Extracting knowledge from simulation models: trends and perspectives from the viewpoint of quantitative geography

J. Raimbault[1,2,*]
juste.raimbault@polytechnique.edu

[1]Complex Systems Institute, Paris, UPS CNRS 3611 ISC-PIF
[2]UMR CNRS 8504 Géographie-cités

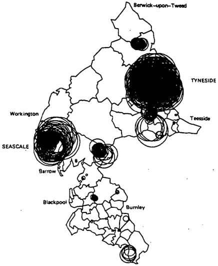
CCS 2018

Satellite Methods and Epistemologies of Simulation

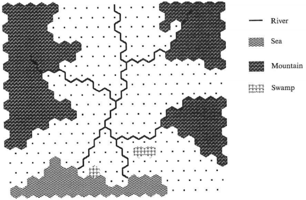
Thessaloniki

September 26th 2018

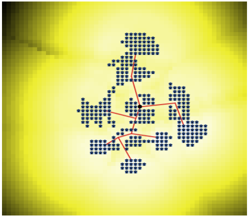
A long history of simulation in TQG



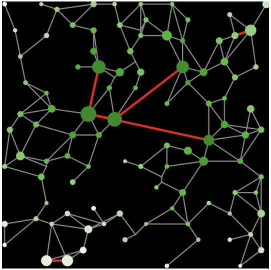
Geographic analysis machine [Openshaw et al., 1987]



Simpop 1 model [Sanders et al., 1997]



Hybrid urban morphogenesis [Raimbault et al., 2014]



SimpopNet model [Schmitt, 2014]

Current trends and challenges in (geo-)simulation

Key domains: quantifying urban growth and form, mining spatio-temporal data, geosimulation, multi-scalar approaches [Behnisch and Meinel, 2018].

Challenges:

- [Perez et al., 2016] key challenges in ABM for planning: addressing complexity in a clean way, addressing multi-dimensionality, feasible trajectories, participatory planning.
- Simulation models [Banos, 2013]: interdiscplinarity, data-driven models, exploration of models, multi-objective issues, reproducibility and reuse of models, coupling models.

Future ? [Banos, 2017] deeper and integrated knowledge; [Arribas-Bel and Reades, 2018] new geographic data science?

OpenMole's positioning

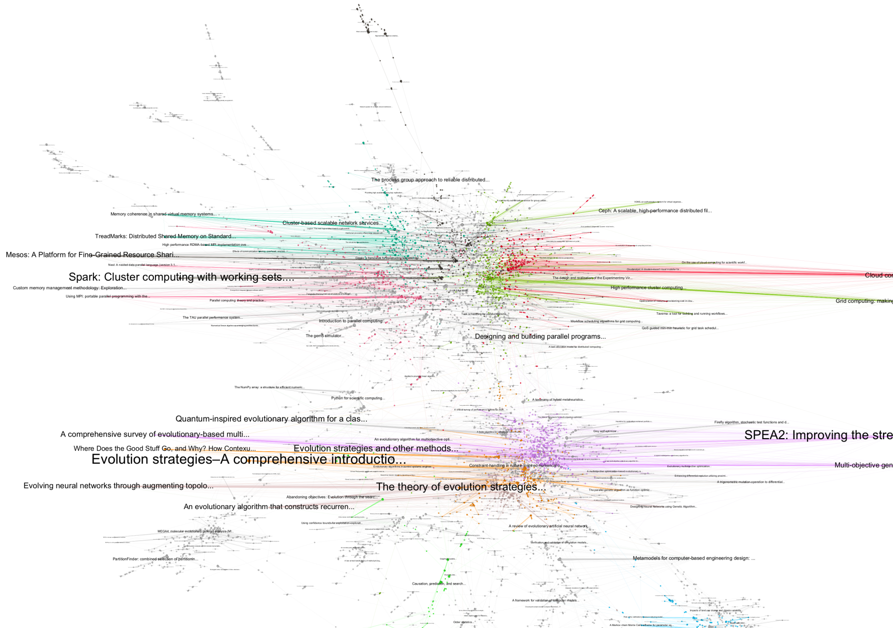
A qualitative shift in knowledge that can be extracted from a simulation model with model exploration methods.

Success stories: SimpopLocal [Schmitt et al., 2015], Marius [Chérel et al., 2015], Ecological modeling [Lavallée et al., 2018], epidemiology [Arduin, 2018], etc.

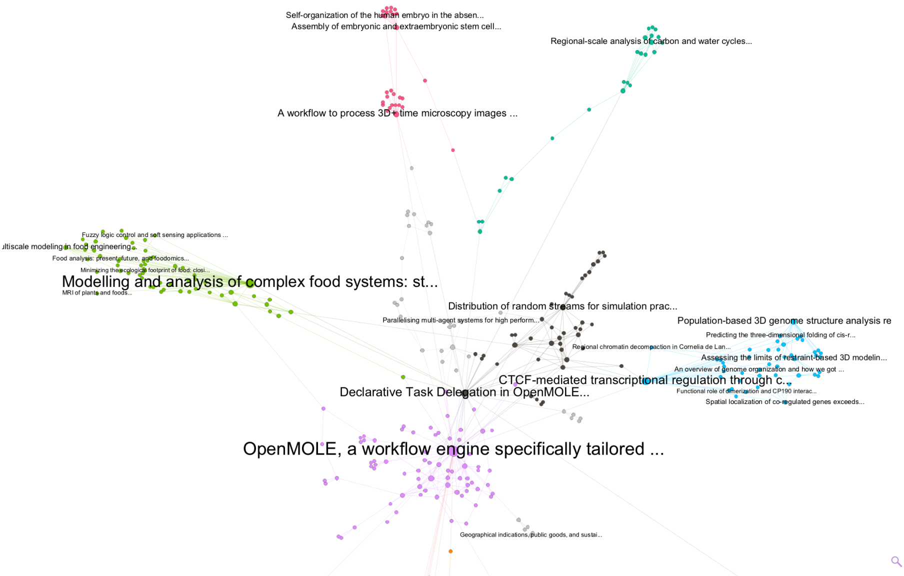
Key features:

- Unique role of complementary axis of computation environment access, methods providing, and model embedding.
- Iterative and integrated construction of models and theories, using all dimensions of knowledge enhanced by simulation and computation (modeling, theory, empirical, data, methods, tools [Raimbault, 2017].
- Coupling models and reproducibility at the core of the workflow approach [Passerat-Palmbach et al., 2017]

Citation network analysis (full)



Citation network analysis (OpenMOLE)



Some of crucial issues :

Generic issues:

- Overfitting in simulation models
- Model coupling
- Direct and inverse mapping
- Stochasticity

Specific issues to spatio-temporal systems:

- Spatio-temporal non-stationarity
- Spatio-temporal synthetic data

Dealing with overfitting

When do additional parameters actually capture new dimensions of the simulated system ? i.e. when does fit improvement is not only due to more degree of freedom ?

$\rightarrow$ Crucial for parcimonious models, which can be used then as building bricks for more complex models.

- Black-box brutal data explanation ?
- Extension of AIC-type measures ?
- Multi-objective optimisation with degrees of freedom ?

Model coupling

Definition/theory/quantification of model coupling ?

$\rightarrow$ Crucial for interdisciplinarity, reproducibility and the reuse of models; crucial for multi-scalar approaches (downward causation); crucial for model benchmarking.

- Model-independent notion of "coupling-strength" ?
- Covariance structures ?
- Causal graphs ?

Direct and inverse mapping

How to have a comprehensive overview of strongly non-linear simulation models mapping ?

$\rightarrow$ Feasible space and unexpected patterns (PSE algorithm [Chérel et al., 2015]); dealing with equifinality.

- Inverse problem heuristics being currently tested in OpenMOLE
- Iterative approach to determine main patterns in the mapping ?

Handling stochasticity

How to handle stochastic models during genetic algorithm calibration or exploration ?

$\rightarrow$ Crucial for more data-driven and "real-world" models; crucial for a robust knowledge extracted from simulation models.

- Deal with "real-world" noise patterns
- An embedding approach tested in OpenMOLE to deal with noisy fitness
- A bayesian approach to calibration (ABC) also currently tested

Spatio-temporal modeling issues: non-stationarity

How to understand and include spatio-temporal non stationarity in empirical analysis / in model simulations and calibration ?

→ Intrinsic complexity of spatial systems; crucial for multi-scalar approaches.

- Link between non-stationarity and non-ergodicity
- Link between ABM approaches and dynamical systems approaches, towards hybrid approaches [Banos et al., 2015]

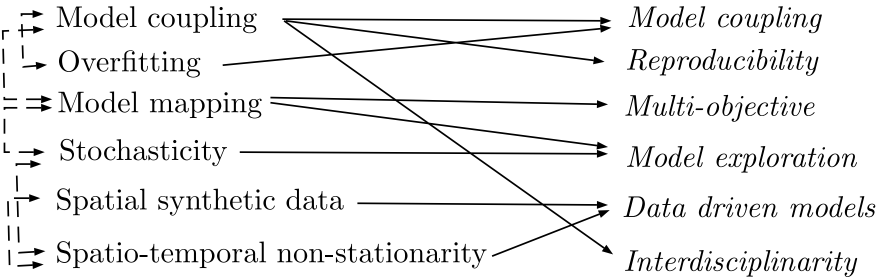
Spatio-temporal synthetic data

How to extend sensitivity analyses to spatial initial conditions ? How to generate spatial synthetic data ?

$\rightarrow$ Crucial for the robustness of spatial simulations

- First work in the case of density grids in territorial systems [Cottineau et al., 2017]
- Extend to other disciplines: ecology, geosciences
- Towards a generic library integrated in OpenMOLE

An integrated view



Applied perspectivism to couple modeling approaches

Giere's perspectivism [Giere, 2010]: a "third way" beyond constructivism-realism: *Any scientific knowledge construction process as a perspective by an agent to answer a purpose with a media (model).*

Applied knowledge framework proposed by [Raimbault, 2017] to study [complex] systems: *co-evolution* of cognitive agents and knowledge domains, through the intermediate of perspectives.

Applied perspectivism principles:

- Foster consistence of perspectives and their communication (Banos' virtuous spiral between disciplinarity and interdisciplinarity)
- Importance of reflexivity to ease the coupling of perspectives
- New model exploration methods increase the integration between knowledge domains
- Coupling of models as a possible medium to couple perspectives (transfer hypothesis)

Still to be formalized, specified as possible implementations, and experimented.

Conclusion

Significant accomplishments beyond disciplines, construction of new research practices (see the satellite presentations) → still much to do ? (e.g. how to put into practice ? how to achieve true integration ? etc.)

"La route est longue mais la voie est libre"

You need OpenMOLE and OpenMOLE needs you ! (win-win interdisciplinary relations): apply to the summer school !
`https://exmodelo.org/`

Related works on epistemological considerations

Raimbault, J. (2017, December). An Applied Knowledge Framework to Study Complex Systems. In Complex Systems Design & Management (pp. 31-45). arXiv:1706.09244.

Raimbault, J. (2018). Caractérisation et modélisation de la co-évolution des réseaux de transport et des territoires (Doctoral dissertation, Université Paris 7 Denis Diderot). `https://halshs.archives-ouvertes.fr/tel-01857741`

Reserve Slides

References I

Arduin, H. (2018).
Modélisation mathématique des interactions entre pathogènes chez l'hôte humain: Application aux virus de la grippe et au pneumocoque.
PhD thesis, Université Paris-Saclay.

Arribas-Bel, D. and Reades, J. (2018).
Geography and computers: Past, present, and future.
Geography Compass, page e12403.

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

References II

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.

Banos, A., Corson, N., Gaudou, B., Laperrière, V., and Coyrehourcq, S. R. (2015).
The importance of being hybrid for spatial epidemic models: a multi-scale approach.
Systems, 3(4):309–329.

Behnisch, M. and Meinel, G. (2018).
Trends in Spatial Analysis and Modelling.
Springer.

References III

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

Cottineau, C., Raimbault, J., Le Texier, M., Le Néchet, F., and Reuillon, R. (2017).
Initial spatial conditions in simulation models: the missing leg of sensitivity analyses?
In *2017 International Conference on GeoComputation: Celebrating 21 Years of GeoComputation*.

Giere, R. N. (2010).
Scientific perspectivism.
University of Chicago Press.

Lavallée, F., Alvarez, I., Dommanget, F., Martin, S., Reineking, B., and Smadi, C. (2018).
A dynamical model for the growth of a stand of japanese knotweed including mowing as a management technique.
In *Conference on Complex Systems 2018*.

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

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.

References V

Perez, P., Banos, A., and Pettit, C. (2016).
Agent-based modelling for urban planning current limitations and future trends.
In *International Workshop on Agent Based Modelling of Urban Systems*, pages 60–69. Springer.

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

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.

References VI

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.

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.