

Self-organized Criticality and Urban Development

MICHAEL BATTY^{a,*} and YICHUN XIE^b

^a*Centre for Advanced Spatial Analysis, University College London, 1-19 Torrington Place, London WC1E 6BT, UK;*

^b*Department of Geography and Geology, Eastern Michigan University, Ypsilanti, MI 48197, USA*

(Received 16 November 1998)

Urban society is undergoing as profound a spatial transformation as that associated with the emergence of the industrial city two centuries ago. To describe and measure this transition, we introduce a new theory based on the concept that large-scale, complex systems composed of many interacting elements, show a surprising degree of resilience to change, holding themselves at critical levels for long periods until conditions emerge which move the system, often abruptly, to a new threshold. This theory is called ‘self-organized criticality’; it is consistent with systems in which global patterns emerge from local action which is the hallmark of self-organization, and it is consistent with developments in system dynamics and their morphology which find expression in fractal geometry and weak chaos theory. We illustrate the theory using a unique space–time series of urban development for Buffalo, Western New York, which contains the locations of over one quarter of a million sites coded by their year of construction and dating back to 1773, some 60 years before the city began to develop. We measure the emergence and growth of the city using urban density functions from which measures of fractal dimension are used to construct growth paths of the way the city has grown to fill its region. These phase portraits suggest the existence of transitions between the frontier, the settled agricultural region, the centralized industrial city and the decentralized postindustrial city, and our analysis reveals that Buffalo has maintained itself at a critical threshold since the emergence of the automobile city some 70 years ago. Our implied speculation is: how long will this kind of urban form be maintained in the face of seemingly unstoppable technological change?

Keywords: Urban growth, Urban density, Self-organized criticality, Phase transitions, Fractal dimension, Buffalo, New York

THE URBAN TRANSITION

Paul Kennedy (1993) in his book *Preparing for the Twentyfirst Century* suggests that during the next 25 years, the rate of world urbanization will be greater than at any time in previous history and probably at any time thereafter. By 2025, the population which

lives in cities will have increased from 37 percent at present to close on 60 percent. New cities will not be formed around manufacturing, nor around services in the conventional sense, but around the confluence of global flows of capital and labor, with a somewhat less local economic rationale, at least in the developing world, from that traditionally

* Corresponding author. E-mail: mbatty@ucl.ac.uk.

ascribed to the developed. Cities in developed societies will also lose their traditional economic focus around their historical cores as they continue to spread and when entire societies become urbanized in this fashion, the very concept of the city will have to be redefined. What physical form the urban future will take is entirely unclear. It may be similar to what already exists but it could equally be very different for there is little sense, as yet, as to how the deep-seated waves of technological change in computation and communication will play themselves out.

These trends suggest that we are about to cross some sort of threshold as we pass from cities whose physical form is still reminiscent of industrial society based on limited physical mobility and abundant urban space, to cities where opportunities for electronic mobility are vast but where available living space will be scarcer. The transition from agricultural to industrial society heralded in an age where cities could be seen as simply larger versions of the market towns and villages from which they grew. But this was short-lived for the arrival of the automobile produced a much more dispersed form which is more characteristic of what, perhaps, we should now refer to as the industrial city. What will follow is unclear, but apart from these initial thoughts, we will not try to answer this. Here we will be much more concerned with how to formally describe and measure the potential transition, and we will do this by attempting to measure the earlier transitions from data which refer to the way industrial cities have grown during the last 200 years.

We will introduce a series of ideas from contemporary systems theory which provide a consistent rationale for integrating the dynamics of systems with their form and function. Per Bak and his colleagues have devised a theory for large-scale, complex systems which suggests that as such systems evolve, they reach a critical threshold embodying a fragile equilibrium which is maintained through a process of self-organization (Bak *et al.*, 1988). This is a theory of weak chaos which they argue has very wide applicability to many

natural and artificial systems whose dynamics consists of the action of local agents which generate highly ordered global patterns. They argue that the spatial and temporal 'fingerprints' or 'signatures' of such systems have no characteristic length scales, and are therefore fractal. Self-organized criticality, as the theory has been called, is being applied to a diverse range of systems; the most persuasive examples involve earthquakes, forest fires, and ecologies, but the theory is finding favor in areas as different as the origins and evolution of life (Bak and Sneppen, 1993), and the dynamics of the stock market (Krugman, 1994). In short, self-organized criticality is a good candidate for explaining the evolutionary dynamics that can lead to any systems whose temporal and spatial characteristics appear fractal (Bak and Paczuski, 1993).

It is this applicability to fractal systems that makes the theory relevant to urban form and structure, to the morphology of cities. During the last decade, fractal ideas have been applied to urban form, exploiting the evident self-similarity which exists within and between cities through their various hierarchical orders such as central place theory, transportation networks, and the scaling laws which are consistent with urban densities (Batty and Longley, 1994; Frankhauser, 1994; Benguigui, 1992). None of this is very surprising given the extensive interest in physical form which fractal theory and computer graphics have jointly stimulated. And many of these ideas concerning cities were already buried deep within urban and regional theory, ready to be rediscovered and extended as soon as interest in linking the physical form of the city to its socio-economic functioning revived.

The prospect now exists, for the first time, of developing a coherent and consistent dynamics of urban evolution which is built around the current fascination with the highly decentralized complex systems whose operation is at the local level, and which generates urban forms which are consistent with the fractal patterns that have been widely observed for cities. The task we set ourselves in this paper is to show how this might be approached, to provide a simple test of the theory's consistency,

and to apply these ideas to the major problem of measuring the urban transition to a postindustrial world. We urgently need to think about the future evolution of urban form in the context of the shift from cities based on energy and industry to those based on information and services. The emergence of 'world cities' and 'edge cities' are two features of this change and the kinds of dynamic theories that we will allude to here have, at least in principle, the ability to make some sense of this type of phase transition.

There is another relevant theme. Digital data are becoming available for cities which record the detailed location and attributes of individual sites or land parcels, and from such sources, the urban morphology of cities can be measured and visualized across many scales. An important attribute of such data, largely due to the fact that it is collected for taxation purposes, is the age at which construction on the parcel first took place. When combined with data from other sources, this is providing varieties of space–time series which have rarely been available hitherto. Theories which purport to link local to global dynamics in space–time such as self-organized criticality, now have a real chance of empirical verification, at least in some part.

We will first introduce the theory, arguing that systems whose morphologies display some stability through time are likely to lie at a critical threshold, which once disturbed, can generate abrupt transitions to new regimes. The fingerprint of criticality used here will be a measure of the fractal dimension which, in the case of cities, is a measure of the rate and extent to which the city fills its available space through urban growth. There are other measures of criticality although for real systems, these are hard to observe from data and thus our test is necessarily a partial one. We present the measures of space-filling and urban density next, and then review the data base which we use to test the theory. This is based on several attributes at the scale of land parcels in Buffalo, Western New York State, amongst which the location and year of construction of all taxable properties in the metropolitan area in 1989, have been recorded. We can construe

an appropriate space–time series from this and thus measure the way the fractal dimension of the city has changed over the last 200 years, from the time when the agricultural frontier was first settled, through the early growth of the industrial city as an entrepot within the Great Lakes economic system, until the emergence of the automobile city during the last 70 years.

In later sections, we estimate and measure various dimensions and densities associated with this growth, and then discuss the extent to which these results are consistent with the idea of self-organized criticality. As we are at pains to imply, our conclusions are tentative but they do suggest that the city has reached a critical point. Our speculation for Buffalo, as an archetype for all emergent postindustrial cities, must be that a major phase transition from the current urban regime to one which is consistent with a new technological era is increasingly likely. Our analysis for Buffalo does not suggest such an imminent change although we argue that elsewhere, in more vibrant urban economies, such a transition might already be detectable through the methods we introduce here.

SELF-ORGANIZED CRITICALITY

Change occurs in cities through the addition of new activities such as births and immigration, the deletion of activities through deaths and emigration, but with most activities changing the pattern of development through processes of redistribution. Whenever an activity changes its location, this sets off a chain reaction in which other activities are motivated to move as economic agents which compose such activities readjust their locations to the changed circumstances. The causes of such reactions need not concern us here. They may be due to life-cycle effects, preferences concerning segregation and clustering, the changing economic accessibility of various parts of the city, the supply of new development sites or the demolition of old.

The fact that the city continues to exist in much the same form while these myriad of reactions play themselves out within its overall fabric means that such reactions do not continue indefinitely. More importantly, they are consistent with maintaining the existing organizational morphology of the city.

Self-organized criticality is a theory built around these interaction effects. In essence, the theory postulates that when activities initiate moves in time and space, the chain reactions which result from such moves, follow distributions in time and space which occur on all scales. In other words, these reactions can range from simply one isolated move to moves which involve all activities in the system; there is no characteristic length of chain in duration, no characteristic number or size of activities involved, and no characteristic distance over which the reaction takes place. In fact, experiments with theoretical systems suggest that the duration and size of activities involved in such reactions follow power laws. The key finding for such systems is that systems evolve to a form which embodies the critical state in which these reactions continually occur in such a manner that the critical state of the system is preserved.

Bak *et al.* (1989) say:

“The canonical example of self-organized criticality is a ‘pile of sand’. Imagine building up the pile by slowly adding particles. As the pile grows, there will be bigger and bigger avalanches. Eventually a statistical stationary state is reached in which avalanches of all sizes occur, that is the correlation length is infinite. Thus in analogy with equilibrium thermodynamical systems, the state is ‘critical’. It is self-organized because no fine-tuning of external fields was needed to take the system to the critical state: the criticality is unavoidable.”

In the sand pile example which immediately generalizes to related geophysical phenomena such as earthquakes, river flow systems, and volcanic eruptions (Turcotte, 1992; Rinaldo *et al.*, 1993), the critical state of the pile is its slope. Once this slope is reached, dropping an additional grain of sand on the top of the pile will cause an avalanche, thus changing the critical slopes elsewhere in the pile. Further grains will cause more avalanches, all of

which will be of different sizes and durations as the pile continues to build back up to its critical value. From experiments, the distributions of avalanches over time – that is, the number of avalanches of duration time t , sometimes called lifetimes – is given as $n(t) \sim t^{-\beta}$ with the exponent $\beta \approx 1.0$ while distribution over size – the number of avalanches of size s – is given as $n(s) \sim s^{-\tau}$ with the exponent $\tau \approx 0.4$. As size varies with time, these exponents can be simply related, the precise relation depending upon the physical configuration of the system under study. The language and methods of the theory are derived from statistical physics and related to the ideas of phase transitions which make their appearance in fractal growth (Barabasi and Stanley, 1995) and percolation theory (Stauffer and Aharony, 1992).

This might seem far removed from urban development although the mechanisms of change which drive the dynamics appear similar. The problem however in applying this theory to large-scale systems which are in some sense remote from physical experimentation is assembling data on the dynamics; in most applications to date, such testing has been with computer simulations of what are essentially idealized systems. Two applications suggest quite close parallels with urban systems. The first is based on an analysis of the ‘Game of Life’, a simple cellular automata which is usually played out on a two-dimensional grid in which cells become ‘live’ if they are surrounded by three live neighbors, or ‘die’ if they are surrounded by more than three (overcrowding) or less than two (isolation). Bak *et al.* (1989), and Alstrom and Leao (1994) show that the distributions $n(t)$ and $n(s)$, which are formed from the chain reactions when a dead cell is made live, follow power laws with exponents $\beta \approx 1.4$ and $\tau \approx 1.6$. This implies that ‘Life’ is a model which generates self-organized criticality. More important from our perspective is that the number of active sites at distance r from the original site, which are set-off in the chain reaction, scale as $n(r) \sim r^{D-1}$ where D is a fractal dimension computed as approximately 1.7. This is the finding that links this type of dynamics to the theory of the

fractal city in which exponents of this value have been widely observed (Batty and Longley, 1994; Frankhauser, 1994).

Fractal growth models based on simple diffusion mechanisms such as diffusion-limited aggregation (DLA) provide very good analogs for city growth in which the growth seed is the central business district (CBD). These models have been widely researched in non-equilibrium physics and the general consensus is that their fractal dimension is approximately 1.7 in two-dimensional systems. Batty (1991) has shown how this model can be used to simulate urban growth and how robust the value of this dimension is as the size and shape of the space within which the city is grown change. Similar models have been used for urban transport networks (Benguigui, 1992). A more direct link between fractal growth models and self-organized criticality has been forged by Alstrom (1990) who using branching theory provides an analysis of the chain reactions contained in the dendrites which make up such fractal growth. His model in fact suggests that the fractal dimension of such growth is $D = \ln 3 / \ln 2 \approx 1.585$ while further experimental evidence by Alstrom *et al.* (1990) suggests that both DLA models and those based on invasion percolation can be conceived of entirely as sets of chains, the growth they simulate thus being consistent with self-organized criticality.

Comprehensive testing of this theory depends upon observations of the dynamics at the most micro level in terms of chain reactions in time and space as well as at the macro level in terms of fractal patterns that are generated. For social systems, such micro level information on chains is almost impossible to collect. It does exist in some housing market research, and it appears consistent with White's (1970) labor market vacancy chains and Schelling's (1978) models of residential segregation and ordering which are based on reactive diffusions of the kind implied by self-organized criticality. But in general, it is unlikely that data sets can be easily assembled from which the lifetime and size distributions, $n(t)$ and $n(s)$, could be estimated. It is much more likely that the macro patterns consistent

with these dynamics can be measured. There is now substantial work on measuring the fractal dimension associated with the population density profiles of different cities (Batty and Longley, 1994; Frankhauser, 1994), and if this data were available through time, then it would be possible to test whether or not these patterns were consistent with self-organized criticality. This of course would be a very partial test of this theory but it would provide some initial support. In fact, very little urban theory ever gets tested in a comprehensive sense for most is only validated at the occasional points where it touches the real world.

Our task then in this paper is very clear. If we can show that the fractal dimension of real cities is comparatively stable over long periods of time, we will have some confidence in thinking that the theory of self-organized criticality has some relevance to the way cities develop. This will force us to consider other implications of the theory and other ways in which it might be tested. It will also force us to think more incisively about the ways in which growing systems manifest such criticality and how such criticality itself might change as technologies which govern behavior in time and space change. In short, the theory looks promising and some bold claims have been made by its proponents. Bak and Chen (1991) say: "To our knowledge, self-organized criticality is the only model or mathematical description that has led to a holistic theory for dynamic systems".

URBAN DENSITY DISTRIBUTIONS AS FRACTAL GROWTH

In western cities, population fills the space around the origin of urban growth, typically the CBD, according to an inverse distance relation reflected in the power function

$$\rho(r) \sim r^{-\alpha}, \quad (1)$$

where $\rho(r)$ is the density of population or development at distance r from the origin (CBD), and α is a

parameter of the distribution controlling the rate at which density declines; strictly this is an elasticity. The area associated with the distance r from the CBD is $A(r) = \pi r^2$ and thus from Eq. (1), the population $n(r)$ associated with an increment of area defined as $dA(r) = 2\pi r$ is

$$n(r) = \rho(r) dA(r) \sim r^{1-\alpha}. \quad (2)$$

We also need to compute the accumulation of densities – the cumulative population $N(R)$ – up to distance R from the center whose origin is arbitrarily set at $r = 1$. Then from Eq. (1) or Eq. (2)

$$\begin{aligned} N(R) &= \int_{A(r)} n(r) dr \sim \int_{r=1}^{r=R} \rho(r) 2\pi r dr \\ &\sim R^{2-\alpha} = R^D, \end{aligned} \quad (3)$$

the cumulative density $P(R)$ thus being defined as

$$P(R) = \frac{N(R)}{A(R)} \sim R^{-\alpha}. \quad (4)$$

We will use these four relations in Eqs. (1)–(4) as our basic measures of urban spatial structure throughout this paper. Note that the variable r measures the distance associated with the density or population at a point r and the variable R , the distance associated with the density or population accumulated up to the point R .

We use a power function rather than a negative exponential because as we argue elsewhere, this function is scale-invariant which we believe is a necessary condition for the analysis of systems which are growing in size and scale. Such functions are consistent with the notion that the morphology and density of the city is fractal, self-similar over several orders of scale, hence applicable at different scales. The density parameter α is thus directly related to the fractal dimension $D (= 2 - \alpha)$ which is a measure of the extent to which two-dimensional space is filled by the city. If all space is filled, $D = 2$ and $\alpha = 0$ (uniform density), while if only a line across the space is filled then $D = 1$ and $\alpha = 1$. There are many arguments which we present elsewhere that $1 < D < 2$ and $0 < \alpha < 1$, and there is now considerable empirical

evidence that $1.6 < D < 1.9$ (Batty and Longley, 1994).

Our test for self-organized criticality is quite straightforward. We will compute the fractal dimension D from the Buffalo data for all the years in the space–time series, using several methods of estimation which we discuss in the Appendix to this paper. Estimating fractal dimension has the same problems as other methods of parameter estimation in that different methods can lead to widely varying values. We will postpone any further discussion of these problems until we have explored the data set but it is important to note that density functions detect two related but often competing aspects of spatial pattern – the degree to which space is filled as well as the rate at which it is filled. Different methods emphasize these two features in different ways.

THE GROWTH OF METROPOLITAN BUFFALO

Our space–time series contains all developed sites by age and location which existed at the end of March, 1989, in the Buffalo region. The data source is from the New York State Equalization and Assessment Board which monitors local property taxation, and which is available digitally as raw data records and in ARC-INFO coverages. These types of file are updated yearly and they contain extensive details of the physical condition and size of all taxable properties. In 1989, there were some 336,334 records in the Erie County file within which the Buffalo metropolitan area is situated. Erie County is a reasonably good approximation to Buffalo’s urban field although development to the north in Niagara County and some very recent suburban development to the east are excluded. However of these records, only 250,455 are usable for the remainder, mainly to the south and south east of the urban area and on the southern border of the region, have not been coded by year of construction. We have used various methods to approximate this missing data, mainly remote sensing, although we will use the original not modified data

set here. We are however confident that the missing data do not invalidate, in any way, the conclusions that we are able to draw here (Batty and Howes, 1996). Clearly, this data set only contains a history of the city from the vantage point of 1989, and given the constraints on missing data and the fixed regional space, all our subsequent analysis is necessarily tentative.

Density models have been traditionally based on population enumerated within large areas such as blocks or tracts, rather than on individual sites for which the actual location of each is known. Thus our densities are densities of development rather than population. The cumulative space-time series aggregated to 30 year intervals from 1820 and

represented at 100m grid square resolution is illustrated in Fig. 1 from which the difficulty of plotting many thousands of points to give a complete impression of historical development is obvious. The entire data are plotted at a larger scale, but with the same resolution, in Fig. 2 where the grey tones code development in 1989 and the underlying topography are illustrated. The missing data are clearly visible although as it is an arbitrary slice through the chronology of the city's development, it is easy to show that its exclusion does not distort the patterns implied within the remaining series (Batty and Howes, 1996).

An explanation of these growth patterns is required. Until around 1810, Buffalo was a frontier

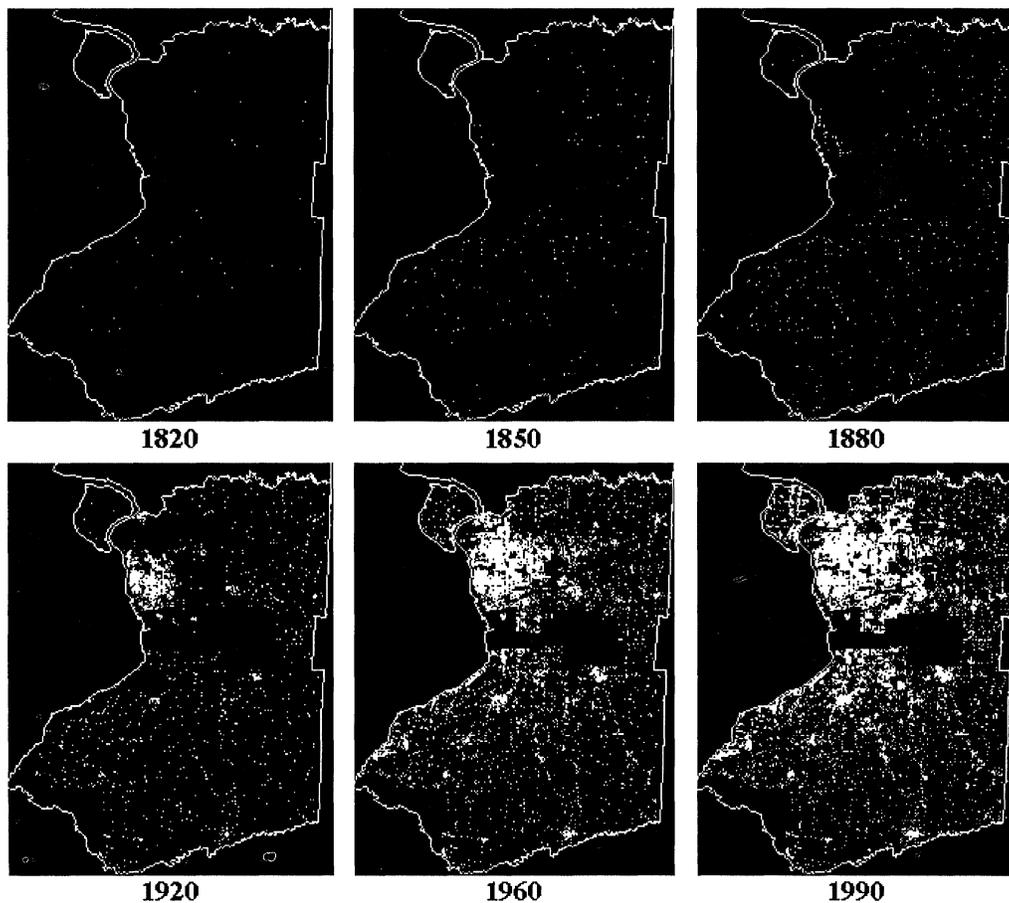


FIGURE 1 The growth of Buffalo from 1750 to 1990.

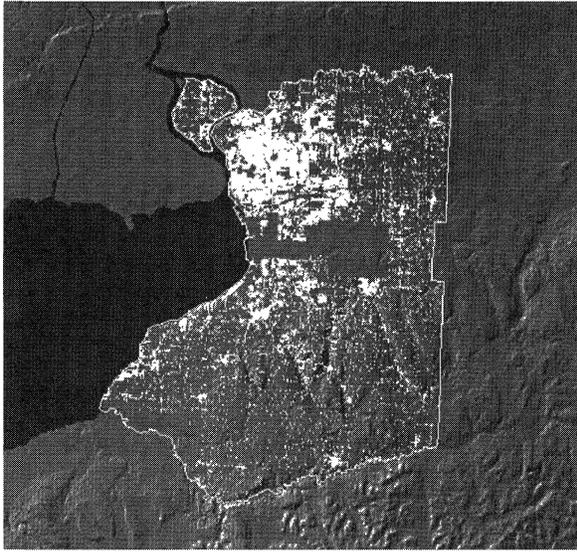


FIGURE 2 Topography and development in 1989.

of urban America (Reps, 1965). The region was sparsely settled and politically unstable in conflict with the indigenous and neighboring population in Canada. In fact, the British burnt the “village” of Buffalo in the war of 1812. At the end of the Napoleonic era, the village began to grow but the dominant characteristic of the region from 1820 to 1840 and beyond, was the intensification of agricultural settlement. By 1850, Buffalo had clearly become a small but significant regional focus and as a Great Lakes port, the town then grew rapidly over the next 70 years until the end of World War I. In the 1920s and 1930s, the city suffered economically as did most cities in north east America, but the economy was buoyant in the 1940s and 1950s with rapid suburbanization. However, deindustrialization began in the 1960s, and the core city fell into long term decline. This was exacerbated in the 1970s and 1980s with massive losses of population from the city and its inner suburbs, coupled with extensive suburbanization to the very edge of the county. These trends are clearly visible in the animation implied by Fig. 1.

From Figs. 1 and 2, we have constructed an abstract picture of this growth around the central core of the region – downtown Buffalo. Figure 3(a)

represents the sparsely populated frontier until about 1820, while Fig. 3(b) implies an intensification of agricultural settlement until 1840 but with Buffalo still remaining a small village. After 1840, the focus of regional settlement is clearly Buffalo (Fig. 3(c)) which continues to accelerate in growth into the early 20th century (Fig. 3(d)). During the middle years of this century, the decline of the central city begins, accompanied by rapid suburbanization (Fig. 3(e)), both trends intensifying dramatically over the last 20 years (Fig. 3(f)). In the subsequent analysis, it is important to note that until about 1840 or so, the region lacked a major focus. Thus the core shown in Fig. 3(a) and (b) is before the city existed and thus any spatial measurements made about this point are arbitrary. Up to about 1850, the mean distance to this point increases rapidly reaching a peak. As the city begins to grow, this compacts the settlement of the region and the mean distance then falls. It only begins to rise again with the onset of urban decentralization from the 1920s reaching 70 percent of its mid-19th century value by 1989. In effect, this mean distance is really a measure of compactness of the urban region, not a measure of the average distance traveled. It shows how the agricultural settlement first diffuses, how the city counters this, and then how the city itself is blown apart by urban decline and suburban growth.

We must treat this data set very carefully. It is a record of what (“still”) existed in 1989, not what existed at 1790, 1800 and so on. In a region with a severe winter climate, much of what has existed over the last 200 years has been demolished and/or rebuilt, and thus our space–time series is just one perspective on Buffalo’s historical development. In fact, Fig. 3 is culled not only from this data set but from what we know more generally about settlement in this region (see Goldman, 1983). To put this space–time series in a wider context, note that the actual Census population of Erie County at the years 1830, 1860, 1890, 1920, 1950, 1980, and 1989–1990 was 36, 142, 323, 634, 899, 1015, and 968 (in thousands) in contrast to the numbers of developed properties in the 1989 data set of 1, 4, 15, 70, 137,

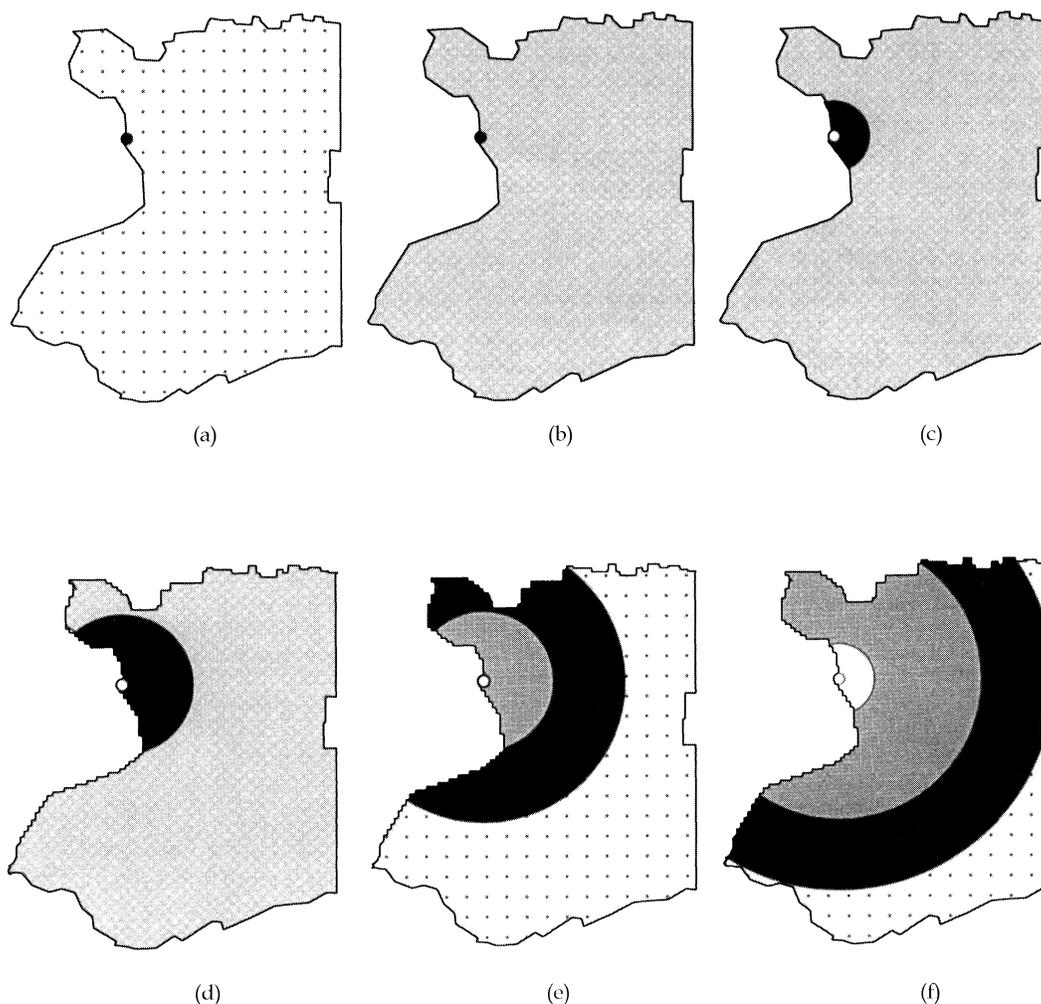


FIGURE 3 The abstracted pattern of urban growth. (a) Up to 1820: the sparsely settled frontier. (b) 1820–1850: intensifying agricultural settlement. (c) 1850–1880: the emergent industrial city. (d) 1880–1920: the maturing industrial city. (e) 1920–1960: the early suburban city. (f) 1960–1990: urban decline and suburban growth.

238, and 250 (also in thousands) at the same years. If we take the ratio of developed sites to Census population as 1 in 1989, then these ratios form the series 0.15, 0.11, 0.18, 0.43, 0.59, 0.91, and 1 which indicates that there is progressive loss of actual development in our 1989 series as we go back through time. This does not take account of changing population density and as densities were higher in the past, then the series probably overestimates the loss of past data.

The population of development sites is counted in rings of width $\sqrt{2}$ at increasing distances r from downtown Buffalo for every year from 1800. The number of rings rises very rapidly from 25 in 1800, to 1008 in 1835, 1183 in 1900, and 1208 in 1989 at which point the entire region is effectively covered. The population associated with each ring i at distance r_i at time t is counted as $n_i(t)$, and the cumulative population up to distance R_i and time T is computed as $N_i(T) = \sum_i \sum_t n_i(t)$ where the

summations are from $t = 1$ to T and $i = 1$ to R_i . Note that we also define the discrete temporal equivalents of the densities $\rho(r)$ and $P(r)$ as $\rho_i(t)$ and $P_i(t)$ respectively. Total population up to each time T , called either $N(T)$ or $N(t)$, is computed taking the i summation over the entire range. The spatial referent for each time period is the mean distance from all sites to downtown Buffalo defined as $\bar{R}(t) = \sum_i n_i(t)r_i/N(t)$, and in the sequel, we will define all growth paths with respect to the space variable $\bar{R}(t)$ and the time variable t . In short, $\bar{R}(t)$ and t define the basic axes of the phase space through which our urban growth paths will flow.

The simplest and perhaps most basic growth path involves the change in total population which we plot in terms of space $\bar{R}(t)$ and time t in Fig. 4(a). What this path shows is that the city establishes itself once the original effects of the agricultural frontier settlement have washed themselves out. This is clearer in Fig. 4(b) when we project the path onto the two-dimensional plane for each of two from the three variables. Comparing $N(t)$ against t reveals the classic growth in the system with no surprises but when $N(t)$ is compared against $\bar{R}(t)$, the arbitrary nature of this mean, prior to Buffalo existing, is clearly seen. As $N(t)$ increases, the mean increases very quickly, reaching a peak around 1860, then quickly falling as the city develops. The classic profile is only established from the 1920s when the population of the region begins to grow dramatically. This effect is seen even more clearly

when the mean $\bar{R}(t)$ is traced against time t . In Fig. 4(b), this is the projection on the horizontal plane which shows that the mean rises to a peak in the 1820s, remains at this level until the 1880s and then begins a steep decline in value as the region compacts around the growing city. From the 1930s on, however, the city begins to spread into its wider region and this mean then begins to rise. In a sense, it is only in the last 70 years, that we can interpret this variable as measuring the spreading of the city due to the automobile. This projection of $\bar{R}(t)$ and t is common to all the growth paths we chart in the next section as it is the basis of the phase space within which various portraits of growth are displayed. Finally, note that in Fig. 4 and the following graphs, the $\bar{R}(t)$ axis varies from 0 to 700 units of distance, t from 1800 to the year 2000, and the population (of sites) $N(t)$ from 0 to 300,000.

URBAN GROWTH PATHS BASED ON FRACTAL DIMENSION

Several ways of defining and estimating the fractal dimensions associated with the scaling relations in Eqs. (1)–(4) are presented in the Appendix. These methods fall into two groups: estimates based on *exact* and *statistical* methods. The dimension D measures two aspects of spatial growth: first, the conventional definition of dimension relates to the amount of space filled, and in two-dimensional systems, this should lie between 1 and 2 as we argued earlier. Second, the rate at which space is filled with respect to the distance from the origin is also picked up by D , and for systems which perfectly accord to the scaling relations, rate and degree of space-filling are consistent with one another. However, for real systems where there is spatial variation which has to be treated as random, the exact estimates remove variation from the data and thus produce a clear measure of space-filling while the statistical estimates pick up the rate of filling which is confused by random variations.

In the Appendix, we show that the simplest measure of dimension is based on an exact estimate

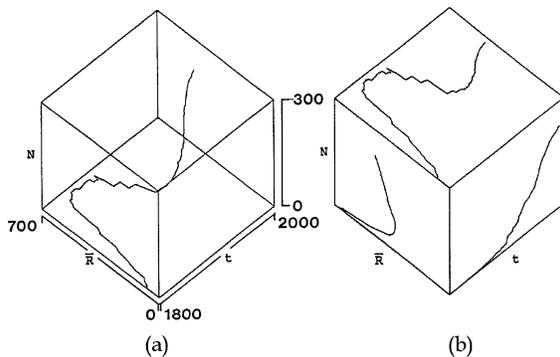


FIGURE 4 The population $N(t)$ growth path.

which we call the fractal signature (Eq. (A10)). This relates the average distance traveled (or the degree of compactness of the region) measured by \bar{R} to the cumulative density of population at the mean $P(\bar{R})$. This is repeated here with a time index as

$$D_{\bar{R}}(t) = 2 + \frac{\log P(\bar{R}(t))}{\log \bar{R}(t)}. \quad (5)$$

We will call this the baseline dimension which we will use as a comparator. We show the path of this variable for the space–time coordinates $\bar{R}(t)$ and t in Fig. 5 where the vertical axis – dimension $D_{\bar{R}}(t)$ – ranges from 0 to 2 (as in all subsequent graphs of dimensional paths). This trajectory is highly correlated with the population $N(t)$ in Fig. 4, because $D_{\bar{R}}(t)$ is a space-filling dimension. With a fixed regional space and a relatively fixed number of locational rings for most of the 200 year series, Eq. (5) measures the increasing density of the region.

What is of profound interest is that the fractal dimension $D_{\bar{R}}(t)$ clearly begins to stabilize once the automobile city takes off in the 1920s. The projection of $D_{\bar{R}}(t)$ on t in Fig. 5 shows how $D_{\bar{R}}(t)$ flattens; its value is 0.668 in 1800, rising to 1.661 by 1920 and 1.711 in 1930. However, over the next 60 years the dimension only increases to 1.752. This is a remarkable result given that there has been enormous change during this latter period. The data show $N(1930) = 96,606$ and $N(1989) = 250,455$, a net increase of 159 percent in contrast to a mere shift of 2.5 percent in the value of the dimension

over this period. This is suggestive evidence that the city reached its critical threshold around the 1920s since which time its space-filling has simply readjusted, hardly changing at all with respect to the mechanisms of urban growth and movement. In terms of the physics of fractals, the system underwent a second-order phase transition of the kind suggested by the theory of self-organized criticality. We now need to see if this stability is repeated for the other estimates of dimension.

We make a direct comparison between the population $N(t)$ and the baseline dimension $D_{\bar{R}}(t)$ (Figs. 4 and 5) in Fig. 6 which also compares these trajectories with the number of rings at each time $k(t)$; these range from 0 to 1300 on the vertical axis which is scaled for each range independently. The similarity of these profiles is obvious. The plots in the $N(T) - D_{\bar{R}}(t) - k(t)$ versus $\bar{R}(t)$ dimension show the same sorts of reversal as noted earlier due to the build-up and subsequent decentralization of the city. But in terms of the time dimension, it is clear that the fractal dimension stabilizes long after the number of rings defining the region becomes fixed while population continues to rise. This suggests that the stability of $D_{\bar{R}}(t)$ is not influenced by the trend in $N(T)$ or $k(t)$.

We are now in a position to chart the differential estimates of dimension based on the order we have introduced them in this paper as summarized in Table I. We will begin with the exact and then move to the statistical, drawing out similarities and differences in their growth paths. The major

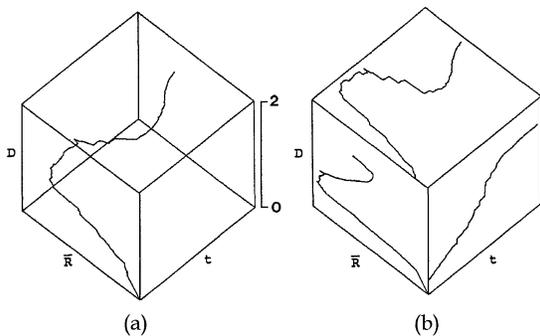


FIGURE 5 The baseline dimension $D_{\bar{R}}(t)$.

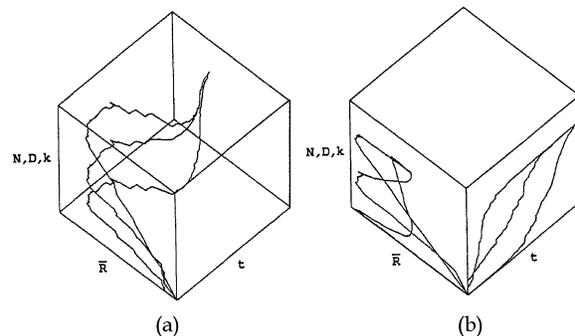


FIGURE 6 Similarities between the population, ring and baseline dimension growth paths.

TABLE I Estimates of dimension

Method of estimate	Class of estimator	Equation type	Related equation number
Exact	Discrete square grid	Cumulative count	(A5)
		Cumulative density	(A6)
		Incremental count	(A7)
		Incremental density	(A8)
	Discrete circular grid	Cumulative count	(A9)
		Cumulative density	(A10) the baseline
Statistical	Unconstrained regression	Incremental density	(A13)
		Incremental count	(A14)
		Cumulative count	(A15)
		Cumulative density	(A16)
	Constrained regression	Incremental density	(A13) ($a = 1$ and $2r_i$)
		Incremental count	(A14) ($b = 1$ and $2r_i$)
		Cumulative count	(A15) ($c = 1$ and $2R_i$)
		Cumulative density	(A16) ($d = 1$ and $2R_i$)

distinction we will find is between the cumulative and the incremental statistical estimates in terms of their magnitude and shape, and these are consistent with previous applications of these functions which we have researched (Batty and Xie, 1996; Mesev *et al.*, 1995).

We will first present the entire array of exact estimates which includes the baseline, and which best describes the space-filling characteristics of urban growth. In Fig. 7, we show the cumulative dimensions for the grid and the circle models separately from the incremental in both two- and three-dimensional form. It is immediately clear that small but significant differences exist between the grid and circle models, with the grid giving slightly higher dimension values in both cumulative and incremental forms. The incremental forms also give rise to much more jagged paths through the phase space due to the fact that incremental data are less smooth anyway, also revealing errors in the year of construction attribute in the dataset (Batty and Howes, 1996). Yet these do follow the cumulative paths quite closely. The correlations between all the paths – the three-dimensional projections in Fig. 7(a) and (c) and the two-dimensional projections in Fig 7(b) and (d) – are very high. Note however that the incremental values do not stabilize to quite so narrow a range as the cumulative during

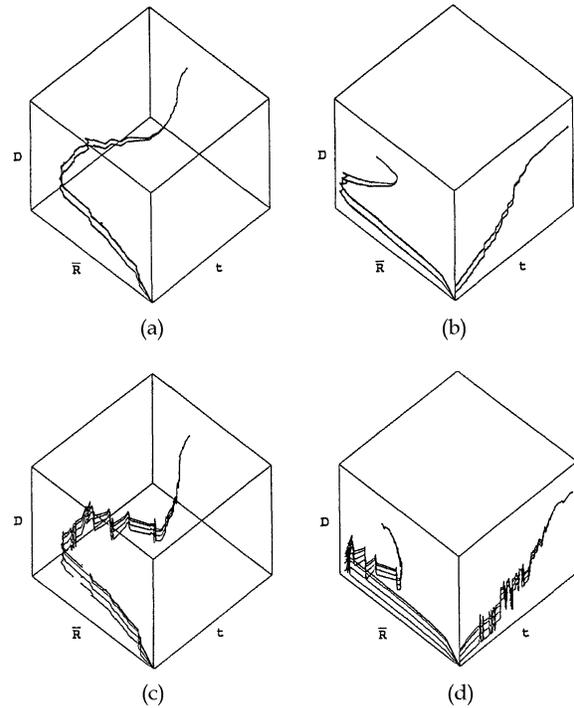


FIGURE 7 Dimensions of the grid and circle functions.

the last 60 years of the series; the evidence for self-organized criticality thus becomes a little less convincing.

We now consider the statistical estimates where we have separated the unconstrained regressions

from the constrained. The unconstrained results show large differences between the cumulative and incremental model forms while these differences virtually disappear once these models are constrained. For the unconstrained models, the dimension paths of the cumulative forms are close to those of the exact space-filling estimates in Fig. 7 but the incremental have quite a different form being closer to the variations in the mean value $\bar{R}(t)$. These results are shown in Fig. 8(a) and (b), and it is clear that the order of magnitude of the cumulative dimensions increases to the expected range while the incremental are much lower. When these regressions are constrained, the cumulative and incremental generate paths which are quite similar as shown in Fig. 8(c) and (d). In one sense, these are the most appropriate estimates for they mix a degree of spatial variation without distorting the role of dimension in capturing the properties of space-filling. The shape of their growth paths and their values are as expected.

It is worth commenting on the entire range of dimension values generated and in the case of the statistical estimates, the performance of the models. In Table II, we present these values for all 16 estimates for the last year 1989. All the cumulative and the constrained regressions generate values in the range $1 < D < 2$ with an average value of 1.68,

very close to the DLA value of 1.71. The incremental models estimated statistically give a lower performance ($r^2 \approx 0.2$) than the cumulative count ($r^2 \approx 0.8$) but these are clearly picking up random

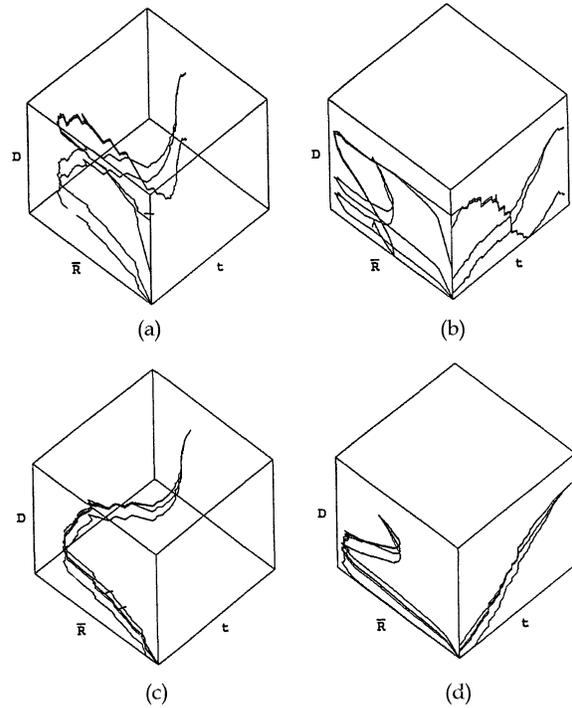


FIGURE 8 Dimensions based on regression.

TABLE II Dimensions and correlations for the end year 1989

Method	Equation type	Equation number	Dimension D	Correlation r^2
Discrete square grid	Cumulative count	(A5)	1.742	na
	Cumulative density	(A6)	1.777	na
	Incremental count	(A7)	1.753	na
	Incremental density	(A8)	1.752	na
Discrete circular grid	Cumulative count	(A9)	1.868	na
	Cumulative density	(A10)	1.751	na
	Incremental count	(A11)	1.779	na
	Incremental density	(A12)	1.723	na
Unconstrained regression	Incremental density	(A13)	0.441	0.122
	Incremental count	(A14)	0.382	0.519
	Cumulative count	(A15)	1.601	0.858
	Cumulative density	(A16)	1.589	0.287
Constrained regression	Incremental density	(A13)	1.589	0.121
	Incremental count	(A14)	1.385	0.198
	Cumulative count	(A15)	1.631	0.882
	Cumulative density	(A16)	1.664	0.236

na – Not applicable.

variation in the observations which, in essence, reflect the lack of fit of the scaling relations to data. We have commented on the difficulties of working with incremental scaling relations elsewhere (Batty and Xie, 1996; Mesev *et al.*, 1995).

To summarize, the various growth paths in Figs. 4–8 confirm the existence of a series of distinct transitions: first between 1820 and 1830 when the region was transformed from the frontier, then again around 1870–1880 when the city began to dominate the region, continuing until the 1920s when the most recent transition to the decentralized city began. However, the most important transition is the passage to a stable fractal dimension paths during the last 70 years. This increases the likelihood of a further transition based, perhaps, on the phenomena of the ‘edge city’ which is occurring in many other North American cities (Garreau, 1991). There are already signs in the last 5 years that such a phenomenon is occurring in the Buffalo metropolitan area but whether this will change the value of the space-filling dimension which has persisted for the last 70 years is unclear.

THE NEXT TRANSITION

The transitions we have noted are clear enough both from our casual knowledge of the history of cities and from the data. An unwelcome comment on our work might suggest that what we have shown is self-evident and therefore hardly surprising. In one sense, of course it is but in another, we are able to point to the long-standing stability of the automobile city. This suggests that it is this particular urban form which is the one that represents the logical outcome of industrialism, rather than the city forms of the late 19th and early 20th centuries which are often taken as exemplars. The automobile city is thus the form that is consistent with the age of energy. In the transition to the information age, it is difficult to know whether the space-filling characteristic of cities like Buffalo will continue, whether cities will dramatically spread out, or grow into much more dense

structures making greater use of vertical space. All these scenarios appear in the fiction of the near future.

An even more arresting view suggests that the transition that is really important is not from the industrial to the postindustrial but from the non-urban to the urban. What we are seeing is a change from sparsely populated agricultural and central place settlement systems to more densely populated regions, entirely urbanized but in the most decentralized ways possible and lacking any single focus. The fact that the fractal dimension of Buffalo has been stable for 70 years might simply be evidence that this is the way all societies and their cities will be, forever. Our speculations are made somewhat more tangible by setting them within the formal framework of self-organized criticality, but there are many technical improvements that might be made to increase the robustness of the analysis. Data on the chain effects of movement in time and space would be extremely valuable in providing additional hypotheses to test for criticality, but there are also ways of improving the analysis through examining missing data, scale and size effects relating to the resolution and size of the region, as well as error and bias in the data. We have begun to tackle these for this data elsewhere (Batty and Howes, 1996; and <http://www.geog.ucl.ac.uk/casa/ur.html>) but we need to explore how our temporal data might be improved and compared against real time series from past to present. Work towards these more modest goals will also help support the more dramatic speculations which have served to guide the project so far.

References

- Alstrom, P. (1990) Self-organized criticality and fractal growth, *Physical Review A*, **41**, 7049–7052.
- Alstrom, P. and Leao, J. (1994) Self-organized criticality in the “Game of Life”, *Physical Review E*, **49**, R2507–R2508.
- Alstrom, P., Trunfio, P.A. and Stanley, H.E. (1990) Spatiotemporal fluctuations in growth phenomena: dynamical phases and $1/f$ noise, *Physical Review A*, **41**, 3403–3406.
- Barabasi, A.L. and Stanley, H.E. (1995) *Fractal Concepts in Surface Growth*, Cambridge University Press, Cambridge, UK.
- Bak, P. and Chen, K. (1991) Self-organized criticality, *Scientific American*, **264**, 46–53.

Bak, P., Chen, K. and Creutz, M. (1989) Self-organized criticality in the "Game of Life", *Nature*, **342**, 780–782.

Bak, P., Chen, K. and Wiesenfeld, K. (1988) Self-organized criticality, *Physical Review A*, **38**, 364–374.

Bak, P. and Paczuski, M. (1993) Why nature is complex, *Physics World*, **6**(12), 39–43.

Bak, P. and Sneppen, K. (1993) Punctuated equilibrium and criticality in a simple model of evolution, *Physical Review Letters*, **71**, 4083–4086.

Batty, M. (1991) Cities as fractals: simulating growth and form, in T. Crilly, R.A. Earnshaw and H. Jones (Eds.) *Fractals and Chaos*, Springer-Verlag, New York, pp. 41–69.

Batty, M. and Howes, D. (1996) Exploring urban development dynamics through visualization and animation, in D. Parker (Ed.) *Innovations in GIS 3*, Taylor and Francis, London, pp. 143–155.

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

Batty, M. and Xie, Y. (1996) Preliminary evidence for a theory of the fractal city, *Environment and Planning A*, **28**, 1745–1762.

Benguigui, L. (1992) Some speculations on fractals and railway networks, *Physica A*, **191**, 75–78.

Frankhauser, P. (1994) *La Fractalité des Structures Urbaines*, Collection Villes, Anthropos, Paris, France.

Garreau, J. (1991) *Edge City: Life on the New Frontier*, Doubleday, New York.

Goldman, M. (1983) *High Hopes: The Rise and Decline of Buffalo*, New York, State University of New York Press, Albany, NY.

Kennedy, P. (1993) *Preparing for the Twentyfirst Century*, Vintage Books, New York.

Krugman, P. (1994) Complex Landscapes in Economic Geography, *American Economic Association, Papers and Proceedings*, **84**, 412–416.

Messey, T.V., Longley, P.A., Batty, M. and Xie, Y. (1995) Morphology from imagery: detecting and measuring the density of urban land use, *Environment and Planning A*, **27**, 759–780.

Reps, J.W. (1965) *The Making of Urban America: A History of City Planning in the United States*, Princeton University Press, Princeton, NJ.

Rinaldo, A., Dietrich, W.E., Rigon, R., Vogel, G.K. and Rodriguez-Iturbe, I. (1993) Geomorphological signatures of varying climate, *Nature*, **374**, 632–635.

Schelling, T. (1978) *Micromotives and Macrobehavior*, W.W. Norton and Co., New York.

Stauffer, D. and Aharony, A. (1992) *Introduction to Percolation Theory*, Taylor and Francis, London.

Turcotte, D.L. (1992) Fractals, chaos, self-organized criticality, and tectonics, *Terra Nova*, **4**, 4–12.

White, H. (1970) *Chains of Opportunity*, Harvard University Press, Cambridge, MA.

APPENDIX

Defining and Estimating Fractal Dimensions

Two methods are used to estimate the fractal dimension D associated with each of the four scaling relations in Eqs. (1)–(4) in the main paper.

These are called *exact estimates* and *statistical estimates*. Exact estimates are based on computing a single statistic from the relevant distribution, and then manipulating this to give an exact value of the fractal dimension. Statistical estimates are based on using all observations in the distribution and computing a value of the dimension which 'best fits' the distribution, for example by least squares in the case of estimation using linear regression.

Exact estimation is based on assuming a simple discrete form for $N(R)$. Density is measured and population counted on a square grid, in incremental or cumulative square bands up to distance R with the first band set as $r = 1$. The grid is of sufficiently fine resolution to detect no more than a single unit of development in each square. The cumulative count (of population) is given as

$$N(R) = (2R)^D, \quad (\text{A1})$$

with the total possible area occupied as $A(R) = (2R)^2$. The cumulative density $P(R)$, the incremental count $n(r)$, and the incremental density $\rho(r)$ are given respectively as

$$P(R) = \frac{N(R)}{A(R)} = (2R)^{D-2}, \quad (\text{A2})$$

$$n(r) = \frac{dN(R)}{dr} = D(2r)^{D-1}, \quad (\text{A3})$$

and

$$\rho(r) = \frac{dN(R)}{dA(R)} = \frac{D}{2} (2r)^{D-2}. \quad (\text{A4})$$

Equations (A1)–(A4) absorb the constant of proportionality into the measure of distance, and for any value of r , an exact estimate of D can be calculated. For Eqs. (A1) and (A2) using the mean \bar{R} (which is known from data) gives

$$D(\bar{R}) = \frac{\log N(\bar{R})}{\log(2\bar{R})}, \quad (\text{A5})$$

and

$$D(\bar{R}) = 2 + \frac{\log P(\bar{R})}{\log(2\bar{R})}. \quad (\text{A6})$$

Calculating D from Eqs. (A3) and (A4) is less simple as these must be solved by iteration. If we assume the DL value of 1.7 for $D(\bar{R})$, then Eqs. (A3) and (A4) yield

$$D(\bar{R}) = 1 + \frac{\log n(\bar{R}) - \log(1.7)}{\log(2\bar{R})}, \quad (\text{A7})$$

and

$$D(\bar{R}) = 2 + \frac{\log \rho(\bar{R}) - \log(\frac{1.7}{2})}{\log(2\bar{R})}. \quad (\text{A8})$$

It is possible to compute other exact estimates if different assumptions governing the underlying space are used. For example for a circular system with $A(R) = \pi R^2$, Eqs. (A1)–(A4) become $N(R) = \pi R^D$, $P(R) = R^{D-2}$, $n(r) = D\pi r^{D-1}$, and $\rho(r) = (D/2) r^{D-2}$, and the equivalents to (A5)–(A8) become

$$D(\bar{R}) = \frac{\log N(\bar{R}) - \log \pi}{\log(\bar{R})}, \quad (\text{A9})$$

$$D(\bar{R}) = 2 + \frac{\log P(\bar{R})}{\log(\bar{R})}, \quad (\text{A10})$$

$$D(\bar{R}) = 1 + \frac{\log n(\bar{R}) - \log(1.7\pi)}{\log(\bar{R})}, \quad (\text{A11})$$

and

$$D(\bar{R}) = 2 + \frac{\log \rho(\bar{R}) - \log(\frac{1.7}{2})}{\log(\bar{R})}. \quad (\text{A12})$$

Equation (A10) is the so-called ‘fractal signature’ relation used extensively in Batty and Longley (1994); we argue that this relation represents the ‘purest’ of our exact estimates which we use in the main paper as our baseline dimension.

For statistical estimation, we will define the incremental and cumulative count and density

relations in (1)–(4) as ρ_i , n_i , N_i and P_i respectively, where i is an index associated with the distance r_i or R_i from the origin (CBD). The basic data are n_i which are counted in incremental rings of area Δx_i associated with r_i . The density is thus defined as $\rho_i = n_i/\Delta x_i$, the cumulative count as $N_i = \sum_i n_i$ and the cumulative density as $P_i = N_i/\sum_i \Delta x_i$, where the summations are from $i=1$ to $i=R_i$. We can linearize the discrete relations based on applying these definitions to Eqs. (1)–(4) by taking logarithms and this gives

$$\log \rho_i = \log a - \alpha \log r_i, \quad (\text{A13})$$

$$\log n_i = \log b - (1 - \alpha) \log r_i, \quad (\text{A14})$$

$$\log N_i = \log c - (2 - \alpha) \log R_i, \quad (\text{A15})$$

and

$$\log P_i = \log d - \alpha \log R_i. \quad (\text{A16})$$

a , b , c , and d are intercepts of the original scaling functions with α , hence D , related to the slopes of the appropriate regression lines. In contrast to the exact estimates based on the grid and circle models, these values will be heavily affected by the shape of the underlying density distribution and any departures from the hypothesized power laws will be reflected in the values of a , b , c and d as well as in α .

In the main text, we argue that statistical estimates are more likely to reflect the rate of space-filling or the attenuation of the density function than the exact estimates which are more likely to reflect the amount of space filled. It is possible however to develop a statistical estimation which reflects the degree of space-filling by constraining the regression to predetermined constant values for a , b , c , and d . Using the grid model which absorbs these constants in the distance variable, we can set their values in Eqs. (A13)–(A16) to unity and use $2r_i$ and $2R_i$ as the independent variables. From this type of constrained regression, we can show the extent to which the values of the constants distort the interpretation of α and D as measures of space-filling. Table I in the main paper summarizes all the estimates used.