

# A scala library for spatial sensitivity analysis

J. Raimbault<sup>\*1,2,3</sup>, J. Perret<sup>4,2</sup> and R. Reuillon<sup>2,3</sup>

<sup>1</sup>Center for Advanced Spatial Analysis, University College London

<sup>2</sup>UPS CNRS 3611 ISC-PIF

<sup>3</sup>UMR CNRS 8504 Géographie-cités

<sup>4</sup>LaSTIG STRUDEL, IGN, ENSG, Univ. Paris-Est

## Summary

The sensitivity analysis and validation of simulation models require specific approaches in the case of spatial models. We describe the `spatialdata` scala library providing such tools, including synthetic generators for urban configurations at different scales, spatial networks, and spatial point processes. These can be used to parametrize geosimulation models on synthetic configurations, and evaluate the sensitivity of model outcomes to spatial configuration. The library also includes methods to perturb real data, and spatial statistics indicators, urban form indicators, and network indicators. It is embedded into the OpenMOLE platform for model exploration, fostering the application of such methods without technical constraints.

**KEYWORDS:** Sensitivity analysis; Geosimulation; Spatial synthetic data; Model validation; Model exploration.

## 1 Introduction

The sensitivity of geographical analyses to the spatial structure of data is well known since the Modifiable Areal Unit Problem was put forward by Openshaw (1984). This type of issue has been generalized to various aspects since, including temporal granularity (Cheng and Adepeju, 2014) or the geographical context more generally (Kwan, 2012). When studying geosimulation models (Benenson and Torrens, 2004), similar issues must be taken into account, extending classical sensitivity analysis methods (Saltelli et al., 2004) to what can be understood as *Spatial Sensitivity Analysis* as proposed by Raimbault et al. (2019).

Several studies showed the importance of that approach. For example, in the case of Land-use Transport interaction models, Thomas et al. (2018) show how the delineation of the urban area can significantly impact simulation outcomes. Banos (2012) studies the Schelling segregation model on networks, and shows that network structure strongly influences model behavior. The spatial resolution in raster configurations can also change results (Singh et al., 2007).

On the other hand, the use of spatial synthetic data generation is generally bound to model parametrization without a particular focus on sensitivity analysis, such as in microsimulation models (Smith et al., 2009), spatialized social networks (Barrett et al., 2009), or architecture (Penn, 2006). Raimbault et al. (2019) however showed that systematically generating synthetic data, with constraints of proximity to real data configuration, can be a powerful tool to evaluate the sensitivity of geosimulation models to the spatial configuration.

This contribution describes an initiative to synthesize spatial sensitivity analysis techniques such as synthetic data generation, real data perturbation, and specific indicators, under a common

---

\*juste.raimbault@polytechnique.edu

operational framework. In practice, methods are implemented in the `spatialdata` scala library, allowing in particular its embedding into the OpenMOLE model exploration platform (Reuillon et al., 2013).

## 2 Spatial sensitivity methods

**Generation of spatial synthetic data** Realistic spatial synthetic configurations can be generated for geographical systems at different scales, and as different data types. Regarding raster data, (i) at the microscopic scale raster representation of building configurations (typical scale 500m) are generated using procedural modeling, kernel mixtures, or percolation processes (Raimbault and Perret, 2019); and (ii) at the mesoscopic scale, population density grids (typical scale 50km) are generated using a reaction-diffusion urban morphogenesis model (Raimbault, 2018a) or kernel mixture. Regarding network data, synthetic generators for spatial networks include baseline generators (random planar network, tree network) and generators tailored to resemble road networks at a mesoscopic scale, following different heuristics including gravity potential breakdown, cost-benefits link construction, and a bio-inspired (slime mould) network generation model (Raimbault, 2018b) (Raimbault, 2019b). Finally, regarding vector data, spatial fields generators can be applied at any scale (points distribution following a given probability distribution, or spatial Poisson point processes), while at the macroscopic scale system of cities with a spatialized network can be generated (Raimbault, 2020).

**Real data perturbation** Real raster data can be loaded with the library and perturbed with random noise or following a Poisson point process. A raster generator at the microscopic scale can be used to load real building configurations from OpenStreetMap. For transportation networks, vector representations can be imported from shapefiles, directly from the OpenStreetMap API, or from a database (MongoDB and PostGIS are supported), and are transformed into a proper graph representation. Network perturbation algorithms include node or link deletion (for resilience studies e.g.) and noise on nodes coordinates.

**Indicators** Finally, various indicators are included in the library, which can be used to characterize generated or real configurations, and compare them. They include spatial statistics measures (spatial moments, Ripley K), urban morphology measures at the microscopic and mesoscopic scale, and network measures (basic measures, centralities, efficiency, components, cycles). Network measures can furthermore take into account congestion effects, as basic network loading algorithms (shortest paths and static user equilibrium) are implemented.

**Implementation and integration in OpenMOLE** The library is implemented in the language scala, which is based on the Java Virtual Machine and can benefit of existing Java libraries, and couples the robustness of functional programming with the flexibility of object-oriented programming. It can therefore easily be combined with one of the numerous Java simulation frameworks (Nikolai

and Madey, 2009), such as for example Repast Symphony for agent-based models (North et al., 2013), JAS-mine for microsimulation (Richiardi and Richardson, 2017), or Matsim for transportation (Horni et al., 2016). The library is open source under a GNU GPL License and available at <https://github.com/openmole/spatialdata/>. A significant part of the library (synthetic raster generation methods) is integrated into the OpenMOLE model exploration platform (Reuillon et al., 2013). This platform is designed to allow seamless model validation and exploration, using workflows making the numerical experiments fully reproducible (Passerat-Palmbach et al., 2017). It combines (i) model embedding in almost any language; (ii) transparent access to high performance computation infrastructures; and (iii) state-of-the-art methods for models validation (including design of experiments, genetic algorithms for calibration, novelty search, etc.). Reuillon et al. (2019) illustrates how this tool can be particularly suited to validate geosimulation models.

### 3 Applications

Different applications of the library have already been described in the literature. Regarding the generation of synthetic data in itself, Raimbault and Perret (2019) show that the building configuration generators are complementary to reproduce a large sample of existing configurations in European cities. Raimbault (2018a) shows that the reaction-diffusion morphogenesis model is flexible enough to capture most existing urban forms of population distributions across Europe also. Raimbault (2019a) shows that it is possible to weakly couple the population density generator with the gravity-breakdown network generator, and that correlations between urban form and network indicators can be modulated this way. Raimbault (2019b) does a similar coupling in a dynamic way and shows that the co-evolution between road network and population distribution can be modeled this way.

For the application of the library to spatial sensitivity analysis, Raimbault et al. (2019) apply the population distribution generator to two textbook geosimulation models (Schelling and Sugarscape models), and show that model outcomes are affected by the spatial configuration not only quantitatively in a considerable way, but also qualitatively in terms of behavior of model phase diagram. Raimbault (2020) shows that the SimpopNet model introduced by Schmitt (2014) for the co-evolution of cities and transportation networks is highly sensitive both to initial population distribution across cities and to the initial transportation network structure.

### 4 Discussion

Beyond the direct application of the library to study the spatial sensitivity of geosimulation models, several developments can be considered. The inclusion of network and vector generation methods into OpenMOLE is currently explored, but remains not straightforward in particular because of the constraint to represent workflow prototypes as primary data structures, to ensure interoperability when embedding different models and languages. More detailed and operational transportation network capabilities are also currently being implemented into the library, including multi-modal

transportation network computation and accessibility computation. Specific methods tailored for the validation of Land-use Transport Models are elaborated, such as correlated noise perturbation across different layers (coupling population and employment for example), or transportation infrastructure development scenarios. The strong coupling of generators into co-evolutive models such as done by Raimbault (2019b) is being more thoroughly investigate in order to provide such coupled generators as primitives. This library and its integration with the OpenMOLE software should thus foster the development of more thorough geosimulation models validation practices, and therein strengthen the confidence in the results obtained with such models.

## References

- 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.
- Barrett, C. L., Beckman, R. J., Khan, M., Anil Kumar, V., Marathe, M. V., Stretz, P. E., Dutta, T., and Lewis, B. (2009). Generation and analysis of large synthetic social contact networks. In *Winter Simulation Conference*, pages 1003–1014. Winter Simulation Conference.
- Benenson, I. and Torrens, P. (2004). *Geosimulation: Automata-based modeling of urban phenomena*. John Wiley & Sons.
- Cheng, T. and Adepeju, M. (2014). Modifiable temporal unit problem (mtup) and its effect on space-time cluster detection. *PloS one*, 9(6):e100465.
- Horni, A., Nagel, K., and Axhausen, K. W. (2016). *The multi-agent transport simulation MATSim*. Ubiquity Press London.
- Kwan, M.-P. (2012). The uncertain geographic context problem. *Annals of the Association of American Geographers*, 102(5):958–968.
- Nikolai, C. and Madey, G. (2009). Tools of the trade: A survey of various agent based modeling platforms. *Journal of Artificial Societies and Social Simulation*, 12(2):2.
- North, M. J., Collier, N. T., Ozik, J., Tataru, E. R., Macal, C. M., Bragen, M., and Sydelko, P. (2013). Complex adaptive systems modeling with repast simphony. *Complex adaptive systems modeling*, 1(1):3.
- Openshaw, S. (1984). The modifiable areal unit problem. *Concepts and techniques in modern geography*.
- Passerat-Palmbach, J., Reuillon, R., Leclaire, M., Makropoulos, A., Robinson, E. C., Parisot, S., and Rueckert, D. (2017). Reproducible large-scale neuroimaging studies with the openmole workflow management system. *Frontiers in neuroinformatics*, 11:21.
- Penn, A. (2006). Synthetic networks-spatial, social, structural and computational. *BT technology journal*, 24(3):49–56.

- Raimbault, J. (2018a). Calibration of a density-based model of urban morphogenesis. *PloS one*, 13(9):e0203516.
- Raimbault, J. (2018b). Multi-modeling the morphogenesis of transportation networks. In *Artificial Life Conference Proceedings*, pages 382–383. MIT Press.
- Raimbault, J. (2019a). Second-order control of complex systems with correlated synthetic data. *Complex Adaptive Systems Modeling*, 7(1):1–19.
- Raimbault, J. (2019b). 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.
- Raimbault, J., Cottineau, C., Le Texier, M., Le Nechet, F., and Reuillon, R. (2019). Space matters: Extending sensitivity analysis to initial spatial conditions in geosimulation models. *Journal of Artificial Societies and Social Simulation*, 22(4):10.
- Raimbault, J. and Perret, J. (2019). Generating urban morphologies at large scales. *Artificial Life Conference Proceedings*, (31):179–186.
- Reuillon, R., Leclaire, M., Raimbault, J., Arduin, H., Chapron, P., Chérel, G., Delay, E., Lavallée, P.-F., Passerat-Palmbach, J., Peigne, P., et al. (2019). Fostering the use of methods for geosimulation models sensitivity analysis and validation. In *European Colloquium on Theoretical and Quantitative Geography 2019*.
- 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.
- Richiardi, M. G. and Richardson, R. E. (2017). Jas-mine: A new platform for microsimulation and agent-based modelling. *International Journal of Microsimulation*, 10(1):106–134.
- Saltelli, A., Tarantola, S., Campolongo, F., and Ratto, M. (2004). Sensitivity analysis in practice: a guide to assessing scientific models. *Chichester, England*.
- Schmitt, C. (2014). *Modélisation de la dynamique des systèmes de peuplement: de SimpopLocal à SimpopNet*. PhD thesis, Université Panthéon-Sorbonne-Paris I.
- Singh, A., Vainchtein, D., and Weiss, H. (2007). Schelling’s segregation model: Parameters, scaling, and aggregation. *arXiv preprint arXiv:0711.2212*.
- Smith, D. M., Clarke, G. P., and Harland, K. (2009). Improving the synthetic data generation process in spatial microsimulation models. *Environment and Planning A*, 41(5):1251–1268.
- Thomas, I., Jones, J., Caruso, G., and Gerber, P. (2018). City delineation in european applications of luti models: review and tests. *Transport Reviews*, 38(1):6–32.