

# Calibration of a Density-based Model of Urban Morphogenesis

Raimbault Juste<sup>1,2,\*</sup>

**1 UMR CNRS 8504 Géographie-cités, Paris, France**

**2 UMR-T 9403 IFSTTAR LVMT, Champs-sur-Marne, France**

**\* Corresponding Author**

**Email : [juste.raimbault@polytechnique.edu](mailto:juste.raimbault@polytechnique.edu)**

## Abstract

We propose a stochastic model of urban growth that generates spatial distributions of population densities, at an intermediate scale between economic models at the macro scale and land-use evolution models focusing on local relations. Integrating simply the two opposite key processes of aggregation (“preferential attachment”) and diffusion (urban sprawl), we show that we can capture the whole spectrum of existing urban forms in Europe. An extensive exploration and calibration of the proposed model allows determining the region of parameter space corresponding morphologically to observed european urban systems, providing an validated thematic interpretation to model parameters, and furthermore determining the effective dimension of the urban system at this scale regarding morphological objectives.

## Introduction

Urban Systems are complex socio-technical systems which can be studied from a large variety of viewpoints and disciplines: Batty has advocated in that sense for the construction of a dedicated science defined by its objects of study more than the methods used [1]. Simulating urban growth is one typical aspect As the short term trend is towards a mostly urbanized

[2]

Macro models (Gibrat and extensions) are partially spatialized (Favaro - Pumain includes distance between cities, Simpop/Marius also) whereas micro (most CA) are ‘over-spatialized’ (neighborhood-based). [3] [4]

[5]

[6] propose a micro-based model of urban growth, with the purpose to replace non-interpretable physical mechanisms with agent mechanisms, including interactions forces and mobility choices. Local correlations are used in [7], which develops the model introduced in [8], to modulate growth patterns to resemble real configurations.

[9] morphogenesis for roads

Cellular automata [10] [11]

[12]

[13]

[14]

[15]

In the same spirit, our model situates at similar scales and can be qualified as a morphogenesis model. What about a simple model at the meso scale, close to macro by the stylized facts (preferential attachment) but producing morphologically consistent spatial distribution of densities ?

The rest of this paper is organized as follows: we first describe formally the model and the morphological indicators; we then develop results of real morphological measures, model exploration and model calibration.

## Material and Methods

### Urban growth model

**Rationale** [16] : how Simon model generates power law (paper more general to be quoted ?) ; first mover : path dependency of obtained shapes.

Our model is based on widely accepted ideas of diffusion-aggregation processes for Urban Processes. The combination of attraction forces with repulsion, due for example to congestion, already yield a complex outcome that has been shown under some simpli-

fying assumptions to be representative of urban growth processes. A model capturing  
 these processes was introduced by [17]. Indeed, the tension between antagonist aggre-  
 gation and sprawl mechanisms may be an important process in urban morphogenesis.  
 For example [18] opposes centrifugal forces with centripetal forces in the equilibrium  
 view of urban spatial systems, what is easily transferable to non-equilibrium systems in  
 the framework of self-organized complexity : a urban structure is a far-from-equilibrium  
 system that has been driven to this point by these opposite forces. The two contra-  
 dictory processes of urban concentration and urban sprawl are captured by the model,  
 what allows to reproduce with a good precision a large number of existing morphologies.

The question at which scale is it possible and relevant to define and try to simulate  
 urban form is rather open, and will in fact depend on which issues are being tackled.

choice of scale : linked to the notion of morphogenesis ; choice of a scale at which  
 spatial form has a sense but city system is “wide enough” : 50-100km, meso-scale ?

**Formalization** We formalize now the model, together with its parameters and their  
 possible interpretations. The simulation model proceeds iteratively the following way.  
 The world is a square grid of width  $N$ , in which each cell is characterized by its pop-  
 ulation  $(P_i(t))_{1 \leq i \leq N^2}$ . initially empty, is represented . At each time step, until total  
 population reaches a fixed parameter  $P_m$ ,

- total population is increased of a fixed number  $N_G$  (growth rate), following a  
 preferential attachment such that

$$\mathbb{P}[P_i(t+1) = P_i(t) + 1 | P(t+1) = P(t) + 1] = \frac{(P_i(t)/P(t))^\alpha}{\sum (P_i(t)/P(t))^\alpha} \quad (1)$$

- a fraction  $\beta$  of population is diffused to cell neighborhood, this operation being  
 repeated  $n_d$  times

**Indicators** [19] : sort of morphological analysis

As our model is only density-based, we propose to quantify its outputs through spa-  
 tial morphology, i.e. characteristics of density spatial distribution. We need therefore  
 quantities having a certain level of robustness and invariance. For example, two poly-  
 centric cities should be classified as morphologically close whereas a direct comparison

of distributions (Earth Mover Distance e.g.) could give a very high distance between configurations depending on center positions. To tackle this issue, we refer to the Urban Morphology Analysis literature which proposes an extensive set of indicators to describe urban form [20]. The number of dimensions can be reduced to obtain a robust description with a few number of independent indicators [21]. For the choice of indicators, we follow the analysis done in [22] where a typology of large European cities is obtained in consistence with qualitative knowledge. Let denote  $(P_i)_{1 \leq i \leq N}$  the population of cells, sorted in decreasing order,  $d_{ij}$  the distance between cells  $i, j$ , and  $P = \sum_{i=1}^N P_i$  total population. The indicators are the following :

1. Rank-size slope  $\gamma$ , expressing the degree of hierarchy in the distribution, computed by fitting with Ordinary Least Squares a power law distribution by  $\ln P_i/P_0 \sim k - \gamma \cdot \ln i/i_0$ .

2. Distribution Entropy

$$\mathcal{E} = \sum_{i=1}^N \frac{P_i}{P} \cdot \ln \frac{P_i}{P} \quad (2)$$

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

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

4. Mean distance between individuals, which captures population concentration

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

## Real Data

We compute morphological measure on real urban density data, namely the population density grid of the European Union at 100m resolution provided openly by Eurostat [23]. The morphological measures used for calibration are the one described above that are the same used to classify model outputs. The calibration of the model is thus done on morphological objectives (entropy, hierarchy, spatial auto-correlation, mean distance). We show in Fig. 1 maps giving values of indicators for France, to ease readability. Maps for the full European union are available in S1 Text. The choice of the resolution, the

space range, and the shape of the window on which indicators are computed, is made according to the thematic specifications of the model : We however tested the sensitivity to window size and shape

## Results

### Generation of urban patterns

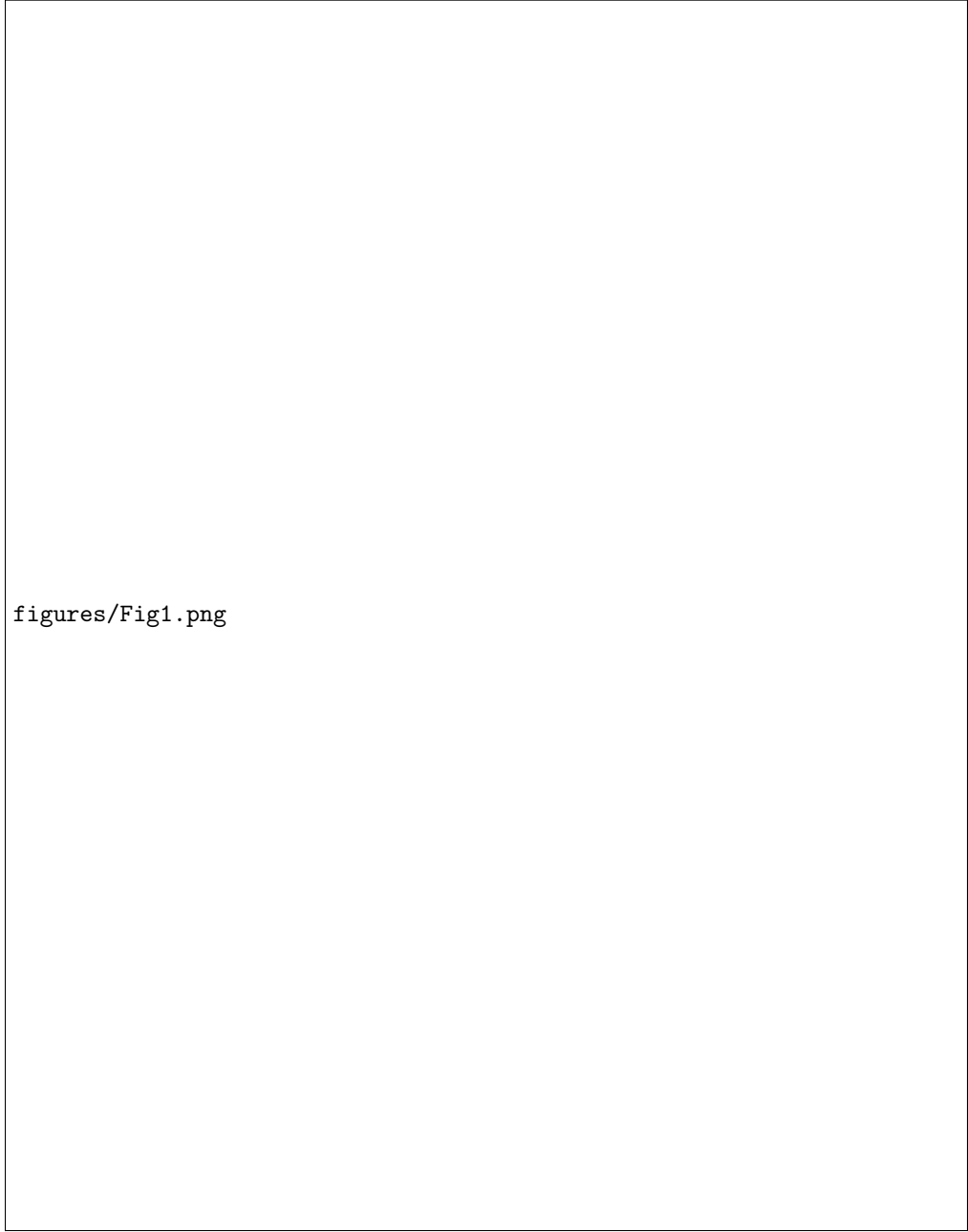
**Implementation** The model is implemented both in NetLogo [24] for exploration and visualization purposes, and in **Scala** for performance reasons and easy integration into OpenMole [25], which allows a transparent access to High Performance Computing environments. Computation of indicator values on geographical data is done in **R** using the **raster** package [26]. Source code and results are available on the open repository of the project at <https://github.com/JusteRaimbault/CityNetwork/tree/master/Models/Synthesis>. Raw datasets for real indicator values and simulation results are available on Dataverse at <http://dx.doi.org/10.7910/DVN/WSUSBA>.

**Generated Shapes** The model has a relatively small number of parameters but is able to generate a large variety of shapes, extending beyond existing forms. In particular, its dynamical nature allows through  $P_m$  parameter to choose between final configurations that can be non-stationarity or stationarity. Fig. 2 shows examples of generated shapes for different parameter values and different regimes.

### Model Behavior

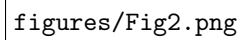
In the study of such a computational model of simulation, the lack of analytical tractability must be compensated by an extensive knowledge of model behavior in the parameter space [?]. This type of approach is typical of what Arthur calls the *Computational shift in modern science* [?].

**Convergence** First of all we need to assess the convergence of the model and its behavior regarding stochasticity. We run for a sparse grid of the parameter space consisting of 81 points, with 100 repetitions for each point. Corresponding histograms are shown in S1 Text. Indicators show good convergence properties: most of indicators



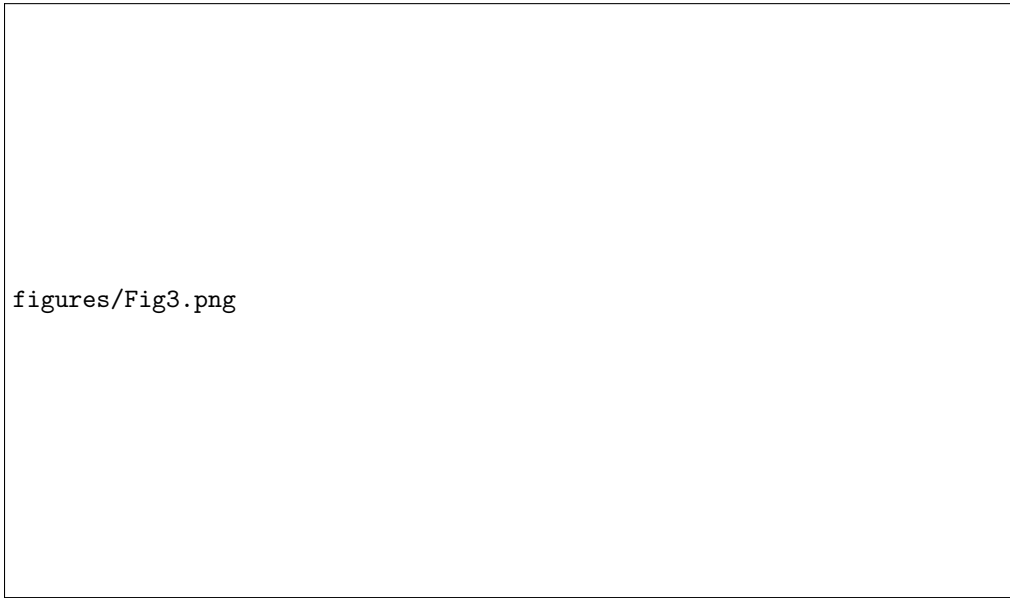
figures/Fig1.png

**Figure 1. Empirical values of morphological indicators.**



figures/Fig2.png

**Figure 2. Example of the variety of generated urban shapes.**



figures/Fig3.png

**Figure 3. Behavior of indicators.**


are easily statistically discernable across parameter points, and these are distinguished without ambiguity when taking into account all indicators. We use this experiment to find a reasonable number of repetitions needed in larger experiments. For each point, we estimate the Sharpe ratios for each indicators as

**Exploration of parameter space** We sample the Parameter space using a Latin Hypercube Sampling, with parameter as  $\alpha \in [0.1, 4]$ ,  $\beta \in [0, 0.5]$ ,  $n_d \in \{1, \dots, 5\}$ ,  $N_G \in [500, 30000]$ ,  $P_m \in [1e4, 1e6]$ . This type of cribbing is a good compromise to have a reasonable sampling without being subject to the dimensionality curse within normal computation capabilities.

## Semi-analytical Analysis

Our model can be understood as a type of reaction-diffusion model, that have been widely used in other fields such as biology: similar processes were used for example by Turing in its seminal paper on morphogenesis [27]. An other way to formulate the model typical to these approaches is by using Partial Differential Equations. We propose to gain insights into long-time dynamics by studying them on a simplified case. We consider the system in one dimension, such that  $x \in [0, 1]$  with  $1/\delta x$  cells of size  $\delta x$ . Each cell is characterized by its population as a random variable  $P(x, t)$ . We work





figures/Fig4.png

**Figure 4. Randomness and frozen accidents.**

on their expected values  $p(x, t) = \mathbb{E}[P(x, t)]$ , and assume that  $n_d = 1$ . As developed  
in Supplementary Material S2 Text, we show that this simplified process verifies the  
following PDE:

$$\delta t \cdot \frac{\partial p}{\partial t} = \frac{N_G \cdot p^\alpha}{P_\alpha(t)} + \frac{\alpha \beta (\alpha - 1) \delta x^2}{2} \cdot \frac{N_G \cdot p^{\alpha-2}}{P_\alpha(t)} \cdot \left( \frac{\partial p}{\partial x} \right)^2 + \frac{\beta \delta x^2}{2} \cdot \frac{\partial^2 p}{\partial x^2} \cdot \left[ 1 + \alpha \frac{N_G p^{\alpha-1}}{P_\alpha(t)} \right] \quad (3)$$

Assuming an asymptotic separable solution of the form  $p(x, t) = f(t) \cdot p(x)$  implies  
that  $f(t) =$ .

This section allow us to learn the following properties, that are important to better  
understand the previous simulation results

1. Existence of a stationary solution for population proportion
2. Characteristic distance of the stationary distribution
3. Existence of bifurcations in the random case

**Figure 5. Calibration of the model.** The principal component analysis is conducted to maximize the spread of the differences between real data and model output, i.e. on the set  $\{|R_i - M_j|\}$  where  $R_i$  is the set of real points,  $M_j$  the set of model outputs. We select then the overlapping cloud at threshold  $\theta$ , by taking models output closer to real point cloud than  $\theta$  in the (PC1,PC2) plan.

## Model Calibration

We use a specific calibration process: a principal component analysis allows to maximize the cumulated distance between generated points and reals points. We select then the point cloud that overlaps real points in the (PC1,PC2) plan, given a distance threshold. Fig. 5 shows the points we obtain for four different values of the threshold ranging from  $10^{-6}$  to  $10^{-3}$ .

## Discussion

more refined model : thresholded aggregation and diffusion (distance to center for diffusion, maximum density for aggregation)

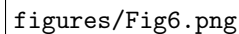
**Calibration refinement and Targeted Exploration** We plan in further work to extract the exact parameter space covering all real situations and provide interpretation of its shape (correlations between parameters). Its volume in different directions should give the relative importance of parameters.

We also use the parameter space exploration algorithm [28] implemented in OpenMole, and obtain in Fig. ?? the lower bound in Moran-entropy plan, that unexpectedly exhibit a scaling relationship that we aim to explore further.

## Integration into a multi-scale growth model

It could be possible to couple this model with a Gibrat (or Favaro-pumain) at Europa scale (macro) (with addition of consistence on migration constraints), where meso growth rates which were exogenous before are top-down determined, and bottom-up feedback is done through local aggregation level, influence importance of each area.

[29]

The figure is a scatterplot showing the relationship between Moran's I and Entropy. It contains two sets of data points: blue points representing results from LHS and red points representing results from PSE exploration. A green line indicates the lower bound. The plot area is currently blank, with only the file path 'figures/Fig6.png' visible.

**Figure 6. PSE exploration.** Scatterplots of Moran against Entropy, with blue points obtained with LHS and red with PSE exploration. Lower bound is in green.

## Conclusion

In conclusion, this first modeling step provide an accurately calibrated spatial urban growth model at the mesoscopic scale that can reproduce any European urban pattern in terms of urban form. Further work is needed for an interpretation of parameter influence and the determination of effective independent dimensions of the urban system at this scale. We will use this model for other purposes in the following.

## Supporting Information

### S1 Text

#### Extended Model Exploration.

### S2 Text

#### Semi-analytical Analysis. Analytical and numerical developments of subsection .

## References

1. Batty M. The new science of cities. Mit Press; 2013.
2. Rozenfeld HD, Rybski D, Andrade JS, Batty M, Stanley HE, Makse HA. Laws of population growth. *Proceedings of the National Academy of Sciences*. 2008;105(48):18702–18707.
3. Bretagnolle A, Mathian H, Pumain D, Rozenblat C. Long-term dynamics of European towns and cities: towards a spatial model of urban growth. *Cybergeo: European Journal of Geography*. 2000;.
4. Favaro JM, Pumain D. Gibrat Revisited: An Urban Growth Model Incorporating Spatial Interaction and Innovation Cycles. *Geographical Analysis*. 2011;43(3):261–286.
5. Gabaix X. Zipf’s law for cities: an explanation. *Quarterly journal of Economics*. 1999;p. 739–767.
6. Andersson C, Lindgren K, Rasmussen S, White R. Urban growth simulation from “first principles”. *Physical Review E*. 2002;66(2):026204.
7. Makse HA, Andrade JS, Batty M, Havlin S, Stanley HE, et al. Modeling urban growth patterns with correlated percolation. *Physical Review E*. 1998;58(6):7054.
8. Makse HA, Havlin S, Stanley H. Modelling urban growth. *Nature*. 1995;377(1912):779–782.
9. Courtat T, Gloaguen C, Douady S. Mathematics and morphogenesis of cities: A geometrical approach. *Physical Review E*. 2011;83(3):036106.
10. Ward DP, Murray AT, Phinn SR. A stochastically constrained cellular model of urban growth. *Computers, Environment and Urban Systems*. 2000;24(6):539–558.
11. Xie Y. A Generalized Model for Cellular Urban Dynamics. *Geographical Analysis*. 1996;28(4):350–373. Available from: <http://dx.doi.org/10.1111/j.1538-4632.1996.tb00940.x>.

12. Clarke KC, Gaydos LJ. Loose-coupling a cellular automaton model and GIS: long-term urban growth prediction for San Francisco and Washington/Baltimore. *International journal of geographical information science*. 1998;12(7):699–714.
13. Clarke KC, Gazulis N, Dietzel C, Goldstein NC. A decade of SLEUTHing: Lessons learned from applications of a cellular automaton land use change model. *Classics in IJGIS: twenty years of the international journal of geographical information science and systems*. 2007;p. 413–427.
14. Frankhauser P. Fractal geometry of urban patterns and their morphogenesis. *Discrete Dynamics in Nature and Society*. 1998;2(2):127–145.
15. Raimbault J, Banos A, Doursat R. A hybrid network/grid model of urban morphogenesis and optimization. In: *Proceedings of the 4th International Conference on Complex Systems and Applications (ICCSA 2014)*, June 23–26, 2014, Université de Normandie, Le Havre, France; M. A. Aziz-Alaoui, C. Bertelle, X. Z. Liu, D. Olivier, eds.: pp. 51–60.; 2014. .
16. Sheridan Dodds P, Rushing Dewhurst D, Hazlehurst FF, Van Oort CM, Mitchell L, Reagan AJ, et al. Simon’s fundamental rich-gets-richer model entails a dominant first-mover advantage. *ArXiv e-prints*. 2016 Aug;.
17. Batty M. Hierarchy in cities and city systems. In: *Hierarchy in natural and social sciences*. Springer; 2006. p. 143–168.
18. Fujita M, Thisse JF. Economics of agglomeration. *Journal of the Japanese and international economies*. 1996;10(4):339–378.
19. Guérois M, Pumain D. Built-up encroachment and the urban field: a comparison of forty European cities. *Environment and Planning A*. 2008;40(9):2186–2203.
20. Tsai YH. Quantifying urban form: compactness versus’ sprawl’. *Urban studies*. 2005;42(1):141–161.
21. Schwarz N. Urban form revisited—Selecting indicators for characterising European cities. *Landscape and Urban Planning*. 2010;96(1):29 – 47. Available from: <http://www.sciencedirect.com/science/article/pii/S0169204610000320>.

22. Le Néchet F. 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. Cybergeog: European Journal of Geography. 2015;.
23. EUROSTAT. Eurostat Geographical Data; 2014. Available from: <http://ec.europa.eu/eurostat/web/gisco/geodata/reference-data/administrative-units-statistical-units>.
24. Wilensky U. NetLogo. 1999;.
25. Reuillon R, Leclaire M, Rey-Coyrehourcq S. OpenMOLE, a workflow engine specifically tailored for the distributed exploration of simulation models. Future Generation Computer Systems. 2013;29(8):1981–1990.
26. Hijmans RJ. Geographic data analysis and modeling. 2015;.
27. Turing AM. The chemical basis of morphogenesis. Philosophical Transactions of the Royal Society of London B: Biological Sciences. 1952;237(641):37–72.
28. Chérel G, Cottineau C, Reuillon R. Beyond Corroboration: Strengthening Model Validation by Looking for Unexpected Patterns. PLoS ONE. 2015 09;10(9):e0138212. Available from: <http://dx.doi.org/10.1371/journal.pone.0138212>.
29. Zhang Z, Su S, Xiao R, Jiang D, Wu J. Identifying determinants of urban growth from a multi-scale perspective: A case study of the urban agglomeration around Hangzhou Bay, China. Applied Geography. 2013;45:193–202.