Extending methods to explore spatial models of simulation

Postdoc project

JUSTE RAIMBAULT[1,2]
[1] UPS CNRS 3611 ISC-PIF
[2] UMR CNRS 8504 Géographie-cités

1 Introduction

The study of complexity has recently observed stunning developments, in particular with the use of modeling and simulation. According to [Rey-Coyrehourcq, 2015], the example of theoretical and quantitative geography is a perfect illustration of how methodological, technical, empirical and theoretical advances necessarily strongly bind together: the use of computation centers in the seventies would be comparable to the current democratisation of grid computing. [Raimbault, 2017] conceptualizes this concept within a knowledge framework, built by induction from the case study of the genesis of the Evolutive Urban Theory developed for twenty years to study systems of cities [Pumain, 1997]. This geographical knowledge is in particular tied with considerable methodological and epistemological innovations.

Indeed, the development of the OpenMole software for model exploration [Reuillon et al., 2013] is accompanied by a progressive shift in simulation practices. According to REUILLON (interview cited in [Raimbault, 2017]), three fundamental innovative axis distinguish this new philosophy and technology compared to existing approaches in simulation: (i) the embedding of models within workflows, making model coupling and multi-modeling easier; (ii) the provision of novel heuristic methods for model exploration; and (iii) the transparent access to various intensive computation infrastructures. [Banos, 2017] emphasizes how this "knowledge accelerator" favors the construction of a robust and experimental social science, by the introduction of tools to deal with main requirements for it coined out by [Banos, 2013]: multiple heterogenous models can be compared and coupled in an interdisciplinar approach within a new incremental methodology introduced by [Cottineau et al., 2015], models and workflows are open to ensure reproducibility, the behavior of models is better known with specific methods such as the Pattern Space Exploration algorithm developed by [Chérel et al., 2015] that provide the output feasible space of a model or the Calibration Profile algorithm [Reuillon et al., 2015], multi-objective approaches to model optimization are implemented in genetic algorithms for model calibration [Schmitt et al., 2015].

Several open issues remain in the development of these new paradigms, in particular for the application to spatial models. Following [Varenne, 2017], epistemological innovations have in history largely been induced by the inclusion of space in simulation models. In other words, the spatial aspect has implications on the nature of knowledge itself beyond the sole models. Indeed, in the case of territorial systems, the multi-scalar and non-stationary nature of processes makes multi-modeling and model coupling highly relevant, what in return raises the open question of model parsimony [Pumain and Reuillon, 2017b]. In a similar vein, sensitivity analysis methods for spatial models are surprisingly underdeveloped as [Cottineau et al., 2017] shows.

This research project aims at contributing to these open issues in the context of the new epistemology for model exploration described above. We describe below the different research axis.

2 Project

2.1 The interaction of complementary research axis

This project has the general goal of improving the way to extract knowledge from spatial simulation models. More precisely, it is structured around three complementary axis that tackle some of the open issues given above. This axis are, in an arbitrary order :

- the development of spatial synthetic data generation methods [Raimbault, 2016], and of associated sensitivity analysis protocols and methods [Cottineau et al., 2017];
- the development and benchmarking of heuristics for model exploration;
- the investigation of how to consider model structure in the evaluation of simulation models, or how to avoid overfitting in multi-modeling practices [Raimbault, 2018].

The insertion within the general dynamic of the OpenMole project, and elements of answer to the overall problematic, arise as much from the *interaction* between each research axis than from each axis itself :

- At the interface of synthetic data generation and of the study of the model structure lies crucial issues on model coupling and meta-modeling: for example, testing the sensitivity of a given model to the spatial structure of data may imply coupling it with an upstream model to generate this structure. This naturally increases the number of parameters and may change the nature of calibration problems. These questions relate to OpenMole's philosophy of embedding model within workflows.
- This last point is also behind the relation between methods benchmarks and model structure, since at the crossroad of these axis we are confronted to establish types or classes of models to be tested: are there particular type of problems more suited to particular heuristic methods ?
- Finally, between synthetic data and methods benchmarks lies the development of methods for the exploration of simulation models, which is one other cornerstone of OpenMole's approach.

Our project could not necessary focus of spatial models, however these will be of particular interest for at least two reasons: (i) methods to generate synthetic data are still underdeveloped on the spatial character; and (ii) a large number of spatial simulation models remain unexplored.

2.2 Possible models to be studied

Several models may be of a particular interest as case studies to explore our research axis.

Spatial large models From a technical viewpoint, a significant challenge remains in embedding models that are relatively large (in terms of conceptual complexity and of practical implementation) and computationally costly. The following models have this property and show interesting thematic features.

- Dynex forcity: its exploration furthermore challenges the idea of "models as black boxes", or how informative exploration heuristics can be to understand model behavior and structure at an aggregated level, what can also been put within a data-driven perspective.
- H24 mobility epidemiological model: synthetic spatial data is implied at different stages of this model.
- The Lutecia model [Le Néchet and Raimbault, 2015] is a co-evolution model for transportation network and urban activities that takes into account network governance processes and is relatively computationally greedy.

Synthetic data generation The following models are example of recently studied models that can be used as synthetic spatial data generation methods.

- Density generation [Raimbault, 2018].
- Network generation [Raimbault, 2018b].
- Correlated synthetic data [Raimbault, 2016].

Model structure and over-fitting A particular interest can be found in comparing models of growth for systems of cities based on spatial interactions, since these have a very similar structure but include different processes. Their integration in a multi-modeling paradigm can shed light on methods and measures of model performance that take into account model structure. Such models include for example:

- Interaction model [Raimbault, 2016] and co-evolution model [Raimbault, 2018a]
- the Marius model family [Cottineau et al., 2015]
- the Favaro-Pumain model for the diffusion of innovation [Favaro and Pumain, 2011]

2.3 Specific issues to explore

2.3.1 Benchmark of methods

In [Pumain and Reuillon, 2017a], a new strategy to evaluate noisy fitness function in the standard NSGA-II genetic algorithm is introduced. The principle is to constantly reevaluate individuals at a fixed rate, and to add as an objective in the optimization the number of evaluation. Several open questions remain on that innovative methodology:

- Does this method perform better than existing strategies in terms of trade-off between convergence and computational burden, on what type of optimization problems ?
- How does the form of the added objective influences the performance ?
- What is the influence of different types of noises, and of non-stationary noises ?

The detailed study of this strategy for noisy fitness function will be the first priority of the benchmark axis. Beside, the sensitivity to the choice of mutation, crossover and selection operators will be important to look at.

2.3.2 Spatial sensitivity

The axis of spatial sensitivity methods is particularly prone to include an open problem in the exploration of simulation models, which is the reconstruction of the inverse image of a given set in the indicator space. Morphological classes of cities are well established objects in the literature [Le Néchet, 2015][Raimbault, 2018], and their sensitivity to equifinality is interesting from a thematic point of view: given some generative models, which processes or set of parameters do produce the same type of cities ? This axis thus raises this methodological question, which can be explored conjointly within the method development axis.

3 Organisation

3.1 Integration within Dynamicity

Two research axis are direct contributions to the issues tackled within Dynamicity: the benchmark of methods and the development of spatial sensitivity methods. The last axis relates to the broader multi-modeling context of OpenMole.

3.2 Deliverables

- Any method developed or tested on a closed model (that should be as few as possible, ideally only Dynex) should be similarly applied to a similar open case.
- Most methods developed should be integrated as OpenMole plugins.

References

[Banos, 2013] Banos, A. (2013). Pour des pratiques de modélisation et de simulation libérées en géographies et shs. *HDR. Université Paris*, 1.

[Banos, 2017] Banos, A. (2017). Knowledge accelerator' in geography and social sciences: Further and faster, but also deeper and wider. In *Urban Dynamics and Simulation Models*, pages 119–123. Springer.

[Chérel et al., 2015] Chérel, G., Cottineau, C., and Reuillon, R. (2015). Beyond corroboration: Strengthening model validation by looking for unexpected patterns. *PLoS ONE*, 10(9):e0138212.

[Cottineau et al., 2017] Cottineau, C., Raimbault, J., Le Texier, M., Le Nechet, F., and Reuillon, R. (2017). Initial spatial conditions in simulation models: the missing leg of sensitivity analyses? In *Geocomputation Conference*.

[Cottineau et al., 2015] Cottineau, C., Reuillon, R., Chapron, P., Rey-Coyrehourcq, S., and Pumain, D. (2015). A modular modelling framework for hypotheses testing in the simulation of urbanisation. *Systems*, 3(4):348–377.

[Favaro and Pumain, 2011] Favaro, J.-M. and Pumain, D. (2011). Gibrat revisited: An urban growth model incorporating spatial interaction and innovation cycles. *Geographical Analysis*, 43(3):261–286.

[Le Néchet, 2015] 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*.

[Le Néchet and Raimbault, 2015] Le Néchet, F. and Raimbault, J. (2015). Modeling the emergence of metropolitan transport autorithy in a polycentric urban region. In *European Colloqueum on Theoretical and Quantitative Geography*, Bari, Italy.

[Pumain, 1997] Pumain, D. (1997). Pour une théorie évolutive des villes. *Espace géographique*, 26(2):119–134.

[Pumain and Reuillon, 2017a] Pumain, D. and Reuillon, R. (2017a). Evaluation of the simpoplocal model. In *Urban Dynamics and Simulation Models*, pages 37–56. Springer.

[Pumain and Reuillon, 2017b] Pumain, D. and Reuillon, R. (2017b). *Urban Dynamics and Simulation Models*. Springer International.

[Raimbault, 2016] Raimbault, J. (2016). Generation of correlated synthetic data. In *Actes des Journées de Rochebrune 2016*.

[Raimbault, 2017] Raimbault, J. (2017). An applied knowledge framework to study complex systems. In *Complex Systems Design & Management*, pages 31–45.

[Raimbault, 2018] Raimbault, J. (2018). Calibration of a density-based model of urban morphogenesis. *PLoS ONE, in revision*.

[Raimbault, 2018] Raimbault, J. (2018). Indirect evidence of network effects in a system of cities. *Environment and Planning B, in revision.*

[Raimbault, 2018a] Raimbault, J. (2018a). Models for the co-evolution of cities and networks. In *forthcoming in Handbook on Cities and Networks, Rozenblat C., Neal Z., eds.*

[Raimbault, 2018b] Raimbault, J. (2018b). An urban morphogenesis model capturing interactions between networks and territories. In *forthcoming in Mathematics of Urban Morphology. D'Acci L., ed. Springer Nature - Birkhäuser Mathematics.*

[Reuillon et al., 2013] 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.

[Reuillon et al., 2015] Reuillon, R., Schmitt, C., De Aldama, R., and Mouret, J.-B. (2015). A new method to evaluate simulation models: The calibration profile (cp) algorithm. *Journal of Artificial Societies and Social Simulation*, 18(1):12.

[Rey-Coyrehourcq, 2015] Rey-Coyrehourcq, S. (2015). *Une plateforme intégrée pour la construction et l'évaluation de modèles de simulation en géographie.* PhD thesis, Université Paris 1 Panthéon-Sorbonne.

[Schmitt et al., 2015] Schmitt, C., Rey-Coyrehourcq, S., Reuillon, R., and Pumain, D. (2015). Half a billion simulations: Evolutionary algorithms and distributed computing for calibrating the simpoplocal geographical model. *Environment and Planning B: Planning and Design*, 42(2):300–315.

[Varenne, 2017] Varenne, F. (2017). *Théories et modèles en sciences humaines. Le cas de la géographie.* Editions Matériologiques.