An Urban Morphogenesis Model Capturing Interactions between Networks and Territories

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

Abstract

Urban systems are composed by complex couplings of several components, and more particularly by the built environment and transportation networks. Their interactions is involved in the emergence of the urban form. We propose in this chapter to introduce an approach to urban morphology grasping both aspects and their interactions. We first define complementary measures and study their empirical values and their spatial correlations on European territorial systems. We then introduce a generative model of urban growth at a mesoscopic scale. Given a fixed exogenous growth rate, population are dispatched following a preferential attachment that depends on a potential controlled by local urban form (density, distance to network) and network measures (centralities and generalized accessibilities), and then diffused in space to capture urban sprawl. Network growth is included through a multi-modeling paradigm: implemented heuristics include biological network generation and gravity potential breakdown. The model is calibrated both at the first (measures) and second (correlations) order, the later capturing indirectly relations between networks and the urban form.

Keywords : *Urban Morphology; Road Network Topology; Spatial Correlations; Urban Morphogenesis; Reaction-diffusion; Co-evolution*

1 Introduction

1.1 Urban morphology

indicators to describe urban form Tsai (2005) at the mesoscale

Le Néchet (2009) recalls the necessity of a multi-dimensional measure of the urban form. It is possible to obtain a robust description with a small number of independant indicators by a reduction of the dimension Schwarz (2010).

It is also possible to use indexes from fractal analysis, such as for example systematically applied by Chen (2016) to classify urban forms. The fractal approach to the city has been largely matured, as it can be tracked back to Batty and Longley (1994)

The link between urban morphology and topology of the underlying relational network has been suggested in a theoretical approach by Badariotti et al. (2007). Other more original indexes can be proposed, such as by Lee et al. (2017) which use the variations of trajectories for routes going through a city to establish a classification and show that it is strongly correlated with socio-economic variables.

Lagesse (2015) for the exploration of new measures for the road network

In operational urbanism, urban morphology is defined as "the characteristics of the material form of cities and fabrics" Paquot (2010). We use this term here for fabrics at a mesoscopic scale, seen through the spatial distribution of populations.

D'Acci (2015) models cities as isobenefit lines ; makes the link in opening between the distribution of amenities and network and flows, highlighting the open question of co-evolutive processes (circular causalities).

The rest of this chapter is organized as follows. We first detail the rationale of our approach to urban form, in particular the scale and objects considered. The second section develops the methodology used to quantify urban form, from the point of view of morphological indicators, network topology indicators, and their correlations. Empirical results for Europe are then described, to then be used to calibrate a model of urban morphogenesis that we detail and explore. We finally discuss potential developments of this work.

1.2 Urban form and interactions between networks and territories

Through relocation processes, sometimes induced by networks, we can expect the latest to influence the distribution of populations in space. Reciprocally, network characteristics can be influenced by this distribution. We propose therefore here to study a coupled approach to the urban form, given by synthetic indicators for these two subsystems, and by correlations between these indicators. At the scale of the system of cities, the spatial nature of the urban system is captured by cities position, associated with aggregated city variables. We work here at the mesoscopic scale, at which the precise spatial distribution of activities is necessary to understand the spatial structure of the territorial system. We will therefore use the term of morphological characteristics for population density and the road network.

The choice of "relevant" boundaries for the territory or the city is a relatively open problem which will often depend on the question we are trying to answer (Páez and Scott, 2005). This way, Guérois and Paulus (2002) show that the entities obtained are different if we consider an entry by the continuity of the built environment (morphological), by urban functions (employment area for example) or by administrative boundaries. We choose here the mesoscopic scale of a metropolitan center, of an order of one hundred kilometers, first for the relevance of the spatial field computed, and secondly because smaller scales become less relevant for the notion of urban form, whereas larger scales induce a too large variability.

At this scale, we can assume that territorial characteristics, for population and network, are locally defined and vary in an approximately continuous way in space. Thus, the construction of fields of morphological indicators will allow to endogenously reconstruct territorial entities through the emergent spatial structure of indicators at larger scales. For examples, cities should be distinguishable within non-urban spaces. The aim of this chapter is thus to study properties of these indicators and their interactions, and thus indirectly interactions between the territory and the network.

2 Measuring morphology

2.1 Urban morphology

We recall here the formal definition of morphological indicators, taken from Raimbault (2017). We consider population data on a grid $(P_i)_{1 \leq i \leq N^2}$, where $M = N^2$ is the number of cells, d_{ij} the distance between cells i, j, and $P = \sum_{i=1}^{M} P_i$ the total population. We measure urban form using:

1. Rank-size slope γ, expressing the degree of hierarchy in the distribution, computed by fitting with Ordinary Least Squares a power law distribution 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 (Le Néchet, 2015), which expresses how uniform the distribution is, what is a way to capture a level of concentration:

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

$\mathcal{E} = 0$ means that all the population is in one cell whereas $\mathcal{E} = 1$ means that the population is uniformly distributed.

3. Spatial-autocorrelation given by Moran index Tsai (2005), 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. Average distance between individuals Le Néchet (2009), which captures a spatial dispersion of population and quantifies a level of acentrism (distance to a monocentric model):

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

where d_M is a normalisation constant taken as the diagonal of the area on which the indicator is computed in our case.

The first two indexes are not spatial, and are completed by the last two that take space into account. Following Schwarz (2010), the effective dimension of the urban form justifies the use of all.

2.2 Network Measures

We consider network aggregated indicators as a way to characterize transportation network properties on a given territory, the same way morphological indicators yielded information on urban structure. We propose to compute some simple indicators on same extents as for morphology, to be able to explore relations between these static measures.

Static network analysis has been extensively documented in the literature, such as for example through the typology of urban street networks obtained by Louf and Barthelemy (2014) on a cross-sectional study of cities. Similarly, ? uses techniques from deep learning to establish a typology of urban road networks for a large number of cities across the world.

The questions behind such approaches are multiple: they can aim at finding typologies or at characterizing spatial networks, at understanding underlying dynamical processes in order to model morphogenesis, or even at being applied in urban planning such as *Space Syntax* approaches (Hillier and Hanson, 1989). We are positioned here more within the two first logics since we aim at characterizing the shape of networks in a first step, and then to include their dynamics in models in a second step. Our significant contribution is the characterization of the road network on large spatial extents, covering Europe and China.

We introduce indicators to have a broad idea of the form of the network, using a certain number of indicators to capture the maximum of dimensions of properties of networks, more or less linked to their use. These indicators summarize the mesoscopic structure of the network and are computed on topological networks obtained through simplification steps that will be detailed later. If we denote the network with $N = (V, E)$, nodes have spatial positions $\vec{x}(V)$ and populations $p(v)$ obtained through an aggregation of population in the corresponding Voronoï polygon, and edges E have *effective distances* $l(E)$ taking into account impedances and real distances (to include the primary network hierarchy). We then use:

- Characteristics of the graph, obtained from graph theory, as defined by Haggett and Chorley (1970): number of nodes $|V|$, number of links $|E|$, density d, average length of links $\bar{d}_l$, average clustering coefficient $\bar{c}$, number of components c_0.

- Measures linked to shortest paths: diameter r, euclidian performance v_0 (defined by ?), average length of shortest paths $\bar{l}$.

- Centrality measures: these are aggregated at the level of the network by taking their average and their level of hierarchy, computed by an ordinary least squares of a rank-size law, for the following centrality measures:

- Betweenness centrality (Crucitti et al., 2006), average $\bar{bw}$ and hierarchy α_{bw}: given the distribution of centrality on all nodes, we take the slope of a rank-size adjustment and the average of the distribution.

- Closeness centrality (Crucitti et al., 2006), average $\bar{cl}$ and hierarchy α_{cl}.

- Accessibility (Hansen, 1959), which is in our case computed as a closeness centrality weighted by populations: average $\bar{a}$ and hierarchy α_a.

Network performance is close to the rectilinearity measure (*straightness*) proposed by Josselin et al. (2016), which show that it efficiently differentiate rectilinear networks and radio-concentric networks, that are both recurring urban networks. Our indicators are conceived to capture network topology but not the use of the network: developments with suited data could extend these analyses to the functional aspect of networks, such as for example performance measures computed by Trépanier et al. (2009) using massive data for a public transportation network.

2.3 Correlations

Local spatial correlations are computed on windows gathering a certain number of observations, and thus of windows on which indicators have been computed. We denote by l_0 (which is equal to 10km in preceding results) the resolution of the distribution of indicators. The estimation of correlations in then done on squares of size $\delta \cdot l_0$ (with δ which can vary typically from 4 to 100).

δ gives simultaneously the number of observations used for the local estimation of correlation, and the spatial range of the corresponding window. Its value thus directly influences the confidence of the estimation. We can indeed derive the behavior of the correlation estimator as a function of the size of the sample. Under the assumption of a normal distribution of two random variables X, Y, then the Fisher transform of the Pearson estimator $\hat{\rho}$ such that $\hat{\rho} = \tanh(\hat{z})$ has a normal distribution. If z is the transform of the real correlation ρ, then a confidence interval for ρ is of size

$$\rho_+ - \rho_- = \tanh(z + k/\sqrt{N}) - \tanh(z - k/\sqrt{N})$$

where k is a constant. As $\tanh z = \frac{\exp(2z)-1}{\exp(2z)+1}$, we can develop this expression and reduce it, to obtain

$$\rho_+ - \rho_- = 2 \cdot \frac{\exp(2k/\sqrt{N}) - \exp(-2k/\sqrt{N})}{\exp(2z) - \exp(-2z) + \exp(2k/\sqrt{N}) + \exp(-2k/\sqrt{N})}$$
$$= 2 \cdot \frac{\sinh\left(2k/\sqrt{N}\right)}{\cosh\left(2z\right) + \cosh\left(2k/\sqrt{N}\right)}$$

Using the fact that $\cosh u \sim_0 1 + u^2/2$ and that $\sinh u \sim_0 u$, we indeed obtain that $\rho_+ - \rho_- \sim_{N \gg 0} k'/\sqrt{N}$.

3 Results

3.1 Urban morphology

The implementation of indicators must be done carefully, since computational complexities can reach $O(N^4)$ for the Moran index for example: we use convolution through Fast Fourier Transform, which is a technique allowing the computation of the Moran index with a complexity in $O(\log^2 N \cdot N^2)$[1].

---

[1] I.e. having an execution time bounded by $\log^2 N \cdot N^2$ if N is the data size, what is a considerable gain compared to N^4: to process a grid of width 100, the asymptotic gain factor will be approximatively 10000.

3.2 Network Topology

3.2.1 Data preprocessing

We work here with the road network, which structure is finely conditioned to territorial configuration of population densities. Furthermore, data for the current road network is openly available through the OpenStreetMap (OSM) project **?**. Its quality was investigated for different countries such as England Haklay (2010) and France Girres and Touya (2010). It was found to be of a quality equivalent to official surveys for the primary road network. Concerning China, although Zheng and Zheng (2014) underlined a quick acceleration of OSM road data completeness and accuracy, its use for computation of network indicators may be questioned at a very fine scale. Zhang et al. (2015) highlights different regimes of data quality, partitioning China into regions among which qualitative behavior of OSM data varies. We will have to keep in mind this variability, and to ensure the robustness of results, we will simplify the network at a sufficient level of aggregation.

The network constituted by primary road segments is aggregated at the fixed granularity of the density grid to create a graph. It is then simplified to keep only the topological structure of the network, normalized indicators being relatively robust to this operation. This step is necessary for a simple computation of indicators and a thematic consistence with the density layer. We keep only the nodes with a degree strictly greater or smaller than two, and corresponding links, by taking care to aggregate the real geographical distance when constructing the corresponding topological link. Given the order of magnitude of data size (for Europe, the initial database has $\simeq 44.7 \cdot 10^6$ links, and the final simplified database $\simeq 20.4 \cdot 10^6$), a specific parallel algorithm is used, with a *split-merge* structure. It separates the space into areas that can be independently processed and then merged. It is detailed in Appendix **??**.

3.2.2 Network Simplification Algorithm

We detail here the road network simplification algorithm from OpenStreetMap data. The general workflow is the following: (i) data import by selection and spatial aggregation at the raster resolution; (ii) simplification to keep only the topological network, processed in parallel through *split/merge*.

OSM data are imported into a `pgsql` database (`Postgis` extension for the management of geometries and to have spatial indexes). The import is done using the software `osmosis` **?**, from an image in compressed `pbf` format of the OpenStreetMap database[2]. We filter at this stage the links (`ways`) which posses the tag `highway`, and keep the corresponding nodes.

The network is first aggregated at a 100m granularity in order to be consistently used with population grids. It furthermore allows to be robust to local coding imperfections or to very local missing data. For this step, roads are filtered on a relevant subset of tags[3]. For the set of segments of corresponding lines, a link is created between the origin and the destination cell, with a real length computed between the center of cells and a speed taken as the speed of the line if it is available.

The simplification is then operated the following way:

1. The whole geographical coverage is cut into areas on which computations will be partly done through parallel computation (*split* paradigm). Areas have a fixed size in number of cells of the base raster (200 cells).

2. On each sub-area, a simplification algorithm is applied the following way: as long as there still are vertices of degree 2, successive sequences of such vertices are determined, and corresponding links are replaced by a unique link with real length and speed computed by cumulation on the deleted links.

---

[2]Dumps were retrieved from `http://download.geofabrik.de`, in July 2016 for Europe, and July 2017 for China.
[3]That we take within `motorway,trunk,primary,secondary,tertiary,unclassified,residential`.

3. As the simplification algorithm keeps the links having an intersection with the border of areas, a fusion followed by a simplification of resulting graphs is necessary. To keep a reasonable computational cost, the size of merged areas has to stay low: we take merge areas composed by two contiguous areas. A paving by four sequences of independent merging allows then to cover the full set of joints between areas[4], these sequences being executed sequentially. The Frame **??** shows the covering of joints by merging areas.

We have then at our disposition a topological graph given by the links between cells of the base raster, having distance and speed attributes corresponding to the underlying real links. Topological graphs for Europe and China are available as open databases on the dataverse repository at `https://doi.org/10.7910/DVN/RKDZMV`.

3.2.3 Results

Network indicators have been computed on the same areas than urban form indicators, in order to put them in direct correspondance and later compute the correlations. We show in Fig. 1 a sample for France.

The spatial behavior of indicators unveils local regimes as for the urban form (urban, rural, metropolitan), but also strong regional regimes. They can be due to the different agricultural practices depending on the region for the rural for example, implying a different partition of parcels and also a particular organization of their serving. For network size, Brittany is a clear outlier and rejoins urban regions, witnessing very fragmented parcels (and a fortiori also of a land property fragmentation in the simplifying assumption of corresponding parcels and properties). This is partly correlated to a low hierarchy of accessibility. The South and the East of the extended *Bassin Parisien* are distinguishable by a strong average betweenness centrality, in accordance with a strong hierarchy of the network.

The same way as for urban form, this spatial variability suggests the search of variables regimes of interactions between indicators, as we will do for later through their correlations.

3.3 Effective static correlations and non-stationarity

3.3.1 Spatial correlations

We show in Fig. 2 examples of correlations estimated with $\delta = 12$ in the case of France. With 20 indicators, the correlation matrix is significantly large in size, but the effective dimension (the number of components required to reach the majority of variance) is reduced: principal components analysis shows that 10 components already capture 62% of variance, and the first component already captures 17%, what is considerable in a space where the dimension is 190.

The Fig. 3 gives the spatial distribution for all Europe, of a sample of correlations between indicators: $\rho[\alpha_{cl}, I]$, $\rho[\gamma, \alpha]$, $\rho[\bar{bw}, \gamma]$, $\rho[\alpha_{bw}, \alpha_{cl}]$, $\rho[|V|, \bar{l}]$, $\rho[\gamma, r_\gamma]$ (with r_γ adjustment coefficient for γ). We see interesting structures emerging, such as the hierarchy and its adjustment which present an area of strong correlation in the center of Europe and negative correlation areas, or the number of nodes and the path length which correlate in mountains and along the coasts (what is expected since roads then do several detours) and have a negative correlation otherwise.

To have an idea of the robustness of the estimation, we investigate for the correlations estimated on the full dataset (corresponding to $\delta = \infty$) the relative size of confidence intervals at the 95% level (Fisher method) given by $\frac{|\rho_+ - \rho_-|}{|\rho|}$, for correlations such that $|\rho| > 0.05$. The median of this rate is at 0.04, the ninth decile at 0.12 and the maximum at 0.19, what means that the estimation is always relatively good compared to the value of correlations.

It is possible to examine the bloc for urban form, for the network, or for crossed correlations, which directly express a link between properties of the urban form and of the network. For example, the relation

---

[4]In the very rare cases of a link between two non-contiguous areas, the remaining link is not simplified. This case was not observed in practice in our data.
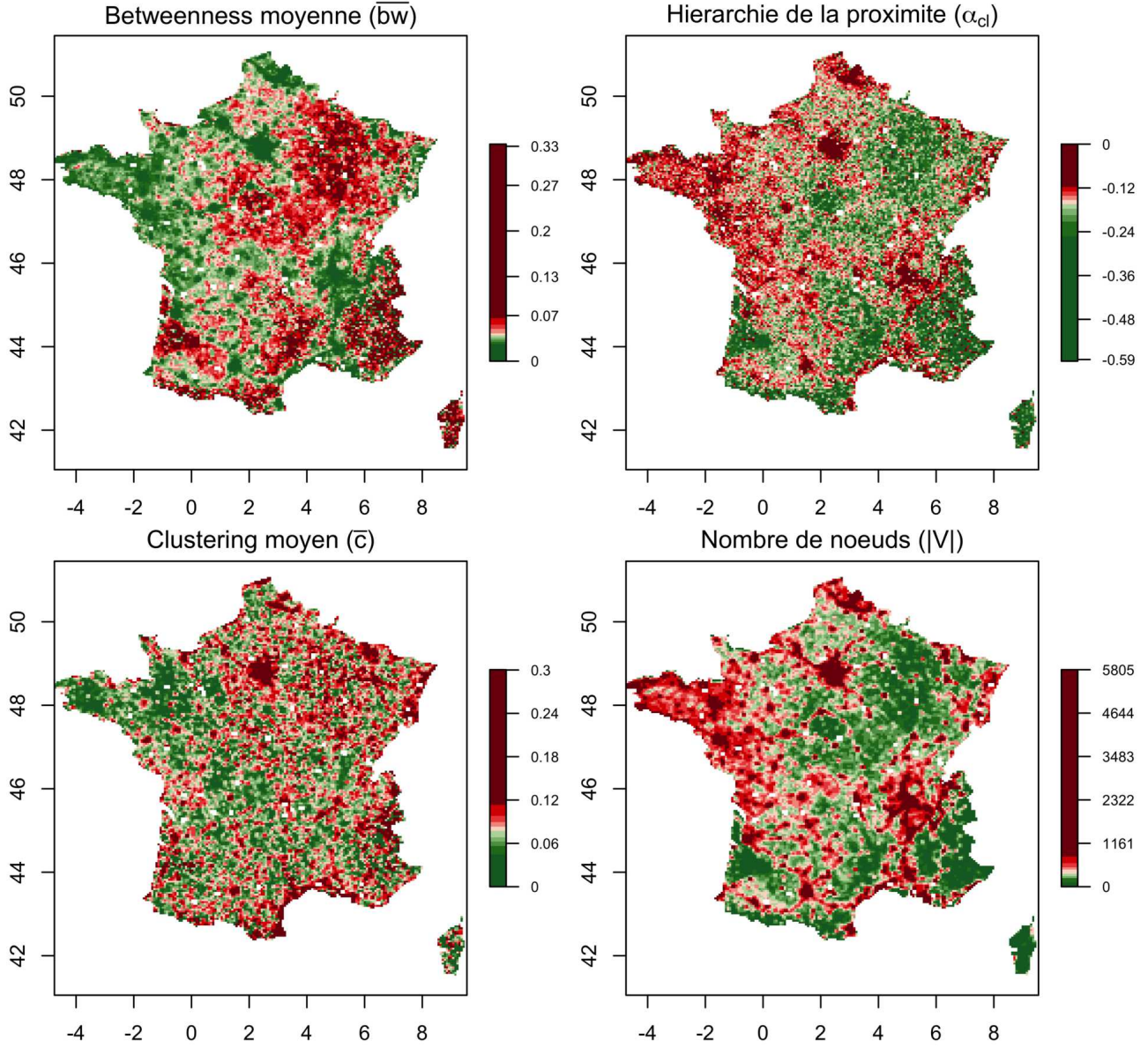
Figure 1: **Spatial distribution of network indicators.** We show indicators for France, in correspon-
dance with morphological indicators described previously. We give here the average betweenness central-
ity $\bar{bw}$, the hierarchy of closeness centrality $\alpha_c l$, the average clustering coefficient $\bar{c}$ and the number of
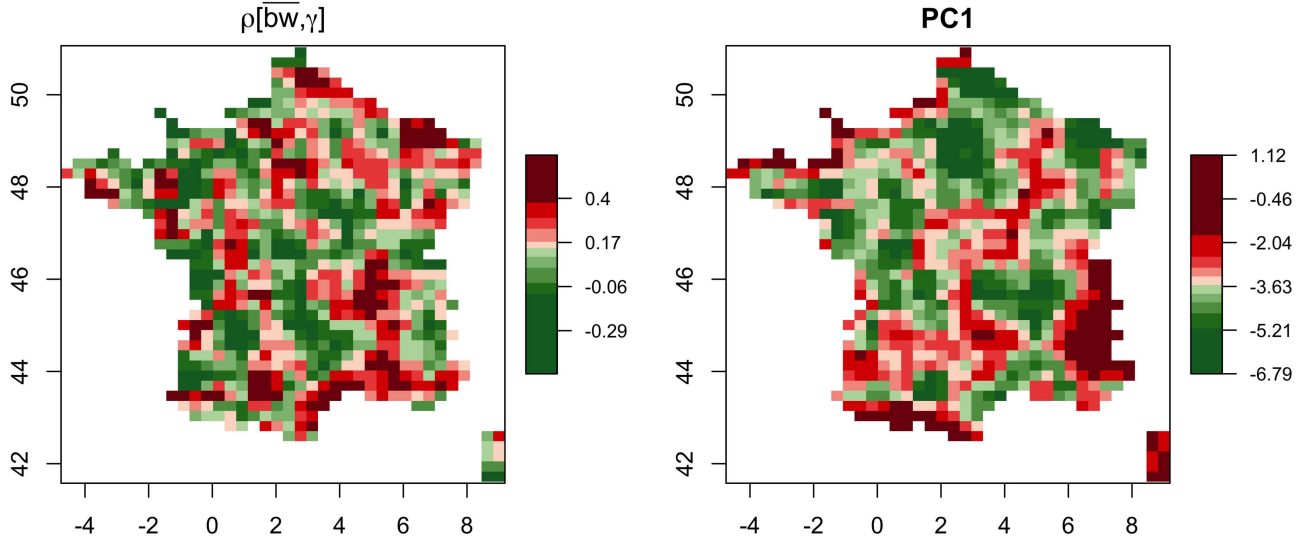nodes $|V|$.

Figure 2: **Examples of spatial correlations.** For France, the maps give $\rho\left[\bar{bw}, \gamma\right]$, correlation between the average betweenness centrality and the hierarchy of population (*Left*) and the first component of the reduced matrix (*Right*).

between average betweenness centrality and morphological hierarchy that we visualize allows to understand the process corresponding to the correspondance of hierarchies: a hierarchical population can induce a hierarchical network or the opposite direction, but it can also induce a distributed network or such a network create a population hierarchy - this must be well understood in terms of correspondence and not causality, but this correspondance informs on different urban regimes. Metropolitan areas seem to exhibit a positive correlation for these two indicators, as shows the Fig. 2, and rural spaces a negative correlation.

In order to give a picture of global relations between indicators, we can refer to the matrix obtained for $\delta = \infty$: for example, a strong population hierarchy is linked to a high and hierarchical betweenness centrality, but is negatively correlated to the number of edges (a diffuse population requires a more spread network to serve all the population). However, it is not possible this way to systematically link indicators, since they especially strongly vary in space.

This suggests a very high variety of interaction regimes. The spatial variation of the first component of the reduced matrix confirms it, what clearly reveals the spatial non-stationarity of interaction processes between forms, since the first and second moments vary in space. The statistical significance of stationarity can be verified in different ways, and there does not exist to the best of our knowledge a generic test for spatial non-stationarity. Zhang and Zhou (2014) develops for example a test for rectangular regions of any dimension, but in the specific case of *point processes*. We use here the method of Leung et al. (2000) which consists in estimating through bootstrap the robustness of Geographically Weighted Regression models. These will be developed below, but we obtain for all tested models a significant non-stationarity without doubt ($p < 10^{-3}$).

3.3.2 Variations of the estimated correlations

We show in Fig. 4 the variation of the estimation of correlation as a function of window size. More precisely, we observe a strong variation of correlations as a function of δ, what is reflected in the average value of the matrix given here (which extends for example from $\rho(4) = 0.22$ to $\rho(80) = 0.12$ for average absolute cross-correlations). An increase of δ leads for all measures a shift towards positive values, but also a narrowing of the distribution, these two effects resulting in a decrease of average absolute correlations, which approximatively stabilize for large values of δ. Such a variation could be a clue of a multi-scalar behavior: a change in window size should not influence the estimation if a single process would be implied,
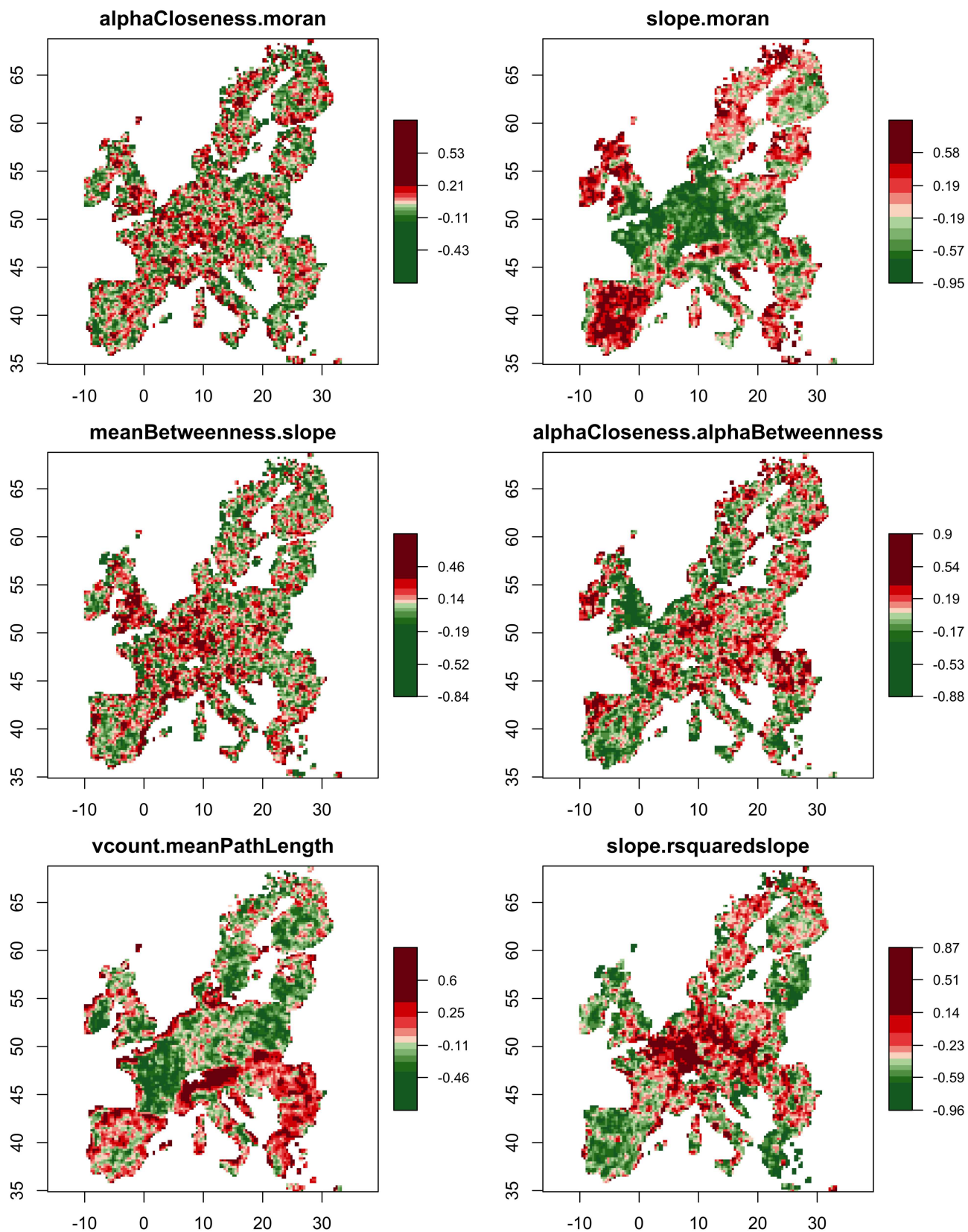
Figure 3: **Spatial correlations for Europe.** The estimation is done here with $\delta = 12$.

it should only change the robustness of the estimation. A formal demonstration of this hypothesis remains however out of the scope of this work.

Furthermore, the variation of the normalized size of the confidence interval for correlations, which in theory under an assumption of normality should lead $\delta \cdot |\rho_+ - \rho-|$ to remain constant, since boundaries vary asymptotically as $1/\sqrt{N} \sim 1/\sqrt{\delta^2}$ (the demonstration is given in Appendix **??**), follows the direction of this hypothesis of processes superposed at different scales as proposed previously.

Thus, processes are both non-stationary, and clues suggest that they result of the superposition of processes at different scales[5].

3.3.3   Typical scales

We also propose to explore the possible property of multi-scalar processes by the extraction of endogenous scales which are present in the data. A Geographically Weighted Principal Component Analysis (GWR-PCA) **?** in exploration suggests weights and importances that vary in space, what is in consistence with the non-stationarity of correlation structures obtained above. There is no reason a priori that the scales of variation of the different indicators are strictly the same. We propose thus to extract typical scales for crossed relations between the urban form and network topology.

We implement therefore the following method: we consider a typical sample of indicators (four for each aspect, see the list in Table 1), and for each indicator we formulate all the possible linear models as a function of opposite indicators (network for a morphological indicator, morphological for a network indicator), aiming at directly capturing the interaction without controlling on the type of form or of network. These models are then adjusted by a Geographically Weighted Regression (GWR) with an optimal range determined by a corrected information criteria (AICc)[6]. For each indicator, we keep the model with the best value of the information criteria. We adjust the models on data for France, with a *bisquare* kernel and an adaptative bandwidth in number of neighbors.

Results are presented in Table 1. It is first interesting to note that all models have only one variable, suggesting relatively direct correspondances between topology and morphology. All morphological indicators are explained by network performance, i.e. the quantity of detours it includes. On the contrary, network topology is explained by Moran index for centralities, and by entropy for performance and the number of vertices. There is thus a dissymmetry in relations, the network being conditioned in a more complex way to the morphology than the morphology to the network. The adjustments are rather good ($R^2 > 0.5$) for most indicators, and *p-values* obtained for all models (for the constant and the coefficient) are lower than 10^{-3}. Concerning the scales corresponding to the optimal model, they are very localized, of the order of magnitude of ten kilometers, i.e a larger variation than the one obtained the correlations. This analysis confirms thus statistically on the one hand the non-stationarity, and on the other hand give a complementary point of view on the question of endogenous scales.

4   Urban morphogenesis model

4.1   Model rationale

Urban settlements and transportation networks have been shown to be co-evolving, in the different thematic, empirical and modeling studies of territorial systems developed up to here. As we saw, modeling approaches of such dynamical interactions between networks and territories are poorly developed. We propose in this section to realize a first entry at an intermediate scale, focusing on morphological and functional properties of the territorial system in a stylized way. We introduce a stochastic dynamical

---

[5]The notion of multi-scalar process is otherwise very broad, and can manifest itself in scaling laws for example **?**. An approach closer to the one we took is given by **?** which measures intrinsic scales to segregation phenomenons by using measures from Information Theory.
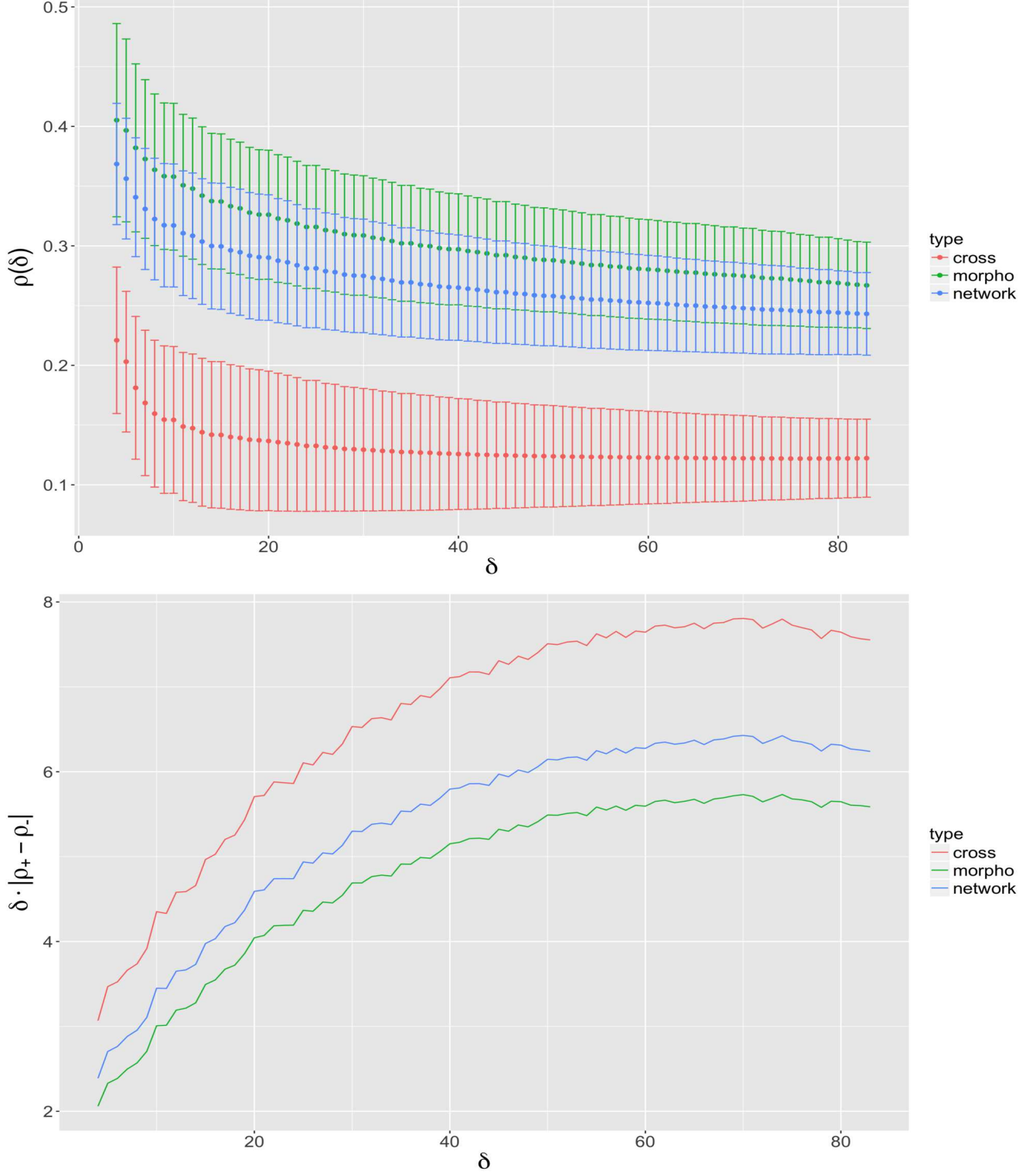
[6]By using the R package GWModel **?**.

Figure 4: **Variation of correlations with scale, for correlations computed on Europe.** (*Top*)
Average absolute correlations and their standard deviations, for the different blocs, as a function of δ;
(*Bottom*) Normalized size of the confidence interval $\delta \cdot |\rho_+ - \rho_-|$ (confidence interval $[\rho_-, \rho_+]$ estimated
by the Fisher method) as a function of δ.

Table 1: **Interrelations between network indicators and morphological indicators.** Each relation is adjusted by a Geographically Weighted Regression, for the optimal range adjusted by AICc.

Indicator	Model	Range (km)	Adjustment (R^2)
Average distance $\bar{d}$	$\bar{d} \sim v_0$	11.6	0.31
Entropy $\mathcal{E}$	$\mathcal{E} \sim v_0$	8.8	0.75
Moran I	$I \sim v_0$	8.8	0.49
Hierarchy γ	$\gamma \sim v_0$	8.8	0.68
Average betweenness $\bar{bw}$	$\bar{bw} \sim I$	12.3	0.58
Average closeness $\bar{cl}$	$\bar{cl} \sim I$	13.9	0.26
Performance v_0	$v_0 \sim \mathcal{E}$	8.6	0.86
Number of nodes $|V|$	$|V| \sim \mathcal{E}$	8.6	0.88

model of urban morphogenesis which couples the evolution of population density within grid cells with a growing road network.

The general principles of the model are the following. With an overall fixed growth rate, new population aggregate preferentially to a local potential, for which parameters control the dependance to various explicative variables. These are in particular local density, distance to the network, centrality measures within the network and generalized accessibility. **?** shows in the case of Stockholm the very strong correlation between centrality measures in the network and the type of land-use, what confirms the inportance to consider centralities as explicative variables for the model at this scale. We generalize thus the morphogenesis model studied in **??**, with aggregation mechanisms similar to the ones used by Raimbault et al. (2014). A continuous diffusion of population completes the aggregation to translate repulsion processes generally due to congestion. Because of the different time scales of evolution for the urban environment and for networks, the network grows at fixed time steps, following the submodel developed in **??**: a first fixed rule ensures connectivity of newly populated patches to the existing network. The different network generation heuristics are then included in the model. We expect the different heuristics to be complementary since for example the gravity model would be more typical of planned top-down network evolution, whereas the biological model will translate bottom-up processes of network growth. The Fig. 5 summarizes the general structure of the morphogenesis model.

4.2   Model description

The model is based on a squared population grid of size N, which cells are defined by populations (P_i). A road network is included as a vectorial layer on top of the grid, similarly to Raimbault et al. (2014). We assume at the initial state a given population distribution and a network.

The evolution of densities is based on a utility function, influenced by local characteristics of the urban form and function, that we call *explicative variables*. Let $x_k(i)$ a local explicative variable for cell i, which will be among the following variables: population P_i; proximity to roads, taken as $\exp(-d/d_n)$ where d is the distance by projection on the closest road, and $d_n = 10$ is fixed; betweenness centrality; closeness centrality; accessibility.

For the last three, they are defined as previously for network nodes, and then associated to cells by taking the value of the closest node, weighted by a decreasing function of the distance to it, i.e. of the form $x_k = x_k^{(n)}(\text{argmin}_j d(i,j)) \cdot \exp\left(-\min_j d(i,j)/d_0\right)$, with $x_k^{(n)}$ the corresponding variable for nodes, the index j being taken on all nodes, and the decay parameter d_0 is in our case fixed at $d_0 = 1$ to keep the property that network variables are essentially significant at close distances from the network. We consider then normalized explicative variables defined by $\tilde{x}_k(i) = x_k(i) - \min_j x_k(j)/(\max_j x_k(j) - \min_j x_k(j))$.

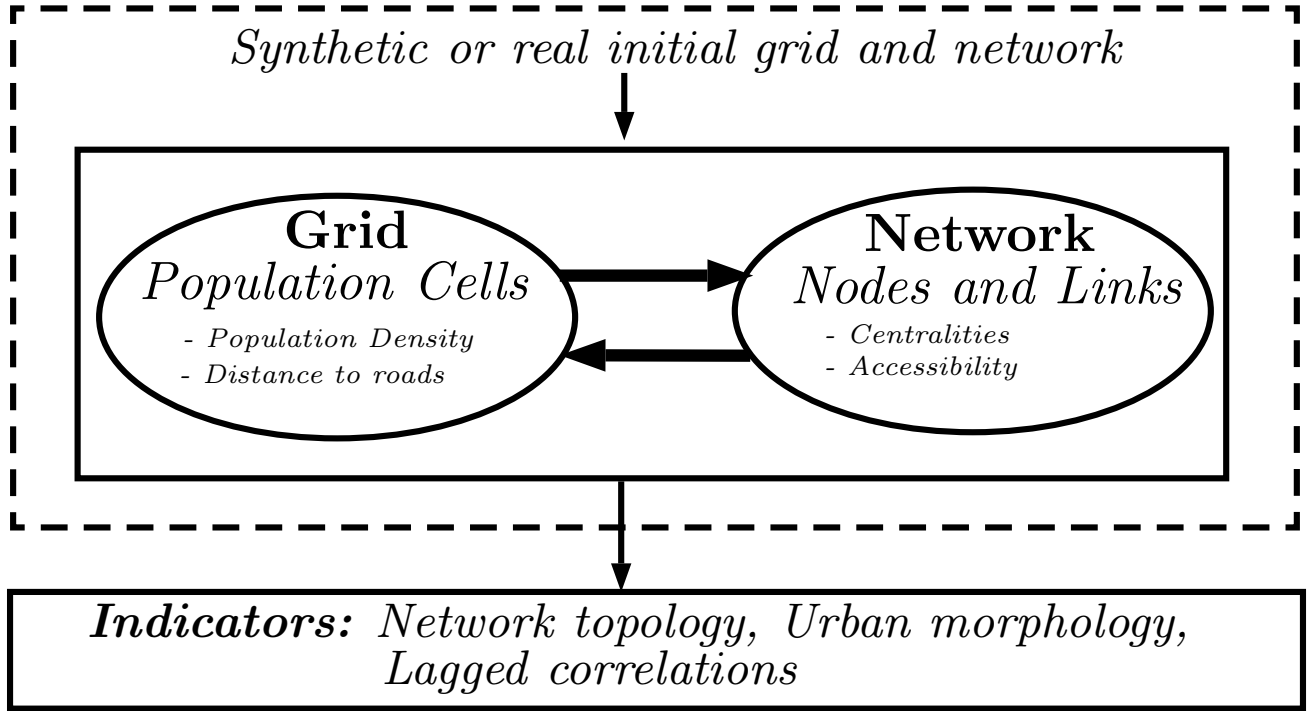The utility of a cell is then given by a linear aggregation

Figure 5: **Structure of the co-evolution model at the mesoscopic scale.**

$$U_i = \sum_k w_k \cdot \tilde{x}_k(i) \tag{2}$$

where $\tilde{x}_k$ are the normalized local explicative variables, and w_k are weight parameters, which allow to weight between the different influences.

A time step of model evolution includes then the following stages.

1. Evolution of the population following rules similar to Raimbault (2017). Given an exogenous growth rate N_G, individuals are added independently following an aggregation done with a probability $U_i^\alpha / \sum_k U_k^\alpha$, followed by a diffusion of strength β to neighbor cells, done n_d times.

2. Network growth is done through multi-modeling as proposed by Raimbault (2017b), knowing that this takes place is the time step is a multiple of a parameter t_N, which allows to integrate a differential between temporal scales for population growth and for network growth.

The aggregation following a power of the utility yields a flexibility in the underlying optimization problem, since as **?** recall, the use of different norms in spatial optimal location problems corresponds to different logics of optimization.

The parameters of the model that we will make vary are then:

- aggregation-diffusion parameters α, β, N_g, n_d, summarized in Table **??**;

- the four weight parameters w_k for the explicative variables, which vary in $[0; 1]$;

- network growth parameters for the different heuristics, summarized in Table **??**.

Output model indicators are the urban morphology indicators, topological network indicators, and lagged correlations between the different explicative variables.

4.3 Simulation results

The model is implemented in NetLogo, given the heterogeneity of aspects that have to be taken into account, and this language being particularly suitable to couple a grid of cells with a network. Urban morphology indicators are computed thanks to a NetLogo extension specially developed (see Appendix ??).

We propose to focus on the ability of the model to capture relations between networks and territories, and more particularly the co-evolution. Therefore, we will try to establish if (i) the model is able to reproduce, beyond the form indicators, the static correlation matrices computed previously; and (ii) the model produces a variety of dynamical relations in the sense of causality regimes developed in ??.

The model is initialized on fully synthetic configurations, with a grid of size 50. Configurations are generated through an exponential mixture in a way similar to ?: $N_c = 8$ centers are randomly located, to which a population is attributed following a scaling law $P_i = P_0 \cdot (i+1)^{-\alpha_S}$ with $\alpha_S = 0.8$ and $P_0 = 200$. The population of each center is distributed to all cells with an exponential kernel of shape $d(r) = P_{max} \exp(-r/r_0)$ where the parameter r_0 is determined to fix the population at P_i, with $P_{max} = 20$ (density at the center)[7]. The initial network skeleton is generated as detailed in ??.

We explore a Latin Hypercube Sampling of the parameter space, with 10 repetitions for around 7000 parameter points, corresponding to a total of around 70000 model repetitions[8], realized on a computation grid by using OpenMole.

4.3.1 Static and dynamical calibration

The model is calibrated at the first order, on indicators for the urban form and network measures, and at the second order on correlations between these. We introduce an *ad hoc* calibration procedure in order to take into account the first two moments, that we detail below. More elaborated procedures are used for example in economics, such as ? which uses the noise of the difference between two variables to obtain the same covariance structure for the two corresponding models, or in finance, such as ? which define a notion on equivalence between latent variables models which incorporates the equality of the interdependence structure between variables. We avoid here to add supplementary models, and consider simply a distance on correlation matrices. The procedure is the following.

- Simulated points are the ones obtained through the sampling, with average values on repetitions.

- In order to be able to estimate correlation matrices between indicators for simulated data, we make the assumption that second moments are continuous as a function of model parameters, and split for each heuristic the parameter space into areas to group parameter points[9], what allows to estimate for each group indicators and the correlation matrix.

- For each estimation done this way, that we write $\bar{S}$ (indicators) and $\rho[S]$ (correlations), we can then compute the distance to real points on indicators $d_I(R_j) = d(\bar{S}, R_j)$ and on correlation matrices $d_\rho(R_j) = d(\rho[S], \rho[R_j])$ where R_j are the real points with their corresponding correlations[10], and d an euclidian distance normalized by the number of components.

- We consider then the aggregated distance defined as $d_A^2(R_j) = d_I^2(R_j) + d_\rho^2(R_j)$. Indeed, as developed empirically and analytically in Appendix ??, the shape of Pareto fronts for the two distances considered suggests the relevance of this aggregation. The real point closest to a simulated point is then the one in the sense of this distance.

---

[7]We have indeed $P_i = \iint d(r) = \int_{\theta=0}^{2\pi} \int_{r=0}^{\infty} d(r) r dr d\theta = 2\pi P_{max} \int_r r \cdot \exp(-r/r_0) = 2\pi P_{max} r_0^2$, and therefore $r_0 = \sqrt{\frac{P_i}{2\pi P_{max}}}$.

[8]For which simulation results are also available at http://dx.doi.org/10.7910/DVN/OBQ4CS.

[9]Each parameter being binned into $15/k$ equal segments, where k is the number of parameters: we empirically observed that this allowed to always have a minimal number of points in each area.

[10]That are estimated in ?? as we recall, with a square window centered around the point, that we take here for $\delta = 4$.
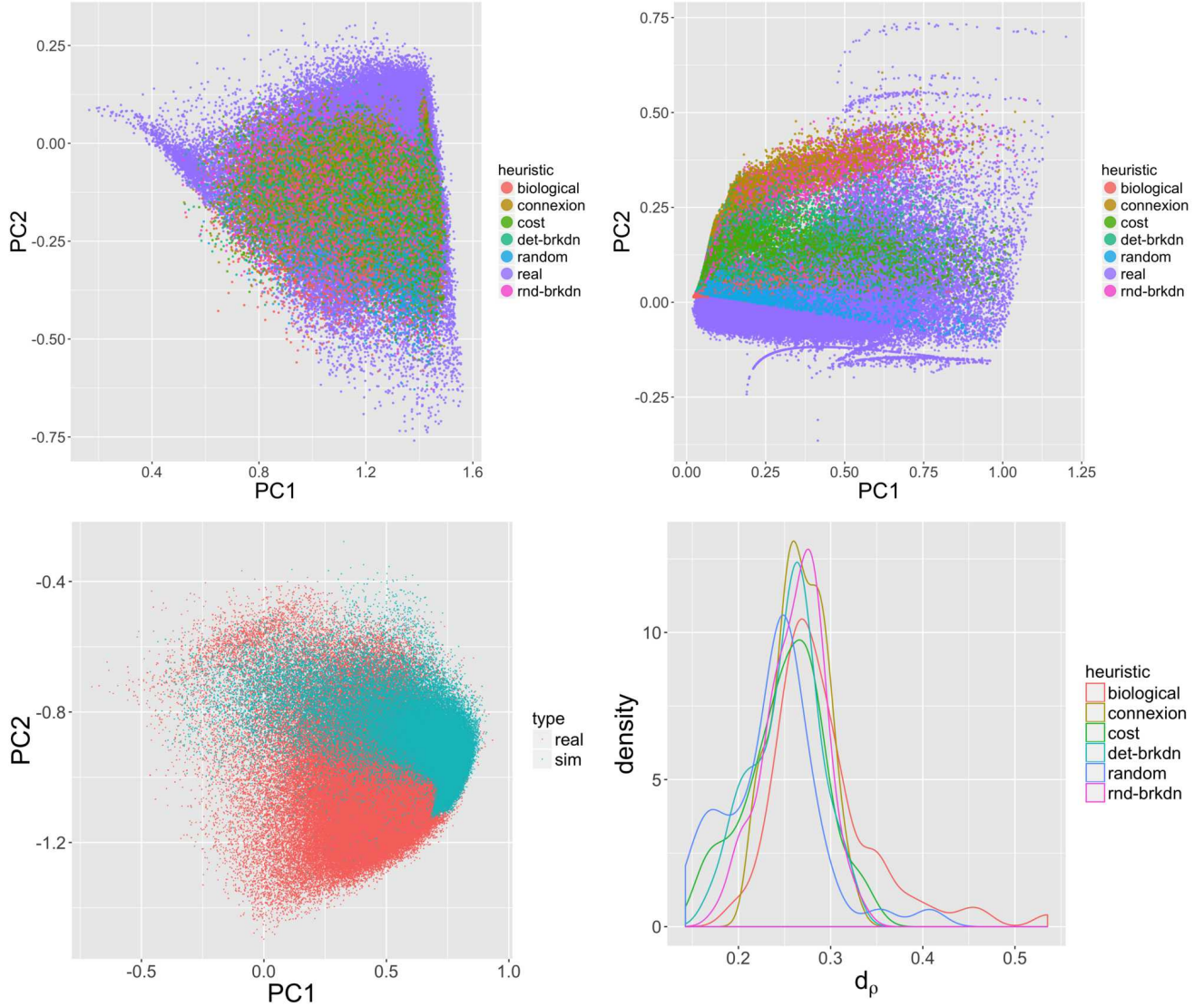
Figure 6: **Calibration of the morphogenesis model at the first and second order.** (*Top Left*) Simulated and observed point clouds in a principal plan for urban morphology indicators. (*Top Right*) Simulated and observed could points in a principal plan for network indicators. (*Bottom Left*) Simulated and observed point clouds in a principal plan for all indicators. (*Bottom Right*) Distributions of distances on correlations d_ρ, for the different heuristics.

The Fig. 6 summarizes calibration results. Morphological indicators are easier to approach than network indicators, for which a part of the simulated clouds does not superpose with observed points. We find again a certain complementarity between network heuristics. When considering the full set of indicators, few simulated points are situated far from the observed points, but a significant proportion of these is beyond the reach of simulation. Thus, the simultaneous capture of morphology and topology is obtained at the price of less precision.

We however obtain a good reproduction of correlation matrices as shown in Fig. 6 (histogram for d_ρ, bottom right). The worse heuristic for correlations is the biological one in terms of maximum, whereas the random produces rather good results: this could be due for example to the reproduction of very low correlations, which accompany a structure effect due to the initial addition of nodes which imposes already a certain correlation. On the contrary, the biological heuristic introduces supplementary processes which can possibly be beneficial to the network in terms of independence (or following the opposed viewpoint be detrimental in terms of correlations). In any case, this application shows that our model is able to resemble real configurations both for indicators and their correlations.

4.3.2 Causality regimes

We furthermore study dynamical lagged correlations between the variations of the different explicative variables for cells (population, distance to the network, closeness centrality, betweenness centrality, accessibility). We apply the method of causality regimes introduced in Raimbault (2017a). The Fig. 7 summarizes the results obtained with the application of this method on simulation results of the co-evolution model. The number of classes inducing a transition is smaller than for the RDB model, translating a smaller degree of freedom, and we fix in that case $k = 4$. Centroid profiles allow to understand to ability of the model to more or less capture a co-evolution.

The regimes obtained appear to be less diverse than the ones obtained by Raimbault (2017a) or for a macroscopic model of co-evolution by Raimbault (2018). Some variables have naturally a strong simultaneous correlation, spurious from their definitions, such as closeness centrality and accessibility, or the distance to the road and the closeness centrality. For all regimes, population significantly determines the accessibility. The regime 1 corresponds to a full determination of the network by the population. The second is partly circular, through the effect of roads on populations. The regime 3 is more interesting, since closeness centrality negatively causes the accessibility: this means that in this configuration, the coupled evolution of the network and the population follow the direction of a diminution of congestion. Furthermore, as population causes the closeness centrality, there is also circularity and thus co-evolution in that case. When we locate it in the phase diagram, this regime is rather sparse and rare, contrary for example to the regime 1 which occupies a large portion of space for a low importance of the road ($w_{road} \leq 0.3$). This confirms that the co-evolution produced by the model is localized and not a characteristic always verified, but that it is however able to generate some in particular regimes.

5 Discussion

5.1 Correlations

We have first shown empirically the non-stationarity of interactions between the morphology of the distribution of populations and the topology of the road network. Various developments of this analysis are possible.

Population density grids exist for all regions of the workd, such as for example the ones provided by ?.The analysis may be repeated on other regions of the world, to compare the correlation regimes and test if urban system properties stay the same, keeping in mind the difficulties linked to the differences in data quality.

The research of local scales, i.e. with an adaptive estimation window in terms of size and shape for
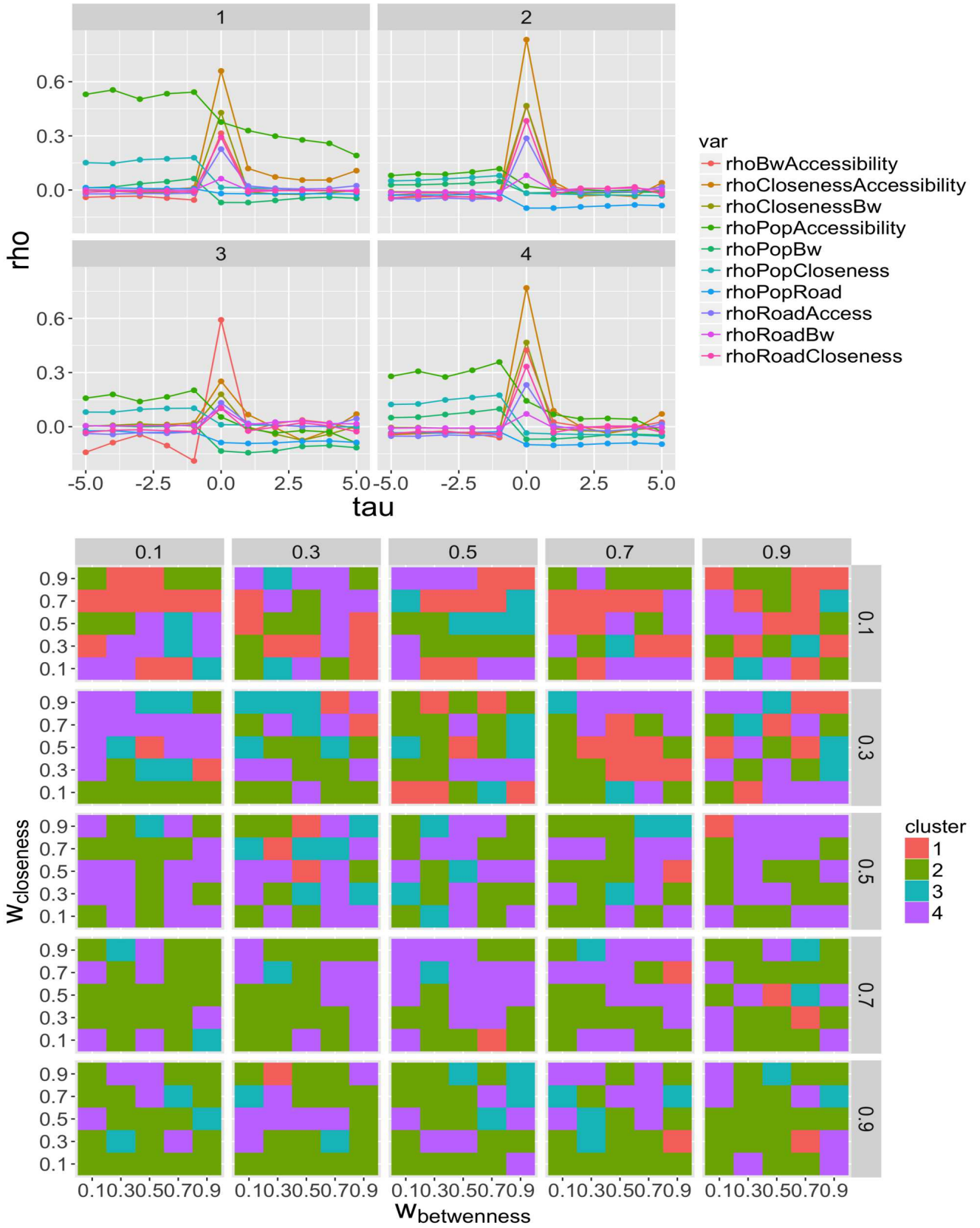
Figure 7: **Causality regimes for the co-evolution model.** (*Top*) Trajectories of classes centers in terms of $\rho[\tau]$ between the different explicative variables. (*Bottom*) Phase diagram of regimes in the parameter space for w_k, represented here as the variation of diagrams for (w_{bw}, w_{cl}), along the variations of w_{road} (in rows) and of w_{pop} (in columns).

correlations, would allow to better understand the way processes locally influence their neighborhood. The validation criteria for window size would still be to determine: it can be as above an optimal range for explicative models that are locally adjusted.

The question of ergodicity, crucial to urban systems (Pumain, 2012), should also be explored from a dynamical point of view, by comparing time and spatial scales of the evolution of processes, or more precisely the correlations between variations in time and variations in space, but the issue of the existence of databases precise enough in time appears to be problematic. The study of a link between the derivative of the correlation as a function of window size and of the derivatives of the processes is also a direction to obtain indirect informations on dynamics from static data.

Finally, the search of classes of processes on which it is possible to directly establish the relation between spatial correlations and temporal correlations, is a possible research direction. It stays out of the scope of this present work, but would open relevant perspectives on co-evolution, since it implies evolution in time and an isolation in space, and therefore a complex relation between spatial and temporal covariances.

5.2   Modeling urban morphogenesis

We have then proposed a co-evolution model at the mesoscopic scale, based on a multi-modeling paradigm for the evolution of the network. The model is able to reproduce a certain number of observed situations at the first and second order, capturing thus a static representation of interactions between networks and territories. It also yields different dynamical causality regimes, being however less diverse than the simple model studied by Raimbault (2017a): therefore, a more elaborated structure in terms of processes must be paid in flexibility of interaction between these. This suggests a tension between a "static performance" and a "dynamical performance" of models.

An open question is to what extent a pure network model with preferential attachment for nodes would reproduce results close to what we obtained. The complex coupling between aggregation and diffusion (shown by Raimbault (2017)) could not be easily included, and the model could in any case not answer to questions on the coupling of the dynamics.

Conclusion

We have thus explored a co-evolution model based on morphogenesis that takes into account multiple processes for the evolution of the network. We studied its calibration on observed data at the first and the second order, and explored the causality regimes it produces.

Acknowledgements

References

Badariotti, D., Banos, A., and Moreno, D. (2007). Conception d'un automate cellulaire non stationnaire à base de graphe pour modéliser la structure spatiale urbaine: le modèle remus. *Cybergeo: European Journal of Geography*.

Batty, M. and Longley, P. A. (1994). *Fractal cities: a geometry of form and function*. Academic press.

Chen, Y. (2016). Normalizing and Classifying Shape Indexes of Cities by Ideas from Fractals. *arXiv preprint arXiv:1608.08839*.

Crucitti, P., Latora, V., and Porta, S. (2006). Centrality measures in spatial networks of urban streets. *Physical Review E*, 73(3):036125.

D'Acci, L. (2015). Mathematize urbes by humanizing them. cities as isobenefit landscapes: psycho-economical distances and personal isobenefit lines. *Landscape and Urban Planning*, 139:63–81.

Girres, J.-F. and Touya, G. (2010). Quality assessment of the french openstreetmap dataset. *Transactions in GIS*, 14(4):435–459.

Guérois, M. and Paulus, F. (2002). Commune centre, agglomération, aire urbaine: quelle pertinence pour l'étude des villes? *Cybergeo: European Journal of Geography*.

Haggett, P. and Chorley, R. J. (1970). *Network analysis in geography*. St. Martin's Press.

Haklay, M. (2010). How good is volunteered geographical information? a comparative study of openstreetmap and ordnance survey datasets. *Environment and planning B: Planning and design*, 37(4):682–703.

Hansen, W. G. (1959). How accessibility shapes land use. *Journal of the American Institute of planners*, 25(2):73–76.

Hillier, B. and Hanson, J. (1989). *The social logic of space*. Cambridge university press.

Josselin, D., Labatut, V., and Mitsche, D. (2016). Straightness of rectilinear vs. radio-concentric networks: modeling simulation and comparison. *arXiv preprint arXiv:1609.05719*.

Lagesse, C. (2015). Read Cities through their Lines. Methodology to characterize spatial graphs. *arXiv preprint arXiv:1512.01268*.

Le Néchet, F. (2009). Quantifier l'éloignement au modèle de bussière: monocentrisme contre" acentrisme". In *Neuvièmes rencontres de Théo Quant*, pages 19–p.

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*.

Lee, M., Barbosa, H., Youn, H., Holme, P., and Ghoshal, G. (2017). Morphology of travel routes and the organization of cities. *Nature communications*, 8(1):2229.

Leung, Y., Mei, C.-L., and Zhang, W.-X. (2000). Statistical tests for spatial nonstationarity based on the geographically weighted regression model. *Environment and Planning A*, 32(1):9–32.

Louf, R. and Barthelemy, M. (2014). A typology of street patterns. *Journal of The Royal Society Interface*, 11(101):20140924.

Páez, A. and Scott, D. M. (2005). Spatial statistics for urban analysis: a review of techniques with examples. *GeoJournal*, 61(1):53–67.

Paquot, T. (2010). *L'abc de l'urbanisme*. IAU - UPEC.

Pumain, D. (2012). Urban systems dynamics, urban growth and scaling laws: The question of ergodicity. In *Complexity Theories of Cities Have Come of Age*, pages 91–103. Springer.

Raimbault, J. (2017). Calibration of a Density-based Model of Urban Morphogenesis. *ArXiv e-prints*.

Raimbault, J. (2017a). Identification de causalités dans des données spatio-temporelles. In *Spatial Analysis and GEOmatics 2017*.

Raimbault, J. (2017b). Modeling the co-evolution of urban form and transportation networks. In *Conference on Complex Systems 2017*.

Raimbault, J. (2018). Modeling the co-evolution of cities and networks. *ArXiv e-prints*.

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.

Schwarz, N. (2010). Urban form revisited—selecting indicators for characterising european cities. *Landscape and Urban Planning*, 96(1):29 – 47.

Trépanier, M., Morency, C., and Agard, B. (2009). Calculation of transit performance measures using smartcard data. *Journal of Public Transportation*, 12(1):5.

Tsai, Y.-H. (2005). Quantifying urban form: compactness versus' sprawl'. *Urban studies*, 42(1):141–161.

Zhang, T. and Zhou, B. (2014). Test for the first-order stationarity for spatial point processes in arbitrary regions. *Journal of agricultural, biological, and environmental statistics*, 19(4):387–404.

Zhang, Y., Li, X., Wang, A., Bao, T., and Tian, S. (2015). Density and diversity of openstreetmap road networks in china. *Journal of Urban Management*, 4(2):135–146.

Zheng, S. and Zheng, J. (2014). Assessing the completeness and positional accuracy of openstreetmap in china. In *Thematic Cartography for the Society*, pages 171–189. Springer.