A scala library for spatial sensitivity analysis

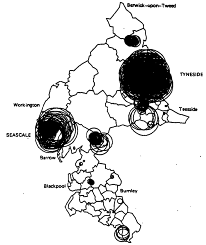
J. Raimbault[1,2,3,*], J. Perret[4,2] and R. Reuillon[2,3]

[1] Center for Advanced Spatial Analysis, UCL
[2] UPS CNRS 3611 ISC-PIF
[3] UMR CNRS 8504 Géographie-cités
[4] LaSTIG STRUDEL, IGN, ENSG, Univ. Paris-Est

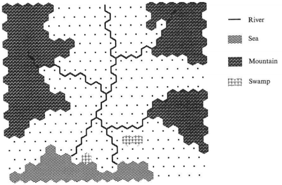
[*] juste.raimbault@polytechnique.edu

GISRUK 2020
July 21st, 2020

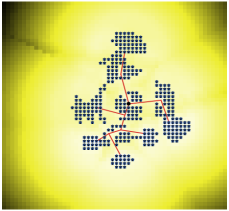
Modeling and simulation in geography



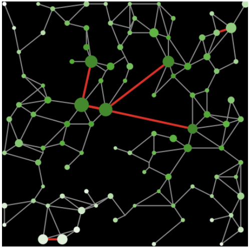
Geographical analysis machine [Openshaw et al., 1987]



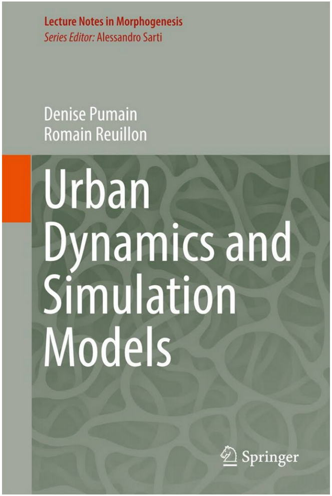
Simpop 1 model[Sanders et al., 1997]



Hybrid urban morphogenesis [Raimbault et al., 2014]



SimpopNet model [Schmitt, 2014]



Development of an evolutionary urban theory [Pumain, 2018]

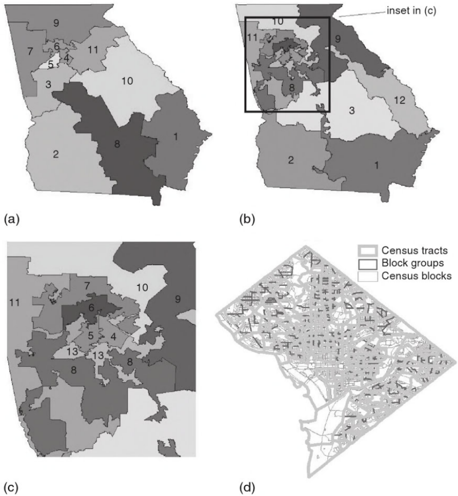
→ Recurrent stylized facts on main systems of cities
→ Construction of simulation models (with an explicative purpose)
→ Tools and methods to explore simulation models

Sensitivity analysis issues in geography

Classical problems in geography and spatial sciences:

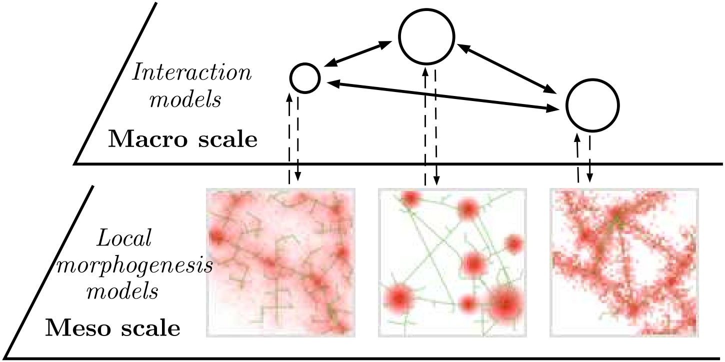
- Modifiable Areal Unit Problem
- Dependancy of processes to scale
- Spatial non-stationarity
- Fuzzy and noisy data
- Genericity/particularity

The Modifiable Areal Unit Problem

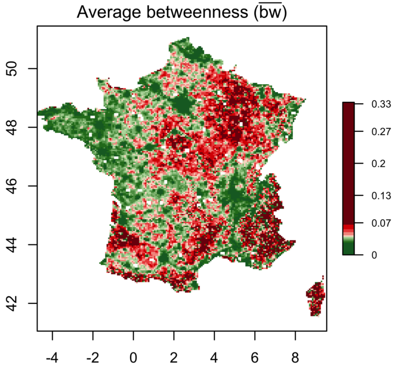


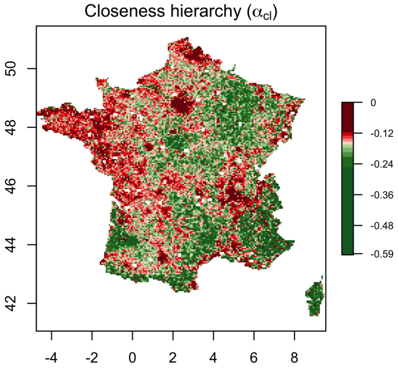
Wong, D. (2009). The modifiable areal unit problem (MAUP). The SAGE handbook of spatial analysis, 105, 23.

Processes specific to scales, coupling implies dedicated ontologies [Raimbault, 2019a]



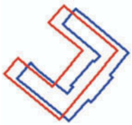
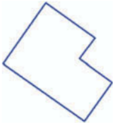
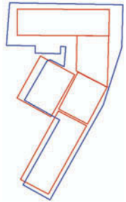
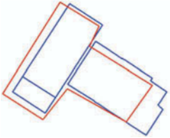
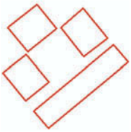
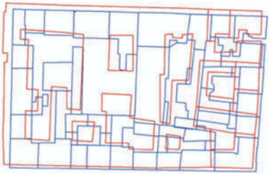
Spatial non-stationarity of road network properties
[Raimbault, 2019c]



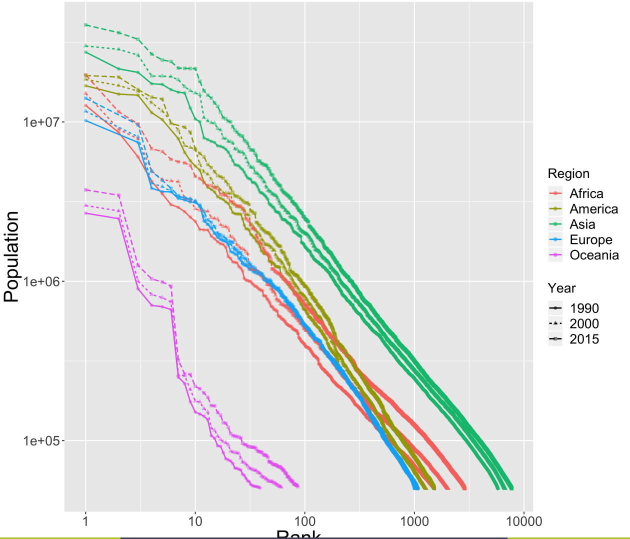


Noisy spatial data

Assessment of data quality in OpenStreetMap [Fan et al., 2014]

Relation	1:1	1:0	1:n
Illustration			
Relation	n:1	0:1	n:m
Illustration			

Urban systems are simultaneously universal and particular
[Raimbault et al., 2020]



Spatial configurations are parameters too!

- "*Space matters*": impact of spatial configuration on model behavior
- Model behaviors which are robust to spatial configuration
- Model behaviors which are robust to noise in parametrization real datasets

$\Longrightarrow$ ***Construction of a generic library for spatial sensitivity analysis, including the generation of synthetic data, perturbation of real data and indicators***

→ coupling models with spatial configuration generators (spatial synthetic data) gives model sensitivity to space through sensitivity analysis of the coupled model

→ synthetic urban forms resembling real configurations

→ at different scales: microscopic (buildings), mesoscopic (population distribution), macroscopic (system of cities)

Generating building layouts

At the microscopic scale (district): generating building layouts

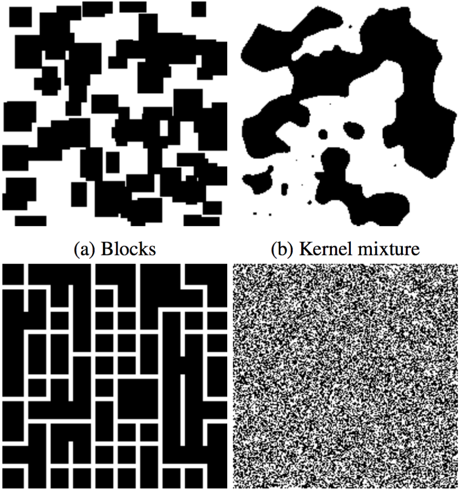
Raimbault, J., & Perret, J. (2019, July). Generating urban morphologies at large scales. In Artificial Life Conference Proceedings (pp. 179-186). MIT Press.

- systematic comparison of simple processual generators
- introduction of morphological indicators
- calibration on sampled layouts from OpenStreetMap

Urban form indicators for building layouts:

- density, number of buildings, average area
- Moran index and average distance on rasterized representation
- average detour in the free space
- mathematical morphology indicators (steps for erosion and dilation) [Serra, 1994]

Complementary generators



(a) Blocks (b) Kernel mixture

(c) Network percolation (d) Random

At the mesoscopic scale: population grids

- a reaction-diffusion model for population distributions
- urban form measures at the mesoscopic scale
- synthetic generators coupling population and road networks

Raimbault, J. (2018). Calibration of a density-based model of urban morphogenesis. PloS one, 13(9), e0203516.

Raimbault, J. (2019). An urban morphogenesis model capturing interactions between networks and territories. In The Mathematics of Urban Morphology (pp. 383-409). Birkhäuser, Cham.

A simple Reaction-diffusion model

→ Crucial role of the interplay between concentration forces and dispersion forces [Fujita and Thisse, 1996] in keeping Urban Systems at the border of chaos

→ Potentiality of aggregation mechanisms (such as Simon model) to produce power laws

→ Link with Reaction-diffusion approaches in Morphogenesis [Turing, 1952]

→ Extension of a DLA-type model introduced by [Batty, 1991], with simple abstract processes of population aggregation and diffusion

Model Formalization

OpenMOLE

$\rightarrow$ Grid world with cell populations $(P_i(t))_{1\leq i\leq N^2}$.

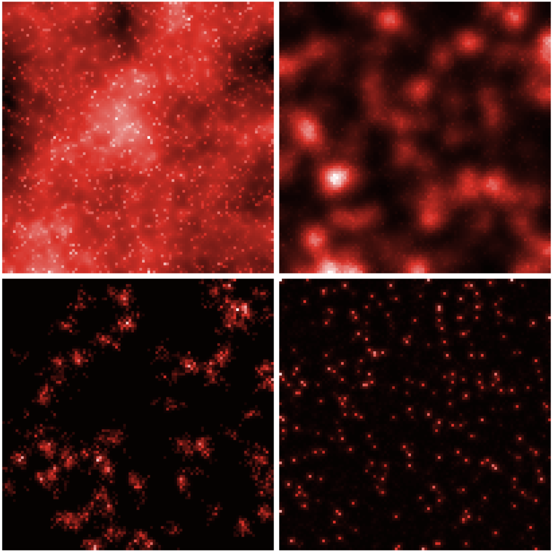
$\rightarrow$ At each time step:

1. Population growth with exogenous rate N_G, attributed independently to a cell following a preferential attachment of strength α
2. Population is diffused n_d times with strength β

$\rightarrow$ Stopping criterion: fixed maximal population P_m.

$\rightarrow$ Output measured by morphological indicators: Moran index, average distance, rank-size hierarchy, entropy.

Generating Population Distributions



Examples of generated territorial shapes

1. Rank-size slope γ, given by $\ln\left(P_{\tilde{i}}/P_0\right) \sim k + \gamma \cdot \ln\left(\tilde{i}/i_0\right)$ where $\tilde{i}$ are the indexes of the distribution sorted in decreasing order.
2. Entropy of the distribution:

$$\mathcal{E} = \sum_{i=1}^{M} \frac{P_i}{P} \cdot \ln \frac{P_i}{P} \qquad (1)$$

$\mathcal{E} = 0$ means that all the population is in one cell whereas $\mathcal{E} = 0$ means that the population is uniformly distributed.

3. Spatial-autocorrelation given by Moran index, with simple spatial weights given by $w_{ij} = 1/d_{ij}$

$$I = M \cdot \frac{\sum_{i \neq j} w_{ij} \left(P_i - \bar{P}\right) \cdot \left(P_j - \bar{P}\right)}{\sum_{i \neq j} w_{ij} \sum_i \left(P_i - \bar{P}\right)^2}$$

4. Mean distance between individuals

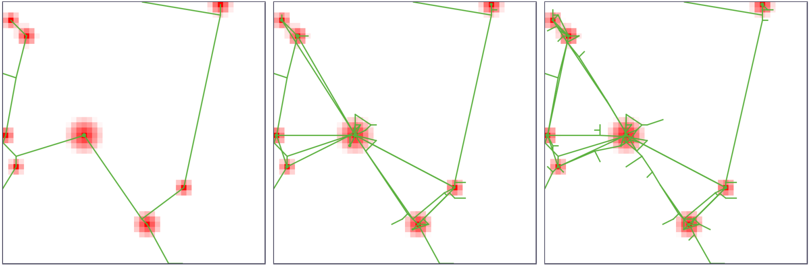
$$\bar{d} = \frac{1}{d_M} \cdot \sum_{i<j} \frac{P_i P_j}{P^2} \cdot d_{ij}$$

where d_M is a normalisation constant

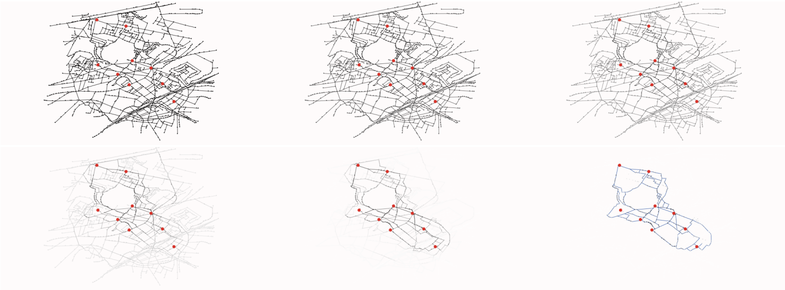
[currently being implemented]

Network generated conditionally to population; at fixed time steps:

1. Add new nodes preferentially to new population and connect them
2. Variable heuristic for new links, among: nothing, random, gravity-based deterministic breakdown, gravity-based random breakdown (from [Schmitt, 2014]), cost-benefits (from [Louf et al., 2013]), biological network generation (based on [Tero et al., 2010])



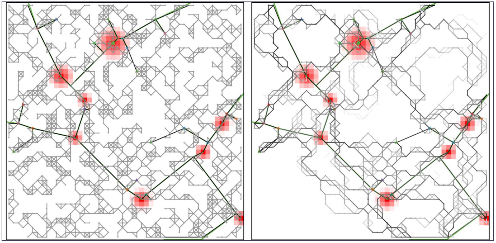
Model studied by [Tero et al., 2010] : exploration and reinforcement by a slime mould searching for ressources



Application to the design of optimal bus routes

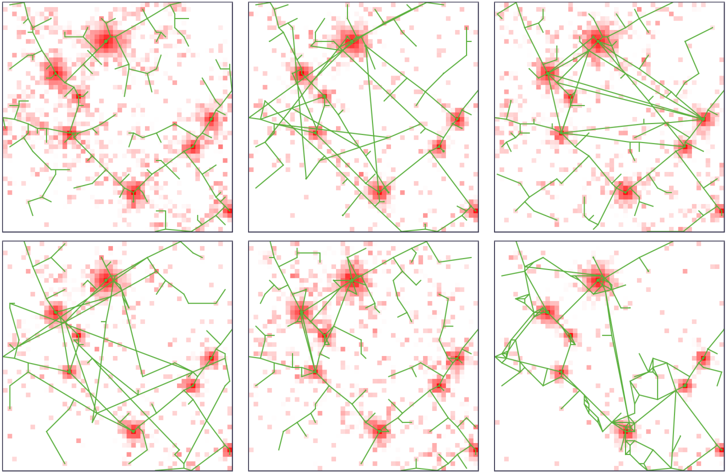
Adding new links with biological heuristic:

1. Create network of potential new links, with existing network and randomly sampled diagonal lattice
2. Iterate for k increasing ($k \in \{1, 2, 4\}$ in practice) :
   - Using population distribution, iterate $k \cdot n_b$ times the slime mould model to compute new link capacities
   - Delete links with capacity under θ_d
   - Keep the largest connected component
3. Planarize and simplify final network



Intermediate stage for biological network generation

Generated Urban Shapes: Network



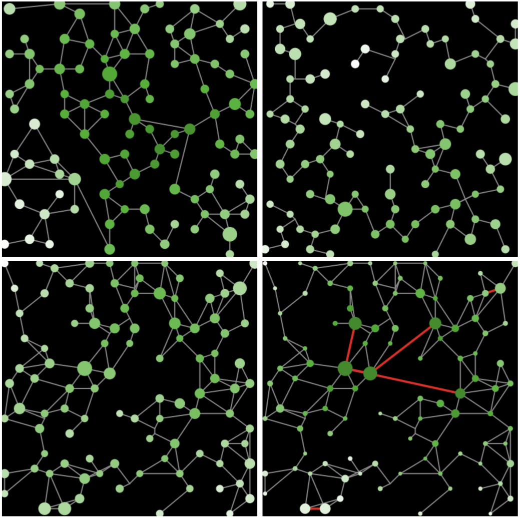
In order: connection; random; deterministic breakdown; random breakdown; cost-driven; biological.

Synthetic systems of cities

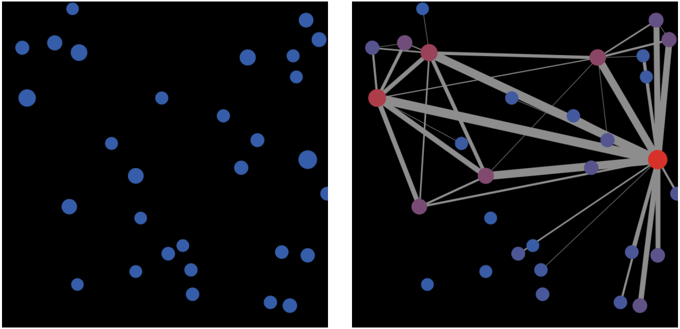
At the macroscopic scale: synthetic systems of cities

- Evolutive urban theory: systems of cities follow general stylized facts [Pumain, 2018]
- Rank-size law [Pumain et al., 2006]
- Central place theory

Synthetic system of cities and network for the SimpopNet model [Raimbault, 2020]



Cities-network co-evolution model explored on synthetic systems of cities [Raimbault, 2018a]



Real data perturbation

[currently being implemented]

$\rightarrow$ *How does noise in real data impacts the result ?*

- Impact of missing elements
- Impact of imprecise coordinates or topology
- Optimal matching between spatial datasets

$\rightarrow$ *How does perturbation of real data allows to explore scenario*

Examples:

- simulating urban projects by modifying population of areas with a given spatial correlation structure
- simulating network disruptions or new transportation lines

Spatial statistics

In the spatial approach, spatial model indicators are also important: what kind of spatial structure does the model produce ?

- previous form indicators at different scales, applied on any spatialized variable or event: quantify level of aggregation, hierarchy, clustering
- spatial statistics indicators and methods
- more complicated approaches: fractals and multifractals, spatial datamining

Implementation: integration into OpenMOLE

Library implemented in `scala`: advantages of functional and object programming; Apache Spark; no widely used GIS library in scala.

`https://github.com/openmole/spatialdata`

→ integration into the OpenMOLE model exploration open source software [Reuillon et al., 2013]

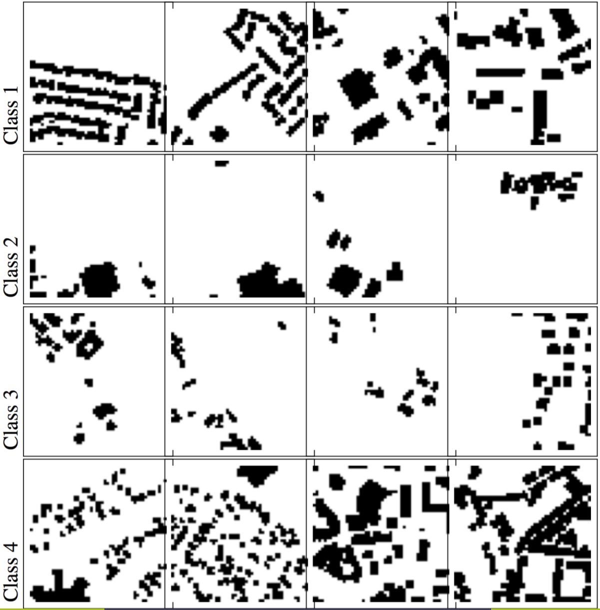
OpenMOLE

Enables seamlessly (i) model embedding; (ii) access to HPC resources; (iii) exploration and optimization algorithms

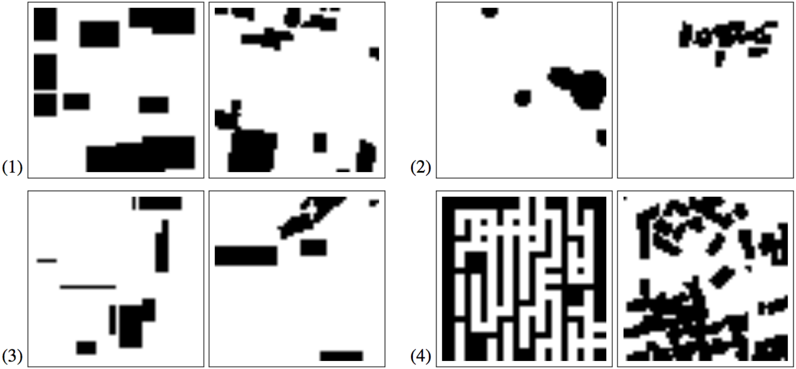
`https://openmole.org/`

Sampled districts from OpenStreetMap

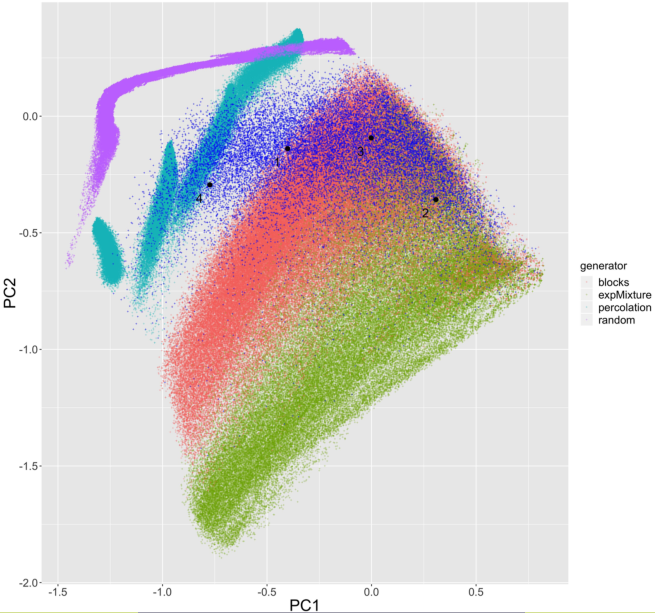
Classification of urban forms



Calibrated forms



Point cloud


0.0
-0.5
-1.0
-1.5
-2.0
-1.5
-1.0
-0.5
0.0
0.5
PC1
PC2
1
2
3
4
generator
blocks
expMixture
percolation
random

Calibration

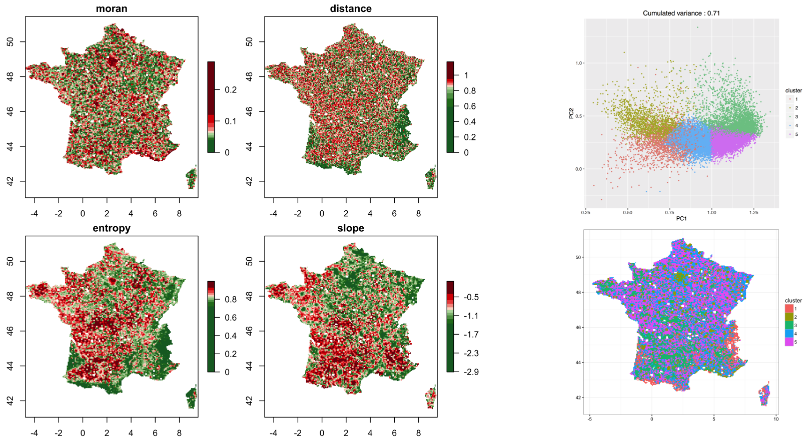
	Random	Blocks	Exp. Mixture	Percolation
Centroid 1	0.424 ± 0.011	0.106 ± 0.063	0.303 ± 0.101	0.325 ± 0.019
Centroid 2	0.809 ± 0.022	0.164 ± 0.099	0.184 ± 0.141	0.947 ± 0.019
Centroid 3	0.428 ± 0.019	0.095 ± 0.054	0.109 ± 0.064	0.541 ± 0.019
Centroid 4	0.515 ± 0.005	0.311 ± 0.077	0.589 ± 0.149	0.083 ± 0.025

Why not use calibration heuristics? Open question of fitting a point cloud; issue of projecting in a reduced dimension space

Implementation in OpenMOLE

```
sampling =
BlocksGridSpatialSampling(
    samples = 2,
    gridSize = 10,
    blocks = (p1gen in Range(1.0,10.0)),
    blockMinSize = (p2gen in Range(2.0,4.0)),
    blockMaxSize = (p3gen in Range(2.0,4.0)),
    prototype = w
)
```

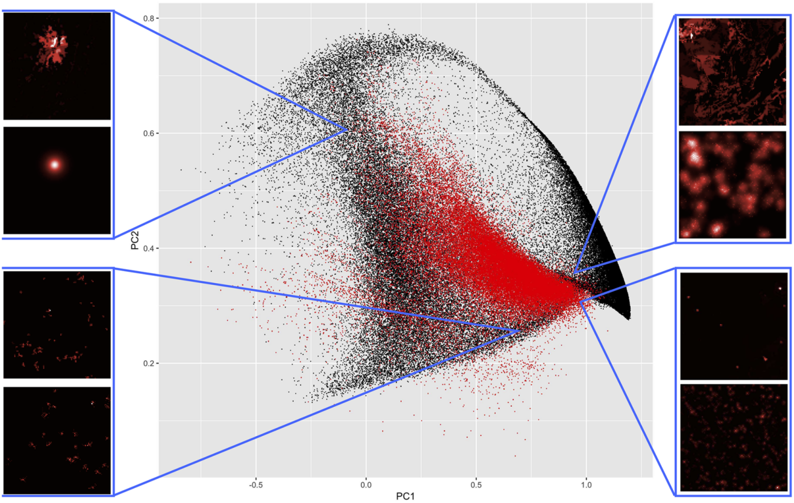
→ other generators have their own primitives (`ExpMixtureThresholdSpatialSampling`, `PercolationGridSpatialSampling`) and arguments (**see the documentation**)



Computation of morphological indicators on population density data for Europe (shown here on France), morphological classification.

Population grids: model Calibration

OpenMOLE

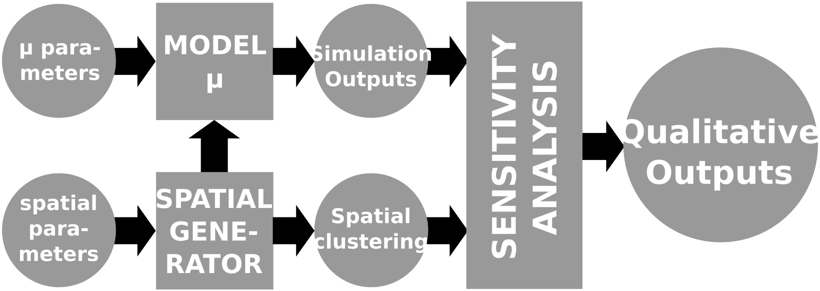


Brute force calibration by exploring the parameter space. Reproduction of most existing configuration in the morphological sense (here in principal plan).

Method flowchart

General workflow to test the spatial sensitivity of simulation models

Raimbault, J., Cottineau, C., Le Texier, M., Le Nechet, F., & Reuillon, R. (2019). Space Matters: Extending Sensitivity Analysis to Initial Spatial Conditions in Geosimulation Models. Journal of Artificial Societies and Social Simulation, 22(4).



Quantification of spatial sensitivity

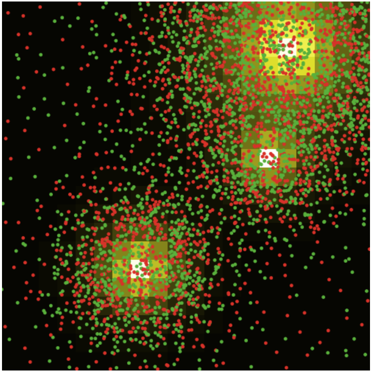
OpenMOLE

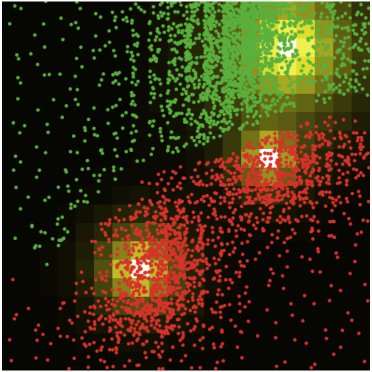
Relative distance of phase diagrams to compare global model behavior when meta-parameters change

$$d_r\left(\mu_{\vec{\alpha}_1},\mu_{\vec{\alpha}_2}\right) = 2\cdot\frac{d(\mu_{\vec{\alpha}_1},\mu_{\vec{\alpha}_2})^2}{Var\left[\mu_{\vec{\alpha}_1}\right]+Var\left[\mu_{\vec{\alpha}_2}\right]}$$

Why could the Schelling model be sensitive to space ?

[Banos, 2012] network effects in Schelling model





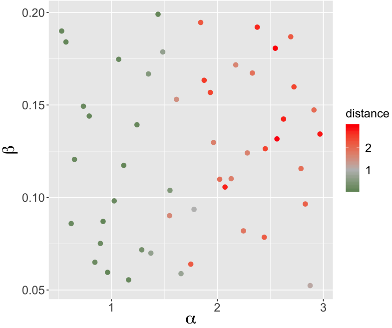
Influence of spatial generator parameters on model outputs

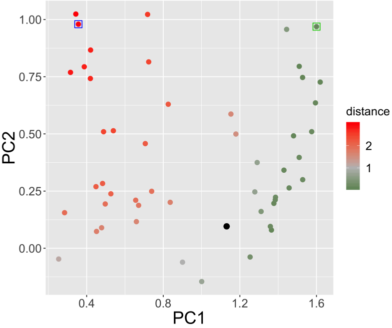
Simulation outcome by segregation index:	Dissimilarity		Entropy		Moran's I	
Intercept	-0.212 ***	-0.141 ***	-0.254 ***	-0.208 ***	-0.036 ***	-0.061 ***
Similarity Wanted (S)	1.212 ***	1.212 ***	1.250 ***	1.250 ***	0.550 ***	0.550 ***
quadratic term (S^2)	-0.942 ***	-0.942 ***	-0.963 ***	-0.963 ***	-0.428 ***	-0.438 ***
Vacancy Rate (V)	0.602 ***	0.602 ***	0.453 ***	0.453 ***	-0.027 ***	-0.027 ***
Minority Index (%Maj - %Min)	0.307 ***	0.307 ***	0.130 ***	0.130 ***	-0.067 ***	-0.067 ***
Density Grid = Polycentric		0.087 ***		0.052 ***		0.001 ***
Density Grid = Discontinuous		0.111 ***		0.068 ***		*0.00*
Attraction generator parameter α		-0.083 ***		-0.053 ***		0.014 ***
Diffusion generator parameter β		0.323 ***		0.218 ***		0.017 ***
R2 (%)	30.6	34.7	24.1	25.6	23.9	24.0
# of observations (sim. runs)	2,106,000	2,106,000	2,106,000	2,106,000	2,106,000	2,106,000
AIC	-70717.68	-198748.2	208213.8	166048.8	-4385990	-4387816

A model of resource collection

- agents collect a spatial resource
- the resource regrows at a certain rate only

Relative distances between phase diagrams





Developments

- more spatial network generative models [Raimbault, 2018b], correlated synthetic data [Raimbault, 2019b]
- domain models: LUTI, urban dynamics
- other disciplines, ecology, geosciences [Mogheir et al., 2004]?
- interaction with data driven disciplines ? (planning, architecture, spatio-temporal datamining)
- genericity of some models? (reaction-diffusion)
- synthetic data generation methods (synthetic populations)
- synthetic data at the core of applied statistics methodology (less in spatial statistics?)
- port the library to more classic languages (python, R)

Synthesis of available tools

	OpenMOLE	spatialdata
Micro grid spatial samplings	✓	✓
Meso grid spatial samplings	✓	✓
Macro spatial samplings	✗	✓
Spatial network generation	✗	✓
Real data import	✗	✗
Real data perturbations	✗	✓
Spatial statistics	✓	✓
Hybrid methods	✗	✓
Domain models (transportation, land-use)	✗	✓

Conclusion

→ **Space matters:** relevance of spatially-explicit models and spatial sensitivity analysis.

→ **Synthetic data:** first experimental samplings included in OpenMOLE, soon more to come.

→ **Disciplinary context:** strong contingency on included models.

Get the library at `https://github.com/openmole/spatialdata`

Open issues at `https://github.com/openmole/spatialdata/issues`

References I

Banos, A. (2012).
Network effects in schelling's model of segregation: new evidence from agent-based simulation.
Environment and Planning B: Planning and Design, 39(2):393–405.

Batty, M. (1991).
Generating urban forms from diffusive growth.
Environment and Planning A, 23(4):511–544.

Fan, H., Zipf, A., Fu, Q., and Neis, P. (2014).
Quality assessment for building footprints data on openstreetmap.
International Journal of Geographical Information Science, 28(4):700–719.

References II

Fujita, M. and Thisse, J.-F. (1996).
Economics of agglomeration.
Journal of the Japanese and international economies, 10(4):339–378.

Louf, R., Jensen, P., and Barthelemy, M. (2013).
Emergence of hierarchy in cost-driven growth of spatial networks.
Proceedings of the National Academy of Sciences, 110(22):8824–8829.

Mogheir, Y., De Lima, J., and Singh, V. (2004).
Characterizing the spatial variability of groundwater quality using the entropy theory: I. synthetic data.
Hydrological processes, 18(11):2165–2179.

References III

Openshaw, S., Charlton, M., Wymer, C., and Craft, A. (1987).
A geographical analysis machine for the automated analysis of point data sets.
International Journal of Geographical Information System, 1(4):335–358.

Pumain, D. (2018).
An evolutionary theory of urban systems.
In *International and Transnational Perspectives on Urban Systems*, pages 3–18. Springer.

Pumain, D., Paulus, F., Vacchiani-Marcuzzo, C., and Lobo, J. (2006).
An evolutionary theory for interpreting urban scaling laws.
Cybergeo: European Journal of Geography.

References IV

Raimbault, J. (2018a).
Modeling the co-evolution of cities and networks.
arXiv preprint arXiv:1804.09430.

Raimbault, J. (2018b).
Multi-modeling the morphogenesis of transportation networks.
In *Artificial Life Conference Proceedings*, pages 382–383. MIT Press.

Raimbault, J. (2019a).
A multi-scalar model for system of cities.
In *Conference on Complex Systems 2019*, Singapore, Singapore.

References V

Raimbault, J. (2019b).
Second-order control of complex systems with correlated synthetic data.
Complex Adaptive Systems Modeling, 7(1):1–19.

Raimbault, J. (2019c).
An urban morphogenesis model capturing interactions between networks and territories.
In *The Mathematics of Urban Morphology*, pages 383–409. Springer.

Raimbault, J. (2020).
Unveiling co-evolutionary patterns in systems of cities: a systematic exploration of the simpopnet model.
In *Theories and Models of Urbanization*, pages 261–278. Springer.

References VI

Raimbault, J., Banos, A., and Doursat, R. (2014).
A hybrid network/grid model of urban morphogenesis and optimization.
In *4th International Conference on Complex Systems and Applications*, pages 51–60.

Raimbault, J., Denis, E., and Pumain, D. (2020).
Empowering urban governance through urban science: Multi-scale dynamics of urban systems worldwide.
Sustainability.

Raimbault, J. and Perret, J. (2019).
Generating urban morphologies at large scales.
In *Artificial Life Conference Proceedings*, pages 179–186. MIT Press.

References VII

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.

Sanders, L., Pumain, D., Mathian, H., Guérin-Pace, F., and Bura, S. (1997).
Simpop: a multiagent system for the study of urbanism.
Environment and Planning B, 24:287–306.

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

References VIII

- Serra, J. (1994).
  Morphological filtering: an overview.
  Signal processing, 38(1):3–11.

- 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.

- Turing, A. (1952).
  The chemical theory of morphogenesis.
  Phil. Trans. Roy. Soc, 13(1).