The Old and the New: Qualifying City Systems in the World with Classical Models and New Data

Cura R., Cottineau C., Swerts E., Ignazzi C., Bretagnolle A., Vacchiani-Marcuzzo C., Pumain D.

Within the context of a growing share of urban population in the world and the continuation of the urban transition for half of the world population during the next four or five decades, it is crucial that social sciences provide sound interpretations of the structure and evolution of urban systems. We are interested here in statistical theories about urban hierarchies and dynamics that are classically approached through Zipf's city size distribution and Gibrat's urban growth models. A huge literature has been published about these models which seem to provide a reasonable explanation to the empirically observed similarities in the hierarchy of city sizes (reviews can be found in Nitsch, 2005, Favaro, Pumain, 2011, Cottineau, 2016). However, this knowledge is far from being cumulative, partly because of a lack of attention to the quality of urban data and partly because of diverging conceptions about the role of these models.

Measuring city sizes indeed is not a purely technical question. According to the focus of the analyses, several relevant definitions of a "city" may be retained, illustrating different aspects of urban life, thus leading to different measures of its size. Questions of governance and policies are better examined in the frames of administrative local units or the more recent metropolitan institutions, whereas technical networks may be better approached within the delimitation of "urban agglomerations" including a continuous built-up area spreading over several administrative local units. As we are here looking for a theoretical interpretation of urban development we chose a definition which is conceptualized in abstract terms of geographical space-time scale for the sake of international and historical comparisons but which keeps concretely the identity of an urban nucleus in terms of a place of concentration of daily activities that usually has been spatially expanding over time.

Regarding the role of urban models this paper is not aiming at establishing which could be considered as "universal", contrary to the ambition of some articles from "hard science" (for instance, Corominas-Murtra and Solé 2010). Our main point is not either to propose more refined testing methods (as for instance in Schaffar and Pavleas, 2014), since their accuracy would probably exceed the precision of the available data. We think that urban models, because of their generality, may well reflect rather universal underlying processes in urban dynamics that are grounded in large numbers of longstanding spatial interaction, obeying the same kind of geographical constraints (Zipf, 1941, Pumain, 2006, Favaro, Pumain, 2011, Hernando et al., 2015). But their application to empirical urban systems also reveal deviations from the general models some of which can be meaningfully related to the demographical and political evolution of the countries under investigation (Moriconi-Ebrard, 1993). This paper aims at investigating variations around these models and establishing if the explanation of some deviations can be found in a diversity of geographical situations. The statistical results

that are obtained are of course depending upon our databases that remain subject to a number of errors and uncertainties but are to our knowledge of the best possible quality for comparing the seven regions of the world that we investigate.

Usually the models of urban size and growth are tested on national subsets of cities. In our conclusive remarks we shall more precisely qualify the possible relevance of this kind of spatial envelope. We want here to test these models on large countries for which data about urban populations are available for many decades. The "BRICS" completed with USA and Europe happens to make a good sample because amounting together to almost half of the world population, being dispersed on all continents and including countries among the most developed as well as from the "global South". In this paper, more than 20 000 urban agglomerations and metropolitan areas larger than 10 000 inhabitants are observed at six points in time during a fifty years period, from the 1960s to the 2010s. This data set offers therefore the opportunity of experimenting models that have been mostly designed for industrialized regions on more recently emerging economies. Moreover, as urban growth is still very rapid in many of these parts of the world, our findings may be useful at forecasting some future trends.

In a previous paper using about the same data sets we have compared the forms of the urban hierarchical organisation in each country or world region and classified the trajectories of cities in terms of the evolution of their population (Pumain et al., 2015). In this paper we concentrate on Zipf's and Gibrat's models and go deeper in the related statistical evidence about urban growth process and its connection with urban hierarchies. We first describe city size distributions on a sample of large urban systems in the world having distinct histories in the urban transition process, aiming at finding which statistical form of distribution can better describe them and how it has evolved over time, comparing the power law (Zipf) and lognormal models. Second, we test the hypotheses of the Gibrat's growth model on each urban system. Third, we investigate the trend towards growing hierarchical inequalities over time by comparing observations and expectations from a pure Gibrat's model and discuss how such a simulation model may help to evaluate predictions about the future evolution of city sizes. In each section we rely on a recent working paper including many references (Cottineau, 2016) to compare our results with those found in the literature, We conclude in a final discussion about what can be expected for the improvement of theories about urban dynamics.

1   Statistical city size distribution: Zipf or log-normal?

Zipf's "rank-size" distribution, analogous to a Paretian model and Gibrat's log-normal distribution, of the exponential type, are often tested as alternative candidates for describing urban hierarchies. However, they are recognized as belonging to the same family of highly skewed statistical distributions (Barbut, 2004) and since long many scholars (as Quandt, 1964) using quantity of empirical testing demonstrated that many kinds of skewed statistical distributions could fit the same distribution of city sizes. A more recent review on this topic can be found in Clauset et al. 2009. In this paper we are not interested in deciding which of

these models would lead towards an ultimate universal statistical explanation or even could be envisaged as a societal norm. We simply want to use them as benchmarks enabling to characterize urban evolution and tools for comparing systems of cities in a variety of historical and geographical contexts. We first recall briefly how we propose to measure city sizes for harmonised urban statistics in space and time and show the results for the sample of world regions that we have chosen. These databases enable to test and characterise the shape of the statistical distributions and how they evolve over time.

1.1 Urban definitions for comparison

Building on a longstanding geographical view on the hierarchical organisation of urban systems (Berry, 1964, Pred, 1977) and its revised version as an evolutionary process (Pumain, 2000 and 2006) we consider the definition of cities as geographical entities to be necessary at mainly two spatio-temporal scales: cities as evolutionary centres of concentration of people and activities keeping a consistent meaning at local scale as daily urban systems (strong interactions connecting in average 3 or 4 different places of activity per person) and cities as attractive nodes within competing networks of cities at higher scales of space and time (less frequent interactions but generating interdependencies leading to inter-urban co-evolution). That is why for international comparison we rely on standardized databases where "cities" are not taken for granted from the official statistical definitions but designed according to similar criteria (Bretagnolle et al., 2015). These harmonised databases have been built for seven regions of the world[1] (Figure 1). Depending to the availability of time series, it was not always possible to apply the same delineation for the urban entities in each region, although the same concept of "daily urban system" guided the data aggregation: functional urban areas have been designed for the USA and South Africa; urban agglomerations of continuous built-up area were constituted from aggregating the smallest possible spatial building blocks providing data on their population and activities in the cases of China and India (Swerts, 2013) and for Brazil (Ignazzi, 2014), as well as for historical databases describing cities in Europe (Moriconi-Ebrard, 1994 and Bretagnolle et al., 2010) and in the former Soviet Union – this territory appears more relevant than present Russia for analyzing urban evolution over the past half-century because of persistent and strong inter-urban linkages from the socialist period (Cottineau, 2014). For the sake of comparison, six dates were selected from each national censuses to present the state of urban hierarchy in each system, every ten years or so between 1960 and 2010, although data are neither entirely comparable nor necessary reliable - e.g. for China in the first period (1964 census). The resulting urban units have variable delineations over time in order to follow their spatial expansion, except European urban agglomerations for which populations were collected within their delineation at final date (Moriconi-Ebrard, 1994).

Although we think that due to the huge effort in making it comparable[2] this material is reasonably reliable and perhaps a good support for international comparisons we cannot claim

---

1 Five of these data bases were built thanks to the funding of the ERC grant GeoDivercity

2 A rough estimation is of about 5 person-years for the input of work that was dedicated at preparing the full dataset (without counting the amount of time already included in the census data themselves).

that better statistics could not be provided. Before these databases were made freely available online to enable the reproducibility of our experiments, we have made a few checks with international databases. Comparisons are easy for the largest cities (table 1). Sometimes, other existing online databases tend to overestimate populations (i.e. populationdata.net or citypop.de which substitute the population of some megacities by the population of the conurbations of megalopolises of which they are the largest centre) while statistics from the UN urbanization department often underestimate the size of some cities.

Table 1 Population of largest cities in seven urban systems in 1960 and 2010

Rank	City	Country	Population around 1960	Population around 2010	Online database*	Online database**
1	Delhi	India	5 952 080	22 424 729	21 753 486	26 400 000
2	Shanghai	China	6 376 930	20 275 662	24 256 800	30 900 000
3	Sao Paulo	Brazil	3 781 446	19 672 582	21 090 791	21 800 000
4	New York	USA	10 694 633	19 567 410	23 723 696	22 000 000
5	Kolkata	India	5 283 105	17 251 493	15 644 040	16 000 000
6	Mumbai	India	4 918 815	16 743 028	21 900 967	23 200 000
7	Beijing	China	4 033 296	15 938 241	21 516 000	21 100 000
8	Moscow	Russia	7 186 148	14 425 208	14 926 513	16 900 000
9	Guangzhou	China	2 134 254	13 944 080	65 701 102	47 700 000
10	Los Angeles	USA	6 038 771	12 828 837	18 679 763	17 600 000
13	Paris	France	7 514 899	10 416 084	12 405 426	11 200 000
17	London	U-K	8 902 770	9 423 910	14 184 776	14 400 000
20	Johannesburg	South Africa	1 753 605	7 613 297	4 949 347	13 400 000

Source: GeoDiverCity databases on urban agglomerations > 10 000 inhab.

NB : we have added the largest city of each region to the ten largest cities of our whole sample.

*https://www.populationdata.net/**http://www.citypopulation.de

But unlike to most of the urban rankings that are now displayed online, we do not limit our investigation to 1,000,000+ cities. Our database includes every urban agglomeration over 10,000 inhabitants in each of the seven world regions, amounting to a total of 20,782 cities (Figure 1). In table 2 these towns and cities are distributed in size classes. These distributions both reflect the general inverse geometric progression of the number of cities with size and the well known peculiarities of the urbanisation in each continent.

Table 2 Number of cities in size classes (thousand inhab.) around 2010

Country	10 - 100	100 - 1 000	1 000 - 10 000	> 10 000
Brazil	1636	142	20	2
China	8002	1107	74	5
Europe	3455	451	39	1
India	4646	489	47	3
FSU	1643	251	26	1
South Africa	223	26	5	0
USA	536	322	49	2

Source: GeoDiverCity databases on urban agglomerations > 10 000 inhab.

FSU = Former Soviet Union

We successively apply Zipf's and Gibrat's models to these distributions. Each model has interesting properties: Zipf's model is a good tool to visualize the shape of the distribution of city sizes, to compare their inequalities and to follow their evolution, whereas the lognormal distribution is well connected with a random growth process which is easy to test empirically.

Figure 1 Populations of cities around 2010 in the seven systems
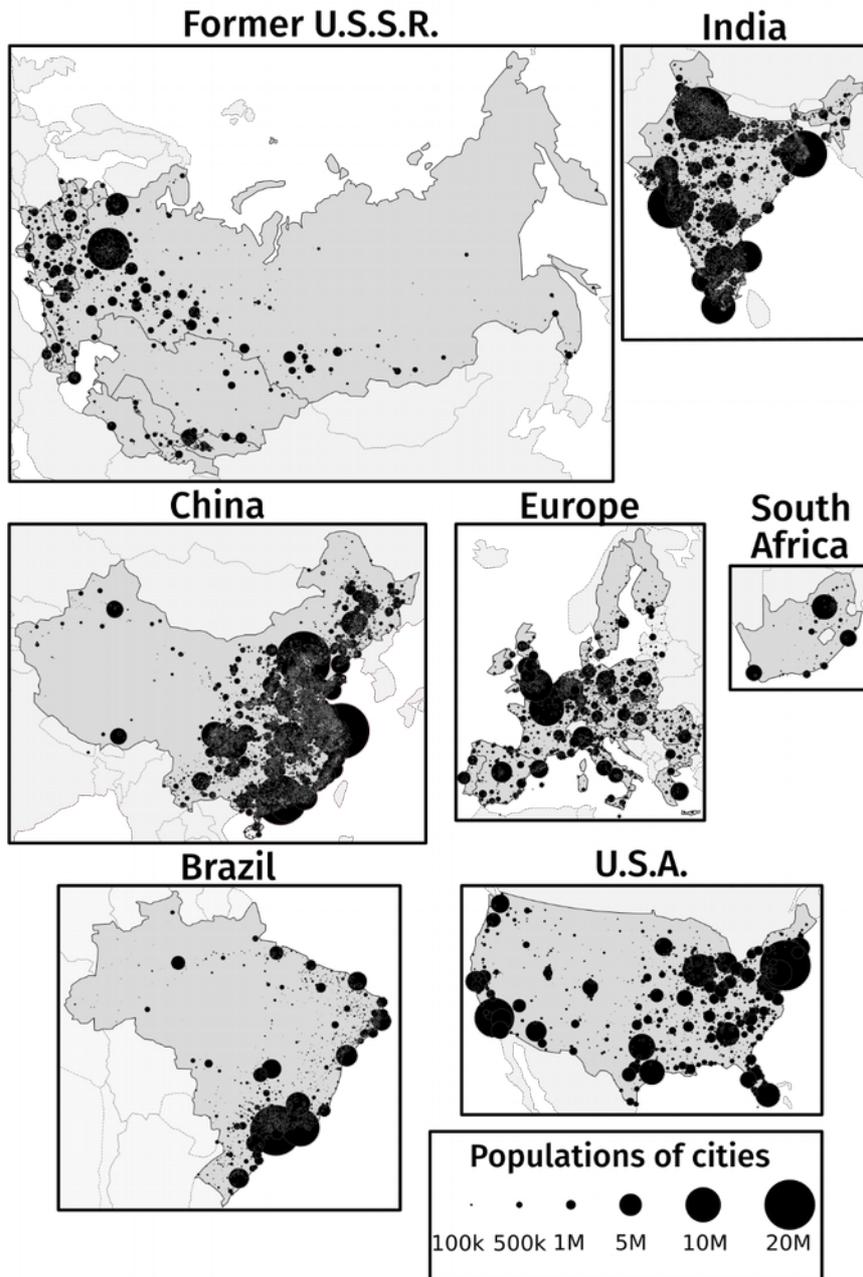
1.2 Zipf rank-size curves and their evolution

The latest city size distribution of each of the seven world regions is represented on a log-log graph in figure 2, enabling to compare the number and size of cities as it varies according to the general urban characteristics of the countries (i.e. size and urbanisation level).

Figure 2 Observed city sizes distributions in seven regions of the world in 2010 or 2011
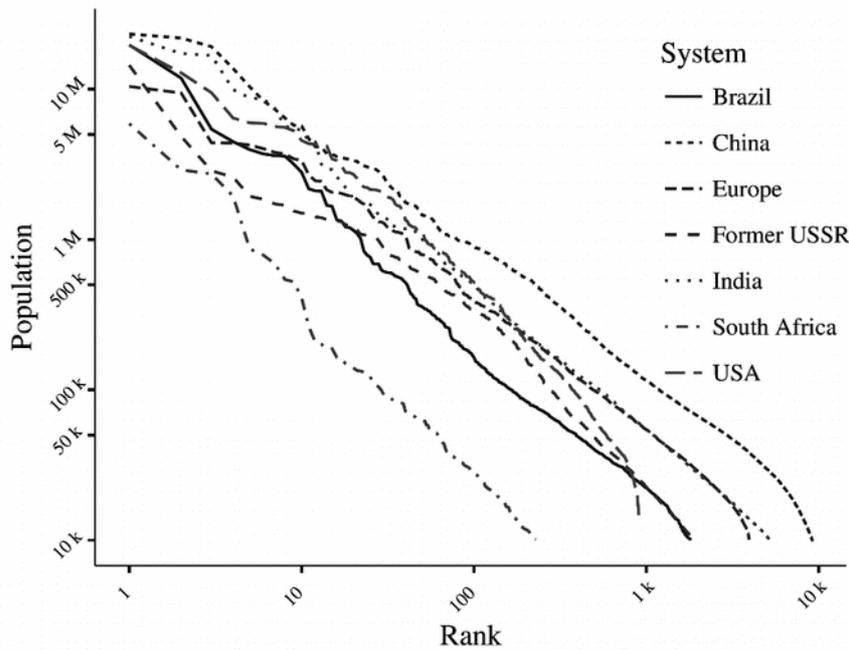
Source: GeoDiverCity databases on urban agglomerations > 10 000 inhab.

Well-known features of urban systems can be recognised on these graphs, for instance the similarity of their shapes, which are close to the straight line which would perfectly fit the model and the sometimes convex or concave patterns for the upper tail of the urban hierarchies compared to the more regular lower part. China and India each have three major cities which are not in the continuity of their rank-size distribution, whereas the distributions of the US and South African systems are steeper than the others. Such visual comparisons have to be complemented by statistics, so adjustments to Zipf's model have been made for each system using an OLS estimation of the Lotka formulation:

$\log (P_i) = \alpha \log (i) + \beta$, where P_i is the population of the city of rank i and α the slope of the adjusted line is considered as a measure of the unevenness of the size distribution. This parameter is independent of the size of the urban system (contrary to what could be an equivalent measure in the case of the lognormal model, i.e. the standard deviation of logarithms of city sizes). Results are presented in table 3 for the latest available date (including the estimated values of α with their upper and lower bound of confidence interval at 1% and quality of adjustment as measured by R^2) whereas details for the other dates are given in Appendix 1.

Table 3 Adjusted slopes of distribution of city sizes to Zipf's model in 2010

Country	Year	Slope	R²	UpperBound	LowerBound	NbCities
Brazil	2010	-0.959	0.992	-0.955	-0.963	1800
China	2010	-0.949	0.978	-0.946	-0.952	9187
Europe	2011	-0.942	0.992	-0.939	-0.944	3946
India	2011	-0.970	0.999	-0.969	-0.971	5185
FSU	2010	-1.104	0.987	-1.098	-1.109	1921
S. Africa	2011	-1.151	0.991	-1.138	-1.165	254
USA	2010	-1.237	0.974	-1.223	-1.250	909

Source: GeoDiverCity databases on urban agglomerations > 10 000 inhab.

Slope values of the adjusted rank-size curves all differ from 1 in a significant way. Their absolute values are above 1 in regions that were settled relatively recently as compared to Europe or Eastern and Southerm Asian countries where urbanization developed well before and in a more continuous way than in "New world" countries such as the USA or South Africa. The explanation is that cities developed there mainly from 19th century on together on less dense rural populations and with more rapid and efficient transportation means that could produce both larger cities and make them more distant from each other. Our results are congruent with what was found by Moriconi-Ebrard (1993) and confirmed in her recent extensive review of the literature and meta-analysis by Cottineau (2016) who concludes: "there is a persistent difference in the size distribution of areas of ancient and recent urbanization, with the latter being more uneven. The structure of urban settlement thus proves rather independent from economic development and also very long to adjust to new transportation features."

The evolution of the distributions of city sizes in each world region over our relatively short period of fifty years is represented on figure 3. Every urban system exhibits a remarkable persistency of the peculiarities of its curve. But there is not a universal trend regarding the evolution of the inequalities of city sizes. In Brazil and India there is a continuous trend over the last fifty years to an increase of inequalities of city sizes which confirms previous observations by Pumain and Moriconi (1997) about the "metropolization" process occurring during the evolution of urban systems. Other observations made on longer and early historical periods – from 1600 to 1980 - for Europe (Bretagnolle et al., 2000) or over the last hundred years for USA (Bretagnolle et al., 2008) also demonstrated such a trend of increasing unevenness of urban size distributions when cities are correctly delineated. Bretagnolle et al. (2008) explain the contradiction of their results with those published by Black and Henderson (2003) or Dobkins and Ioannides (2000) who retropolated within the delineation of metropolitan areas defined in 1950 the population of cities from 1900: "that method overestimates the population of cities in past periods and hides the concentration process due to the spatial expansion of cities" (p. 58).

Between the 1960s and the 2010s the values of the slope in Europe, the USA and South Africa however fluctuated without any clear trend. China appears as an exception with a continuously decreasing value of the inequality index, as already noted by Swerts (2013). It seems thus that a trend toward increasing size inequalities cannot always be acknowledged as

a recurrent stylized fact that could be part of the urban theory (Cottineau, 2016). This will be more precisely measured in the last sections of the paper when using the Gibrat's urban growth model.

**Figure 3 Observed city sizes distributions in seven regions of the world in 2010**
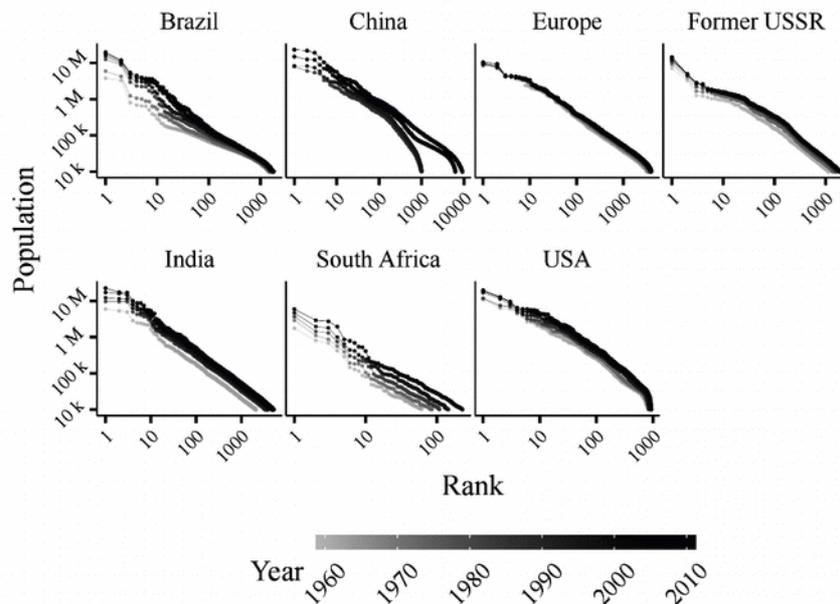
Source: GeoDiverCity data bases on urban agglomerations > 10 000 inhab.

1.**3 Adjustment to a lognormal distribution (with three parameters)**

The similarities between Paretian and exponential models are obvious and the distribution of city sizes can as well be adjusted to a lognormal distribution (Robson, 1973, Parr and Susuki, 1973, Eeckhout, 2004). As this form is a better description of the entire settlement system including villages, hamlets and even isolated houses, rather than of cities only (the growth model is probabilistic), the inventor of the model himself (Gibrat, 1931) suggested the use of a formula with three parameters, adding, to the mean and standard deviation of the lognormal distribution, the population cutoff used for defining cities, in our case 10,000 residents, and subtracting this fixed quantity to the population variable for better adjusting the lognormal distribution. We thus applied the usual tests of normality (Shapiro-Wilk and Kolmogorov-Smirnov according to Gillespie 2015) to the distributions of [log (P_i - 10,000)] for each region, three of which (Brazil, Europe and India) are plotted in figure 4 – these tests are not conclusive for any country according to their values given in Appendix 2. However in all cases the observed distributions are rather symmetric but a little less concentrated than the pure statistical model of the same mean and standard deviation (skewness and kurtosis coefficients are not compatible with lognormality), and small irregularities are visible on both sides of urban hierarchies. These little deviations, compared to the statistical model, appear

more clearly on the graphs of figure 5 where the observed distribution of population sizes is plotted in quantiles against their theoretical values in a lognormal distribution for the same three countries. Whatever the case and the continent, the observed size of the smaller towns tend to be "too small" whereas largest cities are "too large" compared to the values expected in a pure lognormal distribution. We shall try to explain these "anomalies" by analyzing the urban growth process in section 2.

Figure 4 Lognormality comparison for city sizes distributions
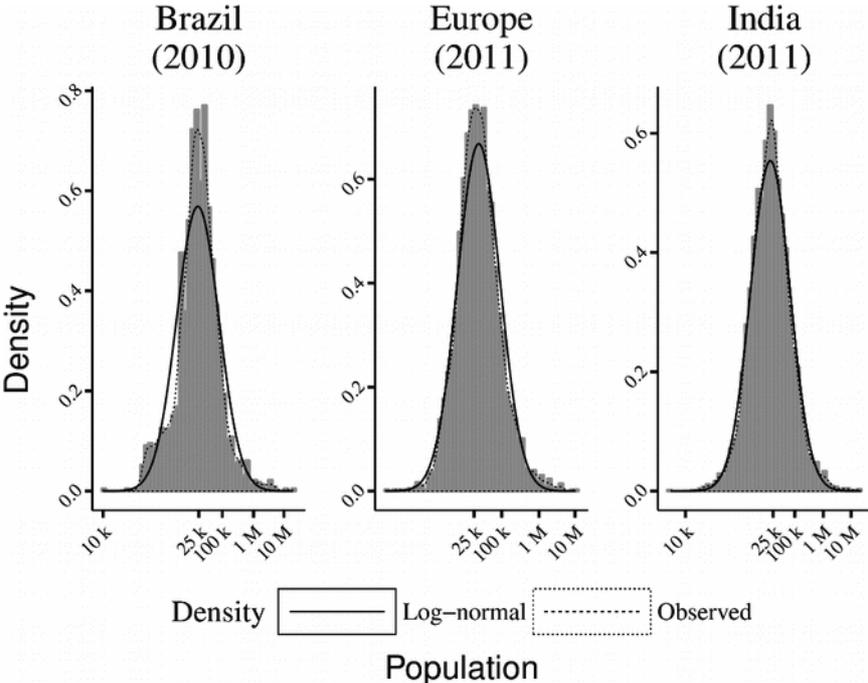
Source: GeoDiverCity data bases on urban agglomerations > 10 000 inhab.

Figure 5: Observed and theoretical (lognormal model) city size distributions
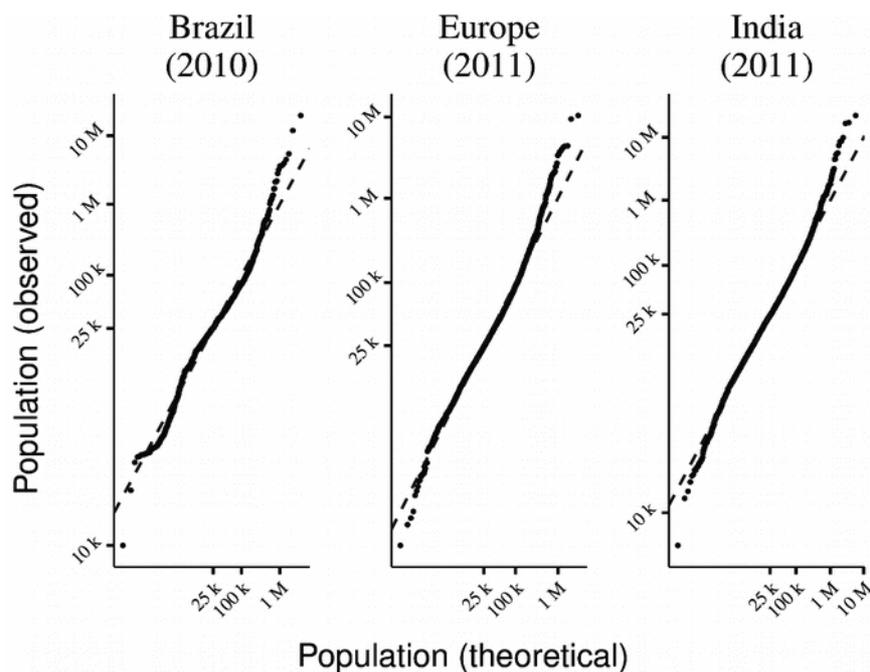
Source: GeoDiverCity data bases on urban agglomerations > 10 000 inhab.

2   Testing Gibrat's urban growth model

If the adjustments to mathematical distributions were always perfect with only a small random noises it would be easy to explain the distribution of city sizes by a stochastic process describing the way urban growth is distributed within the systems of cities. Statisticians and physicians try to imagine which dynamic processes could explain this type of distributions that can be found in a variety of natural or social phenomena (Newman, 2004). Since Simon's paper in 1955 recurrent proposals have been made about possible stochastic processes that would lead to a Paretian or a lognormal distribution of city sizes (for instance Tomita et al., 2008 or Baek et al., 2011). A major advantage of the earlier Gibrat's model (1931) is that is establishes a mathematical connexion between the shape of the city size distribution and the process of growth distribution in the urban system in a way that can be tested with our databases. According to Gibrat (1931) a lognormal distribution of city sizes is generated if on small time intervals the quantity of population added to each city is (in probability) proportional to its size (this is equivalent to say that the probability of any growth rate is the same for all cities of the system) and if there is no correlation between growth rates and size and between growth rates during former or later period. The lognormal distribution (which Gibrat calls the "law of proportional effect") results both from a stochastic process and from the application of the central limit theorem about the addition of independent variables.

2.1 Testing the hypotheses of the urban growth model

The growth model can easily be tested by checking first if urban growth rates are independent of city size and second if they are not auto-correlated in time. After the first test of this model on French cities by Gibrat (1931), a pioneer work on urban growth was achieved by B.

Robson (1973) on a well-designed urban dataset about Britain and Wales cities and towns from 1810 to 1911. He found stylized facts of empirical evidence that were confirmed by analogous studies over two centuries on France (Pumain, 1982, Guérin-Pace, 1993) and Europe (Bretagnolle et al., 2000). Indeed when testing Gibrat's model, attention should be paid to the length of the time period and the fineness of the temporal resolution for which urban growth is analyzed. Even if the mathematical model obviously predicts the emergence of a lognormal distribution after an infinite duration only, testing the validity of the model in the real world requires a sufficient number of time intervals. This is not always the case even in advanced papers using sophisticated methods of data collection, as for example in Rozenfeld et al. 2008 where they test Gibrat's growth model on a ten-year period (1990-2000). The same criticism could be addressed to panel analysis aiming at finding specific growth trends for cities from longitudinal data which often have a much too limited time scope (for instance Resende, 2004 on Brazilian cities from 1980 to 2000). It also must be kept in mind that the measurement of urban growth is highly dependent on the quality of urban data and may lead to different conclusions as demonstrated by Bretagnolle et al. (2008) on the USA case. Ideally, the model should be tested over long periods of time, on large number of towns and cities, and on different delineations, including (variable delineation) or not (constant delineation) the spatial extensions of urban areas for each time interval (Paulus, 2004, Bretagnolle et al., 2015). This last test was not possible for all the data sets examined here and thus our results must be considered cautiously.

In fact, if urban growth always appears to be roughly proportional to city size in absolute terms, and if on the long run the interdependencies between cities belonging to the same system of cities tend to smooth its momentary inequalities, urban growth rates fluctuate much over space and time: in our data bases the standard deviation of growth rates is always at least of the same magnitude as their mean and most of times much higher. This short-term heterogeneity of growth rates distributions among cities of the same system is explained mainly because local urban growth is often a response to the impulses that are generated by innovation waves which are neither constant nor ubiquitous but usually filter down the urban hierarchy or follow networks of urban specialized centers (Hägerstrand, 1952, Pred, 1977, Favaro, Pumain, 2011). This could provide an explanation to the "short- and long-term memories in the social system" as well as the "spatial coherence" in urban growth detected in their spatio-temporal analysis of US cities from 1900 to 2000 by Hernando et al. (2014). Indeed, the observed spatial and temporal fluctuations of urban growth rates reflects a very complex pattern of urban co-evolution (Paulus, 2004) that requires a more disaggregated type of generative modeling than the statistical exercise we are doing here (see for instance the agent-based model MARIUS simulating the urban growth patterns in former USSR by Cottineau et al., 2015).

Our present data set does not cover a very long period of time (fifty years only) but includes a large number of cities of various sizes and functions which should enable a reasonable assessment of the validity of the model. We have computed for each urban system of the seven considered regions the correlations between urban growth rates of five time intervals and the population sizes of cities (a nonlinear relation, using the logarithm of city size was also computed, see Appendix 3) and the temporal auto-correlation of growth rates (table 4).

Our results do confirm observations that were already made by other scholars who analyzed urban growth on several time intervals. The hypotheses of Gibrat's model are not always verified and systematic deviations from a purely stochastic model appear. When it exists the correlation with city size is very slight and more often positive than not (Robson, 1973). It was shown that a consequence of periods of crisis and wars was a tendency to take refuge in smaller towns or countryside, which corresponds to most cases of negative correlation (Pumain, 1982, Brakman, Garretsen, 2004). Such cases could become more frequent in the future if the fate of "shrinking cities" were to expand. A consequence would be a reversal of the historical trend towards an increasing concentration in the distribution of city sizes because negative growth would tend to reduce the variance of city sizes. Indeed, we have demonstrated elsewhere that a lower concentration of settlement sizes already characterizes the French villages whose population has been declining on average for more than a century (Pumain, 2006, p. 208 figure 9) and this process could extend in the future to at least the small towns which are bypassed in actual geographical space and devitalized from many of their former activities and services as the speed of communication and transportation increases.

Regarding the temporal autocorrelation of growth rates, series of positive values appear frequently with low but significant values. These results partly depend on the quality of data and partly reveal different stages in the evolution of urban systems. In table 4 for instance the rather high correlation between city sizes and growth rates in Europe is for one part an artifact explained by the construction of the database (i.e. urban agglomerations delineated in a constant spatial extent at final date) while the temporal autocorrelation of growth rates is due to the regional heterogeneity regarding the stages of urban transition which transported the higher urban growth successively from Northern and Western Europe towards the Southern then Eastern parts during that half century.

Table 4 Correlation coefficients of urban growth with city size (five first columns) and over time (last four columns) for ten years time intervals

	1960-1970	1970-1980	1980-1990	1990-2000	2000-2010	P I*	P II	P III	P IV
Brazil	-0,146	0,109	-0,018	0,056	0,198	0,224	0,237	0,113	0,219
China	-0,197		0,035	-0,135	-0,149	0,01		0,0743	0,089
Europe	-0,309	-0,312	-0,226	-0,207	-0,026	0,45	0,504	0,372	0,413
Former USSR	0,169	0,083	-0,013	-0,066	0,1	0,525	0,549	0,203	0,48
India	0,003		0,023	0,016	0,016	-0,024		0,006	0,203
South Africa	0,19	0,131	0,016	-0,086	-0,150	0,223	0,189	0,070	-0,029
USA	-0,007	-0,075	0,087	-0,232	0,112	0,09	0,173	0,134	0,179

Source: GeoDiverCity databases on urban agglomerations > 10 000 inhab.

* P I corresponds to the correlation between the growth in 1960-1970 and in 1970-1980.

In a rather ancient paper, Vining (1974) demonstrated through simulations that the lognormal distribution was not much affected when the hypotheses of absence of temporal correlation between growth rates and between city size and growth were relaxed. It is thus possible to understand the apparent contradiction of rather good fits of the size distribution as a whole and slight violations of the hypotheses of the model. We discover here an interesting paradox: it is the general quasi-stochastic character of the growth process which explains why the structural differences that we observed on the distribution of city sizes of different countries do persist over very long periods of time, as the small but systematic deviations from that stochastic process reinforce them.

Perhaps our more interesting result is the observation that the consequences on the long run of such small deviations are translated into small towns that are "too small" and large cities that are "too large" than expected from a pure lognormal distribution (figure 5) in all world regions at any moment[3]. Another illustration of this process is provided when we compute transition matrices between size classes of cities at time intervals. A single example is given here but the GibratSim application (Cura 2016) conceived with the software R (see Inset 1) enables to produce them for all time periods and world regions of our database. A huge majority of results show a heavy diagonal with most of cities staying in the same size class for each ten years intervals whereas a few cities 'jump' to other classes, usually of larger size for large cities (above the diagonal), and a feeble proportion descending in lower classes that are almost always among the smallest towns (table 5).

Inset 1 A software for reproducible statistical analysis of urban growth : GibratSim
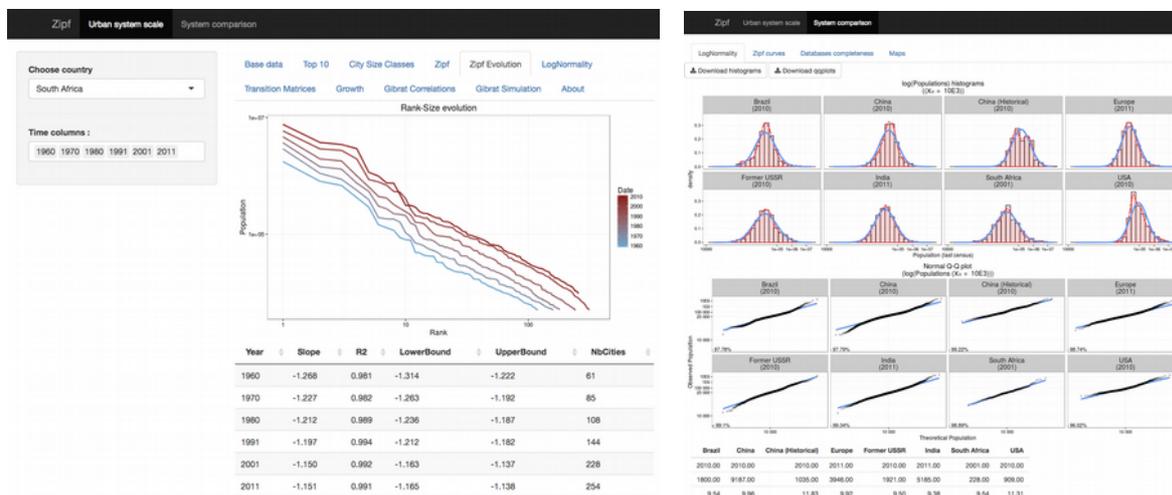
This application is a supplement to this paper, aiming at presenting the described databases and their analyses. It enables the automatic computation of all statistical analyses presented in this paper, simple tables of largest cities or numbers in size classes, adjustments to Zipf's law or a lognormal distribution, computation of transition matrices, growth rates measurements and their correlations with city size and over time, simulation of size

---

3 The USA at the last date are the only case where small cities are slightly larger than expected, due to a specially generous spatial delineation of micropolitan areas at the 2010 census, leading to a sharp decrease of the total number of urban units at that date (909 versus 940 in 2000).

distribution according to Gibrat's growth model. It is accessible online: http://gibratsim.parisgeo.cnrs.fr

---

Table 5: Transition matrix for Indian cities 2001-2011

	< 10k	10k - 50k	50k - 100k	100k - 1M	1M - 10M	> 10M
< 10k	0	0	0	0	0	0
10k - 50k	0	3596	160	5	0	0
50k - 100k	0	3	399	88	0	0
100k - 1M	0	0	3	396	13	0
1M - 10M	0	0	0	0	34	0
> 10M	0	0	0	0	0	3

Source: GeoDiverCity databases on urban agglomerations > 10 000 inhab.

This is another way of revealing a dynamics towards an overconcentration in urban systems which could be predicted from other theories in geography, when combining the hierarchical diffusion of innovations (Hägerstrand, 1952) with the time-space reorganization (Janelle, 1969). Over time, large cities become larger because they capture innovation at an early stage when it provides higher returns, whereas adoption occurs later in medium cities and small towns, inducing on the long run a slight but positive correlation with city sizes in already established urban systems (Pumain et al., 2009). Meanwhile, improvements to the speed and capacity of transportation systems (which may as well be considered as an innovation) enlarge the sphere of influence of large cities whereas smaller towns tend to be bypassed and loose parts of their centrality. Both selective processes, from the top and from below, thus ensure an increasing hierarchical differentiation in urban systems over time.

Another feature of urban systems has to be considered when applying the Gibrat's model to historical evolution: they are open systems in which "new" cities enter when the population of settlements reaches the urban threshold used for urban definition or leave when it falls below it[4]. During the 1950-2010 period of observation, as urban growth is positive, the number of entries is always much higher than the rare cases of disappearance of cities from the sample. We have estimated the shares that these "new cities" represent in the number and population of each urban system (Appendix 5) which may be rather high in periods of rapid urban growth (about 20% in number and stabilizing from 8% to 1% in India for instance). On much longer periods of observation historians have noticed alternating periods of growth in existing cities and of cities creation that may result in more or less steep slopes in the city size distribution (de Vries, 1984).

2.2 Using Gibrat's model for simulating urban growth

In order to better assess the consequences of these observed slight deviations to Gibrat's stochastic model on city size distributions, we simulated the evolution of city sizes under a

---

4 besides the rarer cases of new town creations and accidental destructions.

stochastic distribution of growth rates following its hypothesis and computed the differences between the observed final population of cities in our data set and an average of ten simulations, using the observed mean growth rates and standard deviation of each urban system for each period and applying them to the initial distribution of city sizes. The plots comparing the observed and simulated final populations for each region are represented on figure 6.

Figure 6 Comparing observed final city size and simulated with Gibrat's model
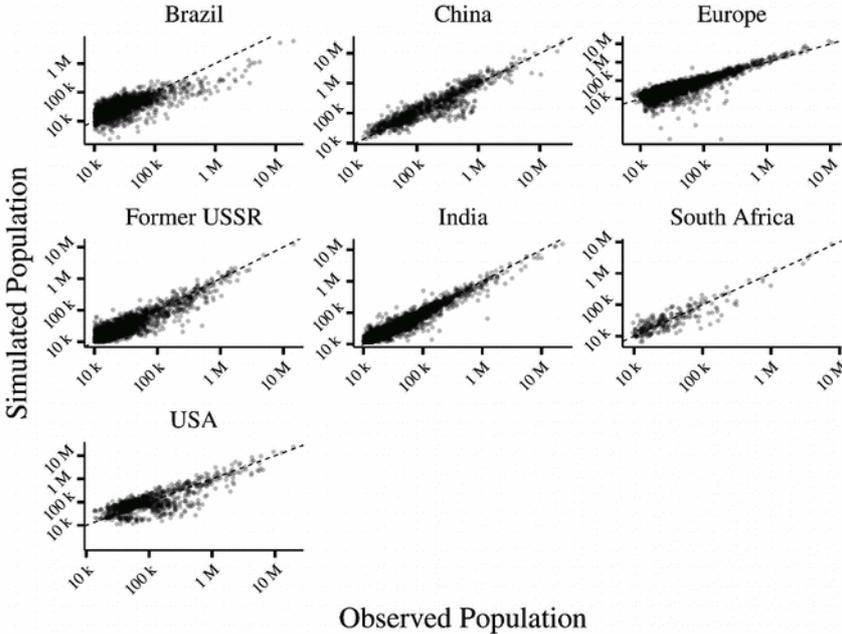
Source: GeoDiverCity data bases on urban agglomerations > 10 000 inhab.

The online application enables as well to compare the evolution of the mean and standard deviation of population sizes over time. In most cases, the observed values are slightly higher than the simulated ones. This indicates a trend to a wider variation in the actual distribution of urban growth compared to a stochastic model. China is closer to the stochastic model, which may seem paradoxical because of the amplitude of political control on the urbanization process, but may as well reveal a deliberate policy for instituting more equity in this distribution. Observed values are lower for Europe during the period (probably because of the already mentioned artifact in data delineation) and in the USA at the end of the period.

Conclusion

In this paper we have used harmonized data bases that are consistent with an evolutionary concept of a "city" over time and that enable systematic international comparisons. A first

conclusion is that our results do confirm the quasi relevance of Zipf's law and Gibrat's urban growth models. From that study on large countries of different regions of the world we can conclude that both models reasonably fit to most observations. Rather than deciding which of them could represent a universal urban dynamics, we think of them as helpful tools to specify different types of urban size distribution and their evolution because the slight departures from these models adequately describe nuances that can be meaningfully interpreted from features of geographical and historical contexts.

Despite all the imperfections of our data bases we think that our results are reliable because they are based on large data sets of the best possible quality: we have applied a consistent evolutionary perspective on urban systems at both scales. But quality of data certainly matters and more tests have to be run and can be done in a reproducible way thanks to open softwares as in Cottineau (2016) or Cura (inset 1). It is time to develop an open science of cities by sharing the effort of completing large databases and reproducing statistical and modeling experiments on them with collectively improved computation and visualization tools.

To summarize, from that study interested in documenting the diversity of urban systems from a geographical point of view, it results that Zipf's law is a good tool for visualization and for comparison of the degree of size inequalities between cities in different countries and over time, using the slope of the adjusted trend line as an index. Regarding the fit of this distribution to the upper part of the hierarchy, there are frequent cases of "urban primacy", especially in the frame of small or medium-sized or formerly colonized national states, whereas in large and less integrated urban systems the population of largest cities may be less than expected by the model. On the lower part of the distribution, this model is not so well adapted and can be replaced by adjusting a lognormal distribution to the whole settlement system. The lognormal distribution receives a more direct explanation with the Gibrat's model of urban growth. With the deviations from that model, we can then identify the specific stages in the historical evolution of urban systems.

A possible objection to our method is related to the territorial extent of national states which is not well adapted to the real urban interactional field. To be in agreement with urban theory in future testing experiments, the systems of cities should be seen in a geo-historical perspective according to the design of networks that shape their development.

However many regulation processes are still operating inside the national boundaries as they have done on sometimes very long time periods, generating more internal interactions than external for most of the cities. Our results confirm that beyond the general common processes in urban dynamics that are explained by competitive interactions within global urbanization trends, there are still relevant geographical specificities in the configuration and evolution of systems of cities that can be widely explained by well identified differences in settlement history and politico-administrative organization of national territories.

References:

Baek S.K., Beernharsson S., Minnhagen P., (2011), Zipf's law unzipped. *ArXiv* 1104.-1789v1.

Barbut M., (2004), A family of distributions: from Paretian to « contra-Paretian ». Applications to the study of urban concentration and its evolution. *Cybergeo, European Journal of Geography*, 266.

Black, D., Henderson V., (2003), Urban evolution in the USA, *Journal of Economic Geography* 3 (4): 321-327.

Bosker M., Brakman S., Garretsen H., Schramm M., (2008), A century of shocks: The evolution of the German city size distribution 1925-1999. *Regional Science and Urban Economics,* 38, 330-347.

Brakman S., Garretsen H., Schramm S., (2004), The strategic bombing on German cities during World War II and its impact on city growth. *Journal of Economic Geography* 4 (2), 201-218.

Bretagnolle A., Mathian H., Pumain D., Rozenblat C., (2000), Long-term dynamics of European towns and cities: towards a spatial model of urban growth. *Cybergeo, European Journal of Geography,* 131, 17 p.

Bretagnolle A., Giraud T., Mathian H., (2008), La mesure de l'urbanisation aux Etats-Unis, des premiers comptoirs coloniaux aux *Metropolitan Areas* (1790-2000), *Cybergeo, European Journal of Geography*, 427., DOI : 10.4000/cybergeo.19683

Bretagnolle A., Delisle F., Mathian H., Vatin G., (2015), Urbanization of the United States over two centuries: an approach based on a longterm database (1790–2010), *International Journal of Geographical Information Science*, DOI:10.1080/13658816.2014.999681, 20 p.

Clauset A., Rohilla Shalizi C., & Newman M.E.J., (2009), Power-law distributions in empirical data. *SIAM review* 51 (4), 661-703.

Corominas-Murtra B., Solé R.V., (2010), Universality for Zip's law. *Physical Review,* E, 82, 1, n° 011102.

Cottineau C., (2014), *L'évolution des villes dans l'espace post-soviétique; observations et modélisation*. Université Paris 1, thèse de doctorat.

Cottineau, C., Reuillon, R., Chapron, P., Rey-Coyrehourcq, S., & Pumain, D. (2015). A modular modelling framework for hypotheses testing in the simulation of urbanisation. *Systems*, *3*(4), 348-377.

Cottineau C. (2016), MetaZipf. (Re)producing knowledge about city size distributions. ArXiv.

Cura R. (2016), GibratSim : Population growth analysis and simulation v1.0. Zenodo. DOI:10.5281/zenodo.60495

Dobkins L.H.., Ioannides Yannis M. (2000), Dynamic evolution of the U.S. city size distribution, in : Huriot, Thisse (eds), *The Economics of Cities*. Cambridge University Press, Cambridge, 217-260.

Eeckhout, J., (2004), Gibrat's Law for (All) Cities, *The American Economic Review* 94(5), 1429-1451.

Favaro J.-M. Pumain D., (2011), Gibrat Revisited: An Urban Growth Model including Spatial Interaction and Innovation Cycles. *Geographical Analysis*, 43, 3, 261-286.

Gabaix X. (1999), Zipf's Law for Cities: An Explanation. *Quarterly Journal of Economics,* 1143, 739-767.

Gabaix X., Ioannides Y. M., (2003), The evolution of city size distributions. In J. V.Henderson, J. F. Thisse Eds.. *Handbook of regional and urban economics* Vol.4. Amsterdam: North-Holland Publishing Company.

Gillespie C. S., (2015), Fitting heavy tailed distributions : the poweRlaw package, *Journal of Statistical Software*, Vol. 64, Issue 2

Gonzalez-Val R., Lanaspa L. and Sanz-Gracia F., (2014), New evidence on Gibrat's law for cities. *Urban Studies*, 51 (1), 93-115.

Hägerstrand, T. (1952), The propagation of innovation waves. *Lund studies in geography: Series B, Human geography,* 4.

Hernando A. Hernando R., Plastino A., Zambrano E., (2014), Memory endowed US cities and their demographic interactions. *J. R. Soc. Interface* 12: 20141185. http://dx.doi.org/10.1098/rsif.2014.1185

Ignazzi C.A., (2015), *Coevolution in the Brazilian system of cities*, thèse de Doctorat, Université Paris 1, Panthéon-Sorbonne, Paris, France.

Janelle D. G., (1969), Spatial reorganisation : a model and concept. *Annals of the Association of American Geographers,* 343-368.

Kadi A.S., Halingali.I., (2010), Growth pattern of cities in India: Application of statistical models. *Canadian Studies in Population*, vol. 37, 1-2, 125-150.

Kuhnert C.D., Helbing D., West G.B., (2006), Scaling laws in Urban Supply Networks. *Physics A, Statistical Mechanics and applications,* 263 1, 96-103.

Mitzenmacher M.,(2004), A brief history of generative models for power law and lognormal distributions. *Internet mathematics* 1 (2): 226-51.

Moriconi-Ebrard F., (1993), *L'urbanisation du monde*. Paris, Anthropos-Economica.

Moriconi-Ebrard F., (1994), Geopolis, pour comparer les villes du monde. Paris, Anthropos-Economica.

Newman M.E.J., (2004), Power laws, Pareto distributions and Zif's law. *ArXiv:* cond-mat/0412004v3.

Nitsch V., (2005), Zipf zipped. *Journal of Urban Economics*, 57, 86-100.

Parr, J.B., and K. Suzuki., (1973), Settlement Populations and the Lognormal Distribution. *Urban Studies* 10(3), 335-352.

Paulus F., (2004), *Coévolution* dans les systèmes de *villes : croissance et spécialisation des aires urbaines françaises de 1950 à 2000.* Université Paris 1, thèse de doctorat.

Pred A., (1977), *City systems in advanced societies*. London, Hutchison.

Pumain D. Moriconi-Ebrard F., (1997), City Size distributions and metropolisation. *Geojournal*, 43, 4, 307-314.

Pumain, D., (2000), Settlement systems in the evolution. *Geografiska Annaler*, Series B: Human Geography, 82B (2), 73–87. doi:10.1111/1468-0467.00075

Pumain D. ,(2006), Alternative explanations of hierarchical differentiation in urban systems, in Pumain D. (ed.) *Hierarchy in natural and social sciences*, Springer, Methodos series 3, 169-222.

Pumain D., Paulus F., Vacchiani-Marcuzzo C., (2009), Innovation Cycles and Urban Dynamics, in : D.Lane, D. Pumain, S. Van der Leeuw, G. West (eds.), *Complexity perspectives on innovation and social change*, ISCOM, Springer, Methodos Series, Berlin, 237-260.

Pumain D., Swerts E., Cottineau C. Vacchiani-Marcuzzo C., Ignazzi A., Bretagnolle A., Delisle F., Cura R., Lizzi L, Baffi S.,(2015), Multi-level comparison of large urban systems. *Cybergeo European Journal of Geography,* 706, http://cybergeo.revues.org/26730 ; DOI : 10.4000/cybergeo.26730.

Quandt R. E., (1964), Statistical discrimination among alternate hypothesis and some economic regularities. *Journal of Regional Science,* 52, 1-23.

Resende M., (2004), Gibrat's Law and the Growth of Cities in Brazil: A Panel Data Investigation.*Urban Studies,* 418, 1537-1549.

Robson B. T., (1973), *Urban Growth, an approach*. London: Methuen.

Rozenfeld, H., Rybski, D., Andrade, J., Batty, M., Stanley, E., Makse, H., (2008), Laws of population growth. *Proc. Natl. Acad. Sci.* 105, 18702–18707.

Schaffar A. Pavleas, (2014), The evolution of the Greek urban centers 1951-2011. *Région et Développement*, 39, 87-104.

Schmidheiny K. Suedekum J., (2015), The pan-European population distribution across consistently defined functional urban areas. *Economic letters*, 133, 10-13.

Swerts E., (2013), Les systèmes de villes en Inde et en Chine. Université Paris 1, thèse de doctorat.

Tomita S., Hayashi Y., (2008), A controllable model of a random multiplicative process for the entire distribution of population. *Physica A*, 387, 5-6, 1345-1351.

Vacchiani-Marcuzzo, C. 2005, *Mondialisation et systèmes de villes : les entreprises étrangères et l'évolution des agglomérations sud-africaines*, Thèse de doctorat, Université Paris 1, 360 p. http://tel.ccsd.cnrs.fr/docs/00/05/44/70/PDF/these.pdf

Vining D., (1974), On the sources of instability in the rank size rule: some simple tests of Gibrat's law. *Geographical Analysis,* 6,4, 313-329.

de Vries J., (1984), *European Urbanization*, 1500-1800. Cambridge, Mass., Harvard University Press.

Zipf C.K., (1941), *National unity and disunity*. Bloomington (Indiana), Principia Press.

Zipf, C.K., (1949). *Human behaviour and the principle of least effort*. Cambridge, MA: Addison Wesley Press.

APPENDIX 1 – Zipf curves parameters

Brazil

Year	Slope	R2	LowerBound	UpperBound	NbCities
1960	-0.664	0.972	-0.669	-0.658	1449
1970	-0.723	0.984	-0.728	-0.718	1501
1980	-0.810	0.984	-0.815	-0.805	1593
1991	-0.861	0.990	-0.866	-0.857	1681
2000	-0.913	0.991	-0.917	-0.909	1731
2010	-0.959	0.992	-0.963	-0.955	1800

China

Year	Slope	R2	LowerBound	UpperBound	NbCities
1964	-1.215	0.942	-1.236	-1.195	832
1982	-1.219	0.935	-1.239	-1.198	981
1990	-1.250	0.925	-1.272	-1.228	1003
2000	-0.877	0.976	-0.881	-0.874	6239
2010	-0.949	0.978	-0.952	-0.946	9187

Europe

Year	Slope	R2	LowerBound	UpperBound	NbCities
1961	-0.945	0.996	-0.947	-0.943	3248
1971	-0.954	0.994	-0.957	-0.952	3548
1981	-0.951	0.992	-0.954	-0.948	3775
1991	-0.948	0.991	-0.950	-0.945	3877
2001	-0.940	0.991	-0.942	-0.937	3953
2011	-0.942	0.992	-0.944	-0.939	3946

Former Soviet Union

Year	Slope	R2	LowerBound	UpperBound	NbCities
1959	-1.035	0.989	-1.041	-1.028	1145
1970	-1.079	0.986	-1.086	-1.072	1378
1979	-1.092	0.986	-1.098	-1.085	1595
1989	-1.089	0.986	-1.095	-1.083	1822
2002	-1.089	0.987	-1.095	-1.083	1883
2010	-1.104	0.987	-1.109	-1.098	1921

India

Year	Slope	R2	LowerBound	UpperBound	NbCities
1961	-0.931	0.999	-0.933	-0.930	2095
1981	-0.940	0.999	-0.941	-0.939	3402
1991	-0.945	0.999	-0.946	-0.944	3823
2001	-0.957	0.999	-0.957	-0.956	4699
2011	-0.970	0.999	-0.971	-0.969	5185

South Africa

Year	Slope	R2	LowerBound	UpperBound	NbCities
1960	-1.268	0.981	-1.314	-1.222	61
1970	-1.227	0.982	-1.263	-1.192	85
1980	-1.212	0.989	-1.236	-1.187	108
1991	-1.197	0.994	-1.212	-1.182	144
2001	-1.150	0.992	-1.163	-1.137	228
2011	-1.151	0.991	-1.165	-1.138	254

USA :

Year	Slope	R2	LowerBound	UpperBound	NbCities
1960	-1.176	0.976	-1.189	-1.164	829
1970	-1.209	0.973	-1.222	-1.196	862
1980	-1.212	0.960	-1.228	-1.195	864
1990	-1.238	0.967	-1.253	-1.223	856
2000	-1.212	0.969	-1.226	-1.198	930
2010	-1.237	0.974	-1.250	-1.223	909

APPENDIX 2 – Test of adjustment to lognormal distribution (2010)

Country	Nb cities	Meanlog	Stlog	pSh-W	pK-S	skewness	kurtosis
Brazil	1800	9.54	1.61	0.00	0.00	0.01	4.40
China	9184	9.96	1.51	0.00	0.00	-0.29	4.89
Europe	3946	9.92	1.37	0.00	0.00	0.36	4.08
FSU	1921	9.50	1.92	0.00	0.00	-0.21	3.82
India	5185	9.38	1.66	0.00	0.00	0.13	3.95
South Africa	254	9.63	1.82	0.00	0.00	-0.08	3.76
USA	909	11.31	1.39	0.00	0.00	0.80	3.71

pSh-W: p value for Shapiro-Wilk test; pK-S: p value for Kolmogoroff-Smirnoff test

Source: GeoDiverCity data bases on urban agglomerations > 10 000 inhab.

APPENDIX 3 – Correlation between Size and Growth rate

Brazil

Period	Correlation	LogCorrelation	NbCities
1960-1970	0.043	-0.146	1800
1970-1980	0.166	0.109	1800
1980-1991	0.039	-0.018	1800
1991-2000	0.061	0.056	1800
2000-2010	0.048	0.198	1800

China

Period	Correlation	LogCorrelation	NbCities
1964-1982	-0.129	-0.197	832
1982-1990	-0.010	0.035	1019
1990-2000	-0.061	-0.135	1019
2000-2010	-0.008	-0.149	6313

Europe

Period	Correlation	LogCorrelation	NbCities
1961-1971	-0.052	-0.309	3982
1971-1981	-0.086	-0.312	3975
1981-1991	-0.066	-0.226	3969
1991-2001	-0.046	-0.207	3966
2001-2011	0.010	-0.026	3962

Former Soviet Union

Period	Correlation	LogCorrelation	NbCities
1959-1970	0.047	0.169	1145
1970-1979	0.026	0.083	1380
1979-1989	-0.018	-0.013	1596
1989-2002	-0.007	-0.066	1822
2002-2010	0.062	0.100	1883

India

Period	Correlation	LogCorrelation	NbCities
1961-1981	0.011	0.003	2095
1981-1991	0.053	0.023	3402
1991-2001	0.024	0.016	3824
2001-2011	0.025	0.016	4700

South Africa

Period	Correlation	LogCorrelation	NbCities
1960-1970	0.129	0.190	120
1970-1980	0.011	0.131	160
1980-1991	-0.030	0.016	180
1991-2001	-0.054	-0.086	243
2001-2011	-0.000	-0.150	314

USA

Period	Correlation	LogCorrelation	NbCities
1960-1970	0.003	-0.007	981
1970-1980	-0.070	-0.075	966
1980-1990	0.103	0.087	935

Period	Correlation	LogCorrelation	NbCities
1990-2000	-0.055	-0.232	913
2000-2010	0.043	0.112	940

APPENDIX 4 – Temporal auto-correlation

Brazil

Period	Label	Correlation	NbCities
1960-1970 -> 1970-1980	I	0.224	1800
1970-1980 -> 1980-1991	II	0.237	1800
1980-1991 -> 1991-2000	III	0.113	1800
1991-2000 -> 2000-2010	IV	0.219	1800

China

Period	Label	Correlation	NbCities
1964-1982 -> 1982-1990	I	0.010	832
1982-1990 -> 1990-2000	II	0.074	1019
1990-2000 -> 2000-2010	III	0.089	1019

Europe

Period	Label	Correlation	NbCities
1961-1971 -> 1971-1981	I	0.450	3969
1971-1981 -> 1981-1991	II	0.504	3966
1981-1991 -> 1991-2001	III	0.372	3962
1991-2001 -> 2001-2011	IV	0.413	3962

Former Soviet Union

Period	Label	Correlation	NbCities
1959-1970 -> 1970-1979	I	0.525	1145
1970-1979 -> 1979-1989	II	0.549	1380
1979-1989 -> 1989-2002	III	0.203	1596
1989-2002 -> 2002-2010	IV	0.480	1822

India

Period	Label	Correlation	NbCities
1961-1981 -> 1981-1991	I	-0.024	2091
1981-1991 -> 1991-2001	II	0.006	3391
1991-2001 -> 2001-2011	III	0.203	3815

South Africa

Period	Label	Correlation	NbCities
1960-1970 -> 1970-1980	I	0.223	114
1970-1980 -> 1980-1991	II	0.189	146
1980-1991 -> 1991-2001	III	0.070	174
1991-2001 -> 2001-2011	IV	-0.029	209

USA

Period	Label	Correlation	NbCities
1960-1970 -> 1970-1980	I	0.090	875
1970-1980 -> 1980-1990	II	0.173	853
1980-1990 -> 1990-2000	III	0.134	811
1990-2000 -> 2000-2010	IV	0.179	785

APPENDIX 4-5 - Share in number and population of cities entering the system

Brazil

	1960-1970	1970-1980	1980-1991	1991-2000	2000-2010
New cities (absolute)	114	112	107	68	69
New cities (relative)	8%	7%	6%	4%	4%
New pop. (abs.)	1 424 273	1 374 623	1 230 326	761 013	739 305
New pop. (rel.)	2%	1%	1%	1%	0%

China

	1964-1982	1982-1990	1990-2000	2000-2010
New cities (absolute)	147	23	29	4
New cities (relative)	15%	2%	3%	0%
New pop. (abs.)	8 209 298	261 619	4 810 578	57 616
New pop. (rel.)	5%	0%	1%	0%

Europe

	1961-1971	1971-1981	1981-1991	1991-2001	2001-2011
New cities (absolute)	315	235	114	86	2
New cities (relative)	9%	6%	3%	2%	0%
New pop. (abs.)	3 861 014	2 968 834	1 435 101	1 038 216	28 396
New pop. (rel.)	1%	1%	0%	0%	0%

Former Soviet Union

	1959-1970	1970-1979	1979-1989	1989-2002	2002-2010
New cities (absolute)	233	217	227	61	40
New cities (relative)	17%	14%	12%	3%	2%
New pop. (abs.)	3 448 867	4 204 643	3 799 610	1 073 278	456 325
New pop. (rel.)	3%	3%	2%	1%	0%

India

	1961-1981	1981-1991	1991-2001	2001-2011
New cities (absolute)	1307	426	885	486
New cities (relative)	38%	11%	19%	9%
New pop. (abs.)	17 415 779	5 048 660	12 647 561	5 544 504
New pop. (rel.)	8%	2%	4%	1%

South Africa

	1960-1970	1970-1980	1980-1991	1991-2001	2001-2011
New cities (absolute)	25	26	39	89	35
New cities (relative)	29%	24%	27%	39%	14%
New pop. (abs.)	318 103	428 267	608 283	1 548 782	450 991
New pop. (rel.)	4%	4%	4%	6%	1%

USA

	1960-1970	1970-1980	1980-1990	1990-2000	2000-2010
New cities (absolute)	71	45	39	124	44
New cities (relative)	8%	5%	5%	13%	5%
New pop. (abs.)	1 686 039	1 276 470	1 224 370	5 406 694	2 657 350
New pop. (rel.)	1%	1%	1%	2%	1%