

Identification de causalités dans des données spatio-temporelles

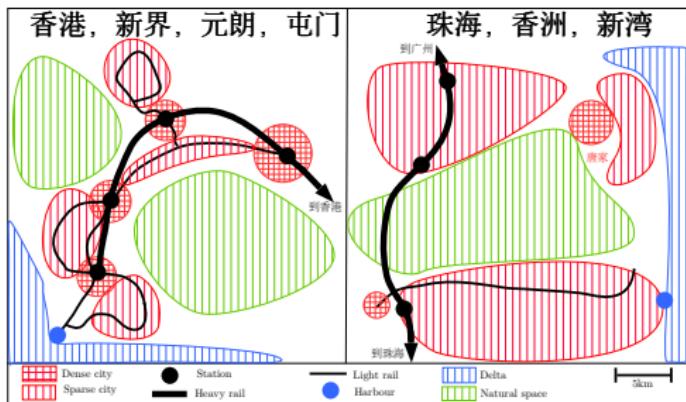
J. Raimbault^{1,2}
juste.raimbault@parisgeo.cnrs.fr

¹UMR CNRS 8504 Géographie-cités

²UMR-T IFSTTAR 9403 LVMT

Séminaire LAET
9 janvier 2018

Interactions entre réseaux et territoires



Un travail de terrain dans le Delta de la Rivière des Perles montre des manifestations locales des interactions entre réseaux de transport et territoires

Introduction

Méthodes et Résultats

Discussion

Effet structurants des infrastructures

De [Bonnafous and Plassard, 1974] à [Offner, 1993] : quels effets de structure des infrastructures de transport sur les territoires ?

- Existence de processus co-évolutifs [Bretagnolle, 2009]
- A petite échelle et sur le temps long, existence de dynamiques structurelles des systèmes urbains [Offner et al., 2014]
- La question des causalités circulaires revient à toutes les échelles (e.g. échelle métropolitaine et mobilité [Cerqueira, 2017]) et dans différents domaines (retombées locales et innovation [Audretsch and Feldman, 1996])

Introduction

Méthodes et Résultats

Discussion

Causalité en Géographie

- La Géographie classique s'intéressait déjà à des liens causaux dans l'espace [Loi, 1985]
- [Claval, 1985] : au delà de la causalité réductionniste en analyse systémique
- La systémogenèse introduite par [Durand-Dastes, 2003] se concentre sur la dynamique et les dépendances au chemin
- Vers une approche complexe de la causalité ? [Morin, 1976]

Introduction

Méthodes et Résultats

Discussion

Approches empiriques existantes

Réseaux de transports et territoires

- Corrélations retardées : [Levinson, 2008] Population et connectivité au réseau à Londres ; [Gargi Chaudhuri and Keith C Clarke, 2015] données historiques en Italie du Nord
- Variables instrumentales : [Duranton and Turner, 2012] Réseau routier et emplois aux Etats-Unis ; [Berger and Enflo, 2017] effet significatif du réseau ferré suédois sur les trajectoires urbaines

Corrélations spatio-temporelles

- Méthode de correspondance pour les flux de trafic [Liu et al., 2011]
- Causalité de Granger généralisée en neurosciences [Ke et al., 2007]
- Corrélations spatio-temporelles en Vision par Ordinateur [Ke et al., 2007]

Introduction

Méthodes et Résultats

Discussion

Contexte

- Généricité et opérationnalité des approches existantes ?
- Capturer la complexité simplement ?
- A l'interface des domaines de connaissance (méthodologie, modélisation et empirique) [Raimbault, 2017]

Objectif de recherche :

Exploration d'une méthode générique basée sur les motifs des corrélations spatio-temporelles retardées : notion de causalité de Granger ; validation sur données synthétiques.

Introduction

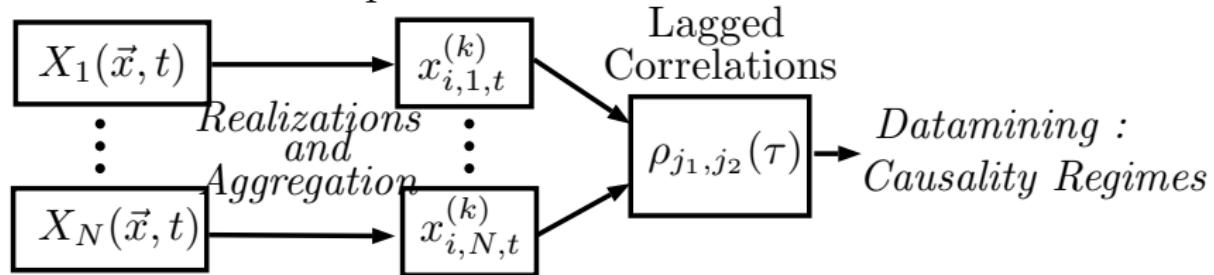
Méthodes et Résultats

Discussion

Aperçu de la méthode

Spatio-temporal
stochastic field

Trajectories of
spatial units



Formalisation de la méthode

Estimateur de corrélation $\hat{\rho}$ s'appliquant dans le temps, l'espace, et les répétitions, i.e. la covariance est estimée par

$$\hat{\text{Cov}}[X, Y] = \hat{\mathbb{E}}_{i,t,k}[XY] - \hat{\mathbb{E}}_{i,t,k}[X]\hat{\mathbb{E}}_{i,t,k}[Y]$$

Corrélations retardées définies comme

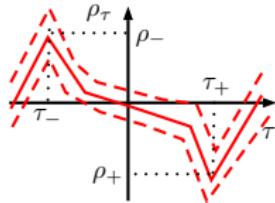
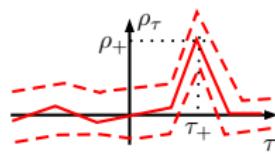
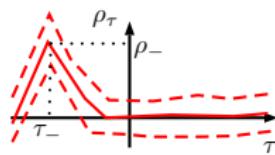
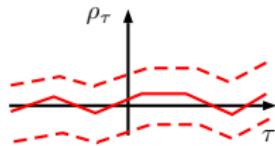
$$\rho_\tau[X_{j_1}, X_{j_2}] = \hat{\rho}\left[x_{i,j_1,t-\tau}^{(k)}, x_{i,j_2,t}^{(k)}\right] \quad (1)$$

Le motifs de $\text{argmax}_\tau \rho_\tau[X_{j_1}, X_{j_2}]$ ou de $\text{argmin}_\tau \rho_\tau[X_{j_1}, X_{j_2}]$ (en les supposant clairement définis : e.g. significativité statistique, valeur seuil) capture le sens de la causalité entre j_1 et j_2

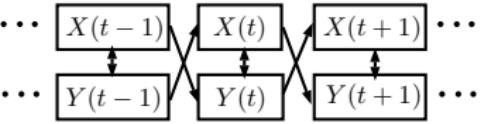
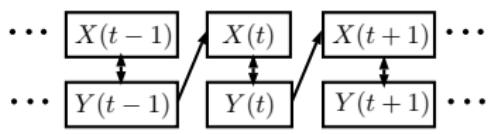
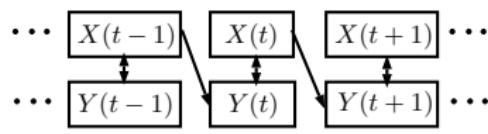
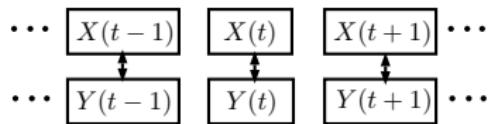
→ Datamining sur ρ_τ (possiblement paramétré comme $\rho_\tau^{(\omega)}$) pour explorer les motifs de causalité.

Illustration pour 2 variables

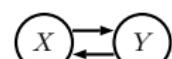
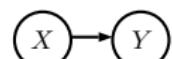
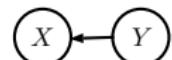
Lagged correlation



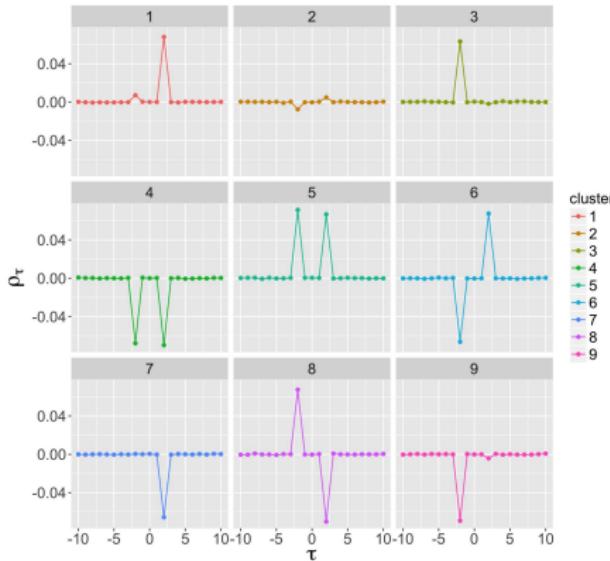
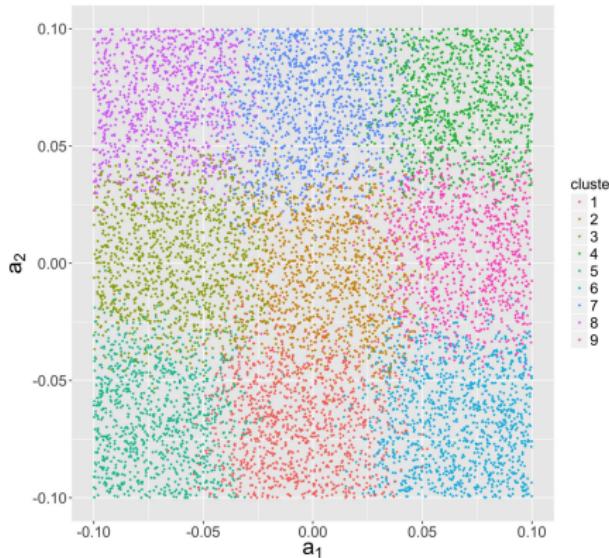
Temporal evolution



Causal graph

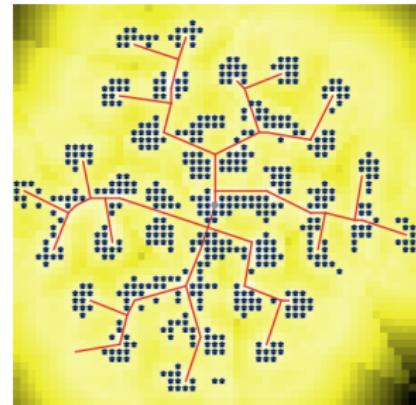
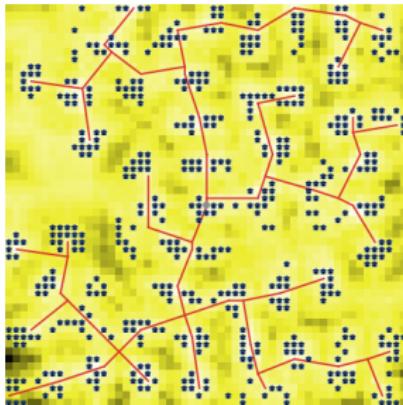
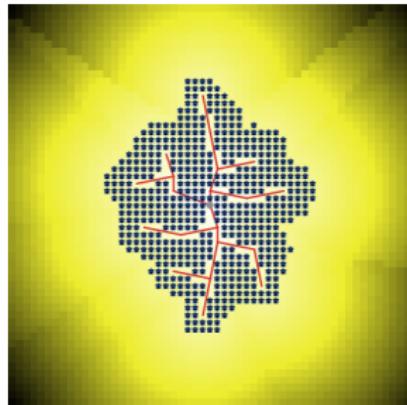


Validation basique



Données synthétiques : processus AR avec retard 2, termes croisés paramétrés par $(a_1, a_2) \in [-0.1, 0.1]$ aléatoires.

Validation sur données synthétiques



Configurations urbaines synthétiques générées par un modèle de morphogenèse hybride issu de [Raimbault et al., 2014]

Description du modèle de morphogenèse

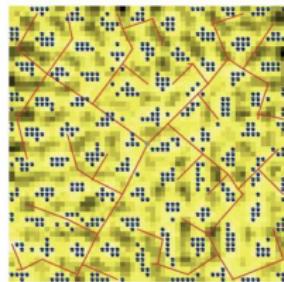
- Automate cellulaire: cellules d'une grille carrée $(L_{i,j})_{1 \leq i,j \leq N}$, occupées ou non (fonction $\delta(i,j,t) \in \{0,1\}$)
- Réseau vectoriel $G(t) = (V(t), E(t))$ qui évolue, incluant des centres urbains fixes $C_0 \subset V(0)$ auxquels des activités $a \in \{1, \dots, a_{max}\}$ sont attribuées (propriétés fonctionnelles de l'environnement urbain).
- Variables explicatives $(d_k)_{1 \leq k \leq K}$ définies sur les cellules, avec des poids associés $(\alpha_k)_{1 \leq k \leq K}$ (paramètres principaux du modèle), qui sont:
 - d_1 densité autour de la cellule (dans un rayon fixé r)
 - d_2 distance à la route la plus proche
 - d_3 distance au centre le plus proche par le réseau
 - $d_4(i,j,t) = \left(\frac{1}{a_{max}} \sum_{a=1}^{a_{max}} d_3(i,j,t; a)^{p_4} \right)^{1/p_4}$: accessibilité intégrée aux activités

Règles d'évolution

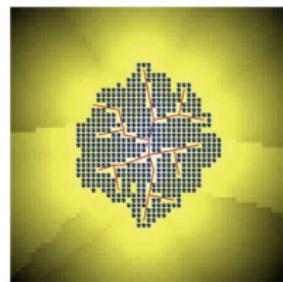
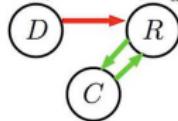
A chaque pas de temps :

- Etalement de la surface occupée : les meilleures N cellules selon la valeur du potentiel $v(i,j,t) = \frac{1}{\sum_k \alpha_k} \sum_{k=1}^K \alpha_k \frac{d_{k,\max}(t) - d_k(i,j,t)}{d_{k,\max}(t) - d_{k,\min}(t)}$ sont construites.
- Adaptation du réseau : quand une nouvelle cellule est construite, si $d_2 > \theta_2$, la cellule est connectée au réseau par une route directe.

Profils de corrélations retardées



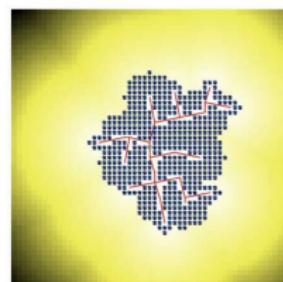
$$w_r = 0, w_c = 0, w_d = 1$$



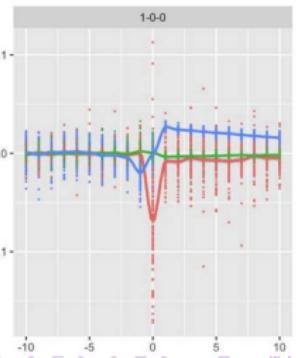
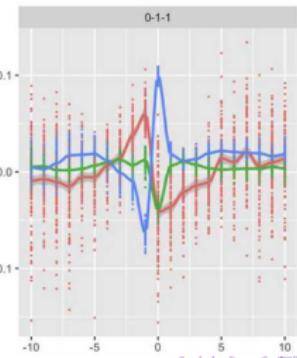
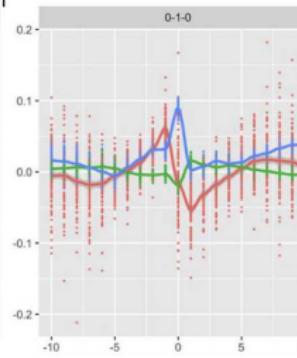
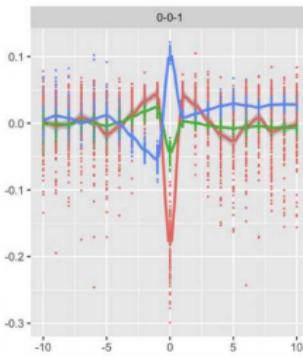
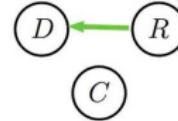
$$w_r = 0, w_c = 1, w_d = 0$$



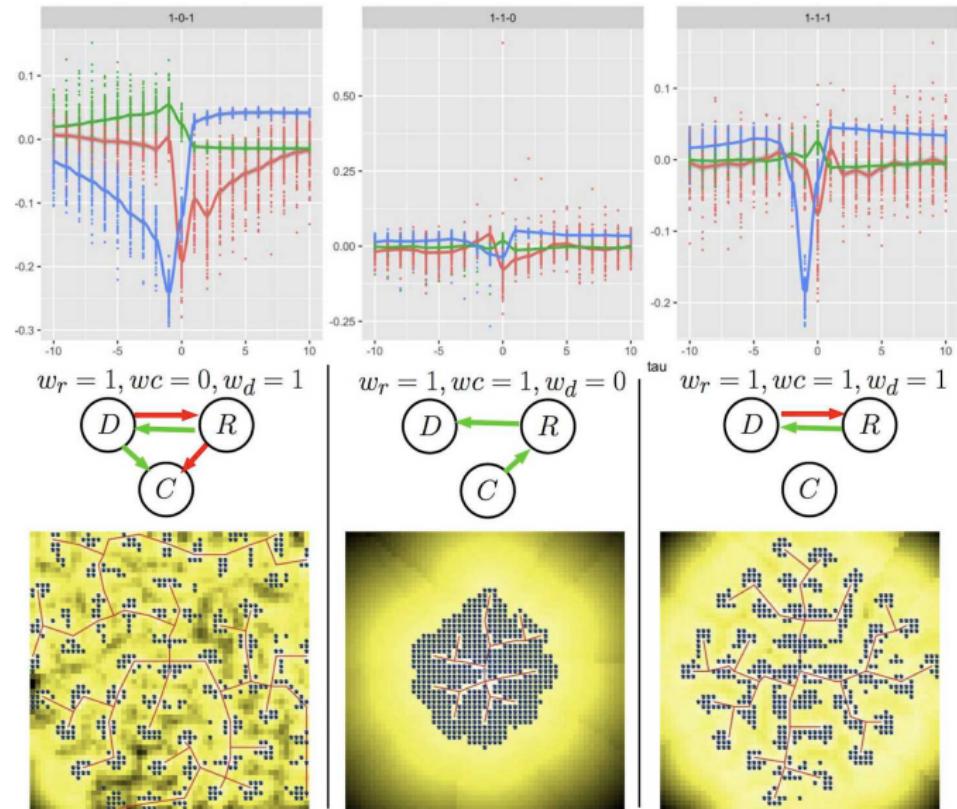
$$w_r = 0, w_c = 1, w_d = 1$$



$$w_r = 1, w_c = 0, w_d = 0$$



Profils de corrélations retardées

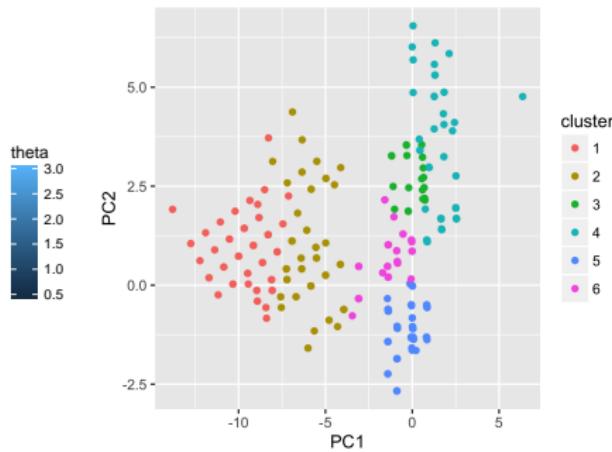
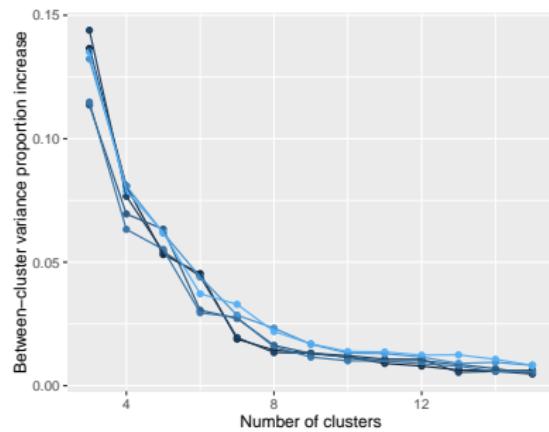


Paramètres

$N_C = 1$
 $\theta_d = 5$
 $r = 5$
 $N_G = 10$
 $t_f = 10$

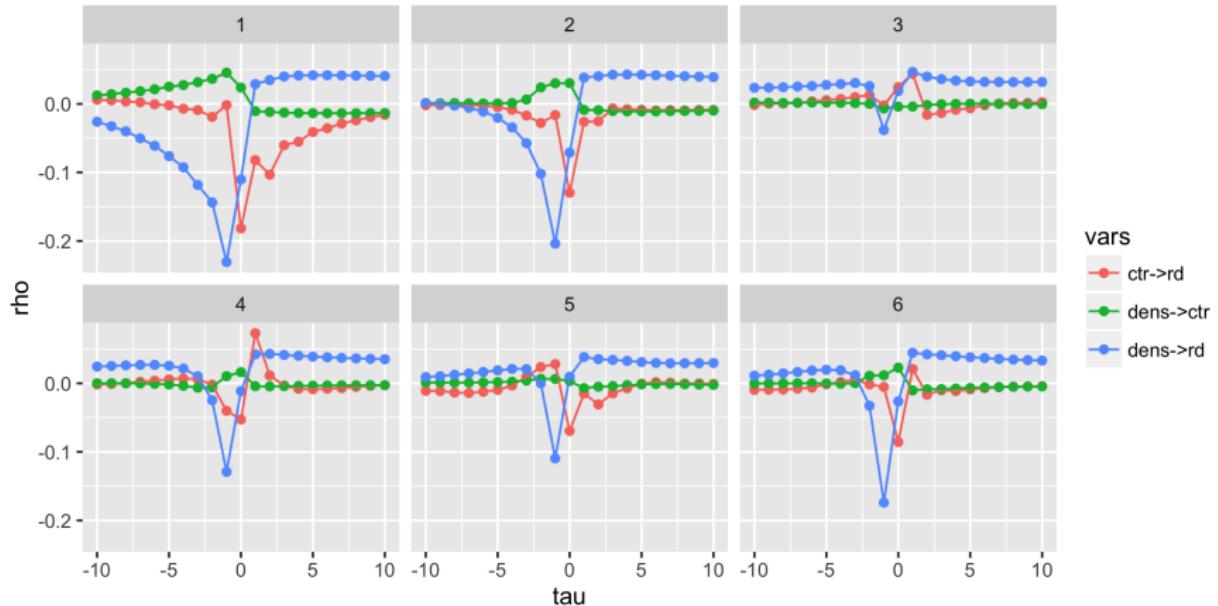
Régimes endogènes de causalité

Exploration intensive de l'espace des paramètres du modèle (1000 points de paramètres x 100 répétitions) avec le logiciel OpenMole
[Reuillon et al., 2013]



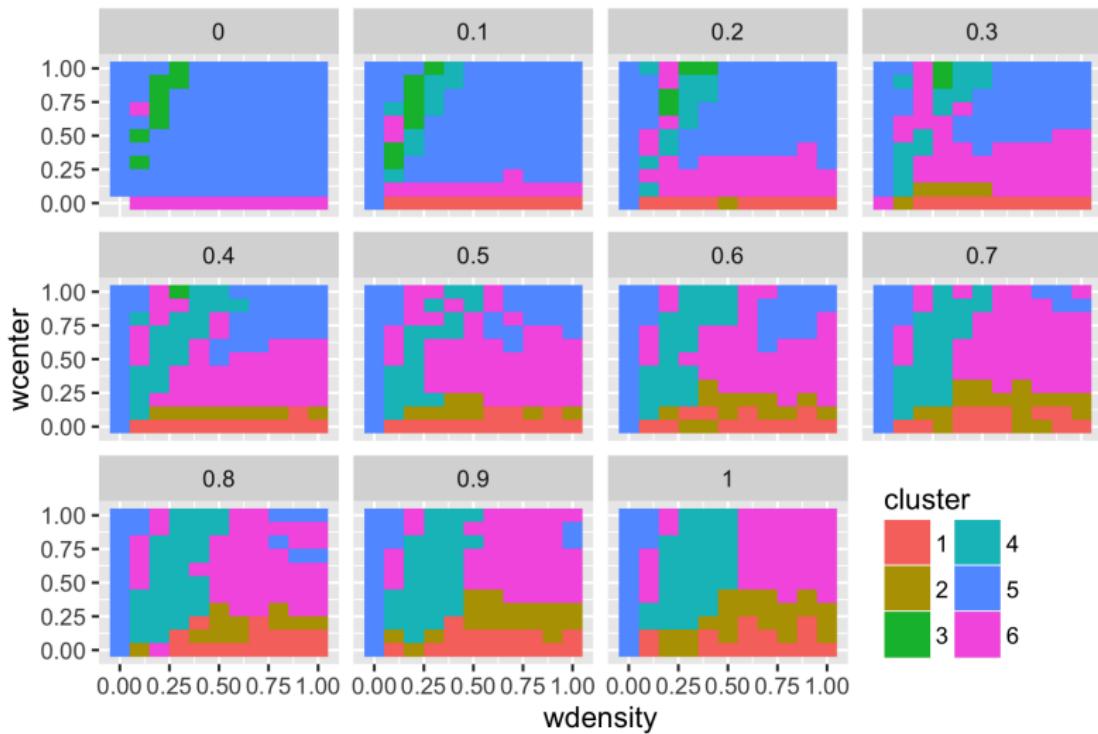
*Classification non-supervisée (k-means robustes) sur les caractéristiques τ_{min}, τ_{max} :
(Gauche) Dérivée du coefficient de clustering en fonction du nombre de clusters k;
(Droite) Visualisation en plan principal pour k = 6.*

Composition des régimes



Valeurs des centres des clusters en termes de ρ_τ

Interprétation des régimes

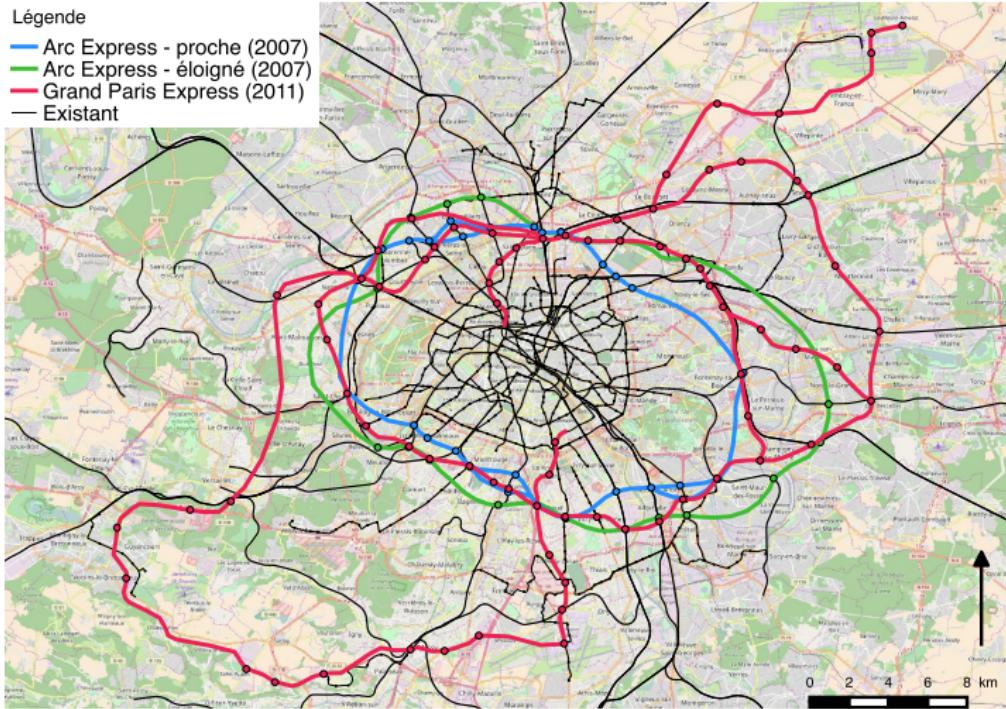


Position des clusters dans l'espace des paramètres w_i

Application: Cas d'étude

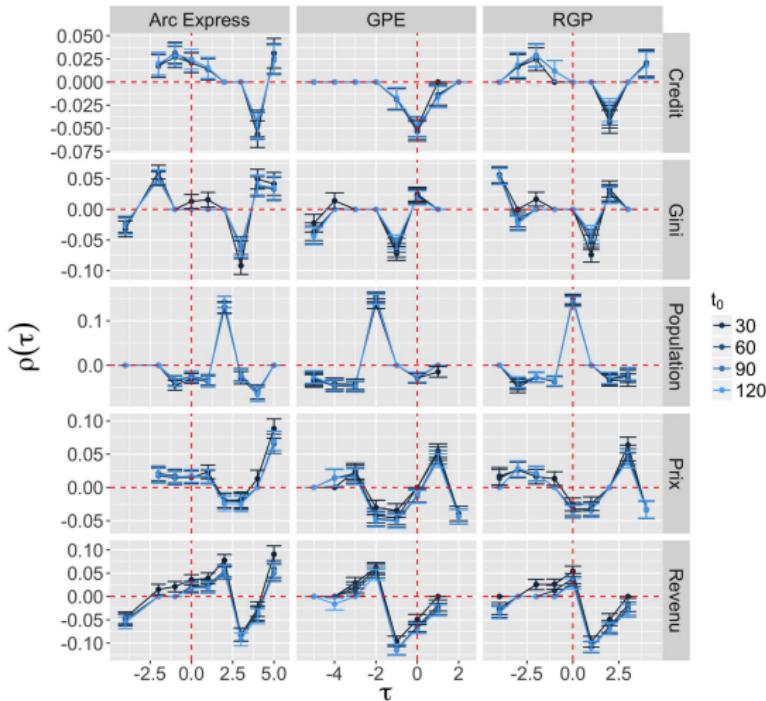
Légende

- Arc Express - proche (2007)
- Arc Express - éloigné (2007)
- Grand Paris Express (2011)
- Existant



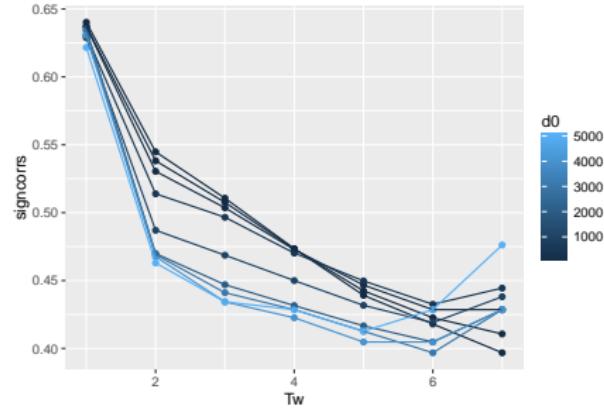
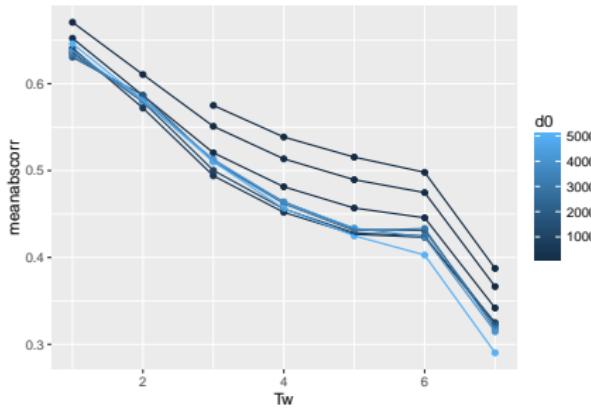
Projet de transport successifs pour la nouvelle infrastructure de transport du Grand Paris

Application: Résultats



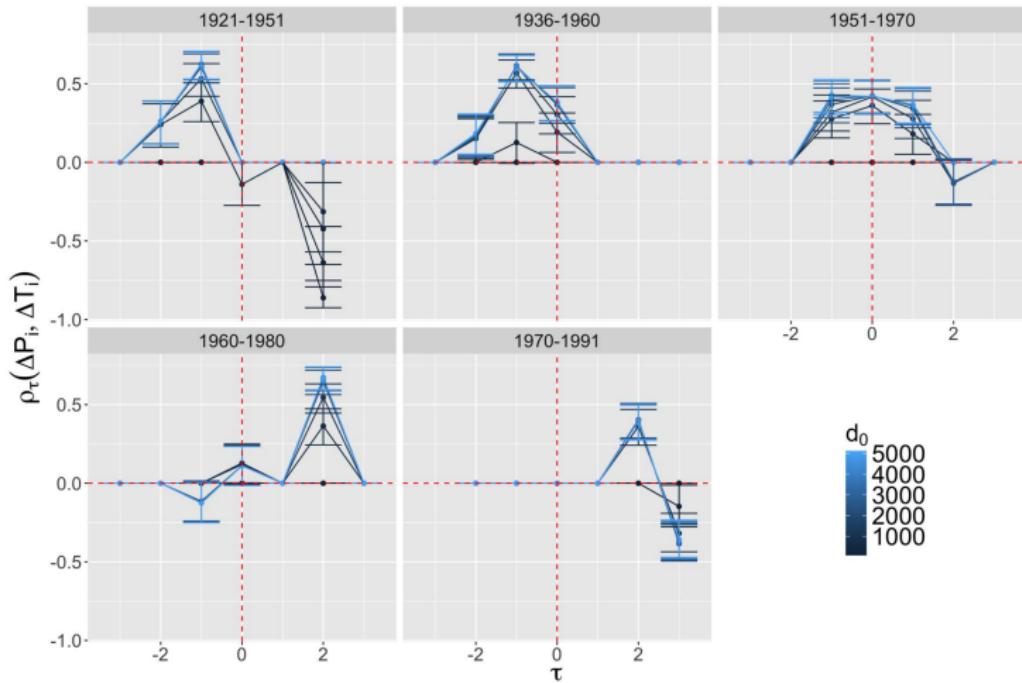
Valeurs de ρ_τ pour les différents projets (colonnes) et les différentes variables (lignes), avec les différentiels d'accessibilité

Application à l'Afrique du Sud



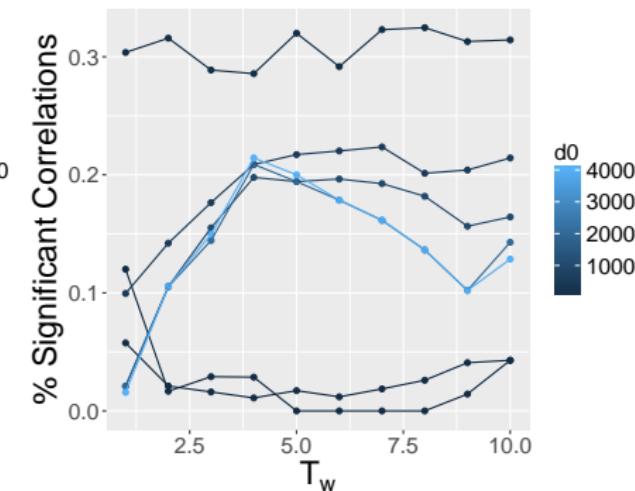
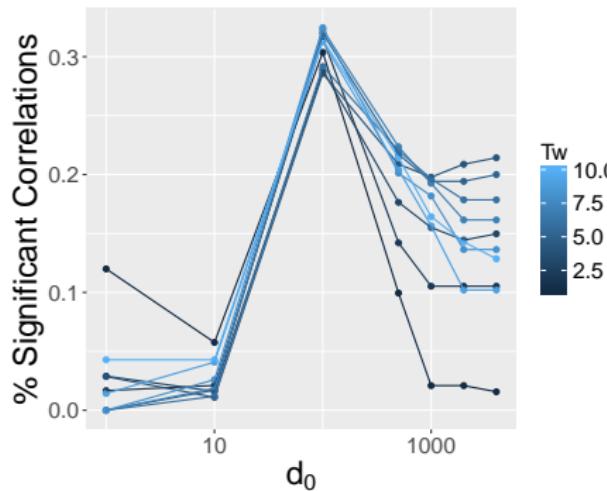
Détermination de la fenêtre temporelle et de la portée spatiale de l'accessibilité

Application à l'Afrique du Sud



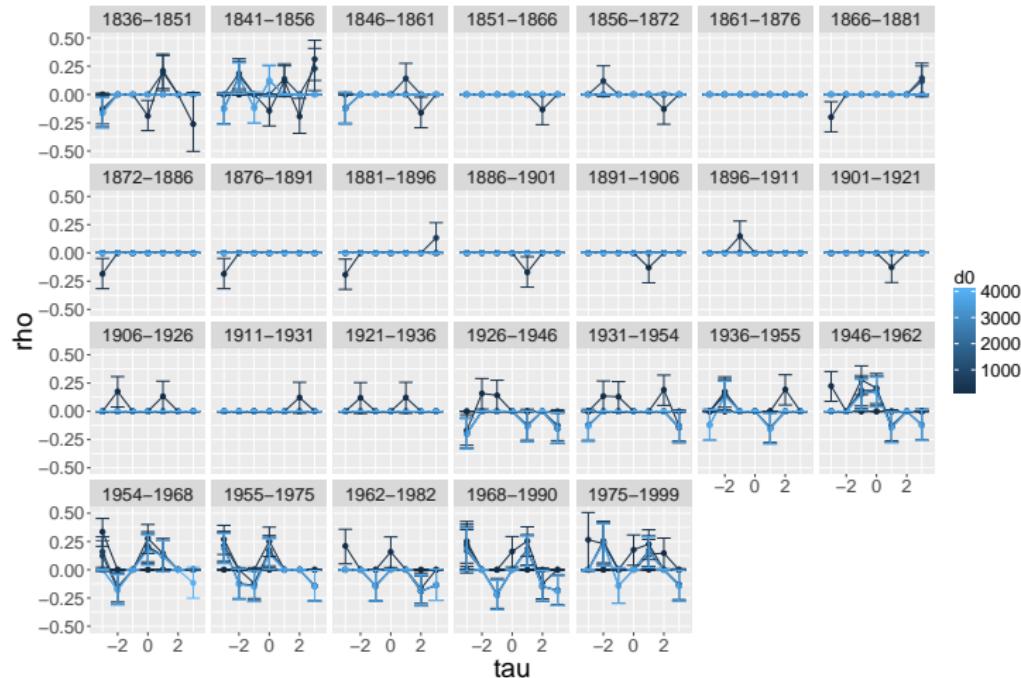
Inversion du sens de la causalité suggère un effet de ségrégation structurelle des politiques d'apartheid

Application à la France



Fenêtre temporelle et portée spatiale optimales (réseau ferré et population sur la période 1830-1999)

Application à la France



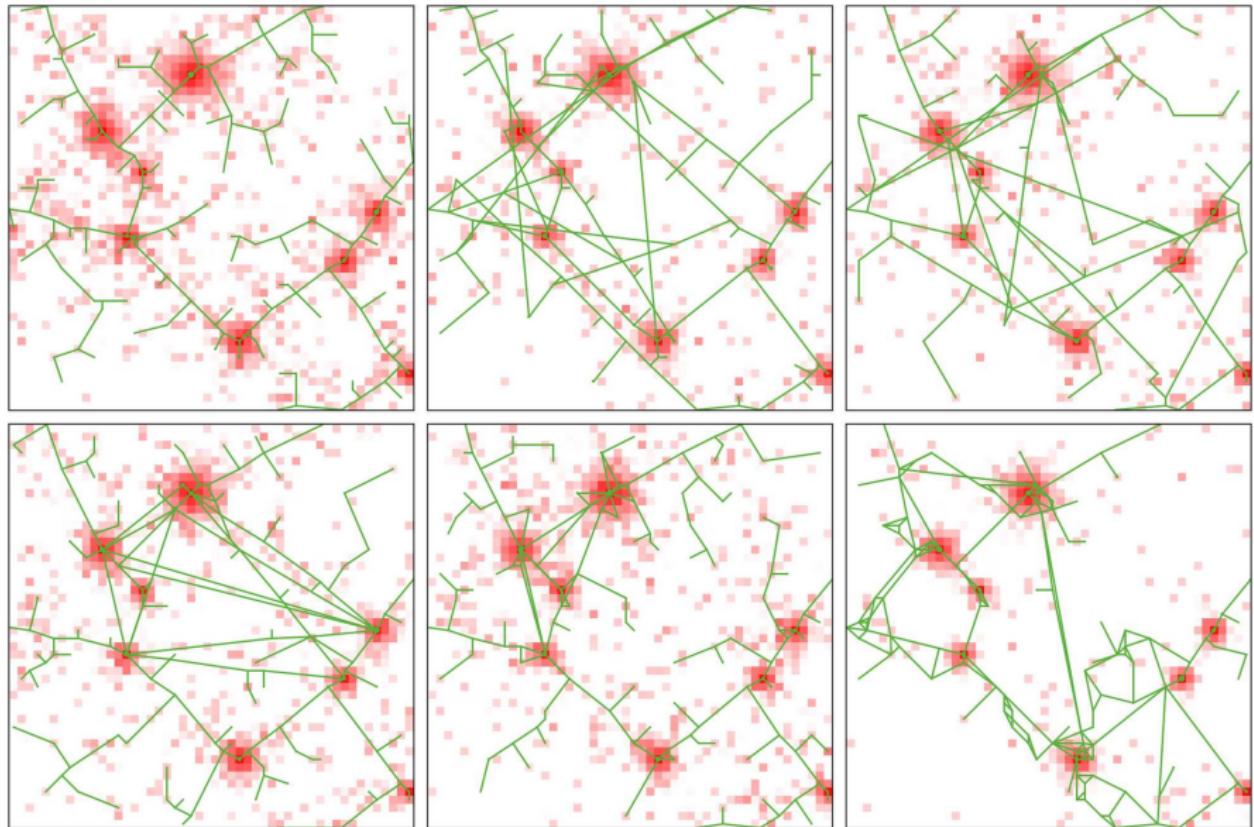
Profils de corrélation : pas de signal significatif

Perspective : co-évolution

Proposition d'une définition de la co-évolution, basée sur une revue multi-disciplinaire :

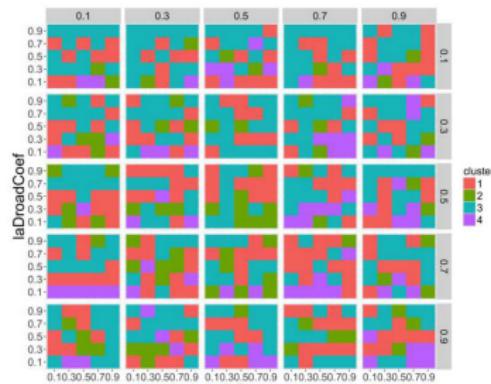
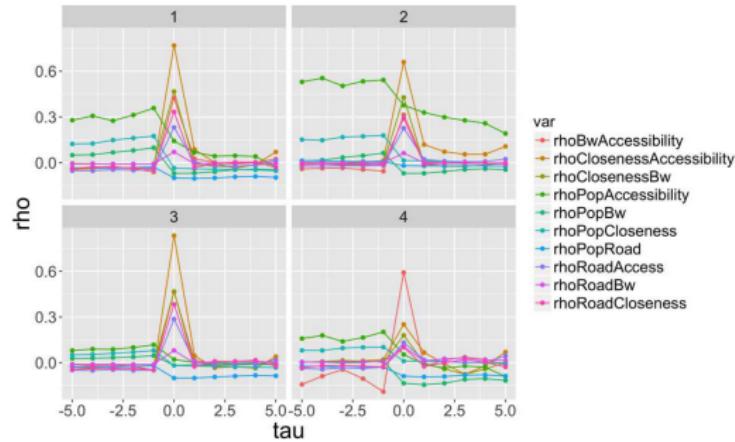
- ① Existence de processus évolutifs : transformations des composantes du système territorial aux différentes échelles
- ② Trois manifestations de la co-évolution à des niveaux emboités :
 - Entités en relations causales circulaires
 - Population d'entités dans une région géographique, identifiable en pratique par les régimes de causalité
 - Niveau global du système, interdépendance forte
- ③ Existence de sous-systèmes en relative isolation spatio-temporelle où s'opèrent différentes co-évolutions (lien avec le concept de morphogenèse)

Modélisation à l'échelle mesoscopique



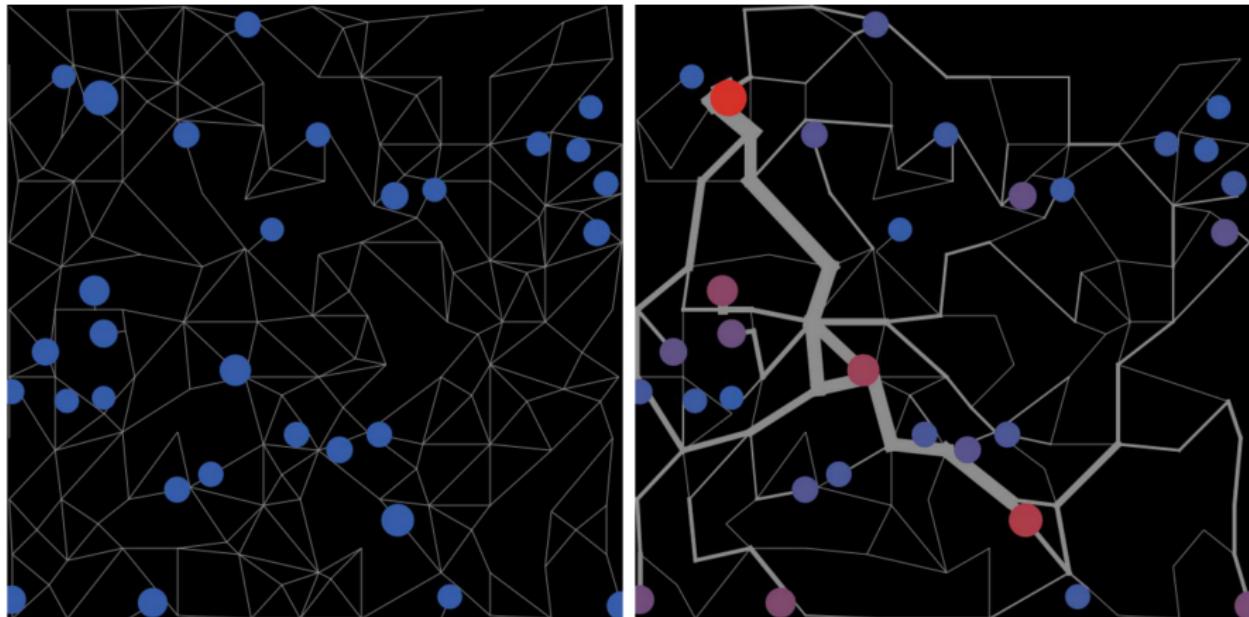
Régimes de causalité

Multi-modélisation pour la croissance du réseau couplée au modèle RBD



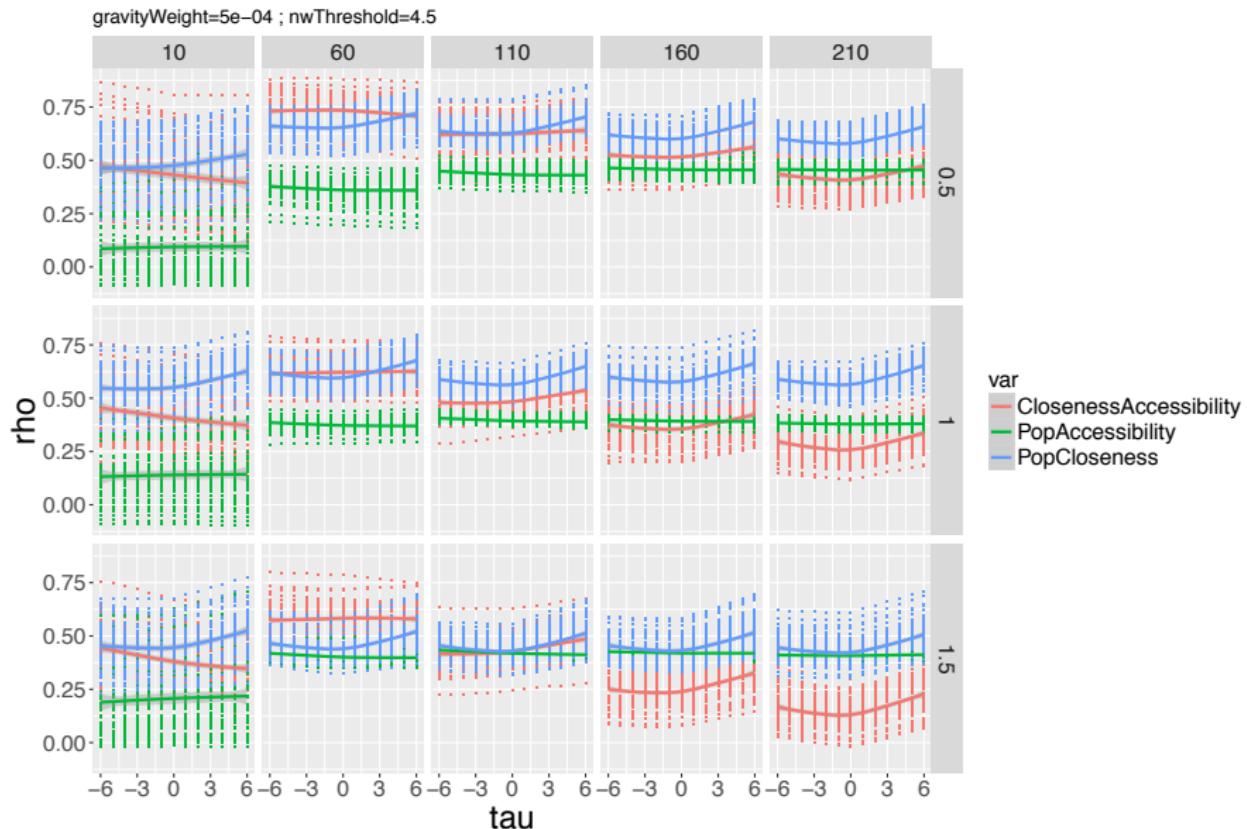
(Gauche) Profils de corrélations retardées ; (Droite) Distribution des régimes dans l'espace des paramètres

Co-évolution à l'échelle macroscopique



Modèle d'interaction entre villes [Pumain and Reuillon, 2017] couplé à un modèle d'auto-renforcement de réseau [Tero et al., 2010]

Co-évolution macroscopique : motifs de corrélation



Discussion

Implications

- Motifs de corrélations retardées pour identifier des “effets structurants” dans des systèmes complexes
- Le concept opérationnel de *Régime de causalité* introduit une nouvelle façon de comprendre la co-évolution dans les modèles de simulation

Développements

- Caractérisation de la diffusion spatio-temporelle : test de l'hypothèse de la diffusion spatiale de l'innovation dans la Théorie Evolutive des Villes [Pumain, 2010]
- Echelles optimales pour la stationnarité : lien avec GWR [Brunsdon et al., 1998]

Conclusion

- Une méthode validée sur données synthétiques et montrée opérationnelle sur des systèmes réels
 - A l'interface des domaines de connaissance : théorie, modélisation, empirique, méthodologique
 - A l'interface des disciplines : analyse spatiale, statistiques, datamining
-
- Code, données et résultats disponibles à
<https://github.com/JusteRaimbault/CityNetwork>
 - Article sur arXiv à <https://arxiv.org/abs/1709.08684>
 - Remerciements : Nous remercions l'*European Grid Infrastructure* et *France-Grilles* en particulier pour le support technique et l'infrastructure.

Reserve slides

Reserve Slides

Granger causality

Granger causality test based on VAR processes :

$$X(t) = \sum_{0 \leq \tau \leq \tau_Y} b_\tau Y(t - \tau)$$

If there exists b_τ such that $|b_\tau| > 0$ significantly, then Y Granger-causes X .

We have then $\rho_\tau(Y, X) > 0$.

Morphogenesis

Morphogenesis (*Oxford dictionary*)

- ① *Biology* : The origin and development of morphological characteristics
- ② *Geology* : The formation of landforms or other structures.

History of the notion

- Started significantly with embryology around 1930 [Abercrombie, 1977]
- Turing's 1952 paper [Turing, 1952], linked to the development of Cybernetics
- first use in 1871, large peak in usage between 1907-1909, increase until 1990, decrease until today. *Scientific fashion* ?

Defining Morphogenesis

Meta-epistemological framework of imbricated notions:
Self-organization \supseteq Morphogenesis \supseteq Autopoiesis \supseteq Life

Properties:

- Architecture links form and function
- Emergence strength [Bedau, 2002] increases with notion depth, as bifurcations [Thom, 1974]

Definition of Morphogenesis : *Emergence of the form and the function in a strongly coupled manner, producing an emergent architecture*
[Doursat et al., 2012]

References I

-  Abercrombie, M. (1977).
Concepts in morphogenesis.
Proceedings of the Royal Society of London B: Biological Sciences,
199(1136):337–344.
-  Audretsch, D. B. and Feldman, M. P. (1996).
R&d spillovers and the geography of innovation and production.
The American economic review, 86(3):630–640.
-  Bedau, M. (2002).
Downward causation and the autonomy of weak emergence.
Principia: an international journal of epistemology, 6(1):5–50.

References II

-  Berger, T. and Enflo, K. (2017).
Locomotives of local growth: The short-and long-term impact of railroads in sweden.
Journal of Urban Economics, 98:124–138.
-  Bonnafous, A. and Plassard, F. (1974).
Les méthodologies usuelles de l'étude des effets structurants de l'offre de transport.
Revue économique, pages 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.

References III

-  Brunsdon, C., Fotheringham, S., and Charlton, M. (1998). Geographically weighted regression. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 47(3):431–443.
-  Cerqueira, E. V. (2017). Les inégalités d'accès aux ressources urbaines dans les franges périphériques de belo horizonte (brésil): quelles évolutions? *EchoGéo*, (39).
-  Claval, P. (1985). Causalité et géographie. *Espace géographique*, 14(2):109–115.

References IV

-  Doursat, R., Sayama, H., and Michel, O. (2012).
Morphogenetic engineering: toward programmable complex systems.
Springer.
-  Durand-Dastes, F. (2003).
Les géographes et la notion de causalité.
-  Duranton, G. and Turner, M. A. (2012).
Urban growth and transportation.
The Review of Economic Studies, 79(4):1407–1440.
-  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, 42(1):133–156.

References V

-  Ke, Y., Sukthankar, R., and Hebert, M. (2007). Spatio-temporal shape and flow correlation for action recognition. In *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*, pages 1–8. IEEE.
-  Levinson, D. (2008). Density and dispersion: the co-development of land use and rail in london. *Journal of Economic Geography*, 8(1):55–77.
-  Liu, W., Zheng, Y., Chawla, S., Yuan, J., and 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*, pages 1010–1018. ACM.

References VI

-  Loi, D. (1985).
Une étude de la causalité dans la géographie classique française.[l'exemple des premières thèses régionales].
Espace géographique, 14(2):121–125.
-  Morin, E. (1976).
La méthode: la nature de la nature.
Le Seuil.
-  Offner, J.-M. (1993).
Les "effets structurants" du transport: mythe politique, mystification scientifique.
Espace géographique, 22(3):233–242.

References VII

-  Offner, J.-M., Beaucire, F., Delaplace, M., Frémont, A., Ninot, O., Bretagnolle, A., and Pumain, D. (2014).
Les effets structurants des infrastructures de transport.
Espace Géographique, (42):p–51.
-  Pumain, D. (2010).
Une théorie géographique des villes.
Bulletin de la Société géographie de Liège, (55):5–15.
-  Pumain, D. and Reuillon, R. (2017).
Urban Dynamics and Simulation Models.
Springer International, ISBN 978-3-319-46497-8.
-  Raimbault, J. (2017).
An applied knowledge framework to study complex systems.
ArXiv e-prints.

References VIII

-  Rimbault, J., Banos, A., and 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.
-  Reuillon, R., Leclaire, M., and Rey-Coyrehourcq, S. (2013).
Openmole, a workflow engine specifically tailored for the distributed exploration of simulation models.
Future Generation Computer Systems, 29(8):1981–1990.

References IX

-  Tero, A., Takagi, S., Saigusa, T., Ito, K., Bebber, D. P., Fricker, M. D., Yumiki, K., Kobayashi, R., and Nakagaki, T. (2010). Rules for biologically inspired adaptive network design. *Science*, 327(5964):439–442.
-  Thom, R. (1974). Stabilité structurelle et morphogénèse. *Poetics*, 3(2):7–19.
-  Turing, A. M. (1952). The chemical basis of morphogenesis. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 237(641):37–72.