

Truncated lognormal distributions and scaling in the size of naturally defined population clustersÁlvaro Corral ^{1,2,3,4} Frederic Udina ^{5,6} and Elsa Arcaute⁷¹*Centre de Recerca Matemàtica, Edifici C, Campus Bellaterra, E-08193 Barcelona, Spain*²*Departament de Matemàtiques, Facultat de Ciències, Universitat Autònoma de Barcelona, E-08193 Barcelona, Spain*³*Barcelona Graduate School of Mathematics, Edifici C, Campus Bellaterra, E-08193 Barcelona, Spain*⁴*Complexity Science Hub Vienna, Josefstädter Straße 39, 1080 Vienna, Austria*⁵*Department of Economics and Business, Universitat Pompeu Fabra, Ramon Trias Fargas 25-27, E-08005 Barcelona, Spain*⁶*Barcelona Graduate School of Economics, Ramon Trias Fargas 25-27, E-08005 Barcelona, Spain*⁷*Centre for Advanced Spatial Analysis, University College London, 90 Tottenham Court Road, London W1T 4TJ, England, United Kingdom*

(Received 26 October 2019; revised manuscript received 9 March 2020; accepted 9 March 2020; published 30 April 2020)

Using population data of high spatial resolution for a region in the south of Europe, we define cities by aggregating individuals to form connected clusters. The resulting cluster-population distributions show a smooth decreasing behavior covering six orders of magnitude. We perform a detailed study of the distributions, using state-of-the-art statistical tools. By means of scaling analysis we rule out the existence of a power-law regime in the low-population range. The logarithmic-coefficient-of-variation test allows us to establish that the power-law tail for high population, characteristic of Zipf's law, has a rather limited range of applicability. Instead, lognormal fits describe the population distributions in a range covering from a few dozen individuals to more than 1×10^6 (which corresponds to the population of the largest cluster).

DOI: [10.1103/PhysRevE.101.042312](https://doi.org/10.1103/PhysRevE.101.042312)**I. INTRODUCTION**

Cities are expected to experience enormous growth in the next decades, but already nowadays they can be considered in some sense as the largest structures built by humankind. However, in contrast to other human constructions, cities develop from social and economic processes combined with top-down planning. Social and economic processes in their turn depend on technological and scientific advances. So, cities are complex systems driven by self-organization, where their fundamental constituents (the individuals) participate in a vast number of different types of interactions that keep the city alive [1]. In fact, the analogy between cities and living organisms is not just a metaphor but a very deep insight [2].

Probably, the first characterization of any entity is in terms of its size, and this also holds for complex entities. It is well known that for cities their size (measured, for instance, in number of inhabitants) is broadly distributed (there are cities of vastly different sizes, taking a broad definition of a city as a “human settlement”). Then, a statistical description is necessary. Several statistical models for city size have been proposed, with the most important one being Zipf's law [3], which states that, given a country or a large region, the probability mass function $f(s)$ of city size s is given by a power-law (PL) distribution:

$$f_{\text{PL}}(s) \propto \frac{1}{s^\beta},$$

with the symbol “ \propto ” denoting proportionality and the exponent β taking values close to 2 (an important requirement is that the exponent has to be larger than 1). The law should apply at least to the largest cities, i.e., for the upper tail

of the size distribution, and so one has in mind cities and towns but not necessarily small villages. It is a remarkable fact that Zipf's law seems to hold in many other systems in which individuals gather into some sort of groups or classes (companies [4], religions [5], etc.), and where the “individuals” can be anything from animals [6] to links in the Internet [7], word tokens in a text [8], or combinations of musical notes [9].

Nevertheless, there have been authors who have argued in favor of other models; in particular, for city-size distribution the lognormal model has been proposed as the most remarkable alternative to Zipf's law, and some debate has arisen [10–12]. This debate can be put in the broader context of the adequacy of power-law fitting procedures [5,13–18], but is certainly different from the controversy about power-law relations in “urban metabolism” or urban allometry [2,19,20]. Nevertheless, at the core of both problems is the proper use of statistical tools, which is in part responsible for the recent, unfortunate, and deep problem known as reproducibility crisis, or replicability crisis [21,22].

In any case, one can realize that there is a degree of arbitrariness in city-statistics research, related to the definition of what a city is. If the usual administrative delimitations (which were established, in general, following criteria developed many decades or centuries ago) are used for modern urban agglomerations, one can wonder to what extent the results for city-size distributions are just an artifact caused by old bureaucracy. Clearly, more realistic and scientific definitions of the concept of city are necessary. This has been attempted by several authors [23–25], who have introduced the concept of naturally defined cities; see in particular the citations in Refs. [19,23,26].

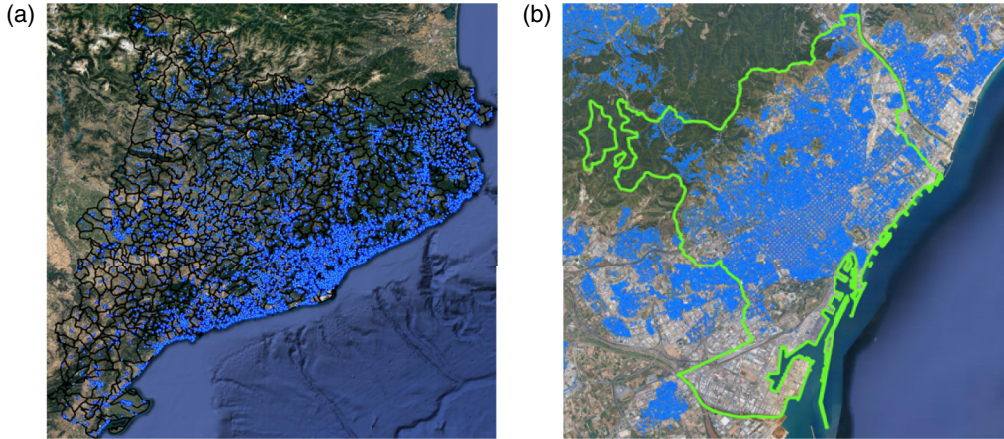


FIG. 1. (a) Whole data set: Latitude and longitude (dots) of the 7 586 888 inhabitants of Catalonia on January 1, 2013. (b) Zoom of the data around the Barcelona zone. Notice that we are representing the coordinates of the residence place of each individual, so the high resolution of the data becomes apparent in this plot. The contour line delimits the municipality of Barcelona.

In the present paper we use population data of high resolution to construct clusters of population, which we identify with cities, the size distribution of which is scrutinized with state-of-the-art statistical tools. In the next section we describe the data; in Sec. III we explain several similar procedures used to construct the population clusters (our definition of cities); and in Sec. IV we present our statistical study of the size of clusters, using scaling analysis, the logarithmic-coefficient-of-variation test, as well as truncated lognormal and power-law fits of the resulting distributions. We anticipate that the lognormal distribution is much more suitable than the power law to describe the cities arising from the analyzed data set and the city definition introduced. Also, the importance of spatial correlations in the number of inhabitants to get a broad cluster-size distribution is clearly established.

II. DATA

Here we approach the problem of city definition and the validity of Zipf's law using high-resolution data for the scatter of a population through a territory. The territory under study is Catalonia (*Catalunya*), located in northeast Spain and the capital of which is the colorful city of Barcelona. Catalonia has a population of about 7 500 000 inhabitants in a area of 32 000 km², which yields an average density around 230 inhabitants per km² and classifies Catalonia as a highly populated area. Note that these figures are similar to those of some small European countries, such as Switzerland, for example.

In Spain, the municipality councils (*ayuntamientos*) collect a population register called *Padrón Municipal de Habitantes*. All citizens are required to be registered in some municipality and actually it is necessary to be registered to access most of the administrative services like health, education, etc. The coordination of the registers of all municipalities in the country is done by the Spanish *Instituto Nacional de Estadística* [27], which sends the information referred to Catalonia to the Catalan *Institut d'Estadística de Catalunya* (IDESCAT [28]). The processing of the registers is an important step because it guarantees their high quality: duplicated entries are removed

as are people deceased or registered in a foreign-country embassy.

In the last years, IDESCAT has undertaken the task of georeferencing each individual's postal address present in the register, by means of the geocoding web service of the *Institut Cartogràfic i Geològic de Catalunya* [29], which assigns geographical coordinates to each postal address. The complete procedure including the imputation for missing data is detailed in Ref. [30].

The data that we have used for our paper are of the georeferenced population of Catalonia on January 1, 2013, with a total population of $M = 7\,586\,888$ inhabitants in 989 997 places of residence (i.e., domicile buildings), and with a 7.6% error in the georeferencing for which the procedure of imputation is applied [30]. This register can be considered as high-resolution population data, even of higher resolution than the data used in Refs. [31,32] (which were 100 and 200 m, respectively; ours is about a few meters, corresponding to the minimum distance between places of residence). The spatial distribution of the complete data set is displayed in Fig. 1.

III. CLUSTERS OF POPULATION

A. Grid approach

In order to construct our aggregations of population, we first work using a simple (equirectangular) projection of longitude and latitude into Cartesian coordinates, which introduces very little distortion due to the small extent of the territory. We cover the resulting projection by a grid composed by identical square cells, each of fixed width ℓ in degrees and projected area $\ell \times \ell$, aligned with the longitude-latitude axes. In a second, more refined approach, we transform longitude and latitude into distances (using that 1° in latitude is equal to 111.1 km and 1° in longitude is about 83 km at latitude 41°), and introduce again a square grid. We call these two approaches grid-in-degrees and grid-in-km, respectively. We advance that both of them will lead to essentially the same results.

Note that a square in longitude-latitude is equivalent to a rectangle in distance, and vice versa, so, in terms of distances,

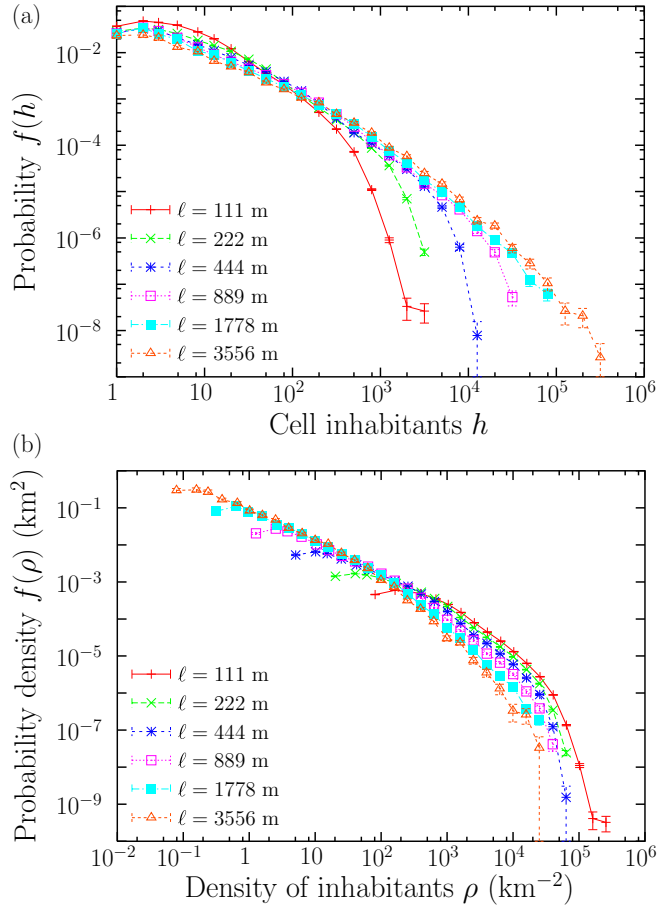


FIG. 2. (a) Empirical probability mass functions $f(h)$ of number of inhabitants h per cell, for several values of cell width ℓ , in the grid approach, using the grid-in-km procedure. (b) Corresponding empirical probability densities of population density per cell. Observe the enormous variability, from less than 0.1 inhabitant per km^2 to more than 10^5 . One can conclude that the concept of average density has a rather limited value. Unpopulated cells are not considered.

our two types of grids are rectangles (with fixed aspect ratio) and squares, respectively. When we report the width ℓ of a cell in degrees it is implicit that we are dealing with the first approach, and when ℓ is in meters or in km we will follow the second one.

The next step is counting the number of inhabitants h in each cell. For reasonable values of the cell width (for instance, $\ell = 0.001^\circ$) the resulting h turns out to be broadly distributed, from one inhabitant per cell to many thousands (we will disregard unpopulated cells, for reasons that will become clear later). For the sake of illustration, we display the corresponding probability mass function $f(h)$ in Fig. 2(a) for different values of the cell width ℓ for the grid-in-km approach, confirming the broadness of the distribution [dependence of $f(h)$ on ℓ is obviated in the notation]. The population density in each cell can be calculated straightforwardly as $\rho = h/\ell^2$, and its probability density $f(\rho)$ (the probability density of the population density) is shown in Fig. 2(b), in units of km^2 . It is obvious that, for the same ℓ , both distributions, $f(h)$ and $f(\rho)$, have the same shape, with the only difference of the scale factor ℓ^2 . The fact that we consider $f(h)$ as a probability

mass function and $f(\rho)$ as a probability density is not relevant, and comes from the consideration of h as a discrete variable and ρ as a continuous one, but this difference does not carry any deep meaning.

Under the present grid approach, our definition of city is based on the aggregation of adjacent occupied cells. This is a natural definition, previously used in Ref. [23]. We will consider a cell as occupied if its population is greater than or equal to a threshold value, and unoccupied otherwise. Naturally, the most immediate value for the population threshold is 1 [23], but other prescriptions are possible; the advantage of our high-resolution data is that the threshold can be made as small as desired, in contrast, for instance, to Ref. [19]. For this reason, the occupation threshold is equal to 1 in this paper.

More concretely, as in the problem of site percolation [33,34], a set of nearest-neighbor occupied cells surrounded by nonoccupied nearest neighbors defines a (connected) cluster [23]. These clusters will constitute a proxy for cities (we identify the clusters by means of a variation of the classic Hoshen-Kopelman algorithm), and we may generically refer to *clusters of population*. As in this framework the definition of what a cluster (or a city) depends on ℓ , and there is no *a priori* way to find an optimum ℓ , different values of this parameter will be considered, in order to test the robustness of the results.

The size s of a cluster is defined as its total population (do not get confused with its total area), i.e., for a cluster i ,

$$s_i = \sum_{j \in i} h_j, \quad (1)$$

where the sum runs for all cells j that are part of cluster i (obviously, the cluster definition implies that no cell can belong to more than one cluster). Note that the cluster sizes can range from 1 to the whole population of the territory (depending on the spatial location and on the selected value of the underlying cell width ℓ). Then, one should not find it strange in this context to talk about cities with just one inhabitant, although it is more proper to refer to them as size-1 clusters. Table I provides, for different values of ℓ , the total number of clusters and the size of the largest one (in terms of numbers of inhabitants) resulting from applying our procedure. As an example, Fig. 3 shows part of the largest cluster arising in the case of cell width $\ell = 0.002^\circ$.

B. Ball approach

An alternative approach to define clusters of population can be done using the CCA (city clustering algorithm [23]); this percolation method has been previously employed by one of the authors (see Ref. [26]). Its implementation can be done using a DBSCAN algorithm (density-based spatial clustering of applications with noise [35]), changing the distance at each iteration. The approach is based on considering “balls” of a fixed radius ℓ , centered on each individual; given a radius value ℓ , a cluster is defined as the set of all balls that overlap with (or touch) at least another ball in the cluster. This means that any individual in the cluster is at a distance smaller than (or equal to) 2ℓ of at least another individual in the cluster. The distance is measured in meters over the Earth’s surface.

TABLE I. Population-cluster properties obtained from the grid-in-degrees, grid-in-km, and ball approaches (N and s_{\max}); results from the logarithmic-coefficient-of-variation test (s_{PL} and n_{PL}), together with a power-law fit (β); and results from the lognormal fit. N is the total number of clusters, s_{\max} is the size of the largest one (always associated to Barcelona), s_{PL} is the cutoff value for the power-law tail, n_{PL} is the number of clusters in that tail, s_{ln} and n_{ln} are the equivalent for the lognormal fit, r is the number of orders of magnitude of that fit [$r = \log_{10}(s_{\max}/s_{\text{ln}})$], μ and σ are the selected lognormal parameters, and p is the p value of the lognormal fit. The last row for each approach corresponds to the aggregation for different ℓ of the rescaled variable $s/(s)$. The results for municipalities are also included, using data from IDESCAT on January 1, 2016. The values of s_{ln} swept are 50 per order of magnitude and the p value is computed from 1000 Monte Carlo simulations.

Approach	ℓ	N	s_{\max}	s_{PL}	n_{PL}	β	s_{ln}	r	n_{ln}	μ	σ	p
Grid-in-degree	1°	8376	1.98×10^6	6893	134	1.95	52.5	4.576	2450	2.120	3.102	0.21
	2°	4465	2.48×10^6	10216	91	1.95	13.2	5.275	2571	3.138	2.975	0.23
	4°	2976	3.05×10^6	13033	57	1.83	8.3	5.565	2082	3.643	2.869	0.40
	8°	1827	4.06×10^6	1473	203	1.72	5.2	5.888	1470	3.933	2.814	0.25
	16°	694	6.00×10^6	1462	101	1.69	2.1	6.458	635	4.665	2.653	0.33
	32°	102	7.27×10^6	236	61	1.47	14.5	5.702	82	6.426	2.415	0.25
	1 to 8°	17644	2.18×10^3				0.02	5.059	7998	-4.735	3.064	0.44
Grid-in-km	111 m	10258	1.94×10^6	7468	125	1.96	50.1	4.587	2810	1.542	3.176	0.23
	222 m	4966	2.11×10^6	9763	92	1.90	10.5	5.304	2972	2.772	3.062	0.26
	444 m	3209	3.05×10^6	6913	98	1.82	12.0	5.404	2047	3.604	2.866	0.21
	889 m	2095	3.62×10^6	1462	228	1.72	6.3	5.758	1613	3.980	2.788	0.32
	1778 m	922	6.00×10^6	1462	123	1.71	3.0	6.298	814	4.532	2.615	0.32
	3556 m	159	7.18×10^6	635	70	1.59	22.9	5.496	126	6.598	2.131	0.27
	111 to 889	20528	2.62×10^3				0.03	4.998	8458	-4.843	3.076	0.27
Ball	100 m	10263	1.88×10^6	6731	135	1.95	26.3	4.853	3875	0.957	3.305	0.28
	200 m	5029	2.41×10^6	9764	95	1.95	14.5	5.222	2833	3.131	2.921	0.32
	400 m	3363	2.66×10^6	11263	74	1.86	7.6	5.545	2358	3.466	2.937	0.23
	800 m	2262	3.38×10^6	2736	163	1.74	6.3	5.729	1744	3.997	2.782	0.39
	1600 m	1059	5.08×10^6	1492	148	1.71	3.0	6.226	941	4.578	2.640	0.30
	3200 m	220	7.14×10^6	486	93	1.61	251.	4.454	129	3.959	2.824	0.27
	100 to 800	20916	2.54×10^3				0.02	5.085	9622	-4.749	3.051	0.21
Municip.	947	1.60×10^6	10870	110	2.01	100	4.205	919	6.445	2.178	0.23	

The cluster population is obtained again as the sum of individuals contained in the cluster. One could still use Eq. (1), but then h_i has to be interpreted as the number of individuals in the center of each ball and the sum has to run for all balls j associated to cluster i . We refer to this procedure as the ball approach, and it has some advantages with respect to the grid approach, as it is not affected by the arbitrariness of setting an origin of coordinates for the grid and avoids problems in defining a grid over a sphere (mainly if one hypothesizes to extend this kind of analysis to much larger regions). Nevertheless, the grid approach can be more useful for other purposes, as we will see below. The properties of the clusters resulting from this approach are included in Table I. The spatial location and shape of the top ten clusters (in terms of population) can be appreciated in Fig. 4 for four different values of ℓ . Once the population has been computed for each cluster, the cluster-size distribution follows immediately; this will be shown in the next section.

Note that both the ball approach and the grid approach use a definition of city based on where people live (and we assume people live in the places they are registered). An alternative definition of cities could be based, for instance, on where people work. It would be extremely interesting to get working places of citizens in high resolution to compare those results with the ones presented here. Unfortunately, we do not have access to such data.

IV. ANALYSIS AND RESULTS

In order to investigate the validity of Zipf's law for the clusters of population, we consider, as in Ref. [5], the point of view of the distribution of sizes (in contrast to using the rank-size relation [36,37], which can lead to confusing interpretations [38]). The advantages of this choice are discussed in Ref. [39] (see also Ref. [8]). Figure 5 displays the corresponding cluster-size distributions in terms of the empirical probability mass function $f(s)$, for different values of ℓ (the notation obviates the dependence on ℓ) and for the three approaches (grid-in-degrees, grid-in-km, and balls). For the sake of comparison, some of these results appear also in the Appendix using the more popular rank-size representation.

We clearly observe the broadness of the distributions, ranging from population 1 to more than 1×10^6 (more than six orders of magnitude). The smoothness of the distributions is also apparent, with no change of behavior for all the range, except, perhaps, in the transition from one inhabitant to two, where the probability of the former value (1) is decreased with respect to the latter (in a sense, one could speculate that a fundamental unit of human population could be the couple, instead of the single individual). In contrast, the usual distribution of population for the municipalities (also included in the plot) shows a clear transition around population 200; thus, the presence of villages with population below this value is greatly diminished. With our definition of population clusters,

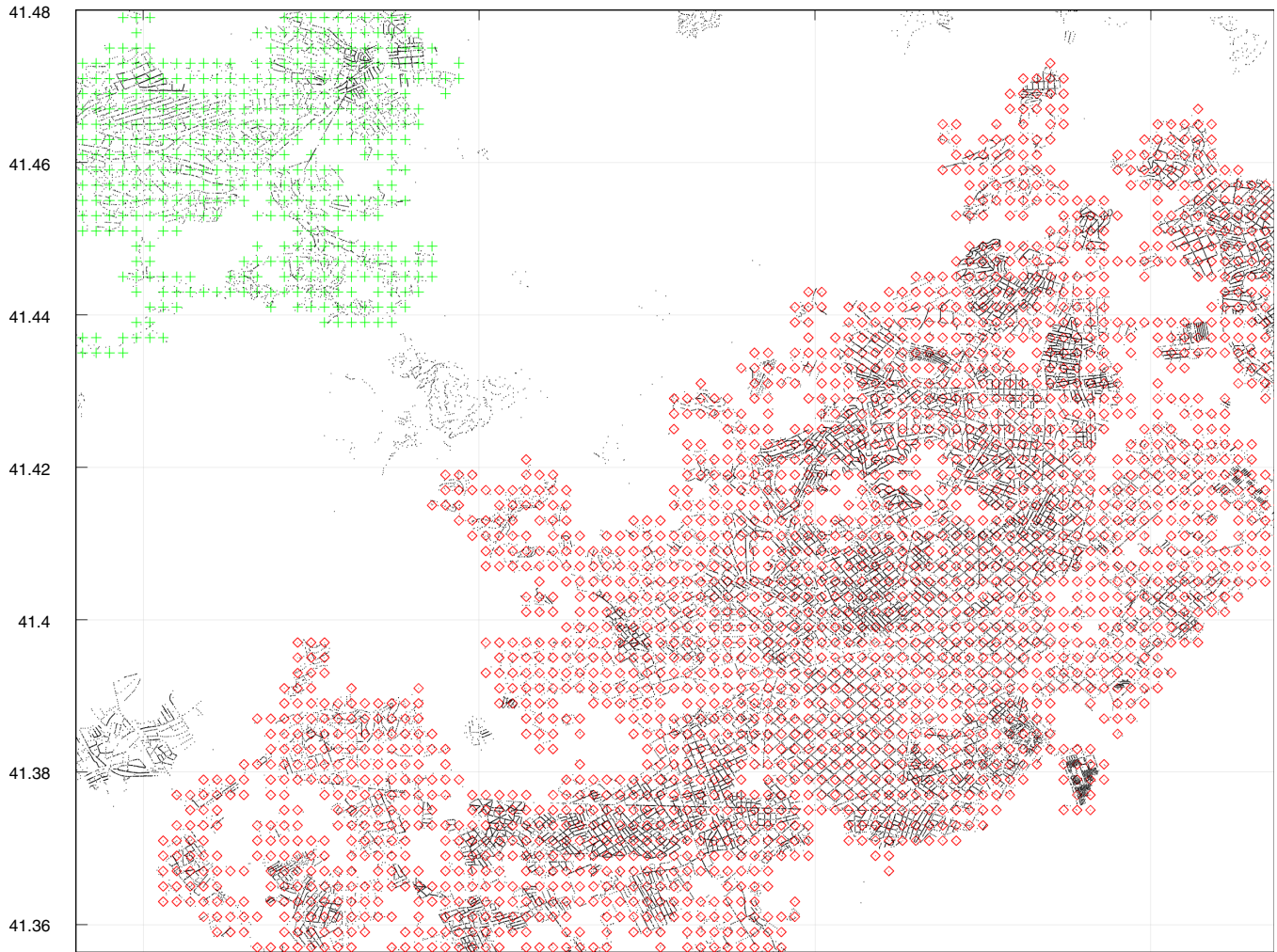


FIG. 3. Part of the cluster associated to the Barcelona urban area (in red diamonds), for cell width $\ell = 0.002^\circ$. The cluster, which is the largest one for this ℓ , includes part of other municipalities in addition to Barcelona, but there are parts of the Barcelona municipality not included in the cluster. Part of the second-largest cluster (in green crosses) is also shown. Smaller clusters are not highlighted.

instead, the broadness and smoothness of the distributions do not allow us to find discontinuity points to distinguish between cities and towns, and between towns and villages (except for the change of behavior between clusters of size 1 and 2, as just mentioned). In addition, the proximity of the distributions to a straight line in log-log representation suggests a power-law behavior. However, it is misleading to use visual information of linear behavior in log-log plots as an indication of power-law behavior [5,13]. Rigorous statistical tools are required [5,15,17,18].

A. Scaling analysis

As a first step before moving to more quantitative methods, we apply scaling analysis to the distributions (do not get confused between scaling and power-law behavior; the distinction will become clear in what follows). When one has several broad distributions, which depend on some parameter (ℓ in our case), scaling analysis can be a very informative tool [40]. At this point we are interested in the behavior of the distributions for small sizes; in particular, we will consider that they can have a power-law shape. We will denote the

corresponding exponent as α , in order to distinguish it from the tail (large-size) power-law exponent β . In other words, the distributions can have two power-law regimes, one for small s and another one for large s , but we do not assume anything in between these two regimes.

Let us assume that, for different values of the cell width ℓ , the cluster-size distributions $f(s)$ scale with some scale parameter θ (which depends on ℓ) as

$$f(s) \simeq \frac{K}{\theta} G\left(\frac{s}{\theta}\right),$$

where K is a normalization “constant” (which could depend on θ) and G is the scaling function (which is the same no matter the value of θ , i.e., of ℓ). It turns out that when, for small arguments, G does not behave as a power law, or behaves as a power law with exponent α smaller than 1, the scaling law can be rewritten as

$$f(s) \propto \frac{1}{\langle s \rangle} G\left(\frac{s}{\langle s \rangle}\right),$$

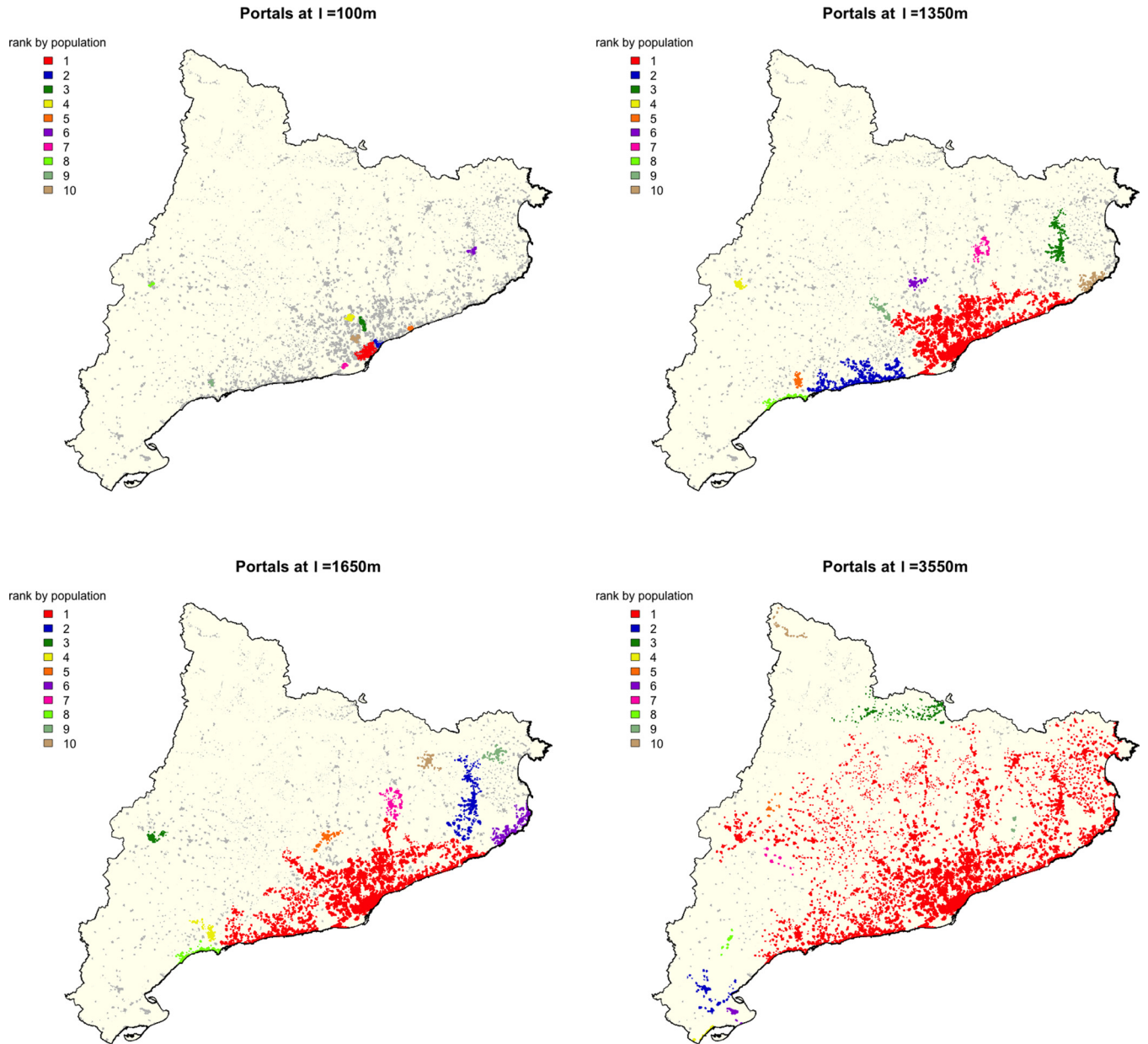


FIG. 4. Top ten clusters following the ball approach. The maps show the result of the percolation on the position of the *portals* (households), at different distance thresholds. At 100 m the red cluster shows a high overlap with Barcelona.

due to the fact that the mean $\langle s \rangle$ scales linearly with θ and the constant K is a true constant (K and the constant relating $\langle s \rangle$ to θ are absorbed into G).

However, when for small arguments G diverges as a power law with exponent α greater than 1 (but smaller than 2), the previous scaling law is not valid and one instead has

$$f(s) \propto \frac{1}{\theta} \left(\frac{m}{\theta} \right)^{\alpha-1} G\left(\frac{s}{\theta} \right) \propto \frac{\langle s \rangle^3}{\langle s^2 \rangle^2} G\left(\frac{\langle s \rangle s}{\langle s^2 \rangle} \right), \quad (2)$$

as $\langle s \rangle \propto \theta^{2-\alpha}$ and $\langle s^2 \rangle \propto \theta^{3-\alpha}$, with m the minimum value of s [below which $f(s)$ is zero] and $\langle s^2 \rangle$ the second moment of the distribution (see Ref. [40]). In fact, this new scaling law is also valid in the other case ($\alpha < 1$ or absence of power law),

due to the trivial scaling $\langle s \rangle \propto \theta$ and $\langle s^2 \rangle \propto \theta^2$ there, but the reciprocal is not true.

Therefore, estimating the moments from the sample and plotting, for different values of the cell width, $\langle s \rangle f(s)$ versus $s/\langle s \rangle$ as well as $\langle s^2 \rangle^2 f(s)/\langle s \rangle^3$ versus $\langle s \rangle s/\langle s^2 \rangle$, one will be able to check not only if a scaling law holds, but if for small arguments the distribution has a power-law shape with exponent in the range $1 < \alpha < 2$. This is done in Fig. 6 for the three representations (grid-in-degrees, grid-in-km, and balls); the data collapse for all analyzed ℓ indicates that both scaling laws are indeed fulfilled, and this implies that the power-law behavior (with $\alpha > 1$) can be discarded. The data collapse also shows that for different cell width ℓ all analyzed cluster-size distributions have (roughly) the same shape, but

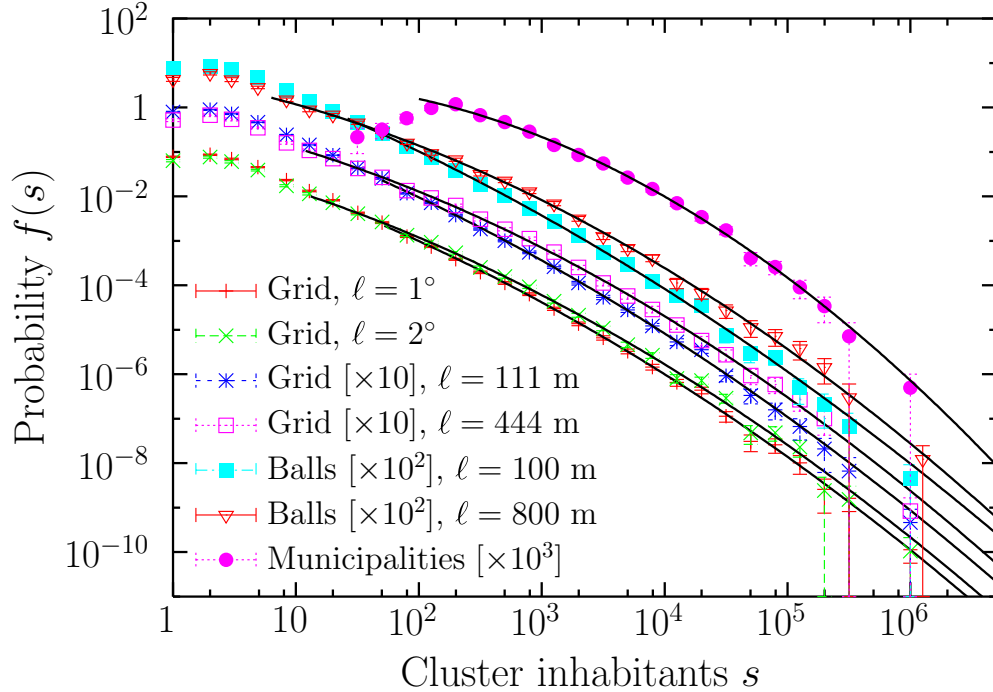


FIG. 5. Empirical probability mass functions $f(s)$ of cluster size s (in number of inhabitants) for several values of cell width ℓ , using square grid in longitude-latitude (grid-in-degrees, bottom curves), square grid in distance (grid-in-km, y axis multiplied by a factor 10, for clarity sake), and ball approach (y axis multiplied by a factor 100). Results for the usual approach based on municipalities are also included, for the sake of comparison. Continuous lines are truncated lognormal fits.

at different characteristic scale, and this shape is not a power law (with $\alpha > 1$), at least for small s .

B. Residual logarithmic coefficient of variation

Still it could happen that we had a power law not for small s but for the tail. In order to investigate this we apply the test proposed in Ref. [12] to compare the performance of a lognormal tail versus a power-law tail. By tail we mean the part of the distribution that is above an arbitrary threshold value s_{cv}

of the random variable; in other words, the tail is given by the domain $s > s_{cv}$. We expect s_{cv} to be relatively large. The test proceeds by computing the (residual) coefficient of variation c_v of $\ln(s/s_{cv})$, which is

$$c_v = \frac{1}{\sqrt{n_{cv} - 1}} \frac{\sqrt{\sum_i (\ln s_i - \overline{\ln s})^2}}{\overline{\ln s} - \ln s_{cv}}, \tag{3}$$

with $\overline{\ln s} = n_{cv}^{-1} \sum_i \ln s_i$, the sums comprising only the values s_i above s_{cv} (the “residual” values), and n_{cv} counting the

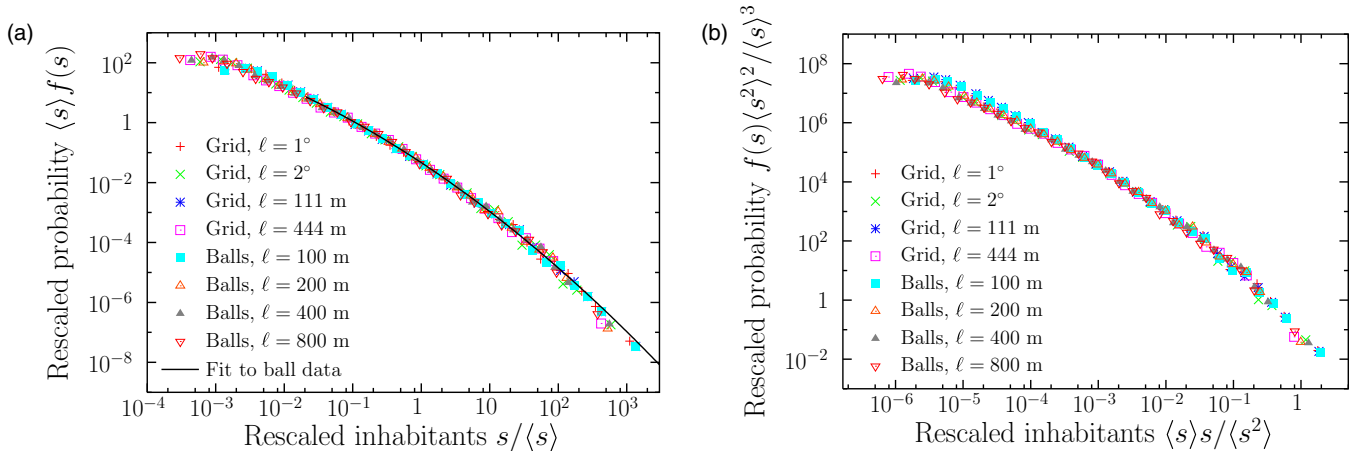


FIG. 6. (a) Empirical probability mass functions $f(s)$ rescaled by their mean value $\langle s \rangle$ for some values of ℓ in the grid-in-degrees, grid-in-km, and ball approaches. The curve is the fit corresponding to the ball approach, as shown in the table. (b) Same probability mass functions using the nontrivial rescaling given by Eq. (2) (see also Ref. [40]). The scalings in (a) and (b) can be considered as visually equivalent, which rules out the existence of a power law with $\alpha > 1$ for small s . This could be made more quantitative using the collapse algorithm of Ref. [41].

number of data fulfilling this condition (in practice, s_{cv} is set equal to an empirical value, which is excluded then from the tail, due to the strict inequality $s > s_{cv}$).

It is a fundamental fact that this residual “logarithmic” coefficient of variation (3) is a decreasing function of the likelihood ratio between the truncated lognormal and the power law [42], so a “large enough” likelihood ratio corresponds to a “small enough” c_v and this is what allows one to replace the likelihood ratio by c_v in the test (which has the clear advantage that one avoids the maximum-likelihood estimation of the parameters). Note also that the distribution of c_v does not depend either on the value of the exponent or on the value of s_{cv} ; it only depends on n_{cv} (this property will be very helpful in simulations).

As the power law can be considered a particular instance of a truncated lognormal (one with $\mu - \ln s_{cv} \rightarrow -\infty$ and $\sigma^2 \rightarrow \infty$, with μ and σ^2 the mean and variance of the associated untruncated normal distribution, which leads to power-law exponent $\beta = 1 + |\mu - \ln s_{cv}|/\sigma^2$; see Refs. [12,42]), the likelihood ratio in this case will correspond to that of nested distributions, i.e., the power law is nested into the lognormal, which constitutes then a more general distribution than the former. So, it should be clear that a truncated lognormal will fit a tail at least as well as a power law. The point is if the improvement given by the lognormal is significant or not. Note that this test constitutes the uniformly most powerful unbiased test for testing the power law hypothesis against untruncated lognormality [12,42]. Considering different values of s_{cv} we will be able to determine if there is a transition between a power-law tail (c_v close to 1) and a lognormal tail (c_v significantly below from 1) as the tail domain is increased, in other words, at which value of s_{cv} a hypothetical power-law tail starts. We refer to such a value, if it exists (the value for which s_{cv} crosses the critical line given by the percentile corresponding to the desired confidence level), as s_{PL} . More details are given in Ref. [17].

Table I and Fig. 7 incorporate the results of this approach. For the critical values of the test we take the 5th and 95th percentiles of the distribution of c_v , which leads to a 90% confidence that the tail is a power law, and a 95% confidence that the tail is a power law in front of the lognormal. The reason behind this is that, given the null hypothesis that the distribution is a power law, if the alternative is that the distribution is not a power law, the test is two-sided (rejection when c_v is significantly far from 1); however, if the alternative is that the distribution is lognormal, the test is one-sided (rejection when c_v is significantly below 1). Note that the percentiles can be computed from Monte Carlo simulations of a simple (unit-scale parameter) exponential distribution [42]. We observe in the table that, except for the largest considered cell width ℓ , the cutoff value of the hypothetical power-law tail s_{PL} is in a range from 1500 to 10 000 (inhabitants). The number of population clusters (cities) covered by that range (number of points in the tail, n_{PL}) turns out to be rather small, from 60 to 230, roughly.

We can use a variation of the logarithmic-coefficient-of-variation test (in fact, its original linear form, essentially) to rule out that the distribution of cluster size has an exponential tail, as was claimed in other contexts [43] (and already criticized in Refs. [44,45]). If we compute the usual residual

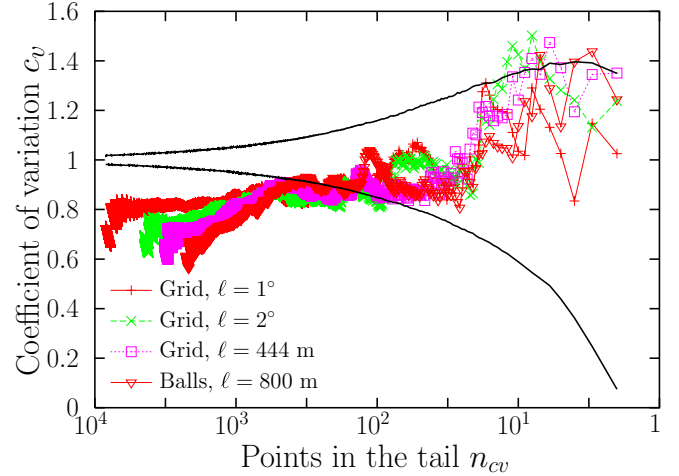


FIG. 7. Empirical value of the “residual logarithmic coefficient of variation” c_v of the cluster population as a function of the number of residual points in the tail of the distribution, n_{cv} . Several examples are shown, corresponding to our three approaches. The solid lines correspond to the simulated 5th and 95th percentiles of the distribution of c_v under the power-law null hypothesis. In all cases the first crossing below the fifth percentile takes place for n_{cv} around 100, which corresponds to n_{PL} in Table I. Note that the horizontal axis is reversed, to display the tail on the right side.

coefficient of variation of the cluster size (just dividing the standard deviation and the mean of the difference between s and a lower cutoff s'_{cv} , i.e., $s - s'_{cv}$) and compare with the results expected for an exponential variable [42], we get that c_v turns out to be above the 95th percentile, which rules out the exponential tail for any value of s'_{cv} . Thus, the tail of the cluster-size distribution is not exponential. Similar conclusions are reached if one uses as a test statistic the mean of $s - s'_{cv}$ divided by its maximum [46].

C. Lognormal fits

As the existence of a power-law tail does not rule out the existence of a lognormal tail, and due to the low range of the power-law tail, and due also to the fact that we have ruled out the existence of a power law for small s , as a next step we explore the performance of a lognormal fit. Concretely, a lower-truncated lognormal (ln) distribution is given by a probability density

$$f_{\ln}(s) = \sqrt{\frac{2}{\pi}} \left[\operatorname{erfc} \left(\frac{\ln s_{\ln} - \mu}{\sqrt{2}\sigma} \right) \right]^{-1} \frac{1}{\sigma s} \exp \left(-\frac{(\ln s - \mu)^2}{2\sigma^2} \right), \quad (4)$$

defined for s above the lower cutoff s_{\ln} , with erfc the complementary error function, and μ and σ the mean and standard deviation of the associated untruncated normal distribution [e^μ turns out to be the scale parameter θ of $f_{\ln}(s)$ and σ its shape parameter].

We fit this truncated lognormal distribution to our population data extending to lognormals the method introduced in Refs. [15,47] for (continuous) power laws, consisting in maximum-likelihood estimation plus a Kolmogorov-Smirnov goodness-of-fit test [17]. Although in recent years the recipe

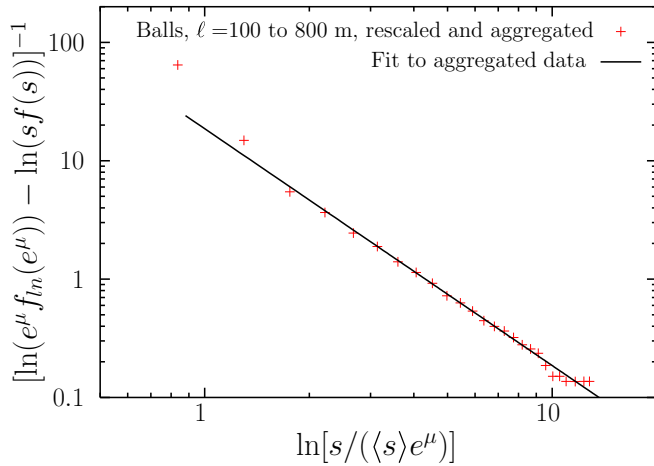


FIG. 8. Empirical probability density of rescaled cluster sizes ($s/\langle s \rangle$) aggregated for diverse values of ℓ (100, 200, 400, and 800 m), in the ball approach. The axes have been transformed to evaluate the goodness of the lognormal fit (Table I, continuous line). Note that both axes are doubly logarithmic. Data are the same as those in Fig. 6(a) for the ball approach.

of Clauset *et al.* [5] has become very popular for power-law tails, we prefer the more intuitive approach of Refs. [15,17,47] (the reasons for our choice derive in part from the results of Refs. [14,18] and also from the present research). The extension of the method to lognormals has also been used in Ref. [17]. At the end, we arrive to optimal values of the three parameters s_{in} , μ , and σ , which are included in Table I.

We see in the table how, in contrast to the power-law tail, the lognormal fit covers a considerable range of data, with a rather small value of s_{in} (from 2 to 50 inhabitants, which leads to fits valid for more than four orders of magnitude in population) and a relatively large n_{in} (either $n_{\text{in}} \simeq N$ or $n_{\text{in}} > 2000$). Inadequacy of the lognormal to fit the smallest values of s is expected due to the fact that the lognormal is a continuous distribution and s is a discrete variable.

The performance of the fits can be visually appreciated from Fig. 5; however, in order to stress the lognormal behavior, we apply a transformation which should lead to a linear plot in the lognormal case (see Fig. 8). This consists in representing

$$-\ln\{\ln[e^{\mu} f_{\text{in}}(e^{\mu})] - \ln[sf(s)]\} \text{ versus } \ln \ln \frac{s}{e^{\mu}}$$

(restricted to $s > e^{\mu}$, to avoid the overlap with the branch $s < e^{\mu}$) or, equivalently, $\{\ln[e^{\mu} f_{\text{in}}(e^{\mu})] - \ln[sf(s)]\}^{-1}$ versus $\ln s - \mu$ in additional logarithmic scale on each axis. Note that $f(s)$ refers to the empirical estimations of the density whereas $f_{\text{in}}(e^{\mu})$ refers to the theoretical distribution evaluated at $s = e^{\mu}$ (so, both μ and σ need to be estimated from data); note also that both axes are doubly logarithmic. Indeed, from Eq. (4) we get

$$-\ln \ln \left[\frac{e^{\mu} f_{\text{in}}(e^{\mu})}{sf(s)} \right] = -2 \ln \ln \frac{s}{e^{\mu}} + \ln(2\sigma^2),$$

which is a straight line with slope -2 , in the variables defined above. We see in Fig. 8 how the straight behavior is more

apparent than in the usual log-log plot of $f(s)$ versus s (Fig. 5), so we have an additional visual support for the lognormal fit in front of the power law. In fact, an additional shift by $2 \ln \ln \langle s \rangle$ is applied in the figure, in order to collapse the different distributions, which are merged into a single one.

D. Scaling to fix the lognormal parameters

From Table I one also realizes that the lognormal scale parameter e^{μ} increases with the grid or ball size ℓ but the shape parameter σ keeps constant, roughly (from 2.5 to 3, except for the largest ℓ). We now take advantage of the fact that the cluster-size distributions display scaling [at least approximately; see Fig. 6(a)]. Therefore, for the different data sets (corresponding to different ℓ) we can rescale s as $s/\langle s \rangle$ (as in Fig. 8), and then merge the different data sets into a single one, to which we can fit the lognormal distribution. The results in Table I confirm that a single lognormal can fit the rescaled distributions corresponding to different ℓ . Denoting the new value of the parameter μ as μ' (corresponding to the rescaled data), and taking into account the simple relation between the original and the rescaled data, we can express the value of μ of the original (not rescaled) distributions as $\mu = \mu' + \ln \langle s \rangle$, which shows indeed that μ increases with ℓ , as $\langle s \rangle$ increases with ℓ . Summarizing, and approximating the results from the table, we can write $e^{\mu} \simeq \langle s \rangle / e^{4.8}$ and $\sigma \simeq 3$. For the dependence of $\langle s \rangle$ on ℓ , using the data-collapse optimization method explained in Ref. [41], we find $\langle s \rangle \propto \ell^{0.8}$; nevertheless, the uncertainty of this exponent is rather large.

E. Power-law fits and comparison with lognormal

If one insists in fitting a power law to the tail of the distribution (taking advantage of the fact that the tail of a lognormal becomes asymptotically a power law, the tail being understood as the power-law tail defined by the logarithmic-coefficient-of-variation test as calculated above) we find exponents roughly close to but below $\beta = 2$ (Zipf's value; see Table I); nevertheless, for the reasons mentioned above, we consider the power-law fit as anecdotic and prefer the lognormal fit, valid for a much larger range (Table I).

In fact, we can make a quantitative comparison between power-law and lognormal fits. For the power-law tail one already knows that the lognormal fit has to yield a likelihood at least as large as the one given by the power law (when both are fitted over the same range), but this difference is not significant (this is what the c_v test is about, allowing one to define the power-law tail in the range $s > s_{\text{PL}}$). In this sense, the power law always wins over the lognormal. However, we are interested in fitting not just a short power-law tail, but as much as possible of the bulk of the distribution (including the tail). So, on the one hand we have (model 1) the lognormal fit, in the range $s \geq s_{\text{in}}$, and on the other hand (model 2) we have to consider the power-law tail complemented by the lognormal fit from s_{in} to s_{PL} . Note that model 2 yields four parameters (μ and σ plus s_{PL} and β , considering s_{in} as fixed), whereas model 1, the single lognormal, only has two parameters (μ and σ). In mathematical terms, the two models

TABLE II. Model comparison between single lognormal fit (model 1, with two parameters) and lognormal bulk plus power-law tail (model 2, with four parameters). Differences are computed as model 2 minus model 1. The outcome always favors the simpler model 1, lognormal. Results for the chi-squared test explained in the text (see also [11]) applied to model 1 are also included and confirm that the lognormal fit cannot be rejected.

ℓ	$\Delta \ln \hat{L}$	ΔAIC	ΔBIC	χ^2 Model 1
1°	-1.269	6.538	18.15	0.0012
2°	-1.105	6.209	17.91	0.0009
4°	-0.632	5.264	16.55	0.4684
8°	-1.381	6.761	17.35	0.8624
16°	-1.603	7.207	16.11	0.9272
32°	-0.610	5.220	10.03	0.8631
111 m	-1.152	6.305	18.19	0.0006
222 m	-1.380	6.760	18.75	0.2616
444 m	-1.158	6.316	17.56	0.6692
889 m	-1.446	6.893	17.66	1.2255
1778 m	-2.179	8.358	17.76	0.6197
3556 m	-1.221	6.441	12.11	0.0000
100 m	-1.030	6.060	18.58	0.1532
200 m	-1.413	6.826	18.72	0.0147
400 m	-1.346	6.692	18.22	0.2874
800 m	-1.364	6.728	17.66	1.3501
1600 m	-1.492	6.983	16.68	0.5327
3200 m	-1.658	7.317	13.04	0.3306
Municip.	-1.559	7.118	16.76	0.0054

can be expressed as $f_1(s) = f_{\ln}(s)$ for $s \geq s_{\ln}$ and

$$f_2(s) = \begin{cases} f_{\ln}(s) & \text{for } s_{\ln} \leq s \leq s_{\text{PL}} \\ f_{\text{PL}}(s)n_{\text{PL}}/n_{\ln} & \text{for } s > s_{\text{PL}}, \end{cases}$$

where the factor n_{PL}/n_{\ln} ensures normalization [note that we do not impose continuity of $f_2(s)$ at $s = s_{\text{PL}}$, that would reduce the likelihood of the resulting fit].

We can compute the difference in log-likelihoods of both models as

$$\Delta \ln \hat{L} = \ln \hat{L}_2 - \ln \hat{L}_1 = \sum_{i=1}^{n_{\ln}} [\ln f_2(s_i) - \ln f_1(s_i)]$$

(we only number the clusters with $s \geq s_{\ln}$, and, in increasing order, $s_1 \leq s_2 \leq \dots \leq s_{n_{\ln}}$); however, as the two models coincide in the range $s_{\ln} \leq s < s_{\text{PL}}$, the comparison of likelihoods only needs to be done at the tail, $s \geq s_{\text{PL}}$, and thus

$$\Delta \ln \hat{L} = \sum_{i=n_{\ln}-n_{\text{PL}}+1}^{n_{\ln}} [\ln f_{\text{PL}}(s_i; \beta, s_{\text{PL}}) - \ln f_{\ln}(s_i; \mu, \sigma, s_{\text{PL}})],$$

where we have made explicit the dependence on the parameters. Note that for the lognormal we have replaced its lower cutoff s_{\ln} by s_{PL} ; this is what allows us to eliminate the factor n_{PL}/n_{\ln} that multiplied $f_{\text{PL}}(s)$ as in this way both distributions are normalized in the range $s \geq s_{\text{PL}}$.

As expected, this difference of log-likelihoods turns out to be negative (see Table II), which means that model 1 (single lognormal) would be favored; however, this comparison does not take into account the different number of parameters. If we introduce the Akaike information criterion, $\text{AIC} = 2k -$

$2 \ln \hat{L}$, where k is the number of parameters, we get $\Delta \text{AIC} = 4 - 2\Delta \ln \hat{L}$, which is positive, favoring more clearly the lognormal model (Table II again). The Bayesian information criterion, $\text{BIC} = k \ln n_{\ln} - 2 \ln \hat{L}$, leads to $\Delta \text{BIC} = 2 \ln n_{\ln} - 2\Delta \ln \hat{L}$, with the same conclusion. In any case, the lognormal fit is preferred.

Note that when only the tail is compared the power law has one parameter whereas the lognormal has two, which favors the former; however, considering the whole range $s \geq s_{\ln}$, the situation is reversed, as the power-law tail combined with the lognormal bulk has four parameters, which favors the single lognormal, despite the fact that the difference in log-likelihood does not change. Note also that the difference in AIC does not depend on the number of data in the fit, whereas ΔBIC does; in fact, the more data, the better the simple (lognormal) fit.

Finally, in order to allow some comparison, we apply the same test as in Ref. [11], which is Pearson's chi-squared test. We consider just two classes, $s_{\ln} \leq s < s_{\text{PL}}$ and $s \geq s_{\text{PL}}$, and compute $x^2 = \sum_{k=1}^2 (E_k - O_k)^2 / E_k$, where O_k is the observed number of clusters in the k th class (either $n_{\ln} - n_{\text{PL}}$ or n_{PL}) and E_k is the expected number of clusters from the lognormal fit, which is either $(1 - q_{\text{tail}})n_{\ln}$ or $q_{\text{tail}}n_{\ln}$, with q_{tail} the probability that the lognormal fit assigns to the tail. This is calculated as

$$q_{\text{tail}} = \text{erfc}\left(\frac{\ln s_{\text{PL}} - \mu}{\sqrt{2}\sigma}\right) / \text{erfc}\left(\frac{\ln s_{\ln} - \mu}{\sqrt{2}\sigma}\right).$$

The results, included in Table II, show that the values of x^2 are in all cases too small to reject the lognormal fit.

F. Dragon-king effect

A remarkable fact is that, although the cluster associated to the city of Barcelona could be considered to have the status of a dragon king (in Sornette's sense of a very large outlier in s [48]), with a population much larger than that of the second largest cluster (at least fivefold larger, depending on ℓ), with a clear deviation therefore with respect to Zipf's law in terms of the rank-size representation, this does not cause the rejection either of the lognormal fit or of the power-law tail for the cluster-size distribution. In other words, in terms of the distribution of sizes, the presence of the dragon-king effect does not preclude the validity of the lognormal and power-law fits (whereas, in terms of the rank-size representation, those fits should be rejected, due to the incompatibility of the sizes for rank equal to 1 and 2).

We can go one step further and study the influence of the largest cluster on the distribution, just removing it. Starting with the lognormal fit, the results show very little change, with very small variation in the parameters s_{\ln} , μ , and σ with respect the original case (dragon king not removed). This is not unexpected, as the lognormal fits contain a large number of data (n_{\ln} is high) and the influence of the dragon king is in any case small. If we try to fit a power law for the tail of the distribution of sizes using the same initial range as the original case ($s > s_{\text{PL}}$), the power-law fit is not rejected, but the values of the parameters show some change (typically the change in the exponent β is smaller than 0.1, but in one case it gets larger than 0.3, not shown). This is also expected, as the power-law range is much smaller than the lognormal range,

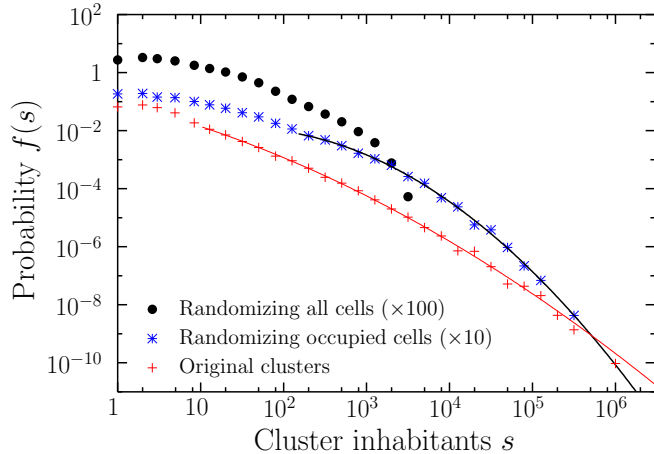


FIG. 9. Comparison of the original distribution of cluster sizes $f(s)$ with the distribution of cluster sizes obtained when the values h of the population of each cell are randomized between occupied cells. The less interesting case in which the values of h are randomized between all cells (occupied and unoccupied) is also shown. The example shown corresponds to the grid-in-km approach, with $\ell = 222$ m. Data in the y axis multiplied by integer powers of 10, for clarity sake.

so in the power-law fit the weight of the dragon king becomes larger.

Nevertheless, if we apply the logarithmic-coefficient-of-variation test to the data with the dragon king removed we obtain a rather different picture. The removal of the largest cluster decreases the variability of the distribution, which yields a smaller coefficient of variation, which causes the coefficient to cross to the significance region before (i.e., with much less data in the tail). So, the data with the dragon king removed have a shorter power-law tail (the new value s_{PL}^* becomes larger than the original s_{PL} , and the number of data in the tail verifies $n_{PL}^* < n_{PL}$). The shorter power-law range leads to a larger value of the exponent, which is an indication in fact of the curvature of the tail and an additional support for the superiority of the lognormal fit.

G. Origin of the large variability in city sizes

Our approach allows us to investigate the origin of the broadness of the distribution of the size of population clusters (i.e., the size of cities, in our definition). In the grid approach, the distribution $f(s)$ arises from the sum of the number of inhabitants h of each cell [with a distribution given by $f(h)$] along each cell; so, $f(s)$ depends both on $f(h)$ and on the distribution of the number of cells per cluster (directly related to cluster area); nevertheless, this is not enough, as one needs to take into account that the values of h are not independent from cell to cell, i.e., there are spatial correlations in the values of h (highly populated cells tend to be surrounded by highly populated cells, and reciprocally). This is in fact an obvious fact, but we can demonstrate the relevance of correlations in the values of h eliminating these correlations and looking for the resulting $f(s)$.

We eliminate the correlations just reshuffling the values of h among occupied clusters; this keeps the distribution $f(h)$

and the spatial extent of clusters unchanged. The results are displayed in Fig. 9, showing that the resulting randomized $f(s)$ is less broad than the original $f(s)$; in particular, no sign of an approximate power-law tail with exponent close to 2 is found, leading to the conclusion that the main cause of the large variability in city size (and the cause of the rough Zipf-like behavior) is spatial correlation. This means that the (fractal) shape of cities (in a two-dimensional projection) is not enough to explain their population distribution. Note also that in both cases (original clusters and clusters with population in cells randomized) the distribution of cluster areas is the same. Above we mentioned the advantages of the ball approach over the grid approach; however, notice that this reshuffling procedure can only be performed under the grid approach, so both approaches can be considered as complementary.

V. CONCLUSIONS

Population data of high spatial resolution allow one to locate individuals in space and to build spatial clusters of them. If the locations of individuals correspond to their residence place (as it is in our paper), these clusters constitute a natural definition of human settlements or, broadly speaking, cities. We have scrutinized the distribution of the number of individuals in the clusters obtained in this way (natural-city size) with up-to-date data-analysis tools.

On the one hand, scaling analysis allows us to rule out the existence of a power-law size distribution with exponent $\alpha > 1$ for the smallest cities; on the other hand, the logarithmic-coefficient-of-variation test shows that a power-law tail has a very limited range of applicability. Instead, a lognormal fit holds for a considerable part of the data and does not only yield a higher likelihood than a model joining a lognormal part in the bulk plus a power-law tail, but also has less parameters, which makes the lognormal to be clearly supported by model comparison using AIC or BIC.

Obviously, we do not dispute that the U.S. population distribution, as measured from its census, is better described by a power law than by a lognormal at its tail, as claimed in other studies [11,12]. But it could be that the bad performance of the lognormal at the tail of the U.S. city distributions is due to the fact that it is a pure (untruncated) lognormal which is fitted, and not a truncated one, as used here. In any case, our example shows clearly that other data sets and/or other definitions of cities can lead to a better characterization by (truncated) lognormal distributions. This implies that universality [49] does not seem to be a characteristic of city-size distributions, as also found for other complex systems, for instance, wildfires [17,50].

Our spatial-grid based approach also allows us to stress the importance of spatial correlations in the broadness of the city-size distribution; in other words, knowledge of the area occupied by cities together with the distribution of individuals in small grids does not allow us to explain the number of inhabitants in cities. Finally, we have seen how, although our largest cluster (associated to the city of Barcelona) has the status of a dragon king, it does not have an important influence on the city-size distribution. The reason is that, under the distribution-of-sizes representation (in contrast to the rank-

size approach), the largest cluster counts just as one single realization of the random variable (in contrast to more than 1×10^6 counts of individuals from that cluster in the other approach), and, thus, it has a very small statistical weight.

In a further step, one can fit to the data a power law truncated also from above (a power law defined in a range $s_{p1} \leq s \leq s_{p2}$, where these parameters are optimized by the fitting [15,17]). This leads to exponents α in the range 1.6 to 1.8, valid for several orders of magnitude (between 3 and 4), starting at values of s around 500 or higher. In some cases, the upper truncation point turns out to be above s_{\max} (the largest cluster), which means that the fitting method is not able to distinguish the power-law tail (which has a somewhat different power-law exponent β). In any case, the lognormal fit turns out to be valid in a wider range, and it is thus preferred.

In a future work one could compare the performance of the lognormal with that of a double power law [17], which has been proposed for other Zipf-like systems [51,52]. We believe that, although our results have been obtained for a single region (due to the availability of high-resolution data), their validity is much more general (there is nothing special about Catalonia, the region under study). However, at this point this is just a conjecture that needs to be confirmed when other high-resolution data sets become available.

ACKNOWLEDGMENTS

We are indebted to E. Suñé from IDESCAT, S. Manrubia, J. del Castillo, and I. Serra. A.C. has received funding from the Spanish MINECO grant awarded to the Barcelona Graduate School of Mathematics under the María de Maeztu Program (Grant No. MDM-2014-0445), as well as from MINECO Projects No. FIS2012-31324 and No. FIS2015-71851-P, MICIU Project No. FIS-PGC2018-099629-B-I00, and AGAUR Project No. 2014SGR-1307. F.U. and E.A. acknowledge funding from MICIU Grant No. PGC2018-101643-B-I00 and EPSRC Grant No. EP/M023583/1, respectively.

APPENDIX

For comparison of visualization procedures, part of the results in Fig. 5 for the distributions of cluster sizes are shown

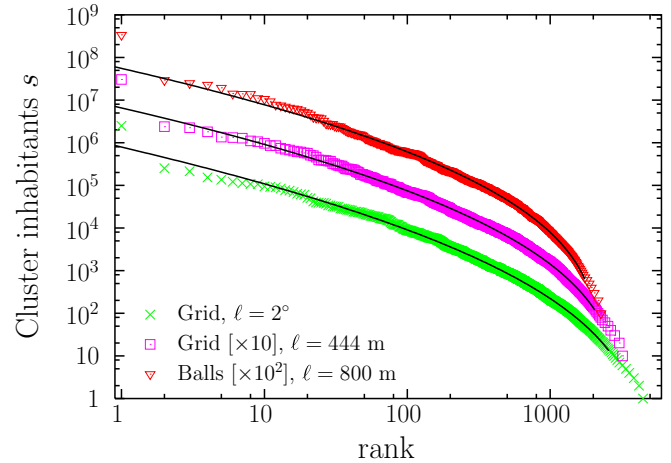


FIG. 10. Same results as in Fig. 5, represented in terms of the rank-size relation. The size (vertical axis) has been multiplied by integer powers of 10, for the sake of clarity. Continuous lines are the truncated lognormal fits translated into the rank-size representation (the mathematical expression is not lognormal anymore but involves the inverse of the error function).

here in terms of the rank-size representation (see Fig. 10). The fits shown are the same as in Fig. 5, translated into the rank-size representation; so, the cluster-size distribution is fitted (and not the rank-size relation). Notice that a good fit for the cluster-size distribution (p value above 0.20) also looks good in the present representation, except for the lowest ranks (largest sizes). The reason is that, when fitting the cluster sizes, the largest cluster (for instance) is just a single point, and does not influence very much the resulting fit (it does not change very much if the largest cluster has size 1×10^6 or 2×10^6 , roughly). However, in the rank-size representation, changes of several thousand (compared to populations of 1×10^6) are critical and can lead to the rejection of any fit. Essentially, to fit the rank-size relation (taking the rank as the random variable instead of the size) one would need one parameter for each value of the rank (for its smallest values, which are the ones containing more statistics). See Ref. [39].

-
- [1] M. Barthelemy, The statistical physics of cities, *Nat. Rev. Phys.* **1**, 406 (2019).
 - [2] L. Bettencourt and G. West, A unified theory of urban living, *Nature (London)* **467**, 912 (2010).
 - [3] G. K. Zipf, *Human Behavior and the Principle of Least Effort* (Addison-Wesley, Reading, MA, 1949).
 - [4] R. L. Axtell, Zipf distribution of U.S. firm sizes, *Science* **293**, 1818 (2001).
 - [5] A. Clauset, C. R. Shalizi, and M. E. J. Newman, Power-law distributions in empirical data, *SIAM Rev.* **51**, 661 (2009).
 - [6] S. Pueyo and R. Jovani, Comment on “A keystone mutualism drives pattern in a power function,” *Science* **313**, 1739c (2006).
 - [7] L. A. Adamic and B. A. Huberman, Zipf’s law and the Internet, *Glottometrics* **3**, 143 (2002).
 - [8] I. Moreno-Sánchez, F. Font-Clos, and A. Corral, Large-scale analysis of Zipf’s law in English texts, *PLoS ONE* **11**, e0147073 (2016).
 - [9] J. Serrà, A. Corral, M. Boguñá, M. Haro, and J. Ll. Arcos, Measuring the evolution of contemporary western popular music, *Sci. Rep.* **2**, 521 (2012).
 - [10] J. Eeckhout, Gibrat’s law for (all) cities, *Amer. Econ. Rev.* **94**, 1429 (2004).
 - [11] M. Levy, Gibrat’s law for (all) cities: Comment, *Amer. Econ. Rev.* **99**, 1672 (2009).
 - [12] Y. Malevergne, V. Pisarenko, and D. Sornette, Testing the Pareto against the lognormal distributions with the uniformly most powerful unbiased test applied to the distribution of cities, *Phys. Rev. E* **83**, 036111 (2011).

- [13] E. P. White, B. J. Enquist, and J. L. Green, On estimating the exponent of power-law frequency distributions, *Ecol.* **89**, 905 (2008).
- [14] A. Corral, F. Font, and J. Camacho, Non-characteristic half-lives in radioactive decay, *Phys. Rev. E* **83**, 066103 (2011).
- [15] A. Deluca and A. Corral, Fitting and goodness-of-fit test of non-truncated and truncated power-law distributions, *Acta Geophys.* **61**, 1351 (2013).
- [16] A.-L. Barabási, Love is all you need. Clauset's fruitless search for scale-free networks, 2018, <https://www.barabasilab.com/post/love-is-all-you-need>.
- [17] A. Corral and A. González, Power law size distributions in geoscience revisited, *Earth Space Sci.* **6**, 673 (2019).
- [18] I. Voitalov, P. van der Hoorn, R. van der Hofstad, and D. Krioukov, Scale-free networks well done, *Phys. Rev. Res.* **1**, 033034 (2019).
- [19] E. Arcaute, E. Hatna, P. Ferguson, H. Youn, A. Johansson, and M. Batty, Constructing cities, deconstructing scaling laws, *J. Roy. Soc. Interf.* **12**, 102 (2015).
- [20] J. C. Leitão, J. M. Miotto, M. Gerlach, and E. G. Altmann, Is this scaling nonlinear? *R. Soc. Open Sci.* **3**, 150649 (2016).
- [21] R. Peng, The reproducibility crisis in science: A statistical counterattack, *Significance* **12**, 30 (2015).
- [22] R. L. Wasserstein and N. A. Lazar, The ASA's statement on p-values: Context, process, and purpose, *Am. Stat.* **70**, 129 (2016).
- [23] H. D. Rozenfeld, D. Rybski, J. S. Andrade, M. Batty, H. E. Stanley, and H. A. Makse, Laws of population growth, *Proc. Natl. Acad. Sci. USA* **105**, 18702 (2008).
- [24] B. Jiang and T. Jia, Zipf's law for all the natural cities in the United States: A geospatial perspective, *Int. J. Geograp. Inform. Sci.* **25**, 1269 (2011).
- [25] B. Jiang, J. Yin, and Q. Liu, Zipf's law for all the natural cities around the world, [arXiv:1402.2965](https://arxiv.org/abs/1402.2965).
- [26] E. Arcaute, C. Molinero, E. Hatna, R. Murcio, C. Vargas-Ruiz, A. P. Masucci, and M. Batty, Cities and regions in Britain through hierarchical percolation, *R. Soc. Open Sci.* **3**, 150691 (2016).
- [27] Instituto Nacional de Estadística, <http://www.ine.es>.
- [28] Institut d'Estadística de Catalunya, <http://www.idescat.cat>.
- [29] Institut Cartogràfic i Geològic de Catalunya, <http://www.icgc.cat>.
- [30] E. Suñé Luis, Hacia un registro estadístico de territorio, *Proceedings of the XXXV Congreso Nacional de Estadística e Investigación Operativa*, IX Jornadas de Estadística Pública, 2015 (unpublished, in Spanish).
- [31] C. D. V. Orozco, J. Golay, and M. Kanevski, Multifractal portrayal of the Swiss population, [arXiv:1308.4038](https://arxiv.org/abs/1308.4038).
- [32] F. Sémécurbe, C. Tannier, and S. G. Roux, Spatial distribution of human population in France: Exploring the modifiable areal unit problem using multifractal analysis, *Geograp. Anal.* **48**, 292 (2016).
- [33] D. Stauffer and A. Aharony, *Introduction To Percolation Theory*, 2nd ed. (CRC, Boca Raton, FL, 1994).
- [34] K. Christensen and N. R. Moloney, *Complexity and Criticality* (Imperial College, London, 2005).
- [35] *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96)*, Portland, Oregon, USA, edited by E. Simoudis, J. Han, and U. M. Fayyad (AAAI, New York, 1996).
- [36] E. G. Altmann and M. Gerlach, Statistical laws in linguistics, in *Creativity and Universality in Language. Lecture Notes in Morphogenesis*, edited by M. D. Esposti, E. G. Altmann, and F. Pachet (Springer, New York, 2016).
- [37] G. Martínez-Mekler, R. A. Martínez, M. B. del Río, R. Mansilla, P. Miramontes, and G. Cocho, Universality of rank-ordering distributions in the arts and sciences, *PLoS ONE* **4**, e4791 (2009).
- [38] M. Cristelli, M. Batty, and L. Pietronero, There is more than a power law in Zipf, *Sci. Rep.* **2**, 812 (2012).
- [39] A. Corral, I. Serra, and R. Ferrer-i-Cancho, The distinct flavors of Zipf's law in the rank-size and in the size-distribution representations, and its maximum-likelihood fitting, [arXiv:1908.01398](https://arxiv.org/abs/1908.01398).
- [40] A. Corral, Scaling in the timing of extreme events, *Chaos. Solit. Fract.* **74**, 99 (2015).
- [41] A. Deluca and A. Corral, Scale invariant events and dry spells for medium-resolution local rain data, *Nonlinear Proc. Geophys.* **21**, 555 (2014).
- [42] J. del Castillo and P. Puig, The best test of exponentiality against singly truncated normal alternatives, *J. Am. Stat. Assoc.* **94**, 529 (1999).
- [43] S. Bernhardtsson, L. E. Correa da Rocha, and P. Minnhagen, The meta book and size-dependent properties of written language, *New J. Phys.* **11**, 123015 (2009).
- [44] F. Font-Clos, G. Boleda, and A. Corral, A scaling law beyond Zipf's law and its relation to Heaps' law, *New J. Phys.* **15**, 093033 (2013).
- [45] A. Corral and F. Font-Clos, Dependence of exponents on text length versus finite-size scaling for word-frequency distributions, *Phys. Rev. E* **96**, 022318 (2017).
- [46] P. Rochet and I. Serra, The mean/max statistic in extreme value analysis, [arXiv:1606.08974](https://arxiv.org/abs/1606.08974).
- [47] O. Peters, A. Deluca, A. Corral, J. D. Neelin, and C. E. Holloway, Universality of rain event size distributions, *J. Stat. Mech.* (2010) P11030.
- [48] D. Sornette, Dragon-kings, black swans and the prediction of crises, *Int. J. Terraspace Sci. Eng.* **1**, 1 (2009).
- [49] H. E. Stanley, Scaling, universality, and renormalization: Three pillars of modern critical phenomena, *Rev. Mod. Phys.* **71**, S358 (1999).
- [50] S. Hantson, S. Pueyo, and E. Chuvieco, Global fire size distribution: From power law to log-normal, *Int. J. Wildl. Fire* **25**, 403 (2016).
- [51] R. Ferrer i Cancho and R. V. Solé, Two regimes in the frequency of words and the origin of complex lexicons: Zipf's law revisited, *J. Quant. Linguist.* **8**, 165 (2001).
- [52] M. Gerlach and E. G. Altmann, Stochastic Model for The Vocabulary Growth in Natural Languages, *Phys. Rev. X* **3**, 021006 (2013).