HAL
open science

Identification de causalités dans des données spatio-temporelles

Juste Raimbault

To cite this version:

Juste Raimbault. Identification de causalités dans des données spatio-temporelles. Spatial Analysis and GEOmatics 2017, INSA de rouen, Nov 2017, Rouen, France. hal-01650531v1

HAL Id: hal-01650531

https://hal.science/hal-01650531v1

Submitted on 28 Nov 2017 (v1), last revised 29 Nov 2017 (v2)

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.

Identification de causalités dans des données spatio-temporelles

Juste Raimbault [1,2]

1. UMR CNRS 8504 Géographie-cités

2. UMR-T 9403 IFSTTAR LVMT
juste.raimbault@polytechnique.edu

RÉSUMÉ. *Cet article contribue à la compréhension des processus spatio-temporels fortement couplés, en proposant une méthode générique basée sur la causalité de Granger. Celle-ci est validée par l'identification robuste de régimes de causalité et de leur diagramme de phase pour un modèle de morphogenèse urbaine couplant croissance du réseau et de la densité. L'application au cas réel des projets de transport du Grand Paris démontre un lien entre les dynamiques territoriales, plus particulièrement socio-économiques et foncières, et la croissance anticipée du réseau. Nous discutons finalement des extensions possibles à d'autres échelles temporelles et spatiales.*

ABSTRACT. *This paper contributes to the understanding of strongly coupled spatio-temporal processes by describing a generic method based on Granger causality. The method is validated by the robust identification of causality regimes and of their phase diagram for an urban morphogenesis model that couples network growth with density. The application to the real case study of Greater Paris transportation projects shows a link between territorial dynamics, more particularly of real estate and socio-economic, and the anticipated network growth. We finally discuss potential extensions to other temporal and spatial scales.*

MOTS-CLÉS : *Causalité Spatio-temporelle ; Interactions Réseaux-Territoires ; Morphogenèse Urbaine ; Grand Paris*

KEYWORDS: *Spatio-temporal Causality ; Network-territories Interactions ; Urban Morphogenesis ; Greater Paris*

1. Introduction

L'étude des processus spatio-temporels fortement couplés implique la prise en compte d'intrications entre ceux-ci généralement difficiles à isoler. Essence même des approches par la complexité, ces interactions qui sont à l'origine du comportement émergent d'un système font sens comme objet d'étude en lui-même, et une séparation des processus paraît alors contradictoire avec une vision intégrée du système. Dans le cas des systèmes territoriaux, l'exemple des interactions entre réseaux de transport et territoires est une excellente allégorie de ce phénomène : des méthodes isolant les "effets structurants" d'une infrastructure développées dans les années 70 (Bonnafous, Plassard, 1974) se sont révélées par la suite de l'instrumentation politique et sans fondement empirique (Offner, 1993). Le débat est toujours d'actualité puisque la question se pose actuellement par exemple pour la construction de lignes à grande vitesse (Crozet, Dumont, 2011). La réalité des processus territoriaux est en fait bien plus compliquée qu'une simple relation causale entre la mise en place d'une infrastructure et les retombées sur le développement local, mais correspond bien à une *co-évolution* complexe (Bretagnolle, 2009). Sur le temps long et à grande échelle, certains effets de renforcement des dynamiques dans les systèmes de villes par l'insertion dans les réseaux, ont été mis en valeur par l'application de la Théorie Evolutive des Villes (L'Espace géographique, 2014), montrant que le démêlage est toutefois possible dans certains cas par une compréhension plus globale du système. A une autre échelle, toujours concernant les relations entre réseaux et territoires, on peut citer les liens entre pratiques de mobilité, étalement urbain et localisation des ressources dans un cadre métropolitain (Cerqueira, 2017) qui s'avèrent tout autant complexes. Ce type de problématique est bien sûr présent dans d'autres domaines : en Economie Géographique, l'exemple des liens entre innovation, impacts locaux de la connaissance et aggregation des agents économiques est une illustration typique de processus économiques spatio-temporels présentant des causalités circulaires difficiles à démêler (Audretsch, Feldman, 1996). Des méthodes spécifiques sont introduites, comme l'utilisation d'instruments statistiques : (Aghion *et al.*, 2015) montre que l'origine géographique des membres du Bureau du Congrès américain attribuant les subventions locales est une bonne variable instrumentale pour lier caractère innovant et inégalités des plus haut salaires, ce qui permet de confirmer que la correlation significative entre les deux est en fait une causalité de l'innovation sur les inégalités.

Le couplage fort spatio-temporel implique généralement l'introduction de la notion de causalité, à laquelle la géographie s'est toujours intéressée : (Loi, 1985) montre que les questions fondamentales que se pose la géographie théorique récente (isolation des objects, lien entre espace et structures causales, etc.) étaient déjà présentes dans la géographie classique de Vidal. (Claval, 1985) critique d'ailleurs les nouveaux déterminismes ayant émergé, notamment celui proposé par certains tenants de l'analyse systémique : dans ses débuts, cette

approche héritait de la cybernétique et donc d'une vision réductionniste impliquant un déterminisme même dans une formulation probabiliste. Claval note que des travaux contemporains à son écriture devraient permettre de capturer la complexité qui fait la particularité des décisions humaines : l'école de Prigogine et la Théorie des Catastrophes de Thom. Ce point de vue est remarquablement visionnaire, puisque comme le rappelle Pumain dans (Pumain, 2003), le glissement de l'analyse des systèmes à l'auto-organisation puis à la complexité a été long et progressif, et ces travaux ont été fondamentaux pour le permettre. François Durand-Dastès résume cette situation plus récemment dans (Durand-Dastes, 2003), en appuyant l'importance des bifurcations et de la dépendance au chemin lors des instants initiaux de la constitution du système qu'il désigne par *systèmogenèse*. Ce type de dynamique complexe implique généralement une co-évolution des composantes du système, qu'on peut interpréter comme des causalités circulaires entre processus : la question de pouvoir les identifier est donc cruciale au regard de la notion de causalité pour la géographie complexe contemporaine.

Les régimes sous lesquels des identifications de causalité sont cohérentes ne sont pas identifiés de manière évidente. Ceux-ci dépendront des définitions utilisées, de la même manière que les méthodes à disposition pour lesquelles nous pouvons donner quelques illustrations. (Liu *et al.*, 2011) propose la detection de relations spatio-temporelles entre perturbations des flots de trafic, introduisant une définition particulière de la causalité basée sur une correspondance de points extrêmes. Les algorithmes associés sont toutefois spécifiques et difficilement applicables à des types de systèmes différents. L'utilisation des correlations spatio-temporelles a été démontrée comme ayant dans certains cas un fort pouvoir prédictif pour les flots de traffic (Min, Wynter, 2011). Egalement dans le domaine des transports et de l'usage du sol, (Xie, Levinson, 2009) applique une analyse par causalité de Granger, qu'on pourra interpréter comme une corrélation retardée, pour montrer dans un cas particulier que la croissance du réseau induit le développement urbain et est elle-même tirée par des externalités comme les habitudes de mobilité. Les neurosciences ont développé de nombreuses méthodes répondant à des problématiques similaires. (Luo *et al.*, 2013) définit une causalité de Granger généralisée prenant en compte la non-stationnarité et s'appliquant à des régions abstraites issues d'imagerie fonctionnelle. Ce genre de méthode est également développée en Vision par Ordinateur, comme l'illustre (Ke *et al.*, 2007) qui exploite les correlations spatio-temporelles de formes et de flux dans des successions d'images pour classifier et reconnaître des actions. Les applications peuvent être très concrètes comme la compression de fichier videos par extrapolation des vecteurs de mouvement (Chalidabhongse, Kuo, 1997). Dans l'ensemble de ces cas, l'étude des correlations spatio-temporelles rejoint les notions faibles de causalité vues précédemment.

Dans le cas particulier des relations entre réseaux et territoires, des études majoritairement en économétrie ont cherché à établir des relations de causalité

entre variables liées a ces deux objets. Par exemple, (Levinson, 2008) explique pour Londres les variables de population et de connectivité au réseau par ces mêmes variables décalées dans le temps, démontrant des effets causaux circulaires. (Gargi Chaudhuri and Keith C Clarke, 2015) utilise des techniques similaires sur une region d'Italie avec des données historiques sur le temps long, mais modère les conclusions d'effets systématiques en rappelant l'importance des événements historiques sur les relations estimées. (Cuthbert *et al.*, 2005) procède à des estimations économétriques des influences réciproques, et conclut que dans le cas d'étude canadien à une échelle sous-régionale, le développement du réseau induit le développement de l'usage du sol mais pas l'inverse. L'échelle de temps et d'espace influencent ainsi significativement les résultats de telles analyses. (Koning *et al.*, 2013) propose une estimation des relations entre existence d'une desserte de Train à Grande Vitesse et variables économiques sur les Unites Urbaines Françaises, et conclut à un effet propre de la desserte négatif, après contrôle du caractère endogène de la desserte par un modèle de sélection, et à un effet significatif des caractéristiques des unités urbaines. Cette étude reste limitée, ne prenant ni en compte un décalage temporel supérieur à une unité de temps, ni des relations spatiales entre entités. Enfin, toujours dans la meme logique mais sans prise en compte explicite de l'espace, (Canning, Pedroni, 2008) montre sur le temps long un lien de causalité entre stock d'infrastructure et croissance économique sur un panel mondial, mais que ces effets sont atténués localement par des sous ou sur-investissements.

Cette contribution cherche à explorer la possibilité d'une méthode analogue pour des données spatio-temporelles présentant a priori des causalités circulaires complexes, et donc de tenter l'exercice d'équilibriste de concilier un certain niveau de simplicité et de caractère opérationnel à une prise en compte de la complexité. Nous introduisons ainsi une méthode d'analyse des correlations spatio-temporelles similaire à une causalité de Granger estimée dans le temps et l'espace, dont la robustesse est démontrée systématiquement par l'application à un modèle de simulation complexe de morphogenèse urbaine et par l'isolation de régimes de causalités distincts dans l'espace des phases du modèle. Notre contribution inclut également l'application à un cas d'étude empirique, ce qui la positionne à l'interface des domaines de la méthodologie, de la modélisation et de l'empirique dans le cadre épistémologique introduit par (Raimbault, 2017).

La suite de cette article est organisée de la façon suivante : le cadre générique de la méthode proposée est décrit dans la section suivante. Nous l'appliquons ensuite à un jeu de données synthétiques afin de la valider partiellement et de tester ses potentialités, ce qui permet de l'appliquer ensuite au cas d'étude réel des réseaux de transport du Grand Paris. Nous discutons finalement la proximité avec d'autres méthodes existantes et des développements possibles.

2. Méthode

Nous formalisons ici de manière générique la méthode, basée sur une formulation faible de la causalité de Granger, pour tenter d'identifier des relations causales dans des systèmes spatiaux. Soit $X_j(\vec{x}, t)$ des processus aléatoires spatiaux unidimensionnels, se réalisant dans le temps et l'espace. On se donne un ensemble d'unités spatiales fondamentales (u_i) qui peuvent être par exemple les cellules d'un raster ou un pavage quelconque de l'espace géographique. On suppose l'existence de fonctions $\Phi_{i,j}$ permettant de faire correspondre les réalisations de chaque composante aux unités spatiales, possiblement par une première agrégation locale, ou par un processus plus élaboré conduit par un réseau par exemple. Une réalisation d'un système est donnée par un ensemble de trajectoires pour chaque processus $x_{i,j,t}$, et on pourra noter un ensemble de réalisations $x^{(k)}_{i,j,t}$ (accessibles dans le cas d'un modèle de simulation par exemple, ou par hypothèse de comparabilité de sous-systèmes territoriaux dans des cas réels). On suppose disposer d'un estimateur de correlation $\hat{\rho}$ s'exerçant dans le temps, l'espace et les répétitions, i.e. $\hat{\rho}\left[X, Y\right] = \hat{\mathbb{E}}_{i,t,k}\left[XY\right] - \hat{\mathbb{E}}_{i,t,k}\left[X\right]\hat{\mathbb{E}}_{i,t,k}\left[Y\right]$. Il est important de noter ici l'hypothèse de stationnarité spatiale et temporelle, qui peut toutefois aisément se relâcher dans le cas d'une stationnarité locale. D'autre part, l'autocorrelation spatiale n'est pas explicitement incluse, mais est prise en compte soit par l'agrégation initiale si l'échelle caractéristique des unités est plus grande que celle des effets de voisinage, soit par un estimateur spatial adéquat (statistiques spatiales pondérées de type *GWR* (Brunsdon *et al.*, 1998) par exemple). Cela nous permet de définir la correlation retardée par

$$\rho_\tau\left[X_{j_1}, X_{j_2}\right] = \hat{\rho}\left[x^{(k)}_{i,j_1,t-\tau}, x^{(k)}_{i,j_2,t}\right] \tag{1}$$

La corrélation retardée n'est pas directement symétrique, mais on a de manière évidente $\rho_\tau\left[X_{j_1}, X_{j_2}\right] = \rho_{-\tau}\left[X_{j_2}, X_{j_1}\right]$. On applique alors cette mesure de manière simple : si $\mathrm{argmax}_\tau \rho_\tau\left[X_{j_1}, X_{j_2}\right]$ ou $\mathrm{argmin}_\tau \rho_\tau\left[X_{j_1}, X_{j_2}\right]$ sont "clairement définis" (les deux pouvant l'être simultanément), leur signe donnera alors le sens de la causalité entre les composantes j_1 et j_2 et leur valeur absolue le retard de propagation. Les critères de significativité dépendront du cas d'application et de l'estimateur utilisé, mais peuvent par exemple inclure la significativité du test statistique (test de Fisher dans le cas d'un estimateur de Pearson), la position des bornes d'un intervalle de confiance à un niveau donné, ou même un seuil exogène θ sur $|\rho_\tau|$ pour forcer un certain degré de correlation.

3. Résultats

3.1. Données Synthétiques

Cette méthode doit dans un premier temps être testée et partiellement validée, ce que nous proposons de faire sur des données synthétiques, méthode qui

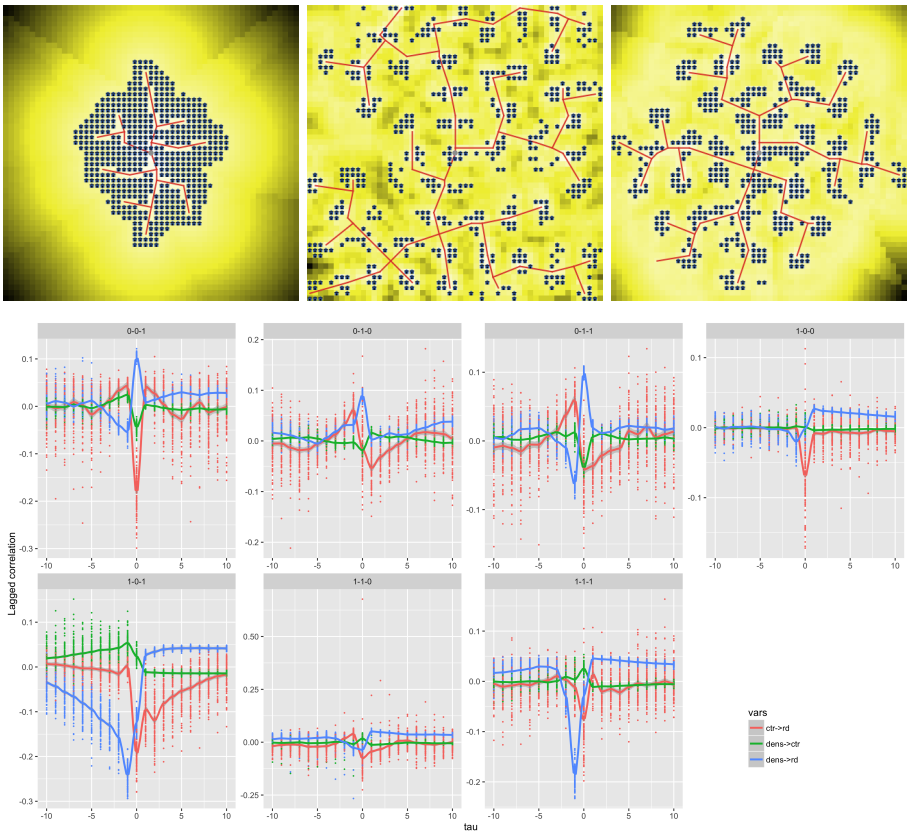

FIGURE *1.* ***Correlations dans le modèle RDB (Première ligne)*** *Exemples de configurations finales variées, obtenues avec* (w_d, w_c, w_r) *valant respectivement* $(0,1,1)$,$(1,0,1)$, *et* $(1,1,1)$. ***(Deuxième ligne)*** *Corrélations retardées, pour chaque combinaison des paramètres, en fonction du retard* τ. *Les différentes couleurs correspondent à chaque couple de variables : distance au centre (`ctr`), densité (`dens`) et distance au réseau (`rd`). Les points montrent l'étendue sur l'ensemble des répétitions du modèle (estimateurs sur* i *et* t*).*

permet une connaissance plus fine des comportements des modèles (Raimbault, 2016). En écho à l'exemple des relations entre réseaux de transport et territoires qui a permis d'introduire notre problématique précédemment, nous proposons de générer des configurations urbaines stylisées dans lesquelles réseau et densité s'influencent mutuellement, et pour lesquelles les causalités ne sont pas évidents *a priori* étant donné les paramètres du modèle génératif. (Raimbault

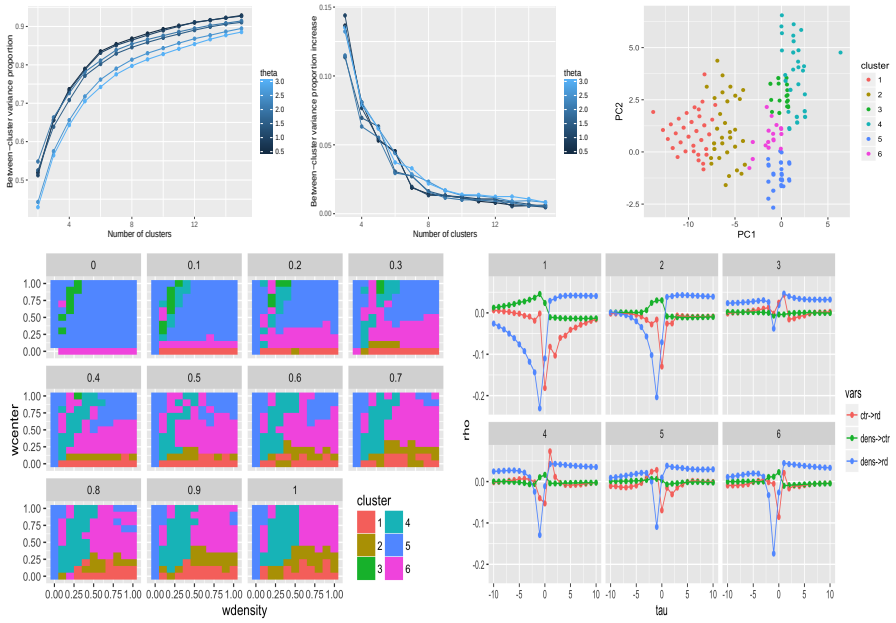

FIGURE 2. ***Identification de régimes d'interactions (Haut Gauche)*** *Variance inter-cluster comme fonction du nombre de clusters.* ***(Haut Milieu)*** *Dérivée de la variance inter-cluster.* ***(Haut Droite)*** Features *dans un plan principal (81% de variance expliquée par les deux premières composantes)* ***(Bas Gauche)*** *Diagramme de phase des régimes dans l'espace* (w_d, w_c, w_r), w_r *variant entre les différents sous-diagrammes de* $(w_d, w_c$. ***(Bas Droite)*** *Trajectoires correspondantes des centroïdes.*

et al., 2014) décrit et explore un modèle simple de morphogenèse urbaine (modèle RBD) répondant parfaitement à ces contraintes. En effet, les variables explicatives de la croissance urbaine, les processus d'extension du réseau et le couplage entre densité urbaine et réseau sont relativement simples. Cependant, hormis dans des cas extrêmes (par exemple lorsque la distance au centre détermine la valeur foncière uniquement, le réseau dépendra de manière causale de la densité, ou lorsque la distance au réseau seule compte, la causalité sera inversée), les régimes mixtes n'exhibent pas de causalités évidentes : c'est donc un parfait cas pour tester si la méthode est capable d'en détecter. Nous utilisons une implémentation adaptée [1] du modèle initial, permettant de capturer les valeurs des variables étudiées pour chaque cellule de l'automate cellulaire et

1. disponible sur le dépôt ouvert du projet à
`https://github.com/JusteRaimbault/CityNetwork/tree/master/Models/Simple/ModelCA`

à chaque pas de temps et de calculer les correlations retardées, au sens donné précédemment, entre variables au sein du modèle. Nous explorons une grille de l'espace des paramètres du modèle RBD, faisant varier les paramètres de poids de la densité, de la distance au centre et de la distance au réseau [2], que l'on note respectivement (w_d, w_c, w_r), dans $[0; 1]$ avec un pas de 0.1. Les autres paramètres sont fixés à leur valeurs par défaut données par (Raimbault *et al.*, 2014). Pour chaque valeur des paramètres, nous procédons à $N = 100$ répétitions ce qui est suffisant pour une bonne convergence des indicateurs. Les explorations sont effectuées via le logiciel OpenMole (Reuillon *et al.*, 2013), le grand nombre de simulations (1,330,000) nécessitant l'utilisation d'une grille de calcul. Nous calculons sur l'ensemble des patches les corrélations retardées par estimateur de Pearson non biaisé entre les variations des variables suivantes [3] : densité locale, distance au centre et distance au réseau. La Fig. 1 montre le comportement de ρ_τ pour chaque couple de variable (non dirigé, τ prenant des valeurs négatives et positives), pour les combinaisons des valeurs extrêmes des paramètres. On peut voir déjà différents régimes émerger : par exemple, $(1, 0, 1)$ conduit à une causalité de la densité sur la distance au centre avec un retard $\tau = 1$, et une causalité négative de la densité sur la distance au réseau avec le même retard, tandis que distance au centre et au réseau sont corrélées de manière synchrone. Afin d'étudier ces comportements de manière systématique, nous proposons d'identifier des régimes de manière endogène, en procédant à un apprentissage non-supervisé. Nous appliquons une classification des *k-means*, robuste à la stochasticité (5000 répétitions), avec les points caractéristiques (*features*) suivants : pour chaque couple de variable, $\text{argmax}_\tau \rho_\tau$ et $\text{argmin}_\tau \rho_\tau$ si la valeur correspondante est telle que $\frac{\rho_\tau - \bar{\rho}_\tau}{|\bar{\rho}_\tau|} > \theta$ avec θ paramètre de seuil, 0 sinon. L'inclusion des *features* supplémentaires des valeurs de ρ_τ n'influence pas significativement les résultats, celles-ci n'ont pas été prises en compte pour réduire la dimension. Le choix du nombre de clusters k est en général épineux dans ce genre de problème (Hamerly *et al.*, 2003), dans notre cas le système possède une structure agréable : les courbes de la proportion de variance inter-cluster et de sa dérivée en Fig. 2, en fonction de k pour différentes valeurs de θ, présentent une transition pour $\theta = 2$, ce qui donne pour cette courbe une rupture à $k = 5$. Un examen visuel des clusters dans un plan principal confirme la bonne qualité de la classification pour ces valeurs. Une classe correspond alors à un *régime de causalité*, dont nous pouvons représenter le diagramme de phase en fonction des paramètres du modèle, ainsi que les trajectoires des centres des clusters (calculées comme barycentre dans l'espace complet initial) en Fig. 2. Le comportement obtenu est particulièrement intéressant : les régions du diagramme correspondant aux régimes sont clairement

2. Le modèle fonctionne de la façon suivante : une valeur des cellules est déterminée par la moyenne pondérée de ces différentes variables explicatives, valeur qui détermine la croissance de nouveaux patches à l'instant suivant.

3. Calculer les corrélations sur les variables directement n'a pas de sens puisque leur valeur n'en a pas en absolu.

délimitées et connexes. Par exemple, on observe l'émergence du régime 6 où la distance au réseau cause fortement la densité de manière négative, mais la distance au centre cause la distance au réseau, régime dont l'étendue maximale sur (w_d, w_r) est pour une valeur intermédiaire $w_r = 0.7$. Ainsi, pour maximiser l'impact du réseau sur la densité, il ne faut pas maximiser le poids correspondant, ce qui peut paraître contre-intuitif en premier abord. Cela illustre l'intérêt de la méthode dans le cas de relations circulaires difficiles à démêler a priori. Le régime 5, où la distance au réseau influence la densité de la même manière, mais la relation entre distance au centre et route est inversée, est tout aussi intéressant, et est prédominant dans les faibles w_r. Le régime 1, extrême, correspond à une situation isolée dans laquelle la distance au centre n'importe pas : cet aspect domine alors totalement les autres processus d'interaction entre densité et réseau. Cette application sur données synthétiques démontre ainsi d'une part la robustesse de la méthode vu la cohérence des régimes obtenus, et constitue aussi une qualification beaucoup plus précise des comportements du modèle que celle réalisée dans l'article initial. Dans ce cas précis, il peut s'agir d'un instrument de connaissance des relations entre réseaux et territoires en lui-même, permettant le test d'hypothèses ou la comparaison de processus dans le modèle stylisé.

3.2. Cas d'étude

3.2.1. Contexte

Nous proposons une application sur un cas d'étude réel, toujours lié aux relations entre réseaux de transport et territoires. La région métropolitaine de Paris est en train de connaître de grandes mutations, avec la mise en place d'une gouvernance métropolitaine et de nouvelles infrastructures de transport par exemple. La construction d'un réseau de métro en rocade permettant des liaisons de banlieue à banlieue est un besoin ancien, et a mené à plusieurs propositions sur lesquelles se sont opposés l'Etat et la Région au tournant des années 2010 (Desjardins, 2010). Le projet Arc Express (STIF, 2010), porté par la Région, axé sur une égalité des territoires, contrastait avec les propositions initiales de Réseau du Grand Paris visant à relier des "clusters d'excellence" en dépit d'un possible effet tunnel. La solution finalement adoptée (voir le dernier schéma directeur (SDRIF, 2013)) est un compromis et permet un rééquilibrage est-ouest de l'accessibilité (Beaucire, Drevelle, 2013). Nous proposons d'étudier les relations entre différentiel d'accessibilité pour chaque projet, et variables liées au foncier (transactions immobilières) et socio-économiques. En effet, les liens entre nouvelles lignes et évolution du foncier sont parfois remarquables (Damm *et al.*, 1980).

3.2.2. Données

Les données des transactions immobilières sont fournies par la base BIENS (Chambre des Notaires d'Ile de France, base propriétaire). Le nombre de tran-

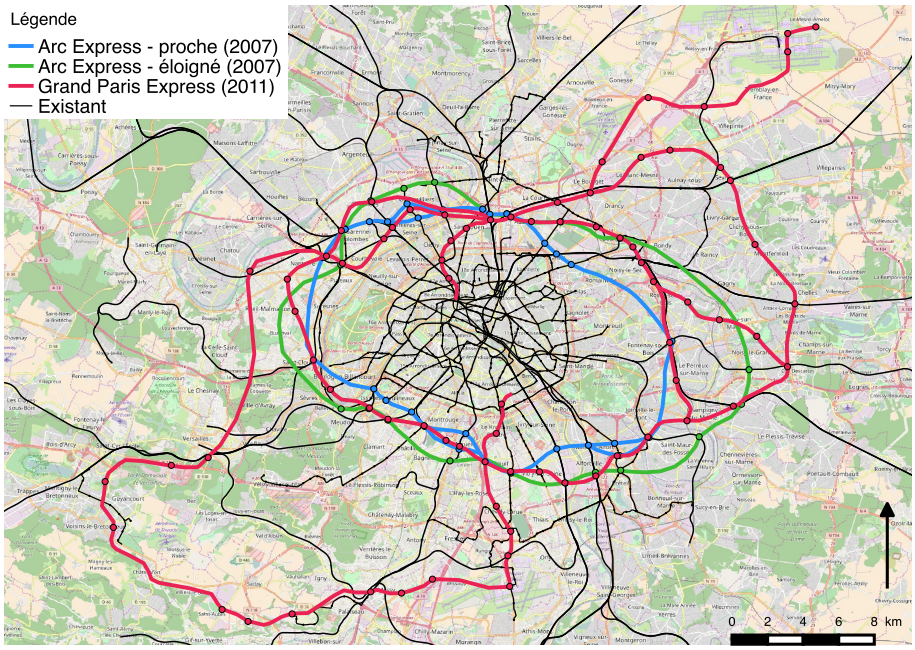

FIGURE 3. ***Projets de transport successifs de la métropole du Grand Paris.*** *Nous montrons les deux alternatives du projet Arc Express porté par la région, et le Grand Paris Express (GPE) porté par l'état. Le Réseau du Grand Paris, précurseur du GPE, n'est pas montré ici pour des raisons de visibilité à cause de sa proximité avec celui-ci.*

sactions utilisables après nettoyage est de 862360, se répartissant sur l'ensemble des IRIS, pour une plage temporelle couvrant des années 2003 à 2012 incluses. Les données par IRIS pour population et revenu (revenu médian et indice de Gini) proviennent de l'INSEE. Les données de réseau ont été vectorisées à partir des cartes des projets (voir Fig. 3 pour les projets). Les temps de trajets sont calculés par transport en commun uniquement, avec des valeurs standard pour les vitesses moyennes des différents modes (RER 60km.h^{-1}, Transilien 100km.h^{-1}, Metro 30km.h^{-1}, Tramway 20km.h^{-1}). La matrice des temps est calculée depuis l'ensemble des centroïdes des IRIS vers l'ensemble des centroïdes des communes. Ceux-ci sont reliés au réseau par des connecteurs à la gare la plus proche, de vitesse 50km.h^{-1} (trajet en voiture). Les analyses sont implémentées

intégralement en langage R (R Core Team, 2015) et l'ensemble des données, du code source et des résultats sont disponibles sur un dépôt git ouvert [4].

3.2.3. Résultats

Nous calculons pour chaque projet, le différentiel ΔT_i d'accessibilité en temps moyen de trajet à partir de chaque IRIS en comparaison à celui dans le réseau sans le projet, défini par $T_i = \sum_k \exp -t_{ik}/t_0$ avec k communes, t_{ik} temps de trajet, et t_0 paramètre d'atténuation. A chaque projet est associée une date [5], correspondant environ à l'année d'annonce mature du projet, restant toutefois arbitraire car difficile d'une part à déterminer précisément, un projet n'émergeant pas d'un coup du jour au lendemain, et d'autre part pouvant correspondre à des réalités différentes d'apprentissage du projet par les différents agents économiques (nous faisons donc l'hypothèse réductrice mais nécessaire d'une diffusion sur la majorité des agents dans un temps inférieur à l'année). Nous étudions les corrélations décalées de cette variable avec les variations ΔY_{ij} des variables socio-économiques suivantes : population, revenu médian, indice de Gini des revenus, prix moyen des transactions immobilières et montant moyen des crédits immobiliers. Un test de Fisher est effectué pour chaque estimation, et la valeur est fixée nulle si celui-ci n'est pas significatif ($p < 0.05$ de manière classique). L'étude avec accessibilité généralisée au sens de Hansen a également été menée mais moins intéressante car, très peu sensible à la composante mobilité (réseau et atténuation) par rapport aux variables elle-mêmes, informe uniquement sur des relations entre celles-ci et n'est donc pas présentée ici. Nous présentons en Fig. 4 les résultats pour l'ensemble des réseaux et variables. Il est remarquable tout d'abord de noter l'existence d'effets significatifs pour l'ensemble des variables. Des valeurs plus basses du paramètre t_0 donnent des corrélations plus fortes en valeur absolue, révélant une possible plus grande importance de l'accessibilité locale sur les dynamiques territoriales. Le comportement de la population montre un pic très détaché correspondant à 2008, laissant supposer un impact du plus vieux projet d'Arc Express sur la croissance de la population, l'effet des autres projets serait alors fallacieux de par leur proximité dans les grands tronçons : cela impliquerait que les zones où ils diffèrent fondamentalement comme le Plateau de Saclay ne soient que très peu sensibles au projet de transport, ce qui confirmerait l'aspect artificiel planifié du développement de ce territoire. Concernant les revenus, on observe un comportement similaire mais négatif, ce qui impliquerait un appauvrissement lié à l'augmentation de l'accessibilité, mais qui semble toutefois s'accompagner d'une baisse des inégalités. Enfin, comme attendu les prix immobiliers sont tirés

4. A l'adresse `https://github.com/JusteRaimbault/CityNetwork/tree/master/Models/SpatioTempCausality/GrandParis`. Les données de la base BIENS ne sont fournies que de manière agrégée à l'IRIS et pour les variables de prix et de crédit, pour des raisons de fermeture contractuelle de la base brute.

5. 2006 pour Arc Express, 2008 pour le Réseau du Grand Paris, 2010 pour le Grand Paris Express

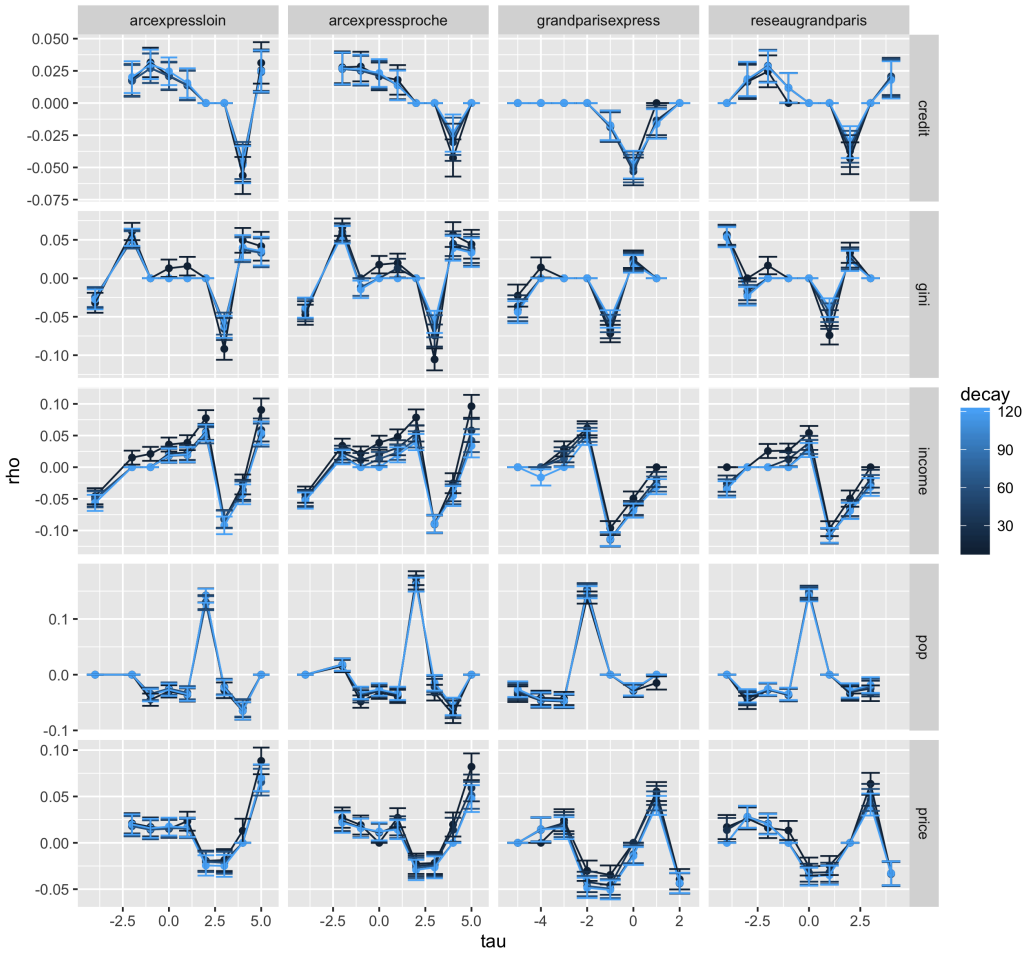

FIGURE 4. ***Corrélations retardées empiriques.*** *Les graphiques donnent la valeur de la correlation entre le différentiel d'accessibilité en temps de trajet moyen ΔT pour chaque projet (en colonnes) et le différentiel des différentes variables socio-economiques et de transactions immobilières (en lignes), pour différentes valeurs du paramètre d'atténuation (`decay`). Les barres d'erreur donnent l'intervalle de confiance à 95%.*

par l'arrivée potentielle des nouveaux réseaux, effet qui disparait à deux ans pour le Grand Paris Express, suggérant une bulle immobilière passagère. Nous démontrons ainsi l'existence de liens de correlations retardées complexes qu'on nomme causalités en ce sens, entre dynamiques territoriales et dynamiques an-

ticipées des réseaux. Une compréhension plus fine des processus à l'oeuvre est au delà de la portée de cet article, car supposerait des études de terrain qualitatives, des études de cas ciblées, etc. Cet exemple illustre cependant le caractère opérationnel de notre méthode sur un cas d'étude réel.

4. Discussion

4.1. Diffusion spatio-temporelle

L'application de notre approche doit être menée précautionneusement concernant le choix des échelles, processus et objets d'étude. Typiquement, elle ne sera pas du tout adaptée à la quantification de processus spatio-temporels dont l'échelle temporelle de diffusion est de l'ordre de celle de la fenêtre d'estimation : l'hypothèse de stationnarité est basique. On peut proposer de procéder à des estimations par fenêtres glissantes, mais il faudrait ensuite élaborer une technique de correspondance spatiale pour traquer la propagation des phénomènes. Un exemple d'application concrète à l'impact thématique fort serait une caractérisation d'une composante fondamentale de la Théorie Evolutive des Villes, la diffusion hiérarchique de l'innovation entre les villes (Pumain, 2010), en analysant les potentielles dynamiques spatio-temporelles des classifications de brevets comme celle introduite par (Bergeaud *et al.*, 2017). Il faut noter toutefois qu'il s'agit de questions méthodologiques relativement ouvertes, dont une des manifestations est le lien potentiel entre le caractère non-ergodique des systèmes urbains (Pumain, 2012) et une caractérisation ondulatoire de ces processus.

4.2. Regression Géographique Pondérée

Une autre direction de développement et d'applications potentielles se révèle en se tournant vers l'échelle plus locale, et d'explorer une hybridation avec les techniques de Regression Géographique Pondérée (Brunsdon *et al.*, 1998). La détermination par validation croisée ou Critère d'Akaike d'une portée spatiale optimale pour la performance de ce type de modèles, comme introduit par (Raimbault, Bergeaud, 2017), pourrait être adaptée dans notre cas pour déterminer une échelle locale optimale sur laquelle les correlations retardées sont les plus significatives, ce qui permettrait de s'extraire du problème de la non-stationnarité prioritairement par l'aspect spatial.

5. Conclusion

Nous avons proposé une méthode générique de causalité de Granger sur des données territoriales spatio-temporelles, puis démontré sa potentialité et son caractère opérationnel sur données synthétiques et un cas réel. Nous postulons

que l'appareillage méthodologique simple est un atout pour une certaine généralité, mais que l'application à ces cas complexes présentant des causalités circulaires démontre le fort potentiel de contribution à la compréhension des dynamiques de ce type de systèmes co-évolutifs.

Remerciements

Les résultats obtenus dans la section 3.1 de cet article ont été calculés sur l'organisation virtuelle `vo.complex-system.eu` de l'European Grid Infrastructure (`http://www.egi.eu`). Nous remercions l'*European Grid Infrastructure* et ses *National Grid Initiatives* (France-Grilles en particulier) pour fournir le support technique et l'infrastructure.

Bibliographie

Aghion P., Akcigit U., Bergeaud A., Blundell R., Hémous D. (2015). *Innovation and top income inequality*. Rapport technique. National Bureau of Economic Research.

Audretsch D. B., Feldman M. P. (1996). R&d spillovers and the geography of innovation and production. *The American economic review*, vol. 86, n° 3, p. 630–640.

Beaucire F., Drevelle M. (2013). «grand paris express»: un projet au service de la réduction des inégalités d'accessibilité entre l'ouest et l'est de la région urbaine de paris? *Revue d'Économie Régionale & Urbaine*, n° 3, p. 437–460.

Bergeaud A., Potiron Y., Raimbault J. (2017, 04). Classifying patents based on their semantic content. *PLOS ONE*, vol. 12, n° 4, p. 1-22. Consulté sur https://doi.org/10.1371/journal.pone.0176310

Bonnafous A., Plassard F. (1974). Les méthodologies usuelles de l'étude des effets structurants de l'offre de transport. *Revue économique*, p. 208–232.

Bretagnolle A. (2009). *Villes et réseaux de transport : des interactions dans la longue durée, France, Europe, États-Unis*. Hdr, Université Panthéon-Sorbonne - Paris I. Consulté sur http://tel.archives-ouvertes.fr/tel-00459720

Brunsdon C., Fotheringham S., Charlton M. (1998). Geographically weighted regression. *Journal of the Royal Statistical Society: Series D (The Statistician)*, vol. 47, n° 3, p. 431–443.

Canning D., Pedroni P. (2008). Infrastructure, long-run economic growth and causality tests for cointegrated panels. *The Manchester School*, vol. 76, n° 5, p. 504–527. Consulté sur http://dx.doi.org/10.1111/j.1467-9957.2008.01073.x

Cerqueira E. V. (2017). Les inégalités d'accès aux ressources urbaines dans les franges périphériques de belo horizonte (brésil): quelles évolutions? *EchoGéo*, n° 39.

Chalidabhongse J., Kuo C. J. (1997). Fast motion vector estimation using multiresolution-spatio-temporal correlations. *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 7, n° 3, p. 477–488.

Claval P. (1985). Causalité et géographie. *Espace géographique*, vol. 14, n° 2, p. 109–115.

Crozet Y., Dumont F. (2011). Retour sur les effets économiques du TGV. Les effets structurants sont un mythe (interview). *Ville, Rail et Transports*, n° 525, p. 48-51. Consulté sur https://halshs.archives-ouvertes.fr/halshs-01094554

Cuthbert A. L., Anderson W. P., Hall F. L. (2005). An empirical analysis of the relationship between road development and residential land development. *Canadian Journal of Regional Science*, vol. 28, n° 1, p. 49–76.

Damm D., Lerman S. R., Lerner-Lam E., Young J. (1980). Response of urban real estate values in anticipation of the washington metro. *Journal of Transport Economics and Policy*, p. 315–336.

Desjardins X. (2010). la bataille du grand paris. *L'Information géographique*, vol. 74, n° 4, p. 29–46.

Durand-Dastes F. (2003). *Les géographes et la notion de causalité.* PUF.

Gargi Chaudhuri and Keith C Clarke. (2015). On the Spatiotemporal Dynamics of the Coupling between Land Use and Road Networks: Does Political History Matter? *Environment and Planning B: Planning and Design*, vol. 42, n° 1, p. 133-156. Consulté sur {http://dx.doi.org/10.1068/b39089}

Hamerly G., Elkan C. *et al.* (2003). Learning the k in k-means. In *Nips*, vol. 3, p. 281–288.

Ke Y., Sukthankar R., Hebert M. (2007). Spatio-temporal shape and flow correlation for action recognition. In *Computer vision and pattern recognition, 2007. cvpr'07. ieee conference on*, p. 1–8.

Koning M., Blanquart C., Delaplace M. (2013, juillet). Dessertes ferroviaires à grande vitesse et dynamisme économique local : Une analyse économétrique exploratoire sur les unités urbaines françaises. In *ASRDLF 2013, 50ème colloque de l'Association des sciences régionales de langue française*, p. 32p. Belgium. Consulté sur https://hal.archives-ouvertes.fr/hal-00962384

L'Espace géographique. (2014). *Les effets structurants des infrastructures de transport, l'espace géographique 2014/1 (tome 43), p. 51-67.*

Levinson D. (2008). Density and dispersion: the co-development of land use and rail in london. *Journal of Economic Geography*, vol. 8, n° 1, p. 55–77.

Liu W., Zheng Y., Chawla S., Yuan J., Xing X. (2011). Discovering spatio-temporal causal interactions in traffic data streams. In *Proceedings of the 17th acm sigkdd international conference on knowledge discovery and data mining*, p. 1010–1018.

Loi D. (1985). Une étude de la causalité dans la géographie classique française.[l'exemple des premières thèses régionales]. *Espace géographique*, vol. 14, n° 2, p. 121–125.

Luo Q., Lu W., Cheng W., Valdes-Sosa P. A., Wen X., Ding M. *et al.* (2013). Spatio-temporal granger causality: A new framework. *NeuroImage*, vol. 79, p. 241–263.

Min W., Wynter L. (2011). Real-time road traffic prediction with spatio-temporal correlations. *Transportation Research Part C: Emerging Technologies*, vol. 19, n° 4, p. 606–616.

Offner J.-M. (1993). Les "effets structurants" du transport: mythe politique, mystification scientifique. *Espace géographique*, vol. 22, n° 3, p. 233–242.

Pumain D. (2003). Une approche de la complexité en géographie. *Geocarrefour*, vol. 78, n° 1, p. 25–31.

Pumain D. (2010). Une théorie géographique des villes. *Bulletin de la Société géographie de Liège*, n° 55, p. 5–15.

Pumain D. (2012). Urban systems dynamics, urban growth and scaling laws: The question of ergodicity. In *Complexity theories of cities have come of age*, p. 91–103. Springer.

R Core Team. (2015). R: A language and environment for statistical computing Manuel de logiciel. Vienna, Austria. Consulté sur http://www.R-project.org/

Raimbault J. (2016, janvier). Génération de Données Synthétiques Corrélées. In *Rochebrune 2016, Journées d'Etude sur les Systèmes Complexes Naturels et Artificiels.* Mégève, France. Consulté sur https://halshs.archives-ouvertes.fr/halshs-01514415

Raimbault J. (2017, juin). An applied knowledge framework to study complex systems. *ArXiv e-prints.*

Raimbault J., Banos A., Doursat R. (2014). A hybrid network/grid model of urban morphogenesis and optimization. In *Proceedings of the 4th international conference on complex systems and applications (iccsa 2014), june 23-26, 2014, université de normandie, le havre, france; m. a. aziz-alaoui, c. bertelle, x. z. liu, d. olivier, eds.: pp. 51-60.*

Raimbault J., Bergeaud A. (2017, juin). The cost of transportation : Spatial analysis of us fuel prices. *ArXiv e-prints.*

Reuillon R., Leclaire M., Rey-Coyrehourcq S. (2013). Openmole, a workflow engine specifically tailored for the distributed exploration of simulation models. *Future Generation Computer Systems*, vol. 29, n° 8, p. 1981–1990.

SDRIF. (2013). *Île-de-france 2030. orientations réglementaires et carte de destination générale des différentes parties du territoire.*

STIF. (2010). *Arcexpress, débat public sur le métro de rocade. dossier du maitre d'ouvrage.*
archived at http://archive.wikiwix.com/cache/?url=http%3A%2F%2Fwww.debatpublic-arcexpress.org%2F_script%2Fntsp-document-file_download.php%3Fdocument_id%3D92%26document_file_id%3D106.

Xie F., Levinson D. (2009). How streetcars shaped suburbanization: a granger causality analysis of land use and transit in the twin cities. *Journal of Economic Geography*, p. lbp031.