

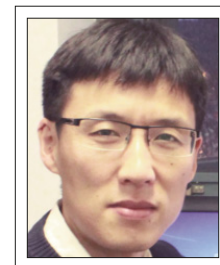
AN ACROSS-COUNTRY COMPARISON OF THE HIERARCHICAL SPATIAL STRUCTURES OF CITIES

Xintao Liu^a, Yifang Ban^b, Songnian Li^a

^aDepartment of Civil Engineering, Ryerson University, Toronto, Ontario, Canada

^bDivision of Geoinformatics, Department of Urban Planning and Environment,
Royal Institute of Technology, Sweden

This paper investigates the hierarchical structures of 29 selected European countries from the perspective of blocks and natural cities and makes an across-country comparison among the countries. Blocks are minimum cycles consisting of road segments in the road network of a whole country; natural cities are defined as the aggregations of small blocks. We test the size distributions of blocks and natural cities at the country level and find that both exhibit heavy-tailed distributions. The power law distribution of city sizes indicates the presence of the scaling property. Therefore, the cities in a country can be repeatedly grouped into a similar two-tier structure of head and tail via the head/tail division rule. The ascending tiers represent the small, medium, large and mega cities. Accordingly, a simple model is developed to evaluate and cross compare the degree of similarity and stability of the scaling properties and hierarchical structures of cities. Moreover, cities and blocks are the functional units of a country, and the correlation coefficient values between city sizes/number of blocks and economic factors (i.e., gross domestic product and population) are up to 0.87. We further conjecture that the compared results of hierarchies can serve as an indicator to assess a country's economic system.

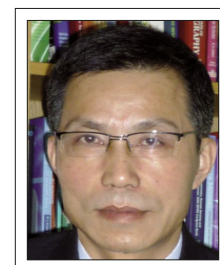


Xintao Liu
xintao@kth.se

Le présent article examine les structures hiérarchiques de vingt-neuf pays européens sélectionnés sous l'angle de blocs et de villes naturelles et fait une comparaison entre les pays. Les blocs sont des cycles minimaux formés de segments de routes dans le réseau routier d'un seul pays; les villes naturelles sont définies comme étant des regroupements de petits blocs. Nous avons vérifié la distribution granulométrique des blocs et des villes naturelles au niveau des pays et nous avons constaté que les deux présentent des distributions à queue lourde. La distribution, selon la loi de puissance, de la taille des villes indique la présence de la propriété de mise à l'échelle. Par conséquent, les villes d'un pays peuvent être groupées itérativement dans une structure similaire à deux niveaux de tête et de queue par le truchement de la règle de division tête/queue. Les niveaux ascendants représentent les petites villes, les villes moyennes, les grandes villes et les villes géantes. Par conséquent, un modèle simple est développé pour évaluer et faire une comparaison croisée du degré de similarité et de stabilité des propriétés de mise à l'échelle et des structures hiérarchiques des villes. En outre, les villes et les blocs sont les unités fonctionnelles d'un pays et les valeurs du coefficient de corrélation entre la taille des villes et le nombre de blocs et les facteurs économiques (c.-à-d., produit intérieur brut et population) peuvent atteindre jusqu'à 0,87. Nous émettons de plus l'hypothèse que les résultats comparés des hiérarchies peuvent servir d'indicateurs pour évaluer le système économique d'un pays.



Yifang Ban



Songnian Li

1. Introduction

Scaling refers to the special non-linear, mathematical relationship between sizes of entities and their probabilities (i.e., a heavy-tailed distribution, especially the power law in statistical physics) in a complex system. Extensive literature has been dedicated to exploring the scaling properties of urban systems from different perspectives, such as urban morphologies, road networks, cellular structures,

urban economies and city sizes (e.g., *Batty and Longley* [1994], *Chen and Zhou* [2003], *Pumain* [2004], *Lämmer et al.* [2006], *Kalapala et al.* [2006], *Shalizi* [2011]), while other studies have emphasized modeling the dynamics, mechanisms and fundamental drivers behind the scaling laws in urban systems (e.g., *Alperovich* [1993], *Blanter et al.* [1997], *Pumain et al.* [2006], *Isalgue et al.*

[2007], *Barthelemy and Flammini* [2008], *Chen and Zhou* [2008]). These studies agree on the fact that the urban or geographic systems exhibit scaling laws and bear hierarchical structures. The shared feature of the distributions (i.e., scaling) describes the division of inner elements into groups [*Adamic* 2011] that constitute what we call hierarchies. The smaller inner elements are interactive components with physical meanings, and the hierarchies indicate different geographic implications in different environmental contexts. Nevertheless, most of these previous studies were limited to the description of scaling phenomena, and a few of the studies attempted to develop quantitative approaches to obtain hierarchical structures for application in urban studies.

Jiang and Liu [2011] illustrated the scaling phenomenon by showing that there are more numerous small geographic entities than large entities from the perspective of blocks (i.e., the minimum cycles decomposed from a road network). The authors also proposed a quantitative way to differentiate the hierarchical structures: if the sizes of all entities (e.g., blocks) follow a heavy-tailed distribution, then the mean value clearly divides them into a high percentage of small entities and a low percentage of large entities. This regularity is termed as the head/tail division rule, which has also been successfully applied in several urban studies (e.g., *Jiang et al.* [2011], *Jiang and Jia* [2011]). However, blocks are generated only in the three largest European countries (France, Germany and the UK), and the hierarchical structures of blocks are identified only at the first level to delineate natural cities in a country from a geospatial perspective. Comparisons among countries are not conducted due to the small number of countries.

Motivated by previous work [*Jiang and Liu* 2011] and the increasingly available volunteered geographic information (VGI), such as OpenStreetMap data, we expand the study area from three to 29 European countries. We extract blocks as basic geographic representations and generate natural cities for each country. We test the scaling properties of blocks and natural cities and iteratively apply the head/tail division rule to the cities in each country. In this way, we can obtain a serial of such two-tier head/tail hierarchical structures at different levels, where the ascending tiers can be represented as the small, medium, large and mega cities. The cities in each level will follow a power law distribution if the number of cities is at a reasonably large scale. The percentages of the cities in the tail and head should also be near 80% and 20%, respectively (i.e., numerous small cities and fewer large cities). Therefore, whether cities in each level follow a power law distribution and how close the percentages

are to the 80/20 rule are set up as the standard for assessing and cross-comparing hierarchical structures among the countries. We also find that there is a strong correlation between city sizes/number of blocks and economic factors (i.e., gross domestic product (GDP) and population), which suggests that the hierarchies can serve as an indicator for assessing a country's economic system.

The major contributions of this paper are multi-fold. First, the city system is naturally categorized into small, medium, large and mega cities at the country level and a uniform standard is set up to assess and cross-compare the hierarchical structures (i.e., cities) at the country level from the bottom up. Second, the iterative application of the head/tail division rule proves to be successful. Third, the generated blocks and natural cities are as large as tens of gigabytes and will be of value to further urban studies. The data sets and the source code developed in this paper are released online at <http://fromto.hig.se/block>.

The remainder of the paper is organized as follows. Section 2 briefly introduces the study area and the data sets we use as well as the concepts of border number, block and natural city according to previous studies. Section 3 provides the identification of heavy-tailed distributions, with a particular focus on the scaling property. A specific example using real data is given to provide an intuitive explanation. In Section 4, the results from 29 selected European countries are visualized and analyzed. The model for comparing hierarchical structures across countries is provided and discussed in detail. Section 5 draws conclusions and points to future work.

2. Study Area and Data Processing

Two types of data sets are involved in this study: the main data set from the OpenStreetMap (OSM) database and the statistical data set from GDP and population reports by the *United Nations* [2010]. The OSM data are used to generate the blocks and natural cities, while the statistical data set is mainly used for comparisons. OSM is a type of VGI that refers to numerous people voluntarily creating, assembling and disseminating geographic information with the support of Web 2.0 technologies [*Goodchild* 2007]. VGI collects geographic data from the bottom up, which is a new method compared with the conventional top-down approach. OSM is perhaps the most successful

example of VGI. OSM inherits the features of VGI and aims to create a set of free and editable maps [Haklay and Weber 2008].

Founded in July 2004 by Steve Coast at University College, London, OSM has experienced rapid growth and now covers most of the world. However, spatial heterogeneity exists in the quality of the OSM data in terms of positional and attribute accuracies, completeness and consistency. As stated above, OSM originated in Europe and most European countries are open and easy for implementation of the OSM project. Thus, the OSM data in Europe possess higher data quality than in other areas, and the data quality among the different European countries is similar. Moreover, the social, economic and cultural backgrounds of these countries are similar, making the countries suitable for a cross comparison. Therefore, the study area (Figure 1) covers the European Zone countries as well as Ireland, Norway and additional countries due to their geographic proximity. Countries such as Andorra, Gibraltar, Monaco, San Marino, Holy See (Vatican City) and Liechtenstein are neglected because their geographic area is too small to be treated at the country level.

As a VGI project, the OSM data set possesses many types of spatial and non-spatial data. OSM data are a type of extensible markup language (XML), which means the database is self-explanatory and easily extended. The road network is characterized by key “highway” and can be easily extracted by traversing all of the items of OSM XML data in a country. Although there are many ways to download and process OSM data, most of these ways are time-consuming and may require large hard disks and large amounts of memory. Fortunately, websites, such as *Cloudmade.com*, divide OSM data by country boundaries and regions for end-users. We download the OSM data country by country directly and generate blocks and natural cities for each country. This process has been well documented in the literature [Jiang and Liu 2011].

Here, we briefly describe the process from Jiang and Liu [2011] (Figure 2). We begin by extracting the road network from OSM data and building the road network into a topological format using ArcGIS coverage (i.e., road segments). After the extraction, we track blocks from the outermost road segment until all road segments are traversed. Correspondingly, the outermost blocks possess a designation of border number one, and the neighbours inward by one step are assigned with border number two and so on, as necessary. Next, we divide these generated blocks into small and large groups using the head/tail division rule: the small blocks are

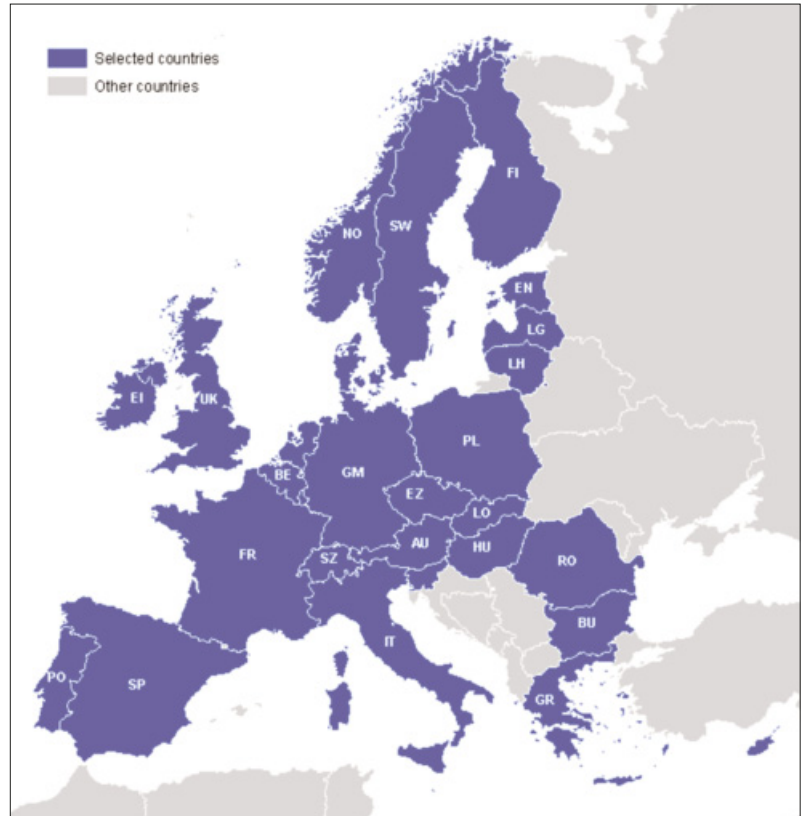


Figure 1: The 29 European countries selected for the study area.

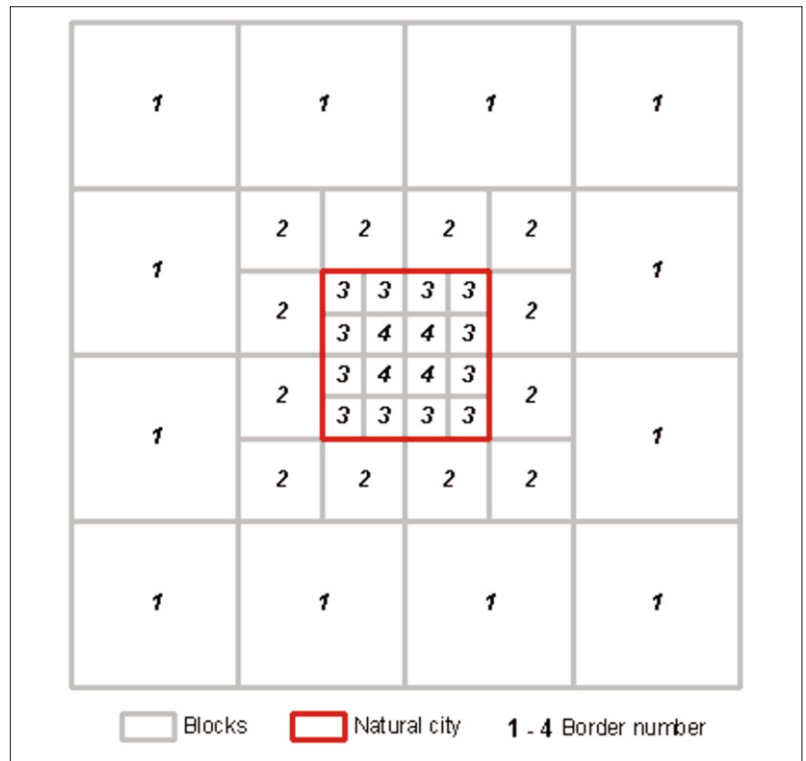


Figure 2: Blocks displayed with border numbers and a clustered natural city in an artificial grid road network.

those with areas less than the mean value of all of the block areas and vice versa for the large blocks. Next, for all of the small blocks, we start from any small block and search its neighbours and its neighbours' neighbours for small blocks until no such small neighbour blocks can be found. The traversed blocks are adjacent and form groups, which constitute what we call natural cities (red rectangle in Figure 2). We can see that the natural cities are generated in a natural way, which is the reason for the term "natural city."

The above process is applied to 29 European countries. The total data volume is up to 40 gigabytes, and the processing time is several days with a 2-terabyte disk and 42-gigabyte memory server computer. To implement the blocks with border numbers and generate natural cities, some isolated islands are removed, including the islands of Portugal and Greece. We directly cite the data from the three largest European countries (France, Germany and the UK) from *Jiang and Liu* [2011].

3. Scaling and Heavy-Tailed Distributions

In this section, we briefly introduce the concepts of the scaling law and heavy-tailed distributions, with particular emphasis on their relationships.

Generally, the term "scale" has different meanings in different fields. For instance, in cartography, scale refers to the ratio of distances between map and earth, whereas in the physical sciences, scale indicates the characteristic length (e.g., size or extent) of physical phenomena [Mark 2003]. The scale-free concept refers to the idea that regardless of the scale used, the power law distribution remains the same [Newman 2005]. This concept is similar to the concept of a fractal in geometry: the spatial form of an object is irregular, and the irregularity is maintained from scale to scale [Batty and Longley 1994]. Both the concepts of scale-free phenomena and fractals refer to the self-similarity of hierarchies in a man-made or natural process. The question encountered here is how to determine the scale in different contexts. In cartography, we can define the scale as 1:500, 1:1000 and so on, which appears a bit arbitrary; fractality in geometry is validated by strict formulas, which can be difficult to observe in natural phenomena. Therefore, finding a meaningful scale among a series of variables is crucial to understanding the hierarchical structures.

Newman [2005] noted that the 80/20 rule is best maintained if the median of the variables is used to divide the variables into a two-tier structure. Similarly, *Jiang and Liu* [2011] proposed that the head/tail division rule be used to obtain the two-tier hierarchical structure, which can also be found with

Table 1: The number of blocks in the head and tail at different levels of detail (LOD).

LOD	# of total	# (%) in Head (>Mean)	# (%) in Tail (<=Mean)	Mean
1	113 858	22 647 (20%)	91 211 (80%)	7
2	22 647	5 279 (23%)	17 368 (77%)	24
3	5 279	1 224 (23%)	4 055 (77%)	64

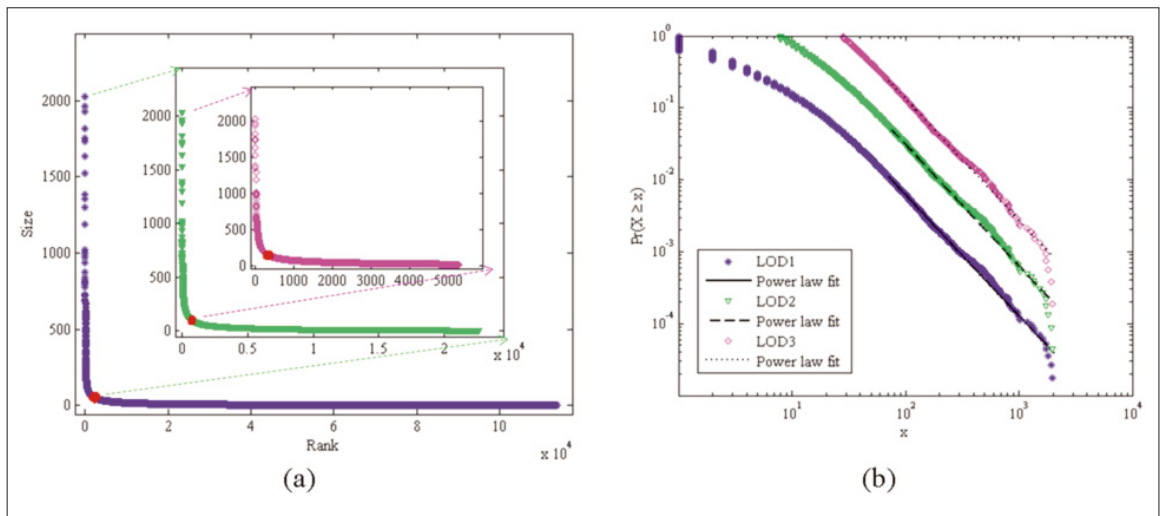


Figure 3: (a) Rank-size and (b) power law distributions at different levels of detail.

the 80/20 rule. To understand this process, we examine a series of data sets that use actual roads in Texas, USA (Table 1). The data are for the number of dangling road segments inside each block [Liu and Jiang 2012]. We repeatedly apply the head/tail division rule to the data set, and we obtain the two-tier hierarchical structures at different levels of detail (Table 1). We plot the head part (where the values are greater than the mean value) in a rank-size plot (Figure 3a) and a power law distribution (Figure 3b) at different levels of detail. We note that the scaling property and hierarchies are very well maintained. More importantly, there are always meaningful geographic implications behind the hierarchies when the information is viewed from the appropriate urban context. To this extent, we believe that the head/tail division rule provides deep insights into the hierarchies of scaling phenomena.

Now the question is how to mathematically detect the presence of a heavy-tailed distribution. As mentioned previously, a heavy-tailed distribution describes the special nonlinear relationships between a variable and its probability. There are many probability distribution functions to fit the data, such as the power law, power law with an exponential cutoff, exponential, stretched exponential distribution and lognormal [Clauset et al. 2009]. Clauset et al. [2009] also provide the models and source code in different versions to detect heavy-tailed distributions, which also have been extended by Jiang and Jia [2011]. Next, we briefly introduce the models adopted to detect heavy-tailed distributions (see Clauset et al. [2009] and Jiang and Jia [2011] for more information). To explain the parameters of these five distributions, we first provide their mathematical formulas below, where x denotes the given variable, and $f_{(x)}$ represents the probability function:

$$f(x) = C_1 x^{-a\sigma}, \quad C_1 = (a-1)x_{min}^{a-1} \quad (1)$$

$$f(x) = C_2 x^{-a\sigma} e^{-\lambda x}, \quad C_2 = \frac{\lambda^{1-\alpha}}{r(1-a, \lambda x_{min})} \quad (2)$$

$$f(x) = C_3 e^{-\lambda x\sigma}, \quad C_3 = \lambda e^{\lambda x_{min}} \quad (3)$$

$$f(x) = C_4 x^{\beta-1} e^{-\lambda x^\beta \sigma}, \quad C_4 = \beta \lambda e^{\lambda x_{min}^\beta} \quad (4)$$

$$f(x) = C_5 \frac{1}{x} e^{-\frac{[\ln x - \mu]^2}{2\sigma^2}}, \quad C_5 = \beta \lambda e^{\lambda x_{min}^\beta} \quad (5)$$

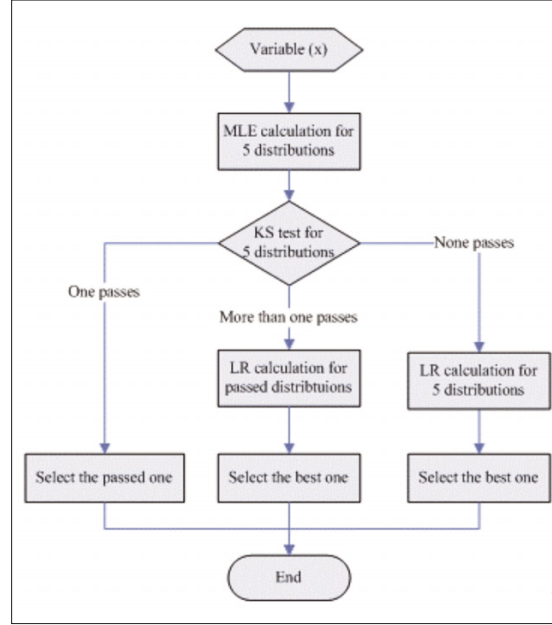


Figure 4: Flow chart for the detection of heavy-tailed distributions.

The above formulas denote (1) the power law distribution, (2) power law distribution with an exponential cutoff (i.e., the degenerated version of the power law distribution), (3) exponential distribution, (4) stretched exponential distribution (i.e., the degenerated version of the exponential distribution) and (5) lognormal distribution. The detection methods can be described as follows. First, maximum likelihood estimation (MLE) is adopted to calculate the parameters (x_{min} , α , β , λ , σ and μ) in the five mathematical functions above to describe or define the functions. Next, the Kolmogorov-Smirnov (KS) [Chakravarti et al. 1967] test is used to verify the validation of the distributions in the form of the p value. For example, if the p value is greater than a threshold (e.g., 0.05), we can assume that the variables follow the distribution [Clauset 2009]. The complete process is illustrated by the flow chart in Figure 4.

First of all, we calculate parameters for each distribution by MLE and then do a KS test for each type of distribution. If only one distribution passes the KS test, this indicates that the distribution fits the data best, and we will certainly select this distribution. Secondly, if more than one distribution passes the KS test, then we calculate LR for the passed distributions to decide the best one. If not any distributions pass the KS test, then LR calculation is the only way to decide the best-fitted distribution. Other methods could be used to decide the best-fit distribution when there are no passes for the KS test. The reason why we use LR is

because this method is popular and frequently used in literature with the support of open source codes.

In the process of identifying whether the sizes of blocks and natural cities follow basic heavy-tailed distributions, in most instances we depend on the calculated parameters of their functions. However, when a serial of blocks or natural cities follow more than one type of heavy-tailed distribution, we generally only compare the likelihood ratio (LR) rather than calculating the parameters of their degenerated versions such that the computing complexity is dramatically reduced. We use this method because the degenerated versions are in accordance with their basic distribution to a certain extent.

4. Results and Discussion

We first individually extracted the blocks from each of the 29 European countries. Next, we conducted the mathematical detection for each country and found that they all follow heavy-tailed distributions (Table 2). For Cyprus and Malta, both countries follow exponential and lognormal distributions. According to the method in Section 3, we calculated the LR for exponential versus lognormal distributions. The results show that the exponential distribution is a better fit for the block sizes in these two countries. For Germany and the UK, which do not follow any of the three basic heavy-tailed distributions according

Table 2: The number of blocks and size distributions for the selected 29 European countries.

	Country	# of blocks	Power law		Exponential		Lognormal			Distribution
			α	p	α	p	u	std	p	
1	Austria	189 295	2.0	0.0	4.7E-09	0.0	-8.3	4.9	1.0	Lognormal
2	Belgium	123 863	2.3	0.0	7.1E-08	1.0	11.8	1.7	0.0	Exponential
3	Bulgaria	21 639	1.4	0.0	2.5E-09	0.9	18.4	1.2	1.0	Lognormal
4	Cyprus	15 371	1.5	0.0	2.0E-08	1.0	-6.5	6.1	1.0	Exponential
5	Czech	142 779	3.8	0.0	7.1E-08	0.0	15.2	1.0	1.0	Lognormal
6	Denmark	75 659	2.4	0.0	3.3E-09	0.0	13.0	1.7	1.0	Lognormal
7	Estonia	15 323	2.4	0.0	7.0E-09	0.7	16.4	1.5	1.0	Lognormal
8	Finland	149 710	2.8	0.0	1.9E-09	0.0	16.3	1.7	0.4	Lognormal
9	France	569 739	1.5	0.0	5.5E-09	0.0	16.7	1.3	1.0	Lognormal
10	Germany	2 095 388	3.0	0.0	9.1E-08	0.0	10.7	2.0	0.0	Lognormal
11	Greece	88 780	1.8	0.0	2.8E-09	0.6	8.5	3.4	1.0	Lognormal
12	Hungary	42 734	3.8	0.0	1.4E-08	0.8	17.3	1.0	0.7	Exponential
13	Ireland	29 077	1.9	0.0	1.2E-08	0.4	5.7	3.5	1.0	Lognormal
14	Italy	356 210	2.2	0.0	4.7E-09	0.0	12.7	2.1	1.0	Lognormal
15	Latvia	14 935	1.4	0.0	5.9E-09	0.8	16.5	1.5	0.2	Exponential
16	Lithuania	12 040	1.4	0.0	5.8E-09	0.6	16.9	1.4	1.0	Lognormal
17	Luxembourg	10 382	2.9	0.0	2.4E-07	0.0	12.4	1.4	1.0	Lognormal
18	Malta	4 943	1.8	0.0	1.5E-06	0.8	11.9	1.1	0.8	Exponential
19	Netherlands	389 767	2.9	0.0	9.1E-07	0.5	12.9	1.2	0.0	Exponential
20	Norway	56 157	1.5	0.0	1.3E-09	0.7	5.8	4.5	1.0	Lognormal
21	Poland	201 756	2.1	0.0	4.7E-09	0.9	9.5	3.2	0.0	Exponential
22	Portugal	31 656	1.5	0.0	9.3E-10	0.3	16.0	2.0	0.0	Exponential
23	Romania	44 126	2.3	0.0	2.0E-09	0.7	16.7	2.0	0.0	Exponential
24	Slovakia	44 760	1.5	0.0	1.2E-08	0.5	14.6	1.7	0.9	Lognormal
25	Slovenia	13 465	1.9	0.0	1.6E-08	0.7	11.7	2.5	0.9	Lognormal
26	Spain	339 798	1.5	0.0	1.1E-09	0.0	16.1	1.8	1.0	Lognormal
27	Sweden	146 081	1.5	0.0	1.0E-09	0.0	14.2	2.4	1.0	Lognormal
28	Switzerland	157 772	1.8	0.0	1.1E-08	0.0	4.9	3.3	1.0	Lognormal
29	UK	586 809	2.4	0.0	2.4E-07	0.0	12.7	1.7	0.0	Lognormal

to obtained parameters, we calculate their LR to compare among the three distributions and found lognormal to be the best distribution.

According to the strategy in Section 2, we divide the blocks into small ones and large ones for each country using the head/tail division rule, because the blocks in each country exhibit heavy-tailed distributions (Table 2). We then generate natural cities based on the small blocks for all of the countries one by one. Figure 5 shows the blocks and cities for all of the selected 29 countries. All of the natural cities and border numbers for the blocks are generated at the country level. *Jiang and Liu* [2011] illustrated the spatial pattern in a single country as a mental image of the country for different people, which is quite different with geometric media axes inside the polygon boundary of a country. In this study, the patterns of visualized natural cities and blocks with border numbers in Figure 5 provide some insights into the spatial patterns from a special perspective at the continental level.

A closer look at the distributions of sizes of natural cities showed that the natural cities at the country level demonstrate a power law distribution (Table 3), which is in agreement with previous studies, but it is difficult to differentiate and assess these cities in terms of their statistical distributions. However, as stated above, the scale-free property allows us to use any scale for the power law distribution. Consequently, the scaling property will also be

well maintained in the two-tier hierarchical structures that are obtained via the head/tail division rule. If we apply the head/tail division rule to the head part of the natural cities in each country iteratively, we will obtain the two-tier hierarchical structures at different levels in a natural way (Table 3). This hierarchical structure is easy to understand; however, three questions arise: (1) why do we use cities instead of blocks, (2) why do we use three levels, and (3) how do we assess and make sense of the hierarchical structures?

To answer the first question, although blocks and cities indicate the geographic carrying capacity in a country (because human activities are mainly constrained within blocks and cities), cities are the real functional units for a country when compared with blocks. Moreover, on average, 85% of blocks in a country belong to cities. Therefore, the hierarchical structure of cities at the country level can better represent a country than blocks. In response to the second question, the scaling property (e.g., the power law distribution from the perspective of statistical physics) and 80/20 rule in cities cannot be well maintained after the third level in most of the countries. Moreover, three levels of division are in agreement with the common concept of small, medium, large and mega cities in a country. The cities in the tails of three hierarchies are called small, medium and large cities, and the cities in the head of the third level correspond to mega cities (Table 3). The physical meanings behind the

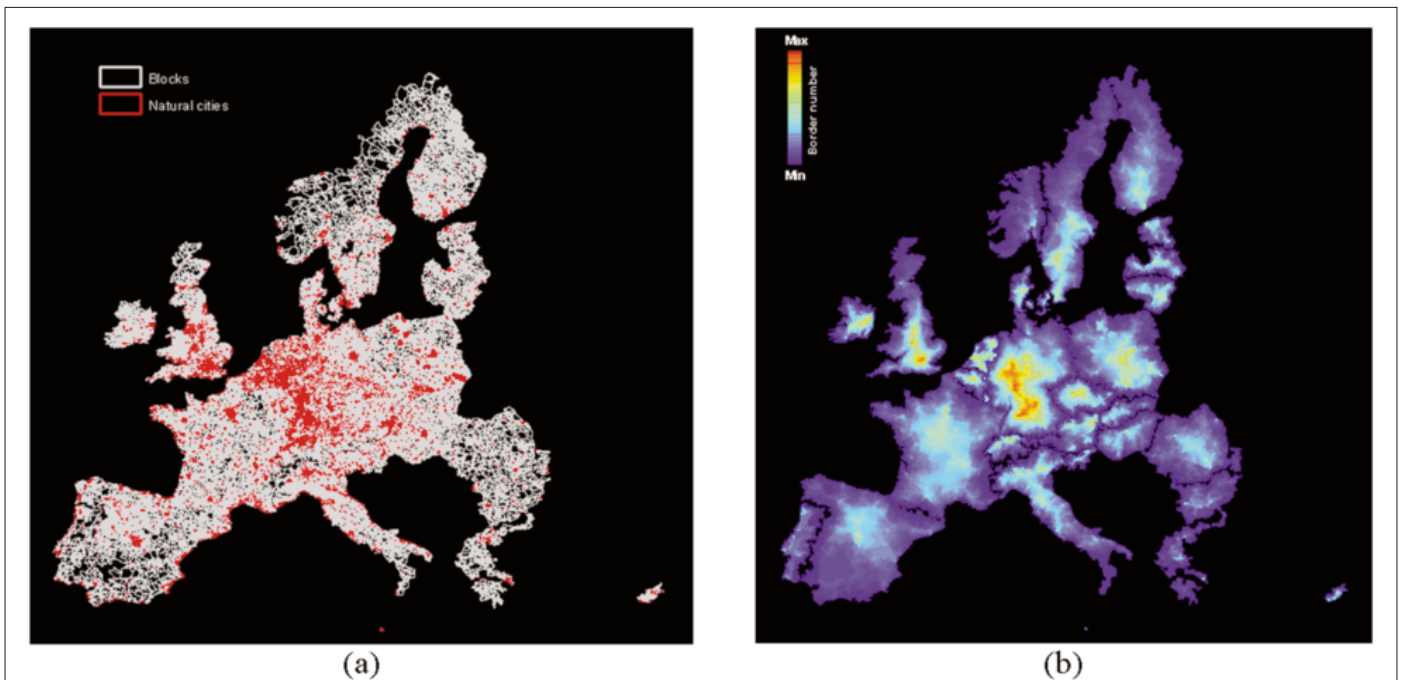


Figure 5: Visualization of (a) blocks and natural cities and (b) the border number of blocks at the country level for the selected 29 European countries.

obtained hierarchies of cities correspond to the levels of cities, and this finding is undoubtedly correct. For example, cities such as Paris, London and Berlin are all included in mega cities.

For the third question, previous studies show that urban systems or geographic environments evolve in a self-organized manner, which is recognized as the underlying reason for the equilibrium state of a complex system. Statistically, the self-organized behavior of cities demonstrates the scaling property. According to the mathematical definitions of scaling, as one looks deeper or “zooms in” on the complex system, a mini-structure similar to the large structure will emerge, which is termed self-similarity. Here, as we “zoom in” on the city

system using the head/tail division rule, a series of two-tier head/tail hierarchical structures can be obtained in the three levels (Table 3). Consequently, the self-similarity and stability of obtained hierarchies can be evaluated by whether they exhibit scaling property and how they are in agreement with the 80/20 rule. For example, if the percentages are far away from 80/20, such as 50/50 percent, we can judge that the hierarchical structure is not well held. Thus, the more often the percentages agree with the 80/20 rule, the more intact the hierarchical structure remains through the three levels.

Based on the above analysis, we can assess the degree of self-similarity and stability of each country by checking the scaling property and proximity

Table 3: The power law distributions and two-tier hierarchical structures of natural cities for the selected 29 European countries in three levels. (Note: the heads and tails are the percentages of whole levels, whereas the ratios are percentages of heads divided by tails).

	Country	Cities	Power law		Hierarchy1			Hierarchy2			Hierarchy3		
			α	p	Head1	Tail1	Ratio1	Head2	Tail2	Ratio2	Head3	Tail3	Ratio3
1	Austria	601	2.0	0.60	0.12	0.88	0.14	0.15	0.85	0.18	0.18	0.82	0.22
2	Belgium	403	2.0	0.50	0.10	0.90	0.11	0.14	0.86	0.16	0.33	0.67	0.49
3	Bulgaria	68	1.9	0.80	0.19	0.81	0.23	0.15	0.85	0.18	0.50	0.50	1.00
4	Cyprus	29	1.7	0.40	0.14	0.86	0.16	0.50	0.50	1.00	0.50	0.50	1.00
5	Czech	432	2.1	0.90	0.12	0.88	0.14	0.21	0.79	0.27	0.18	0.82	0.22
6	Denmark	143	1.8	0.50	0.13	0.87	0.15	0.16	0.84	0.19	0.33	0.67	0.49
7	Estonia	37	1.8	0.60	0.19	0.81	0.23	0.14	0.86	0.16	1.00	0.00	N/A
8	Finland	68	1.6	0.20	0.13	0.87	0.15	0.11	0.89	0.12	1.00	0.00	N/A
9	France	1 240	1.8	0.18	0.13	0.87	0.15	0.22	0.78	0.28	0.21	0.79	0.27
10	Germany	5 160	2.1	0.52	0.13	0.87	0.15	0.15	0.85	0.18	0.24	0.76	0.32
11	Greece	74	2.0	0.20	0.07	0.93	0.08	0.20	0.80	0.25	1.00	0.00	N/A
12	Hungary	74	2.0	0.50	0.11	0.89	0.12	0.12	0.88	0.14	1.00	0.00	N/A
13	Ireland	53	1.9	0.60	0.09	0.91	0.10	0.20	0.80	0.25	1.00	0.00	N/A
14	Italy	1 036	1.8	0.20	0.16	0.84	0.19	0.18	0.82	0.22	0.24	0.76	0.32
15	Latvia	32	1.7	0.00	0.22	0.78	0.28	0.14	0.86	0.16	1.00	0.00	N/A
16	Lithuania	24	1.7	0.80	0.12	0.88	0.14	0.67	0.33	2.03	0.50	0.50	1.00
17	Luxembourg	66	2.2	0.50	0.12	0.88	0.14	0.12	0.88	0.14	1.00	0.00	N/A
18	Malta	26	3.5	0.00	0.42	0.58	0.72	0.36	0.64	0.56	0.25	0.75	0.33
19	Netherlands	853	1.9	0.30	0.14	0.86	0.16	0.22	0.78	0.28	0.23	0.77	0.30
20	Norway	69	1.7	1.00	0.16	0.84	0.19	0.36	0.64	0.56	0.25	0.75	0.33
21	Poland	475	1.8	0.40	0.15	0.85	0.18	0.19	0.81	0.23	0.36	0.64	0.56
22	Portugal	86	2.0	1.00	0.12	0.88	0.14	0.20	0.80	0.25	0.50	0.50	1.00
23	Romania	70	1.7	0.30	0.20	0.80	0.25	0.29	0.71	0.41	0.25	0.75	0.33
24	Slovakia	106	1.9	0.90	0.10	0.90	0.11	0.18	0.82	0.22	0.50	0.50	1.00
25	Slovenia	48	2.1	0.03	0.10	0.90	0.11	0.40	0.60	0.67	0.50	0.50	1.00
26	Spain	605	1.8	0.10	0.13	0.87	0.15	0.22	0.78	0.28	0.24	0.76	0.32
27	Sweden	185	1.8	0.90	0.16	0.84	0.19	0.21	0.79	0.27	0.33	0.67	0.49
28	Switzerland	276	1.8	0.20	0.17	0.83	0.20	0.17	0.83	0.20	0.25	0.75	0.33
29	UK	1 251	1.90	0.00	0.15	0.85	0.18	0.18	0.82	0.22	0.26	0.74	0.35

to the 80/20 rule of their hierarchies. As mentioned in Section 3, each hierarchy in Table 3 demonstrates a power law distribution (the scaling property). Therefore, the assessment standard is how far each hierarchy deviates away from the 80/20 rule. The farther the hierarchy deviates, the worse the structure is. Some of the worst cases can easily read according to the normalized ratios in Table 3. Malta bears the worst hierarchical structures, because the ratio deviates too much in the first level. Similarly, the ratios of Cyprus, Lithuania, Norway and Slovenia deviate much in the second level; the ratios of Bulgaria, Estonia, Finland, Hungary, Ireland, Latvia, Poland, Portugal and Slovakia deviate much in the third level. The distributions of city sizes for

Malta, Latvia and the UK do not follow power law distributions. The cause for this finding is a matter of continuing research. For instance, we conjecture that one of the reasons may be because the UK and Malta are island countries.

As stated above, blocks and cities are generated from road networks and are the functional units of a country. We conjecture that there exist types of relationships between the spatial patterns or geographic configurations of blocks and the economic or social structures of the country. To investigate this idea, we conducted a correlation analysis of the total physical area, number of blocks and natural cities between GDP and population for the countries. We found that the correlation coefficients

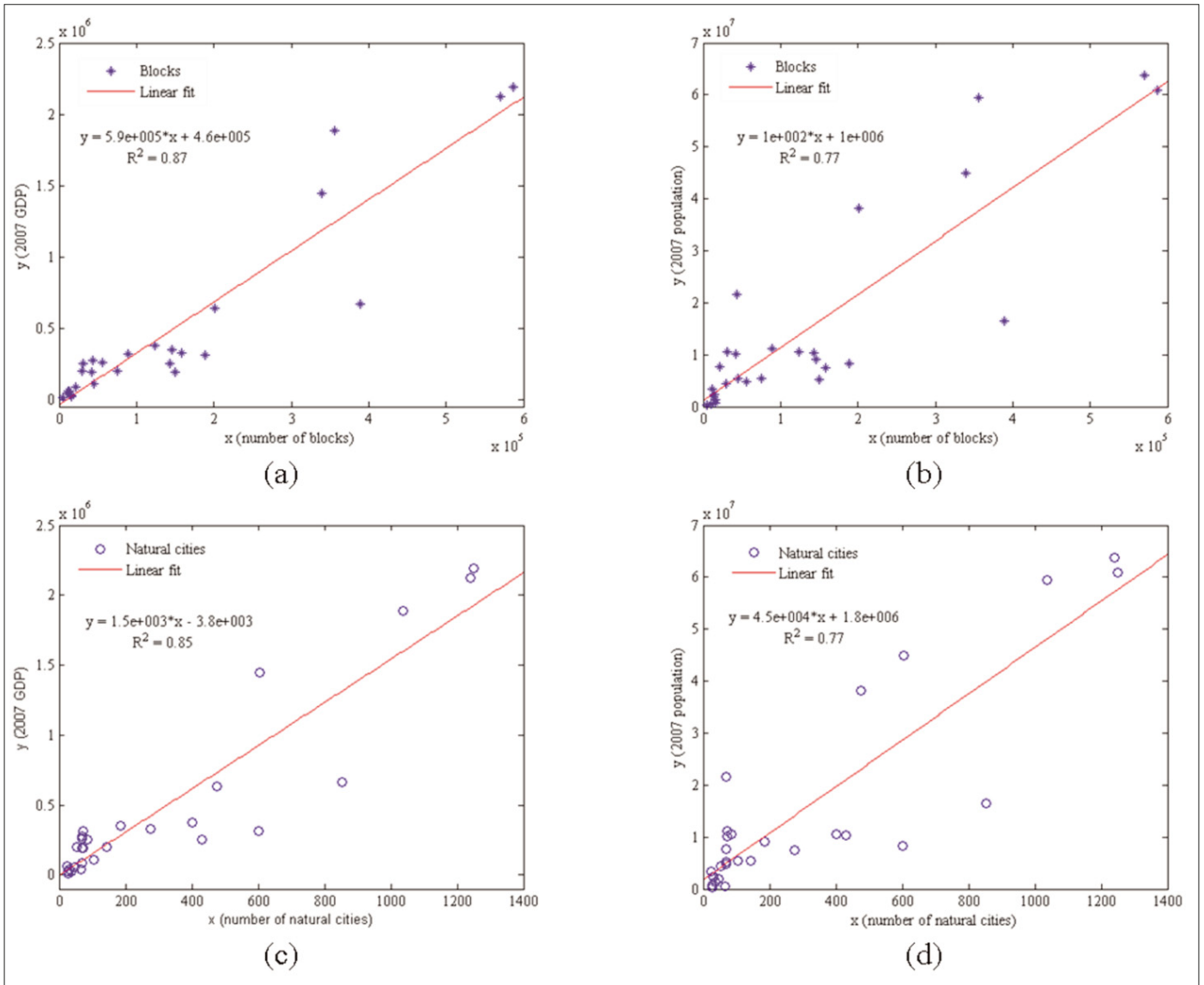


Figure 6: Correlation plots for the selected 29 European countries in 2007 between (a) the number of blocks and GDP, (b) the number of blocks and population, (c) the number of natural cities and GDP and (d) the number of natural cities and population.

between the number of blocks and cities and GDP and population are up to 0.87 (Figure 6). Germany has too many blocks compared with other countries, and thus Germany is eliminated when calculating the correlation coefficient. The statistical data for the year 2007 were chosen because this year contained the most up-to-date information for all of the countries.

Although the cities have strong relationships with economic factors (i.e., GDP and population) and play an important role in a country's economy and other aspects, the extension of hierarchical spatial structure to economic or social systems requires the verification of real economic indices. However, such economic data sets are very difficult to obtain. Nevertheless, the recent economic crisis in Greece could be a proof of its unstable structure in the third level as shown in Table 3. At the same time, Germany and France maintained a stable hierarchical structure. In addition, Switzerland has only 276 natural cities, but its hierarchical structure is very stable.

5. Conclusions

This study develops an across-country comparison of hierarchical spatial structures using blocks and cities as the basic geographic representations at the country level. Blocks and cities are extracted from volunteered geographic information data (OpenStreetMap) for 29 European countries via data intensive geo-computation. This study presents a complete model for the detection of heavy-tailed distributions. The sizes of blocks and cities of the 29 European countries are all tested to follow heavy-tailed distributions. Based on the power law distribution of natural cities at the country level and the head/tail division rule, we obtain the two-tier hierarchical structure of cities at three different levels. We contend that the heads and tails at three levels correspond to small, medium, large and mega cities. Furthermore, proximity to the 80/20 rule at each level of hierarchy is chosen as the standard to assess the degree of its self-similarity and stability in terms of a self-organized complex system. This method is a bottom-up approach to categorize and assess city systems. Moreover, cities correlate strongly with two economic factors: population and GDP at the country level. We conjecture that the hierarchical structure and spatial arrangement of cities can be indicators of the human and social systems and structures in a country.

To maintain the spirit of open source and volunteered geographic information, the main data sets (the road networks and blocks) and the core

algorithm source codes we develop in this paper are shared directly online via the Internet: <http://from-to.hig.se/block>. For future work, different types of geographic representations are to be extracted from volunteered geographic information to find more spatial patterns upon which additional social and economic patterns can be explored.

Acknowledgements

This study was financially supported by the Lars Erik Lundberg Scholarship.

References

- Adamic, L. 2011. Unzipping Zipf's law. *Nature*. 474:164-165.
- Alperovich, G. 1993. An explanatory model of city-size distribution: evidence from cross-country data. *Urban Studies*. 30(9): 1591-1601.
- Barthelemy, M., and A. Flammini. 2008. Modeling urban street patterns. *Physical Review Letters*. 100(13): 138702.
- Batty, M., and P. Longley. 1994. *Fractal Cities: A Geometry of Form and Function*. San Diego, CA and London: Academic Press.
- Blanter, E.M., M.G. Shnirman, J.L. Le Mouël, and C.J.Allègre. 1997. Scaling laws in blocks dynamics and dynamic self-organized criticality. *Physics of the Earth and Planetary Interiors*. 99: 295-307.
- Chakravarti, L., and L. Roy. 1967. *Handbook of Methods of Applied Statistics*. John Wiley and Sons. 1: 392-394.
- Chen, Y., and Y. Zhou. 2003. The rank-size rule and fractal hierarchies of cities: mathematical models and empirical analyses. *Environment and Planning B: Planning and Design*. 30(6): 799-818.
- Chen, Y., and Y. Zhou. 2008. Scaling laws and indications of self-organized criticality in urban systems. *Chaos, Solitons & Fractals*. 35(1): 85-98.
- Clauset, A., C.R. Shalizi, and M.E.J. Newman. 2009. Power-law distributions in empirical data. *SIAM Review*. 51: 661-703.
- Goodchild, M.F. 2007. Citizens as sensors: the world of volunteered geography. *GeoJournal* 69(4): 211-221.
- Haklay, M., and P. Weber. 2008. OpenStreetMap: user-generated street maps. *IEEE Pervasive Computing*. October-December. 12-18.
- Isalgue, A., H. Coch, and R. Serra. 2007. Scaling laws and the modern city. *Physica A*. 382: 643-649.
- Jiang, B., and T. Jia. 2011. Exploring human mobility patterns based on location information of US flights. [Preprint]. doi: arxiv.org/abs/1104.4578
- Jiang, B., and X. Liu. 2011. Scaling of geographic space from the perspective of city and field blocks and using volunteered geographic information. *International Journal of Geographical Information Science*. 10: 22-22. doi: arxiv.org/abs/1009.3635
- Jiang, B., X. Liu, and T. Jia. 2011. Scaling of geographic

- space as a universal rule for mapping or cartographic generalization. [Preprint]. doi: arxiv.org/abs/1102.156.1
- Kalapala, V., V. Sanwalani, A. Clauset, and C. Moore. 2006. Scale invariance in road networks. *Physical Review E*. 73(2).
- Lämmer, S., B. Gehlsen, and D. Helbing. 2006. Scaling laws in the spatial structure of urban road networks. *Physica A*. 363(1): 89-95.
- Liu, X., and B. Jiang. 2012. A novel approach to the identification of urban sprawl patches based on the scaling of geographic space. *International Journal of Geomatics and Geosciences*. 2(2): 415-429.
- Mark, D.M. 2003. Geographic information science: defining the field. In: *Foundations of Geographic Information Science*. Ed. M. Duckham, M.F. Goodchild, and M.F. Worboys. Taylor and Francis: New York, 2003, pp. 3-18.
- Newman, M.E.J. 2005. Power laws, Pareto distributions and Zipf's law. *Contemporary Physics*. 46: 323.
- Pumain, D. 2004. Scaling laws and urban systems. Santa Fe Institute (SFI) working paper. 04-02-002.
- Pumain, D., F. Paulus, C. Vacchiani, and J. Lobo. 2006. An evolutionary theory for interpreting urban scaling laws. *Cybergeo*. 343.
- Shalizi, C.R. 2011. Scaling and hierarchy in urban economies. Proceedings of the National Academy of Sciences (USA). [Preprint]. doi: [arxiv:1102.4101](https://arxiv.org/abs/1102.4101)
- United Nations. 2010. Demographic Yearbook. [Viewed June 2, 2011]. Available from: <http://unstats.un.org/unsd/demographic/products/dyb/dyb2008.htm>

Authors

Dr. Xintao Liu completed his graduate studies with a PhD from the KTH Royal Institute of Technology in Stockholm, Sweden, in 2012. Now Dr. Liu works as a Postdoctoral Fellow at Ryerson University in Toronto, Canada. Dr. Liu's current research interests are in the areas of space syntax, human mobility, Web 3D, big data, and transportation. He is also a member of the Canadian Institute of Geomatics. He has published publications in top international peer-refereed journals and books.

Dr. Yifang Ban is the Chair Professor of Geoinformatics at KTH Royal Institute of Technology in Stockholm, Sweden. Before joining KTH in 2004, Dr. Ban was a tenured Associate Professor at York University in Toronto, Canada. She received a BSc and MSc from Nanjing University, China, in 1984 and 1987 respectively, and a PhD from University of Waterloo, Canada, in 1996.

Prof. Ban's research interests include multi-temporal remote sensing, SAR image analysis, multisensor data fusion, image segmentation and classification, change detection, land cover mapping, urban analysis, 3D visualization, spatial analysis and modelling, and environmental impact assessment. She is a Council Member of the

European Association of Remote Sensing Laboratories (EARSeL) and co-chairs the EARSeL Special Interest Group "Temporal Analysis of Satellite Images." She also chairs the ISPRS Inter-Commission II/IV/VIII Working Group on Global Land Cover Mapping and Services, and is a co-lead of GEO SB-02 Global Land Cover. Prof. Ban is the European Lead PI for the Urbanization and Climate Project within ESA and Chinese Ministry of Science and Technology's Dragon III Program (2012-2016). She has been a scientific committee member for a number of major international remote sensing conferences and serves as reviewer for key remote sensing and GIS journals. She is an EU Expert for evaluation of Horizon 2020 ERC grant and FP7 Marie Curie Proposals. She serves as an Associate Editor for IEEE Journal on Selected Topics in Applied Earth Observations and Remote Sensing (JSTARS). She is also a Guest Editor for the JSTARS Special Issue on the Analysis of Multitemporal Remote Sensing Data and ISPRS Journal of Photogrammetry and Remote Sensing Special Issue on Global Land Cover Mapping. Prof. Ban has published over 100 publications with approximately 50 papers in international peer-refereed journals and books.

Songnian Li is a full professor in geomatics engineering in the Department of Civil Engineering at Ryerson University, Toronto. His research interests lie in the areas of geographical information sciences with focus on geospatial collaboration, dynamic geospatial web services, spatial decision support, moving object data, and web/real-time applications. Some of his recent projects include urban solar modeling and mapping, event-driven sense/response GIS, web-based sea-ice information portal, and road database updating and traffic modeling using GPS trace data collected from smart phones. Songnian has been involved with the International Society for Photogrammetry and Remote Sensing (ISPRS) since 2004. He served as Chair of ISPRS Working Group IV/5 on "Web-based Geoinformation Services and Applications" from 2004 to 2012, and is currently the President of ISPRS Technical Commission II. As a member of the Canadian Institute of Geomatics (CIG), Songnian has served the Institute in a number of capacities including associate editor of *Geomatica*, Treasurer and Interim Chair of the Toronto Branch, Chair of the Canadian National Committee for ISPRS, and council member. Songnian received his Bachelor of Engineering degree from Wuhan Technical University of Surveying and Mapping (now merged into Wuhan University), China, and his PhD degree from University of New Brunswick, Canada. □