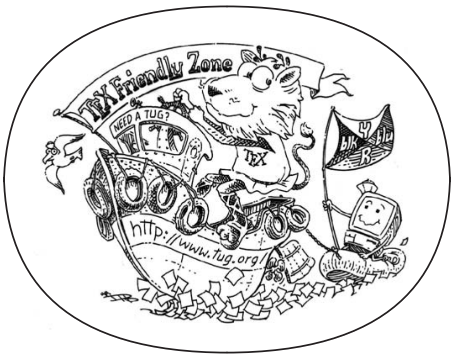
TOWARDS MODELS COUPLING URBAN GROWTH AND TRANSPORTATION NETWORK GROWTH

JUSTE RAIMBAULT



An Homage to The Elements of Typographic Style

February 2016 – version 0.1

ABSTRACT

Short summary of the contents... a great guide by Kent Beck how to write good abstracts can be found here:

https://plg.uwaterloo.ca/~migod/research/beckOOPSLA.html

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LISTINGS

ACRONYMS

DRY Don't Repeat Yourself

API Application Programming Interface

UML Unified Modeling Language

INTRODUCTION

Part I

1 QUANTITATIVE EPISTEMOLOGY

1.1 INTRODUCTION

1.2 ALGORITHMIC SYSTEMATIC REVIEW

1.2.1 *In search of models of co-evolution*

A broad bibliographical study suggests a scarcity of quantitative models of simulation integrating both network and urban growth. This absence may be due to diverging interests of concerned disciplines, resulting in a lack of communication. We propose to proceed to an algorithmic systematic review to give quantitative elements of answer to this question. A formal iterative algorithm to retrieve corpuses of references from initial keywords, based on text-mining, is developed and implemented. We study its convergence properties and do a sensitivity analysis. We then apply it on queries representative of the specific question, for which results tend to confirm the assumption of disciplines compartmentalization .

Transportation networks and urban land-use are known to be strongly coupled components of urban systems at different scales [6]. One common approach is to consider them as co-evolving, avoiding misleading interpretations such as the myth of structural effect of transportation infrastructures [19]. A question rapidly arising is the existence of models endogeneizing this co-evolution, i.e. taking into account simultaneous urban and network growth. We try to answer it using an algorithmic systematic review. We propose in this section, after a brief state of the art of existing literature, to present such an approach by formalizing the algorithm, which results are then presented and discussed.

1.2.2 *Modeling Interactions between Urban Growth and Network Growth : An Overview*

1.2.2.1 *Land-Use Transportation Interaction Models.*

A wide class of models that have been developed essentially for planning purposes, which are the so-called Land-use Transportation Interaction Models, is a first type answering our research question. See recent reviews (Chang, 2006, Iacono et al., 2008), (Wegener and Furst, 2004) to get an idea of the heterogeneity of included approaches, that exist for more than 30 years. Recent models with diverse refinements are still developed today, such as (Delons et al., 2008) which includes housing market for Paris area. Diverse aspects of the same system can be translated into many models (see e.g. (Wegener et al., 1991)), and traffic, residential and employment dynamics, resulting land-use evolution, influenced also by a static transportation network, are generally taken into account.

1.2.2.2 *Network Growth Approaches.*

On the contrary, many economic literature has done the opposite of previous models, i.e. trying to reproduce network growth given assumptions on the urban landscape, as reviewed in (Zhang and Levinson, 2007) . In (Xie and Levinson, 2009), economic empirical studies are positioned within other network growth approaches, such as work by physicists giving model of geometrical network growth (Barthélemy and Flammini, 2008). Analogy with biological networks was also done, reproducing typical robustness properties of transportation networks (Tero et al., 2010).

1.2.2.3 *Hybrid Approaches.*

Fewer approaches coupling urban growth and network growth can be found in the literature. [4] couples density evolution with network growth in a toy model. In [23], a simple Cellular Automaton coupled with an evolutive network reproduces stylized facts of human settlements described by Le Corbusier. At a smaller scale, [1] proposes a model of co-evolution between roads and buildings, following geometrical rules. These approaches stay however limited and rare.

1.2.3 *Bibliometric Analysis*

Literature review is a crucial preliminary step for any scientific work and its quality and extent may have a dramatic impact on research quality. Systematic review techniques have been developed, from qualitative review to quantitative meta-analyses allowing to produce new results by combining existing studies (Rucker, 2012) . Ignoring some references can even be considered as a scientific mistake in the context of emerging information systems [16]. We aim to take advantage of such techniques to tackle our issue. Indeed, observing the form of the bibliography obtained in previous section raises some hypothesis. It is clear that all components are present for co-evolutive models to exist but different concerns and objectives seem to stop it. As it was shown by (Commenges, 2013) for the concept of mobility, for which a "small world of actors" relatively closed invented a notion ad hoc, using models without accurate knowledge of a more general scientific context, we could be in an analog case for the type of models we are interested in. Restricted interactions between scientific fields working on the same objects but with different purposes, backgrounds and

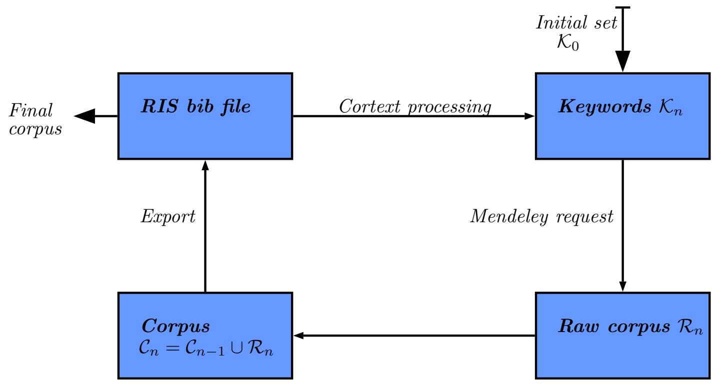


Figure 1: Global workflow of the algorithm, including implementation details : catalog request is done through Mendeley API ; final state of corpuses are RIS files.

at different scales, could be at the origin of the relative absence of co-evolving models. While most of studies in bibliometrics rely on citation networks (Newman, 2013) or co-autorship networks (Sarigöl et al., 2014), we propose to use a less explored paradigm based on text-mining introduced in [7], that obtain a dynamic mapping of scientific disciplines based on their semantic content. For our question, it has a particular interest, as we want to understand content structure of researches on the subject. We propose to apply an algorithmic method described in the following. The algorithm proceeds by iterations to obtain a stabilized corpus from initial keywords, reconstructing scientific semantic landscape around a particular subject.

1.2.3.1 *Description of the Algorithm*

Let A be an alphabet, A^* corresponding words and $T = \cup_{k \in \mathbb{N}} A^{*k}$ texts of finite length on it. A reference is for the algorithm a record with text fields representing title, abstract and keywords. Set of references at iteration n will be denoted $\mathcal{C} \subset T^3$. We assume the existence of a set of keywords $\mathcal{K}_n$, initial keywords being $\mathcal{K}_0$. An iteration goes as follows :

1. A raw intermediate corpus $\mathcal{R}_n$ is obtained through a catalog request providing previous keywords $\mathcal{K}_{n-1}$.
2. Overall corpus is actualized by $\mathcal{C}_n = \mathcal{C}_{n-1} \cup \mathcal{R}_n$.
3. New keywords $\mathcal{K}_n$ are extracted from corpus through Natural Language Processing treatment, given a parameter N_k fixing the number of keywords.

The algorithm stops when corpus size becomes stable or a user-defined maximal number of iterations has been reached. Fig. 1 shows the global workflow.

1.2.3.2 *Results*

IMPLEMENTATION Because of the heterogeneity of operations required by the algorithm (references organisation, catalog requests, text processing), it was found a reasonable choice to implement it in Java. Source code is available on the Github repository of the project1. Catalog request, consisting in retrieving a set of references from a set of keywords, is done using the Mendeley software API (Mendeley, 2015) as it allows an open access to a large database. Keyword extraction is done by Natural Language Processing (NLP) techniques, following the workflow given in (Chavalarias and Cointet, 2013), calling a Python script that uses (NLTK, 2006). Convergence and Sensitivity Analysis. A formal proof of algorithm convergence is not possible as it will depend on the empirical unknown structure of request results and keywords extraction. We need thus to study empirically its behavior. Good convergence properties but various sensitivities to Nk were found as presented in Fig. 2. We also studied the internal lexical consistence of final corpuses as a function of keywords number. As expected, small number yields more consistent corpuses, but the variability when increasing stays reasonable.

Fig. 2 : Convergence and sensitivity analysis. Left : Plots of number of references as a function of iteration, for various queries linked to our theme (see further), for various values of Nk (from 2 to 30). We obtain a rapid convergence for most cases, around 10 iterations needed. Final number of references appears to be very sensitive to keyword number depending on queries, what seems logical since encountered landscape should strongly vary depending on terms. Right : Mean lexical consistence and standard error bars for various queries, as a function of keyword number. Lexical consistence is defined though co-occurrences of keywords by, with N final number of keywords, f final step, and c(i) co-occurrences in references, . The stability confirms the consistence of final corpuses.

Once the algorithm is partially validated, we apply it to our question. We start from five different initial requests that were manually extracted from the various domains identified in the manual bibliography (that are "city system network", "land use transport interaction", "network urban modeling", "population density transport", "transportation network urban growth"). We take the weakest assumption on parameter $N_k = 100$, as it should less constrain reached domains. After having constructed corpuses, we study their lexical distances as an indicator to answer our initial question. Large distances would go in the direction of the assumption made in section 2, i.e. that discipline self-centering may be at the origin of the lack of interest for co-evolutive models. We show in Table 1 values of relative lexical proximity, that appear to be significantly low, confirming this assumption.

Table 1 : Symmetric matrix of lexical proximities between final corpuses, defined as the sum of overall final keywords co-occurrences between corpuses, normalized by number of final keywords (100). We obtain very low values, confirming that corpuses are significantly far.

Further work is planned towards the construction of citation networks through an automatic access to Google Scholar that provides backward citations. The confrontation of inter-cluster coefficients on the citation network for the different corpuses with our lexical consistence results are an essential aspect of a further validation of our results.

The disturbing absence of models simulating the co-evolution of transportation networks and urban land-use, confirmed through a state-of-the-art covering many domain, may be due to the absence of communication between scientific disciplines studying different aspects of that problems. We have proposed an algorithmic method to give elements of answers through text-mining-based corpus extraction. First numerical results seem to confirm the assumption. However, such a quantitative analysis should not be considered alone, but rather come as a back-up for qualitative studies that will be the object of further work, such as the one lead in (Commenges, 2013), in which questionnaires with historical actors of modeling provide highly relevant information.

1.3 TOWARDS MODELING PURPOSE AND CONTEXT AUTOMATIC EXTRACTION

2 METHODOLOGICAL DEVELOPMENTS

2.1 AN UNIFIED FRAMEWORK FOR STOCHASTIC MODELS OF URBAN GROWTH

2.1.1 *Introduction*

Various stochastic models aiming to reproduce population patterns on large temporal and spatial scales (city systems) have been discussed across various fields of the litterature, from economics to geography, including models proposed by physicists. We propose a general framework that allows to include different famous models (in particular Gibrat, Simon and Preferential Attachment model) within an unified vision. It brings first an insight into epistemological debates on the relevance of models. Furthermore, bridges between models lead to the possible transfer of analytical results to some models that are not directly tractable.

2.1.1.1 *Context*

General biblio.

Precise type of models : mathematical models ; stay to a certain level of tractability as essence of our approach is link between models. No clear definition, includes all models that can be linked in the sense of *Generalization/Particularization/Limit case/?*.

2.1.1.2 *Notations*

2.1.2 *Framework*

2.1.2.1 *Formulation*

PRESENTATION What we propose as a framework can be understood as a meta-model in the sense of [10], i.e. an modular general modeling process within each model can be understood as a limit case or as a specific case of another model. More simply it shoud be a diagram of formal relations between models. The ontological aspect is also tackled by embedding the diagram into an ontological state space (which discretization corresponds to the "bricks" of the incremental construction of [10]). It constructs a sort of model classification.

2.1.2.2 *Models Included*

The following models are included in our framework. The list is arbitrary but aims to offer a broad view of disciplines concerned

2.1.2.3 *Thematic Classification*

2.1.2.4 *Framework Formulation*

Diagram linking various models ; first embedded into time/population plane, cases Discrete/Continous. Other aspects more sparse (ex. spatialization) ; how represent it ?

2.1.3 *Models formulation*

2.1.4 *Derivations*

2.1.4.1 *Generalization of Preferential Attachment*

See [24].

2.1.4.2 *Link between Gibrat and Preferential Attachment Models*

Let consider a stricly positive growth Gibrat model given by $P_i(t) = R_i(t) \cdot P_i(t-1)$ with $R_i(t) > 1$, $\mu_i(t) = \mathbb{E}[R_i(t)]$ and $\sigma_i(t) = \mathbb{E}\left[R_i(t)^2\right]$. On the other hand, we take a simple preferential attachment, with fixed attachment probability $\lambda \in [0,1]$ and new arrivants number $m > 0$. We derive that Gibrat model can be statistically equivalent to a limit of the preferential attachment model, assuming that the moment-generating function of $R_i(t)$ exists. Classical distributions that could be used in that case, e.g. log-normal distribution, are entirely defined by two first moments, making this assumption reasonable.

Lemma 1 *The limit of a Preferential Attachment model when $\lambda \ll 1$ is a linear-growth Gibrat model, with limit parameters $\mu_i(t) = 1 + \frac{\lambda}{m\cdot(t-1)}$.*

Proof Starting with first moment, we denote $\bar{P}_i(t) = \mathbb{E}[P_i(t)]$. Independance of Gibrat growth rate yields directly $\bar{P}_i(t) = \mathbb{E}[R_i(t)] \cdot \bar{P}_i(t-1)$. Starting for the preferential attachment model, we have $\bar{P}_i(t) = \mathbb{E}[P_i(t)] = \sum_{k=0}^{+\infty} k\,\mathbb{P}[P_i(t) = k]$. But

$$\{P_i(t) = k\} = \bigcup_{\delta=0}^{\infty} \left(\{P_i(t-1) = k - \delta\} \cap \{P_i \leftarrow P_i + 1\}^{\delta}\right)$$

where the second event corresponds to city i being increased δ times between $t-1$ and t (note that events are empty for $\delta \geqslant k$). Thus, being careful on the conditional nature of preferential attachment formulation, stating that $\mathbb{P}[\{P_i \leftarrow P_i + 1\} | P_i(t-1) = p] = \lambda \cdot \frac{p}{P(t-1)}$ (total population $P(t)$ assumed deterministic), we obtain

$$\begin{aligned}\mathbb{P}[\{P_i \leftarrow P_i+1\}] &= \sum_p \mathbb{P}[\{P_i \leftarrow P_i+1\}|P_i(t-1)=p]\cdot\mathbb{P}[P_i(t-1)=p]\\ &= \sum_p \lambda\cdot\frac{p}{P(t-1)}\,\mathbb{P}[P_i(t-1)=p] = \lambda\cdot\frac{\bar{P}_i(t-1)}{P(t-1)}\end{aligned}$$

It gives therefore, knowing that $P(t-1) = P_0 + m\cdot(t-1)$ and denoting $q = \lambda\cdot\frac{\bar{P}_i(t-1)}{P_0+m\cdot(t-1)}$

$$\begin{aligned}\bar{P}_i(t) &= \sum_{k=0}^{\infty}\sum_{\delta=0}^{\infty} k\cdot\left(\lambda\cdot\frac{\bar{P}_i(t-1)}{P_0+m\cdot(t-1)}\right)^{\delta}\cdot\mathbb{P}[P_i(t-1)=k-\delta]\\ &= \sum_{\delta'=0}^{\infty}\sum_{k'=0}^{\infty}\left(k'+\delta'\right)\cdot q^{\delta'}\cdot\mathbb{P}\left[P_i(t-1)=k'\right]\\ &= \sum_{\delta'=0}^{\infty} q^{\delta'}\cdot\left(\delta'+\bar{P}_i(t-1)\right) = \frac{q}{(1-q)^2}+\frac{\bar{P}_i(t-1)}{(1-q)} = \frac{\bar{P}_i(t-1)}{1-q}\left[1+\frac{1}{\bar{P}_i(t-1)}\frac{q}{(1-q)}\right]\end{aligned}$$

As it is not expected to have $\bar{P}_i(t) \ll P(t)$ (fat tail distributions), a limit can be taken only through λ. Taking $\lambda \ll 1$ yields, as $0 < \bar{P}_i(t)/P(t) < 1$, that $q = \lambda\cdot\frac{\bar{P}_i(t-1)}{P_0+m\cdot(t-1)} \ll 1$ and thus we can expand in first order of q, what gives $\bar{P}_i(t) = \bar{P}_i(t-1)\cdot\left[1+\left(1+\frac{1}{\bar{P}_i(t-1)}\right)q+o(q))\right]$

$$\bar{P}_i(t) \simeq \left[1+\frac{\lambda}{P_0+m\cdot(t-1)}\right]\cdot\bar{P}_i(t-1)$$

It means that this limit is equivalent in expectancy to a Gibrat model with $\mu_i(t) = \mu(t) = 1+\frac{\lambda}{P_0+m\cdot(t-1)}$.

For the second moment, we can do an analog computation. We have still $\mathbb{E}\left[P_i(t)^2\right] = \mathbb{E}\left[R_i(t)^2\right]\cdot\mathbb{E}\left[P_i(t-1)^2\right]$ and $\mathbb{E}\left[P_i(t)^2\right] = \sum_{k=0}^{+\infty} k^2\,\mathbb{P}[P_i(t)=k]$. We obtain the same way

$$\begin{aligned}\mathbb{E}\left[P_i(t)^2\right] &= \sum_{\delta'=0}^{\infty}\sum_{k'=0}^{\infty}\left(k'+\delta'\right)^2\cdot q^{\delta'}\cdot\mathbb{P}\left[P_i(t-1)=k'\right] = \sum_{\delta'=0}^{\infty} q^{\delta'}\cdot\left(\mathbb{E}\left[P_i(t-1)^2\right]+2\delta'\bar{P}_i(t-1)+\delta'^2\right)\\ &= \frac{\mathbb{E}\left[P_i(t-1)^2\right]}{1-q}+\frac{2q\bar{P}_i(t-1)}{(1-q)^2}+\frac{q(q+1)}{(1-q)^3} = \frac{\mathbb{E}\left[P_i(t-1)^2\right]}{1-q}\left[1+\frac{q}{\mathbb{E}[P_i(t-1)^2]}\left(\frac{2\bar{P}_i(t-1)}{1-q}+\right.\right.\end{aligned}$$

2.1.4.3 *Link between Simon and Preferential Attachment*

2.1.4.4 *Link between Favaro-Pumain and Gibrat*

[11]

2.1.4.5 *Link between Bettencourt-West and Simon*

[5]

2.1.4.6 *Other Models*

[12] : Economic model giving a Simon equivalent formulation. Finds that in upper tail, proportional growth process occurs. We find the same result as a consequence of 2.1.4.3.

2.1.5 *Application*

2.1.6 *Discussion*

Conclusion

2.2 ANALYTICAL SENSITIVITY OF URBAN SCALING LAWS TO SPATIAL EXTENT

2.2.1 *Introduction*

Scaling laws have been shown to be universal of urban systems at many scales and for many indicators. Recent studies question however the consistence of scaling exponents determination, as their value can vary significantly depending on thresholds used to define urban entities on which quantities are integrated, even crossing the qualitative border of linear scaling, from infralinear to supralinear scaling. We use a simple theoretical model of spatial distribution of densities and urban functions to show analytically that such behavior can be derived as a consequence of the type of spatial distribution and the method used. Numerical simulation confirm the theoretical results and reveals that results are reasonably independant of spatial kernel used to distribute density.

Scaling laws for urban systems, starting from the well-known rank-size Zipf's law for city size distribution [12], have been shown to be a recurrent feature of urban systems, at many scales and for many types of indicators. They reside in the empirical constatation that indicators computed on elements of an urban system, that can be cities for system of cities, but also smaller entities at a smaller scale, do fit relatively well a power-law distribution as a function of entity size, i.e. that for entity i with population P_i, we have for an integrated quantity A_i, the relation $A_i \simeq A_0 \cdot \left(\frac{P_i}{P_0}\right)^{\alpha}$. Scaling exponent α can be smaller or greater than 1, leading to infra- or supralinear effects. Various thematic interpretation of this phenomena have been proposed, typically under the form of processes analysis. The economic literature has produced abundant work on the subject (see [13] for a review), but that are generally poorly spatialized, thus of poor interest to our approach that deals precisely with spatial organization. Simple economic rules such as energetic equilibria can lead to simple power-laws [5] but are difficult to fit empirically. A interesting proposition by Pumain is that they are intrinsically due to the evolutionnary character of city systems, where complex emergent interaction between cities generate such global distributions [21]. Although a tempting parallel can be done with self-organizing biological systems, Pumain insists on the fact that the ergodicity assumption for such systems is not reasonable in the case of geographical systems and that the analogy cannot be exploited [20]. Other explanations have been proposed at other scales, such as the urban growth model at the mesoscopic scale (city scale) given in [17] that shows that the congestion within transportation networks may be one reason for city shapes and corresponding scaling laws. Note that "classic" urban growth models such as Gibrat's model do provide first order approximation of scaling systems, but that interactions between agaents have to be incorporated into the model to obtain better fit on real data, such as the Favaro-Pumain model for innovation cycles propagation proposed in [11], that generalize a Gibrat model and provide better fits on data for French cities.

However, the blind application of scaling exponents computations was recently pointed as misleading in most cases [18], confirmed by empirical works such as [3] that showed the variability of computed exponents to the parameters defining urban areas, such as density thresholds. An ongoing work by Cottineau & *al.* presented at [9], studies empirically for French Cities the influence of 3 parameters playing a role in city definition, that are a density threshold θ to delimitate boundaries of an urban area, a number of commuters threshold θ_c that is the proportion of commuters going to core area over which the unity is considered belonging to the area, and a cut-off parameter P_c under which entities are not taken into account for the linear regression providing the scaling exponent. Remarquable results are that exponents can significantly vary and move from infralinear to supralinear when threshold varies. A systematic exploration of parameter space produces phase diagrams of exponents for various quantities. One question raising immediately is how these variation can be explained by the features of spatial distribution of variables. Do they result from intrinsic mechanisms present in the system or can they be explained more simply by the fact that the system is particularly spatialized ? We propose to prove by the tractation of a toy analytical model that even simple distributions can lead to such significant variations in the exponents, along one dimension of parameters (density threshold), directing the response towards the second explanation. The rest of the paper is organized as follows : we formalize the simple framework used and derive an analytical relation between estimated exponent and density threshold parameter. We then present a numerical implementation of the model that confirms numerically theoretical results, explore other form of kernels that would be less tractable, and study the sensitivity along two parameters. We finally discuss the implications of our results and further work needed.

2.2.2 *Formalization*

We formalize the simple theoretical context in which we will derive the sensitivity of scaling to city definition. Let consider a polycentric city system, which spatial density distributions can be reasonably constructed as the superposition of monocentric fast-decreasing spatial kernels, such as an exponential mixture model [2]. Taking a geographical space as $\mathbb{R}^2$, we take for any $\vec{x} \in \mathbb{R}^2$ the density of population as

$$d(\vec{x}) = \sum_{i=1}^{N} d_i(\vec{x}) = \sum_{i=1}^{N} d_i^0 \cdot \exp\left(\frac{-\|\vec{x} - \vec{x}_i\|}{r_i}\right) \tag{1}$$

where r_i are spread parameters of kernels, d_i^0 densities at origins, $\vec{x}_i$ positions of centers. We furthemore assume the following constraints :

1. To simplify, cities are monocentric, in the sense that for all $i \neq j$, we have $\|\vec{x}_i - \vec{x}_j\| \gg r_i$.
2. It allows to impose structural scaling in the urban system by the simple constraint on city populations P_i. One can compute by integration that $P_i = 2\pi d_i^0 r_i^2$, what gives by injection into the scaling hypothesis $\ln P_i = \ln P_{max} - \alpha \ln i$, the following relation between parameters : $\ln\left[d_i^0 r_i^2\right] = K' - \alpha \ln i$.

To study scaling relations, we consider a random scalar spatial variable $a(\vec{x})$ representing one aspect of the city, that can be everything but has the dimension of a spatial density, such that the indicator $A(D) = \mathbb{E}\left[\iint_D a(\vec{x}) d\vec{x}\right]$ represents the expected quantity of a in area D. We make the assumption that $a \in \{0; 1\}$ ("counting" indicator) and that its law is given by $\mathbb{P}[a(\vec{x}) = 1] = f(d(\vec{x}))$. Following the empirical work done in [9], the integrated indicator on city i as a function of θ is given by

$$A_i(\theta) = A(D(\vec{x}_i, \theta))$$

where $D(\vec{x}_i, \theta)$ is the area centered in $\vec{x}_i$ where $d(\vec{x}) > \theta$. Assumption 1 ensures that the area are roughly disjoint circles. We take furthermore a simple amenity such that it follows a local scaling law in the sense that $f(d) = \lambda \cdot d^\beta$. It seems a reasonable assumption since it was shown that many urban variable follow a fractal behavior at the intra-urban scale [15] and that it implies necessarily a power-law distribution [8]. We make the additional assumption that $r_i = r_0$ does not depend on i, what is reasonable if the urban system is considered from a large scale. This assumption should be relaxed in numerical simulations. The estimated scaling exponent $\alpha(\theta)$ is then the result of the log-regression of $(A_i(\theta))_i$ against $(P_i(\theta))_i$ where $P_i(\theta) = \iint_{D(\vec{x}_i, \theta)} d$.

2.2.3 *Analytical Derivation of Sensitivity*

With above notations, let derive the expression of estimated exponent for quantity a as a function of density threshold parameter θ. The quantity computed for a given city i is, thanks to the monocentric assumption and in a spatial range and a range for θ such that $\theta \gg \sum_{j\neq i} d_j(\vec{x})$, allowing to approximate $d(\vec{x}) \simeq d_i(\vec{x})$ on $D(\vec{x}_i, \theta)$, is computed by

$$A_i(\theta) = \lambda \cdot \iint_{D(\vec{x}_i,\theta)} d^{\beta} = 2\pi\lambda {d_i^0}^{\beta} \int_{r=0}^{r_0 \ln \frac{d_i^0}{\theta}} r \exp\left(-\frac{r\beta}{r_0}\right) dr$$
$$= \frac{2\pi {d_i^0}^{\beta} r_0^2}{\beta^2} \left[1 + \beta \ln \frac{\theta}{d_i^0} \left(\frac{\theta}{d_i^0}\right)^{\beta} - \left(\frac{\theta}{d_i^0}\right)^{\beta}\right]$$

We obtain in a similar way the expression of $P_i(\theta)$

$$P_i(\theta) = 2\pi d_i^0 r_0^2 \left[1 + \ln\left[\frac{\theta}{d_i^0}\right] \frac{\theta}{d_i^0} - \frac{\theta}{d_i^0}\right]$$

The Ordinary-Least-Square estimation, solving the problem $\inf_{\alpha,C} \|(\ln A_i(\theta) - C - \alpha \ln P_i(\theta))_i\|^2$, gives the value $\alpha(\theta) = \frac{\mathrm{Cov}[(\ln A_i(\theta))_i, (\ln P_i(\theta))_i]}{\mathrm{Var}[(\ln P_i(\theta))_i]}$. As we work on city boundaries, threshold is expected to be significantly smaller than center density, i.e. $\theta/d_i^0 \ll 1$. We can develop the expression in the first order of θ/d_i^0 and use the global scaling law for city sizes, what gives $\ln A_i(\theta) \simeq K_A - \alpha \ln i + (\beta - 1) \ln d_i^0 + \beta \ln \frac{\theta}{d_i^0} \left(\frac{\theta}{d_i^0}\right)^{\beta}$ and $\ln P_i(\theta) = K_P - \alpha \ln i + \ln\left[\frac{\theta}{d_i^0}\right] \frac{\theta}{d_i^0}$. Developping the covariance and variance gives finally an expression of the scaling exponent as a function of θ, where $k_j, {k_j}'$ are constants obtained in the development :

$$\alpha(\theta) = \frac{k_0 + k_1\theta + k_2\theta^{\beta} + k_3\theta^{\beta+1} + k_4\theta\ln\theta + k_5\theta^{\beta}\ln\theta + k_6\theta^{\beta}(\ln\theta)^2 + k_7\theta^{\beta+1}(\ln\theta)^2 + k_8\theta^{\beta+1}\ln\theta}{k'_0 + k'_1\ln\theta + k'_2\theta\ln\theta + k'_3\theta^2 + k'_4\theta^2\ln\theta + k'_5\theta^2(\ln\theta)^2} \tag{2}$$

This rational fraction predicts the evolution of the scaling exponent when the threshold varies. We study numerically its behavior in the next section, among other numerical experiments.

2.2.4 *Numerical Simulations*

IMPLEMENTATION We implement empirically the density model given in section 2.2.2. Centers are successively chosen such that in a given region of space only one kernel dominates in the sense that the sum of other contributions are above a given threshold θ_e. In practice, adapting N to world size allows to respect the monocentric

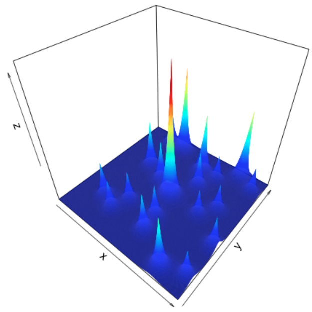


Figure 2: Example of a synthetic density distribution obtained with the exponential mixture, with a grid of size 400×400 and parameters $N = 20$, $r_0 = 10$, $P_{max} = 200$, $\alpha = 0.5$, $\theta_C = 0.01$.

Figure 3: Validation of theoretical result through numerical simulation.

condition. Population are distributed in order to follow the scaling law with fixed α and r_i (arbitrary choice) by computing corresponding d_i^0. Technical details of the implementation done in R [22] and using the package `kernlab` for efficient kernel mixture methods [14] are given as comments in source code[1]. We show in figure 2 example of synthetic density distributions on which the numerical study is conducted. Theoretical result obtained in Eq. 2 are studied and confronted to emprically computed values for various parameter as shown in Fig. 3.

RANDOM PERTURBATIONS The simple model used is quite reducing for maximal densities and radius distribution. We proceed to an empirical study of the influence of noise in the system by fixing d_i^0 and r_i the following way :

- d_i^0 follows a reversed log-normal distribution with maximal value being a realistic maximal density
- Radiuses are computed to respect rank-size law and then perturbated by a white noise.

Results shown in Fig. 4 are quantitatively different from previous one, as expected, but the same qualitative behavior is reproduced.

KERNEL TYPE We test the influence of the type of spatial kernel used on results. We test gaussian kernels and quadratic kernels with parameters within reasonable ranges analog to the exponential kernel.

1 available at `https://github.com/JusteRaimbault/CityNetwork/tree/master/Models/Scaling`

Figure 4: Variation of exponents with variable origin density and radius.

Figure 5: Scaling exponents for other kernels.

Figure 6: Two parameters phase diagram.

As shown in Fig. 5, we obtain the same qualitative results that is the significant variation of $\alpha(\theta)$ as a function of θ.

TWO-PARAMETERS PHASE DIAGRAM We introduce now a second spatial variable that has also an influence on the definition of urban entities, that is the proportion of actives working in city center, as done on empirical data in [9]. To simplify, it is used only to define urban parameter but assumed as having no influence on the local probability distribution of the amenity which stays the same function of the density. We write

2.2.5 *Discussion*

Part II

MODELING AND EMPIRICAL ANALYSIS

3 EMPIRICAL ANALYSIS

4 MODELING

4.1 A SIMPLE MODEL OF URBAN GROWTH

4.2 CORRELATED GENERATION OF TERRITORIAL CONFIGURATIONS

4.2.1 *Application : geographical data of density and network*

4.2.1.1 *Context*

The use of synthetic data in geography is generally directed towards the generation of synthetic populations within agent-based models (mobility, *LUTI* models) [**pritchard2009advances**]. We can make a weak link with some Spatial Analysis techniques. The extrapolation of a continuous spatial field from a discrete spatial sample through a kernel density estimation for example can be understood as the creation of a synthetic dataset (even if it is not generally the initial view, as in Geographically Weighted Regression [**brunsdon1998geographically**] in which variable size kernels do not interpolate data *stricto sensu* but extrapolate abstract variables representing interaction between explicit variables). In the field of modeling in quantitative geography, *toy-models* or hybrid models require a consistent initial spatial configuration. A set of possible initial configurations becomes a synthetic dataset on which the model is tested. The first Simpop model [**sanders1997simpop**], precursor of a large family of models later parametrized with real data, could enter that frame but was studied on an unique synthetic spatialization. Similarly underlined was the difficulty to generate an initial transportation infrastructure in the case of the SimpopNet model [**schmitt2014modelisation**] although it was admitted as a cornerstone of knowledge on the behavior of the model. A systematic control of spatial configuration effects on the behavior of simulation models was only recently proposed [**cottineau2015revisiting**], approach that can be interpreted as a statistical control on spatial data. The aim is to be able to distinguish proper effects due to intrinsic model dynamics from particular effects due to the geographical structure of the case study. Such results are essential for the validation of conclusions obtained with modeling and simulation practices in quantitative geography.

4.2.1.2 *Formalization*

We propose in our case to generate territorial systems summarized in a simplified way as a spatial population density $d(\vec{x})$ and a transportation network $n(\vec{x})$. Correlations we aim to control are correlations between urban morphological measures and network measures. The question of interactions between territories and networks is already well-studied [**offner1996reseaux**] but stays highly complex and difficult to quantify [19]. A dynamical modeling of implied processes should shed light on these interactions ([**bretagnolle:tel-00459720**], p. 162-163). We develop in that frame a *simple* coupling (i.e. without any feedback loop) between a density distribution model and a network morphogenesis model.

DENSITY MODEL We use a model D similar to aggregation-diffusion models [**batty2006hierarchy**] to generate a discrete spatial distribution of population density. A generalization of the basic model is proposed in [**raimbault2016calibration**], providing a calibration on morphological objectives (entropy, hierarchy, spatial auto-correlation, mean distance) against real values computed on the set of 50km sized grid extracted from european density grid [**eurostat**]. More precisely, the model proceeds iteratively the following way. An square grid of width N, initially empty, is represented by population $(P_i(t))_{1 \leqslant i \leqslant N^2}$. 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 $\mathbb{P}[P_i(t+1) = P_i(t) + 1 | P(t+1) = \frac{(P_i(t)/P(t))^\alpha}{\sum (P_i(t)/P(t))^\alpha}$
- a fraction β of population is diffused to four closest neighbors is operated n_d times

The two contradictory 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.

NETWORK MODEL On the other hand, we are able to generate a planar transportation network by a model N, at a similar scale and given a density distribution. Because of the conditional nature to the density of the generation process, we will first have conditional estimators for network indicators, and secondly natural correlations between network and urban shapes should appear as processes are not independent. The nature and modularity of these correlations as a function of model parameters are still to determine by exploration of the coupled model.

The heuristic network generation procedure is the following :

1. A fixed number N_c of centers that will be first nodes of the network si distributed given density distribution, following a similar law to the aggregation process, i.e. the probability to be distributed in a given patch is $\frac{(P_i/P)^\alpha}{\sum (P_i/P)^\alpha}$. Population is then attributed according to Voronoi areas of centers, such that a center cumulates population of patches within its extent.

2. Centers are connected deterministically by percolation between closest clusters : as soon as network is not connected, two closest connected components in the sense of minimal distance between each vertices are connected by the link realizing this distance. It yields a tree-shaped network.

3. Network is modulated by potential breaking in order to be closer from real network shapes. More precisely, a generalized gravity potential between two centers i and j is defined by

$$V_{ij}(d) = \left[(1-k_h) + k_h \cdot \left(\frac{P_i P_j}{P^2}\right)^{\gamma}\right] \cdot \exp\left(-\frac{d}{r_g(1+d/d_0)}\right)$$

   where d can be euclidian distance $d_{ij} = d(i,j)$ or network distance $d_N(i,j)$, $k_h \in [0,1]$ a weight to modulate role of populations, γ giving shape of the hierarchy across population values, r_g characteristic interaction distance and d_0 distance shape parameter.

4. A fixed number $K \cdot N_L$ of potential new links is taken among couples having greatest euclidian distance potential ($K = 5$ is fixed).

5. Among potential links, N_L are effectively realized, that are the one with smallest rate $\tilde{V}_{ij} = V_{ij}(d_N)/V_{ij}(d_{ij})$. At this stage only the gap between euclidian and network distance is taken into account : $\tilde{V}_{ij}$ does indeed not depend on populations and is increasing with d_N at constant d_{ij}.

6. Planarity of the network is forced by creation of nodes at possible intersections created by new links.

We insist on the fact that the network generation procedure is entirely heuristic and result of thematic assumptions (connected initial network, gravity-based link creation) combined with trial-and-error during first explorations. Other model types could be used as well, such biological self-generated networks [**TeroAl10**], local network growth based on geometrical constraints optimization [**barthelemy2008modeling**], or a more complex percolation model than the initial one that would allow the creation of loops for example. We could thus in the frame of a modular architecture, in which the choice between different implementations of a functional brick can be seen as a meta-parameter [10], choose network generation function adapted to a specific need (as e.g. proximity to real data, constraints on output indicators, variety if generated forms, etc.).

PARAMETER SPACE Parameter space for the coupled model[1] is constituted by density generation parameters $\vec{\alpha}_D = (P_m/N_G, \alpha, \beta, n_d)$

1 Weak coupling allows to limit the total number of parameters as a strong coupling would involve retroaction loops and consequently associated parameters to deter-

(we study for the sake of simplicity the rate between population and growth rate instead of both varying, i.e. the number of steps needed to generate the distribution) and network generation parameters $\vec{\alpha}_N = (N_C, k_h, \gamma, r_g, d_0)$. We denote $\vec{\alpha} = (\vec{\alpha}_D, \vec{\alpha}_N)$.

INDICATORS Urban form and network structure are quantified by numerical indicators in order to modulate correlations between these. Morphology is defined as a vector $\vec{M} = (r, \bar{d}, \varepsilon, a)$ giving spatial auto-correlation (Moran index), mean distance, entropy and hierarchy (see **[le2015forme]** for a precise definition of these indicators). Network measures $\vec{G} = (\bar{c}, \bar{l}, \bar{s}, \delta)$ are with network denoted (V, E)

- Mean centrality $\bar{c}$ defined as average *betweeness-centrality* (normalized in $[0, 1]$) on all links.
- Mean path length $\bar{l}$ given by $\frac{1}{d_m}\frac{2}{|V|\cdot(|V|-1)}\sum_{i<j} d_N(i,j)$ with d_m normalization distance taken here as world diagonal $d_m = \sqrt{2}N$.
- Mean network speed **[banos2012towards]** which corresponds to network performance compared to direct travel, defined as $\bar{s} = \frac{2}{|V|\cdot(|V|-1)}\sum_{i<j}\frac{d_{ij}}{d_N(i,j)}$.
- Network diameter $\delta = \max_{ij} d_N(i,j)$.

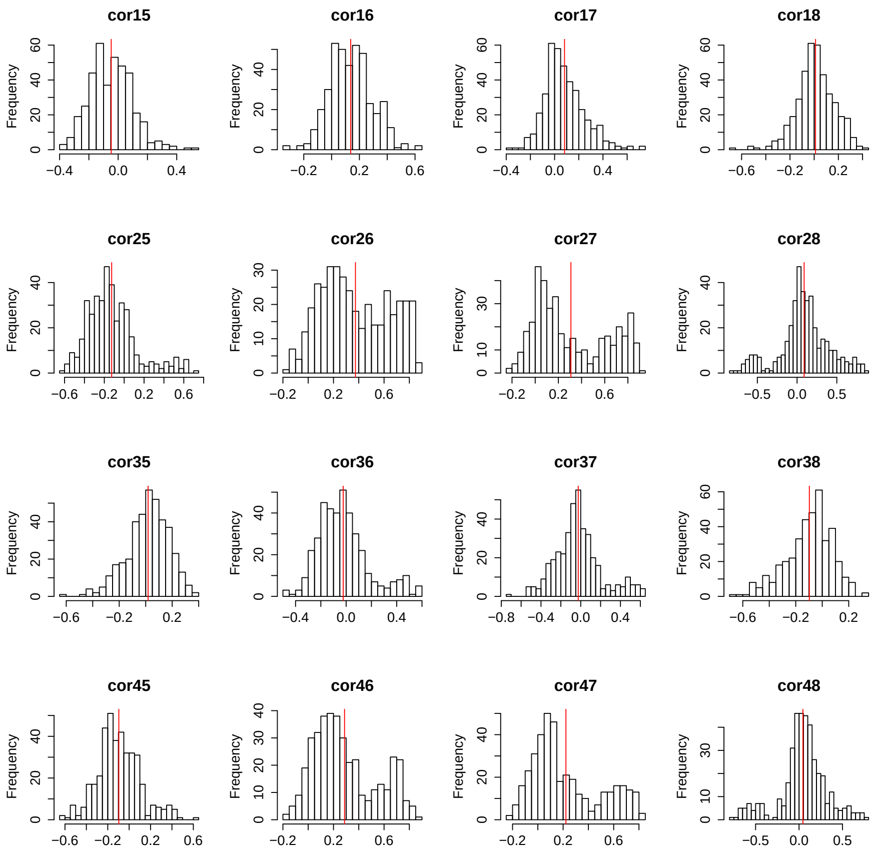
COVARIANCE AND CORRELATION We study the cross-correlation matrix $\text{Cov}\left[\vec{M}, \vec{G}\right]$ between morphology and network. We estimate it on a set of n realizations at fixed parameter values $(\vec{M}\left[D(\vec{\alpha})\right], \vec{G}\left[N(\vec{\alpha})\right])_{1\leqslant i\leqslant n}$ with standard unbiased estimator. We estimate correlation with associated Pearson estimator.

4.2.1.3 *Implementation*

Coupling of generative models is done both at formal and operational levels. We interface therefore independent implementations. The OpenMole software **[reuillon2013openmole]** for intensive model exploration offers for that the ideal frame thanks to its modular language allowing to construct *workflows* by task composition and interfacing with diverse experience plans and outputs. For operational reasons, density model is implemented in `scala` language as an OpenMole `plugin`, whereas network generation is implemented in agent-oriented language `NetLogo` **[wilensky1999netlogo]** because of its possibilities for interactive exploration and heuristic model construction. Source code is available for reproducibility on project repository[2].

mine their structure and intensity. In order to diminish it, an integrated model would be preferable to a strong coupling, what is slightly different in the sense where it is not possible in the integrated model to freeze one of the subsystems to obtain a model of the other subsystem that would correspond to the non-coupled model.

2 at `https://github.com/JusteRaimbault/CityNetwork/tree/master/Models/Synthetic`



0.35

(a) t

Figure 7

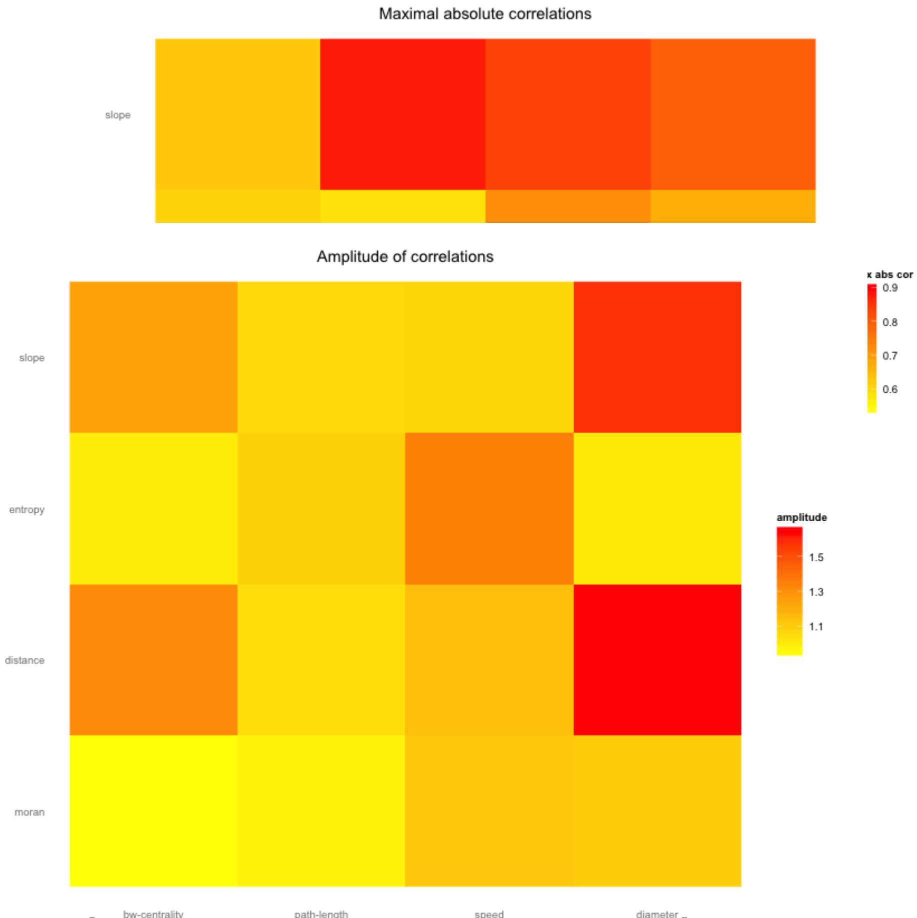


Figure 8

4.2.1.4 *Results*

The study of density model alone is developed in [**raimbault2016calibration**]. It is in particular calibrated on European density grid data, on 50km width square areas with 500m resolution for which real indicator values have been computed on whole Europe. Furthermore, a grid exploration of model behavior yields feasible output space in reasonable parameters bounds (roughly $\alpha \in [0.5, 2], N_G \in [500, 3000], P_m \in [10^4, 10^5], \beta \in [0, 0.2], n_d \in \{1, \ldots, 4\}$). The reduction of indicators space to a two dimensional plan through a Principal Component Analysis (variance explained with two components $\simeq 80\%$) allows to isolate a set of output points that covers reasonably precisely real point cloud. It confirms the ability of the model to reproduce morphologically the set of real configurations.

At given density, the conditional exploration of network generation model parameter space suggest a good flexibility on global indicators $\vec{G}$, together with good convergence properties. For a precise study of model behavior, see appendice giving regressions analysis capturing the behavior of coupled model. In order to illustrate synthetic data generation method, the exploration has been oriented towards the study of cross-correlations.

Given the large relative dimension of parameter space, an exhaustive grid exploration is not possible. We use a Latin Hypercube sampling procedure with bounds given above for $\vec{\alpha}_D$ and for $\vec{\alpha}_N$, we take $N_C \in [50, 120], r_g \in [1, 100], d_0 \in [0.1, 10], k_h \in [0, 1], \gamma \in [0.1, 4], N_L \in [4, 20]$. For number of model replications for each parameter point, less than 50 are enough to obtain confidence intervals at 95% on indicators of width less than standard deviations. For correlations a hundred give confidence intervals (obtained with Fisher method) of size around 0.4, we take thus $n = 80$ for experiments. Figure 12 gives details of experiment results. Regarding the subject of correlated synthetic data generation, we can sum up the main lines as following :

- Empirical distributions of correlation coefficients between morphology and network indicators are not simple and some are bimodal (for example $\rho_{46} = \rho[r, \bar{l}]$ between Moran index and mean path length).

- it is possible to modulate up to a relatively high level of correlation for all indicators, maximal absolute correlation varying between 0.6 and 0.9. Amplitude of correlations varies between 0.9 and 1.6, allowing a broad spectrum of values. Point cloud in principal plan has a large extent but is not uniform : it is not possible to modulate at will any coefficient as they stay themselves correlated because of underlying generation processes. A more refined study at higher orders (correlation of correlations) would be necessary to precisely understand degrees of freedom in correlation generation.

- Most correlated points are also the closest to real data, what confirms the intuition and stylized fact of a strong interdependence in reality.
- Concrete examples taken on particular points in the principal plan show that similar density profiles can yield very different correlation profiles.

4.2.1.5 *Possible developments*

This case study could be refined by extending correlation control method. A precise knowledge of N behavior (statistical distributions on an exhaustive grid of parameter space) conditional to D would allow to determine $N^{<-1>}|D$ and have more latitude in correlation generation. We could also apply specific exploration algorithms to reach exceptional configurations realizing an expected correlation level, or at least to obtain a better knowledge of the feasible space of correlations [**10.1371/journal.pone.0138212**].

4.2.2 *Discussion*

Scientific positioning

Our overall approach enters a particular epistemological frame. On the one hand the multidisciplinary aspect, and on the other hand the importance of empirical component through computational exploration methods, make this approach typical of Complex Systems science, as it is recalled by the roadmap for Complex Systems having a similar structure [**2009arXiv0907.2221B**]. It combines transversal research questions (horizontal integration of disciplines) with the development of heterogeneous multi-scalar approaches which encounter similar issues as the one we proposed to tackle (vertically integrated disciplines). The combination of empirical knowledge obtained from data mining, with knowledge obtained by modeling and simulation is generally central to the conception and exploration of multi-scalar heterogeneous models. Results presented here is an illustration of such an hybrid paradigm.

Direct applications

Starting from the second example which was limited to data generation, we propose examples of direct applications that should give an overview of the range of possibilities.

- Calibration of network generation component at given density, on real data for transportation network (typically road network given the shape of generated networks ; it should be straightforward to use OpenStreetMap open data[3] that have a reasonable quality for Europe, at least for France [**girres2010quality**], with however adjustments on generation procedure in order to avoid edge effects due its restrictive frame, for example by generating on an extended surface to keep only a central area on which calibration would be done) should theoretically allow to unveil parameter sets reproducing accurately existing configurations both for urban morphology and network shape. It could be then possible to derive a "theoretical correlation" for these, as an empirical correlation is according to some theories of urban systems not computable as a unique realization of stochastic processes is observed. Because of non-ergodicity of urban systems [20], there are strong chances that involved processes are different across different geographical areas (or from an other point of view that they are in an other state of meta-parameters, i.e. in an other regime) and that their interpretation as different realizations of the same stochastic process makes no sense, the impossibility of covariation estimation following. By attributing a synthetic dataset similar to a given real configuration, we would be able to compute a sort of *intrinsic correlation* proper to this configuration. As territorial configurations emerge from spatio-temporal interdependences between components of territorial systems, this intrinsic correlation emerges the same way, and its knowledge gives information on these interdependences and thus on relations between territories and networks.

- As already mentioned, most of models of simulation need an initial state generated artificially as soon as model parametrization is not done completely on real data. An advanced model sensitivity analysis implies a control on parameters for synthetic dataset generation, seen as model meta-parameters [**cottineau2015revisiting**]. In the case of a statistical analysis of model outputs it provides a way to operate a second order statistical control.

- We studied in the first example stochastic processes in the sense of random time-series, whereas time did not have a role in the second case. We can suggest a strong coupling between the two model components (or the construction of an integrated model) and to observe indicators and correlations at different time steps during the generation. In a dynamical spatial models we have because of feedbacks necessarily propagation effects and therefore the existence of lagged interdependences in space and time [**pigozzi1980interurban**]. It would drive our

3 https://www.openstreetmap.org

field of study towards a better understanding of dynamical correlations.

Generalization

We were limited to the control of first and second moments of generated data, but we could imagine a theoretical generalization allowing the control of moments at any order. However, as shown by the geographical example, the difficulty of generation in a concrete complex case questions the possibility of higher orders control when keeping a consistent structure model and a reasonable number of parameters. The study of non-linear dependence structures as proposed in [**chicheporticheze2013nested**] is in an other perspective an interesting possible development.

4.2.3 *Conclusion*

We proposed an abstract method to generate synthetic datasets in which correlation structure is controlled. Its rapid implementation in two very different fields shows its flexibility and the broad range of possible applications. More generally, it is crucial to favorise such practices of systematic validation of computational models by statistical analysis, in particular for agent-based models for which the question of validation stays an open issue.

Part III

TOWARDS OPERATIONAL MODELS

Part IV

CONCLUSION

BIBLIOGRAPHY

[1] Merwan Achibet, Stefan Balev, Antoine Dutot, and Damien Olivier. “A Model of Road Network and Buildings Extension Co-evolution.” In: *Procedia Computer Science* 32 (2014), pp. 828–833.

[2] Alex Anas, Richard Arnott, and Kenneth A. Small. “Urban Spatial Structure.” English. In: *Journal of Economic Literature* 36.3 (1998), pp. 1426–1464. ISSN: 00220515. URL: http://www.jstor.org/stable/2564805.

[3] E. Arcaute, E. Hatna, P. Ferguson, H. Youn, A. Johansson, and M. Batty. “Constructing cities, deconstructing scaling laws.” In: *ArXiv e-prints* (Jan. 2013). arXiv: 1301.1674 [physics.soc-ph].

[4] Marc Barthélemy and Alessandro Flammini. “Co-evolution of density and topology in a simple model of city formation.” In: *Networks and spatial economics* 9.3 (2009), pp. 401–425.

[5] Luís MA Bettencourt, José Lobo, and Geoffrey B West. “Why are large cities faster? Universal scaling and self-similarity in urban organization and dynamics.” In: *The European Physical Journal B-Condensed Matter and Complex Systems* 63.3 (2008), pp. 285–293.

[6] Anne Bretagnolle, Denise Pumain, and Céline Vacchiani-Marcuzzo. “The organization of urban systems.” In: *Complexity perspectives in innovation and social change*. Springer, 2009, pp. 197–220.

[7] David Chavalarias and Jean-Philippe Cointet. “Phylomemetic patterns in science evolution—the rise and fall of scientific fields.” In: *Plos One* 8.2 (2013), e54847.

[8] Yanguang Chen. “Characterizing growth and form of fractal cities with allometric scaling exponents.” In: *Discrete Dynamics in Nature and Society* 2010 (2010).

[9] Clémentine Cottineau. *Urban scaling: What cities are we talking about?* Presentation of ongoing work at Quanturb seminar, April 1st 2015. 2015.

[10] Clémentine Cottineau, Paul Chapron, and Romain Reuillon. “An incremental method for building and evaluating agent-based models of systems of cities.” In: (2015).

[11] Jean-Marc Favaro and Denise Pumain. “Gibrat Revisited: An Urban Growth Model Incorporating Spatial Interaction and Innovation Cycles.” In: *Geographical Analysis* 43.3 (2011), pp. 261–286.

[12] Xavier Gabaix. "Zipf's law for cities: an explanation." In: *Quarterly journal of Economics* (1999), pp. 739–767.

[13] Xavier Gabaix and Yannis M. Ioannides. "Chapter 53 The evolution of city size distributions." In: *Cities and Geography*. Ed. by J. Vernon Henderson and Jacques-François Thisse. Vol. 4. Handbook of Regional and Urban Economics. Elsevier, 2004, pp. 2341 –2378. DOI: http://dx.doi.org/10.1016/S1574-0080(04)80010-5. URL: http://www.sciencedirect.com/science/article/pii/S1574008004800105.

[14] Alexandros Karatzoglou, Alex Smola, Kurt Hornik, and Achim Zeileis. "kernlab – An S4 Package for Kernel Methods in R." In: *Journal of Statistical Software* 11.9 (2004), pp. 1–20. URL: http://www.jstatsoft.org/v11/i09/.

[15] Marie-Laurence Keersmaecker, Pierre Frankhauser, and Isabelle Thomas. "Using fractal dimensions for characterizing intra-urban diversity: The example of Brussels." In: *Geographical analysis* 35.4 (2003), pp. 310–328.

[16] Michael Lissack. "Subliminal influence or plagiarism by negligence ? The Slodderwetenschap of ignoring the internet." In: *Journal of Academic Ethics* (2013).

[17] R. Louf and M. Barthelemy. "How congestion shapes cities: from mobility patterns to scaling." In: *ArXiv e-prints* (Jan. 2014). arXiv: 1401.8200 [physics.soc-ph].

[18] Rémi Louf and Marc Barthelemy. "Scaling: lost in the smog." In: *arXiv preprint arXiv:1410.4964* (2014).

[19] Jean-Marc Offner. "Les "effets structurants" du transport: mythe politique, mystification scientifique." In: *Espace géographique* 22.3 (1993), pp. 233–242.

[20] Denise Pumain. "Urban systems dynamics, urban growth and scaling laws: The question of ergodicity." In: *Complexity Theories of Cities Have Come of Age*. Springer, 2012, pp. 91–103.

[21] Denise Pumain, Fabien Paulus, Céline Vacchiani-Marcuzzo, and José Lobo. "An evolutionary theory for interpreting urban scaling laws." In: *Cybergeo: European Journal of Geography* (2006).

[22] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria, 2015. URL: http://www.R-project.org/.

[23] J. Raimbault, A. Banos, and R. Doursat. "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.

[24] Kazuko Yamasaki, Kaushik Matia, Sergey V Buldyrev, Dongfeng Fu, Fabio Pammolli, Massimo Riccaboni, and H Eugene Stanley. "Preferential attachment and growth dynamics in complex systems." In: *Physical Review E* 74.3 (2006), p. 035103.

Part V

APPENDIX