Thesis Progress Meeting

J. Raimbault[1,2]

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

September 29th 2016

Achieved Work (by projects)

- SFICSSS [4.7w] (+ holidays [2.5w])
- Network-Density (RGS) [1.7w] (ETA 1w)
- Interaction Gibrat (CCS) [1w] (//th. paper, ETA 1w)
- China [0.2w]
- Reviewing [0.3w]
- Transportation Eq [0.3w]
- Biblio/reading [1.2w] ; Conference [1.9w] ; Misc [1.4w]

For a Cautious Use of Big Data and Computation (RGS-AC 2016)

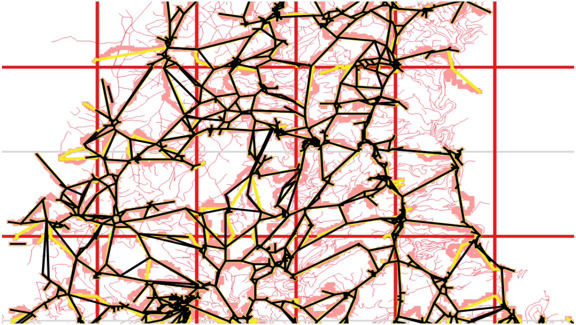
Presentation Rationale : *Large scale computation and Big Data make no sense (are endangered) without theory and/or analytical preliminaries*

Case Study : *Static Correlations between urban form and network measures ; insights into ergodicity and stationarity*

Dataset construction

Computation of topological road network for all Europe, at 100m granularity scale (to be used consistently with population grid [EUROSTAT, 2014])

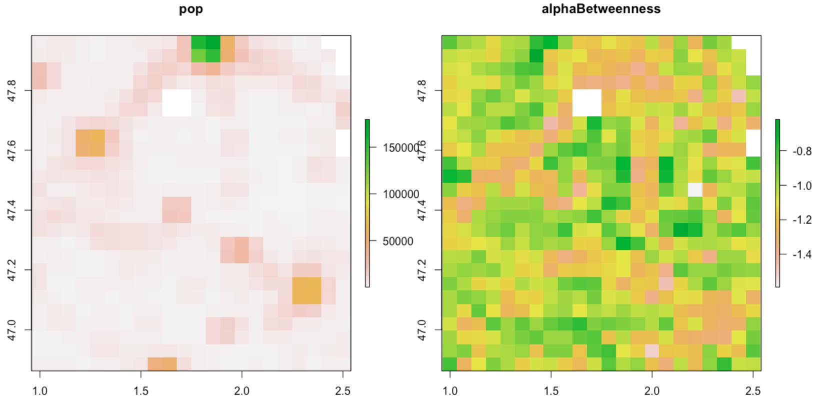
$\rightarrow$ Import of OSM into `pgsql`, simplification at 100m granularity, topological simplification with split/merge algorithm



$\simeq 44 \cdot 10^6$ *links in initial OSM db,* $\simeq 61 \cdot 10^6$ *in first simplified layer,* $\simeq 21 \cdot 10^6$ *in final database*

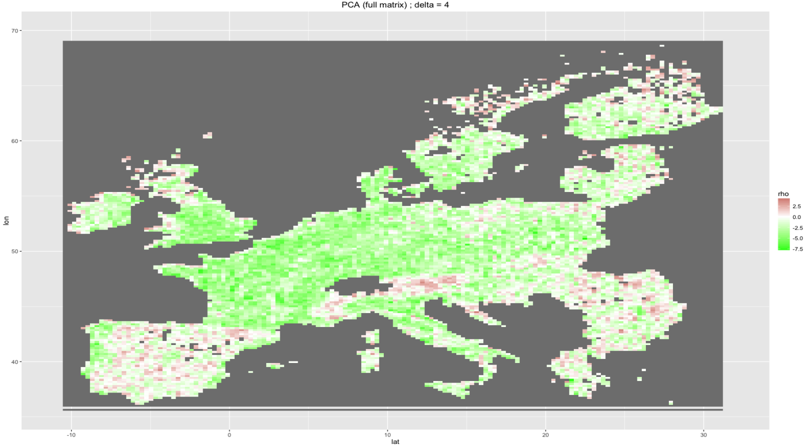
Results : Computation of Indicators

Computation of urban form indicators [Le Néchet, 2015] and network indicators on $l_0 = 10km$ side square



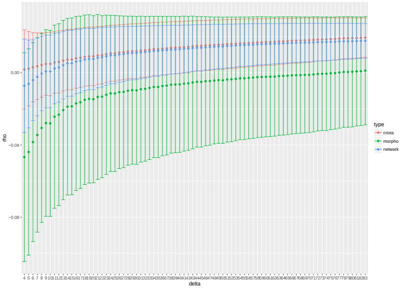
Results : Spatial Correlations

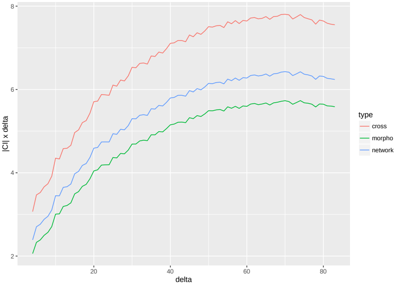
Computation of spatial correlation on square areas of width $\delta \cdot l_0$ *(with typically* $\delta = 4, \ldots, 16$*)*



$\rightarrow$ *local spatial stationarity of processes*

Results : Multi-scale Processes





$\rightarrow$ Significant variation of mean correlation with δ (Left) and of normalized confidence interval (Right) given by $|\rho_+ - \rho_-| \cdot \delta$, as bounds theoretically vary as $\sqrt{N} \sim \sqrt{\delta^2}$: implies multi-scalarity

Empirical Findings (Formalization)

$Y_i\,[\vec{x}, t]$ spatio-temporal stochastic process, verifies empirically :

1. Local spatial autocorrelation is present and bounded by l_ρ (in other words the processes are continuous in space) : at any $\vec{x}$ and t, $\left|\rho_{\|\Delta\vec{x}\|<l_\rho}\left[Y_i(\vec{x}+\Delta\vec{x}, t), Y_i(\vec{x}, t)\right]\right| > 0$.
2. Processes are locally parametrized : $Y_i = Y_i\,[\alpha_i]$, where $\alpha_i(\vec{x})$ varies with l_α, with $l_\alpha \gg l_\rho$ and weakly locally stationary in space.
3. Processes are multi-scalar : since $\rho(\delta = \infty) > \rho(\delta = 0)$, a necessary non-linear correction on processes spatial averages in correlation computation is present.

Analytical Deductions

1. **Regimes of temporal correlations.** Let assume local ergodicity in $\vec{x_0}$ at scale $\delta \cdot l_0$ (reasonable with urban growth and network extension in recent times). The Ergodic theorem implies that $\exists \mathcal{T}$ such that

$$< Y_i(t) >_{\|\vec{x}-\vec{x_0}\|<\delta \cdot l_0} = < Y_i(\vec{x_0}) >_{t \in \mathcal{T}}$$

With spatial stationarity, $< Y_i >_{\vec{x_0}} = < Y_i >_{\vec{x_1}}$, thus $\mathcal{T}$ must be constant to be invariant by translation. By contraposition and (2), processes have different dynamical characteristics.

2. **Global non-ergodicity.** Let X_k a partition of space into local areas. We have $< \cdot >_x = \sum_k w_k < \cdot >_{x_k} =_{(1)} \sum_k w_k < \cdot >_{\mathcal{T}_k}$. On the other hand, global ergodicity would give $< \cdot >_t = < \cdot >_{\mathcal{T}} = \sum_k w_k < \cdot >_{\mathcal{T}}$ and $\sum_k w_k \left(< \cdot >_{\mathcal{T}} - < \cdot >_{\mathcal{T}_k}\right) = 0$. Being true on each subset implies $\mathcal{T} = \mathcal{T}_k$, what contradicts (1).

Simple Models of Growth for System of Cities (CCS 2016)

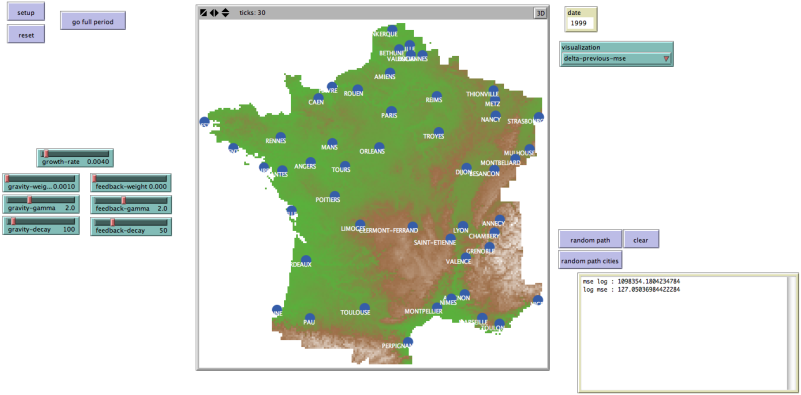
Presentation Content :

→ *Gravity/Network feedback spatial interactions models of growth : presentation and current results*

→ *Proposition of an empirical AIC methodology, application to the model*

Implementation

On the importance of visualization in spatial models : complementary implementations in NetLogo/R/Scala



Quantifying overfitting : Empirical AIC

Not clear nor well theorized how to deal with overfitting in models of simulation. **Intuitive idea :** Approximate gain of information by approaching models of simulation by statistical models.

Let $M_k^* = M_k [\alpha_k^*]$ computational models heuristically fitted to the same dataset. With $S_k \simeq M_k^*$, we show that $\Delta D_{KL}(M_k^*, M_{k'}^*) \simeq \Delta D_{KL}(S_k, S_{k'})$ if fits of S_k are negligible compared to fit difference between computational models and models have same parameter number.

Application M_1 : gravity only model with $(r_0 = 0.0133, w_G = 1.28e-4, \gamma_G = 3.82, d_G = 4e12)$; M_2 : full model with $(r_0 = 0.0128, w_G = 1.30e-4, \gamma_G = 3.80, d_G = 8.4e14, w_N = 0.603, \gamma_N = 1.148, d_N = 7.474)$ Fitting of independent polynomial models ($\tilde{P}_i(t) = Q\left[\tilde{P}_i(t-1)\right]$) with 4 and 7 parameters) gives $\Delta D_{KL} \simeq 19.7 \rightarrow$ fit improvement without overfitting

Extension/Application of Lutecia model

→ Empirical description of governance structure for transportation within the MCR, associated issues.

→ Application of the model to the real case (after refined exploration, sensitivity analysis and internal validation).

→ Focus on specific questions involving the medium-sized city of Zhuhai : position and influence within MCR processes ; specific governance configuration (the city is a Special Economic Zone e.g.).

Towards a SimpopSino model

→ Extension of Gibrat-Interaction model, taking into account at least one economic dimension

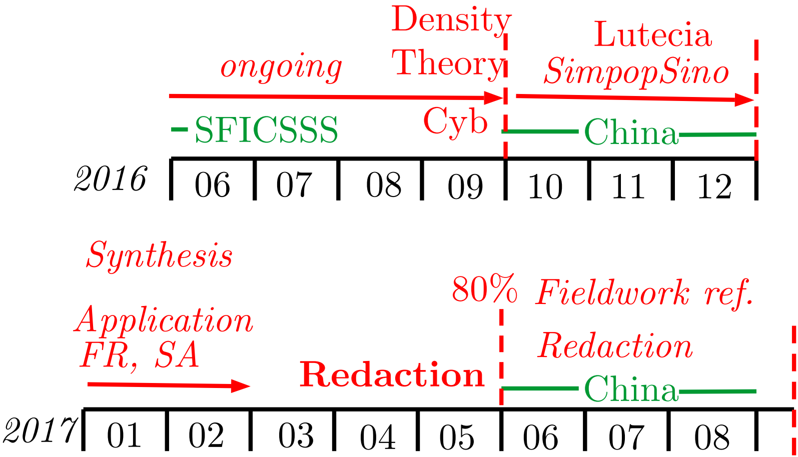
→ Meeting with Denise : ok for Elfie dataset ; importance of link meso-macro in transportation

→ Rapidly growing network in China
[Kenworthy and Hu, 2002, Lyu et al., 2016] :
construction of a dynamical database ; test static/dynamic/co-evolution network extensions for SimpopSino

Next steps (until December 15th 2016)

- Lutecia : model [3w] ; fieldwork [2w]
- SimpopSino : theoretical/modeling [3w]
- Thesis/papers writing [2w]

Thesis Organisation (remainder)



References I

EUROSTAT (2014).
Eurostat geographical data.

Kenworthy, J. and Hu, G. (2002).
Transport and urban form in chinese cities: an international comparative and policy perspective with implications for sustainable urban transport in china.
disP-The Planning Review, 38(151):4–14.

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.

Lyu, G., Bertolini, L., and Pfeffer, K. (2016).
Developing a tod typology for beijing metro station areas.
Journal of Transport Geography, 55:40–50.