Spatial Multiscale entropy

Martin Barner[1,*] **and Elsa Arcaute**[2]

[1]funemployed
[2]Centre for Advanced Spatial Analysis, London, postcode, country
[*]barner.mar@gmail.com

ABSTRACT

We propose a multiscale measure of entropy for spatial configurations in the context of cities. Existing approaches usually use entropy as a proxy for spatial evenness or disorder, and translate the phase space from statistical mechanics literally to the geographical space. In contrast, we focus on building an interpretation of entropy that is conceptually consistent with statistical mechanics as well as the view of cities as complex systems. For this, we focus on the characteristics of places and, most importantly, take into account interactions between them.

By comparing synthetic patterns, we show that if elements of a system interact spatially, complex spatial patterns display higher entropy. We can thus partly explain morphological complexity in cities as simply the most probable configuration. Further, we conduct a case study of entropy in the spatial distribution of buildings with different functions in West London over a period of 130 years. We find that the polycentric sprawl increases entropy compared to spatially random or segregated patterns.

This work offers a new approach to explaining how morphological complexity in cities emerges from individual behaviour. Furthermore, it may be helpful in dealing with uncertainty in planning.

Introduction

Existing measures of entropy in an urban context are either non-spatial, or use the mathematical properties of entropy as a proxy for spatial evenness. In this paper, we attempt to formulate a measure of entropy that is instead conceptually consistent with entropy in statistical mechanics and an understanding of cities as complex systems:
Viewing cities as complex systems brought a fundamental shift in our understanding of how cities function, grow, and change over time[1–4]. Essential to this view is that a city "self-organizes out of millions of individual decisions, a global order built out of local interactions" [5, p.38]. What people do in the city has an impact on its spatial structure over long periods of time[6,7]. It follows that in terms of state probabilites, a measure of morphological entropy should relate to how the city is used. Instead of measuring spatial evenness - the uncertainty about where things are - we attempt to measure the uncertainty about urban life that is built into the spatial structure of the city.

Two fundamental aspects differentiate our approach from existing spatial measures of entropy: First, Boltzman's entropy assumes that interactions between particles are neglectable. Existing measures of entropy inherit this for cities, and assume that there were no interactions between places. We recognise that different places are highly dependent on each other due to flows and interactions between the people in them. This interdependence between places is widely recognised in geography in general[8], urban theory[1,9] and quantitatively demonstrated for example in the success of spatial interaction models[10–13], and further in the study of agglomeration economics[14] and neighbourhood effects[15]. Our new measure takes relationships between places into account with a multiscale approach.
Second, existing measures mostly use the geographical space directly as the phase space. The phase space in statistical mechanics is the space of possible states of observed elements[16]. In thermodynamics location makes sense as part of the phase space, because particles actually do float around in space randomly. Buildings do not float around in space randomly. We are not interested in the randomness of geographic coordinates, but in the randomness of how people could use the city depending on its physical structure. Our phase space dimensions describe the characteristics of different places, based on the assumption that what people can do in two places differs if the places have different characteristics.

We show that the randomness in what people can do in a city is not necessarily maximised if urban structures are completely randomised spatially. Instead, spatially complex patterns give the highest randomness in the variety of available types of interdependent places. In our case study we measure the change in entropy in land use patterns in west London in seven time steps from 1875 to 2005, and the results suggest that the observed polycentric sprawl can be explained as the growth pattern with the highest entropy.

A multiscale approach to entropy in cities

We differentiate the existing measures of spatial entropy based on their phase space definitions. After identifying their limitations, we give a combinatorical explanation of how the number of microstates changes if we take into account interdependencies between places at different scales. Finally we introduce the phase space of our multiscale entropy.

Common phase space definitions for spatial entropy

Entropy is defined as[17]:

$$H = \int f(x)\log(f(x)) \tag{1}$$

Where $f(x)$ is the probability density of a continuous phase space. The equivalent to equation 1 in the discrete phase space is[17]:

$$H = -\sum p \log p \tag{2}$$

which reduces to the Boltzman entropy S if all probabilites p are the same:

$$S = k_B \log(\Omega) \tag{3}$$

"where Ω is the total number of microstates available to the system" [18, p.44]. More intuitively, it relates directly to the "number of ways [...] by which a given macroscopic state can be realized"[19]. All microstates can be allocated a location in the phase space. If the phase space is discrete, we can count the positive integer number of accessible microstates.

The highest entropy is always given by a uniform distribution in the phase space, leading to sometimes misleading but common metaphors for entropy[20]: If the entropy of a system is high, the state of a randomly selected element is unpredictable and "uncertain", and if we are uncertain about where things are in a system, one might describe it as "disordered". In thermodynamics, the word entropy refers to an agreed definition of the phase space unless stated otherwise. For cities on the other hand, fundamental conceptual differences between interpretations of entropy come down to different definitions of the phase space dimensions.

- The first essential approach in the literature takes the word "space" literally and defines the phase space as the geographical space.[21–23] The highest entropy is then given by a pattern with a uniform distribution in geographical space as in pattern a) of figure 2. This interpretation answers the question: How uncertain is the absolute location of a place with a given characteristic?
- The second basic phase space uses a characteristic of places or objects in space as the phase space.[24] All patterns in figure 2 have the same global proportions of black and white pixels and therefore, according to this phase space definition, the same entropy. This interpretation answers the question: How uncertain is the characteristic of a given place or observed element in space in general, independent from the spatial configuration?
- Most reviewed approaches to spatial entropy use a combination of the two phase spaces above. They are measures of spatial evenness widely discussed in the literature on measures of segregation[25–36]. They have the highest entropy if entropy is maximised in both phase spaces above at the same time. It answers the question: How evenly are observations of different types or characteristics distributed geographically?

Apart from these, entropy was introduced for spatial interaction models by Wilson[37], and there are attempts to discuss the energy and recources entering and exiting an urban systems in releation to entropy.[38] These approaches to entropy in cities are very remote from our approach and not discussed here.

None of the reviewed approaches simultaneously satisfy both main requirements that we identified above for a conceptually consistent interpretation of entropy that reflects the idea of cities as emergent phaenomena:

- It should observe how places are distributed across characteristics, to reflect the certainty about what people do in them.
- It should reflect that the characteristics of places spatially depend on each other.

In contrast to the existing measures of entropy, we want to answer the question: How uncertain is what a randomly selected resident does, based on the structure of the city? For places in the city this would mean: How uncertain are the characteristics of places a person could be in, considering that the characteristics of a place are defined not only by its own value, but also by the characteristics of the places around it?

Spatial dependence

We demonstrate with a combinatorical example how spatial relationships can be taken into account by observering configurations at multiple scales.

As a thought experiment, we construct a pattern by arranging individual pixels into groups of two, then arranging these groups into groups of four and so on. We observe the randomness involved in each step of the process.
Individual pixels can be white or black. The color of a random pixel can be very predictable (low entropy), for example if all pixels are black. Or it can be distributed in a less predictable, more random way, for example if half of the pixels are black and the other half white. So far this corresponds to the non spatial Shannon entropy, or group 2 in the common phase space definitions above.
The next step takes into account that each pixel's characteristic should also depend on what kind of pixel is next to it. We form pairs of two pixels that will be arranged next to each other in the final pattern. this can be done in a very predictable way: Entropy is low if all 2-pixel combinations are exactly the same, for example if every single white pixel is next to one black pixel and vice versa. They could instead be combined in a less predictable way: for example if one third of the combinations has two black pixels, one third has one white and one black pixel and one third has two black pixels. There is clearly less certainty and higher entropy in the spatial combinations of the second case.
These groups of two can then be combined into groups of four, again in more or less predictable ways and so on, until the pattern is complete. If we make the least predictable, or, the maximum entropy choice in every step of building the spatial configuration, we must end up with a nontrivial pattern like the linear pattern in figure 1.
Again: The phase space here is *not* the geographical space. We associate high entropy *not* with randomness in absolut positions in a pattern, but instead with randomness in the *spatial configurations*.

The Multiscale Entropy Phase Space

Imagine the patterns in figure 2 were real cities, and black and white pixels would refer to residential and commercial buildings. Taking into account the surroundings of each pixel, pattern a) only has two different types of places: residential or commercial buildings, but always in mixed blocks in mixed neighbourhoods in mixed districts of a homogeneously mixed city. Pattern g) has a much larger variety of spatial configurations.
We want to extend the description used to compare individual buildings from "a residential building" to something like "a residential building in a mixed use block in a mainly commercial district that is surrounded by residential areas".
Therefore we define for our quantitative measure the phase space like this: The first dimension of the phase space is the value of a place's own characteristic. We then add further phase space dimensions describing the place's surroundings at different scales. When we consider only one characteristic, the state of each place x_i in the city is given by the vector

$$\vec{x}_i = (x_i^{d_0}, x_i^{d_1}, \ldots, x_i^{d_n}) \tag{4}$$

Where $x_i^{d_0}$ is the value of the place itself, and $x_i^{d_n}$ is given by the values of all places within distance d_n from x_i, aggregated by a function:

$$x_i^{d_n} = f(x_{k_1}^{d_0}, x_{k_2}^{d_0}, \ldots, x_{k_m}^{d_0}) \tag{5}$$

for all $x_k^{d_0}$ within distance d_n from x_i. What this achieves is that we can distinguish between locally identical places based on what kind of area they are in, because the state of a place is literally a function of its surroundings.

Extending this to c scalar characteristics, the whole state of a place in the system is given by the matrix

$$\Psi_i = \begin{pmatrix} x_i^{d_0,1} & x_i^{d_1,1} & \dots & x_i^{d_n,1} \\ x_i^{d_0,2} & x_i^{d_1,2} & \dots & x_i^{d_n,2} \\ \vdots & \vdots & \vdots & \vdots \\ x_i^{d_0,c} & x_i^{d_1,c} & \dots & x_i^{d_n,c} \end{pmatrix} \tag{6}$$

Imagine a number of identical offices. One of them is in a central business district, one of them on the countryside, one on an oil platform and another in the dessert. The way they are or can be used is fundamentally different because the spatial context is. The matrix in 6 takes that context into account. In reverse, the way the ocean / countryside / business district / dessert around them are used is altered as well, and the presence of the office building appears in their state matrices.

Theoretically, the continuous phase space $\phi^{(n+1)*k}$ is then the space of all possible place state matrices $x \in \phi$. The states of all places in the system are distributed in ϕ with the probability density distribution g. The entropy is then

$$\int_{a\in\phi} -g(a)\log g(a) \tag{7}$$

for all a where $p(a) > 0$.

There are strong conceptual parallels to the difference between Boltzman's and Gibbs' phase space in statistical mechanics. For c characteristics observed at n scales, the phase space has $c*n$ dimensions, signifcantly more than the c dimensional space that could be used if we ignored interactions between places. A similar increase in the number of dimensions happens when moving from the Boltzman phase space to the Gibbs phase space that deals with interacting particles in statistical mechanics. Nonetheless, the relationship between the Boltzmann entropy and the Gibbs entropy is far more complex[39], and their precise interpretations still debated[40].

We found an approach to a similar problem by Zhang[41] and Costa et al.[42–45] to measure complexity in medical time series. This "has become a prevailing method to quantify the complexity of signals. It has been used successfully"[46] in numerous fields of research, but to our knowledge not in a spatial context. There are significant differences to our method: Scales are not viewed as different phase space dimensions. Instead, a single value is produced from integrating over the scale dependent entropies. By adding the entropies, different scales are viewed as independent systems. In practice this would capture that for example pattern e) in 2 has a wide variety of values across all neighbourhood sizes, but ignore that small low value neighbourhoods lie very predictably in large low value neighbourhoods and vice versa.

Apart from that, there are methodological and conceptual parallels of our approach to methods for estimating fractal dimensions[47], specifically box counting[48] which could be worth exploring further.

Multiscale Entropy Estimation

The practical entropy estimation in our example patterns and case study is as simple as possible without compromising the general concept. We use simplified square neighbourhoods with varying side length because it makes the results easy to trace, is computationally convenient and is sufficient to demonstrate the concept.

We then calculate the place state matrix 6 using the mean as the aggregation function in equation 5 for the same reasons.

The most important simplification is that we not only use a finite number of discrete scales, but also set a finite number of discrete values for all elements in the place state matrix 6 by binning its values equidistantly after the aggregation. Places are assumed to have the same state if and only if their state matrices are exactly identical. Because this simplified phase space is discrete, we can estimate the probability of discrete states directly from their frequency and estimate the system's entropy directly with equation 2.

Discretising the phase space has multiple advantages. First, we avoid properties of the unit dependent[49] continuous entropy such as negative entropy[50,51] that are difficult to interpret in terms of statistical mechanics. Furthermore, it removes the difficulty of evaluating eucledian distances between values of different place characteristics for equation 7. Finally, it allows us to avoid discussing complicated estimators for multivariate continuous data[52,53]. They are unreliable for high dimensional data because they work with the spaces between observations, and the number of data points on the edges of the phase space increases exponentially with increasing dimensions.

Example patterns

Here we use synthetic spatial patterns to first show how the results of our new method are inherently different from existing phase space definitions. We then compare the multiscale entropies of patterns with different structures and show that if interactions between places are accounted for, complex patterns have a higher entropy than simple ones. Randomised simulated patterns corresponding to the patterns in figure 2 have the multiscale entropies shown in figure 3. Each cell corresponds to a "place", and each cell's colour defines the only characteristic of that place. The patterns are 512 pixels wide and high. For "Neighbourhoods" that go over the edge of the pattern, the invisible part is assumed to have the same proportion of values as the visible part. We bin the mean values in three categories: mainly low values (mean 0-0.33), mixed (mean 0.33-0.66) and mainly high values (mean 0.66-1.0). We use 5 different scales with neighbourhood side lengths with 1, 3, 9, 27 and 81 pixels.

comparison to existing measures

In the non spatial phase space observing only global characteristic proportions, we can directly tell that all patterns would display the same entropy as long as they differ only in the spatial configuration. Because in all patterns approximately half the pixles are black, their entropy according to equation 2 is $H_{nonspatial} = \log(2)$. We could reduce the multiscale entropy phase space to this by using only the $x_i^{d_0,c}$ column on the left of matrix 6.

Measures of entropy using the geographical space directly as the phase space are essentially measures of how evenly elements are distributed across different zones. We split the patterns into square zones with a side length of 32 pixels, and count the number of black pixels as in figure 4.
This approach is inherently different from our measure in its goals and results. As expected from a measure of spatial evenness, the geographical phase space entropy (figure 5) is highest for the uniform distribution (figure 2 a)), and lowest for patterns segregated spatially at a larger scale than the used zones (figure 2 c) and d)).
The frequencies in the discrete phase space in figure 6 show the conceptual difference to our measure. When the geographical space is used directly as the phase space, the spatially even distribution of pattern a) also gives an even distribution in the phase space. In contrast, we see an *even distribution of frequencies* for the sorted patterns e) and f) and for the additive cascade g), which are favoured by a measure that is focused on how much places differ from each other.

entropy and complexity

To show how spatially complex patterns have the highest entropy if interactions between locations are taken into account, we take a closer look at the distributions in the multiscale phase space for the synthetic patterns. Figure 7 shows two dimensions of the multiscale phase space, specifically at the scales of 9 and 81 pixels neighbourhood side length.
The relatively complex additive cascade is most evenly distributed in the phase space. The uniform probability in the geographical space is distributed relatively evenly on the very local scale, because locally, we are likely to find all possible combinations of pixel colors. However, all pixels lie in very similar mixed neighbourhoods, and so the distribution has little variation on the y axis. The patterns b) to d), segregated on different scales, have increased variance on scales of observation close to their scale of segregation, but are fail to maintain variance across multiple scales. The sorted uniform distribution - the only pattern not binary from the beginning to demonstrate this point - are very evenly distributed on all scales. However, there is no variation in which type of small scale neighbourhood is combined with which type of larger scale neighbourhood. This effect also applies to the 1/f noise pattern: While there is some variation everywhere on the very local scale, small white pixel neighbourhoods are systematically more likely to lie and larger white pixel neighbourhoods. The corresponding entropies are shown in figure 3.

Imagine we would try to change any of these patterns to spread the observations more evenly in the phase space and increase the entropy. We would need to add more and more layers of variation on different scales, while simultaneously trying to avoid creating simple random noise, and the result would be a spatially complex configuration similar to the additive cascade.
This may seem rather abstract. However, it should apply to any system in which elements interact with and influence each other over multiple scales of some type of "nearness", to a degree at which they fundamentally change each others meaning. As discussed in the introduction this is certainly the case for places in cities. Under these

circumstances, complex patterns have a higher entropy. Therefore, we can and should expect the whole system to eventually arrange in a complex pattern, simply because that is the most probable configuration.

Case Study: London 1875 - 2005

data

In the case study, we analyse the spatial patterns of land use in west London from 1875 - 2005. The dataset used in the analysis was originally built and provided by Stanilov et al.[54]. It covers 200 square kilometers, spanning 20km from east to west, from London's green belt in the west to the west end hyde park, and roughly 10km from north to south. The data provides the land use of individual building in 32 categories for seven moments in time; 1875, 1895, 1915, 1935, 1960, 1985 and 2005. For further details on the data collection see Stanilov et al.[54]

entropy estimation

To keep the number of dimensions reasonably low, the 32 land uses are grouped into three categories of "business", "residential" and "leisure" and we use 5 scales of observation at 50m, 150m, 450m, 1350m, 4050m. We discretise the values in the place state matrix 6 equidistantly in three bins. The data is rasterized at a resolution of 50m. Neighbourhood parts outside the bounding box are assumed to have the same proportion as the parts within.

results

Figure 8 shows the development of entropy over time in comparison to three null models and non spatial entropy. For all cases, entropy increases until 1935, stagnates around 1965 and then slightly decreases until 2005. This is based on the non spatial entropy of the global distribution of functions, as almost the entire area is undeveloped in the beginning and almost filled entirely in the end. We compare the observed patterns with three null models that are constructed to preserve the global amount of different land uses and differ only in the spatial configuration. The configurations for comparison are shown in figure 9

- spatially random uniform spread: The pixels of the original data are reallocated in a random order. This would be the maximum entropy distribution if the phase space was directly taken from the geographical space.
- compact mixed-use growth: The pixels of the original data are redistributed in a fully random fashion, but separated between developed and undeveloped land and fit compactly to the east edge, corresponding to the general direction of growth in the original data.
- compact segregated growth: The pixels of the original are sorted by function and fit compactly to the east edge.

The observed multiscale entropy of West London is significantly higher than all three null models. Especially between 1915 and 1960, entropy increases in the observed data, while the null models stagnate. The grayscale images in figure 9 show the probability of each pixel's state to investigate which places contribute to the total entropy.

In the spatially uniform randomised case, unique places appear only beyond a certain global density, where only very small segregated clusters appear by chance. In the early stages entropy would be higher if growth was more concentrated, and later if there were also larger segregated and non segregated local concentrations.
In the compact mixed use growth case, the only unique places are on the city edge, while most places are either completely undeveloped or evenly mixed. Entropy could be increased by a less stringent city edge and partial concentration of the less frequent commercial functions. In the compactly segregated case, the most unique places are along the edges between functions, as well as along the city edge. Entropy could be increased by a less stringent city edge, as well as more smaller clusters of segregated or mixed functions.
All of these alterations would change the null model patterns closer to what we actually observe:
First, clusters of different sizes with varying degree of functional segregation. Second, no strict city edge. In the language of urbanists, we could call this *polycentricity*[55] and *sprawl*[56]. From this perspective, we can give an explanation of the polycentric sprawl that dominates the growth patterns of the observed area in terms of entropy: Unless significant restrictions are in place, there are simply overwhelmingly more combinations of individual choices that lead to polycentric sprawl, making it the most likely pattern to occur.

There are great limitiations in terms of data and methodology that make any conclusions or generalisations speculative. First of all, we are only observing a small window of the city, and as the city grows the city edge passes through our

field of view. Furthermore, the results may be biased towards higher entropy because in the original data collection, the area was selected specifically for it's high functional diversity.[54]
In terms of methodology the functional categories, the aggregation function, the scale of rasterisation, the selection of neighbourhood scales and their rectangular shape are all rather arbitrary. While sufficient to demonstrate the basic ideas, neighbourhood sizes and shapes as well as the aggregation function could use a network based measure of distance, take into account subjective travel cost and relate to insights into the actual connectivity between places.

discussion

The ambition of this work is to make a contribution to explaining how individual actions shape cities and establish a more coherent relationship between entropy and complexity. Further, the general framework of thinking may be used as a strategy to deal with uncertainty and unpredictability in planning practice.
Batty recognises not only a "literature gap", but "an entirely new research agenda" and states that "substantive interpretations of entropy measures [. . .] have not been well developed. [. . .] I sketch the need for new and different entropy measures that enable us to see how equilibrium spatial distributions can be generated as the outcomes of dynamic processes"[22]. Entropy in thermodynamics is a concept relating the fast, microscopic behaviour to the slow, macroscopic dynamics of a system. We therefore see it as a suitable tool to be used in studying the relationship between the fast dynamics of individual behaviour and the slow, larger scale dynamics of change in urban structures. The case study - that is arguably too small in scale and too simplified in its methodology to be generalisable in any way - suggest that West London did in fact display a higher entropy than the more extreme toy scenarios.

The understanding of the relationship between entropy and complexity is highly incoherent in the literature.[57] Attempts have been made to associate complexity with decreasing thermodynamic entropy[58,59], regarding the occuring order as higher complexity than the original randomness. Others regard fully unpredictable signals such as white noise as "fully complex"[60] in contrast to fully ordered signals such as strictly periodical signals. This view is also adopted by Batty et al. for the context of cities.[23] Costa et al. conclude that "if one supposes that greater entropy is characteristic of greater complexity, such results are profoundly misleading."[42].

We show that almost arbitrary results can be obtained depending on how the phase space is defined. The key is to define a phase space that is conceptually grounded in how the macroscopic state of the system is produced. We argue that in a system in which the microstates are spatially dependent, this must be considered. The analysis of synthetic patterns with multiscale spatial entropy shows that in that case, complex patterns have the highest entropy. We can thus partly explain the spatial complexity that is frequently observed in cities[61–64] - and more generally the complexity of patterns with interdependent observations - as simply the kind of pattern we are most likely to observe because they can occur in more ways than others.

What is ignored so far entirely, except for a vague notion of some interaction between different places, is essentially everything else we already know about cities: How people use them, or how social and economic processes shape their structure. Paradoxically, that is precisely why this might be a powerful concept. It allows us to make *the statistically best guess about what we do not know.* From a planners perspective, we would try to optimise our planning effort based on some assumptions about people and societies, how they should or want to use cities, and beyond that based on some prediction about the future and an assessment of what should be considered a "good" city. There is a limit to how certain we can be about these assumptions. If we believe to know a number of things with varying certainty, a conceptually consistent theory of urban entropy could be used to *physically express that uncertainty in the structures we build.* That way we could increase the probability to have a positive result even if our assumptions were wrong.

References

1. Jacobs, J. *The death and life of great American cities* (Vintage, 1961).
2. Batty, M. *The new science of cities* (Mit Press, 2013).
3. Batty, M. *Cities and complexity: understanding cities with cellular automata, agent-based models, and fractals* (The MIT press, 2007).
4. Allen, P. M. Cities and regions as evolutionary, complex systems. *Geographical systems* **4**, 103–130 (1997).
5. Johnson, S. *Emergence: The connected lives of ants, brains, cities, and software* (Simon and Schuster, 2002).
6. Portugali, J. What makes cities complex? *Complexity, Cognition, Urban Planning And Design. Springer* (2014).
7. Allen, P. M. Cities: the visible expression of co-evolving complexity. In *Complexity Theories of Cities Have Come of Age*, 67–89 (Springer, 2012).
8. Tobler, W. R. A computer movie simulating urban growth in the detroit region. *Economic geography* 234–240 (1970).
9. Alexander, C. A city is not a tree. *Ekistics* **23**, 344–348 (1967).
10. Wilson, A. G. A family of spatial interaction models, and associated developments. *Environment and Planning* **3**, 1–32 (1971).
11. Haynes, K. E. & Fotheringham, A. S. *Gravity and spatial interaction models*, vol. 2 (Sage publications Beverly Hills, 1984).
12. Maguire, D. J., Batty, M. & Goodchild, M. F. *GIS, spatial analysis, and modeling* (Esri Press, 2005).
13. Harris, B. & Wilson, A. G. Equilibrium values and dynamics of attractiveness terms in production-constrained spatial-interaction models. *Environment and planning A* **10**, 371–388 (1978).
14. Fujita, M. & Thisse, J. *Economics of Agglomeration: Cities, Industrial Location, and Globalization* (Cambridge University Press, 2013). URL https://books.google.co.uk/books?id=w-Jm5K_PSGcC.
15. Galster, G. C. The mechanism(s) of neighbourhood effects: Theory, evidence, and policy implications. In *Neighbourhood effects research: New perspectives*, 23–56 (Springer, 2012).
16. Wehrl, A. General properties of entropy. *Reviews of Modern Physics* **50**, 221 (1978).
17. Shannon, C. A mathematical theory of communication. *Bell System Technical Journal, The* **27**, 379–423 (1948). DOI 10.1002/j.1538-7305.1948.tb01338.x.
18. Cápek, V. & Sheehan, D. P. *Challenges to the second law of thermodynamics* (Springer, 2005).
19. Kleidon, A. & Lorenz, R. D. *Non-equilibrium thermodynamics and the production of entropy: life, earth, and beyond* (Springer Science & Business Media, 2005).
20. Leff, H. S. Entropy, its language, and interpretation. *Foundations of Physics* **37**, 1744–1766 (2007). URL http://dx.doi.org/10.1007/s10701-007-9163-3. DOI 10.1007/s10701-007-9163-3.
21. Batty, M. Spatial entropy. *Geographical analysis* **6**, 1–31 (1974).
22. Batty, M. Space, scale, and scaling in entropy-maximising. *Geographical Analysis* **42, 4**, 395–421 (2010).
23. Batty, M., Morphet, R., Masucci, P. & Stanilov, K. Entropy, complexity, and spatial information. *Journal of geographical systems* **16**, 363–385 (2014).
24. Gudmundsson, A. & Mohajeri, N. Entropy and order in urban street networks. *Nature Scientific Reports* **3**, Article number: 3324 (2013).
25. Henri Theil, A. J. F. A note on the measurement of racial integration of schools by means of informational concepts. *The Journal of Mathematical Sociology* **1:2**, 187–193 (1971). Http://dx.doi.org/10.1080/0022250X.1971.9989795.
26. Duncan, O. D. & Duncan, B. A methodological analysis of segregation indexes. *American sociological review* 210–217 (1955).
27. Jahn, J., Schmid, C. F. & Schrag, C. The measurement of ecological segregation. *American Sociological Review* 293–303 (1947).
28. Williams, J. J. Another commentary on so-called segregation indices. *American Sociological Review* 298–303 (1948).

29. Bell, W. A probability model for the measurement of ecological segregation. *Social Forces* 357–364 (1954).

30. White, M. J. The measurement of spatial segregation. *American journal of sociology* 1008–1018 (1983).

31. Morgan, B. S. An alternate approach to the development of a distance-based measure of racial segregation. *American Journal of Sociology* 1237–1249 (1983).

32. Stearns, L. B. & Logan, J. R. Measuring trends in segregation three dimensions, three measures. *Urban Affairs Review* **22**, 124–150 (1986).

33. Lieberson, S. & Carter, D. K. Temporal changes and urban differences in residential segregation: a reconsideration. *American Journal of Sociology* 296–310 (1982).

34. Massey, D. S. & Denton, N. A. The dimensions of residential segregation. *Social forces* **67**, 281–315 (1988).

35. Wong, D. W. Spatial dependency of segregation indices. *The Canadian Geographer/Le Géographe canadien* **41**, 128–136 (1997).

36. Taeuber, K. E. & Taeuber, A. F. *Negroes in cities: Residential segregation and neighborhood change* (Atheneum, 1969).

37. Wilson, A. G. *Entropy in urban and regional modelling* (Pion Ltd, 1970).

38. Zhang, Y., Yang, Z. & Li, W. Analyses of urban ecosystem based on information entropy. *Ecological Modelling* **197**, 1–12 (2006).

39. Jaynes, E. T. Gibbs vs boltzmann entropies. *American Journal of Physics* **33**, 391–398 (1965).

40. Swendsen, R. H. & Wang, J.-S. The gibbs" volume" entropy is incorrect. *arXiv preprint arXiv:1506.06911* (2015).

41. Zhang, Y.-C. Complexity and 1/f noise. a phase space approach. *Journal de Physique I* **1**, 971–977 (1991).

42. Costa, M., Goldberger, A. L., Peng, C. *et al.* Multiscale entropy analysis (mse) (2000).

43. Costa, M., Goldberger, A. L. & Peng, C.-K. Multiscale entropy analysis of complex physiologic time series. *Physical review letters* **89**, 068102 (2002).

44. Costa, M., Goldberger, A. L. & Peng, C.-K. Multiscale entropy analysis of biological signals. *Physical review E* **71**, 021906 (2005).

45. Costa, M. D. & Goldberger, A. L. Generalized multiscale entropy analysis: Application to quantifying the complex volatility of human heartbeat time series. *Entropy* **17**, 1197–1203 (2015).

46. Humeau-Heurtier, A. The multiscale entropy algorithm and its variants: A review. *Entropy* **17**, 3110 (2015). URL http://www.mdpi.com/1099-4300/17/5/3110. DOI 10.3390/e17053110.

47. Mandelbrot, B. B. *The fractal geometry of nature*, vol. 173 (Macmillan, 1983).

48. Addison, P. *Fractals and Chaos: An illustrated course* (Taylor & Francis, 1997).

49. Hazewinkel, M. *Encyclopaedia of Mathematics: Coproduct — Hausdorff — Young Inequalities*. Encyclopaedia of Mathematics (Springer US, 2013). URL https://books.google.co.uk/books?id=thf3BwAAQBAJ.

50. Bartz-Beielstein, T., Chiarandini, M., Paquete, L. & Preuss, M. *Experimental methods for the analysis of optimization algorithms* (Springer, 2010).

51. Michalowicz, J., Nichols, J. & Bucholtz, F. *Handbook of Differential Entropy* (CRC Press, 2013). URL https://books.google.co.uk/books?id=k2fOBQAAQBAJ.

52. Miller, E. G. A new class of entropy estimators for multi-dimensional densities. In *Acoustics, Speech, and Signal Processing, 2003. Proceedings.(ICASSP'03). 2003 IEEE International Conference on*, vol. 3, III–297 (IEEE, 2003).

53. Stowell, D. & Plumbley, M. D. Fast multidimensional entropy estimation by k-d partitioning. *IEEE Signal Processing Letters* **16**, 537–540 (2009). URL http://dx.doi.org/10.1109/LSP.2009.2017346. DOI 10.1109/lsp.2009.2017346.

54. Stanilov, K. & Batty, M. Exploring the historical determinants of urban growth patterns through cellular automata. *Transactions in GIS* **15**, 253–271 (2011).

55. Kloosterman, R. C. & Musterd, S. The polycentric urban region: Towards a research agenda. *Urban Studies* **38**, 623–633 (2001). URL http://dx.doi.org/10.1080/00420980120035259. DOI 10.1080/00420980120035259. http://dx.doi.org/10.1080/00420980120035259.

56. Brueckner, J. K. & Fansler, D. A. The economics of urban sprawl: Theory and evidence on the spatial sizes of cities. *The Review of Economics and Statistics* **65**, 479–482 (1983). URL http://www.jstor.org/stable/1924193.

57. Shalizi, C. R., Shalizi, K. L. & Haslinger, R. Quantifying self-organization with optimal predictors. *Physical Review Letters* **93**, 118701 (2004).

58. Wolfram, S. Statistical mechanics of cellular automata. *Reviews of modern physics* **55**, 601 (1983).

59. Klimontovich, I. *Turbulent motion and the structure of chaos: a new approach to the statistical theory of open systems.*

60. Garland, J., James, R. & Bradley, E. Model-free quantification of time-series predictability. *Physical Review E* **90**, 052910 (2014).

61. Bettencourt, L. M., Lobo, J., Helbing, D., Kühnert, C. & West, G. B. Growth, innovation, scaling, and the pace of life in cities. *Proceedings of the national academy of sciences* **104**, 7301–7306 (2007).

62. Batty, M. & Longley, P. A. *Fractal cities: a geometry of form and function* (Academic Press, 1994).

63. Chen, Y., Wang, J. *et al.* Multifractal characterization of urban form and growth: the case of beijing. *Environment and Planning B: Planning and Design* **40**, 884–904 (2013).

64. Chen, Y. & Zhou, Y. Multi-fractal measures of city-size distributions based on the three-parameter zipf model. *Chaos, Solitons & Fractals* **22**, 793–805 (2004).

65. Anderson, J. R. *A land use and land cover classification system for use with remote sensor data*, vol. 964 (US Government Printing Office, 1976).

66. Harrison, A. National land use database: land use and land cover classification. *O. Office of the Deputy Prime Minister, Ed., ed. London, UK: Queen's Printer and Controller of Her Majesty's Stationery Office* 81 (2006).

67. R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria (2008). URL http://www.R-project.org. ISBN 3-900051-07-0.

supplementary material

Data and Preprocessing

Stanilov et al.[54] provide detailed information on the collection process of the data that involved digitising historical maps from Ordnance Survey (OS) on the scale 1:2,500. The land use classes were categorised manually: "The process of land use classification involved the interpretation of building footprints from the OS maps; verification of building type (for buildings still in existence) in Microsoft Virtual Earth (now Bing Maps 2D and 3D) and Google Street View; and cross-referencing the results with several land use databases for Greater London"[54], while for some categories label information was given in the OS maps. It is evalutated to be a "representative sample of Londons metropolitan fabric"[54], but differences are pointed out between different parts of London on the scale of the study area. As in this study the diversity of land uses is itself at the focus of attention, the biased choice of the study area towards the section with the largest variety may propagate to make it less representative of London for this works purpose.

The functions in the original data are grouped into three main groups: Residential, business and leisure (table). The chosen categories broadly reflect classifications used by geoinformation sciences[65], urban planning theory[1] and practice[66], altered to cover a broader range of uses. "Leisure" is taken in the broadest sense of activities not related to workplace or home, including most categories not covered by the former two. A virtual fourth category appears in the data's empty space.

Business		Residential		Leisure	
INS	institutional	APT	apartments	GEN	mixed/commercial
INSL	large institutional	APTH	high rise apt	RET	big box retail
O	office	COT	cottages	OLD	old fabric/mix
GAR	garages	DET	detached housing	AGR	alotment gardens
IND	industrial	DETH	high density detached	CEM	cemeteries
UTL	utilities	MEW	mews	EST	land estates
AIR	airport	SDT	semi-detached housing	FRM	farm structures
RRS	rail stations	TER	terraced housing	NRS	tree nurseries
		LDG	lodges / hotel	PRK	parks
				REC	recreational
				CHR	religious
				WAT	water
				STA	stadia
				SCH	schools
				CLR	cleared

All parts of the analysis are performed in R.[67] The spatial data, provided in Shapefile format (.shp), is imported and transformed from the global positioning system (GPS) into the Universal Transverse Mercator coordinate system (UTM). It is then rasterized as a grid with at a resolution of 50m.

Method and Computation

Then, the local values of all neighbourhood sizes are calculated and turned into a matrix in which there is a row for every point in space, and a column for the binned value of each category at each scale of observation. Identical rows are grouped, counted and translated to probabilities from which entropy is calculated.
The rasterised data is split into subsets containing each only one of the categories. We calculate the mean number of pixels of each category within a square moving window with the selected neighbourhood sizes:50m, 150m, 450m, 1350m, 4050m. The choice of neighbourhood size requires further investigation in the future. For now, the number of scales is picked to keep a reasonable proportion between unique values for the place state matrix and the number of observed pixels. The largest size is selected to cover a substantial area of the total space, while still allowing for a sufficient number of non overlapping large scale areas. For neighbourhoods close to the edge, the missing neighbourhood area is assumed to have the same proportion of functions than in the available part. While this is preferable to a wrap-around torus, due to the asymetric nature of the data, it can lead to edge effects, making extreme values more likely along the edges. In a visual analysis of the probabilities of pixels in synthetic patterns and the study data, we found no considerable anomalies near the edges.

We now have a spatial matrix for each category at each scale of observation. The values are transformed into a single matrix with a row for each pixel, and column for each scale, giving a vector for each pixel with the total proportion of each category at each scale as a value between 0 and 1. The values are discretised in 3 equidistant bins, split at 1/3 and 2/3. The number of bins changes the total number of possible states and should return a reasonable proportion between possible states and observations. The binning is more than a technical question; One could ask how different two places need to be in their proportion of functions to have different states. We then count how many vectors of each unique combination we find in the data and calculate the entropy from the probabilities directly.

We repeat the randomised null models as well as the synthetic patterns 200 times. The confidence intervals of the null models are too small to be visible in the plot because the number of pixels is a large sample within each randomised run already.
The synthetic patterns are produced as follows:

pattern a) Uniformly random:
All pixels are set to 0 or 1 with equal probability.

pattern b)-d) Segregated:
We start with a smaller matrix of which a side length for which the target size is a multiple of. All pixels are set to 0 or 1 with equal probability, and each pixel is expanded to multiple pixels so that the target size is produced. This

process can lead to a varying number of 0's and 1's with increasing scale of segregation.

pattern e) Sorted:
A matrix is filled with a uniform distribution between 0 and 1 and sorted in x and y direction.

pattern f) 1/f noise:
A matrix is filled with 0's an 1's, with the probability to receive a 1 increasing linearly from 0 to 1 with increasing x position.

pattern g) additive cascade:
four different values are set and stored, their exact values are not relevant. A 2 by 2 matrix is created. Each pixel of the matrix is randomly set to one of the values. Each pixel is then expanded to 4 pixels, and randomly one of the four values is added to each of these. The process is repeated until the matrix target size is reached. Then, all values lower than the median are set to zero, all others to 1.

sensitivity analysis
NECESSARY? -sensitivity to rasterisation resolution
-sensitivity to scales
-sensitivity to binning

List of Figures

Figure 1. a linear pattern with equal probability to find any combination of 2 or 4 adjacent pixels (when divided into non overlapping segments)

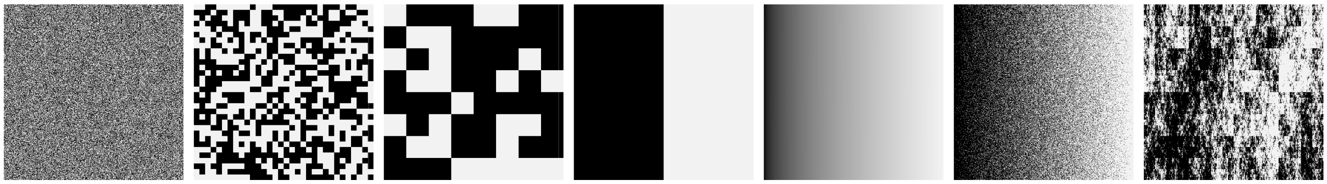

Figure 2. patterns a) - g)

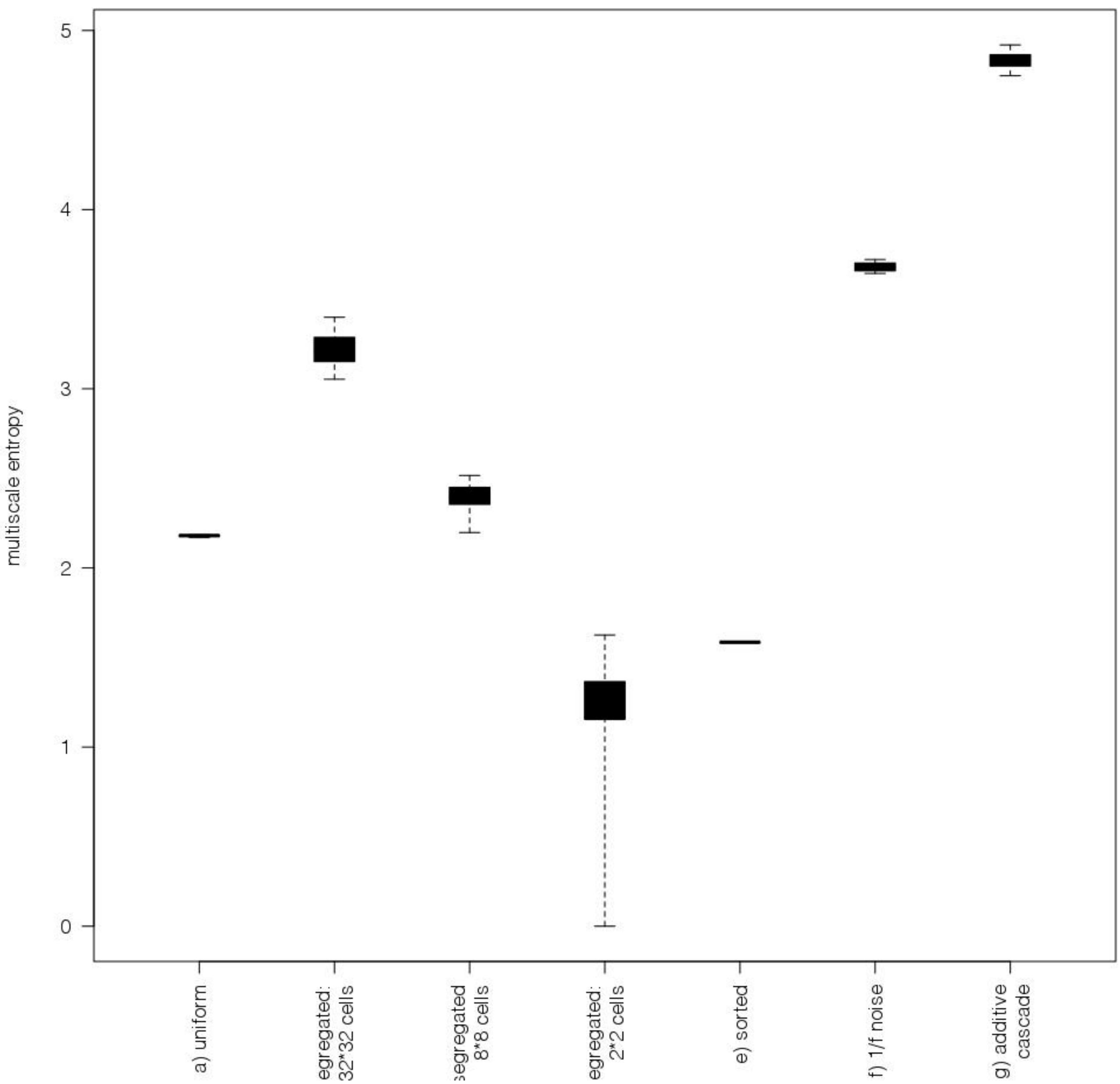

Figure 3. multiscale entropy for patterns a) - g)

Figure 4. probability to find black pixels in zones for the geographical phase space

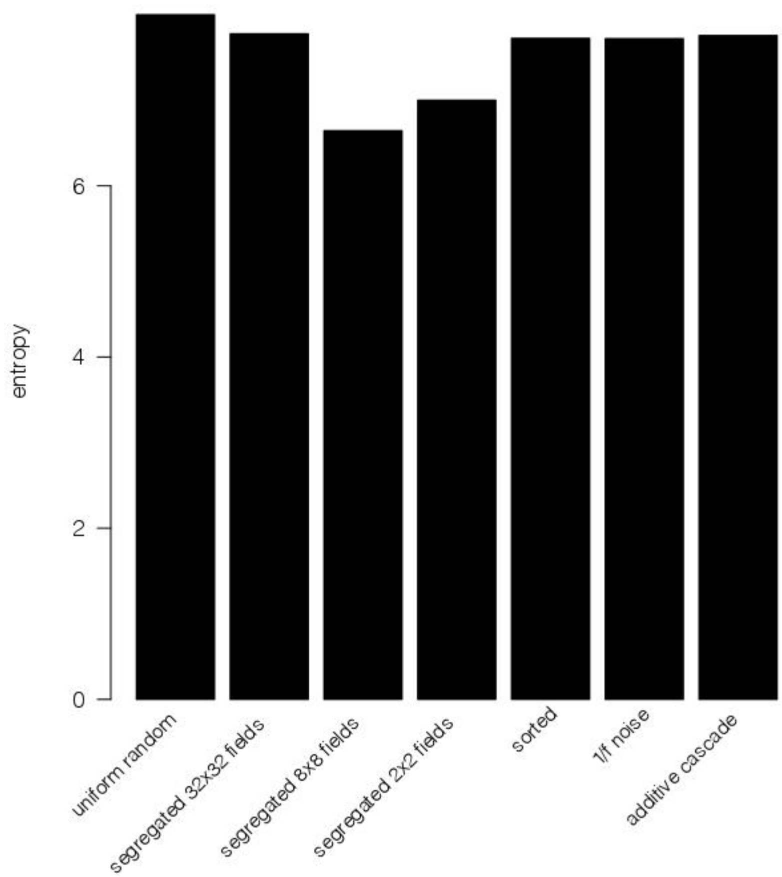

Figure 5. Entropies for patterns in figure 2 in the geographical phase space

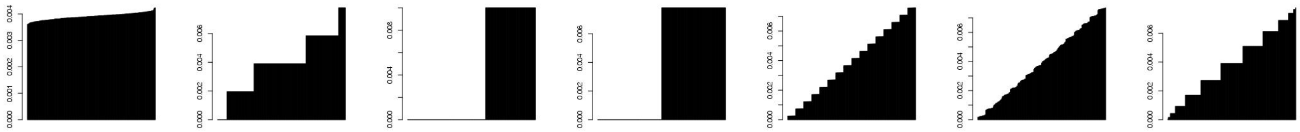

Figure 6. black pixel frequencies in the zones of the geographical phase space, patterns a)-g) in figure 2

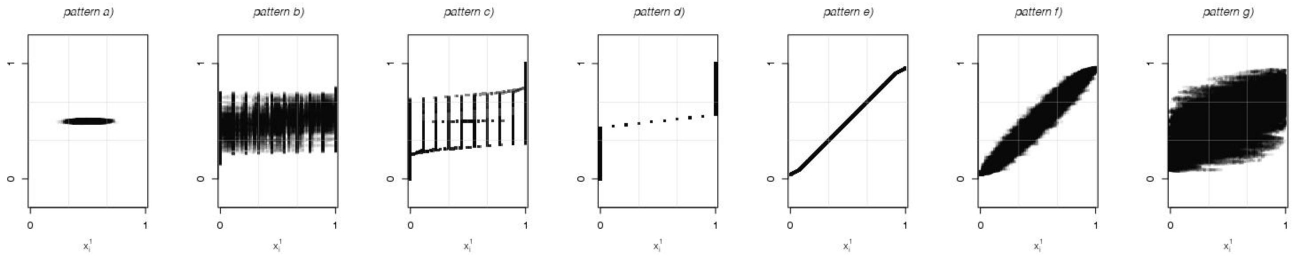

Figure 7. phase space distributions for patterns a) - g) at scales of 9 and 81 pixels neighbourhood side length

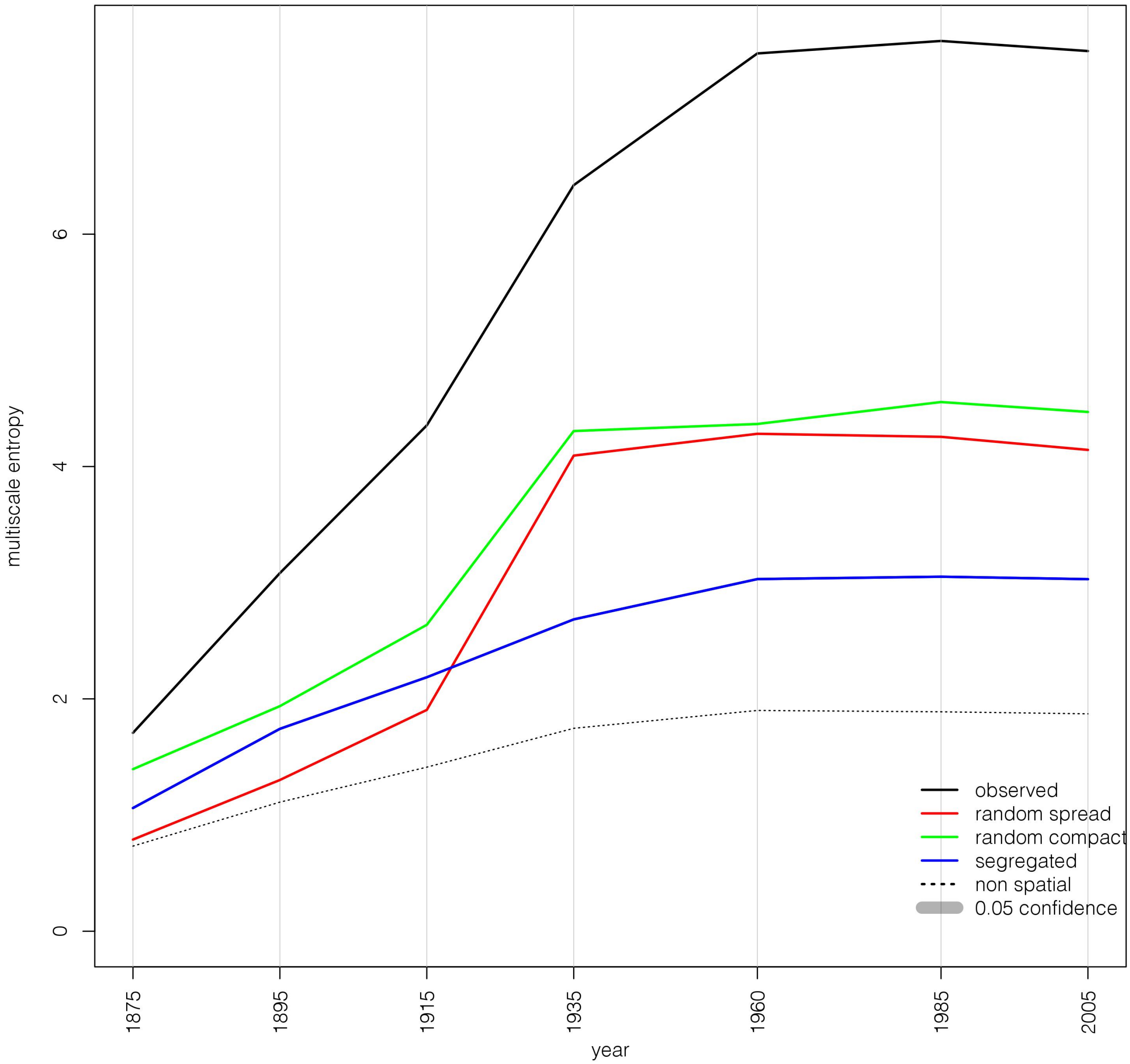

Figure 8. Multiscale entropy in West London over time compared to 3 null models

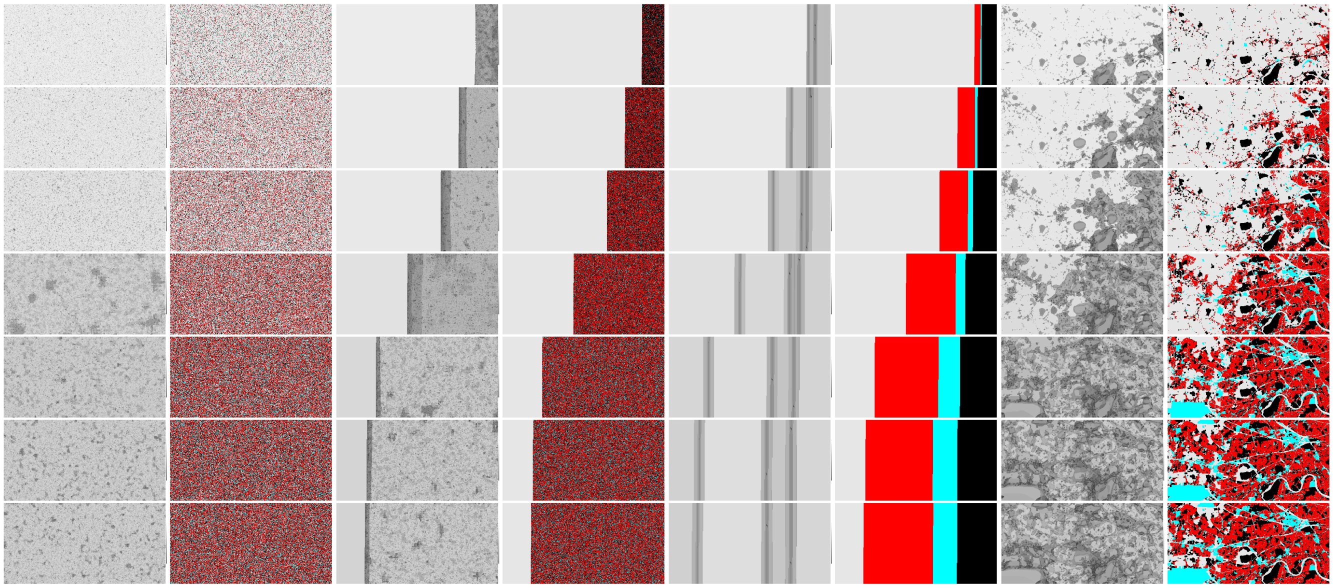

Figure 9. The probability of each pixel's state, and the corresponding spatial distribution of functions. From left to right: Random pixel allocation, compact mixed use growth, compact segregated growth and observed data. Global proportion of functions and observed data correspond to 1875, 1895, 1915, 1935, 1960, 1985, 2005 from top to bottom. Grey: Undeveloped or no data. Red: Residential. Blue: Commercial. Black: Leisure. Grayscale images decreasing probability with increasing brightness (logarithmic)