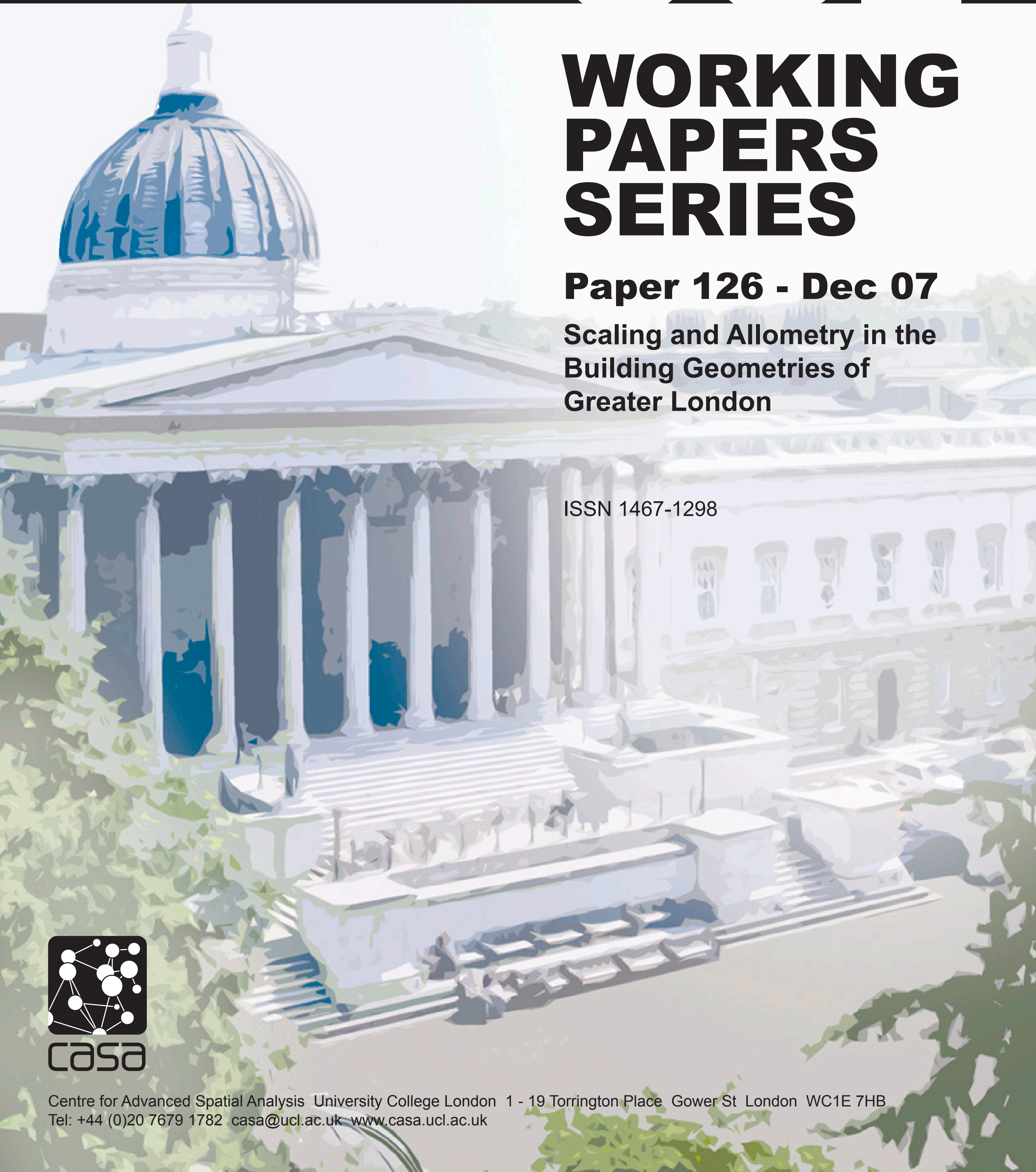
UCL CENTRE FOR ADVANCED SPATIAL ANALYSIS

UCL



WORKING PAPERS SERIES

Paper 126 - Dec 07

Scaling and Allometry in the Building Geometries of Greater London

ISSN 1467-1298

casa

Centre for Advanced Spatial Analysis University College London 1 - 19 Torrington Place Gower St London WC1E 7HB
Tel: +44 (0)20 7679 1782 casa@ucl.ac.uk www.casa.ucl.ac.uk

Scaling and Allometry in the Building Geometries of Greater London

Michael Batty, Rui Carvalho, Andy Hudson-Smith,
Richard Milton, Duncan Smith and Philip Steadman

Centre for Advanced Spatial Analysis, and Bartlett Graduate School, University College London, 1-19 Torrington Place, London WC1E 6BT

m.batty@ucl.ac.uk, rui.carvalho@ucl.ac.uk, asmith@geog.ucl.ac.uk,
richard.milton@ucl.ac.uk, duncan.a.smith@ucl.ac.uk, j.p.steadman@ucl.ac.uk

December 7, 2007

Abstract

Many aggregate distributions of urban activities such as city sizes reveal scaling but hardly any work exists on the properties of spatial distributions within individual cities, notwithstanding considerable progress on their fractal structure. We redress this here by examining scaling relationships in a world city using data on the geometric properties of individual buildings. We first outline Gibrat's (1931) model which gives rise to lognormal distributions and show how we use power laws to approximate their form. We illustrate this for population densities in Greater London and we then extend this analysis to allometric relationships between buildings in terms of their different geometric size properties. We present some preliminary analysis of building heights from the **Emporis** database which suggests very strong scaling in world cities. The data base for Greater London is then introduced and we extract 3.6 million buildings whose scaling properties we explore. We then examine key allometric relationships between these different properties illustrating how building shape changes according to size, and we then extend this analysis to the classification of buildings according to land use types. We conclude with an analysis of two-point correlation functions of building geometries which supports our non-spatial analysis of scaling.

This research was supported by the UK Engineering and Physical Sciences Research Council (EPSRC under grant EP/C513703/1 and by the UK National Centre for e-Social Science under the GeoVUE Project (ESRC under grant RES-149-25-1023).

Introduction

Cities are structured according to the rules of spatial competition which manifest themselves in self-similar patterns which are fractal. Populations tend to cluster around market locations which reflect a hierarchy of needs from the essential to the specialist, ordered spatially according to the strength of demand while densities tend to reflect economies of agglomeration which generate a small number of very high density locations and a large number of lower ones. The patterns that emerge are sustained by transportation routes that tend to fill space in the most economical way, minimising length and maximizing capacity, whose spatial organization is usually hierarchical and tree-like. Cities are thus composed of fractal-like clusters on many spatial levels whose order appears to follow well-defined numerical rules of scaling.

Most demonstrations of such order in fact pertain to systems of cities rather the spatial organization of the city itself, focusing on size distributions in which spatial order is implicit (Berry, 1964). The size distribution of cities in fact is scaling with Zipf's (1949) rank-size rule acting as the bench-mark against which many other spatial distributions are compared and contrasted. Most of the work to date on city size distributions throws away any spatial structure that exists. Cities measured by their populations, incomes or employment, are considered as dimensionless points with the argument always turning on the competition between cities rather than the competition between their component parts. In essence, the fact that there are a small number of large cities and a large number of small and that this distribution manifests a regularity which appears persistent through time, reflects the consequence of competitive processes under resource limits: there is simply never the supply of resources or demand to sustain large numbers of large cities, and thus most cities remain small. The same mechanisms clearly exist at the more local scale, within cities with the competition perhaps being less fierce but regular ordering of populations and other activities by size being the norm rather than the exception.

Inside cities, the predominant theory of ordering is based on microeconomics that suggests that densities of population, rent, and employment decline with increasing transport costs from intensive hubs or clusters of economic activity (Anas, Arnott, and

Small, 1998). As we will show later, it is easy to speculate that such order is consistent with a regular size distribution of population densities for hypothetical models where transport cost or distance from any point is equated with the rank order. But such research has never been attempted with work on scaling distributions barely touching the spatial structure of the city where the focus has been much more on fractal patterns rather than their scaling structure (Batty and Longley 1994). In this paper, we will extend the study of size distributions to the internal structure of cities first throwing away spatial structure, demonstrating that scaling orders are as strong within cities as between, and then reintroducing space to show its relative importance.

There is an additional twist here to our analysis of intra-city size distributions. Although we will begin by examining the distribution of population densities, our focus here is on geometric rather than economic or demographic attributes of the city. We consider that scaling in cities is strongly related to the constraints that geometry imposes on density and nearness and thus we will examine the size distributions of buildings in terms of their Euclidean footprint – area, perimeter, height and so on – making the rather loose argument that these sizes reflect demand for population and employment. Moreover, as buildings grow in size, their shape must change to enable them to function and thus their scaling can be linked to their allometry. In fact a sound theory of urban allometry should relate social and economic activity to building geometry and in this paper, we hope to set the agenda for further work in this area.

To date, work on the scaling of activities in cities has been sparse. As remarked, the study of rank-size distributions across cities has been extensive and work on urban density profiles has been significant. But there has barely been any work on building geometries with the exception of Bon (1973) and Steadman (2006). There has been some on transport and infrastructure supply networks (Carvalho and Penn, 2004; Kuhnert, Helbing and West, 2006; Lammer, Gehlsen and Helbing, 2006) and some on the allometry of transport networks (Cowan and Fine, 1969; Bon, 1979; Samaniego and Moses, 2007). However West, Brown, and Enquist (1999) are beginning to apply their theory of metabolic scaling to cities and social systems (Bettencourt, Lobo**,** Helbing**,** Kühnert, and West, 2007; Isalgue, Coch, and Serra, 2007), thus providing a marker for a better understanding of the way cities scale as they grow.

In the next section, we will introduce the idea of scaling as an approximation to some underling order in the size of things, relating this to ways of representing this order as densities and distributions, and illustrating this with the distribution of population densities in Greater London. We follow this with a statement of the key allometric relationships that characterize building geometry in terms of their volume, the area of their footprint, their height and their perimeter. Our first foray into analysis looks at scaling in the height of buildings world-wide and in three cities – London, Tokyo and New York from the **Emporis** data base (http://www.skyscraper.com/). We show quite conclusively that these distributions can be well approximated by rank-size distributions that imply power laws. We then outline the main database that we are working with for Greater London which contains some 3.6 million building blocks. Analysis of this data then proceeds, first for rank-size scaling of building geometries, then for allometric relationships. We finally introduce two-point correlation measures of the spatial distribution of these building geometries demonstrating that the strong scaling relations already detected, are not completely destroyed in terms of their spatial extent.

Approximating Urban Order Through Rank-Size Scaling

One of the simplest models generating a size distribution that accords to the way agglomeration economies and competitive effects play themselves out is a process whose the elements of growth are proportionate but random. This is Gibrat's (1931) law which generates a lognormal distribution which is skewed to the left and whose mean is greater than its mode. It is worth repeating a simple demonstration of how this arises as there is substantial evidence that many such size distributions are lognormal (at least when not in their steady state) and therefore any scaling that is present in such distributions is an approximation. We first define population density in location i at time t as $p_i(t)$ and the change in density from time $t-1$ to t as $\Delta p_i(t-1)$. The growth dynamic is thus

$$p_i(t) = p_i(t-1) + \Delta\, p_i(t-1) \qquad (1)$$

with the succession of changes in density from $t=1$ to $t=T$, that is from 0 to T, and $p_i(t-1)>0$. Change is hypothesized as being proportionate. Thus from (1)

$$\Delta\, p_i(t-1) = p_i(t) - p_i(t-1) = \varepsilon_i(t-1)\, p_i(t-1) \qquad (2)$$

where $\varepsilon_i(t-1)$ is the variation around the unit rate that simply reproduces the population. Combining (1) and (2), the growth rate is defined from

$$p_i(t) = [1+\varepsilon_i(t-1)]\, p_i(t-1) \qquad (3)$$

where $-1<\varepsilon_i(t-1)<1$ to ensure that the total population is never negative. The rate $\varepsilon_i(t-1)$ is random but clearly proportionate as equations (1) to (3) imply, that is

$$\varepsilon_i(t-1) = \frac{\Delta\, p_i(t-1)}{p_i(t)} = \frac{p_i(t) - p_i(t-1)}{p_i(t)} \quad . \qquad (4)$$

If we sum the increments or decrements of growth, and if we assume that each is small which is likely to be the case in a well-specified growth process, we can approximate their sum as follows (Aitchison and Brown, 1957)

$$\sum_{t-1}^{T} \varepsilon_i(t-1) = \sum_{t-1}^{T} \frac{p_i(t) - p_i(t-1)}{p_i(t)} \sim \sum_{t-1}^{T} \frac{d\ p_i(t)}{p_i(t)} \sim \int_{p_i(0)}^{p_i T)} \frac{d\ p_i(t)}{p_i(t)} \ . \qquad (5)$$

The integral in (5) simplifies to

$$\log p_i(T) - \log p_i(0) = \sum_{t-1}^{T} \varepsilon_i(t-1) \qquad (6)$$

from which the growth equation is

$$\log p_i(T) = \log p_i(0) + \sum_{t-1}^{T} \varepsilon_i(t-1) \quad . \tag{7}$$

$p_i(T)$ is clearly a variate whose logarithm is normally distributed by the central limit theorem. We can easily show that there is a class of models whose growth rates are random but proportionate to the size of the population generated so far, that is $\varepsilon_i(t-1) \propto p_i(t-1)$ and these lead to much more skewed distributions than the lognormal. Although there is no explicit competition between locations in this model, the model is still valid if each population is placed on a lattice composed of points around which there is local diffusion mirroring a degree of spatial competition or trade (Batty, 2006).

To illustrate the lognormality of urban distributions, we use population density defined as $p_i(2005) = P_i(2005)/A_i(2005)$ where P_i is the population at time $t = 2005$ in each ward i in Greater London and A_i is the land area. We have scaled the densities so that they sum to 1000 over $N = 633$ wards, and we order the locations from the smallest to the largest densities and thus change the index from i to k. The density and distributions that we plot are thus based on p_k where $\sum_{k=1}^{N} p_k = 1000$. To plot the probability density function (PDF), we need to bin the data, choosing 50 bins of width 0.1 units where the first bin is from $0 \le p_k \le 0.1$ and the last from $4.9 \le p_k \le 5.0$. In Figure 1(a), we plot these densities as a histogram where the frequencies on the vertical axis are approximated by $f(p_k)$. It is clear that the density is skewed to the left; its mean is 1.579, its median 1.365, its mode by inspection is about 1, its minimum 0.037 and its maximum 4.793. We have not tested this data to derive the closeness of its fit to the lognormal or any related distribution for our purpose here is only to show skewness and the fact that such distributions have asymmetric tails.

The cumulative distribution function (CDF) defined as $F(p_K \le p_k)$ can be computed from the raw data as $\sum_{k=1}^{K} p_k$ without binning and it is thus preferable to work with the data in this form. This is equivalent to the integral of the continuous density and we show the plot in Figure 1(b) where compared to the classic ‘S’ shaped distribution function of the normal variate, this shows a high degree of lognormality. In fact, the normal practice in examining such size distributions is to use the counter or complementary-cumulative distribution function (CCDF) which we define here as $r = F(p_K \geq p_k) = N - F(p_K \leq p_k)$ shown in Figure 1(c). The CCDF is none other than the rank-size distribution defined by Zipf (1949) and used extensively in approximating the fat tail as a power law. Note that henceforth r will be used to define rank in terms of the ordered sizes defined by k.

To illustrate how we assume scaling in such distributions, we plot the CCDF on a logarithmic scale as in Figure 1(d). This form of plot gives greater visual weight to the larger values of density and it is intuitively clear that the relationship can be approximate rather well by a straight line which is the signature of a power law. In fact examining Figure 1(a), the right tail which is often called the fat or heavy tail, so called because it defines densities that are not exponentially bounded, can clearly be approximated by a power law. For the CDF and the CCDF in Figures 1(b) to 1(d), we show the area of the curve which we ‘intuitively’ judge to be the best approximation to such a power law in black with the excluded set of observations in grey. We now approximate the power law in Figure 1(a) as a continuous density suppressing the index k as

$$f(p) \sim p^{-\alpha} \qquad (8)$$

where α is the power of the density. The cumulative (and counter-cumulative which has the same functional form) is the integral of (8) and is

$$F(p) \sim p^{-\alpha+1} \qquad (9)$$

which we can also write explicitly in rank-size terms as

$$r \sim p(r)^{-(\alpha-1)} \quad . \qquad (10)$$

The usual form however is as in Figure 1(c) where population density is written as a function of rank. Then from (10)

$$p(r) \sim r^{-\frac{1}{\alpha-1}} = r^{-\beta} \qquad , \qquad (11)$$

where we now define β as the (inverse) power.

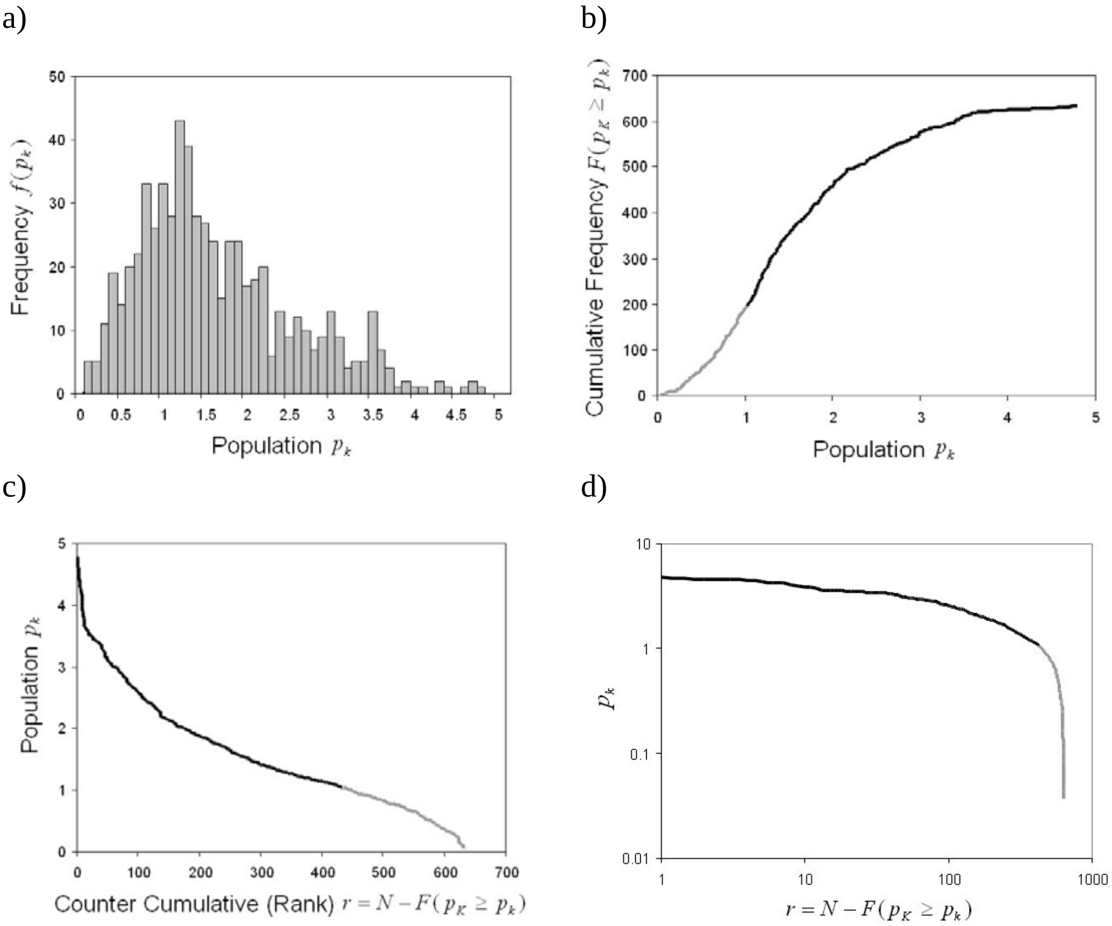


Figure 1: Approximating Lognormal Distributions as Power Laws: The Distribution of Population Densities in Greater London

a) The probability density function as a binned histogram; b) the cumulative distribution function from the data in a); c) the counter-cumulative (Pareto) or rank-size distribution formed by twisting b) through 90 degrees; d) the rank-size distribution in c) plotted on logarithmic scales. Note that range of these distributions that can be approximated by a power law are drawn in black with the excluded portions of the curves in grey

From equations (8) to (11), it is clear that if such an approximation is warranted, then the parameter of the density function α must be greater than 1 for the cumulative distribution function to be defined. If we logarithmically transform (11), we produce the linear equation

$$\log p(r) = \log G - \beta \log r \qquad (12)$$

which can be estimated in a straightforward manner using regression. The linear scaling implied in those observations in Figure 1(d) shown in black yields a correlation of 0.869 and a parameter $\beta = 0.345$. The parameter $\alpha = 1 + 1/\beta = 3.898$ which implies that the slope of the heavy tail is rather flat with relatively modest competition between the locations relative to what we see in city size distributions where β is often greater than unity. We will say more about this later but a note is warranted on approximating heavy tails with power laws and methods of estimation. In various applications, we have used the Hill maximum-likelihood estimator favored by Newman (2005) although here we have kept to the traditional method of regression because as we will see the scaling for the Greater London building geometries is so clear than we consider regression quite robust. We have not tested either the degree to which these distributions are lognormal or scaling but Clauset, Shalizi, and Newman (2007) have introduced a series of tests to enable this. In future work, we will follow this best practice but as this paper is simply to explore the extent to which scaling might be present in building geometries and allometry, we stick with current practice.

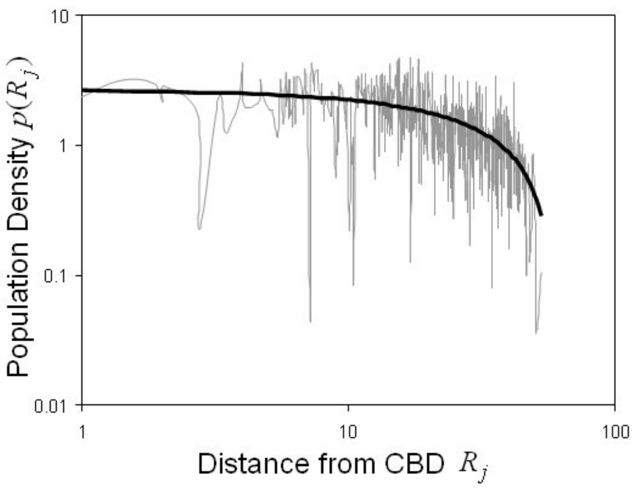


Figure 2: Population Densities in Greater London with respect to Distance from the Central Business District (CBD)

The grey curve shows the distribution and the black line shows an approximate non-linear fit of density to distance which ‘resembles’ a lognormal signature

The way we have ordered population densities as $\{p_k\}$ destroys the spatial-locational organization which is contained in the initial distribution of densities $\{p_i\}$. It is in fact usual to examine the density with respect to the distance or travel cost order from the CBD (central business district) and in urban economic models, densities decline monotonically with distance from the CBD (Clark, 1951). These models imply size distributions that at least would co-vary with their equivalent rank-size distribution although the usual specification is to assume densities decline as a negative exponential with distance rather than a power function. If we assume that distance (radius) from the CBD is defined as R, then we hypothesize that $p(R) \sim R^{-\gamma}$ where γ is the parameter on distance. We have plotted the same data set with respect to distance in Figure 2 where it is immediately clear that there are considerable departures from this scaling. The value of γ is 0.531 which is the same order of magnitude as β, the correlation-squared is 0.228 and the trend in the data in Figure 2 appears lognormal. There are many important links between density scaling by size and density by location contained in these data but these remain unexplored and thus represent an important area for future research.

Allometry in Urban Size Distributions

The focus of this analysis however is not on population density but on the geometric properties of buildings. One reason why the analysis of size distributions within cities is much more complicated than between cities is that the unit of analysis for population is not stable. Populations are composed of individuals and when they are spread out over an area, a critical issue is the spatial unit in which they need to be composed for appropriate analysis. As populations are aggregated, then their relative distribution changes. In the last section, we used 633 wards but had we used a finer scale of unit, the density profile and size distribution would certainly have changed, perhaps becoming more peaked but eventually being destroyed entirely once the data is represented at individual level. When we turn to buildings, this problem changes because the objects in question are distinct. There are still definitional problems but buildings are individual objects that vary in their size whereas individuals in a population do not do so in the same way.

It is hard to argue from the population density profile in Figure 2 that the profile of building sizes containing the same population would follow the same distance or rank order. Population is grouped into buildings and it is likely that the more intense the demand for space, the higher the population density of an individual building with a possible consequence that the higher the density, the greater the volume of space in any building. Any speculation of this kind however is obscured by the fact that overall, we do not make any distinction between type of use of a building. In the sequel, we will deal with different land uses but our first analysis will be for all buildings and thus it is not possible to speculate in advance what the overall scaling might be. In fact, before we began this analysis, we differed between ourselves as to whether or not we considered scaling to be likely in the buildings data with some of us arguing that this would be unlikely. In the event, the scaling is strong and regular as we will see.

We now need to define the geometric properties of buildings that we will use to measure their size. Each building is a rectangular block defined in terms of the lengths of its three dimensions. For each building, height H_j, the area of its footprint A_j, the perimeter of this area L_j, and the building volume $V_j = A_j H_j$ can be calculated directly although volume which is probably the best measure of size, is a product of all three dimensions, in turn a function of the area and height measures. Each of these has a rank order r which we will test for scaling using power law approximations $H(r) \sim r^{-\beta_H}$, $A(r) \sim r^{-\beta_A}$ $L(r) \sim r^{-\beta_L}$ and $V(r) \sim r^{-\beta_V}$. However what is of particular interest is the way these geometric measures relate to one another as their overall sizes change. This is allometry. The critical hypothesis is that as the size of the typical elements change, these relations may well depart from the standard geometric relations that characterize length, area, and volume. The allometric hypothesis suggests that there are critical ratios between geometric attributes that are fixed by the functioning of the element in question and if the element changes in size, these ratios need to remain fixed for the element to still function. Often the geometry has to change if these ratios are fixed (Thompson, 1917). A good example relates to natural light penetrating buildings. As natural light depends on the surface area, then to preserve a given ratio of natural light for the volume of the building, the shape of the building has to change if the building is increased in size. In short, the surface area does not change at the same rate as volume and if the ratio has to be fixed to make the building function, then the volume has to change. This implies a change in shape as the building increases in size.

As yet there is no well worked out theory of urban allometry; indeed there is no complete theory of size in biological systems from whence these ideas arise (Bonner, 2006) although there are various theories in the making (West, Brown and Enquist, 1999). We will begin by stating basic geometric relations assuming the building to be a cube with its basic linear unit as L. L first determines the area A as L^2 and then volume V as L^3 from which it is clear that $V = A\,L$. Standard allometric relations first proposed by Huxley (1932) can be immediately derived which imply changes in the volume, area or length relative to each other of these measures. For our cube (which can be easily generalized to a less uniform geometry), $A = V^{2/3}$, $L = A^{1/2}$, and $L = V^{1/3}$. These imply that as the volume grows, the area grows at a rate 2/3rd's the rate of volume growth. This can easily be seen in the relative growth rate or ratio of dA/A to dV/V (assuming a unit of time) as follows

$$\frac{dA}{dV} = \frac{2}{3} V^{(2/3)-1} = \frac{2}{3}\frac{V^{2/3}}{V} = \frac{2}{3}\frac{A}{V} \quad . \qquad (13)$$

Rearranging terms in (13), we get the ratio, the relative growth of dA/A to dV/V, as

$$\frac{dA}{A} \Big/ \frac{dV}{V} = \frac{2}{3} \quad , \qquad (14)$$

which can be easily generalized for any scaling parameter λ. The general allometric relation relating some physical property y of an object to another x is thus

$$y = G\,x^{\lambda} \quad , \qquad (15)$$

where the scaling parameter is the relative growth rate of y to x

$$\lambda = \frac{dy}{y} \bigg/ \frac{dx}{x} \quad . \tag{16}$$

λ is also the elasticity as defined in economics. (15) and (16) can thus be applied to any relationship which might be scaling with respect to different measures of size where these sizes imply differential relative growth (Von Bertalanffy, 1973).

To simplify our treatment, we assume that the entire array of buildings can be represented as rectangular blocks. In fact this is the case as we will see in our buildings data base where buildings are constructed from plot area and mean height and where more complex buildings are glued together from simpler rectangular blocks. Then in terms of building blocks, linear dimension will involve heights H_j in the (**z**) dimension and vector lengths in the (**x, y**) plane from which the area of the plot A_j, its perimeter L_j, and its volume (or mass) V_j can be computed. We will not compute surface area of the building, or any internal measures of circulation or areas of interior space for these are not possible although currently the databases are being augmented to deal with such complexities. These four measures are defined for each building which is located at a point or centroid (toid) j. We are interested in their scaling with respect to rank-size which we have hypothesized above but we are also interested in how they scale with respect to each other, allometrically. The following scaling relations are stated:

$$\left.\begin{array}{r} H_j = Z_1 L_j^{\kappa}; \quad A_j = Z_2 L_j^{\eta}; \quad V_j = Z_3 L_j^{\mu} \\ A_j = Z_4 H_j^{\varphi}; \; V_j = Z_5 H_j^{\chi} \\ V_j = Z_6 A_j^{\vartheta} \end{array}\right\} \quad . \tag{17}$$

where the Z_* are the constants of proportionality and the power symbols are the appropriate allometric parameters – relative rates of change.

Our key interest in urban allometry is to find out whether the scaling between area and volume implies changes in the shape of buildings. In terms of the relations in (9), we would expect the volume to scale as the cube of height and perimeter, and as the square of the plot area. Plot area is likely to scale as the square of height and perimeter while perimeter and height scale with each other linearly and these are the baseline allometries that we might expect. However if there are changes of shape, then these will reflected in the parameter values that are estimated from the equations in (9). In fact, as it is likely that there will be considerable variation around these forms for all buildings, we will disaggregate the set of all buildings into different land use types which should reveal differences, particularly between buildings in commercial and residential use.

Currently we are not able to measure the surface area of building from the database and this is unfortunate as this may scale quite differently from the 2/3rd's ratio that pertains to the standard pure allometric equations. This is because the skin of the building is the conduit for light and energy and buildings cannot maintain their volume indefinitely through increasing their floor areas because such areas cannot be serviced through natural light and other forms of externally supplied energy. Thus there are limits on shape in this regard. This is why it is likely that as buildings increase in size, they expand vertically rather than horizontally which are the kind of deviations from standard allometry that we are seeking. Our ultimate concern in this work is to count the number of buildings types by land use and to link these counts and their shapes to energy emission in buildings as well as issues involving circulation both within and between buildings.

Building Data and the Preliminary Analysis of Heights

To show that scaling exists in the size of buildings, we begin by selecting height data for the top 200 buildings worldwide and compare these with the same number for London. These data are from the **Emporis** database (http://www.skyscraper.com/) which contains quite detailed information about the largest buildings in 50,000 cities worldwide with up to 3000 of the largest buildings from the largest cities. We will also examine three cities in more detail – London, New York and Tokyo – as a prelude to our work with the Greater London buildings database which as we will outline below, is taken from our Virtual London model which consists of building blocks constructed from digital data sources. In Figure 3(a), we show Zipf plots of the top 200 buildings by height worldwide, for London from the **Emporis** database and for London from our own database. We have also graphed the top 200 cities by population in the year 2006 taken from UN sources (http://unstats.un.org/unsd/) to show that scaling in population is a little more extreme than for high buildings. In all the Zipf plots that we introduce henceforth, we normalize the data in the following way. We normalize the rank r by dividing by its maximum $r_{\max}$ and for the size variable, height H_j say, we divide by its mean $< H_j >$. Our plots are then based on graphing $H_j / < H_j >$ against $r/r_{\max}$, thus enabling us to directly compare data by collapsing all the plots onto one another.

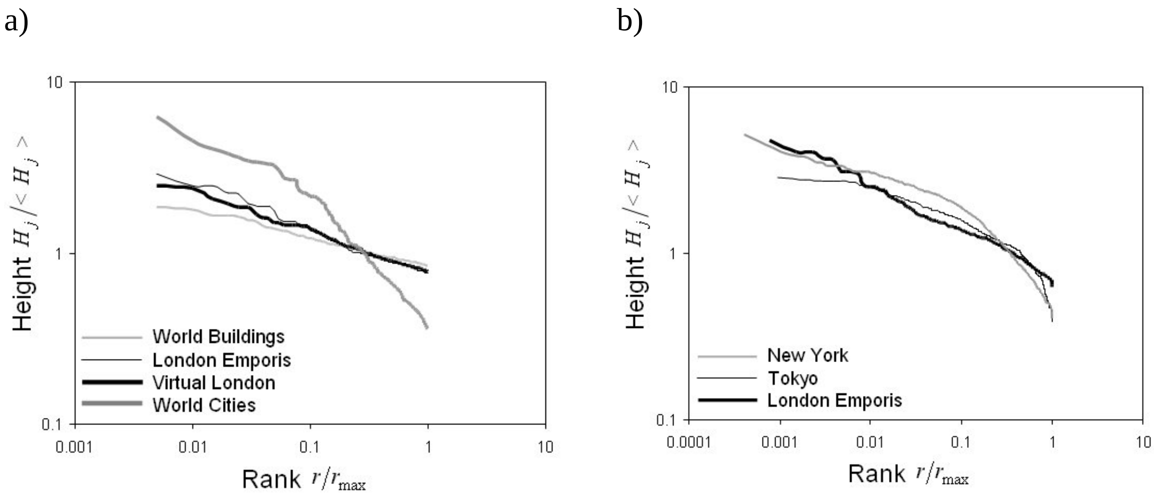


Figure 3: Initial Analysis of Building Heights

a) Top 200 Buildings by Height in the World and London, and Top 200 City Populations b) Top Building Heights in New York, Tokyo and London

There is very clear scaling in all four data sets and we present the parameters of these in Table 1. The slope of the world cities data is steeper than the buildings data which implies that there is less competition for activity inside a city than between them. We have also examined the same scaling in building heights for three world cities from the **Emporis** database and in Figure 3(b) we show building heights over a wider range of magnitudes for Tokyo, London and New York. The results which are also shown in Table 1 imply that New York has greater competition than Tokyo and that London has the flattest profile in terms of rank-size scaling. Although the fit of the power law to the London and World data sets is good, this is less so for Tokyo and New York where there is clear evidence of lognormality in the plots even at their upper end. This simply confirms the observations made above about needing to exercise care in approximating such urban distributions by power laws.

	World Cities	*World Buildings*	*London **Emporis***	*Virtual London*	*Tokyo*	*London*	*New York*
N	200	200	200	200	1036	1302	2424
Scaling Parameter β_H	0.652	0.162	0.262	0.234	0.377	0.288	0.478
Correlation Squared	0.970	0.995	0.983	0.992	0.827	0.979	0.919
Density Parameter α	2.534	7.159	4.823	5.269	3.650	4.477	3.094

Table 1: Scaling Parameters for the Preliminary Analysis of Building Heights

This preliminary analysis gives us some confidence that there is scaling in building geometries and led us to develop a large database for London based on our 3-D GIS/CAD model of London which we refer to as Virtual London (Batty and Hudson-Smith, 2005). This is a digital model of all building blocks within about 40 kilometers of the CBD – the City of London or 'square mile' – covering the 33 boroughs comprising the Greater London Authority (GLA) area which has an extent of 1579 square kilometers. The data set is unique in that it has been created automatically from two main sources of data: first vector parcel files part of Ordnance Survey's **MasterMap** (http://www.ordnancesurvey.co.uk/oswebsite/products/osmastermap/) which code all land parcels and streets to at least one meter accuracy; and second a data set of buildings heights constructed from InfoTerra's LIDAR data which produces a massive cloud of 3-D **x-y-z** data points which when used in association with the vector parcel data, can be used to extrude all buildings. In this data set, there are some 3,595,689 ($\cong$ 3.6m) distinct buildings centroids (or toids as they are called). We are currently dealing with all 3.6 million although we only use a subset of these in our scaling and allometric analysis. In future work, we will be aggregating toids to ensure that we are dealing with appropriate blocks. This becomes critical when land use is to be assigned to each building block because land use is tagged to street addresses which are a subset of all toids.

To give some idea of the range of this data set, the maximum height of any block is 204.06 meters, the Canary Wharf Tower in the London Docklands. The mean height is 5.76 meters and the standard deviation is 3.29 meters which shows that the frequency of building heights is very skewed to the left, reflecting the fact that this distribution is likely to follow a power law. For illustrative purposes only the top 10 blocks by height in London are 204, 197, 169, 160, 151, 150, 138, 130, 128, and 123 meters in comparison with the top 10 from the **Emporis** world database which are 509, 452, 452, 442, 421, 415, 391, 384, and 381. London's highest building is in fact not in the top 200 in the world and from the regression in Table 1 associated with the plot in Figure 3(a), we can estimate its rank as about 400. London is not a city of tall buildings.

From the data set, we are currently working with the perimeter of each plot which is computed directly from the **MasterMap** data, and the mean height of a plot which is important as there are many different heights from the LIDAR data reflecting complex roof shapes, masts, air conditioning units and so on. Other measures of height such as median and mode do not change the results below substantially. We compute volume by taking the area of the plot and multiplying it by its height. This does not take account of course of the fact that some buildings will taper but currently we are not able to do much about this as we do not have elaborate algorithms in place to construct intricate roofing shapes. We also are able to classify these buildings by land use from the **MasterMap** Layer 2 where we have land uses associated with each street address for which there is a toid. However there are many blocks that do not have street addresses and these tend to be part of other building complexes and/or are very small and somewhat idiosyncratic in their form, such as sheds, lean-to's and such-like bric-a-brac. We have various algorithms for joining unclassified polygons to those which are already classified and currently we consider the data set to be robust. There are over a thousand different land use types in the **MasterMap** data and we have classified these into nine major types which we list as the toids classified with at least one residential, office, retail, services, industrial, educational, hotel, transport, and general-commercial land use. We have not yet broached the difficult question of multiple uses for if we have a building with more than one land use classifier, we simply include it in the appropriate analysis. We have not yet tackled this double counting.

Rank-Size Distributions and Allometric Analysis of Building Geometries

We begin with the aggregate scaling relations which result from ranking the area $\{A_j\}$, perimeter $\{L_j\}$, height $\{H_j\}$ and volume $\{V_j\}$ data for a slightly reduced data set of about 3.59 million buildings. We show these in Figure 4 which also contains the same scaling for each of the land uses which we will describe below. What this figure reveals is remarkably strong linearity over many orders of magnitude with the plots collapsing dramatically for the million or so smallest buildings (which are less than about 25 square meters in volume) and quite definitely represent the bric-a-brac of urban construction picked up from the remote sensing. These plots do not show any lognormality which is perhaps surprising and when the right tail is excluded from the data, the linearity is even more apparent. In fact what we have done in fitting power laws to these data is fit the generic equation to only the top 10 percent of buildings.

The aggregate plots are shown in the thick black line in Figures 4(a) to (d) with the excluded data points in grey. We have estimated the scaling parameters β_A, β_L, β_H, and β_V from the appropriate rank-size equations using log-linear regression but we must note that as volume is a simple product of area and height $V_j = A_j H_j$, then this is a derived variable that does not have the same status as the raw data variables area, perimeter and height. In fact area and perimeter are confounded too as perimeter and area are both formed from the same two linear dimensions defining the rectangular blocks that make up the buildings data set. We include volume and area because these are two variables that are usually used in describing cities, notwithstanding the fact that they are composed of more basic geometric primitives. To illustrate the interdependence between these results, if the rank order r, for height and for area, were identical, that is for $A(r) \sim r^{-\beta_A}$ and $H(r) \sim r^{-\beta_H}$, then volume could be predicted as $V(r) \sim r^{-\beta_A} r^{-\beta_H}$. This is unlikely to be the case for we know that height is likely to increase faster than area as buildings seek space upwards. In short this is why we need to examine the allometric relations which relate the various quantities. Thus we might expect volume to decline more steeply with rank than area, which in turn is likely to fall more steeply than height or perimeter for this is the sequence of objects from 3 to 2 to 1 dimension.

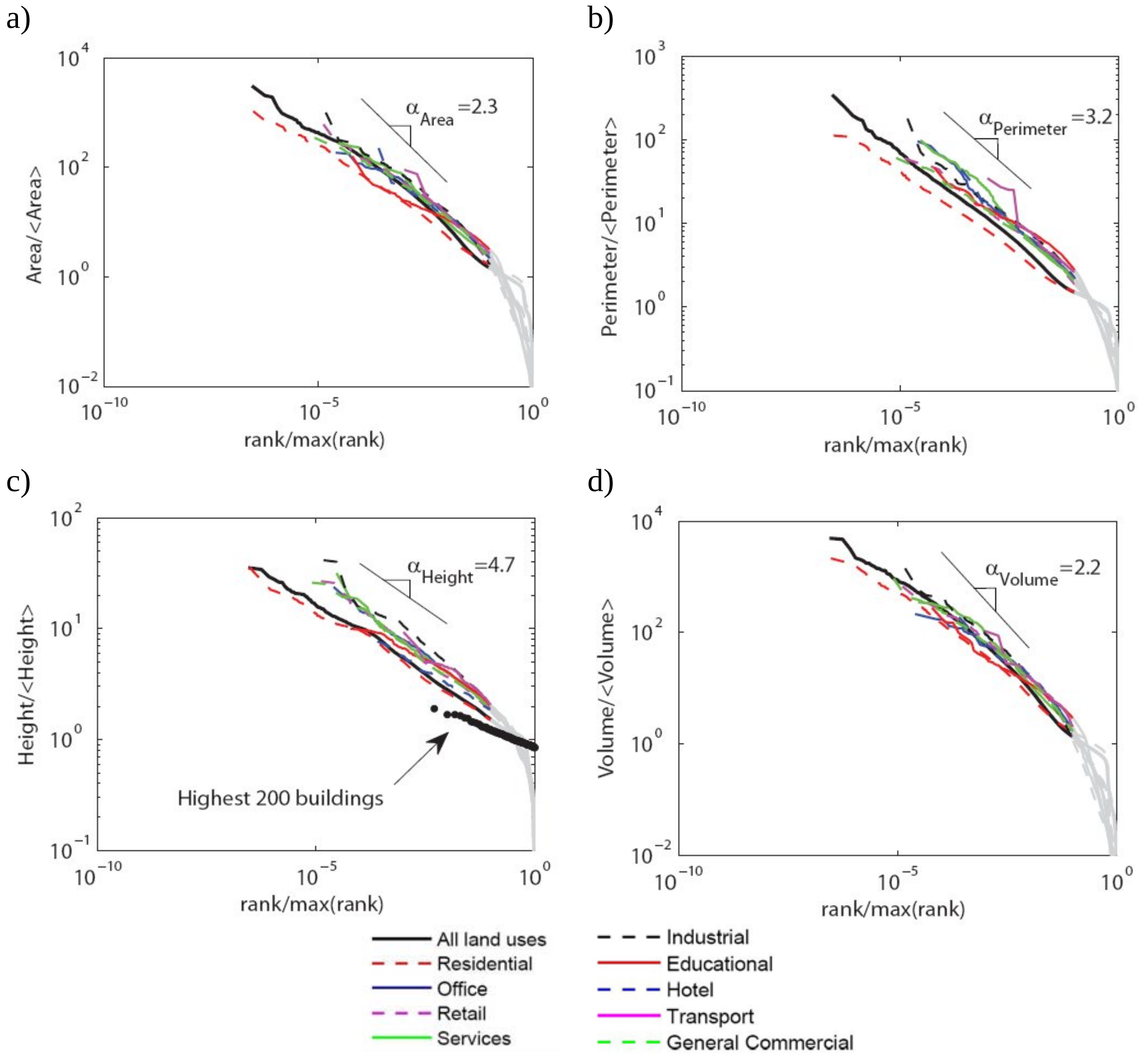


Figure 4: Normalized Rank-Order Plots
a) Building Area, b) Perimeter, c) Height and d) Volume

We plot all buildings (solid curves in black) and buildings classified by their land use (dashed and dotted curves). We also plot fits to the rank-size distribution for all buildings (all land uses) on each panel and compute the corresponding regression coefficients applying the least squares method to the top 10% ranks in each curve. Panel d) includes the rank-order plot of the height for the highest 200 buildings from the **Emporis** worldwide database

In Table 2, we present the results which also show the data for same scaling relations for the land uses. We have very dramatic linearity in the log-log plots over several orders of magnitude for volume from 10^7 to 10^2 after which the plot falls very steeply, implying that buildings less than 25 square meters in volume behave quite differently. These are really sheds and bric-a-brac referred to earlier and in future work will be discounted to an extent as we construct better building blocks (Steadman et al., 2000). These regressions are striking in their linearity and such rank-size relations are amongst the best we have come across. In fact this bears out the remarkable linearity of the rank-size of the heights of the top 200 buildings in the world which enabled us to make such good predictions of building heights further down the scale. The rank-size plots for the nine land use categories – residential, office, retail, services, industrial, educational, hotel, transport, and general-commercial – are also shown in Figure 4 with respect to area, perimeter, height and volume. We expected these plots to show rather different scaling from the aggregate (although 90 percent of the buildings in the database are classified as residential land use) but they are all close to the aggregate relations. From Figure 4, it is clear that their linearity tends to be over a lesser number of orders of magnitude. Any differences that do occur in these slopes are highlighted in Figure 5 which compares the β coefficients and their error bars

	All Land Uses	*Resid-ential*	*Office*	*Retail*	*Services*	*Indus-trial*	*Educa-tional*	*Comm-ercial*
N	3595689	3320579	39587	77075	33949	67270	16257	122874
Area β_A	0.763	0.559	0.711	0.802	0.664	0.840	0.486	0.711
Perimeter β_L	0.272	0.251	0.294	0.305	0.308	0.352	0.272	0.287
Height β_H	0.457	0.352	0.457	0.461	0.469	0.477	0.393	0.432
Volume β_V	0.861	0.688	0.834	0.923	0.841	1.007	0.570	0.841

Table 2: Scaling Parameters for Buildings in the London Database

Note that only the top 10 percent of these building numbers are used in the regressions and that Transport and Hotel have been excluded due to their small numbers

The six sets of allometric relations stated earlier in equations (17) are plotted in logarithmic form as two-dimensional surfaces in Figure 6. Only three of these relationships show the kind of linearity that we might expect from our earlier analysis, and these involve area v. perimeter, volume v. perimeter and volume v. area, that is those based on $A_j = Z_2 L_j^{\eta}$, $V_j = Z_3 L_j^{\mu}$, and $V_j = Z_6 A_j^{\vartheta}$. The other relationships involving height are quite scattered and require different techniques for extracting their allometry for clearly the set of data points must be culled to extract those that reflect the densest parts. As there are almost 3.6 million points in this scatter, their representation as surfaces coloured by their density after appropriate binning into a relatively fine scale set of categories is the most useful way of assessing these relationships. In Table 3, we present results from estimating the three allometric regression lines to the data in its logarithmic form.

.

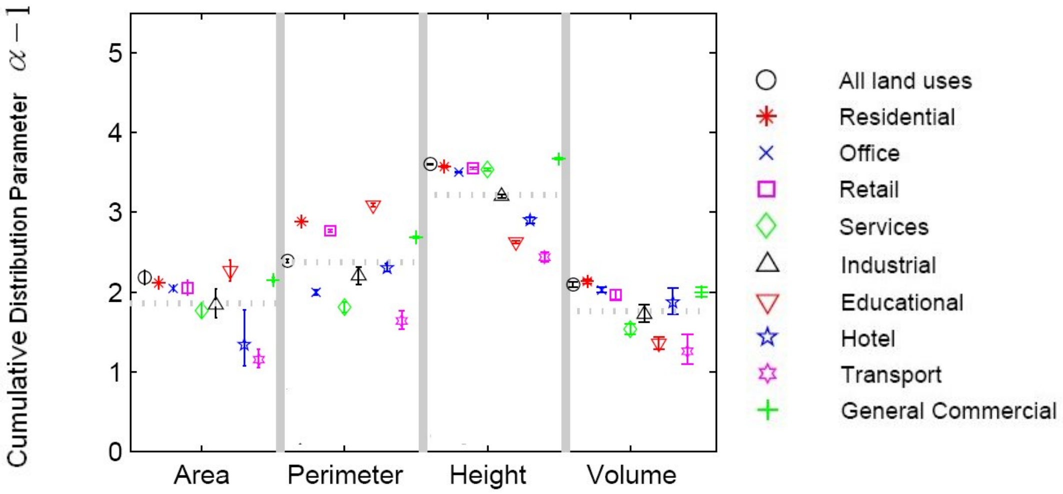


Figure 5: Scaling Coefficients for the Plots in Figure 4.

The least squares method is applied to the top 10% ranks in each curve. Error bars are 95% confidence intervals. Horizontal dashed lines (in grey) are mean values of $\alpha - 1$.

It is immediately clear that the value of the parameters is consistent with the order of the geometric scaling. That is, the parameter of area on perimeter is less the square while the value of the relation between volume and area is less than 3/2. This means that as the perimeter increases, the area increases less than the normal geometric relation implying that shape is changing, probably becoming more crennelated – implying a longer perimeter – as the area grows. In terms of volume, this increases at less than 3/2 of the area which suggests that the volume must get proportionately less as the area grows. This bears out the implied observation that as the surface grows, the shape must change.

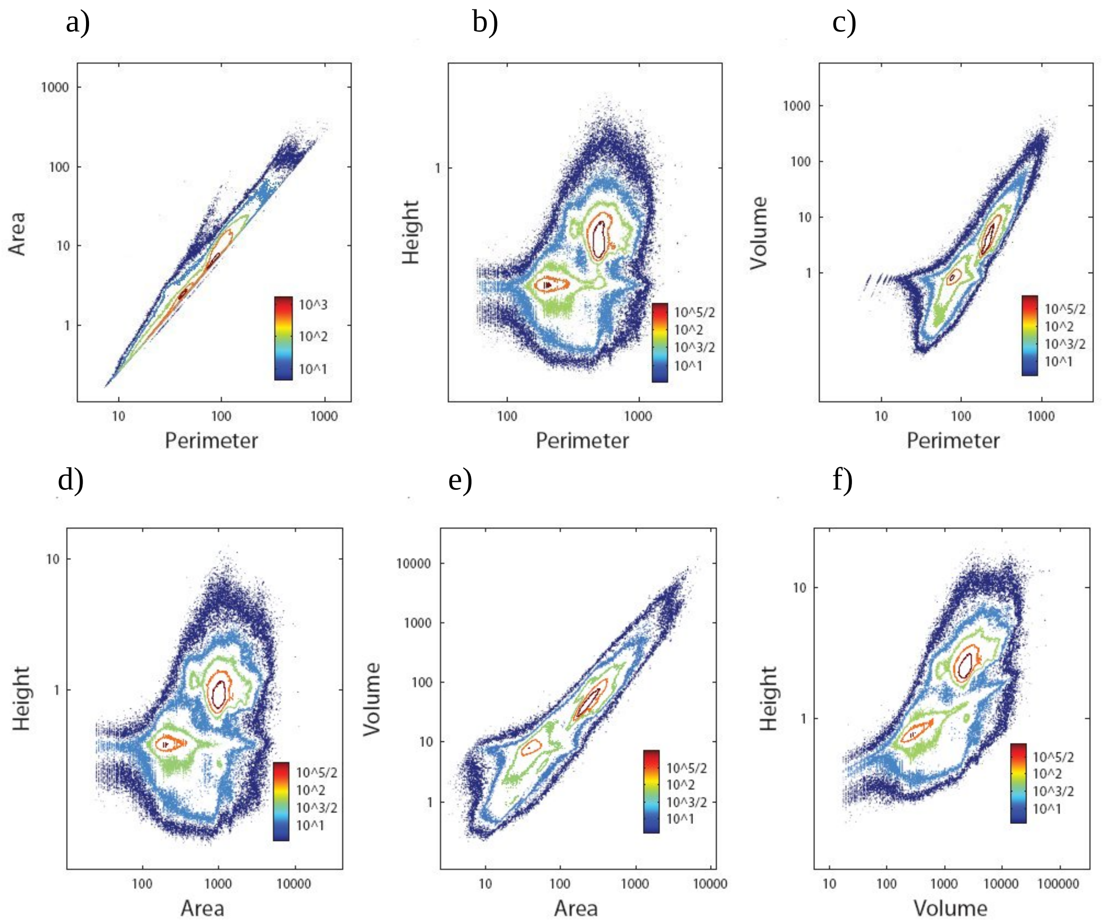


Figure 6: Two-Dimensional Surface Plots of Allometric Relations

a) Perimeter against Area, b) Perimeter against Height, c) Perimeter against Volume, d) Area against Height, e) Area against Volume and f) Volume against Height. Panels implying contour lines of logarithmically-binned histograms (frequency counts) on a logarithmic scale. Colour bars display the range of histogram values on each panel. We have found an approximate linear relation between the variables in panels a), c) and e).

Table 3 also contains all the parameter estimates for these three relationships for each individual land use. Remarkably these are all consistent with the aggregate and show that building volumes grow proportionately less than their increases in area as we might expect. It is even more urgent now however to extend the analysis of allometry to height as this is a key variable in defining volume and it is the weakest aspect of our work, largely because we cannot assume that usable building volumes are the same as geometric volumes. Moreover the height data itself is highly variable due to the fact that we have used mean height which is not necessarily a good measure for computing volume. This requires considerable further research as it is central to some of the notions in this paper which relate to how volume scales with plot area and to questions of surface area that define building skins. This is the research we will develop next when we link the buildings database to related databases of floorspace and energy emissions.

	Area v Perimeter		**Volume v Perimeter**		**Volume v Area**	
	Allometric Coefficient	R-Square	Allometric Coefficient	R-Square	Allometric Coefficient	R-Square
Euclidean Scaling	2		3		3/2	
All Land Uses	1.832 (1.832, 1.833)	0.962	2.386 (2.385, 2.387)	0.811	1.296 (1.296, 1.297)	0.835
Residential	1.846 (1.846, 1.846)	0.963	2.463 (2.461, 2.464)	0.825	1.325 (1.324, 1.326)	0.845
Office	1.783 (1.779, 1.787)	0.952	2.152 (2.141, 2.162)	0.808	1.199 (1.194, 1.204)	0.838
Retail	1.811 (1.808, 1.814)	0.958	2.215 (2.207, 2.222)	0.816	1.216 (1.212, 1.219)	0.842
Services	1.773 (1.769, 1.777)	0.964	2.129 (2.118, 2.140)	0.814	1.195 (1.189, 1.200)	0.836
Industrial	1.788 (1.786, 1.792)	0.957	2.052 (2.042, 2.062)	0.706	1.148 (1.142, 1.153)	0.738
Educational	1.679 (1.673, 1.684)	0.959	1.901 (1.888, 1.914)	0.828	1.132 (1.125, 1.139)	0.862
Hotel	1.770 (1.760, 1.780)	0.969	2.143 (2.115, 2.172)	0.849	1.207 (1.193, 1.222)	0.870
Transport	1.775 (1.749, 1.801)	0.948	1.991 (1.928, 2.053)	0.797	1.116 (1.085, 1.147)	0.833
General Commercial	1.813 (1.811, 1.815)	0.956	2.179 (2.173, 2.185)	0.813	1.194 (1.191, 1.197)	0.840

Table 3: Coefficients and Correlations for the Allometric Relations

The numbers in brackets in the coefficient columns give the 95% confidence intervals

The Spatial Distribution of Building Geometries

To put space back into the argument, we can examine the two-dimensional distribution of building geometries in Greater London by computing the correlation functions with respect to how properties of a building – area, height and so on – vary with respect to every other building. From our earlier analysis of the spatial distribution of population densities with respect to how density varies around one point – the CBD, we might except that these correlation functions would imply power laws with respect to increasing distance from any building in question. In this section, we will compute a composite correlation function in the following way, assuming that building properties meet the definitions of a point process.

The first moment of such a point process can be specified by a single number, the intensity ρ giving the expected number of points per unit area. The second moment can be specified by Ripley's K function (Ripley, 1977) where $\rho K(R)$ is the expected number of points within distance R of an arbitrary point of the pattern. The product density

$$\rho_2(x,y)dA(x)dA(y) = \rho^2 g(R)dA(x)dA(y) \tag{18}$$

describes the probability of finding a point in the area element $dA(x)$ and another point in $dA(y)$, at distance $R = |x-y|$, and $g(R)$ is the two-point correlation function. Ripley's K function is related to $g(R)$ as

$$K(R) = 2\pi \int g(R)dR \quad . \tag{19}$$

In other words, $g(R)$ is the density of $K(R)$ with respect to the radial measure RdR (Stoyan, 2000). The benchmark of complete randomness is the spatial Poisson process, for which $g(R) = 1$ and $K(R) = \pi R_2$, the area of the search region for the points. Values larger than this indicate clustering on that distance scale, and smaller values indicate regularity.

The two-point correlation function can be estimated from N data points $x \in D$ inside a sample window W as

$$g(R) = \frac{|W|}{N(N-1)} \sum_{x \in D} \sum_{y \in D} \frac{\Phi_R(x,y)}{2\pi R\Delta} \omega(x,y) \quad , \tag{20}$$

where $2\pi R\Delta$ is the area of the annulus centred at x with radius R and thickness Δ (Kerscher, Szapudi, and Szalay, 2000). Here $|W|$ is the area of the sample window, and the sum is restricted to pairs of different points $x \neq y$. The function Φ_R is symmetric in its argument and $\Phi_R(x, y) = [R \le d(x, y) \le R + \Delta]$ where $d(x, y)$ is the Euclidean distance between the two points and the condition in brackets equals 1 when true and 0 otherwise. The function $\omega(x, y)$ accounts for a bounded W by weighting points where the annulus intersects the edges of W. There are a number of edge-corrections available, but that developed by Ripley (1976) has a long tradition both in human geography and physics (Carvalho and Batty, 2006). Here we approximate $\omega(x, y) = 1$ as the city does not have clear spatial boundaries.

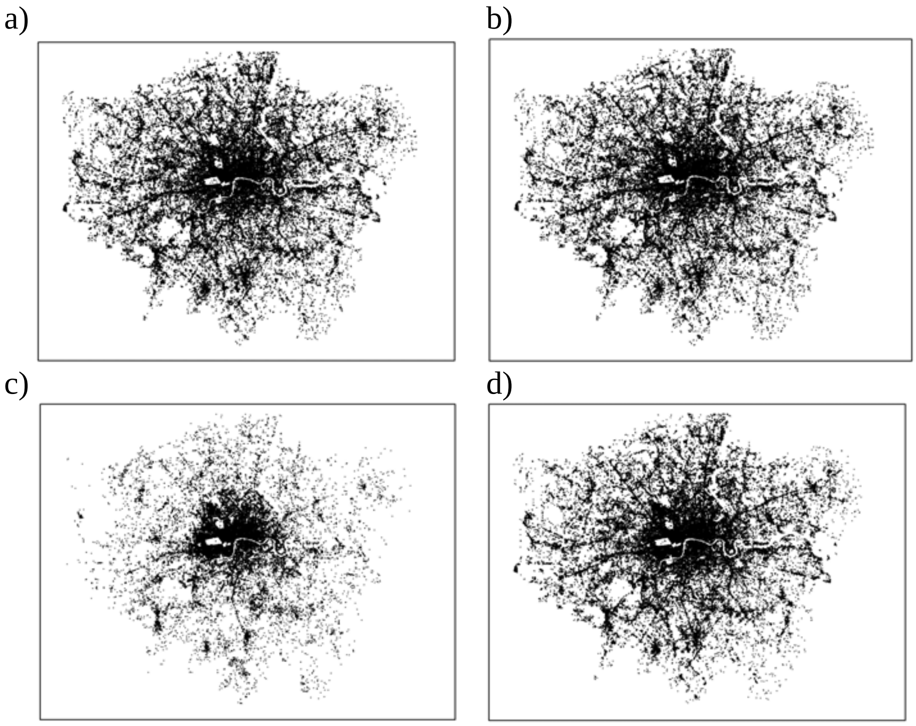


Figure 7: Spatial Distribution of the Geometric Properties of the Highest 100,000 Buildings

a) Area b) Perimeter c) Height, and d) Volume

Of special physical interest is whether the two-point correlation is scale-invariant. A scale-invariant $g(R)$ is an indicator of a fractal distribution of points, and is expected in critical phenomena (Kerscher, Szapudi, and Szalay, 2000). Figure 7 shows the distribution of geometric building properties over Greater London for the largest 100,000 selected by height for a range of distances up to $r \cong 3.3$ km and Figure 8 is a plot of the two-point correlation function on a double logarithmic scale. We observe a power-law decay of $g(R) \sim r^{-0.230} \sim r^{-\gamma}$ for these largest 100,000 buildings.

Interestingly, the two-point correlation function does not display scaling behaviour if we select the 100,000 largest buildings by perimeter size or area. This suggests that building height is a major variable which has so far been overlooked in studies of the fractality of cities and this supports our preliminary analysis of height from related databases.

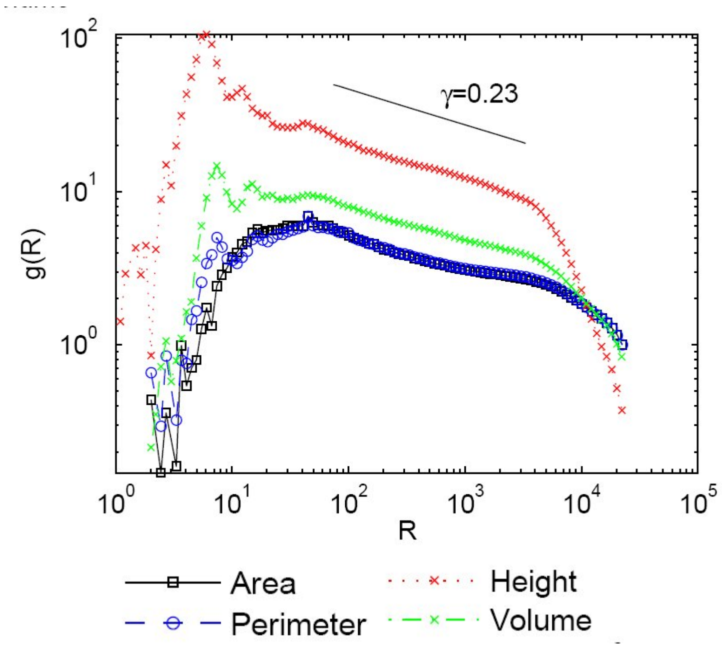


Figure 8: Two-Point Correlation Functions of the Building Geometries with Respect to Distance R

.

Conclusions and Next Steps

Our analysis represents a first step in developing scaling and allometry for spatial distributions within cities and this suggests a research program complementary to that being developed for equivalent relationships between cities (Bettencourt, Lobo, Helbing, Kuhnert, and West, 2007). The link between the rank-size scaling of spatial attributes which suppresses the spatial pattern and the scaling of the spatial patterns with respect to distance which we examined in this paper initially as a demonstration of lognormality and lastly in terms of two-point correlation functions, needs to be explored in considerably more depth. We also need to investigate the relationship between geometric and socio-economic attributes as reflected in the link between building geometries and population densities as this serves to link the physical form of the city to its functioning.

Definitional problems abound when data which is spatial are explored. Data based on individual objects such as people frequently does not display spatial pattern until it is aggregated. Although attributes such as income do accord to scaling for individuals, many others are only retrieved when the data is aggregated to some specific level and thus the degree to which it is aggregated is critical. We need to revisit these definitional issues in more detail and in the case the database used here, iron out many of the problems of building size and type that we have identified. The analysis should be extended to deal with different rank-size and allometric relations in different areas of the city, showing how these relations might change as implied in the distributions pictured, for example, in Figure 8.

We are much encouraged by the very strong scaling implicit in this data. Of course to confirm this, we need more examples from other cities. We need to relate the physical geometry to other measures, particularly linear measures such as utilities and street systems as well as socio-economic activity volumes as proposed by Kuhnert, Helbing, and West (2006) amongst others. We need to link the analysis much more strongly to fractal geometry (Batty, 2005) and we need to link it to circulation patterns in buildings (Bon, 1973; Steadman, 2006). The extension which we are about to do is to examine the surface areas of buildings linking these to energy emissions and related phenomena and when we do this, the variations in these relations with respect to different locations and districts within the city will take on new meaning. In time, we hope that such work will add to our growing knowledge of how efficient cities are in terms of their geometry and in this sense, provide a much more considered position on issues such as urban sprawl and the compact city.

References

Aitchison, J., and Brown, J. A. C. (1957) **The Lognormal Distribution: with special reference to its uses in economics**, Cambridge University Press, Cambridge, UK.

Anas, A., Arnott, R., and Small, K. (1998) Urban Spatial Structure, **Journal of Economic Literature**, **36**, 1426-1464.

Batty, M. (2005) **Cities and Complexity: Understanding Cities with Cellular Automata, Agent-Based Models, and Fractals**, The MIT Press, Cambridge, MA.

Batty, M. (2006) Hierarchy in Cities and City Systems, in D. Pumain (Editor) **Hierarchy in the Natural and Social Sciences**, Springer, Dordrecht, Netherlands, 143-168.

Batty, M. and Hudson-Smith, A. (2005) Urban Simulacra, **Architectural Design**, **75** (6), 42-47.

Batty, M. and Longley, P. A. (1994) **Fractal Cities: A Geometry of Form and Function**, Academic Press, London and San Diego, CA.

Berry, B. J. L. (1964) Cities as Systems within Systems of Cities, **Papers and Proceedings of the Regional Science Association**, **13**, 147–164.

Bettencourt, L. M. A., Lobo, J., Helbing, D., Kühnert, C., and West, G. B. (2007) Growth, Innovation, Scaling, and the Pace of Life in Cities, **Proceedings of the National Academy of Sciences**, **104**, 7301-7306.

Bon, R. (1973) Allometry in the Topologic Structure of Architectural Spatial Systems, **Ekistics**, **36** (215), 270-276.

Bon, R. (1979) Allometry in the Topologic Structure of Transportation Networks, **Quality and Quantity, 13**, 307-326.

Bonner, J. T. (2006) **Why Size Matters: From Bacteria to Blue Whale**, Princeton University Press, Princeton, NJ.

Carvalho, R. and Batty, M. (2006) The Geography of Scientific Productivity: Scaling in US Computer Science, **Journal of Statistical Mechanics**, **10**, P10012, 1-11.

Carvalho, R., and Penn, A. (2004) Scaling and Universality in the Micro-Structure of Urban Space, **Physica A**, **332**, 539–547.

Clarke, C. (1951) Urban Population Densities, **Journal of the Royal Statistical Society**, Series A, **114**, 490-496.

Clauset, A., Shalizi, C. R., and Newman, M. E. J. (2007) Power-Law Distributions in Empirical Data, **arXiv.org**, http://arxiv.org/abs/0706.1062v1 (accessed December 3rd 2007)

Cowan, P., and Fine, D. (1969) On the Number of Links in a System, **Regional Studies**, **3**, 235-242.

Gibrat, R. (1931) **Les Inégalités Économiques**, Librarie du Recueil Sirey, Paris.

Huxley, J. S. (1932, 1993) **Problems of Relative Growth**, The Johns Hopkins University Press, Baltimore, MD.

Isalgue, A., Coch, H., and Serra, R. (2007) Scaling Laws and the Modern City, **Physica A**, **382**, 643-649.

Kerscher, K., Szapudi, I., and Szalay, A. S. (2000) A Comparison of Estimators for the Two-Point Correlation Function, **Astrophysics Journal**, **535**, L13-L16.

Kuhnert, C., Helbing, D., and West, G. (2006) Scaling Laws in Urban Supply Networks, **Physica A**, **363**, 96-103.

Lammer, S., Gehlsen, B., and Helbing, D. (2006) Scaling Laws in the Spatial Structure of Urban Road Networks, **Physica A**, **363**, 89-95.

Newman, M. E. J. (2005) Power Laws, Pareto Distributions and Zipf's Law, **Contemporary Physics**, **46**, 323–351.

Ripley, B. D (1976) The Second-Order Analysis of Stationary Point Processes, **Journal of Applied Probability**, **13**, 255-266.

Ripley, B. D. (1977) Modelling Spatial Patterns, **Journal of the Royal Statistical Society**, **Series B**, **Statistical Methodology**, **39**, 172-212.

Samaniego, H., and Moses, M. E. (2007) Cities as Organisms: Allometric Scaling of Urban Road Networks, Department of Computer Science, University of New Mexico, Albuquerque, NM.

Steadman, P. (2006) Allometry and Built Form: Revisiting Ranko Bon's Work with the Harvard Philomorphs, **Construction Management and Economics**, **24**, 755-765.

Steadman, P., Bruhns, H. R., Holtier, S., Gakovic, B., Rickaby, P. A., and Brown, F. E. (2000) A Classification of Built Forms, **Environment and Planning B**, **27**, 73 – 91.

Stoyan, D. (2000) Basic Ideas of Spatial Statistics, in: K. Mecke, D. Stoyan (Editors) **Statistical Physics and Spatial Statistics**, Springer-Verlag, Heidelberg, DE, 3-21.

Thompson, D'Arcy, W. (1917, 1971) **On Growth and Form**, Abridged Edition, Cambridge University Press, Cambridge, UK.

Von Bertalanffy, L. (1973) **General Systems Theory**, Penguin Books, Harmondsworth, UK.

West, G. B., Brown, J. H., and Enquist, B. J. (1999) The Fourth Dimension of Life: Fractal geometry and the Allometric Scaling of Organisms, **Science**, **284**, 4 June, 1677-1679.

Zipf, G. K. (1949) **Human Behavior and the Principle of Least Effort**, Addison-Wesley, Cambridge, MA.