

Chapter 2

Geospatial Analysis Requires a Different Way of Thinking: The Problem of Spatial Heterogeneity

Bin Jiang

Abstract Geospatial analysis is very much dominated by a Gaussian way of thinking, which assumes that things in the world can be characterized by a well-defined mean, i.e., things are more or less similar in size. However, this assumption is not always valid. In fact, many things in the world lack a well-defined mean, and therefore there are far more small things than large ones. This paper attempts to argue that geospatial analysis requires a different way of thinking – a Paretian way of thinking that underlies skewed distribution such as power laws, Pareto and lognormal distributions. I review two properties of spatial dependence and spatial heterogeneity, and point out that the notion of spatial heterogeneity in current spatial statistics is only used to characterize local variance of spatial dependence or regression. I subsequently argue for a broad perspective on spatial heterogeneity, and suggest it be formulated as a scaling law. I further discuss the implications of Paretian thinking and the scaling law for better understanding geographic forms and processes, in particular while facing massive amounts of social media data. In the spirit of Paretian thinking, geospatial analysis should seek to simulate geographic events and phenomena from the bottom up rather than correlations as guided by Gaussian thinking.

Keywords Big data • Scaling of geographic space • Head/tail breaks • Power laws • Heavy-tailed distributions

B. Jiang (✉)

Faculty of Engineering and Sustainable Development, Division of GIScience, University of Gävle, SE-801 76 Gävle, Sweden
e-mail: bin.jiang@hig.se

2.1 Introduction

Geospatial analysis, or spatial statistics in particular, has been dominated by a Gaussian way of thinking, which assumes that things are more or less similar in size, and can be characterized by a well-behaved mean. Based on this assumption, extremes are rare; if extremes do exist, they can be mathematically transformed into normal things (e.g., by taking logarithms or square roots). This Gaussian thinking is widespread, and has dominated the sciences for a very long time. However, Gaussian thinking has been challenged and been accused of misrepresenting our world (Mandelbrot and Hudson 2004; Taleb 2007). Indeed, many things in the world are not well behaved or lack of a well-behaved mean. This can be seen from the extreme events such as the September 11 attacks. The extent of devastation of such events was enormous and beyond any predictions and estimations. This is the same for many geographic features, which exhibit a pretty skewed or heavy-tailed distribution such as power laws and lognormal distributions. The heavy-tailed distributions imply that there are far more small geographic features than large ones, namely scaling of geographic space.

A power law distribution is often referred to as scale free, literally meaning a lack of average for characterizing the sizes of things (Barabási and Albert 1999). The power law distribution has been given different formats for it was discovered by different scientists in different disciplines over the past 100 years. Among several alternatives, Zipf's law (1949) and the Pareto distribution (Pareto 1897) are the two formats most frequently referred to in the literature. Zipf's law, with respect to city sizes, implies that there are far more small cities than large ones, while the Pareto distribution indicates that there are far more poor people than rich people, or equivalently far more ordinary people than extraordinary people. The Pareto distribution has been popularized as the 80/20 principle (Koch 1999) or the long tail theory (Anderson 2006) in the popular science and business literature. The heavy-tailed distribution, including power laws, lognormal and others similar, is what underlies the new way of thinking I want to advocate in this paper. The central argument is that geospatial analysis requires a new way of thinking radically different from Gaussian thinking, and spatial heterogeneity should be formulated as a scaling law of geography.

This is an unprecedented time when we face increasing rich geographic data sources based not only on the legacy of traditional cartography and remote sensing imagery, but also emerging from various social media such as Flickr, Twitter, and OpenStreetMap, collectively known as volunteered geographic information (Goodchild 2007). Today, one can amass gigabytes of an entire country's data for geospatial analysis and computing, for both data volumes and computing capacity have increased dramatically. However, our mindsets, subsequently our analysis methods, have been relatively slower to adapt the rapid changes (Mayer-Schonberger and Cukier 2013). For example, we tend to sample data rather than take all data for geospatial analysis; we tend to transform skewed data into "normal" by taking logarithms for example. The sampling and logarithm transformation have distorted the

underlying property of the data before the data can yield insights. The old way of thinking, Gaussian thinking, that relies on a well-defined mean to characterize geographic features, is a major barrier to achieving deep and new insights into geographic forms and processes.

Many analytical techniques have been developed, in both standard and spatial statistics, to address outliers, to measure skewness and autocorrelation, and to test significance. However, what I want to argue in this paper is not these techniques per se, but something radical in the way of thinking. Gaussian thinking, based on the assumption of independent things in a simple, static, and equilibrium world, is essentially a typical linear thinking, which implies that small cause small effect, large cause large effect, and the whole is equal to the sum of its parts. This linear thinking is a simple way of thinking guided by the reductionism philosophy, and for understanding a simple world in essence (see Sect. 2.2.2 for more details). The reader may argue that spatial statistics differs from standard statistics in spatial dependence or spatial autocorrelation. It is indeed true, but the notion of spatial dependence or autocorrelation does not help us to go beyond Gaussian thinking assumed by standard statistics, for we tend to characterize things by a well-defined mean with a limited variance. It is well recognized that geographic forms are fractal rather than Euclidean, and geographic processes are nonlinear rather than linear (Batty and Longley 1994; Chen 2009). In other words, a geographic system is a complex nonlinear world, in which there is the butterfly effect, and the whole is greater than the sum of its parts. In this paper, I attempt to argue that the Paretian way of thinking, founded on the assumption of interdependent things in a complex, dynamic, and nonequilibrium world, is more appropriate for geospatial analysis, and for better understanding geographic forms and processes. Geospatial analysis, while facing increasing amounts of social media data, should seek to uncover the underlying mechanisms through simulations from the bottom up rather than simple causality or correlations.

The remainder of this paper is organized as follows. Section 2.2 introduces, in a pedagogic manner, two distinct statistic distributions, namely Gaussian- and Paretian-like distributions, with a particular focus on the underlying ways of thinking. Section 2.3 reviews two unique properties of spatial dependence and spatial heterogeneity, and points out that the notion of spatial heterogeneity in current spatial statistics is only used to characterize local variance of spatial dependence. I therefore argue, in Sect. 2.4, that spatial heterogeneity should be formulated as a scaling law, and suggest some effective ways of detecting and revealing the scaling law and pattern for geographic features. I further discuss, in Sect. 2.5, some deep implications of Paretian thinking and the scaling law before draw a summary in Sect. 2.6.

2.2 Two Distinct Distributions and the Underlying Ways of Thinking

In this section, I first illustrate statistical differences between a homogenous Gaussian-like distribution and a heterogeneous Paretian-like distribution (Note the ‘homogenous’ is relative to the ‘heterogeneous’; see Sect. 2.2.1 for more details), using temperature and population of major US cities, and based respectively on histograms and rank-size plots. The temperature is the annual average maximum during 1981–2010, taken from the site: <http://www.prism.oregonstate.edu/products/matrix.phtml>, while the population is according to the 2010 US census. I then elaborate on the underlying ways of thinking or world views associated with the two categories of distributions.

2.2.1 Gaussian- Versus Paretian-Like Distributions

If we carefully examine two variables – temperature and population – of 720 major U.S. cities with population greater than 50,000 people, we can see that the two variables are very distinct. Although not a normal distribution, the temperature can be well characterized by its mean 20.6 (Fig. 2.1a). One can estimate a city’s temperature fairly accurate and precise based on the mean value, since the highest is 31.6, and the lowest is 9.3. In other words, the mean 20.6 is a typical temperature for US cities. The distribution that can be characterized by a well-defined mean is referred to as a Gaussian-like distribution including for example the binomial and Poisson distributions. This temperature distribution can be further assessed from the detailed statistics as shown in Table 2.1 (the temperature column). The range between the highest (31.6) and the lowest (9.3) is not very big (22.3), and the ratio of the highest to the lowest is as little as 3.4. The two measures of central tendency – mean and median – are the same. The standard deviation is

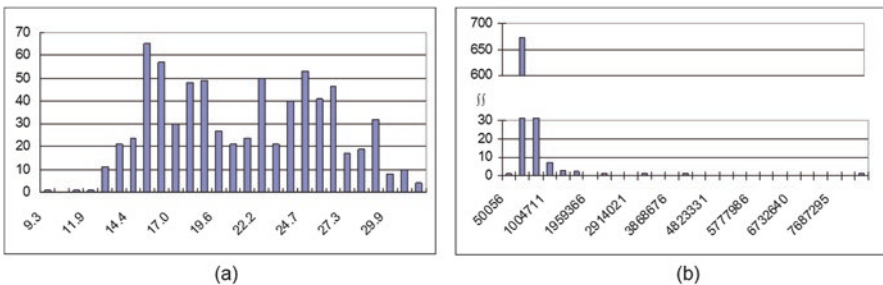


Fig. 2.1 Histograms of (a) the temperature, and (b) the population of U.S. cities (Note: the two distinct distributions indicate respectively Gaussian-like and Paretian-like distributions)

Table 2.1 Statistics about temperature and population of U.S. cities

Statistics	Temperature	Population
Minimum	9.3	50,056
Maximum	31.6	8,323,732
Range	22.3	8,273,676
Ratio	3.4	166
Mean	20.6	157,467
Median	20.6	82,115
Mode	14.9	62,820
St. Dev.	4.9	393,004

4.9, about one quarter of the range. This statistical picture of the temperature is very distinct from that of the city size or population.

The histogram of the population is extremely right skewed (Fig. 2.1b). This extreme skewness is reflected in several parameters: a wide range (8,273,676), a huge ratio (166), and a large standard deviation (393,004). In such a significantly skewed distribution, the mean of 157,467 make little sense for characterizing the population. In other words, the mean of 157,467 does not represent a typical size of the U.S. cities, since the largest city is as big as 8 millions, while the smallest city is as small as 50,000. The right skewed histogram indicates that there are far more small cities than large ones in the U.S. No wonder that the two measures of central tendency – mean and median – differ from each other significantly; refer to Table 2.1 (the population column) for more details. The standard statistics, or the histogram in particular, is little effective for describing data with a heavy-tailed distribution such as city sizes. Instead, power law based statistics, or rank-size plots in particular, should be adopted for characterizing this kind of data.

Instead of plotting temperature and population on the x-axis (as in the histograms), they are plotted on the y-axis, while the x-axis is the ranking order. This way of plotting is called rank-size plot, or rank-size distribution (Zipf 1949). The largest city (in terms of population) ranks number one, followed by the second largest, and so on. The same arrangement is made for the temperature; the highest temperature city ranks number one, followed by the second highest, and so on. The two distribution lines look very different; the temperature curve drops gradually, and then reaches quickly the minimum, while the population curve drops quickly and then gradually approaches the minimum (Fig. 2.2). Note that the red parts in the figure are those above the averages, called the head, while those below the averages, called the tail, are shown in blue. More specifically, 362 cities (approximately 50%) are above the average temperature 20.6, while only 146 cities (approximately 20%) are above the average city size 157,467. Clearly, a heavy or long tail (80% in the tail) exists for the population distribution, but a short tail (50%) for the temperature distribution. Generally, a heavy-tailed distribution possesses an inbuilt imbalance between the head and the tail (e.g., a 70/30 or 80/20 relationship). This imbalance indicates a nonlinear relationship between the head and the tail. Such an inbuilt

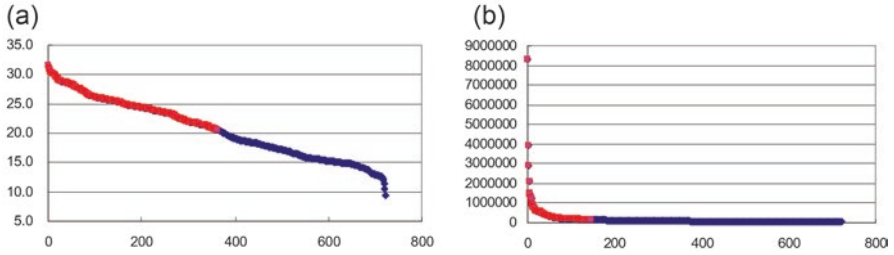


Fig. 2.2 Rank-size plots of (a) the temperature, and (b) the population of the U.S. cities (Note: values above and below the averages are respectively in red and blue; clearly, there is a short head and a long tail for the population, forming an unbalanced contrast, while the values above and below the averages are more or less the same for the temperature)

Table 2.2 Comparison between the two ways of thinking

Gaussian thinking	Paretian thinking
With a mean (or scale)	Without a mean (or scale-free)
Static	Dynamic
Simple	Complex
Equilibrium	Non-equilibrium
Linear	Nonlinear
Predictable	Unpredictable

imbalance, or nonlinearity, is clearly missing in a Gaussian-like distribution with a well-balanced relationship between the head and the tail (e.g., 50/50).

2.2.2 The Underlying Ways of Thinking

The differences between the two distributions lie fundamentally in different ways of thinking, or different ways of viewing the world, rather than different techniques associated with each distribution. Technically, data with a Paretian-like distribution can be easily transformed into a Gaussian-like distribution, e.g., by taking logarithms. Gaussian thinking implies more or less similar things in a simple, static, and equilibrium world, while Paretian thinking believes in far more small things than large ones in a dynamic, complex, and nonequilibrium world (McKelvey and Andriani 2005; see Table 2.2). Standard statistics teaches us that if the probability of an event is small, then the event occurs rarely. The event can be considered an outlier that is literally distant from the rest of the data. However, in Paretian thinking, an event of small probability, or the highly improbable, has a significant impact (e.g., the September 11 attacks) and thus be ranked highly.

In Gaussian thinking, the world does not change much, and all changes occur around a stable and well-defined mean. Thus, the presumed Gaussian world is static,

simple, linear, and predictable. The Newtonian physics is sufficient to understand and deal with the Gaussian world. Why does such a predictable world exist? Such a world reflects a lack of interaction and competition among individual agents; every agent acts independently without influences upon or affects from others. This assumption is fundamental to standard statistics, and of course appropriate for many events in the world like human heights. Spatial statistics differentiate it from standard statistics in spatial dependence, but it does not change fundamentally the underlying way of thinking – Gaussian thinking with a well-defined mean for characterizing things. On the other hand, in Paretian thinking, the world is full of surprises, and changes are often dramatic and unexpected. Thus, there is no stable and well-defined mean for characterizing the surprises and changes. The presumed Paretian world is essentially dynamic, complex, nonlinear, and unpredictable. This unpredictable world is founded on the assumption that everything is related to, or interdependent with, everything else. This interdependence assumption implies that cooperation and competition would eventually lead to unbalanced results characterized by a long-tail distribution (c.f. Sect. 2.3 for more discussions).

Nature is awash with phenomena such as trees, rivers, mountains, clouds, coastlines, and earthquakes that exhibit power laws or heavy-tailed distributions in general (Mandelbrot 1982; Schroeder 1991; Bak 1996). Accordingly, power law has been formulated as a fundamental law in various disciplines such as physics, biology, economics, computer science, and linguistics. People’s daily activities are also governed by power laws (Barabási 2010), indicating bursty behaviors of human mobility or activities in general. Power laws are a signature of complex systems that are evolved in nonlinear manners, i.e., small causes often have disproportional large effects. For instance, the top 10% of the most connected streets account for 90% of traffic flows (Jiang 2009). In a 21-block area of Philadelphia, 70% of the marriages occurred between people who lived no more than 30% of that distance apart (Zipf 1949).

The examination of the two ways of thinking suggests that Paretian-like distribution, or Paretian thinking in general, appears more appropriate for understanding geographic forms and processes, for dependence is a key property of spatial statistics (c.f., Sect. 2.3 for more details). In spite of spatial dependence being its key property, spatial statistics is still unfortunately very much dominated by Gaussian thinking. The very notion of spatial heterogeneity refers to local variance of spatial dependence, but from global to local, or from one single correlation coefficient to multiple coefficients (c.f., Sect. 2.3 for more details). In the remainder of this paper, we review two spatial properties of dependence and heterogeneity, and argue that spatial heterogeneity is ubiquitous, and it should be formulated as a scaling law. And we further discuss some deep implications of the scaling law and Paretian thinking for better understanding of geographic forms and processes in the era of big data.

2.3 Spatial Properties of Dependence and Heterogeneity

It is well known that in contrast to the independence assumption of standard statistics, geographic phenomena or events are not random or independent. Geographic events are more likely to occur in some locations than others (spatial heterogeneity), and nearby events are more similar than distant events (spatial dependence). Both spatial heterogeneity and spatial dependence are referred to as spatial properties, indicating respectively that geographic events are related to their locations and to their neighboring events. Spatial dependence is widely known or formulated as the first law of geography: “*Everything is related to everything else, but near things are more related than distant things*” (Tobler 1970). For example, your housing price is likely to be similar (positive correlation) to those of your neighbors. Similarly, the elevations of two locations 10 m apart are likely to be more similar than the elevations of two locations of 100 m apart. Note that “likely” indicates a statistical rather than a deterministic property; one can always find exceptions in statistical trends.

Spatial heterogeneity refers to no average location that can characterize the Earth’s surface (Anselin 1989; Goodchild 2004). This is indeed true, while for example referring to the diversity of landscapes and species (animals and plants) on the Earth’s surface (Wu and Li 2006; Bonner 2006). This diversity or heterogeneity indicates uneven geographic and statistical distributions involving both landscapes and species – that is, a mix of concentrations of multiple species (biological), terrain formations (geological), environmental characteristics (such as rainfall, temperature, and wind) on the one hand, and various concentrations of various types of species on the other. A variety of habitats such as different topographies, soil types, and climates can accommodate a greater number of species. These are the natural environments of the Earth’s surface. Spatial heterogeneity in geography also concerns human-made built environments created by human activities such as industrialization and urbanization, and in particular, for example, the diversity of human settlements or cities in particular. Given the diversity or spatial heterogeneity of the Earth’s surface, homogeneous Gaussian-like distribution is unlikely to be the right means to characterize complex geographic features.

Spatial dependence and spatial heterogeneity are properties to spatial data and geospatial analysis (Anselin 1989; Griffith 2003), and probably the two most important principles of geographic information science (GIScience). Goodchild (2004) has been a key advocator for formulating general principles for GIScience. On several occasions, he has made insightful remarks on spatial heterogeneity or spatial properties in general. His definition of spatial heterogeneity as “no average location” is in effect the notion of scale-free used to characterize things that exhibit a power law or heavy-tailed distribution (Barabási and Albert 1999). On the other hand, he stated, with respect to spatial heterogeneity, that all locations are unique, due to which geography might be better considered as an idiographic science, studying the unique properties of places (Goodchild 2004). However, I argue, in contrast to Goodchild, that spatial heterogeneity makes geography a nomothetic science. This is because spatial heterogeneity itself is a law – the scaling law, implying that

there are far more small geographic features than large ones. Spatial heterogeneity is a kind of hidden order, which appears disordered on the surface, but possesses a deep order beneath. This kind of hidden order can be characterized by a power law or a heavy-tailed distribution in general.

Current spatial statistics suffers from what I call ‘spatial heterogeneity paradox’. Spatial heterogeneity is defined as no average location, but we tend to use a well-defined mean or average to characterize locations. This paradox implies that our mindsets are still constrained by Gaussian thinking. The current notion of spatial heterogeneity refers to local variance of spatial dependence or regression. This can be seen from the development of local spatial statistics and local statistical models that initially brought spatial heterogeneity into spatial statistics (Anselin 1989). Local spatial statistics concern local variants of spatial autocorrelation or regression, a measure to spatial dependence, including, for example, the local statistical models (Getis and Ord 1992), the LISA techniques (Anselin 1995), and geographically-weighted regression (Fotheringham et al. 2002). The shifting perspective of spatial autocorrelation from global to local brings new insights into spatial dependence, or the heterogeneity of spatial dependence. However, all these techniques and models are essentially based on Gaussian statistics, using a well-defined mean with a limited variance. To paraphrase Mandelbrot (Mandelbrot and Hudson 2004), spatial heterogeneity refers to ‘wild’ variances, but Gaussian-like distribution can only characterize ‘mild’ variances.

Human activities are the major forces behind spatial heterogeneity in the built environments. While carrying out activities, human beings (and their interventions) must respect the spatial heterogeneity of Nature – that is, harmonize with rather than damage the natural environments. Geographic information concerning urban and human geography captures essentially spatial variations of the built environments, which demonstrate ‘wild’ heterogeneity as well. For example, Zipf’s law on city sizes (Zipf 1949) mainly concerns such a spatial variation. Thus, I argue, in contrast to the conventional view, that dependence, or more precisely interdependence, is a first-order effect, while heterogeneity is a second-order effect. Let us do a thought experiment. Imagine that once upon a time, there were no cities, only scattered villages. Over time, large cities gradually emerge through the interactions of villages, so do mega cities through the interactions of cities. The interactions (competition and cooperation) of villages and cities are actually those of people acting individually and/or collectively. These interactions are what we mean by dependence and interdependence. Eventually, there are far more small cities than large ones through for example the mechanism of “the rich get richer.” This observation is the same for the wealth distribution among individuals in a country; far more poor people than rich people, or far more ordinary people than extraordinary people (Epstein and Axtell 1996). The interactions among people and cities reflect the interdependence effect in the formation and evolution of cities and city systems, and the built environments in general.

2.4 Spatial Heterogeneity as a Scaling Law

The subtitle of this paper ‘*The Problem of Spatial Heterogeneity*’ is an homage to the classic work ‘*The Problem of Spatial Autocorrelation*’ (Cliff and Ord 1969), which popularized the concept of spatial dependence. Similarly, spatial heterogeneity under Gaussian thinking is indeed a problem because the Earth’s surface cannot be characterized by a well-defined mean. However, in the Paretian way of thinking, spatial heterogeneity is not a problem, but the norm. Spatial heterogeneity should be formulated as a scaling law in geography.

2.4.1 *Ubiquity of the Scaling Law in Geography*

Geographic features are unevenly or abnormally distributed, so the scaling pattern of far more small things than large ones is widespread in geography (Pumain 2006). The scaling pattern has another name called fractal (Mandelbrot 1982). Fractal-related research in geography has concentrated too much on concepts such as fractal dimension and self-similarity. In fact, the scaling law is fundamental to all of these concepts. In this regard, Salingaros and West (1999) formulated a universal rule for city artifacts; there are far more small city artifacts than large ones, due to which the image of the city can be formed in human minds (Jiang 2013b). With the increasing availability of geographic information, the scaling law has been observed and examined in a wide range of geographic phenomena including, for example, street lengths and connectivity (Carvalho and Penn 2004; Jiang 2009), building heights (Batty et al. 2008), street blocks (Lämmer 2006; Jiang and Liu 2012), population densities (Schaefer and Mahoney 2003; Kyriakidou et al. 2011), and airport sizes and connectivity (Guimerà et al. 2005). Interestingly, the scaling of geographic space has had an enormous effect on human activities; human activities and interactions in geographic space exhibit power law distributions as well (Brockmann et al. 2006; Gonzalez et al. 2008; Jiang et al. 2009). Table 2.3 provides a synoptic view of the ubiquity of power laws in geography, noting that the references listed are non-exhaustive, but for example only.

Despite its ubiquity, ironically the scaling law, or the Paretian way of thinking in general, has not been well received in geospatial analysis as elaborated earlier in the text. Current geospatial analysis adopts a well-defined mean or average to characterize spatial heterogeneity. The two closely related concepts of scale and scaling must be comprehended together, i.e., many different scales, ranging from the smallest to the largest, form a scaling hierarchy. This comprehension should be added, as a fourth one, into the three meanings of scale in geography: cartographic, analysis, and phenomenon (Montello 2001); see more elaborations in this recent paper (Jiang and Brandt 2016). The essence of power laws is the scaling pattern, in which there are far more small scales than large ones. This scaling pattern reflects the true picture of spatial heterogeneity or that of the Earth’s surface.

Table 2.3 Power laws in geographic features or phenomena

Geographic phenomena	References (for example)
City sizes	Zipf (1949), Krugman (1996), and Jiang and Jia (2011)
Fractals in cities or geographic space	Goodchild and Mark (1987) and Batty and Longley (1994)
Coast lines and mountains	Mandelbrot (1967) and Bak (1996)
Hydrological networks	Hack (1957), Horton (1945), Maritan et al. (1996), and Pelletier (1999)
Urban and architectural space	Salingaros and West (1999)
Street lengths and connectivity	Carvalho and Penn (2004) and Jiang (2009)
Building heights	Batty et al. (2008)
Street blocks	Lämmer (2006) and Jiang and Liu (2012)
Population density	Schaefer and Mahoney (2003) and Kyriakidou et al. (2011)
Airport sizes and connectivity	Guimerà et al. (2005)
Human mobility	Brockmann et al. (2006), Gonzalez et al. (2008), and Jiang et al. (2009)

2.4.2 Detecting the Scaling Law

What were claimed to be power laws in the literature could be actually lognormal, exponential, or other similar distributions, because the detection of power laws can be very tricky. Given a power law relationship $y=x^a$, it can be transformed into the logarithm scales, i.e., $\ln(y)=a \cdot \ln(x)$, indicating that the logarithms of the two variables x and y have a linear relationship. Conventionally, an ordinary least squares (OLS) based method was widely used for the detection. In the fractal literature, the box-counting method is usually used to compute the fractal dimension, which is the de facto power law exponent, and its computation is also based on OLS. There are at least two issues surrounding the power law detection. The first is that the OLS based method is found to be less reliable for detecting a power law, so a maximum likelihood method has been developed (Clauset et al. 2009). It was found that many claims on power laws in the literature are likely to be lognormal or other degenerated formats such as a power law with an exponential cutoff. For the sake of readability, this paper does not cover mathematical details on heavy-tailed distributions and their detection; interested readers can refer to Clauset et al. (2009) and the references therein. The second is that even with the OLS-based method, the definition of fractal dimension is so strict that many geographic features are excluded from being fractal (Jiang and Yin 2014). Given the circumstance, the authors have recently provided a rather relaxed definition of fractals, i.e., a geographic feature is fractal if and only if the scaling pattern of far more small things than large ones recurs multiple times. The number of times plus one is referred to as ht-index (Jiang and Yin 2014), an alternative index of fractal dimension for characterizing complexity of fractals or geographic features in particular.

The idea behind the relaxed definition of fractals, or the ht-index, is pretty simple and straightforward. It is based on the head/tail breaks (Jiang 2013a), a new classification scheme for data with a heavy-tailed distribution. Given a variable whose distribution is right skewed, compute its arithmetic mean, and subsequently split its values into two unbalanced parts: those above the mean in the head, and those below the mean in the tail. The values above the mean are a minority, while the values below are a majority. The ranking and breaking process continues for the head part progressively and iteratively until the values in the head no longer meet the condition of far more small things than large ones. This way both the number of classes and the class intervals are naturally and automatically derived based on the inherent hierarchy of data. Eventually, the number of classes, or equivalently the ht-index, indicates hierarchical levels of the values. One can simply rely on an Excel sheet for the computation of the ht-index. As an example, Fig. 2.3 illustrates the scaling pattern of the US cities, discussed earlier in Sect. 2.2.1, and it has the ht-index of 7.

2.4.3 *Revealing the Scaling Pattern*

The head/tail breaks can effectively reveal or visualize the scaling pattern if the data itself exhibits a heavy-tailed distribution. This is because the head/tail breaks was developed initially for revealing the inherent scaling hierarchy or the scaling pattern. In this regard, conventional classification methods, mainly guided by Gaussian thinking, failed to reveal the scaling pattern. For example, the most widely used classification natural breaks (Jenks 1967), which is set as a default in ArcGIS, is based on the principle of minimizing within-classes variance, and maximizing between-classes variance. It sounds very natural. In some case like the US cities, the classification result of the natural breaks may look very similar to the one by natural breaks (Fig. 2.3). However, this is just by chance. Essentially, the natural breaks is motivated by Gaussian thinking; each class is characterized by a well-defined mean with a limited or minimized variance. In a contrast, the head/tail breaks is motivated by Paretian thinking, and for data with a Paretian-like or heavy-tailed distribution. The iteratively or recursively defined averages are used as meaningful cutoffs for differentiating hierarchical levels.

The reader probably has got used to the US terrain surface (Fig. 2.4a), which is based on the natural breaks. It is commonly seen in geography and cartography textbooks and atlases. However, I want to challenge this conventional wisdom, arguing that the natural breaks based visualization is little natural. I contend that the head/tail breaks derived visualization is more natural, since it reflects the underlying scaling pattern of far more small things than large ones (Fig. 2.4b). The things here are referred to individual locations, or more specifically, far more low locations than high locations. The left visualization, which distorted the scaling pattern, appears having far more high locations than the visualization to the right, or equivalently far

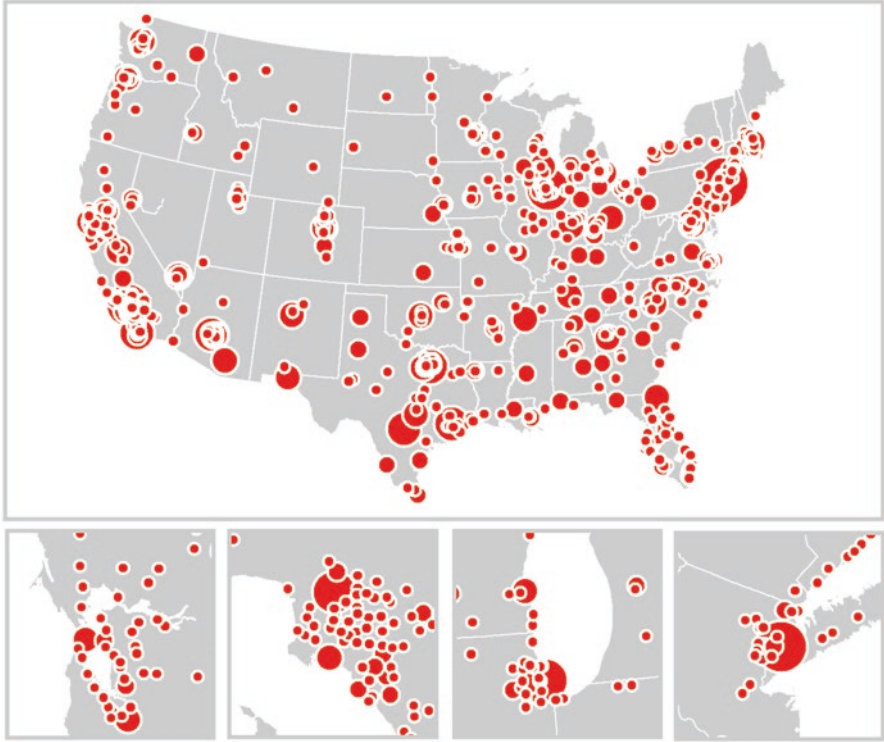


Fig. 2.3 Scaling pattern of US cities with ht-index equal 7 (Note: the four *insets* from the *left* to the *right* provide the enlarged view respectively for San Francisco, Los Angeles, Chicago and New York regions)

more high locations than what it actually has. The visualization to the right reflects well the underlying scaling pattern. This can be further seen from the corresponding histograms of the individual classes of the two classifications (Fig. 2.4c, d). What is illustrated by the left histogram is “more low locations than high ones” which is a linear relationship, rather than “far more low locations than high ones”, which is a nonlinear relationship. For the left histogram, each pair of the adjacent bars from left to right does not constitute an unbalanced contrast of majority versus minority. For example, the first pair of bars of the left histogram shows a well-balanced contrast of 7–6; in a contrast, the first pair of the right histogram is unbalanced, 14–5. Therefore, the right histogram indicates clearly “far more low locations than high ones”. Interestingly, the scaling pattern remains unchanged with respect to different scales of digital elevation models (Lin 2013).

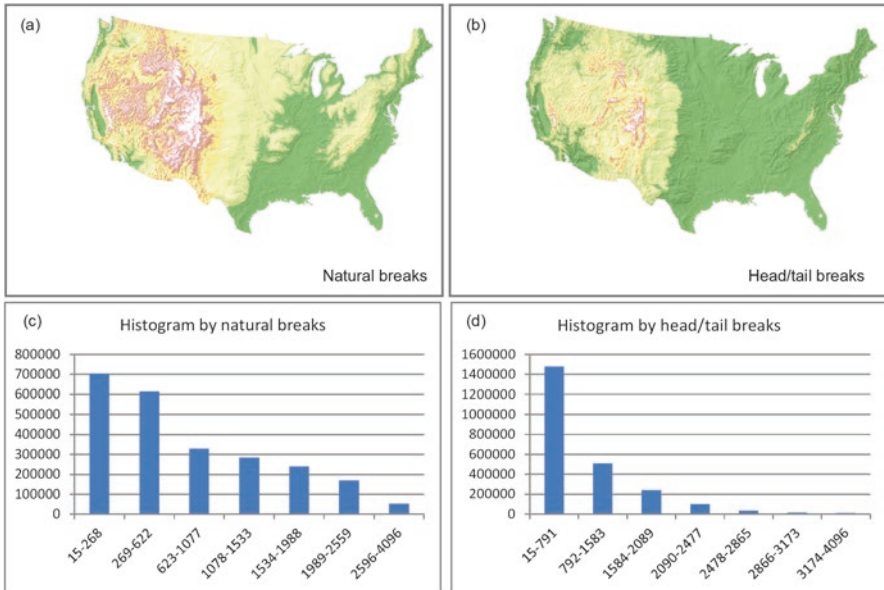


Fig. 2.4 The scaling pattern of US terrain surface is distorted by the natural breaks, but revealed by the head/tail breaks

2.5 Implications of Paretian Thinking and the Scaling Law

Current geospatial analysis concentrates more on geographic forms, but less on why the forms. The forms illustrated are mostly limited to whether they are random, or to what extent they are auto-correlated. As to why the forms, it is usually ended up with simple regressions and causalities. This way of geospatial analysis is much like short-term weather forecasting. Despite its usefulness, the short-term weather forecast adds little to understanding the complex behavior of weather – the long-term weather in general, is unpredictable, just like earthquakes and many other events in Nature and society (Bak 1996). If the real world is unpredictable, what can we do as scientists? We can simulate interactions of things from the bottom up in order to understand the underlying mechanisms, which would help improve predictions. In this regard, the emerging social media, in particular location-based social media, provide valuable data for validating the simulation results (Jiang and Miao 2015). The data, unlike traditional statistical or census data that are mainly aggregated, are not only big in size, but are collected at individual levels. The data are not only at individual levels, but linked in time and among individuals. The data can help track the trajectories of individuals and their associations in space and over time. For this kind of social media data, the scaling law and fractals should be the norm.

Current spatial statistics constrained by Gaussian thinking show critical limitations for analyzing or getting insights into big data (Mayer-Schonberger and Cukier

2013). What are illustrated by spatial statistics, either patterns or associations, can be compared to the mental images of the elephant in the minds of the blind men. These images reflect local truths, and are indeed correct partially, but they did not reflect the whole of the elephant. Geospatial analysis should go beyond illustrating spatial autocorrelation, either globally or locally, but towards uncovering the underlying scaling or fractal patterns. Geographic features are essentially and ultimately scaling or fractal. Therefore, any patterns deviating from the scaling pattern or that can be characterized by a well-defined mean are either wrong or biased.

Geographic forms (or phenomena) are not the outcomes of simple processes but the results of complex processes with positive feedbacks. In the built environments, human interventions (interdependence and interactions) of various kinds are the major effects of spatial heterogeneity. As famously stated by Winston Churchill (1874–1965), “*we shape our buildings, and thereafter they shape us*”. This statement should be comprehended in a progressive and recursive manner. This comprehension, which underlies Paretian thinking, is essentially a complex system perspective for exploring the underlying processes related to geographic forms (e.g., Benguigui and Czamanski 2004; Blumenfeld-Lieberthal and Portugali 2010). In this regard, complexity science has developed a range of tools such as discrete models, complex networks, scaling hierarchy, fractal geometry, self-organized criticality, and chaos theory (Newman 2011). All these modeling tools attempt to reveal the underlying mechanisms, linking surface complex forms (or complexity) to the underlying mechanisms (or deep simplicity) through simulations from the bottom up, rather than simple descriptions of forms or of geographic forms in particular.

Paretian thinking represents a paradigm shift. Shifting from the sands to the avalanches (Bak 1996), and from the street segments to the natural streets (Jiang 2009), enables us to see something interesting and exciting, i.e., from the things of limited sizes to the things of all sizes. The things of all sizes imply a scaling pattern across all scales. Recognition of the scaling pattern helps us to better understand the underlying universal form of geographic features. This scaling pattern can further be linked to the underlying geographic processes that are dynamic, nonlinear, and bottom-up in nature. This view would position geography in the family of science, since we geographers are interested in not only what things look like (the forms) but also why things look that way (the processes). Spatial heterogeneity is thus not a problem but an underlying scaling law of geography.

2.6 Concluding Summary

This paper argues that geospatial analysis requires a different way of thinking, or world view in general, that underlies the Paretian-like distribution of geographic features. We put the two distinct views in comparison: more or less similar things in a simple, static, and equilibrium world on the one hand, and far more small things than large ones in a complex, dynamic, and non-equilibrium world on the other. Geospatial analysis has been dominated by Gaussian statistics with a well-defined

mean for characterizing spatial variation ('mild' variance so to speak). Despite its ubiquity in geography, the Paretian-like heavy-tailed distribution, or the underlying way of thinking in general, has not been well received in geospatial analysis. The current geospatial analysis mainly focuses on how spatial variation deviates from a random pattern, and measuring spatial auto-correlation from global to local (the current spatial heterogeneity), but leaves the underlying processes unexplored. This way of geospatial analysis is inadequate for understanding geographic forms and processes, in particular while facing increasing amounts of social media data.

No average location exists on the Earth's surface. Instead, there are far more small things than large ones in geographic space; small things are a majority while large things are a minority. Importantly, the pattern of far more small things than large ones recurs multiple times (Jiang and Yin 2014). This recurring scaling pattern reflects the true image of spatial heterogeneity that lacks a well-defined mean ('wild' variance so to speak). Spatial heterogeneity is indeed a problem in Gaussian thinking, but it is a law or scaling law in Paretian thinking. In the spirit of Paretian thinking and the scaling law, geospatial analysis should seek to simulate individuals and individual interactions from the bottom up rather than simple correlations and causalities. In this connection, complexity tools such as complex networks, agent-based modeling, and fractal/scaling provide effective means for geospatial analysis of complex geographic phenomena.

Acknowledgments This chapter is a reprint of original paper by Jiang (2015) with the permission of the publisher. The author would like to thank the anonymous referees and the editor Daniel Z. Sui for their valuable comments. However, any shortcoming remains the responsibility of the author.

References

- Anderson C (2006) *The Long tail: why the future of business is selling less of more*. Hyperion, New York
- Anselin L (1989) What is special about spatial data: alternative perspectives on spatial data analysis. National Center for Geographic Information and Analysis, Santa Barbara
- Anselin L (1995) Local indicators of spatial association – LISA. *Geogr Anal* 27:93–115
- Bak P (1996) *How nature works: the science of self-organized criticality*. Springer, New York
- Barabási A (2010) *Bursts: the hidden pattern behind everything we do*. Dutton Adult, Boston
- Barabási A-L, Albert R (1999) Emergence of scaling in random networks. *Science* 286:509–512
- Batty M, Longley P (1994) *Fractal cities: a geometry of form and function*. Academic, London
- Batty M, Carvalho R, Hudson-Smith A, Milton R, Smith D, Steadman P (2008) Scaling and allometry in the building geometries of Greater London. *Eur Phys J B* 63:303–314
- Benguigui L, Czamanski D (2004) Simulation analysis of the fractality of cities. *Geogr Anal* 36(1):69–84
- Blumenfeld-Lieberthal E, Portugali J (2010) Network cities: a complexity-network approach to urban dynamics and development. In: Jiang B, Yao X (eds) *Geospatial analysis of urban structure and dynamics*. Springer, Berlin, pp 77–90
- Bonner JT (2006) *Why size matters: from bacteria to blue whales*. Princeton University Press, Princeton

- Brockmann D, Hufnagel L, Geisel T (2006) The scaling laws of human travel. *Nature* 439:462–465
- Carvalho R, Penn A (2004) Scaling and universality in the micro-structure of urban space. *Phys A* 332:539–547
- Chen Y (2009) Spatial interaction creates period-doubling bifurcation and chaos of urbanization. *Chaos, Solitons Fractals* 42(3):1316–1325
- Clauset A, Shalizi CR, Newman MEJ (2009) Power-law distributions in empirical data. *SIAM Rev* 51:661–703
- Cliff AD, Ord JK (1969) The problem of spatial autocorrelation. In: Scott AJ (ed) *London papers in regional science*. Pion, London, pp 25–55
- Epstein JM, Axtell R (1996) *Growing artificial societies: social science from the bottom up*. Brookings Institution Press, Washington, DC
- Fotheringham AS, Brunson C, Charlton M (2002) *Geographically weighted regression: the analysis of spatially varying relationships*. Wiley, Chichester
- Getis A, Ord JK (1992) The analysis of spatial association by distance statistics. *Geogr Anal* 24:189–206
- Gonzalez M, Hidalgo CA, Barabási A-L (2008) Understanding individual human mobility patterns. *Nature* 453:779–782
- Goodchild M (2004) The validity and usefulness of laws in geographic information science and geography. *Ann Assoc Am Geogr* 94(2):300–303
- Goodchild MF (2007) Citizens as sensors: the world of volunteered geography. *GeoJournal* 69(4):211–221
- Goodchild MF, Mark DM (1987) The fractal nature of geographic phenomena. *Ann Assoc Am Geogr* 77(2):265–278
- Griffith DA (2003) *Spatial autocorrelation and spatial filtering: gaining understanding through theory and scientific visualization*. Springer, Berlin
- Guimerà R, Mossa S, Turtschi A, Amaral LAN (2005) The worldwide air transportation network: anomalous centrality, community structure, and cities' global roles. *Proc Natl Acad Sci U S A* 102(22):7794–7799
- Hack J (1957) *Studies of longitudinal stream profiles in Virginia and Maryland*. US Geol Surv Prof Pap:294-B
- Horton RE (1945) Erosional development of streams and their drainage basins: hydrological approach to quantitative morphology. *Bull Geogr Soc Am* 56:275–370
- Jenks GF (1967) The data model concept in statistical mapping. *Int Yearb Cartogr* 7:186–190
- Jiang B (2009) Street hierarchies: a minority of streets account for a majority of traffic flow. *Int J Geogr Inf Sci* 23(8):1033–1048
- Jiang B (2013a) Head/tail breaks: a new classification scheme for data with a heavy-tailed distribution. *Prof Geogr* 65(3):482–494
- Jiang B (2013b) The image of the city out of the underlying scaling of city artifacts or locations. *Ann Assoc Am Geogr* 103(6):1552–1566
- Jiang B (2015) Geospatial analysis requires a different way of thinking: the problem of spatial heterogeneity. *GeoJournal* 80(1):1–13
- Jiang B, Brandt A (2016) A fractal perspective on scale in geography. *ISPRS Int J Geo-Inf* 5(6):95. doi:[10.3390/ijgi5060095](https://doi.org/10.3390/ijgi5060095)
- 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(8):1269–1281
- Jiang B, Liu X (2012) Scaling of geographic space from the perspective of city and field blocks and using volunteered geographic information. *Int J Geogr Inf Sci* 26(2):215–229
- Jiang B, Miao Y (2015) The evolution of natural cities from the perspective of location-based social media. *Prof Geogr* 67(2):295–306
- Jiang B, Yin J (2014) Ht-index for quantifying the fractal or scaling structure of geographic features. *Ann Assoc Am Geogr* 104(3):530–541
- Jiang B, Yin J, Zhao S (2009) Characterizing human mobility patterns in a large street network. *Phys Rev E* 80:021136

- Koch R (1999) *The 80/20 principle: the secret to achieving more with less*. Crown Business, New York
- Krugman P (1996) *The self-organizing economy*. Blackwell, Cambridge, MA
- Kyriakidou V, Michalakelis C, Varoutas D (2011) Applying Zipf's power law over population density and growth as network deployment indicator. *J Serv Sci Manag* 4:132–140
- Lämmer S, Gehlsen B, Helbing D (2006) Scaling laws in the spatial structure of urban road networks. *Phys A* 363(1):89–95
- Lin Y (2013) *A comparison study on natural and head/tail breaks involving digital elevation models*. Bachelor thesis at University of Gävle, Sweden
- Mandelbrot B (1967) How long is the coast of Britain? Statistical self-similarity and fractional dimension. *Science* 156(3775):636–638
- Mandelbrot BB (1982) *The fractal geometry of nature*. W. H. Freeman, San Francisco
- Mandelbrot BB, Hudson RL (2004) *The (mis)behavior of markets: a fractal view of risk, ruin and reward*. Basic Books, New York
- Maritan A, Rinaldo A, Rigon R, Giacometti A, Rodríguez-Iturbe I (1996) Scaling laws for river networks. *Phys Rev E* E53:1510–1515
- Mayer-Schonberger V, Cukier K (2013) *Big data: a revolution that will transform how we live, work, and think*. Eamon Dolan/Houghton Mifflin Harcourt, New York
- McKelvey B, Andriani P (2005) Why Gaussian statistics are mostly wrong for strategic organization. *Strateg Organ* 3(2):219–228
- Montello DR (2001) Scale in geography. In: Smelser NJ, Baltes (eds) *International encyclopedia of the social & behavioral sciences*. Pergamon Press, Oxford, pp 13501–13504
- Newman M (2011) Resource letter CS-1: complex systems. *Am J Phys* 79:800–810
- Pareto V (1897) *Cours d'économie politique*. Rouge, Lausanne
- Pelletier JD (1999) Self-organization and scaling relationships of evolving river networks. *J Geophys Res* 104:7359–7375
- Pumain D (2006) *Hierarchy in natural and social sciences*. Springer, Dordrecht
- Salingaros NA, West BJ (1999) A universal rule for the distribution of sizes. *Environ Plan B Plan Des* 26:909–923
- Schaefer JA, Mahoney AP (2003) Spatial and temporal scaling of population density and animal movement: a power law approach. *Ecoscience* 10(4):496–501
- Schroeder M (1991) *Chaos, fractals, power laws: minutes from an infinite paradise*. Freeman, New York
- Taleb NN (2007) *The black swan: the impact of the highly improbable*. Allen Lane, London
- Tobler W (1970) A computer movie simulating urban growth in the Detroit region. *Econ Geogr* 46(2):234–240
- Wu J, Li H (2006) Concepts of scale and scaling. In: Wu J, Jones KB, Li H, Loucks OL (eds) *Scaling and uncertainty analysis in ecology*. Springer, Berlin, pp 3–15
- Zipf GK (1949) *Human behavior and the principles of least effort*. Addison Wesley, Cambridge, MA