Thesis Progress Meeting

J. Raimbault[1,2]

[1]Géographie-cités (UMR 8504 CNRS)
[2]LVMT (UMR-T 9403 IFSTTAR)

January 26th 2016

Achieved Work (by projects)

- Biblio/Meetings/Organisation [0.7w]
- Conference [0.7w]
- Reading Records (*Synergetics*) [0.2w]
- Monitorat [1,3w]
- Cybergeo Project [1w]
- Correlated Synthetic data [3w]
- Theory construction [0.2w]
- BP Case Study / Spatial Econometrics [0,3w]

Context

Def. : *Synthetic Data* are output of generative models (and possibly inputs of models using them).

Methodology used in various fields, e.g. therapeutic evaluation [Abadie et al., 20 territorial systems analysis [Moeckel et al., 2003, Pritchard and Miller, 2009], machine learning [Bolón-Canedo et al., 2013] or bio-informatics [Van den Bulcke

Few examples at the second order : specific examples as [Ye, 2011] for discrete choices ; methods that can be interpreted this way : generation of complex networks [Newman, 2003].

Generic Method

$\vec{X_l}$ multidimensional stochastic process, $\mathbf{X} = (X_{i,j})$ realizations.

Aim : Generate a statistical population $\tilde{\mathbf{X}} = \tilde{X}_{i,j}$ such that:

1. proximity to data : given a precision ε and an indicator f, $\|f(\mathbf{X}) - f(\tilde{\mathbf{X}})\| < \varepsilon$
2. control of the estimated correlation structure : $\hat{\text{Var}}\left[(\tilde{X}_i)\right] = \Sigma R$ with R fixed.

Geographical data : Context and Objective

- In geography, generation of synthetic populations for agent-based models [Pritchard and Miller, 2009].
- Generation of spatial synthetic configuration not used (Geo. Weighted Regression [Brunsdon et al., 1998] can be interpreted this way) ; however crucial for abstract models [Schmitt, 2014]
- [Cottineau et al., 2015] recently proposed to estimate the sensitivity of spatial models of simulation to initial configuration (application to Schelling model).
- Case study : city-transportation interactions, complex to understood quantitatively [Offner, 1993, Bretagnolle, 2009] $\rightarrow$ simple model of population density and transportation network morphogenesis.

Modle

Simple coupling between

- Iterative generation of a density grid by preferential attachment/diffusion [Raimbault, 2016] calibrated on morphological objectives on european density grid.

- Heuristic network generation conditional to density :
  - Distribution of a fixed number of centers preferentially following density
  - Deterministic percolation between closest neighbors
  - Breaking of interaction potentials

$$V_{ij}(d) = \left[(1-k_h) + k_h \cdot \left(\frac{P_i P_j}{P^2}\right)^{\gamma}\right] \cdot \exp\left(-\frac{d}{r_g(1+d/d_0)}\right)$$

    for a fixed number of couples N_L such that $V_{ij}(d_N)/V_{ij}(d_{ij})$ is minimal among $K \cdot N_L$ strongest euclidian potentials ($K = 5$ fixed)
  - Planarization

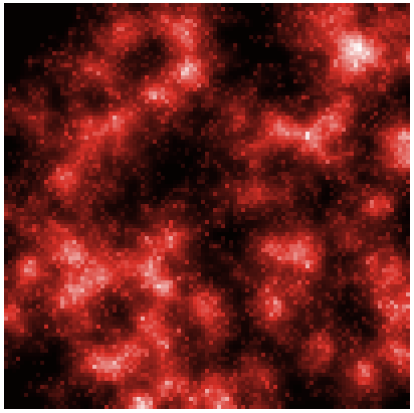
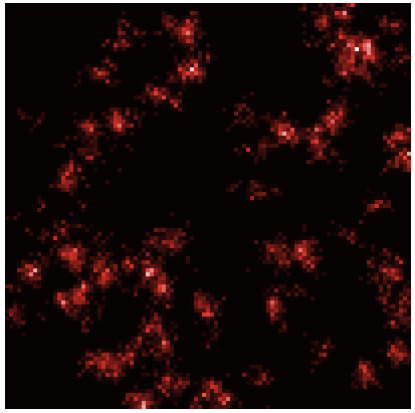
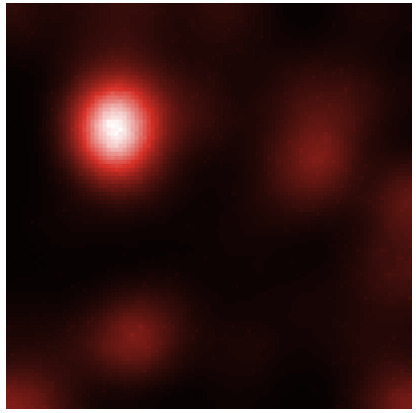
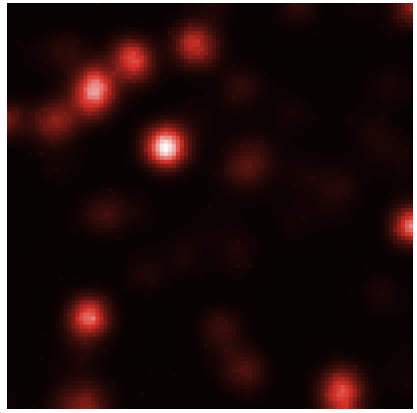
Indicators : morphology [Le Néchet, 2015] (Moran, mean distance, entropy, hierarchy) and network (centrality, mean width, speed, diameter).

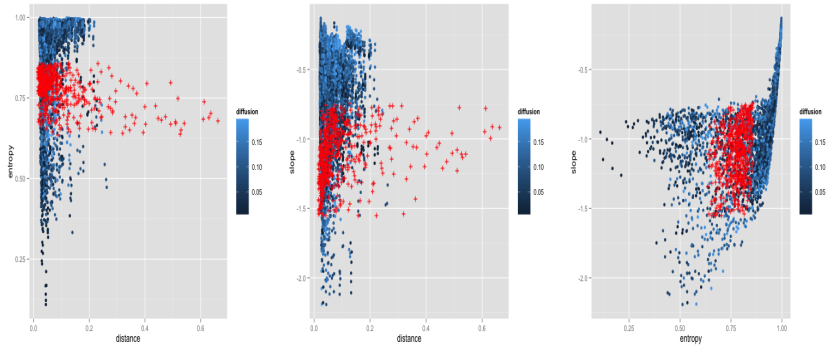
Implementation and Exploration

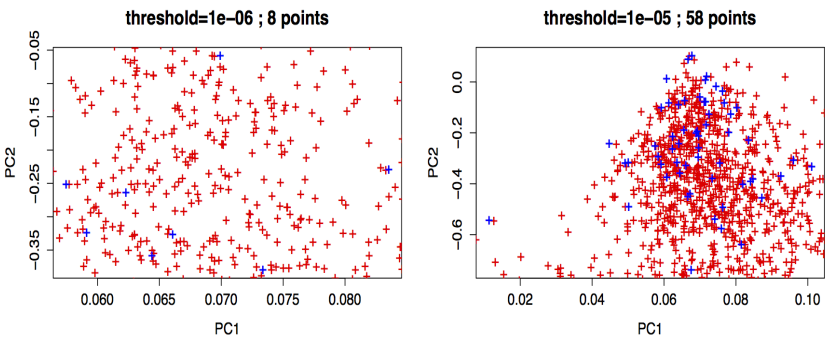
→ Formal and Operational coupling : modular implementation (`scala`/`NetLogo` encapsulated by `OpenMole` [Reuillon et al., 2013]

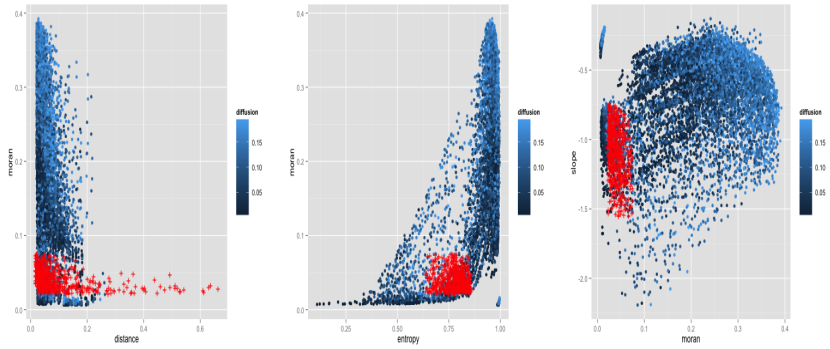
→ Exploration by intensive computation on grid via `OpenMole` : calibration of density model alone ($\sim 1.5 \cdot 10^6$ runs) ; brutal exploration by LHS sampling for feasible correlations ($\sim 5 \cdot 10^4$ runs)

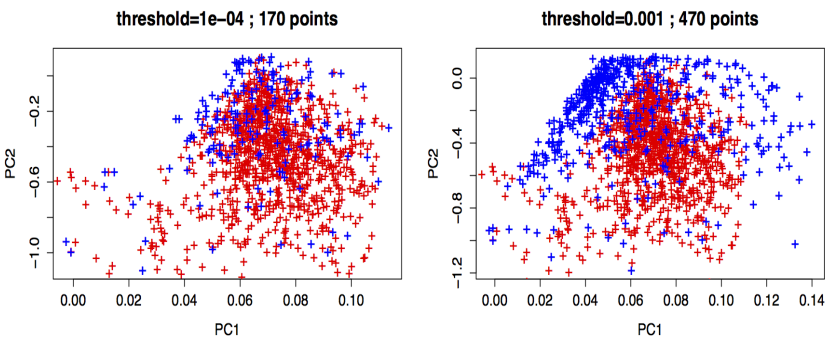
Results : Density Model alone

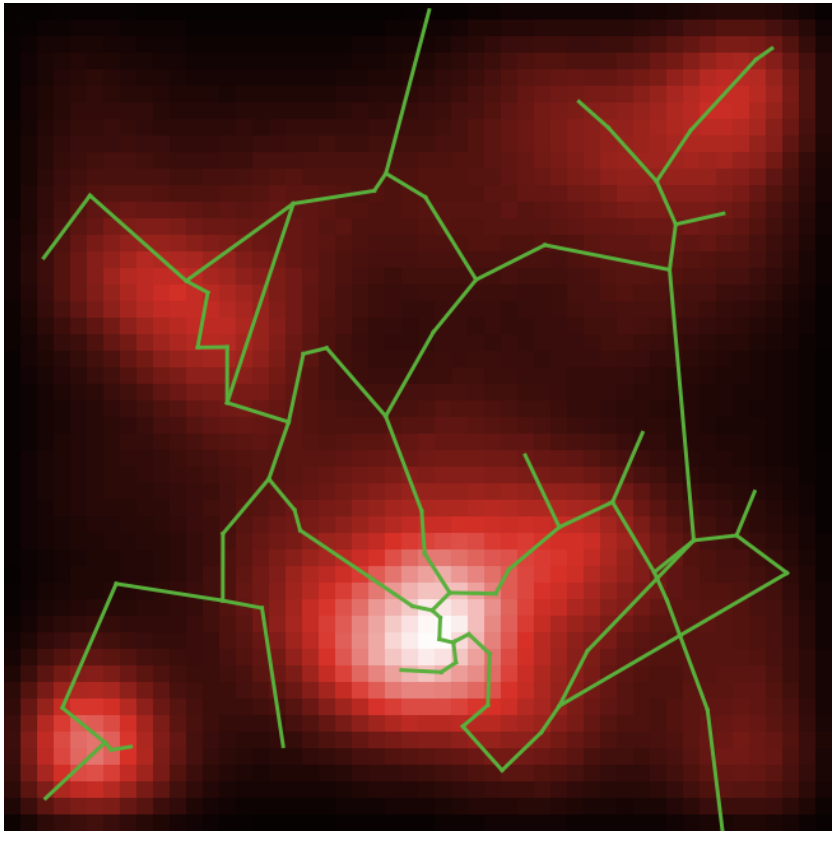


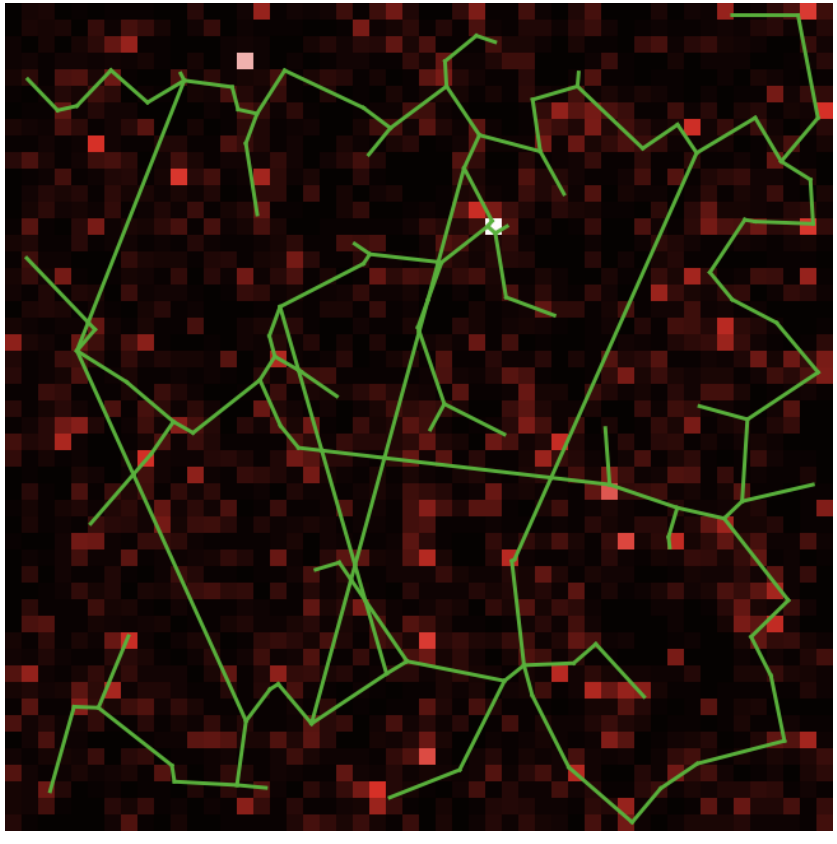




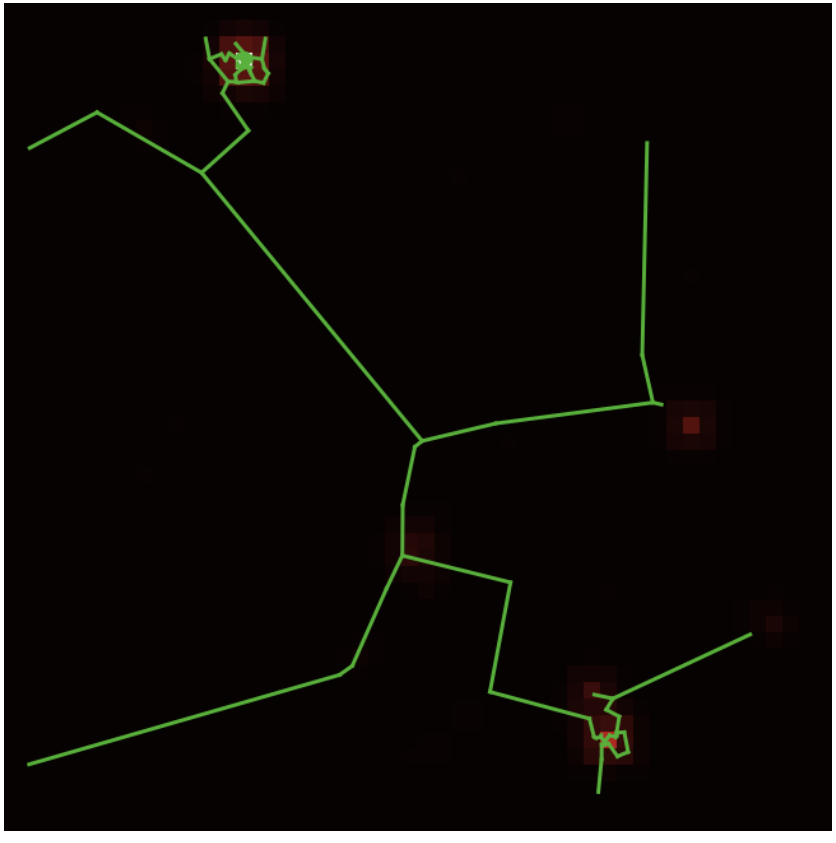


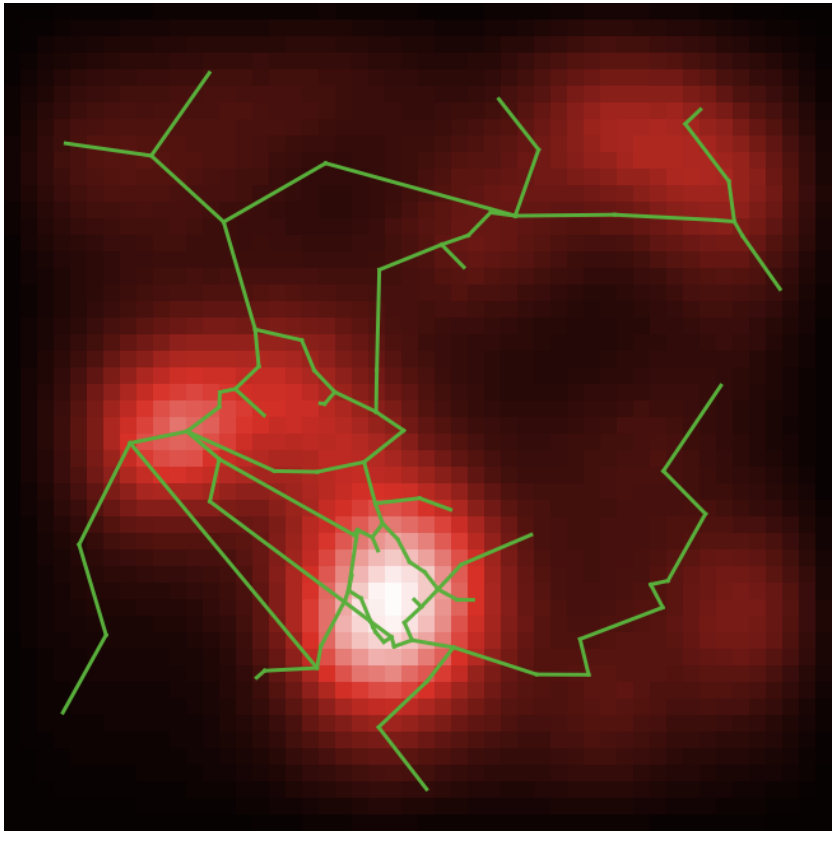
Results : examples of configurations



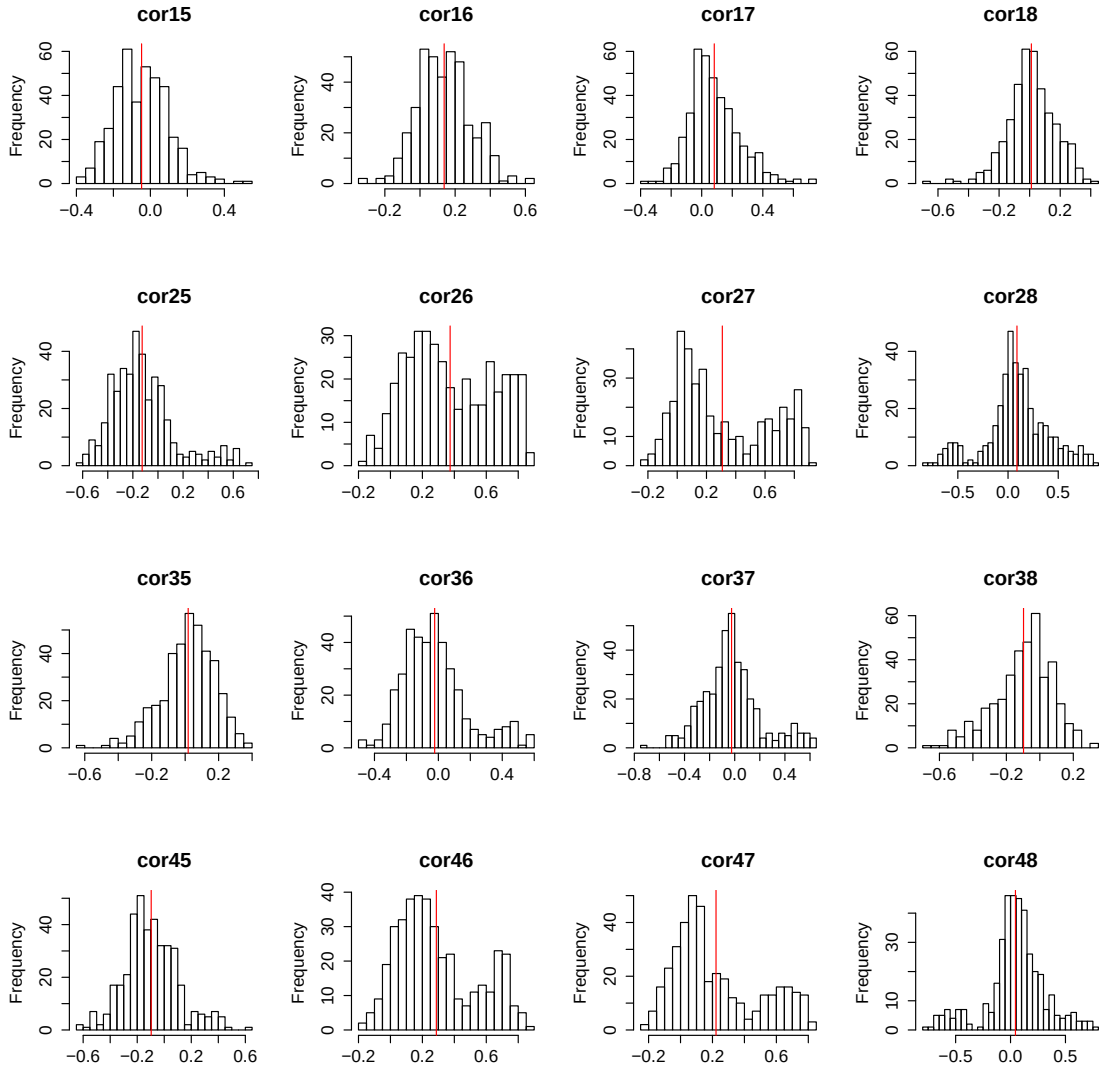


Results : examples of configurations



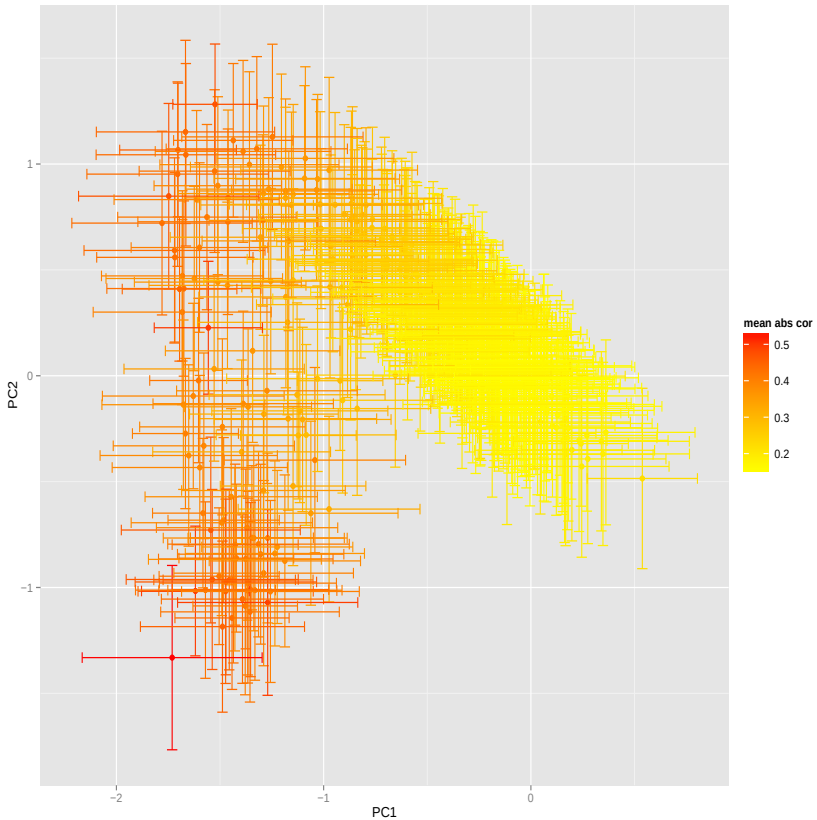


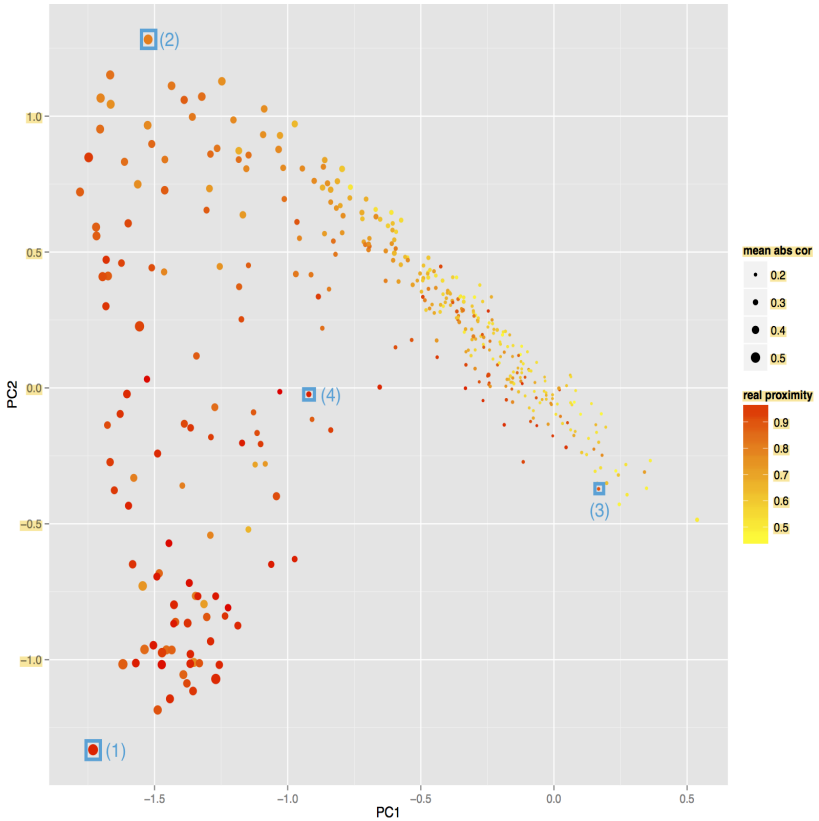
Results : cross-correlations



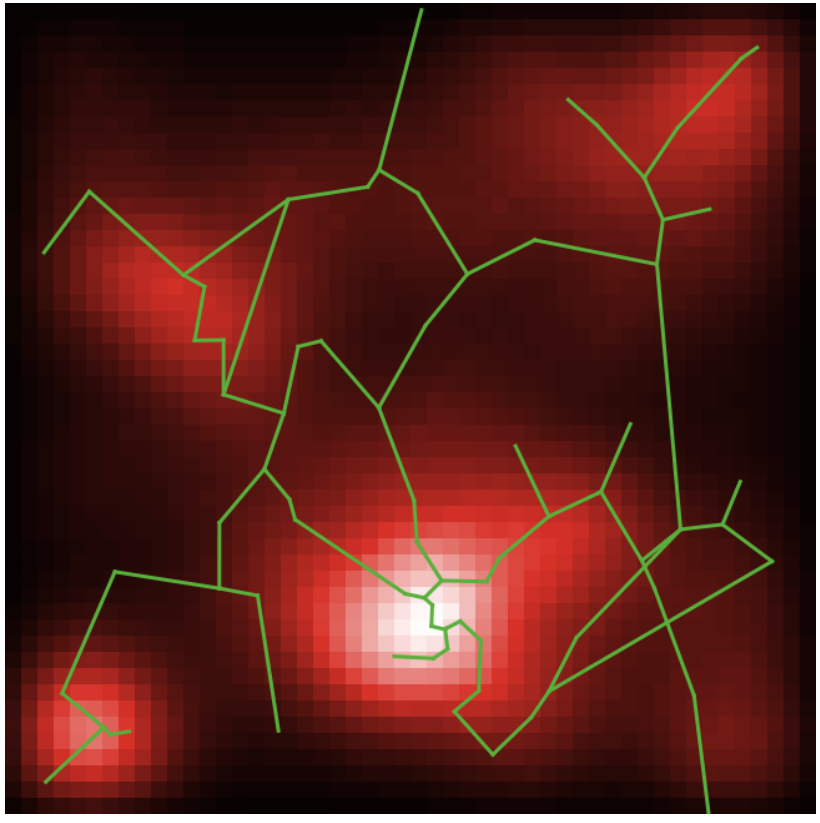
Results : feasible correlations

Mean matrices in a principal plan

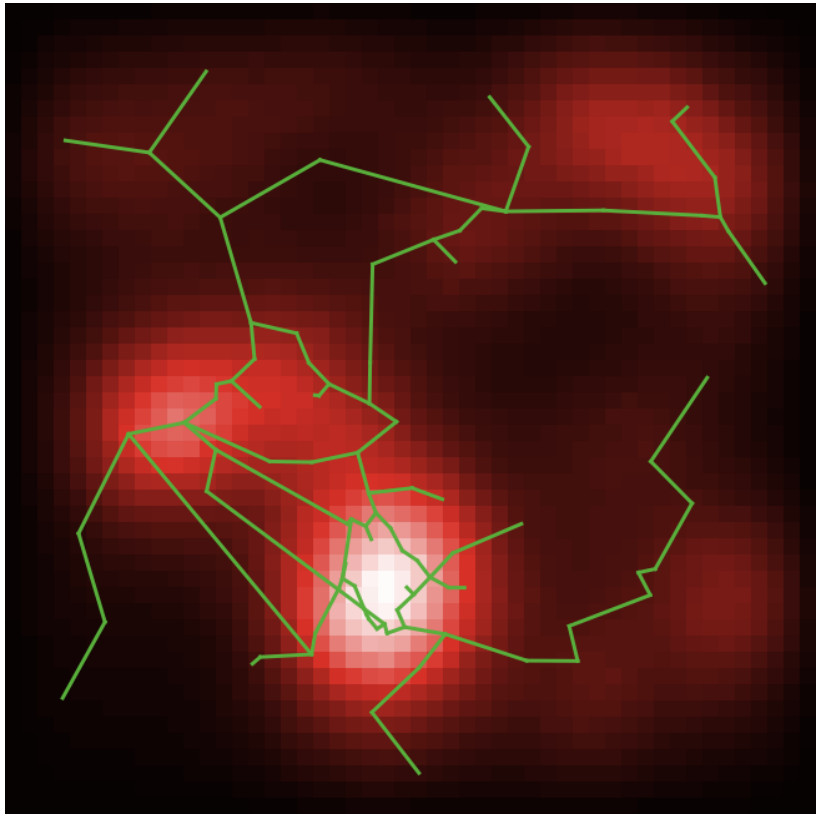




Results : exemples of correlations



$\rho[\bar{d}, \bar{c}] \simeq 0.34$



$\rho[\bar{d}, \bar{c}] \simeq -0.41$

$\rightarrow$ *gravity hierarchy more important in (1)* $\gamma = 3.9, k_h = 0.7$ *against* $\gamma = 1.07, k_h = 0.25$ *for (2)*

Applications

1. Calibration of the coupled model, street network data(edge effects !) $\rightarrow$ generation of correlated synthetic data corresponding to a given urban system $\rightarrow$ intrinsic correlations to be compared to estimated correlations between different states : non-ergodicity of urban systems [Pumain, 2012]).
2. Dynamical correlations in a strongly coupled model / spatio-temporal correlations in a strong spatial coupling.

Context

On Accessibility

Statistical Analysis

P. Bourgine framework for Complex Adaptive Systems

Next steps (until February 15th 2016)

- Theory exemplification, paper finalization [1w]
- Spatial Econometrics / Case study [0.5w]
- Cybergeo [0.5w]
- Wrap everything within a 1-year Memoire [1w]

References I


Abadie, A., Diamond, A., and Hainmueller, J. (2010).
Synthetic control methods for comparative case studies: Estimating the effect of california's tobacco control program.
Journal of the American Statistical Association, 105(490).


Bolón-Canedo, V., Sánchez-Maroño, N., and Alonso-Betanzos, A. (2013).
A review of feature selection methods on synthetic data.
Knowledge and information systems, 34(3):483–519.


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 II

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.

Cottineau, C., Le Néchet, F., Le Texier, M., and Reuillon, R. (2015).
Revisiting some geography classics with spatial simulation.
In *Plurimondi. An International Forum for Research and Debate on Human Settlements*, volume 7.

Le Néchet, F. (2015).
De la forme urbaine à la structure métropolitaine: une typologie de la configuration interne des densités pour les principales métropoles européennes de l'audit urbain.
Cybergeo: European Journal of Geography.

References III


Moeckel, R., Spiekermann, K., and Wegener, M. (2003).
Creating a synthetic population.
In *Proceedings of the 8th International Conference on Computers in Urban Planning and Urban Management (CUPUM)*.


Newman, M. E. (2003).
The structure and function of complex networks.
SIAM review, 45(2):167–256.


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

References IV


Pritchard, D. R. and Miller, E. J. (2009).
Advances in agent population synthesis and application in an integrated land use and transportation model.
In *Transportation Research Board 88th Annual Meeting*, number 09-1686.


Pumain, D. (2012).
Urban systems dynamics, urban growth and scaling laws: The question of ergodicity.
In *Complexity Theories of Cities Have Come of Age*, pages 91–103. Springer.


Raimbault, J. (2016).
Calibration of a spatialized urban growth model.
Working Paper, draft at https://github.com/JusteRaimbault/CityNetwork/tree/master/Docs/Papers

References V


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.


Schmitt, C. (2014).
Modélisation de la dynamique des systèmes de peuplement: de SimpopLocal à SimpopNet.
PhD thesis, Paris 1.


Van den Bulcke, T., Van Leemput, K., Naudts, B., van Remortel, P., Ma, H., Verschoren, A., De Moor, B., and Marchal, K. (2006).
Syntren: a generator of synthetic gene expression data for design and analysis of structure learning algorithms.
BMC bioinformatics, 7(1):43.

References VI

Ye, X. (2011).
Investigation of underlying distributional assumption in nested logit model using copula-based simulation and numerical approximation.
Transportation Research Record: Journal of the Transportation Research Board, (2254):36–43.