciliter l'identification d'entités remarquables, on utilisera le principe de la multi-sélection interactive dans un environnement multi-fenêtres : l'utilisateur pourra afficher plusieurs cartes et graphiques (dans des vues différentes), chacun fournissant une représentation d'un ou de plusieurs indicateurs différents, et sur chaque vue, les sélections effectuées seront reportées sur les autres. Ce mode de forte interactivité permet de réduire une sélection d'un niveau pour chaque indicateur, et donc d'observer le comportement d'entités sur plusieurs plans différents, et notamment dans leur localisation. La plate-forme devant prendre en compte la temporalité, il faudra définir comment reporter ces sélections au cours du temps.

Cette multi-sélection se fera, on l'a dit, sur des fenêtres multiples étant capables de représenter l'ensemble des variables présentes dans les données, mais aussi sur des indicateurs construits au sein de la plate-forme. En mettant à disposition de l'utilisateur un environnement permettant de créer une chaîne de traitement, on lui permet d'accéder à des statistiques de plus haut niveau, et ainsi, par un jeu de composition/transformation, d'identifier des structures plus complexes que celles qui seraient décelables sur les données brutes. On peut s'inspirer, pour la manipulation de ces chaînes de traitement, de ce qui se fait en statistiques exploratoires (par exemple SPAD (Lebart et al, 1988) ou SAS WorkFlow Studio<sup>12</sup>), ou pour l'enchaînement des traitements en SIG (ArcGIS Model Builder<sup>13</sup>).

La logique de construction par le dessin d'une chaîne de traitement rend accessible à chacun des modes d'analyse complexes, et avec le mode de visualisation présentée dans les paragraphes précédents, offre la possibilité de procéder par essais et ajustements successifs afin d'obtenir des indicateurs structurels pertinents et adaptés spécifiquement aux données. Cette liberté laissée à l'utilisateur permet de plus de gérer la diversité des cas d'application, sans forcer l'utilisateur thématique à utiliser un ensemble d'indicateurs préconçus et monolithiques.

L'un des enjeux de cette plate-forme sera la gestion des sources de données variées et composites que l'on a mentionnées. Les formats de stockage de l'information spatiale sont ainsi très divers et hétérogènes, et ce constat est bien plus développé quand on ajoute la dimension temporelle à ces données. Il faudra donc, en amont de la création de la plate-forme, définir un ensemble de formats de données spatio-temporelles, ensemble qui devra constituer le mode d'intégration de données privilégié dans la plate-forme exploratoire, mais devra aussi être suivi dans les modèles générés, afin que les données empiriques et simulées soient comparables par le biais de l'outil.

Pour aller plus loin dans la comparaison empirique/simulé, on devra concevoir un couplage plus direct entre les modèles et la plate-forme d'exploration, celle-ci acquérant alors un rôle d'outil de visualisation et de contrôle des simulations en temps réel, ce qui permettrait de diriger et corriger les simulations au fur et à mesure de leurs exécutions par les thématiques. La conception de ce couplage soulève des problématiques complexes dans la formalisation du cheminement récursif nécessaire entre le modèle, les expérimentations qu'il implique et l'exploration des sorties de celles-ci. Concevoir une connexion directe entre ces composants permettrait de guider en temps réel les simulations, et ainsi d'améliorer la représentation qu'ils constituent de nos cas d'application.

#### **4.3 Application : analyse des transitions du système de peuplement modélisées.**

La modélisation des transitions et la création de la plate-forme d'exploration, s'ils constituent des investissements temporels conséquents, ne sont ici pas une finalité, mais visent à permettre l'analyse, par la comparaison, des transformations dans les systèmes de peuplement de manière plus générique et transverse que les cas d'applications du corpus.

Pour cela, on vise en premier lieu à trouver des faits stylisés répétés ou récurrents dans l'ensemble des zones du monde. Si l'on peut aisément identifier des réactions à certaines typologies de phénomènes, en particulier quand ceux-ci sont exogènes au système, il convient ici d'aller plus loin,

12 <http://support.sas.com/documentation/cdl/en/wfsug/64870/HTML/default/viewer.htm#titlepage.htm>

13 <http://help.arcgis.com/fr/arcgisdesktop/10.0/help/index.html#//002w0000001000000>

de manière systématique, en cherchant à trouver l'ensemble des éventails de possibilité qui mènent, depuis l'état d'un système de peuplement connu, à un autre système. A partir de là, il sera possible d'isoler les processus qui pèsent sur ces transformations, et de les catégoriser selon les rôles qu'ils jouent en fonction du contexte des autres processus en cours.

La diversité des transitions sur lesquels ce projet prend appui devrait permettre de plus d'acquérir des connaissances supplémentaires sur la manière dont ces dynamiques interagissent, et donc sur leurs modalités d'inter-causalités. A titre d'exemple, les sociétés Pueblo étudiées par Timothy Kohler (Kohler et Varien, 2010) ont subi une invasion extérieure qui a profondément modifié leur organisation sociale et conséquemment spatiale, débouchant sur une migration progressive de l'habitat depuis les vallées aux hauteurs des canyons et réduisant la surface des terres cultivées, tandis qu'une autre invasion, romaine cette fois-ci, dans la Gaule du Ier siècle, provoque elle une densification du peuplement et une augmentation et rationalisation des surfaces cultivées (Favory et al, 2011). On peut donc s'interroger sur les raisons contextuelles qui amènent deux transformations opposées à partir d'événements globalement de même nature.

Une autre finalité, plus hypothétique, serait de parvenir à reconstruire de l'information à partir de sources trop incertaines ou lacunaires. Les migrations Bantu (3000 – 2000 B.P.) constituent un exemple de ces transitions dans le système de peuplement reconstituées a posteriori grâce à des données linguistiques et phylogénétiques, et pour lesquelles le manque d'éléments archéologiques complique la compréhension (Coupé et Hombert, 2012). A l'aide d'un modèle établi sur les faits connus, il serait possible d'interpoler des étapes dans la migration, ce qui pourrait rendre possible une prospection plus ciblée de sources archéologiques.

Dans le même esprit, la modélisation peut servir d'outil de projection, objectif assumé de transitions comme celle qui vise à modéliser l'émergence de métropoles polycentriques (Hall et Pain, 2006, d'après F. Le Néchet<sup>14</sup>). Pour cette transition à l'échelle mondiale, la raréfaction des énergies fossiles engendrera une hausse du coût de l'énergie, ce qui, dans un contexte spatial d'habitat étalé, provoquerait l'apparition de ces « Mega-city regions ». La temporalité ici observée s'étend de 1960 au milieu du XXIème siècle, comportant donc une importante période de prospective, et une exploration plus poussée des données existantes ainsi que des résultats de simulation permettrait de poser des hypothèses plus appuyées sur la morphologie des systèmes de villes dont on prévoit l'apparition. On aurait donc ici aussi une construction d'information, non pour reconstituer une source de données trop rare, mais permettant de mener des projections sur l'avenir des systèmes de peuplement.

Ce projet de thèse est animé par l'ambition de parvenir à un méta-modèle de l'évolution des structures spatiales. Celui-ci profiterait et s'appuierait sur la diversité des exemples de transition disponibles pour concevoir des hypothèses générales sur l'évolution des systèmes de peuplement sur le temps long. De la même manière que l'étude des systèmes de villes (Pumain, 1997), cadre conceptuel, peut permettre des projections par exemple sur l'évolution de réseaux de recherches européens (Comin, 2010), une théorie englobante, pourrait amener à spécialisation afin d'étendre, depuis le corpus d'application, cette même analyse rétro- et prospective à d'autres systèmes de peuplement. Pour aller plus loin, on pourra sortir du cadre d'application des transitions, et tester l'environnement d'exploration et de modélisation sur d'autres objets spatio-temporels, par exemple sur les trajectoires de systèmes de villes<sup>15</sup>, ou encore sur l'étude temporelle de la constitution des réseaux de transports<sup>16</sup>, qui font l'objet de recherches au sein de l'UMR Géographie-cités.

## **5 Faisabilité et temporalités**

### **5.1 Étapes clés du projet**

Le projet de thèse débiterait en Octobre 2013, pour une durée de 3 ans, et s'achèverait donc à l'automne 2016. On présente ici les différentes phases prévisionnelles qui devraient jalonner le déroulement de la thèse, phases qui portent autant sur la dimension conceptuelle que thématique dans son application aux cas d'études.

<sup>14</sup> [www.transmondyn.parisgeo.cnrs.fr/transitions-etudiees/t12](http://www.transmondyn.parisgeo.cnrs.fr/transitions-etudiees/t12)

<sup>15</sup> Objet du programme ERC GeoDiverCity, porté par Denise Pumain, qui sert de cadre analytique à plusieurs thèses en cours, menées par Clara Schmitt, Sébastien Rey-Coyrehourcq, Clémentine Cottineau et Antonio Ignazzi.

<sup>16</sup> Anne Bretagnolle sur la France et les États-Unis, Céline Vacchiani-Marcuzzo et Solène Baffi pour l'Afrique du Sud.

## **Phase 1 – Conception d'un environnement d'exploration de données spatio-temporelles**

Cette étape, basée sur l'analyse exhaustive de la bibliographie, devra définir les modalités d'une ré-organisation du corpus existant. Elle sera aussi l'occasion de définir l'architecture conceptuelle de la plate-forme d'exploration à laquelle on souhaite aboutir.

### ***Phase 2 – Choix et modélisation de transitions-types.***

Il conviendra ici de définir des transitions exemples, en fonction des ensembles de connaissances et données à disposition et de l'avancement des modèles conceptuels existant. Cette phase visera aussi à choisir les environnements de modélisation, ainsi qu'à définir la structure des données simulées en cohérence avec le corpus existant. On pourra alors implémenter les modèles de ces transitions exemples.

### ***Phase 3 – Développement de la plate-forme d'exploration par l'analyse des modèles***

L'exploration des modèles constituera un cas d'application permettant de finaliser le choix des méthodes d'exploration proposées par la plate-forme. Lors de cette étape, les retours de thématiciens, sur l'exploration des transitions qu'ils portent, garantiront l'adéquation de la plate-forme avec leurs besoins.

### ***Phase 4 – Conception d'un méta-modèle de transformation des systèmes de peuplement.***

Par l'ouverture de la plate-forme à un ensemble plus large de chercheurs et les retours d'expériences en résultant, on pourra étendre l'environnement d'exploration à d'autres thématiques. La conception, création et validation d'un modèle générique s'appuieront sur une comparaison générale du corpus des cas d'application avec les résultats des modélisations.

## **5.2 Contexte de travail et collaborations possibles**

Ce projet de thèse s'inscrit dans un cadre conceptuel géomatique, mais par l'application à un corpus interdisciplinaire, ce cadre bénéficiera forcément des compétences et angles d'analyse apportés par les géographes et archéologues. L'encadrement pressenti incarne d'ailleurs cette vision interdisciplinaire<sup>17</sup>.

Afin de mener à bien l'étendue des tâches décrites dans ce projet, on s'appuiera fortement sur des dynamiques déjà présentes, et depuis de nombreuses années, au sein de l'UMR Géographie-cités. Les concepts théoriques et une veille scientifique de l'analyse exploratoire de données spatio-temporelles sont en effet étudiés depuis de nombreuses années, tant par des ingénieurs comme Hélène Mathian<sup>18</sup>, qu'au sein de projets auxquels j'ai collaboré, comme l'ERC GeoDiverCity ou l'ANR MIRO<sup>2</sup>.

Dans le domaine de la modélisation et en particulier des modèles orientés agents, on pourra s'appuyer sur son application, innovante en géographie, qui essaime au sein de l'UMR depuis les programmes SimPop (Sanders et al, 1997). Ces connaissances et retours d'expérience pourront être croisées avec la vision plus informatique des systèmes complexes qu'amènent les ingénieurs en informatique du projet GeoDiverCity, notamment sur les problématiques de calibration et de validation automatisées.

En dehors de l'UMR d'accueil, les champs de collaboration possibles sont larges, notamment au sein des Groupements de Recherche MAGIS et MODYS, dans lesquels l'interdisciplinarité joue un rôle clé, centré sur une thématique commune, et qui visent à renforcer les collaborations entre scientifiques de domaines différents, collaborations que ce projet de thèse ambitionne de faciliter par la mise en place d'une méthodologie et d'outils communs aux géographes, archéologues, linguistes et autres chercheurs intéressés par la représentation et l'analyse de phénomènes spatio-temporels sur le temps long.

17 Co-direction de Lena Sanders (Géographe, UMR Géographie-cités) et Anne Ruas (Géomaticienne, IFSTTAR), comité de pilotage enrichi de Laure Nuninger (Archéologue, UMR Chrono-environnement) et Hélène Mathian (Géomaticienne et statisticienne, UMR Géographie-cités).

18 Dont une large partie des recherches consiste à étudier les méthodes et outils d'analyse et de visualisation de données spatio-temporelles, ainsi qu'à mener une veille attentive sur les recherches dans ce domaine.

## 6 Bibliographie

- AMBLARD, Frédéric, ROUCHIER, Juliette et BOMMEL, Pierre, 2006. Evaluation et validation de modèles multi-agents. In : *Modélisation et simulation multi-agents. Applications pour les Sciences de l'Homme et de la Société*, Hermès. 2006. p. 103–140.
- AMOUROUX, Edouard, CHU, Thanh-Quang, BOUCHER, Alain et DROGOUL, Alexis, 2009. GAMA: an environment for implementing and running spatially explicit multi-agent simulations. In : *Agent computing and multi-agent systems*. S.l. : Springer. pp. 359–371.
- ANDRIENKO, Natalia et ANDRIENKO, Gennady, 2006. *Exploratory analysis of spatial and temporal data*. S.l. : Springer Berlin, Germany.
- ANSELIN, Luc, SYABRI, Ibnu et KHO, Youngihn, 2006. GeoDa: An introduction to spatial data analysis. In : *Geographical analysis*. 2006. Vol. 38, n° 1, pp. 5–22.
- BRETAGNOLLE, Anne, LIZZI, Liliane, MATHIAN, Hélène et VAN HAMME, Adrien, 2011. Animating the cities...Dynamic exploration of harmonized urban databases (United-States, France 1800-2000). In : *25th International Cartographic Conference* [en ligne]. Paris, Palais des Congrès, France : s.n. janvier 2011. Disponible à l'adresse : <http://halshs.archives-ouvertes.fr/halshs-00624473>.
- CHARDONNEL, Sonia, 2001. La time-geography : les individus dans le temps et dans l'espace. In : SANDERS, Lena (éd.), *Modèles en analyse spatiale* [en ligne]. S.l. : HERMES Lavoisier. IGAT. pp. 129-156. Disponible à l'adresse : <http://halshs.archives-ouvertes.fr/halshs-00081533>.
- COMIN, Marie-Noëlle, 2010. *Réseaux de villes et réseaux d'innovation en Europe: structuration du système des villes européennes par les réseaux de recherche sur les technologies convergentes*. S.l. : Atelier national de Reproduction des Thèses.
- COSTA, Laurent, 2012. La construction de référentiels géohistoriques: un enjeu pour l'interdisciplinarité dans les sciences historiques. In : *L'Espace géographique*. 2012. n° 4, p. 340–351.
- COUPÉ, Christophe et HOMBERT, Jean-Marie, 2012. Impact of climatic and ecological contexts on sociolinguistic factors among Bantu populations. In : *Workshop « Impact of a major environmental crisis on species, populations and communities: the fragmentation of African forests at the end of the Holocene »*. Paris, March. S.l. : s.n. 2012. pp. 1–2.
- DURAND-DASTES, François, FAVORY, François, FICHES, Jean-Luc, MATHIAN, Hélène, PUMAIN, Denise, RAYNAUD, Claude, SANDERS, Lena et VAN DER LEEUW, Sander, 1998. Des oppida aux métropoles. In : *Anthropos*, Paris. 1998.
- EDSALL, Robert et PEUQUET, Donna, 1997. A graphical user interface for the integration of time into GIS. In : *Proceedings of the 1997 American Congress of Surveying and Mapping Annual Convention and Exhibition, Seattle, WA*. S.l. : s.n. 1997. pp. 182–189.
- FAVORY, François, OURIACHI, Marie-Jeanne et NUNINGER, Laure, 2011. The transformation of rural structures in Southern Gaul between the first century BC and the first century AD. The case of eastern Languedoc (France). In : *Fines imperii - imperium sine fine ? Römische Okkupations- und Grenzpolitik im frühen Prinzipat* [en ligne]. S.l. : s.n. 2011. pp. 157-184.

- GUO, Diansheng, CHEN, Jin, MACEACHREN, Alan M et LIAO, Ke, 2006. A visualization system for space-time and multivariate patterns (vis-stamp). In : *Visualization and Computer Graphics, IEEE Transactions on*. 2006. Vol. 12, n° 6, pp. 1461–1474.
- HALL, Peter Geoffrey et PAIN, Kathy, 2006. *The polycentric metropolis: learning from mega-city regions in Europe*. S.l. : Earthscan/James & James.
- KOHLER, Timothy A. et VARIEN, Mark D., 2010. A scale model of seven hundred years of farming settlements in southwestern Colorado. In : *Becoming villagers: Comparing early village societies*. 2010. Vol. 6, pp. 37.
- KRAAK, Menno-Jan, 1998. The Cartographic Visualization Process: From Presentation to Exploration. In : *Cartographic Journal, The*. 1 juin 1998. Vol. 35, n° 1, pp. 11-15. DOI doi:10.1179/000870498787074100.
- KWAN, Mei-Po, 2004. GIS Methods in Time-Geographic Research: Geocomputation and Geovisualization of Human Activity Patterns. In : *Geografiska Annaler: Series B, Human Geography*. 2004. Vol. 86, n° 4, pp. 267–280.
- LEBART, L, MORINEAU, A, LAMBERT, T et PLEUVRET, P, 1988. Spad. In : *N Système portable pour l'analyse des données. Manuel de référence. (SPAD. N a software for Correspondence Analysis User's. guide). CISIA, Saint-Mandé, France*. 1988.
- L'HOSTIS, Alain, 2003. De l'Espace contracté à l'espace chiffonné: Apports de l'animation à la cartographie en relief des distances-temps modifiées par les réseaux de transport rapides. In : *Revue internationale de géomatique*. 2003. n° 13, pp. 69–80.
- PISON, G, MATHIAN, H, PLUMEJEAUD, C, GENSEL, J, INED, WEB TEAM G et DHERES, FRANCE, 2011. Exploration de la démographie mondiale en ligne. In : 2011.
- PUMAIN, Denise, 1997. Pour une théorie évolutive des villes. In : *Espace géographique*. 1997. Vol. 26, n° 2, pp. 119–134.
- SANDERS, L., PUMAIN, D., MATHIAN, H., GUÉRIN-PACE, F. et BURA, S., 1997. SIMPOP: a multiagent system for the study of urbanism. In : *Environment and Planning B*. 1997. Vol. 24, pp. 287–306.
- SANDERS, Lena, 2006. Les modèles agent en géographie urbaine. In : *Modélisation et simulation multi-agents; applications pour les Sciences de l'Homme et de la Société*. 2006. p. 151–168.
- TAKATSUKA, Masahiro et GAHEGAN, Mark, 2002. GeoVISTA Studio: a codeless visual programming environment for geoscientific data analysis and visualization. In : *Computers & Geosciences*. 2002. Vol. 28, n° 10, pp. 1131–1144.
- THEUS, M, 2005. Statistical Data Exploration and Geographical Information Visualization. In : *Exploring Geovisualization, J. Dykes, AM MacEachren, and M.-J. Kraak, eds*. 2005. pp. 127–142.
- TISUE, S. et WILENSKY, U., 2004. NetLogo: A simple environment for modeling complexity. In : *International Conference on Complex Systems*. S.l. : s.n. 2004. pp. 16–21.