Generation of Correlated Synthetic Data

Actes des Journées de Rochebrune 2016

Juste Raimbault[1,2]
[1] UMR CNRS 8504 Géographie-cités
[2] UMR-T IFSTTAR 9403 LVMT

Abstract

Generation of hybrid synthetic data resembling real data to some criteria is an important methodological and thematic issue in most disciplines which study complex systems. Interdependencies between constituting elements, materialized within respective relations, lead to the emergence of macroscopic patterns. Being able to control the dependance structure and level within a synthetic dataset is thus a source of knowledge on system mechanisms. We propose a methodology consisting in the generation of synthetic datasets on which correlation structure is controlled. The method is applied in a first example on financial time-series and allows to understand the role of interferences between components at different scales on performances of a predictive model. A second application on a geographical system is then proposed, in which the weak coupling between a population density model and a network morphogenesis model allows to simulate territorial configurations. The calibration on morphological objective on european data and intensive model exploration unveils a large spectrum of feasible correlations between morphological and network measures. We demonstrate therein the flexibility of our method and the variety of possible applications.

Keywords : *Synthetic Data ; Statistical Control ; Correlations ; Financial Time-series ; Land-use Transportation Interactions*

1 Introduction

The use of synthetic data, in the sense of statistical populations generated randomly under constraints of patterns proximity to the studied system, is a widely used methodology, and more particularly in disciplines related to complex systems such as therapeutic evaluation [Abadie et al., 2010], territorial science [Moeckel et al., 2003, Pritchard and Miller, 2009], machine learning [Bolón-Canedo et al., 2013] or bio-informatics [Van den Bulcke et al., 2006]. It can consist in data desegregation by creation of a microscopic population with fixed macroscopic properties, or in the creation of new populations at the same scale than a given sample, with criteria of proximity to the real sample. These criteria will depend on expected applications and can for example vary from a restrictive statistical fit on given indicators, to weaker assumptions of similarity in aggregated patterns. In the case of chaotic systems, or systems where emergence plays a strong role, a microscopic property does not directly imply given macroscopic patterns, which reproduction is indeed one aim of modeling and simulation practices in complexity science. With the rise of new computational paradigms [Arthur, 2015], data (simulated, measured or hybrid) shape our understanding of complex systems. Methodological tools for data-mining and modeling and simulation (including the generation of synthetic data) are therefore crucial to be developed.

Whereas first order (in the sense of distribution moments) is generally well used, it is not systematic nor simple to control generated data structure at second order, i.e. covariance structure between generated variables. Some specific examples can be found, such as in [Ye, 2011] where the sensitivity of discrete choices models to the distributions of inputs and to their dependance structure is examined. It is also possible to interpret complex networks generative models [Newman, 2003] as the production of an interdependence structure for a system, contained within link topology. We introduce here a generic method taking into account dependance structure for the generation of synthetic datasets, more precisely with the mean of controlled correlation matrices.

The rest of the paper is organized as follows. The generic method is formally described, to be then applied on very different examples both entering application frame. Each example can be read independently and illustrates potentialities of the method and possible technical limitations. We discuss then possible further developments and applications, in particular for a geographical system.

2 Method Formalization

Domain-specific methods aforementioned are too broad to be summarized into a same formalism. We propose a framework as generic as possible, centered on the control of correlations structure in synthetic data.

Let $\vec{X}_I$ a multidimensional stochastic process (that can be indexed e.g. with time in the case of time-series, but also space, or discrete set abstract indexation). We assume given a real dataset $\mathbf{X} = (X_{i,j})$, interpreted as a set of realizations of the stochastic process. We propose to generate a statistical population $\tilde{\mathbf{X}} = \tilde{X}_{i,j}$ such that

1. a given criteria of proximity to data is verified, i.e. given a precision ε and an indicator f, we have $\|f(\mathbf{X}) - f(\tilde{\mathbf{X}})\| < \varepsilon$
2. level of correlation is controlled, i.e. given a matrix R fixing correlation structure (symmetric matrix with coefficients in $[-1, 1]$ and unity diagonal), we have $\hat{\text{Var}}\left[(\tilde{X}_i)\right] = \Sigma R \Sigma$, where the standard deviation diagonal matrix Σ is estimated on the synthetic population.

The second requirement will generally be conditional to parameter values determining generation procedure, either generation models being simple or complex (R itself is a parameter). Formally, synthetic processes are parametric families $\tilde{X}_i[\vec{\alpha}]$. We propose to apply the methodology on very different examples, both typical of complex systems : financial high-frequency time-series and territorial systems. We illustrate the flexibility of the method, and claim to help building interdisciplinary bridges by methodology transposition and reasoning analogy. In the first case, proximity to data is the equality of signals at a fundamental frequency, to which higher frequency synthetic components with controlled correlations are superposed. It follows a logic of hybrid data for which hypothesis or model testing is done on a more realistic context than on purely synthetic data. In the second case, morphological calibration of a population density distribution model allows to respect real data proximity. Correlations of urban form with transportation network measures are empirically obtained by exploration of coupling with a network morphogenesis model. The control is in this case indirect as feasible space is empirically determined.

3 Applications

3.1 Application : financial time-series

3.1.1 Context

Our first field of application is that of financial complex systems, of which captured signals, financial time-series, are heterogeneous, multi-scalar and highly non-stationary [Mantegna et al., 2000]. Correlations have already been the object of a broad bunch of related literature. For example, Random Matrix Theory allows to undress signal of noise, or at least to estimate the proportion of information undistinguishable from noise, for a correlation matrix computed for a large number of asset with low-frequency signals (daily returns mostly) [Bouchaud and Potters, 2009]. Similarly, Complex Network Analysis on networks constructed from correlations, by methods such as Minimal Spanning Tree [Bonanno et al., 2001] or more refined extensions developed for this purpose [Tumminello et al., 2005], yielded promising results such as the reconstruction of economic sectors structure. At high frequency, the precise estimation of of interdependence parameters in the framed of fixed assumptions on asset dynamics, has been extensively studied from a theoretical point of view aimed at refinement of models and estimators [Barndorff-Nielsen et al., 2011]. Theoretical results must be tested on synthetic datasets as they ensure a control of most parameters in order to check that a predicted effect is indeed observable *all things equal otherwise*. For example, [Potiron and Mykland, 2015] obtains a bias correction for the *Hayashi-Yoshida* estimator (used to estimate integrated covariation between two brownian at high frequency in the case of asynchronous observation times) by deriving a central limit theorem for a general model that endogeneize observation times. Empirical confirmation of estimator improvement is obtained on a synthetic dataset at a fixed correlation level.

3.1.2 Formalization

Framework We consider a network of assets $(X_i(t))_{1 \leq i \leq N}$ sampled at high-frequency (typically 1s). We use a multi-scalar framework (used e.g. in wavelet analysis approaches [Ramsey, 2002] or in multi-fractal signal processing [Bouchaud et al., 2000]) to interpret observed signals as the superposition of components at different time scales : $X_i = \sum_\omega X_i^\omega$. We denote by $T_i^\omega = \sum_{\omega' \leq \omega} X_i^\omega$ the filtered signal at a given frequency ω. A recurrent problem in the study of complex systems is the prediction of a trend at a given scale. It can be viewed as the identification of regularities and their distinction from components considered as random[1]. For the sake of simplicity, we represent such a process as a trend prediction model at a given temporal scale ω_1, formally an estimator $M_{\omega_1} : (T_i^{\omega_1}(t'))_{t' < t} \mapsto \hat{T}_i^{\omega_1}(t)$ which aims to minimize error on the real trend $\|T_i^{\omega_1} - \hat{T}_i^{\omega_1}\|$. In the case of autoregressive multivariate estimators, the performance will depend among other parameters on respective correlations between assets. It is thus interesting to apply the method to the evaluation of performance as a function of correlation at different scales. We assume a Black-Scholes dynamic for assets [Jarrow, 1999], i.e. $dX = \sigma \cdot dW$, with W Wiener process. Such a dynamic model allows an easy modulation of correlation levels.

[1] see [Gell-Mann, 1995] for an extended discussion on the construction of *schema* to study complex adaptive systems (by complex adaptive systems).

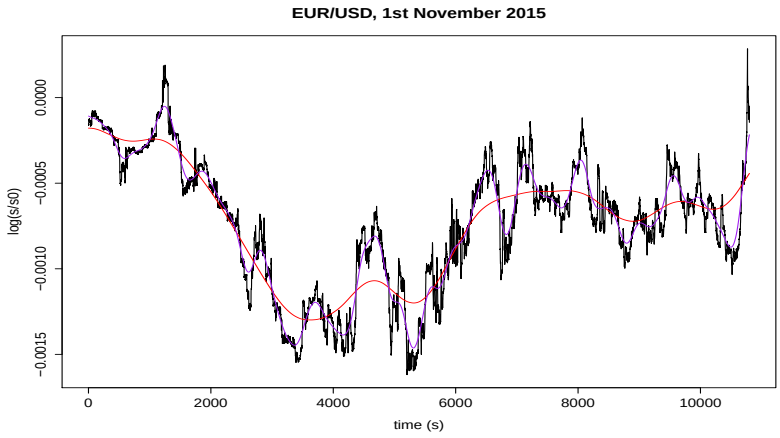


Figure 1: **Example of the multi-scalar structure of the signal, basis of the construction of synthetic signals |** *Log-prices* are represented on a time window of around 3h for November 1st 2015 for asset EUR/USD, with 10min (purple) and 30min trends.

Data generation We can straightforward generate $\tilde{X}_i$ such that $\mathrm{Var}\left[\tilde{X}_i^{\omega_1}\right] = \Sigma R \Sigma$ (with Σ estimated standard deviations and R fixed correlation matrix) and verifying $X_i^{\omega \leq \omega_0} = \tilde{X}_i^{\omega \leq \omega_0}$ (data proximity indicator : components at a lower frequency than a fundamental frequency $\omega_0 < \omega_1$ are identical). We use therefore the simulation of Wiener processes with fixed correlation. Indeed, if $dW_1 \perp\!\!\!\perp dW_1^{\perp\!\!\!\perp}$ (and $\sigma_1 < \sigma_2$ indicatively, assets being interchangeable), then

$$W_2 = \rho_{12} W_1 + \sqrt{1 - \frac{\sigma_1^2}{\sigma_2^2} \cdot \rho_{12}^2} \cdot W_1^{\perp\!\!\!\perp}$$

is such that $\rho(dW_1, dW_2) = \rho_{12}$. Next signals are constructed the same way by Gram orthonormalization. We isolate the component at the desired frequency ω_1 by filtering the signal, i.e. $\tilde{X}_i^{\omega_1} = W_i - \mathcal{F}_{\omega_0}[W_i]$ (with $\mathcal{F}_{\omega_0}$ low-pass filter with cut-off frequency ω_0). We reconstruct then the hybrid synthetic signals by

$$\tilde{X}_i = T_i^{\omega_0} + \tilde{X}_i^{\omega_1} \tag{1}$$

3.1.3 Implementation and Results

Methodology The method is tested on an example with two assets from foreign exchange market (EUR/USD and EUR/GBP), in a six month period from June 2015 to November 2015. Data[2] cleaning, starting from original series sampled at a frequency around 1s, consists in a first step to the determination of the minimal common temporal range (missing sequences being ignored, by vertical translation of series, i.e. $S(t) := S(t) \cdot \frac{S(t_n)}{S(t_{n-1})}$ when t_{n-1}, t_n are extremities of the "hole" and $S(t)$ value of the asset, what is equivalent to keep the constraint to have returns at similar temporal steps between assets). We study then *log-prices* and *log-returns*, defined by $X(t) := \log \frac{S(t)}{S_0}$ and $\Delta X(t) = X(t) - X(t-1)$. Raw data are filtered at a maximal frequency $\omega_m = 10$min (which will be the maximal frequency for following treatments) for concerns of computational efficiency[3]. We use a non-causal gaussian filter of total width ω. We fix the fundamental frequency $\omega_0 = 24$h and we propose to construct synthetic data at frequencies $\omega_1 = 30\text{min}, 1\text{h}, 2\text{h}$. See Fig. 1 for an example of signal structure at these different scales.

It is crucial to consider the interference between ω_0 and ω_1 frequencies in the reconstructed signal : correlation indeed estimated is

$$\rho_e = \rho\left[\Delta\tilde{X}_1, \Delta\tilde{X}_2\right] = \rho\left[\Delta T_1^{\omega_0} + \Delta\tilde{X}_1^{\omega}, \Delta T_2^{\omega_0} + \Delta\tilde{X}_2^{\omega}\right]$$

what yields in the reasonable limit $\sigma_1 \gg \sigma_0$ (fundamental frequency small enough), when $\mathrm{Cov}\left[\Delta\tilde{X}_i^{\omega_1}, \Delta X_j^{\omega}\right] = 0$ for all $i, j, \omega_1 > \omega$ and returns centered at any scale, the correction on effective correlation due to interferences : we have at first order the expression of effective correlation

[2] obtained from `http://www.histdata.com/`, without specified licence. For the respect of copyright, only cleaned and filtered at ω_m data are made openly available.

[3] as time-series are then sampled at $3 \cdot \omega_m$ to avoid aliasing, a day of size 86400 for 1s sampling is reduced to a much smaller size of 432.

$$\rho_e = [\varepsilon_1\varepsilon_2\rho_0 + \rho] \cdot \left[1 - \frac{1}{2}\left(\varepsilon_1^2 + \varepsilon_2^2\right)\right] \tag{2}$$

what gives the correlation that we can effectively simulate in synthetic data.

Correlation is estimated by Pearson method, with estimator for covariance corrected for bias, i.e.

$$\hat{\rho}[X1, X2] = \frac{\hat{C}[X1, X2]}{\sqrt{\hat{\mathrm{Var}}\,[X1]\hat{\mathrm{Var}}\,[X2]}}$$

, where $\hat{C}[X1, X2] = \frac{1}{(T-1)}\sum_t X_1(t)X_2(t) - \frac{1}{T\cdot(T-1)}\sum_t X_1(t)\sum_t X_2(t)$ and $\hat{\mathrm{Var}}\,[X] = \frac{1}{T}\sum_t X^2(t) - \left(\frac{1}{T}\sum_t X(t)\right)^2$.

The tested predictive model M_{ω_1} is a simple *ARMA* for which parameters $p = 2, q = 0$ are fixed (as we do not create lagged correlation, we do not expect large orders of auto-regression as these kind of processes have short memory for real data ; furthermore smoothing is not necessary as data are already filtered). It is however applied in an adaptive way[4]. More precisely, given a time window T_W, we estimate for any t the model on $[t - T_W + 1, t]$ in order to predict signals at $t + 1$.

Implementation Experiments are implemented in `R` language, using in particular the `MTS` [Tsay, 2015] library for time-series models. Cleaned data and source code are openly available on the `git` repository of the project[5].

Results Figure 2 gives effective correlations computed on synthetic data. For standard parameter values (for example $\omega_0 = 24\text{h}$, $\omega_1 = 2\text{h}$ and $\rho = -0.5$), we find $\rho_0 \simeq 0.71$ et $\varepsilon_i \simeq 0.3$ what yields $|\rho_e - \rho| \simeq 0.05$. We observe a good agreement between observed ρ_e and values predicted by 2 in the interval $\rho \in [-0.5, 0.5]$. On the contrary, for larger absolute values, a deviation increasing with $|\rho|$ and as ω_1 decreases : it confirms the intuition that when frequency decreases and becomes closer to ω_0, interferences between the two components are not negligible anymore and invalidate independence assumptions for example.

We apply then the predictive model described above to synthetic data, in order to study its mean performance as a function of correlation between signals. Results for $\omega_1 = 1\text{h}, 1\text{h}30, 2\text{h}$ are shown in Fig. 3. The a priori counter-intuitive result of a maximal performance at vanishing correlation for one of the assets confirms the role of synthetic data to better understand system mechanisms : the study of lagged correlations shows an asymmetry in the real data that we can understand at a daily scale as an increased influence of EUR/GBP on EUR/USD with a rough two hours lag. The existence of this *lag* allows a "good" prediction of EUR/USD thanks to fundamental component. This predictive power is perturbed by added noises in a way that increases with their correlation. The more noises correlated are, the more he model will take them into account and will make false predictions because of the markovian character of simulated brownian[6].

Our case study stays a *toy-model* and has no direct practical application, but demonstrates however the relevance of using simulated synthetic data. Further developments can be directed towards the simulation of more realistic data (presence of consistent *lagged correlation* patterns, more realistic models than Black-Scholes) and apply it on more operational predictive models.

[4] adaptation level staying low, as parameters T_W, p, q and model type do not vary. We are positioned within the framework of [Potiron, 2016] which assumes a locally parametric dynamic but for which meta-parameters are fixed. We could imagine a variable T_W which would adapt for the best local fit, the same way parameters are estimated in bayesian signal processing by augmentation of the state with parameters.

[5] at `https://github.com/JusteRaimbault/SynthAsset`

[6] the model used has theoretically no predictive power at all on pure brownian

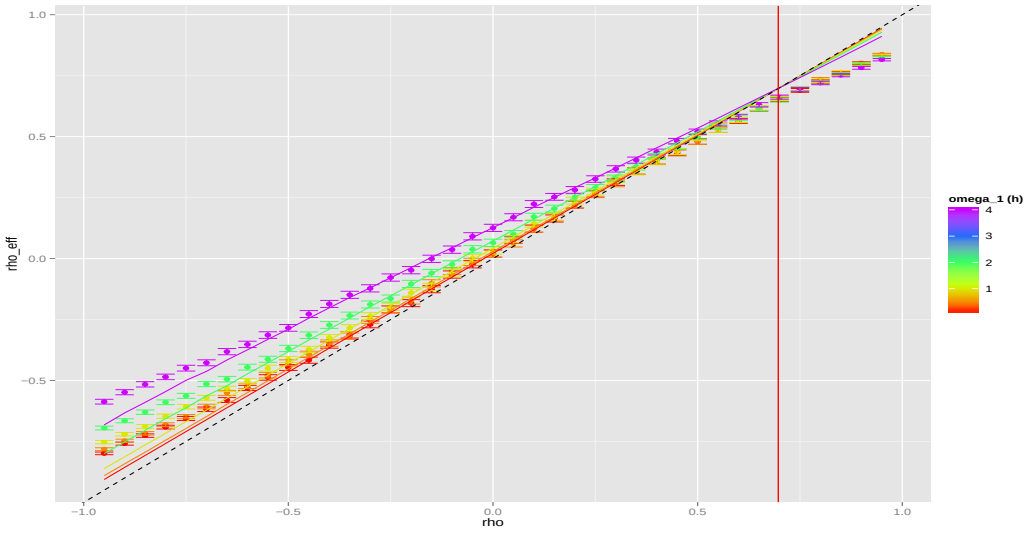


Figure 2: **Effective correlations obtained on synthetic data |** Dots represent estimated correlations on a synthetic dataset corresponding to 6 months between June and November 2015 (error-bars give 95% confidence intervals obtained with standard Fisher method) ; scale color gives filtering frequency $\omega_1 = 10\text{min}, 30\text{min}, 1\text{h}, 2\text{h}, 4\text{h}$; solid lines give theoretical values for ρ_e obtained by 2 with estimated volatilities (dotted-line diagonal for reference) ; vertical red line position is the theoretical value such that $\rho = \rho_e$ with mean values for ε_i on all points. We observe for high absolute correlations values a deviation from corrected values, what should be caused by non-verified independence and centered returns assumptions. Asymmetry is caused by the high value of $\rho_0 \simeq 0.71$.

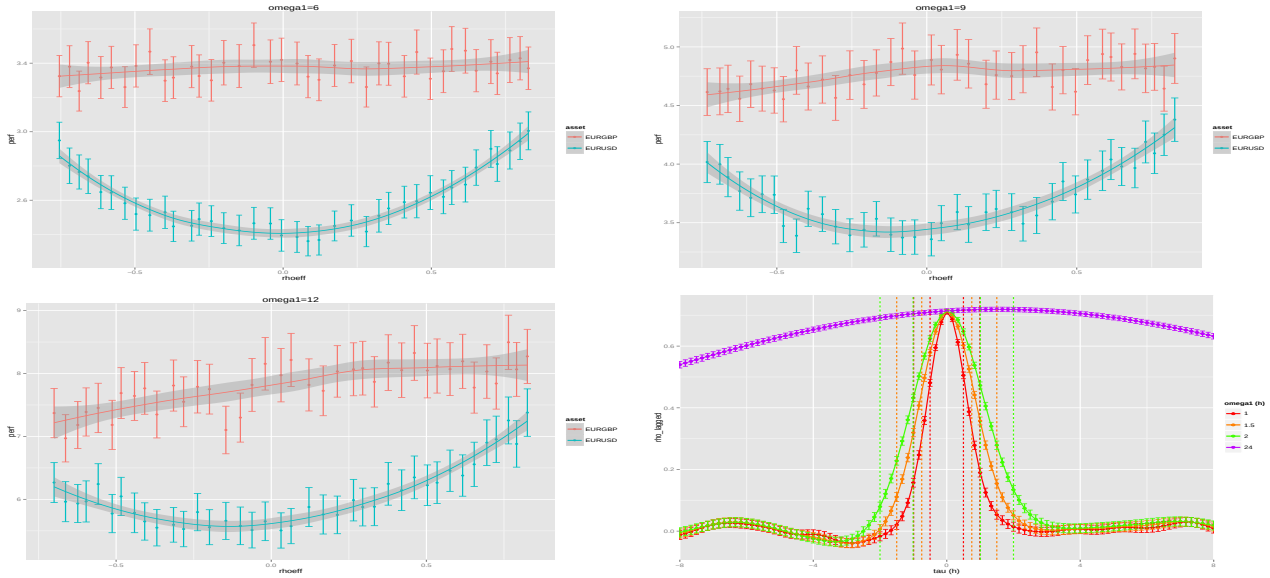


Figure 3: **Performance of a predictive model as a function of simulated correlations |** From left to right and top to bottom, three first graphs show for each asset the normalized performance of an ARMA model ($p = 2, q = 0$), defined as $\pi = \left(\frac{1}{T} \sum_t \left(\tilde{X}_i(t) - M_{\omega_1} \left[\tilde{X}_i \right](t) \right)^2 \right) / \sigma \left[\tilde{X}_i \right]^2$ (95% confidence intervals computed by $\pi = \bar{\pi} \pm (1.96 \cdot \sigma[\pi])/\sqrt{T}$, local polynomial smoothing to ease reading). It is interesting to note the U-shape for EUR/USD, due to interference between components at different scales. Correlation between simulated noises deteriorates predictive power. The study of *lagged correlations* (here $\rho[\Delta X_{\text{EURUSD}}(t), \Delta X_{\text{EURGBP}}(t-\tau)]$) on real data clarifies this phenomenon : fourth graph show an asymmetry in curves at any scale compared to zero lag ($\tau = 0$) what leads fundamental components to increase predictive power for the dollar, amelioration then perturbed by correlations between simulated components. Dashed lines show time steps (in equivalent τ units) used by the ARMA at each scale, what allows to read the corresponding lagged correlation on fundamental component.

3.2 Application : geographical data of density and network

3.2.1 Context

The use of synthetic data in geography is generally directed towards the generation of synthetic populations within agent-based models (mobility, *LUTI* models) [Pritchard and Miller, 2009]. 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 [Brunsdon et al., 1998] 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 [Sanders et al., 1997], 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 [Schmitt, 2014] 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 [Cottineau et al., 2015b], 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.

3.2.2 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 [Offner and Pumain, 1996] but stays highly complex and difficult to quantify [Offner, 1993]. A dynamical modeling of implied processes should shed light on these interactions ([Bretagnolle, 2009], 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 [Batty, 2006] to generate a discrete spatial distribution of population density. A generalization of the basic model is proposed in [Raimbault, 2016], 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, 2014]. More precisely, the model proceeds iteratively the following way. An square grid of width N, initially empty, is represented by population $(P_i(t))_{1\leq i\leq N^2}$. 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 $\mathbb{P}[P_i(t+1) = P_i(t)+1 | P(t+1) = P(t)+1] = \frac{(P_i(t)/P(t))^\alpha}{\sum (P_i(t)/P(t))^\alpha}$
- 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

Le réseau est alors modulé par ruptures de potentiels afin de se rapprocher de formes réelles. Plus précisément, un potentiel d'interaction gravitaire généralisé entre deux centres i et j est défini par

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

où d peut être la distance euclidienne $d_{ij} = d(i,j)$ ou la distance par le réseau $d_N(i,j)$, $k_h \in [0,1]$ un poids permettant de changer le rôle des population dans le potentiel, γ régissant la forme de la hiérarchie selon les valeurs des populations, r_g distance caractéristique de décroissance et d_0 paramètre de forme.

Un nombre $K \cdot N_L$ de nouveaux liens potentiels est pris comme les couples ayant le plus grand potentiel pour la distance euclidienne ($K = 5$ est fixé).

Parmi les liens potentiels, N_L sont effectivement réalisés, qui sont ceux ayant le plus faible rapport $V_{ij}(d_N)/V_{ij}(d_{ij})$: à cette étape seul l'écart entre distance euclidienne et distance par le réseau compte, ce rapport ne dépendant plus des populations et étant croissant en d_N à d_{ij} fixé.

Le réseau est planarisé par création de noeuds aux intersections éventuelles créées par les nouveaux liens.

Notons que la construction du modèle de génération est heuristique, et que d'autres types de modèles comme un réseau biologique auto-généré [Tero et al., 2010], une génération par optimisation locale de contraintes géométriques [Barthélemy and Flammini, 2008] ou un modèle de percolation plus complexe que celui utilisé, peuvent le remplacer. Ainsi, dans le cadre d'une architecture modulaire où le choix entre différentes implémentations d'une brique fonctionnelle peut être vue comme méta-paramètre [Cottineau et al., 2015a], on pourrait choisir la fonction de génération adaptée à un besoin donné (par exemple proximité à des données réelles, contraintes sur les relations entre indicateurs de sortie, variété de formes générées, etc.).

Espace des paramètres L'espace des paramètres du modèle couplé[7] est constitué des paramètres de génération de densité $\vec{\alpha}_D = (P_m/N_G, \alpha, \beta, n_d)$ (on s'intéresse pour simplifier au rapport entre population et taux de croissance, i.e. le nombre d'étapes nécessaires pour générer) et des paramètres de génération de réseau $\vec{\alpha}_N = (N_C, k_h, \gamma, r_g, d_0)$. On notera $\vec{\alpha} = (\vec{\alpha}_D, \vec{\alpha}_N)$.

Indicateurs On quantifie la forme urbaine et la forme du réseau, dans le but de moduler la corrélation entre ces indicateurs. La forme est définie par un vecteur $\vec{M} = (r, \bar{d}, \varepsilon, a)$ donnant auto-corrélation spatiale (indice de Moran), distance moyenne, entropie, hiérarchie (voir [Le Néchet, 2015] pour une définition précise de ces indicateurs). Les mesures de la forme du réseau $\vec{G} = (\bar{c}, \bar{l}, \bar{s}, \delta)$ sont, avec le réseau noté (V, E),

- Centralité moyenne $\bar{c}$, définie comme la moyenne de la *betweeness-centrality* (normalisée dans $[0,1]$) sur l'ensemble des liens.
- Longueur moyenne des chemins $\bar{l}$ définie par $\frac{1}{d_m}\frac{2}{|V|\cdot(|V|-1)}\sum_{i<j} d_N(i,j)$ avec d_m distance de normalisation prise ici comme la diagonale du monde $d_m = \sqrt{2}N$.
- Vitesse moyenne [Banos and Genre-Grandpierre, 2012], qui correspond à la performance du réseau par rapport au trajet à vol d'oiseau, définie par $\bar{s} = \frac{2}{|V|\cdot(|V|-1)}\sum_{i<j}\frac{d_{ij}}{d_N(i,j)}$.
- Diamètre du réseau $\delta = \max_{ij} d_N(i,j)$

Covariance et correlation On s'intéressera à la matrice de covariance croisée $\mathrm{Cov}\left[\vec{M}, \vec{G}\right]$ entre densité et réseau, estimée sur un jeu de n réalisations à paramètres fixés $(\vec{M}\left[D(\vec{\alpha})\right], \vec{G}\left[N(\vec{\alpha})\right])_{1\leq i\leq n}$ par l'estimateur standard non-biaisé. On prend comme correlation associée la correlation de Pearson estimée de la même façon.

[7] Le couplage faible permet de limiter le nombre total de paramètres puisqu'un couplage fort incluant des boucles de retroaction comprendrait nécessairement des paramètres supplémentaires pour régler la forme et l'intensité de celles-ci. Pour espérer le diminuer, il faudrait concevoir un modèle intégré, ce qui est différent d'un couplage fort dans le sens où il n'est pas possible de figer l'un des sous-systèmes pour obtenir un modèle de l'autre correspondant au modèle non-couplé.

3.2.3 Implémentation

Le couplage des modèles génératifs est effectué à la fois au niveau formel et au niveau opérationnel, c'est à dire qu'on fait interagir des implémentations indépendantes. Pour cela, le logiciel OpenMole [Reuillon et al., 2013] utilisé pour l'exploration intensive, offre le cadre idéal de par son langage modulaire permettant de construire des *workflows* par composition de tâches à loisir et de les brancher sur divers plans d'expérience et sorties. Pour des raisons opérationnelles, le modèle de densité est implémenté en langage `scala` comme un `plugin` d'OpenMole, tandis que la génération de réseau est implémentée en langage basé-agent `NetLogo` [Wilensky, 1999], ce qui facilite l'exploration interactive et construction heuristique interactive. Le code source est disponible pour reproductibilité sur le dépôt du projet[8].

3.2.4 Résultats

L'étude du modèle de densité seul est développée dans [Raimbault, 2016]. Il est notamment calibré sur les données de la grille européenne de densité, sur des zones de 50km de côté et de résolution 500m pour lesquelles les valeurs réelles des indicateurs ont été calculées pour l'ensemble de l'Europe. D'autre part, une exploration brutale du modèle permet d'estimer l'ensemble des sorties possibles dans des bornes raisonnables pour les paramètres (grossièrement $\alpha \in [0.5, 2], N_G \in [500, 3000], P_m \in [10^4, 10^5], \beta \in [0, 0.2], n_d \in \{1, \dots, 4\}$). La réduction à un plan de l'espace des objectif par une Analyse en Composantes Principales (variance expliquée à deux composantes $\simeq 80\%$) permet d'isoler un nuage de points de sorties recouvrant assez fidèlement le nuage des points réels, ce qui veut dire que le modèle est capable de reproduire morphologiquement l'ensemble des configurations existantes.

A densité donnée, l'exploration de l'espace des paramètres du modèle de réseau suggèrent une assez bonne flexibilité sur des indicateurs globaux $\vec{G}$, ainsi que de bonnes propriétés de convergence. Pour une étude du comportement précis, voir l'appendice donnant les regressions traduisant le comportement du modèle couplé. Dans le but d'illustrer la méthode de génération de données synthétiques, l'exploration a été orientée vers l'étude des correlations.

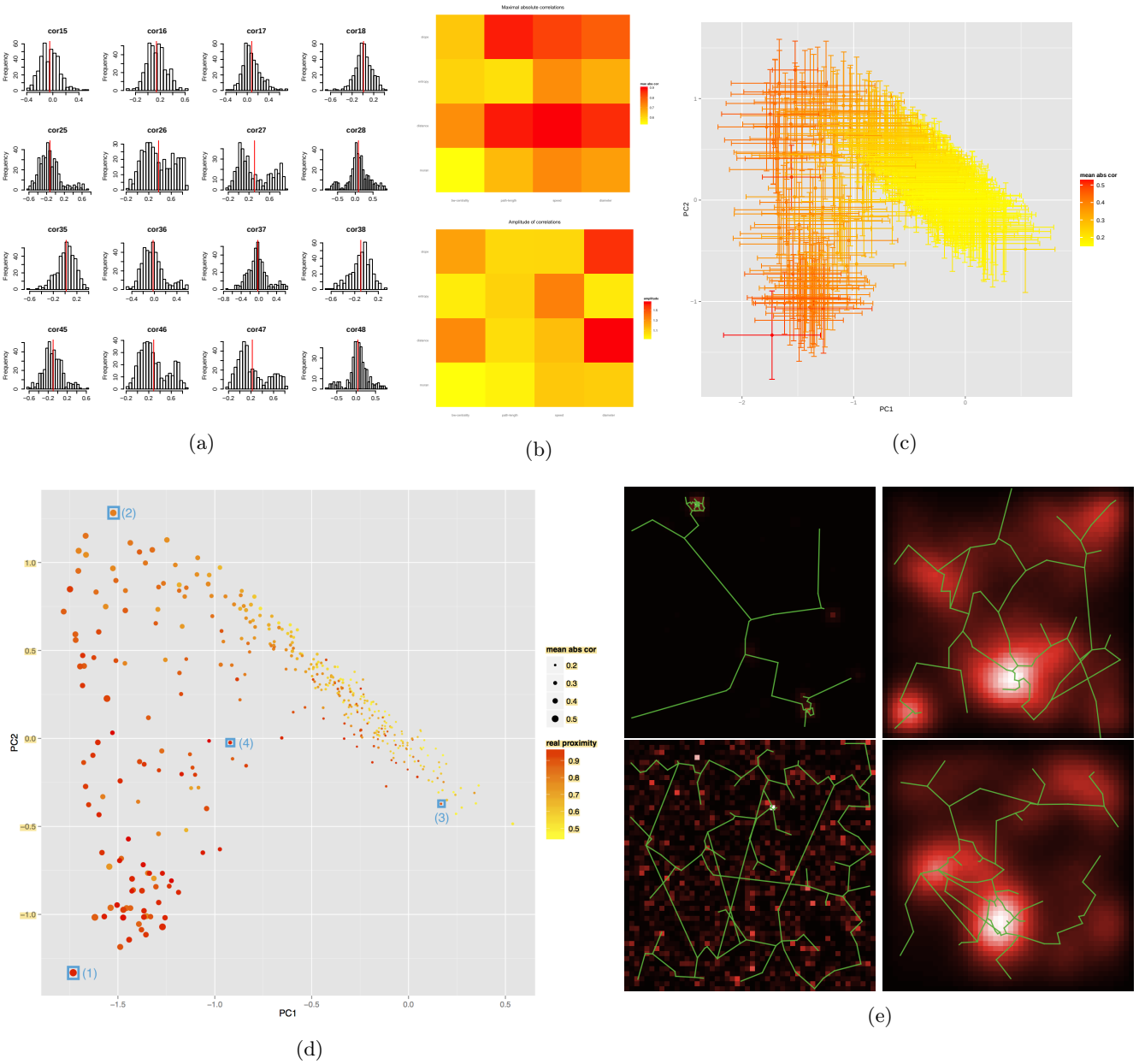
Etant donné la grande dimension relative de l'espace des paramètres, une exploration par grille exhaustive est impossible. On utilise un plan d'expérience par criblage (hypercube latin), avec les bornes indiquées ci-dessus pour $\vec{\alpha}_D$ et pour $\vec{\alpha}_N$, on a $N_C \in [50, 120], r_g \in [1, 100], d_0 \in [0.1, 10], k_h \in [0, 1], \gamma \in [0.1, 4], N_L \in [4, 20]$. Concernant le nombre de réplications du modèle pour chaque valeur des paramètres, moins de 50 sont nécessaires pour obtenir sur les indicateurs des intervalles de confiance à 95% de taille inférieure aux déviations standard. Pour les correlations, une centaine donne des IC (obtenus par méthode de Fisher) de taille moyenne 0.4, on fixe donc $n = 80$ pour l'expérience. La figure 4 donne le détail des résultats de l'exploration. On retiendra les résultats marquants suivants au regard de la génération de données synthétiques corrélées :

- les distributions empiriques des coefficients de correlations entre indicateurs de forme et indicateurs de réseaux ne sont pas simples, pouvant être bimodales (par exemple $\rho_{46} = \rho[r, \bar{l}]$ entre l'index de Moran et le chemin moyen).
- On arrive à générer un assez haut niveau de correlation pour l'ensemble des indicateurs, la correlation absolue maximale variant entre 0.6 et 0.9 ; l'amplitude varie quant à elle entre 0.9 et 1.6, ce qui permet un large spectre de valeurs. L'espace couvert dans un plan principal a une étendue certaine mais n'est pas uniforme : on ne peut pas moduler à loisir n'importe quel coefficients, ceux-ci étant liés par les processus de génération sous-jacent. Une étude plus fine aux ordres suivants (correlation des correlations) serait nécessaire pour cerner exactement la latitude dans la génération.
- les points les plus corrélés en moyenne sont également ceux les plus proches des données réelles, ce qui confirme l'intuition d'une forte interdépendance en réalité.
- Des exemples concrets pris sur des points particuliers distants dans le plan principal montre que des configurations de densité proches peuvent présenter des profils de correlations très différents.

3.2.5 Extensions possibles

Il est possible de raffiner cette étude en étendant la méthode de contrôle des correlations. La connaissance très fine du comportement de N (distribution statistiques sur une grille fine de l'espace des paramètres) conditionnée à D devrait permettre de déterminer exhaustivement $N^{<-1>}|D$ et avoir plus de latitude dans la génération des correlations. On pourra également appliquer des algorithmes spécifiques d'exploration pour essayer atteindre des configurations exceptionnelles réalisant un niveau de corrélation voulu, ou au moins pour découvrir l'espace des correlations atteignables par la méthode de génération [Chérel et al., 2015].

[8] à l'adresse `https://github.com/JusteRaimbault/CityNetwork/tree/master/Models/Synthetic`



(a) (b) (c)

(d) (e)

Figure 4: **Exploration de l'espace des corrélations entre forme urbaine et réseau |** **(a)** Distribution des correlations croisées entre les vecteurs $\vec{M}$ des indicateurs morphologiques (dans l'ordre index de moran, distance moyenne, entropie, hiérarchie) et $\vec{N}$ des mesures de réseau (centralité, longueur moyenne, vitesse, diamètre). **(b)** Amplitude des corrélations, définie comme $a_{ij} = \max_k \rho_{ij}^{(k)} - \min_k \rho_{ij}^{(k)}$, et correlation absolue maximale, $c_{ij} = \max_k |\rho_{ij}^k|$. **(c)** Représentation des matrices dans un plan principal obtenu par Analyse en Composantes Principales sur la population des matrices (variances cumulées: PC1=38%, PC2=68%). Les barres d'erreur sont calculées initialement par des intervalles de confiance à 95% sur chaque élément de la matrice (par une méthode de Fisher asymptotique standard), puis les bornes supérieures sont prises dans le plan principal. L'échelle de couleur donne la correlation absolue moyenne sur l'ensemble des variables. **(d)** Représentation dans le plan principal, l'échelle de couleur donnant la proximité aux données réelles définie comme $1 - \min_r \|\vec{M} - \vec{M}_r\|$ où $\vec{M}_r$ est l'ensemble des mesures morphologiques réelles, la taille des points donne la correlation absolue moyenne. **(e)** Configurations obtenues sur les 4 points particuliers surlignés en (d), dans l'ordre de gauche à droite et de haut en bas. On reconnait des configurations de villes polycentriques (2 et 4), d'habitat rural diffus (3) et de zone à faible densité (1). Voir en appendice pour l'ensemble des valeurs des paramètres, indicateurs et correlations correspondants. Par exemple, $\bar{d}$ est très corrélé à $\bar{l}, \bar{s}$ ($\simeq$0.8) pour (1) mais pas pour (3) alors qu'il s'agit pour les deux d'un environnement rural ; dans un cas urbain on a également beaucoup de variabilité : $\rho[\bar{d}, \bar{c}] \simeq 0.34$ pour (4) mais $\simeq -0.41$ pour (2), ce qui s'explique par une plus forte prégnance de la hiérarchie gravitaire dans (2) $\gamma = 3.9, k_h = 0.7$ (pour (4), $\gamma = 1.07, k_h = 0.25$), alors que la configuration de densité est la même.

4 Discussion

Positionnement

Notre démarche s'inscrit dans un cadre épistémologique particulier. En effet, d'une part la volonté de multi-disciplinarité et d'autre part l'importance de la composante empirique couplée aux méthodes d'exploration computationelles, en font une approche typique des sciences de la complexité, comme le rappelle la structure de la feuille de route pour les systèmes complexes [Bourgine et al., 2009] qui croise des grandes questions transversales aux disciplines à une intégration verticale de celles-ci, qui implique la construction de modèles multi-échelles hétérogènes présentant souvent les aspects précédent. Le croisement de connaissances empiriques issues de la fouille de données avec celles issues de la simulation est souvent central dans leur conception ou leur exploration, et les résultats présentés ici en sont un exemple typique pour le cas de l'exploration.

Applications directes

En partant du deuxième exemple, qui s'est arrêté à la génération des données synthétiques, on peut proposer des pistes d'application directe qui donneront un aperçu de l'éventail des possibilités.

- La calibration de la composante de génération de réseau, à densité donnée, sur des données réelle de réseau de transport (typiquement routier vu les formes heuristiques obtenues, il devrait par exemple être aisé d'utiliser les données ouvertes d'OpenStreetMap[9] qui sont de qualité raisonnable pour l'Europe, du moins pour la France [Girres and Touya, 2010], avec toutefois des ajustements à faire sur le modèle pour supprimer les effets de bord du à sa structure, par exemple en le faisant générer sur une surface étendue pour ne garder qu'une zone centrale sur laquelle la calibration aurait lieu) permettrait en théorie d'isoler un jeu de paramètres représentant fidèlement des situations existantes à la fois pour la forme urbaine et la forme du réseau. Il serait alors possible de dériver une "correlation théorique" pour celles-ci, étant donné qu'une correlation empirique n'est en théorie pas calculable puisqu'une seule instance des processus stochastiques est observée. Vu la non-ergodicité des système urbains [Pumain, 2012], il y a de fortes chances pour que ces processus soient différents d'une zone géographique à l'autre (ou selon un autre point de vue qu'ils soient dans un autre état des meta-paramètres, dans un autre régime) et que leur interprétation en tant que réalisations d'un même processus stochastique n'ait aucun sens, entrainant l'impossibilité du calcul des covariations. En attribuant un jeu de données synthétiques similaire à une situation donnée, on serait capable de calculer une sorte de *correlation intrinsèque* propre à la situation, qui émerge en fait en réalité des interdépendances temporelles des composantes. Connaitre celle-ci renseigne alors sur ces interdépendances, et donc sur les relations entre réseaux et territoires.
- Comme déjà évoqué, la plupart des modèles de simulation nécessitent un état initial, généré artificiellement à partir du moment où la paramétrisation n'est pas effectuée totalement à partir de données réelles. Une analyse de sensibilité avancée du modèle implique alors un contrôle sur les paramètres de génération du jeu de données synthétique, vu comme méta-paramètre du modèle [Cottineau et al., 2015b]. Dans le cas d'une analyse statistique des sorties du modèle, on est alors capable d'effectuer un contrôle statistique au second ordre.
- On a étudié des processus stochastiques dans le premier exemple, au sens de séries temporelles aléatoires, alors que le temps ne jouait pas de rôle dans le second. On peut suggérer un couplage fort entre les deux composantes du modèle (ou la construction d'un modèle intégré) et observer les indicateurs et correlations à différents pas de temps de la génération. Dans le cas d'une dynamique, de par les rétroactions, on a nécessairement des effets de propagation et donc l'existence d'interdépendances décalées dans l'espace et le temps [Pigozzi, 1980], étendant le domaine d'étude vers une meilleure compréhensions des correlations dynamiques.

Généralisation

On s'est limité au contrôle des premiers et second moments des données générées, mais il est possible d'imaginer une généralisation théorique permettant le contrôle des moments à un ordre arbitraire. Toutefois, la difficulté de génération dans un cas concret complexe, comme le montre l'exemple géographique, questionne la possibilité de contrôle aux ordres supérieurs tout en gardant un modèle à la structure cohérente au au nombre de paramètres relativement faibles. Par contre, l'étude de structures de dépendances non-linéaires comme celles utilisées dans [Chicheportiche and Bouchaud, 2013] est une piste de développement intéressante.

[9] `https://www.openstreetmap.org`

5 Conclusion

On a ainsi proposé une méthode abstraite de génération de données synthétiques corrélées à un niveau contrôlé. Son implémentation partielle dans deux domaines très différents montre sa flexibilité et l'éventail des applications potentielles. De manière générale, il est essentiel de généraliser de telles pratiques de validation systématique de modèles par étude statistique, en particulier pour les modèles agents pour lesquels la question de la validation reste encore relativement ouverte.

References

[Abadie et al., 2010] Abadie, A., Diamond, A., and Hainmueller, J. (2010). Synthetic control methods for comparative case studies: Estimating the effect of california's tobacco control program. *Journal of the American Statistical Association*, 105(490).

[Arthur, 2015] Arthur, W. B. (2015). Complexity and the shift in modern science. Conference on Complex Systems, Tempe, Arizona.

[Banos and Genre-Grandpierre, 2012] Banos, A. and Genre-Grandpierre, C. (2012). Towards new metrics for urban road networks: Some preliminary evidence from agent-based simulations. In *Agent-based models of geographical systems*, pages 627–641. Springer.

[Barndorff-Nielsen et al., 2011] Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., and Shephard, N. (2011). Multivariate realised kernels: consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading. *Journal of Econometrics*, 162:149–169.

[Barthélemy and Flammini, 2008] Barthélemy, M. and Flammini, A. (2008). Modeling urban street patterns. *Physical review letters*, 100(13):138702.

[Batty, 2006] Batty, M. (2006). Hierarchy in cities and city systems. In *Hierarchy in natural and social sciences*, pages 143–168. Springer.

[Bolón-Canedo et al., 2013] Bolón-Canedo, V., Sánchez-Maroño, N., and Alonso-Betanzos, A. (2013). A review of feature selection methods on synthetic data. *Knowledge and information systems*, 34(3):483–519.

[Bonanno et al., 2001] Bonanno, G., Lillo, F., and Mantegna, R. N. (2001). Levels of complexity in financial markets. *Physica A Statistical Mechanics and its Applications*, 299:16–27.

[Bouchaud and Potters, 2009] Bouchaud, J. P. and Potters, M. (2009). Financial Applications of Random Matrix Theory: a short review. *ArXiv e-prints*.

[Bouchaud et al., 2000] Bouchaud, J.-P., Potters, M., and Meyer, M. (2000). Apparent multifractality in financial time series. *The European Physical Journal B-Condensed Matter and Complex Systems*, 13(3):595–599.

[Bourgine et al., 2009] Bourgine, P., Chavalarias, D., and al. (2009). French Roadmap for complex Systems 2008-2009. *ArXiv e-prints*.

[Bretagnolle, 2009] Bretagnolle, A. (2009). *Villes et réseaux de transport : des interactions dans la longue durée, France, Europe, États-Unis*. Hdr, Université Panthéon-Sorbonne - Paris I.

[Brunsdon et al., 1998] Brunsdon, C., Fotheringham, S., and Charlton, M. (1998). Geographically weighted regression. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 47(3):431–443.

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

[Chicheportiche and Bouchaud, 2013] Chicheportiche, R. and Bouchaud, J.-P. (2013). A nested factor model for non-linear dependences in stock returns. *arXiv preprint arXiv:1309.3102*.

[Cottineau et al., 2015a] Cottineau, C., Chapron, P., and Reuillon, R. (2015a). An incremental method for building and evaluating agent-based models of systems of cities.

[Cottineau et al., 2015b] Cottineau, C., Le Néchet, F., Le Texier, M., and Reuillon, R. (2015b). Revisiting some geography classics with spatial simulation. In *Plurimondi. An International Forum for Research and Debate on Human Settlements*, volume 7.

[EUROSTAT, 2014] EUROSTAT (2014). Eurostat geographical data.

[Gell-Mann, 1995] Gell-Mann, M. (1995). *The Quark and the Jaguar: Adventures in the Simple and the Complex*. Macmillan.

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

[Jarrow, 1999] Jarrow, R. A. (1999). In honor of the nobel laureates robert c. merton and myron s. scholes: A partial differential equation that changed the world. *The Journal of Economic Perspectives*, pages 229–248.

[Le Néchet, 2015] Le Néchet, F. (2015). De la forme urbaine à la structure métropolitaine: une typologie de la configuration interne des densités pour les principales métropoles européennes de l'audit urbain. *Cybergeo: European Journal of Geography*.

[Mantegna et al., 2000] Mantegna, R. N., Stanley, H. E., et al. (2000). *An introduction to econophysics: correlations and complexity in finance*, volume 9. Cambridge university press Cambridge.

[Moeckel et al., 2003] Moeckel, R., Spiekermann, K., and Wegener, M. (2003). Creating a synthetic population. In *Proceedings of the 8th International Conference on Computers in Urban Planning and Urban Management (CUPUM)*.

[Newman, 2003] Newman, M. E. (2003). The structure and function of complex networks. *SIAM review*, 45(2):167–256.

[Offner, 1993] Offner, J.-M. (1993). Les "effets structurants" du transport: mythe politique, mystification scientifique. *Espace géographique*, 22(3):233–242.

[Offner and Pumain, 1996] Offner, J.-M. and Pumain, D. (1996). Réseaux et territoires-significations croisées.

[Pigozzi, 1980] Pigozzi, B. W. (1980). Interurban linkages through polynomially constrained distributed lags. *Geographical Analysis*, 12(4):340–352.

[Potiron, 2016] Potiron, Y. (2016). Estimating the integrated parameter of the locally parametric model in high-frequency data. *Working Paper*.

[Potiron and Mykland, 2015] Potiron, Y. and Mykland, P. (2015). Estimation of integrated quadratic covariation between two assets with endogenous sampling times. *arXiv preprint arXiv:1507.01033*.

[Pritchard and Miller, 2009] Pritchard, D. R. and Miller, E. J. (2009). Advances in agent population synthesis and application in an integrated land use and transportation model. In *Transportation Research Board 88th Annual Meeting*, number 09-1686.

[Pumain, 2012] 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, 2016] Raimbault, J. (2016). Calibration of a spatialized urban growth model. *Working Paper, draft at https://github.com/JusteRaimbault/CityNetwork/tree/master/Docs/Papers/Density*.

[Ramsey, 2002] Ramsey, J. B. (2002). Wavelets in economics and finance: Past and future. *Studies in Nonlinear Dynamics & Econometrics*, 6.

[Reuillon et al., 2013] Reuillon, R., Leclaire, M., and Rey-Coyrehourcq, S. (2013). Openmole, a workflow engine specifically tailored for the distributed exploration of simulation models. *Future Generation Computer Systems*, 29(8):1981–1990.

[Sanders et al., 1997] Sanders, L., Pumain, D., Mathian, H., Guérin-Pace, F., and Bura, S. (1997). Simpop: a multiagent system for the study of urbanism. *Environment and Planning B*, 24:287–306.

[Schmitt, 2014] Schmitt, C. (2014). *Modélisation de la dynamique des systèmes de peuplement: de SimpopLocal à SimpopNet*. PhD thesis, Paris 1.

[Tero et al., 2010] Tero, A., Takagi, S., Saigusa, T., Ito, K., Bebber, D. P., Fricker, M. D., Yumiki, K., Kobayashi, R., and Nakagaki, T. (2010). Rules for biologically inspired adaptive network design. *Science*, 327(5964):439–442.

[Tsay, 2015] Tsay, R. S. (2015). *MTS: All-Purpose Toolkit for Analyzing Multivariate Time Series (MTS) and Estimating Multivariate Volatility Models*. R package version 0.33.

[Tumminello et al., 2005] Tumminello, M., Aste, T., Di Matteo, T., and Mantegna, R. N. (2005). A tool for filtering information in complex systems. *Proceedings of the National Academy of Sciences of the United States of America*, 102:10421–10426.

[Van den Bulcke et al., 2006] Van den Bulcke, T., Van Leemput, K., Naudts, B., van Remortel, P., Ma, H., Verschoren, A., De Moor, B., and Marchal, K. (2006). Syntren: a generator of synthetic gene expression data for design and analysis of structure learning algorithms. *BMC bioinformatics*, 7(1):43.

[Wilensky, 1999] Wilensky, U. (1999). Netlogo.

[Ye, 2011] Ye, X. (2011). Investigation of underlying distributional assumption in nested logit model using copula-based simulation and numerical approximation. *Transportation Research Record: Journal of the Transportation Research Board*, (2254):36–43.

Appendice : Comportement du modèle couplé, analyse statistique

Les tableaux suivants donnent les résultats quantitatifs de l'analyse du modèle territorial couplé. Il s'agit de régression linéaires simples pour les différents indicateurs ainsi que pour les correlations. On donne à chaque fois les déviations standard ainsi que la significativité des coefficients, codée par la *p-valeur* : (***) $p \sim 0$, (**) $p < 0.001$, (*) $p < 0.01$, (*) $p < 0.05$.

Indicateurs

	BW	Pathlength	relspeed	diameter
R2	0.6098	0.638	0.7049	0.5855
(Intercept)	2.106e-01± 4.728e-03 (***)	1.0939160± 0.0514692 (***)	6.211e-01± 9.722e-03 (***)	2.4956084± 0.1172324 (***)
α	-1.127e-02± 2.012e-03 (***)	-0.530583± 0.021907 (***)	8.469e-02± 4.138e-03 (***)	-1.0366776± 0.04989 (***)
β	9.430e-03± 6.803e-03 ()	0.1738349± 0.0740571 (*)	-6.516e-02± 1.399e-02 (***)	0.2937746± 0.1686813 (.)
n_d	6.786e-04± 6.534e-04 ()	0.0048631± 0.0071129 ()	-4.046e-03± 1.344e-03 (**)	0.0003039± 0.0162012 ()
N_C	-3.005e-04± 2.887e-05 (***)	0.0012026± 0.0003143 (***)	-1.214e-03± 5.937e-05 (***)	0.0040004± 0.0007159 (***)
N_G	4.800e-02± 1.348e-02 (***)	1.4969583± 0.1467356 (***)	-1.400e-01± 2.772e-02 (***)	3.3021779± 0.3342227 (***)
γ	4.615e-03± 4.997e-04 (***)	0.0132129± 0.0054394 (*)	-7.990e-03± 1.027e-03 (***)	0.0389784± 0.0123895 (**)
d_0	2.743e-04± 1.971e-04 ()	-0.0029289± 0.0021453 ()	-5.688e-04± 4.052e-04 ()	-0.0075661± 0.0048863 ()
r_g	-2.726e-05± 2.038e-05 ()	0.0001532± 0.0002219 ()	5.329e-05± 4.191e-05 ()	0.0002358± 0.0005054 ()
k_h	-1.035e-02± 1.952e-03 (***)	-0.095491± 0.021250 (***)	1.992e-02± 4.014e-03 (***)	-0.227141± 0.048402 (***)
N_L	-2.390e-03± 1.243e-04 (***)	-0.0021779± 0.0013533 ()	1.287e-03± 2.556e-04 (***)	-0.0044759± 0.0030824 ()

	moran	distance	entropy	slope
R2	0.7762	0.4753	0.5516	0.6587
(Intercept)	1.106e-02± 1.348e-02 ()	1.158e+00± 3.862e-02 (***)	1.135e+00± 3.024e-02 (***)	0.0839654± 0.0706085 ()
α	-1.631e-02± 5.739e-03 (**)	-2.549e-01± 1.644e-02 (***)	-2.396e-01± 1.287e-02 (***)	-0.7543377± 0.0300528 (***)
β	5.009e-01± 1.940e-02 (***)	-2.497e-02± 5.557e-02 ()	4.580e-01± 4.351e-02 (***)	1.0974897± 0.1015959 (***)
n_d	2.850e-02± 1.863e-03 (***)	-1.069e-02± 5.337e-03 (*)	1.630e-02± 4.179e-03 (***)	0.0322148± 0.0097579 (**)
N_C	-8.494e-05± 8.234e-05 ()	-5.408e-05± 2.358e-04 ()	-5.802e-05± 1.847e-04 ()	0.0001797± 0.0004312 ()
N_G	-7.710e-01± 3.844e-02 (***)	1.222e+00± 1.101e-01 (***)	4.276e-01± 8.622e-02 (***)	1.0692012± 0.2013007 (***)
γ	6.214e-04± 1.425e-03 ()	2.219e-03± 4.081e-03 ()	2.026e-03± 3.196e-03 ()	-0.0015095± 0.0074621 ()
d_0	2.457e-04± 5.620e-04 ()	-3.896e-03± 1.610e-03 (*)	-2.459e-03± 1.261e-03 (.)	-0.0046572± 0.0029430 ()
r_g	4.413e-05± 5.813e-05 ()	1.048e-04± 1.665e-04 ()	1.069e-04± 1.304e-04 ()	0.0003902± 0.0003044 ()
k_h	6.379e-03± 5.567e-03 ()	-5.308e-02± 1.595e-02 (***)	-3.215e-02± 1.249e-02 (*)	-0.0407186± 0.0291524 ()
N_L	5.183e-04± 3.545e-04 ()	-4.340e-04± 1.015e-03 ()	8.870e-06± 7.952e-04 ()	0.0003567± 0.0018565 ()

Correlations

Auto-correlations

	$\rho[r,d]$	$\rho[r,e]$	$\rho[r,s]$
R2	0.5774	0.5093	0.5483
(Intercept)	1.226986±0.034715 (***)	-1.104021±0.039630 (***)	1.21541±0.04161 (***)
α	-0.497699±0.021858 (***)	0.472135±0.024953 (***)	-0.55056±0.02620 (***)
β	0.334188±0.073967 (***)	-0.606526±0.084439 (***)	0.46942±0.08866 (***)
n_d	0.008372±0.007066 ()	-0.023801±0.008066 (**)	0.01310±0.00847 ()
N_G	0.665346±0.146472 (***)	-0.483421±0.167209 (**)	0.99527±0.17557 (***)

	$\rho[d,e]$	$\rho[d,s]$	$\rho[e,s]$
R2	0.5141	0.5383	0.5449
(Intercept)	-1.67032±0.08598 (***)	1.043888±0.016697 (***)	-1.50417±0.07318 (***)
α	1.04667±0.05414 (***)	-0.218800±0.010513 (***)	0.95046±0.04608 (***)
β	-0.91864±0.18320 (***)	0.153942±0.035577 (***)	-0.80453±0.15592 (***)
n_d	-0.04144±0.01750 (*)	0.007833±0.003399 (*)	-0.03624±0.01489 (*)
N_G	-2.23787±0.36278 (***)	0.355478±0.070450 (***)	-2.00208±0.30875 (***)

	$\rho[\bar{c}, \bar{l}]$	$\rho[\bar{c}, \bar{s}]$	$\rho[\bar{c}, \delta]$
R2	0.5892	0.5084	0.581
(Intercept)	1.1594972±0.0508300 (***)	-1.099e+00±5.891e-02 (***)	1.2163505±0.0595172 (***)
α	-0.4989055±0.0216346 (***)	4.722e-01±2.507e-02 (***)	-0.5547464±0.0253321 (***)
β	0.3285271±0.0731373 (***)	-6.064e-01±8.476e-02 (***)	0.4737039±0.0856371 (***)
n_d	0.0079990±0.0070246 ()	-2.399e-02±8.141e-03 (**)	0.0123499±0.0082251 ()
N_C	0.6966627±0.1449132 (***)	-4.869e-01±1.679e-01 (**)	1.0369074±0.1696801 (***)
N_G	0.0011097±0.0003104 (***)	-4.617e-04±3.597e-04 ()	0.0011038±0.0003634 (**)
γ	0.0026346±0.0053719 ()	3.586e-03±6.225e-03 ()	0.0074892±0.0062900 ()
d_0	-0.0009406±0.0021186 ()	9.630e-05±2.455e-03 ()	-0.0009453±0.0024807 ()
r_g	0.0001819±0.0002191 ()	9.951e-05±2.539e-04 ()	0.0002772±0.0002566 ()
k_h	-0.0298275±0.0209863 ()	4.081e-02±2.432e-02 (.)	-0.0833935±0.0245731 (***)
N_L	-0.0015975±0.0013365 ()	1.384e-04±1.549e-03 ()	-0.0060473±0.0015649 (***)

	$\rho[\bar{l}, \bar{s}]$	$\rho[\bar{l}, \delta]$	$\rho[\bar{s}, \delta]$
R2	0.522	0.5494	0.5546
(Intercept)	-1.768e+00±1.266e-01 (***)	1.073e+00±2.450e-02 (***)	-1.5699072±0.1075144 (***)
α	1.043e+00±5.390e-02 (***)	-2.210e-01±1.043e-02 (***)	0.9481902±0.0457610 (***)
β	-9.441e-01±1.822e-01 (***)	1.550e-01±3.525e-02 (***)	-0.8269920±0.1546984 (***)
n_d	-3.719e-02±1.750e-02 (*)	7.778e-03±3.385e-03 (*)	-0.0324681±0.0148582 (*)
N_C	-2.229e+00±3.610e-01 (***)	3.579e-01±6.984e-02 (***)	-2.0137775±0.3065172 (***)
N_G	8.575e-05±7.734e-04 ()	2.392e-05±1.496e-04 ()	-0.0002524±0.0006565 ()
γ	-4.901e-03±1.338e-02 ()	4.571e-03±2.589e-03 (.)	-0.0032340±0.0113624 ()
d_0	4.492e-03±5.279e-03 ()	-6.360e-04±1.021e-03 ()	0.0036384±0.0044813 ()
r_g	-3.880e-04±5.459e-04 ()	-3.725e-05±1.056e-04 ()	-0.0005394±0.0004635 ()
k_h	1.728e-01±5.229e-02 (**)	-1.419e-02±1.011e-02 ()	0.1492597±0.0443898 (***)
N_L	9.444e-04±3.330e-03 ()	-2.201e-03±6.441e-04 (***)	0.0022415±0.0028269 ()

Correlations croisées

	$\rho[r, \bar{c}]$	$\rho[r, \bar{l}]$	$\rho[r, \bar{s}]$	$\rho[r, \delta]$
R2	0.08052	0.1734	0.174	0.1187
(Intercept)	-0.0262848±0.0582459 ()	-1.830e-01±5.801e-02 (**)	-2.800e-01±6.381e-02 (***)	-0.0396938±0.0609953 ()
α	0.0425576±0.0247910 (.)	2.239e-01±2.469e-02 (***)	2.369e-01±2.716e-02 (***)	-0.0546067±0.0259612 (*)
β	-0.3076801±0.0838078 (***)	2.970e-02±8.347e-02 ()	2.487e-02±9.182e-02 ()	0.4311665±0.0877639 (***)
n_d	-0.0225112±0.0080494 (**)	-2.231e-03±8.017e-03 ()	8.874e-03±8.819e-03 ()	0.0418159±0.0084294 (***)
N_C	0.0001178±0.0003557 ()	2.270e-04±3.542e-04 ()	3.001e-04±3.897e-04 ()	-0.0004806±0.0003725 ()
N_G	0.4455534±0.1660556 (**)	-5.238e-01±1.654e-01 (**)	-6.272e-01±1.819e-01 (***)	-0.2879462±0.1738941 (.)
γ	-0.0153085±0.0061556 (*)	4.880e-03±6.131e-03 ()	-3.259e-03±6.744e-03 ()	0.0076671±0.0064462 ()
d_0	0.0019859±0.0024277 ()	2.813e-03±2.418e-03 ()	-3.898e-05±2.660e-03 ()	0.0009132±0.0025423 ()
r_g	0.0001959±0.0002511 ()	5.983e-05±2.501e-04 ()	9.584e-05±2.751e-04 ()	-0.0001899±0.0002629 ()
k_h	0.0412129±0.0240482 (.)	2.120e-02±2.395e-02 ()	5.927e-02±2.635e-02 (*)	-0.0178174±0.0251834 ()
N_L	-0.0011826±0.0015315 ()	5.317e-05±1.525e-03 ()	7.111e-04±1.678e-03 ()	0.0006752±0.0016037 ()

	$\rho[\bar{d}, \bar{c}]$	$\rho[\bar{d}, \bar{l}]$	$\rho[\bar{d}, \bar{s}]$	$\rho[\bar{d}, \delta]$
R2	0.09957	0.726	0.6635	0.2736
(Intercept)	-7.875e-03±1.000e-01 ()	-5.920e-01±5.909e-02 (***)	-0.8352368±0.0769938 (***)	-0.2136136±0.1134114 (.)
α	1.143e-02±4.257e-02 ()	7.915e-01±2.515e-02 (***)	0.8797555±0.0327706 (***)	-0.0164318±0.0482709 ()
β	-6.376e-01±1.439e-01 (***)	-9.049e-02±8.502e-02 ()	0.0652982±0.1107835 ()	1.6099651±0.1631834 (***)
n_d	-4.068e-02±1.382e-02 (**)	-2.081e-03±8.166e-03 ()	0.0208547±0.0106403 (.)	0.0904758±0.0156732 (***)
N_C	3.524e-04±6.108e-04 ()	5.379e-05±3.608e-04 ()	0.0001255±0.0004702 ()	-0.0006020±0.0006926 ()
N_G	1.228e+00±2.851e-01 (***)	-1.668e+00±1.685e-01 (***)	-1.9878616±0.2195048 (***)	-1.1801054±0.3233292 (***)
γ	6.252e-03±1.057e-02 ()	-3.111e-03±6.244e-03 ()	-0.0036648±0.0081369 ()	0.0051939±0.0119857 ()
d_0	8.292e-04±4.169e-03 ()	2.533e-03±2.463e-03 ()	0.0003107±0.0032092 ()	-0.0029674±0.0047271 ()
r_g	4.879e-05±4.311e-04 ()	8.729e-05±2.547e-04 ()	-0.0001203±0.0003319 ()	0.0002683±0.0004889 ()
k_h	3.077e-02±4.129e-02 ()	2.040e-02±2.440e-02 ()	0.0403760±0.0317887 ()	-0.0958810±0.0468245 (*)
N_L	-5.190e-03±2.630e-03 (*)	1.228e-03±1.554e-03 ()	0.0017721±0.0020244 ()	0.0011045±0.0029819 ()

	$\rho[e, \bar{c}]$	$\rho[e, \bar{l}]$	$\rho[e, \bar{s}]$	$\rho[e, \delta]$
R2	0.08321	0.1137	0.04931	0.2212
(Intercept)	1.229e-01±6.394e-02 (.)	-0.2235482±0.0796783 (**)	-1.069e-01±9.653e-02 ()	0.2851561±0.0647508 (***)
α	-1.406e-01±2.721e-02 (***)	0.2003760±0.0339132 (***)	1.098e-01±4.109e-02 (**)	-0.2898858±0.0275597 (***)
β	3.474e-01±9.200e-02 (***)	-0.3847257±0.1146461 (***)	-4.382e-01±1.389e-01 (**)	0.1606966±0.0931675 (.)
n_d	1.386e-02±8.836e-03 ()	-0.0089831±0.0110113 ()	-2.277e-02±1.334e-02 (.)	-0.0058289±0.0089484 ()
N_C	-2.203e-04±3.904e-04 ()	0.0004494±0.0004866 ()	4.578e-04±5.895e-04 ()	0.0001080±0.0003954 ()
N_G	9.209e-02±1.823e-01 ()	-0.5779639±0.2271581 (*)	-4.211e-01±2.752e-01 ()	0.6379269±0.1846008 (***)
γ	6.198e-03±6.757e-03 ()	0.0024670±0.0084206 ()	3.531e-03±1.020e-02 ()	-0.0024161±0.0068431 ()
d_0	-1.465e-03±2.665e-03 ()	-0.0024864±0.0033211 ()	1.782e-03±4.023e-03 ()	-0.0017413±0.0026989 ()
r_g	-7.161e-05±2.756e-04 ()	-0.0002354±0.0003435 ()	-4.484e-04±4.161e-04 ()	-0.0001984±0.0002791 ()
k_h	-1.346e-02±2.640e-02 ()	0.0841865±0.0328970 (*)	9.604e-02±3.985e-02 (*)	-0.0061427±0.0267339 ()
N_L	9.124e-04±1.681e-03 ()	-0.0011866±0.0020950 ()	-8.891e-05±2.538e-03 ()	-0.0024391±0.0017025 ()

	$\rho[s, \bar{c}]$	$\rho[s, \bar{l}]$	$\rho[s, \bar{s}]$	$\rho[s, \delta]$
R2	0.05977	0.6849	0.5995	0.2071
(Intercept)	6.096e-02±8.136e-02 ()	-0.5291554±0.0585027 (***)	-6.684e-01±7.185e-02 (***)	-0.1327534±0.1006121 ()
α	-7.487e-02±3.463e-02 (*)	0.7027361±0.0249003 (***)	7.155e-01±3.058e-02 (***)	-0.0529688±0.0428232 ()
β	-2.238e-01±1.171e-01 (.)	-0.1743332±0.0841774 (*)	-1.466e-01±1.034e-01 ()	1.1937758±0.1447669 (***)
n_d	-1.854e-02±1.124e-02 (.)	-0.0047797±0.0080849 ()	1.271e-02±9.929e-03 ()	0.0711684±0.0139043 (***)
N_C	2.594e-04±4.968e-04 ()	-0.0004555±0.0003573 ()	-5.483e-05±4.388e-04 ()	-0.0004681±0.0006144 ()
N_G	9.746e-01±2.319e-01 (***)	-1.5677015±0.1667878 (***)	-1.554e+00±2.048e-01 (***)	-0.7514017±0.2868391 (**)
γ	2.456e-03±8.598e-03 ()	0.0023678±0.0061827 ()	3.711e-03±7.593e-03 ()	0.0089657±0.0106330 ()
d_0	-1.572e-03±3.391e-03 ()	0.0047638±0.0024384 (.)	1.066e-03±2.995e-03 ()	-0.0017588±0.0041936 ()
r_g	5.791e-05±3.507e-04 ()	-0.0003370±0.0002522 ()	-4.783e-04±3.097e-04 ()	0.0000946±0.0004337 ()
k_h	3.367e-02±3.359e-02 ()	0.0492108±0.0241542 (*)	6.732e-02±2.966e-02 (*)	-0.0728540±0.0415401 (.)
N_L	-5.539e-03±2.139e-03 (**)	0.0017712±0.0015382 ()	9.329e-04±1.889e-03 ()	-0.0004517±0.0026454 ()

	$\rho[e, \bar{c}]$	$\rho[e, \bar{l}]$	$\rho[e, \bar{s}]$	$\rho[e, \delta]$
R2	0.08321	0.1137	0.04931	0.2212
(Intercept)	1.229e-01±6.394e-02 (.)	-0.2235482±0.0796783 (**)	-1.069e-01±9.653e-02 ()	0.2851561±0.0647508 (***)
α	-1.406e-01±2.721e-02 (***)	0.2003760±0.0339132 (***)	1.098e-01±4.109e-02 (**)	-0.2898858±0.02756 (***)
β	3.474e-01±9.200e-02 (***)	-0.3847257±0.1146461 (***)	-4.382e-01±1.389e-01 (**)	0.1606966±0.0931675 (.)
n_d	1.386e-02±8.836e-03 ()	-0.0089831±0.0110113 ()	-2.277e-02±1.334e-02 (.)	-0.0058289±0.0089484 ()
N_C	-2.203e-04±3.904e-04 ()	0.0004494±0.0004866 ()	4.578e-04±5.895e-04 ()	0.0001080±0.0003954 ()
N_G	9.209e-02±1.823e-01 ()	-0.5779639±0.2271581 (*)	-4.211e-01±2.752e-01 ()	0.6379269±0.1846008 (***)
γ	6.198e-03±6.757e-03 ()	0.0024670±0.0084206 ()	3.531e-03±1.020e-02 ()	-0.0024161±0.0068431 ()
d_0	-1.465e-03±2.665e-03 ()	-0.0024864±0.0033211 ()	1.782e-03±4.023e-03 ()	-0.0017413±0.0026989 ()
r_g	-7.161e-05±2.756e-04 ()	-0.0002354±0.0003435 ()	-4.484e-04±4.161e-04 ()	-0.0001984±0.0002791 ()
k_h	-1.346e-02±2.640e-02 ()	0.0841865±0.0328970 (*)	9.604e-02±3.985e-02 (*)	-0.0061427±0.0267339 ()
N_L	9.124e-04±1.681e-03 ()	-0.0011866±0.0020950 ()	-8.891e-05±2.538e-03 ()	-0.0024391±0.0017025 ()

Appendice : Configurations particulières

Les tableaux suivants résument les valeurs des paramètres, indicateurs et correlations pour les 4 configurations présentées en figure 4. On donne les déviations standard pour les indicateurs et les intervalles de confiance à 95% pour les correlations.

Configuration	(1)	(2)	(3)	(4)
α	1.516501e+00	1.555684e+00	1.174289e+00	1.323203e+00
β	4.522689e-02	2.369011e-01	2.878834e-02	2.457331e-02
n_d	1.523121e+00	2.447648e+00	1.005678e+00	1.682232e+00
N_C	9.841719e+01	7.605296e+01	1.176334e+02	9.128398e+01
N_G	2.712064e-02	2.388306e-02	5.093294e-02	1.854230e-02
γ	1.618555e+00	3.917224e+00	2.136885e+00	1.070508e+00
d_0	1.300443e+00	9.939404e+00	2.947665e+00	1.324972e+00
r_g	8.576246e+01	8.394425e+01	8.232540e+01	7.438451e+01
k_h	2.200645e-01	7.177931e-01	8.064035e-01	2.510558e-02
N_L	1.846435e+01	1.803941e+01	9.706505e+00	1.190123e+01
P_m	3.657749e+04	9.526370e+04	4.923410e+04	6.140057e+04
$\bar{c}$	1.433701e-01±1.410993e-02	0.1450925±0.01732669	0.15007478±0.016914043	0.15753385±0.016885300
$\bar{l}$	1.902390e-01±1.490413e-01	0.5264323±0.10235672	0.78084633±0.091161648	0.61213918±0.144466259
$\bar{s}$	6.567478e-01±6.161807e-02	0.6266877±0.03632790	0.55060490±0.030011454	0.63238422±0.049560104
δ	1.026908e+00±4.323170e-01	1.5387645±0.28635511	2.09662379±0.333273034	1.82876419±0.287441536
r	9.582111e-03±1.303745e-03	0.1870607±0.02129171	0.00658906±0.001676165	0.01108170±0.002666821
$\bar{d}$	7.225749e-01±1.475284e-01	0.7894275±0.03793392	0.91603864±0.008777011	0.89253127±0.044771296
e	6.529959e-01±7.448434e-02	0.9656237±0.01121815	0.94523444±0.001576657	0.85650444±0.033640178
s	-1.009041e+00±5.755951e-02	-0.6915159±0.07571627	-0.86066650±0.011647444	-1.04113626±0.022954670
	1.927581e-02	0.1949022	0.08736587	0.01814215

Configuration	(1)	(2)	(3)	(4)
$\rho[r, d]$	-0.4679513 [-6.235562e-01 , -0.2766820]	-0.6021740 [-0.7258547 , -4.407751e-01]	-1.836459e-01 [-0.3877183 , 3.758763e-02]	1.535686e-01 [-0.0684582 , 3.611013e-01]
$\rho[r, e]$	-0.4969249 [-6.460989e-01 , -0.3111840]	-0.1625078 [-0.3690476 , 5.932741e-02]	-1.848848e-01 [-0.3888074 , 3.630699e-02]	5.693427e-01 [0.3996153 , 7.013276e-01]
$\rho[r, s]$	0.3081761 [9.488394e-02 , 0.4944154]	-0.3421209 [-0.5225573 , -1.323529e-01]	-3.313264e-01 [-0.5136505 , -1.203723e-01]	-4.066184e-01 [-0.5749672 , -2.052376e-01]
$\rho[\bar{d}, e]$	0.8508536 [7.762552e-01 , 0.9019531]	0.5705894 [0.4011659 , 7.022647e-01]	-5.995528e-05 [-0.2197740 , 2.196598e-01]	4.344970e-01 [0.2374484 , 5.972012e-01]
$\rho[\bar{d}, s]$	-0.8023042 [-8.688589e-01 , -0.7072636]	0.6692694 [0.5270560 , 7.750198e-01]	6.791720e-02 [-0.1540993 , 2.834050e-01]	-9.559466e-02 [-0.3088251 , 1.267852e-01]
$\rho[e, s]$	-0.9009695 [-9.354970e-01 , -0.8493979]	0.8781442 [0.8158375 , 9.202963e-01]	8.468024e-01 [0.7704290 , 8.992143e-01]	-1.789705e-01 [-0.3836031 , 4.241422e-02]
$\rho[\bar{c}, l]$	0.4234633 [2.246481e-01 , 5.884313e-01]	0.7235166 [0.5990261 , 0.8138558]	0.8431956 [0.7652527 , 0.8967725]	0.4006936 [0.1984478 , 0.5702097]
$\rho[\bar{c}, \bar{s}]$	-0.1023031 [-3.149408e-01 , 1.201137e-01]	-0.7950977 [-0.8638952 , -0.6971747]	-0.7872919 [-0.8585037 , -0.6862905]	-0.4496771 [-0.6092037 , -0.2551716]
$\rho[\bar{c}, \delta]$	0.2568220 [3.932404e-02 , 4.510850e-01]	0.7049855 [0.5742125 , 0.8006794]	0.8299035 [0.7462630 , 0.8877455]	0.6332603 [0.4803829 , 0.7487918]
$\rho[\bar{l}, \bar{s}]$	0.4219260 [2.228701e-01 , 5.872064e-01]	-0.7921187 [-0.8618395 , -0.6930156]	-0.8942859 [-0.9310592 , -0.8395275]	-0.8546133 [-0.9044912 , -0.7816736]
$\rho[\bar{l}, \delta]$	0.6948818 [5.607839e-01 , 7.934560e-01]	0.8158636 [0.7263519 , 0.8781621]	0.8882018 [0.8305739 , 0.9270100]	0.6689605 [0.5266519 , 0.7747964]
$\rho[\bar{s}, \delta]$	0.3805439 [1.755009e-01 , 5.539444e-01]	-0.6657526 [-0.7724742 , -0.5224595]	-0.7760599 [-0.8507179 , -0.6707077]	-0.6680160 [-0.7741129 , -0.5254169]

Configuration	(1)	(2)	(3)	(4)
$\rho[r, \bar{c}]$	-0.0958671 [-0.3090738 , 0.12651465]	-0.2846750 [-0.47470041 , -0.06929222]	0.106860362 [-0.11557001 , 0.3190854424]	0.166141774 [-0.05560493 , 0.372269294]
$\rho[r, \bar{l}]$	0.3184592 [0.1061712 , 0.50298200]	0.3611645 [0.15363946 , 0.53817505]	0.120291668 [-0.10212439 , 0.3312534969]	0.112524033 [-0.10991022 , 0.324224975]
$\rho[r, \bar{s}]$	0.3510671 [0.1423288 , 0.52990926]	0.3864280 [0.18217892 , 0.55870790]	-0.207507373 [-0.40859524 , 0.0127927828]	-0.006745464 [-0.22612722 , 0.213287544]
$\rho[r, \delta]$	-0.3419770 [-0.5224388 , -0.13219278]	0.4010223 [0.19882412 , 0.57047399]	-0.199327994 [-0.40146255 , 0.0213227474]	-0.101075026 [-0.31382258 , 0.121336498]
$\rho[\bar{d}, \bar{c}]$	-0.1833004 [-0.3874145 , 0.03794465]	-0.4143381 [-0.58114893 , -0.21411352]	0.006860986 [-0.21317727 , 0.2262368358]	0.348614959 [0.13959020 , 0.527896760]
$\rho[\bar{d}, \bar{l}]$	0.8121174 [0.7210643 , 0.87559646]	0.7363382 [0.61633582 , 0.82291849]	0.191569016 [-0.02938471 , 0.3946737203]	0.607176014 [0.44710613 , 0.729564080]
$\rho[\bar{d}, \bar{s}]$	0.8175828 [0.7287819 , 0.87933822]	0.7718832 [0.66493682 , 0.84781438]	-0.219254065 [-0.41879596 , 0.0004862563]	0.731832543 [0.61023974 , 0.819738748]
$\rho[\bar{d}, \delta]$	-0.8199313 [-0.8809437 , -0.73210512]	0.7272964 [0.60411696 , 0.81653203]	-0.205542803 [-0.40688430 , 0.0148444885]	-0.146644197 [-0.35492533 , 0.075505502]
$\rho[e, \bar{c}]$	-0.2355562 [-0.4328696 , -0.01670384]	0.1482222 [-0.07390147 , 0.35633436]	-0.028013851 [-0.24621526 , 0.1928903073]	-0.213237762 [-0.41357774 , 0.006797615]
$\rho[e, \bar{l}]$	0.4856866 [0.2977430 , 0.63738583]	-0.3440356 [-0.52413307 , -0.13448447]	0.005838058 [-0.21415354 , 0.2252660012]	-0.259589628 [-0.45344380 , -0.042284479]
$\rho[e, \bar{s}]$	0.5700927 [0.4005481 , 0.70189139]	-0.5411854 [-0.68004036 , -0.36485222]	0.206851807 [-0.01347763 , 0.4080244658]	-0.520820100 [-0.66449667 , -0.340011761]
$\rho[e, \delta]$	-0.5105626 [-0.6566203 , -0.32759502]	-0.4972353 [-0.64633902 , -0.31155624]	0.180368653 [-0.04097196 , 0.3848345410]	0.007300570 [-0.21275760 , 0.226653900]
$\rho[s, \bar{c}]$	-0.5398119 [-0.6789960 , -0.36316900]	-0.3711488 [-0.54631495 , -0.16487721]	0.066233469 [-0.15575000 , 0.2818487302]	0.191933347 [-0.02900679 , 0.394992994]
$\rho[s, \bar{l}]$	0.8100685 [0.7181768 , 0.87419174]	0.5960875 [0.43309300 , 0.72133128]	0.103616019 [-0.11880563 , 0.3161357092]	0.293165083 [0.07850495 , 0.481844758]
$\rho[s, \bar{s}]$	0.8166515 [0.7274653 , 0.87870120]	0.5501248 [0.37583582 , 0.68682409]	-0.214962957 [-0.41507543 , 0.0049896034]	0.326893808 [0.11547047 , 0.509981729]
$\rho[s, \delta]$	-0.7869970 [-0.8582997 , -0.68588011]	0.5902345 [0.42572786 , 0.71697130]	-0.234261611 [-0.43175552 , -0.0153339732]	-0.053173625 [-0.26973912 , 0.168512022]