



CrossMark
click for updates

Research

Cite this article: Hernando A, Hernando R, Plastino A, Zambrano E. 2015 Memory-endowed US cities and their demographic interactions. *J. R. Soc. Interface* **12**: 20141185. <http://dx.doi.org/10.1098/rsif.2014.1185>

Received: 29 October 2014

Accepted: 13 November 2014

Subject Areas:

environmental science, mathematical physics

Keywords:

urban growth, social dynamics, space–time correlations

Author for correspondence:

A. Hernando

e-mail: alberto.hernandodecastro@epfl.ch

Memory-endowed US cities and their demographic interactions

A. Hernando¹, R. Hernando², A. Plastino³ and E. Zambrano¹

¹Laboratory of Theoretical Physical Chemistry, Institut des Sciences et Ingénierie Chimiques, École Polytechnique Fédérale de Lausanne, Lausanne 1015, Switzerland

²Social Thermodynamics Applied Research (SThAR), Madrid, Spain

³National University La Plata, Physics Institute, (IFLP-CCT-CONICET) La Plata, Argentina and Physics Department, Universitat de les Illes Balears, Palma de Mallorca, Spain

A quantitative understanding of cities' demographic dynamics is becoming a potentially useful tool for planning sustainable growth. The concomitant theory should reveal details of the cities' past and also of its interaction with nearby urban conglomerates for providing a reasonably complete picture. Using the exhaustive database of the Census Bureau in a time window of 170 years, we exhibit here empirical evidence for time and space correlations in the demographic dynamics of US counties, with a characteristic memory time of 25 years and typical distances of interaction of 200 km. These correlations are much larger than those observed in a European country (Spain), indicating more coherent evolution in US cities. We also measure the resilience of US cities to historical events, finding a demographical *post-traumatic amnesia* after wars (such as the American Civil War) or economic crisis (such as the 1929 Stock Market Crash).

1. Introduction

One half of the human population lives in urban areas [1]. Asking whether the present population's growth rates are economically and ecologically sustainable is a recurrent question [2] that justifies efforts directed towards the development of quantitative unified theories of urban living [3]. A countless number of degrees of freedom is involved in a city's evolution, that is, a host of individual contributions, involving millions of people, acting on their own free will. Devising a unified theory constitutes a formidable challenge. However, despite this intrinsic difficulty, many advances have been made in recent years. During the twentieth century, many regularities were reported, such as Zipf's law in the city population rank distribution [4–8], or the celebrated Gibrat's law of proportional growth applicable to cities [9–14]. In addition, empirical data show that the scaling with population of internal degrees of freedom in an urban body such as (i) the structure of road networks or urban sprawl patterns [10,15,16] and (ii) metrics such as wages or crime rates [17,18], do follow predictable tendencies that can be mathematically described. In addition to a city's growth, electoral processes and many other social phenomena have been successfully modelled as well [6,8,19–23]. Also, collective modes emerge when cities are considered as entities that evolve and interact in a coherent fashion. The interaction between cities [14] (as measured by, for instance, the number of crossed phone calls [24] or human mobility [25]) displays predictable characteristics. Indeed, an analogy between the evolution of the population of an ensemble of cities and the random movement of particles in a fluid was made by conjoining Gibrat's law of proportional growth with Brownian motion. In such an approach, the size X at time t of the i th geometrical Brownian walker follows the dynamical equation [9–13,26]

$$\dot{X}_i(t) = X_i(t)v_i(t), \quad (1.1)$$

where the dot denotes a time derivative and $v_i(t)$ is the growth rate, which is described as a Wiener coefficient with covariance $\text{Cov}[v_i(t), v_j(t')] = \sigma_v^2 \delta_{ij} \delta(t - t')$ — δ standing for the delta function and σ_v for the standard deviation of

the growth rates—i.e. uncorrelated and memoryless dynamics. Considering a new variable defined as $u_i(t) = \ln X_i(t)$ [26], we recover all the properties of the physical ideal gas in what one may call the scale-free ideal gas [13,27]. Exhaustive empirical observations of the dynamics of Spain's population demonstrate that this analogy can be used to formulate a thermodynamics of population flows [28]. Moreover, we have recently shown that this assumption of uncorrelated evolution fails for cities that are close neighbours [14]. Additionally, the evolution of cities' populations exhibits memory, indicating that we deal with a non-Markovian process. These results are indicative of (i) a rich and complex phenomenology underlying population flows, and (ii) that our models should go beyond the ideal gas to include pairwise interactions and inertia, as in the case of real gases in statistical physics. The presence of some kind of correlations has also been independently proposed in [4], where it is shown, via numerical experiments, that it is necessary to introduce conditioned sampling to reproduce Zipf's law, simulating what the authors call 'coherence'. These advances, both in the internal structure and in the functionality of cities, as well as in the properties of an ensemble of interacting cities, encourage the search for a unified theory.

We present in this work empirical evidence of such correlations in the population dynamics of US counties and compare our results with previous observations on the population of Spanish cities. In §2.1, we first analyse the statistical properties of cities' growth rates, confirming Gibrat's law with finite size effects for small populations. We pass next to analyse, in §2.2, the *memory* of the growth rates, using data from a time window of 170 years, which includes relevant historical events, such as the American Civil War, both world wars and the economic crisis after the 1929 Stock Market Crash. We continue, in §2.3, with the analysis of city–city correlations as functions of the distance between them, which leads us to define a characteristic correlation distance. Finally, a discussion and conclusion are given in §3.

2. Results

2.1. Stochastic properties of US growth rates

We look for quantitative space–time patterns underlying US demographics and ascertain which trends are universal and which are local by comparison with [14], related to Spanish cities. An exhaustive analysis of the US population is made using the Census Bureau database of counties' populations [29]. We have used data from 1830 to 2000 (170 years), in a time window that covers relevant historical events such as the American Civil War, both world wars, and the 1929 Stock Market Crash. More than 3000 counties (all of them with available data) are considered in our study, whose spatial distribution is depicted in figure 1.

Administrative and legal boundaries do not reflect, in all cases, an exact definition of what can be considered as a coherent urban nucleus. Indeed, many counties may include several independent population centres, or split big cities into several units. There have been many efforts to find an accurate and natural definition of city through clustering (e.g. [31,32]). Indeed, these definitions become relevant for predicting the shape of the city size rank distribution, for which Zipf's law generally emerges when natural boundaries are used instead of legal ones. However, these definitions do not become so relevant with respect to the

dynamics described by equation (1.1), as shown in [28]. As we are here interested in the dynamical patterns exhibited by populations and not in size distributions, we can safely extend our findings beyond the particular definition of city or population nucleus.

We first verified, using the counties database, the validity of Gibrat's law, including the correction for smaller populations, discovered in [13]. To this effect, we added a new term to the proportional law, a 'finite size noise' (FSN), of the form

$$\dot{X}_i(t) = X_i(t)v_i(t) + \sqrt{X_i(t)}w_i(t), \quad (2.1)$$

where $w_i(t)$ is an independent Wiener coefficient. This term is a direct consequence of the Central Limit Theorem, as shown in [13,28], due to the independent nature of the $w_i(t)$. The variation of the population X_i is much smaller than the variation of the growth rates. Thus, for the later, the standard deviation $\sigma_{\dot{x}}$, in our time windows, becomes

$$\sigma_{\dot{x}}(X_i) \cong X_i\sigma_v + \sqrt{X_i}\sigma_w, \quad (2.2)$$

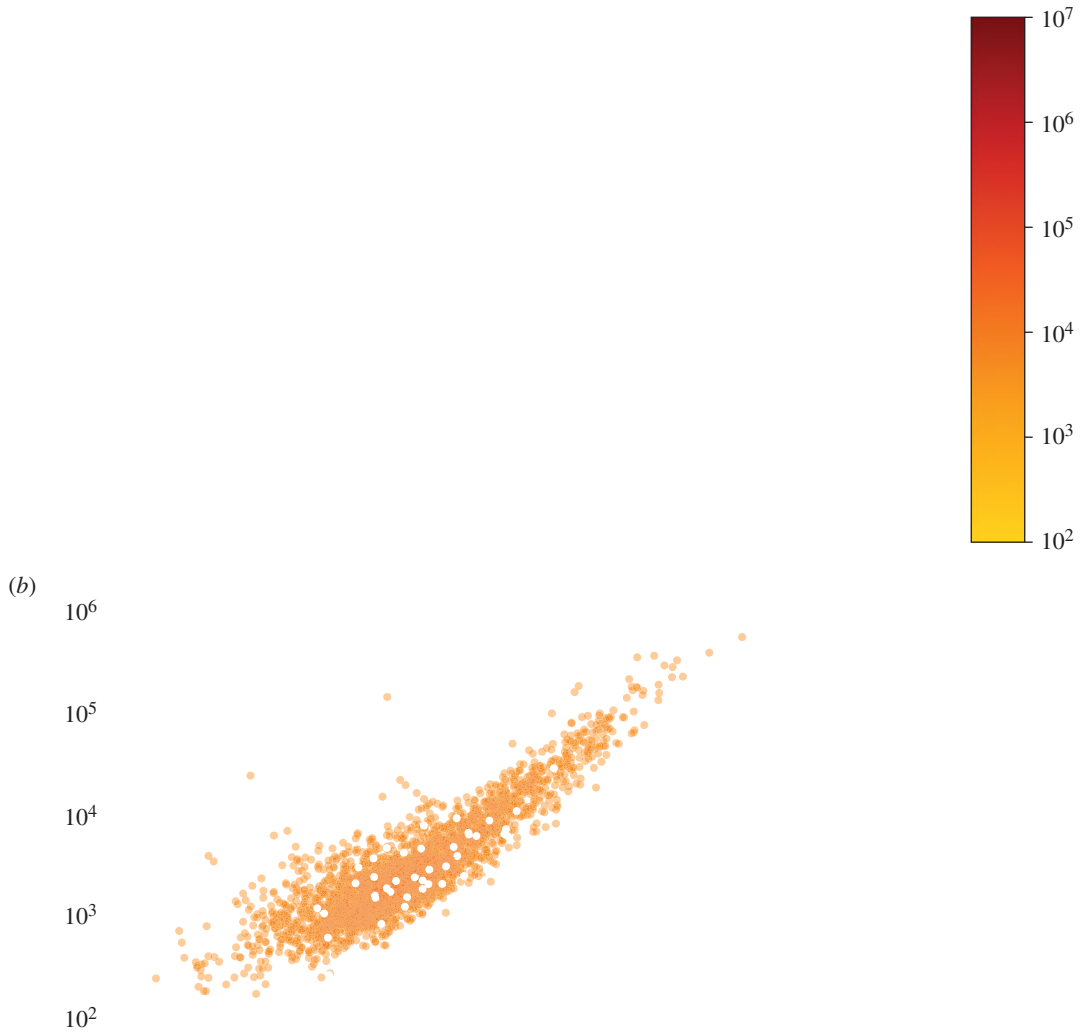
where σ_v and σ_w are the deviations of the proportional and FSN, respectively.

For comparing the raw data with our predictions, we take into account that population data have a strong scale-free behaviour and are characterized by long-tailed distributions. In these circumstances, *mean values* do not converge or do so very slowly, their estimation becoming numerically unstable. We recommend working with *medians* instead, as they are numerically more stable and also invariant under several transformations—indeed, $\text{med}[\log(X)] = \log(\text{med}[X])$ but $\langle \log(X) \rangle \neq \log(\langle X \rangle)$. Following this recipe, we produced the results in figure 1, depicting the empirical $\sigma_{\dot{x}}$ versus population, for every county, in a log–log scale, and also the median of the former. Two trends are immediately observed: one for lower populations ($X < 35\,000$ inhabitants) and a second one for larger populations ($X > 35\,000$). Remarkably, a linear fit to the log–log representation—where power laws become straight lines with slope values linked to the exponents—gives for the exponents 0.508 ± 0.025 (with $\log(\sigma_w) = 2.6 \pm 0.2$) for the first trend and 1.04 ± 0.02 (with $\log(\sigma_v) = -3.1 \pm 0.2$) for the second one, with a coefficient of determination R^2 of 0.96 and 0.994, respectively. As these values almost coincide with the expected ones (1/2 for FSN at lower population and 1 for proportional grow for larger population), we can regard the Gibrat plus FSN law as verified, a rather significant observation. We find, however, that the exponent for proportional growth is slightly larger than 1 according to the confidence interval. In [13], similar cases of exponents larger than 1 are also reported. They emerge as a consequence of the massive migration from small villages to big cities. Even if this deviation from Gibrat's law is mild, it is persistent for our fitting-procedure. Thus, we expect it to also be the signature of fast-growing urban population in US demographics.

2.2. Measuring memory

To check whether US counties exhibit memory effects, we appeal to the Pearson's product–moment correlation coefficient $c_y(t) = \text{Corr}[\dot{X}(y), \dot{X}(t)]$, using as samples the list of counties' growth rates for (i) the year y and (ii) any precedent time t for which data are available. We find that the averaged time correlation over $n_y = 12$ samples (from 1890 to 2000),

(a)



defined as $\langle c(\Delta t) \rangle = n_y^{-1} \sum_y c_y(y + \Delta t)$ and calculated for consecutive census instantiations ($\Delta t = 10$ years), exhibits a behaviour similar to that found for Spanish cities [14]: large cities show greater inertia. We can attribute the smaller counties' loss of memory to the FSN term, which becomes important for them (figure 2a). For larger intervals of time Δt , we find a clear decay of the averaged time correlation. Remarkably, the correlations are much larger than those found for Spanish cities (figure 2b). Considering only the first 40 years, a fit to an exponential decay (figure 2b, inset)—in a similar fashion as that effected for Spanish cities—gives us a characteristic time of 25 ± 7 years ($R^2 = 0.990$), but surprisingly, the correlation eventually becomes negative after approximately 60 years.

In order to gain a deeper understanding of this unexpected trend, and also to check whether it is caused by a non-homogeneous behaviour of the correlations, we have independently studied all the contributions $c_y(t)$ for several years y (and precedent times t (figure 3)). We find that, although for all y a decay of the correlation with time is always present, historical events clearly modulate these

correlations. For the growth rates from 1890 to 1950, we find that, irrespective of the year, no memory remains in the demographics of US counties from the years that precede the American Civil War, in a kind of 'post-traumatic amnesia' (figure 3a). For the second half of the twentieth century, we find, in general, a slower decay—larger memory—than for other time periods. The most important historical event that one immediately detects (by simple inspection), regarding cities' memory, is the economic crisis after the 1929 Stock Market Crash. Again, irrespective of the year, one still encounters (i) a correlation's fall and (ii) loss of memory regarding precedent decades (figure 3b). Thus, instead of a homogeneous year-independent decay of the correlations with time, we find a decay with a strong dependence on historical events.

However, some unanswered questions remain, such as (i) the reason for the strong fluctuations between the years 1990 and 2000 or (ii) why do the years 1940 and 1950 seem to be not strongly afflicted by amnesia? Regarding the latter question, figure 3 indicates that for small time intervals, inertia is more important than historical events. Only after a few decades, the amnesia becomes apparent. Whether this period

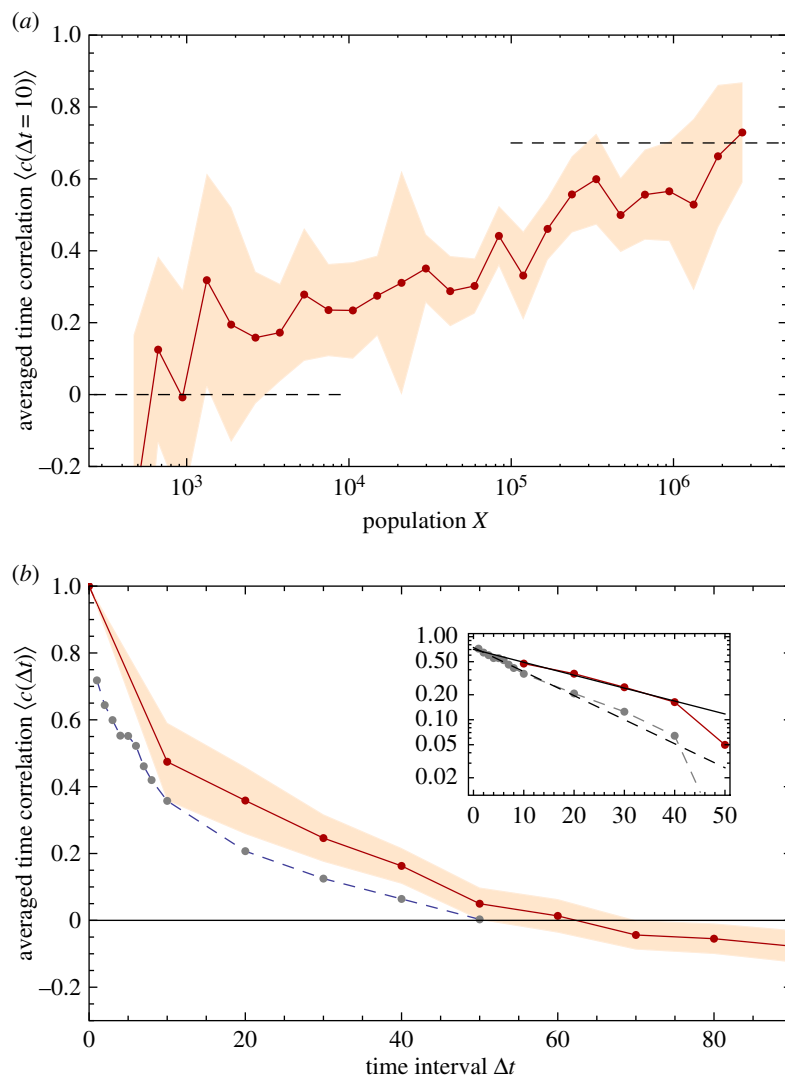


Figure 2. Averaged time correlations in population growth. (a) Averaged time correlations after 10 years $\langle c(\Delta t = 10) \rangle$ versus town sizes X (dot-line). The shaded area represents the width determined by 1 s.d. Horizontal dashed lines are asymptotic levels to follow the decrease of the correlation at lower populations. (b) Averaged time correlation $\langle c(\Delta t) \rangle$ for the relative growth of towns populated by more than 10 000 inhabitants. Shaded areas represent widths determined by 1 s.d. The time correlation for Spanish cities is also shown for comparison's sake (grey dashed line and dots). Inset: same data in semi-log scale, compared with the exponential fits (straight lines).

of inertia is related to the characteristic memory time of approximately 25 years, is something that cannot be tested with the present data, but it is a reasonable hypothesis for future research. If such were the case, two mechanisms would take place defining short- and long-term memories, only the latter being affected by historical events. Deeper insights into how these inertial mechanisms work would indeed shed some light on how information is stored in collective social systems. Accordingly, more research along this line remains to be undertaken.

2.3. Measuring interactions

We consider now spatial correlations. The pairwise Pearson product–moment correlation coefficient $C_{ij} = \text{Corr}[\dot{X}_i, \dot{X}_j]$ of the i th and j th counties is obtained using as samples the evolution of the growth rates of each county in a given time window. We speak here of the twentieth century, from 1900 to 2000 (10 sample sets). We compare the value obtained per each pair with the distance between counties d_{ij} . The averaged value—obtained as $C(d) = \sum_{ij} C_{ij} \delta(d - d_{ij}) / \sum_{ij} \delta(d - d_{ij})$ —exhibits a clear dependence on distance, demonstrating the entanglement between US population

nuclei. The tail of the decay displays a long-range behaviour (figure 4), and the pertinent curve can be nicely fitted to an analytical expression of the form

$$C(d) = \frac{C(0)}{1 + |d/d_0|^\alpha}, \quad (2.3)$$

with $C(0) = 0.62 \pm 0.02$, $d_0 = 215 \pm 32$ km and $\alpha = 0.71 \pm 0.04$, for a R^2 coefficient of 0.997. Remarkably, the correlations are much larger than those observed for Spain [14], with a larger characteristic distance (215 km versus 80 km for Spanish cities) and a much slower decay ($\alpha = 0.71$). Note that for an inverse-square law, $\alpha = 2$.

The comparison between USA and Spain is illustrated by figure 4. Results confirm the conjecture that US cities evolve in a more coherent fashion than the cities in Spain, notwithstanding the fact that the US surface is 20 times larger than Spain's, while its population is 6.5 times larger. We may speak of an integration coherence for US cities that seems to be lacking in Europe, as has also been proposed in [4]: Zipf's law emerges when the largest US cities are considered but not when this is done on a state-by-state basis, whereas in Europe, Zipf's law emerges for each country as a whole, and

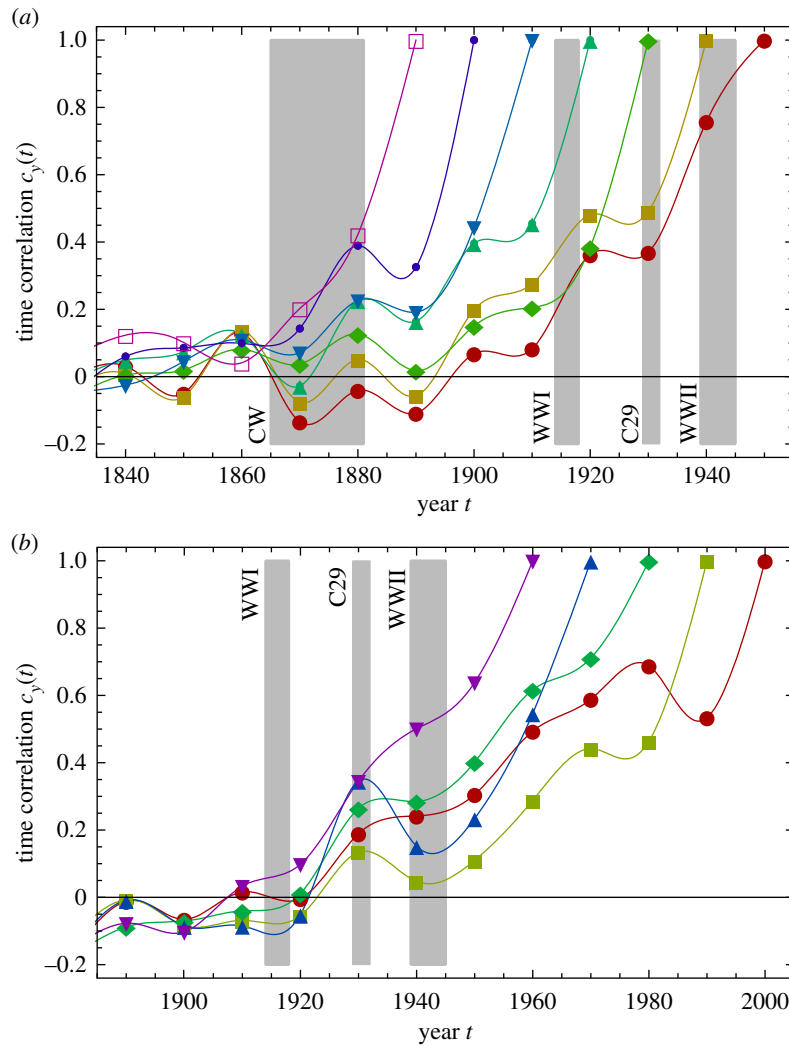


Figure 3. Time correlations in population growth $c_y(t)$ of the year y with all previous years denoted by t . Historical events as the American Civil War (CW), the First World War (WWI), the 1929 Stock Market Crash (C29) and the Second World War (WWII), are marked via shadows. (a) Correlations for years $y = 1950$ (red circles), 1940 (yellow squares), 1930 (green diamonds), 1920 (turquoise up-triangles), 1910 (blue down-triangles), 1900 (navy small circles) and 1890 (purple empty squares). In all cases, the correlation becomes almost zero after the American Civil War. (b) Correlations for the last half of twentieth century: years $y = 2000$ (red circles), 1990 (yellow squares), 1980 (green diamonds), 1970 (blue up-triangles) and 1960 (purple down-triangles). A general pattern of decay is apparent for all years, and remarkably, the five curves drop (or suffer larger decay) from 1930 to 1920 after the Stock Market Crash of 1929 (however, note that years 1950 and 1940 in (a) do not drop to zero after the crash, see text).

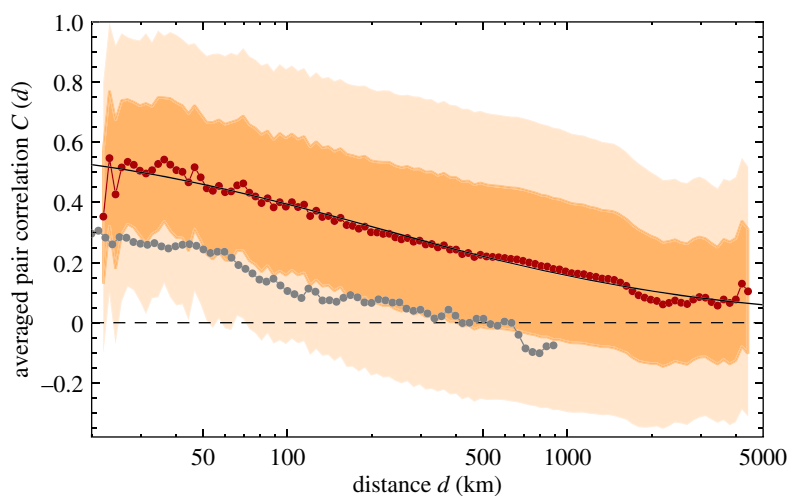


Figure 4. Pairwise correlation C versus distance between counties d . The mean value for a given distance is indicated (red line and dots) together with 1 s.d. (darker shadow) and by two of them (lighter shadow). The black line follows the analytical fit of equation (2.3). The spatial correlation for Spanish cities is also shown for comparison's sake (grey line and dots).

not when all the European continent is considered. The standard deviation observed for the US is approximately 0.4 and does not change with distance. The expected theoretical width

for a bivariate normal distribution for the same number of samples is $1/3$ [33], 20% smaller than the measured one. Thus, we gather that additional factors are involved in the US

pairwise correlation. One can attribute to the distance factor 80% of the city–city entanglement. We expect that some of these extra contributions could be associated with local factors, such as the transportation network, the particular socio-economical status of the city and/or special historical links between some population nuclei. A detailed analysis of the pairwise correlations of a selected county—instead of the coarse-grained viewpoint adopted here—when crossed with other relevant metrics, may help to gain a deeper understanding of the particular demographic and/or economic status, present and future, of a given urban area.

3. Summary and conclusion

Demographic US patterns display a rich and complex phenomenology, including both space and time correlations. US cities exhibit a strong link with their past. In an exercise of quantitative history, we have found that relevant historical events, such

as the American Civil War and the 1929 economic crisis, leave a strong imprint in the demographic dynamics, which one may call ‘post-traumatic amnesia’. Remarkably, this amnesia only takes place after a few decades, indicating the potential existence of short- and long-term memories in the social system. The mechanisms underlying this inertia are still unknown. On the other hand, the spatial correlations, much larger than those observed in Europe, indicate a high level of coherence and suggest that the evolution of any single city cannot be understood without taking the whole collective of cities into account. We feel that these empirical findings are relevant to understanding the country at a collective macroscopic level. Also, some microscopic insights are gained that may help city planners to improve their panoply of tools [29,34].

Acknowledgement. We greatly appreciate the insightful comments of the two reviewers and their help in the improvement of the paper.

Funding statement. This work was partially supported by Social Thermodynamics Applied Research (SThAR).

References

- UN-Habitat. 2010 *State of the World's cities 1010/2011—cities for all: bridging the urban divide*. See www.unhabitat.org.
- Schellnhuber HJ, Molina M, Stern N, Huber V, Kadner S. 2010 *Global sustainability: a noble cause*. Cambridge, UK: Cambridge University Press.
- Bettencourt L, West G. 2010 A unified theory of urban living. *Nature* **467**, 912–913. (doi:10.1038/467912a)
- Cristelli M, Batty M, Pietronero L. 2012 There is more than a power law in Zipf. *Sci. Rep.* **2**, 812. (doi:10.1038/srep00812)
- Zipf GK. 1949 *Human behavior and the principle of least effort*. Cambridge, MA: Addison-Wesley.
- Newman MEJ. 2005 Power laws, Pareto distributions and Zipf's law. *Contemp. Phys.* **46**, 323–353. (doi:10.1080/00107510500052444)
- Baek SK, Bernhardtsson S, Minnhagen P. 2011 Zipf's law unzipped. *New J. Phys.* **13**, 043004. (doi:10.1088/1367-2630/13/4/043004)
- Hernando A, Puigdomènech D, Villuendas D, Vesperinas C, Plastino A. 2009 Zipf's law from a Fisher variational-principle. *Phys. Lett. A* **374**, 18–21. (doi:10.1016/j.physleta.2009.10.027)
- Gibrat R. 1931 *Les Inégalités économiques*. Sirey, Paris: Librairie du Recueil.
- Rozenfeld H, Rybski D, Andrade JS, Batty M, Stanley HE, Makse HA. 2008 Laws of population growth. *Proc. Natl Acad. Sci. USA* **105**, 18 702–18 707. (doi:10.1073/pnas.0807435105)
- Gabaix X, Ioannides YM. 2004 The evolution of city size distributions. In *Handbook of regional and urban economics* (eds JV Henderson, J-F Thisse), pp. 2341–2378. Amsterdam, The Netherlands: North-Holland.
- Blank A, Solomon S. 2000 Power laws in cities population, financial markets and internet sites (scaling in systems with a variable number of components). *Phys. A* **287**, 279–288. (doi:10.1016/S0378-4371(00)00464-7)
- Hernando A, Hernando R, Plastino A, Plastino AR. 2013 The workings of the maximum entropy principle in collective human behavior. *J. R. Soc. Interface* **10**, 20120758. (doi:10.1098/rsif.2012.0758)
- Hernando A, Hernando R, Plastino A. 2013 Space–time correlations in urban sprawl. *J. R. Soc. Interface* **11**, 20130930. (doi:10.1098/rsif.2013.0930)
- Makse HA, Andrade JS, Batty M, Havlin S, Stanley HE. 1998 Modelling urban growth patterns with correlated percolation. *Phys. Rev. E* **58**, 7054–7062. (doi:10.1103/PhysRevE.58.7054)
- Masucci AP, Stalino K, Batty M. 2013 Limited urban growth: London's street network dynamics since the 18th century. *PLoS ONE* **8**, e69469. (doi:10.1371/journal.pone.0069469)
- Bettencourt LMA. 2013 The origins of scaling in cities. *Science* **340**, 1438–1441. (doi:10.1126/science.1235823)
- Bettencourt LMA, Lobo J, Helbing D, Kuehnert C, West GB. 2007 Growth, innovation, scaling, and the pace of life in cities. *Proc. Natl Acad. Sci. USA* **104**, 7301–7306. (doi:10.1073/pnas.0610172104)
- Chatterjee A, Mitrović M, Fortunato S. 2013 Universality in voting behavior: an empiric analysis. *Sci. Rep.* **3**, 1049. (doi:10.1038/srep01049)
- Axtell RL. 2001 Zipf distribution of U.S. firm sizes. *Science* **293**, 1818–1820. (doi:10.1126/science.1062081)
- Kemeny J, Snell JL. 1978 *Mathematical models in the social sciences*. Cambridge, MA: MIT Press.
- Castellano C, Fortunato S, Loreto V. 2009 Statistical physics of social dynamics. *Rev. Mod. Phys.* **81**, 591–646. (doi:10.1103/RevModPhys.81.591)
- Hernando A *et al.* 2010 Unravelling the size distribution of social groups with information theory in complex networks. *Eur. Phys. J. B* **76**, 87–97. (doi:10.1140/epjb/e2010-00216-1)
- Krings G *et al.* 2009 Urban gravity: a model for inter-city telecommunication flows. *J. Stat. Mech.* **2009**, L07003. (doi:10.1088/1742-5468/2009/07/L07003)
- González MC, Hidalgo CA, Barabási AL. 2008 Understanding individual human mobility patterns. *Nature* **453**, 779–782. (doi:10.1038/nature06958)
- Hernando A, Plastino A, Plastino AR. 2012 MaxEnt and dynamical information. *Eur. Phys. J. B* **85**, 147. (doi:10.1140/epjb/e2012-30009-3)
- Hernando A, Vesperinas C, Plastino A. 2010 Fisher information and the thermodynamics of scale-invariant systems. *Phys. A* **389**, 490–498. (doi:10.1016/j.physa.2009.09.054)
- Hernando A, Plastino A. 2012 The thermodynamics of urban population flows. *Phys. Rev. E* **86**, 066105. (doi:10.1103/PhysRevE.86.066105)
- Census bureau website. Government of USA. See www.census.gov.
- Wikimedia Commons. See commons.wikimedia.org/wiki/File:USA_Counties_with_FIPS_and_names.svg.
- Rozenfeld H, Rybski D, Gabaix X, Makse HA. 2011 The area and population of cities: new insights from a different perspective on cities. *Am. Econ. Rev.* **101**, 2205–2225. (doi:10.1257/aer.101.5.2205)
- Jiang B, Jia T. 2011 Zipf's law for all the natural cities in the United States: a geospatial perspective. *Int. J. Geogr. Inf. Sci.* **25**, 1269–1281. (doi:10.1080/13658816.2010.510801)
- Weisstein EW. Bivariate normal distribution. From MathWorld—A Wolfram Web Resource. See mathworld.wolfram.com/BivariateNormalDistribution.html.
- DEMIFER—Demographic and Migratory Flows Affecting European Regions and Cities. ESPON. See www.espon.eu/main/Menu_Projects/Menu_AppliedResearch/demifer.html.