Measuring heterogeneity in urban expansion via spatial entropy

Published Version:
Linda Altieri, D.C. (2019). Measuring heterogeneity in urban expansion via spatial entropy. ENVIRONMETRICS, 30(2), 1-16 [10.1002/env.2548].

Availability:
This version is available at: https://hdl.handle.net/11585/677249 since: 2022-02-22

Published:
DOI: http://doi.org/10.1002/env.2548

(Article begins on next page)

Measuring heterogeneity in urban expansion via spatial entropy

L. Altieri[1], D. Cocchi, G. Roli

Abstract

The lack of efficiency in urban diffusion is a debated issue, important for biologists, urban specialists, planners and statisticians, both in developed and new developing countries. Many approaches have been considered to measure urban sprawl, roughly identified as chaotic urban expansion; such idea of chaos is here linked to the concept of entropy. Entropy, firstly introduced in information theory, has rapidly become a standard tool in ecology, biology and geography to measure the degree of heterogeneity among observations; in such contexts, entropy measures should include spatial information. The aim of this paper is to employ a rigorous spatial entropy based approach to measure urban sprawl associated to the diffusion of metropolitan cities. In order to assess the performance of the considered measures, a comparative study is run over archetypical urban scenarios; afterwards, measures are used to quantify the degree of disorder in the urban expansion of three cities in Europe. Results are easily interpretable and can be used both as absolute measures of urban sprawl and for comparison over space and time.

Keywords: urban sprawl, environmental heterogeneity, spatial entropy, categorical variables.

[1]Corresponding author

Department of Statistical Sciences, University of Bologna, via Belle Arti, 41, 40126, Bologna, Italy

linda.altieri@unibo.it

1 Introduction

Urban sprawl is characterized by an uncontrolled development of cities into surrounding areas, and has aroused wide social focus because the urbanization of a sprawled city is inefficient, dispersed and may impede sustainable development. Rapid urban growth is quite alarming worldwide; the phenomenon has different implications between developed and developing countries, thus the importance of conducting research on this topic is strongly felt (Johnson, 2001; Ewing, 2008; Rosni and Noor, 2016). Although an accurate definition of urban sprawl is still debated, the general consensus is that urban sprawl is characterized by an 'unplanned and uneven pattern of growth, driven by a multitude of processes and leading to an inefficient resource utilization' (Bhatta et al., 2010). More definitions appear in Jaeger et al. (2010) and focus on the negative consequences of sprawl. The negative impacts of urban sprawl concern many aspects, not only for human life quality (e.g. increased costs and time for transportation), but also for the environment. The dispersion of urban areas increases pollution, waste of soil and soil consumption. This endangers ecosystems and species, and reduces the availability of land for rural areas, water bodies, forests and other natural areas (EEA and FOEN, 2016). In addition, urban sprawl does not foster climate changes mitigation, even if variations in climate do not immediately fit with the velocity of an uncontrolled urbanization. Indeed, any spatial planning strategy has a different impact on climate change (Bart, 2010; Stone, 2012), but the standard consequences of an uncontrolled urbanization concern strong precipitation events, additional heat due to increased emissions of carbon dioxide and, in particular, heat island effects. For the above reasons, urban sprawl is a major environmental issue as regards land misuse, which ought to be associated with space. The environmental footprint of a city is a broad concept that includes non-spatial aspects as well. Therefore, measurements of urban sprawl do not result in a general evaluation of the, usually negative, impact of a city on the environment. A global assessment would require an integration of such measures with other environmental quality indicators, and is beyond the scope of this work.

In Europe, urban sprawl is an increasingly serious issue linked to soil consumption and sealing (EEA, 2006; Couch et al., 2007; EEA and FOEN, 2016). This drawback can be evaluated according to several viewpoints. For instance, EEA and FOEN (2016) stress that the spatial configuration of the built up areas is a fundamental component of urban proliferation. Different arguments in EEA reports point out the impact of urban sprawl: the negative effects mentioned before are evident if the costs for future generations are taken into account, and are related to the ideas of fragmentation, degradation and consequences on ecosystems.

The literature about sprawl measurements is voluminous (e.g. Torrens, 2008; Bhatta et al., 2010; Cabral et al., 2013; Ewing and Hamidi, 2015; Oueslati et al., 2015); the quantification of the phenomenon develops according to different routes that keep into account alternative formulations of demographic, social and economic variables. This is partly due to the difficulties of a unique definition. Moreover, characterisation of sprawl in the literature is often narrative and subjective, to the point that existing studies yield contrary results for the same cities in several cases (Torrens, 2008; Bhatta et al., 2010). A sprawl indicator should detect whether an unnecessary waste of urbanised land occurs; one example of such indicator is the low level of population density over an area. Sprawl may also be defined in cost terms, as in Benfield et al. (1999), or by ratios of urban growth indicators (Ewing and Hamidi, 2015). A lot of sprawl measures are indeed based on ratios: relative measures quantify attributes of urban growth and can be compared among cities, among different zones within a city, or across different time points (Bhatta et al., 2010). Such ratios are easy to interpret and receive a lot of discussion, but they are statistically poor. In order to capture different aspects that are related to sprawl, Jiang et al. (2007) proposed an integrated urban sprawl measure that combines 13 indices; unfortunately, the final measure requires extensive inputs of temporal data, and does not mention any threshold to characterise a city as sprawling.

Among the proposals for urban sprawl measurements, there is a number of spatial or landscape metrics, that have long been used in landscape ecology. Landscape metrics aim at evaluating the spatial pattern of land cover classes or entire landscape mosaics of a geographic area. Indeed, the urbanization of a territory can be assessed according to the exhibited pattern of land cover classes: a sprawled city is in contrast with a compact one, with 'empty' (i.e. non-urban) spaces and scattered urban areas denoting inefficient development. Consequently, land cover and land use data are particularly suitable for urban sprawl measurements. Such data usually are vector (polygonal) or raster (pixel) spatial data coming from remote sensing images, where the territory is classified according to a finite number of categories defining the prevailing land use. A definition about what land use classes are considered urban or non-urban is needed, then the pattern of urban areas and its evolution over time can be exploited to quantify urban dispersion as lack of spatial clustering (compactness) of the urban patch. For an approach to sprawl measurement based on a comparative use of Moran's Index with land use data, see Altieri et al. (2014).

One approach to measuring sprawl through landscape metrics is based on entropy measure, for two main reasons: the need to deal with categorical variables and the detection of heterogeneity, i.e. lack of compactness, in the territory. Shannon's entropy is used in several fields, such as geography, ecology, biology, to assess the heterogeneity of a population over an area. Ecological concepts, such as evenness and richness, are strictly related to heterogeneity, and entropy represents the utmost index to measure heterogeneity in a dataset. In the context of urban sprawl, entropy has proved to be a rigorous measurement tool (Yeh and Li, 2001) and is still a widely used technique, suitable for integration of remote sensing and GIS (Bhatta et al., 2010; Chong, 2017; Liu and Chen, 2018). Entropy measures, being entirely based on the occurrences of realizations of predefined categories, can be proposed and computed also in situations, like remote sensing, where information is limited but extensively available. While entropy succeeds in working with qualitative variables and quantifying the heterogeneity of a dataset, in its most known formulation it suffers from the drawback of not considering the role of space as a source of heterogeneity in determining the variable outcomes. Indeed, Shannon's entropy is computed based on the proportions of the land use classes (the common choice for estimating land use probabilities), not on their spatial configurations, and two territories with the same proportions and very different degrees of compactness for the urban tissue share the same entropy value. The urban sprawl issue is tightly bonded to the spatial distribution of land use classes. Therefore, appropriate studies of sprawl which make use of entropy measures should introduce spatial information.

Over the past decades, two main approaches have been adopted to include spatial information into an entropy measure. Extending Theil's work (1972), Batty (1974, 1976, 2010) introduced the first approach by defining a spatial entropy measure accounting for unequal space partition into sub-areas. In 2002, this proposal was modified by Karlström and Ceccato to satisfy the property of additivity, i.e. to decompose the global index into local terms. Three main drawbacks of this approach can be underlined: such entropy can only be computed for a binary variable, the local terms do not possess the properties of the global one, and results are heavily affected by the selected area partition. Nevertheless, in the present work the approach proves to be informative in the context of urban sprawl. The second approach to spatial entropy is based on a suitable transformation of the study variable that accounts for the Euclidean distance between realizations (co-occurrences). The main proposals are authored by O'Neill et al. (1988), Li and Reynolds (1993), Leibovici (2009) and Leibovici et al. (2014), but such distance-based measures do not enjoy the additivity property and rely on the choice of a single distance without capturing the behaviour of the studied variable across distances. A recent work by Altieri et al. (2018a) fulfils desirable properties by proposing a set of spatial entropy measures starting from the co-occurrence approach and focusing on pairs of realizations. The resulting entropy is decomposed into the information due to space and the remaining information brought by the variable itself once space is considered. The proposal preserves additivity and disaggregates results, allowing for partial and global syntheses. The properties of spatial entropy measures make them an appealing tool to evaluate urban sprawl from a spatial perspective. A spatial entropy measure is sensitive to the spatial dispersion of urban patches over an area and may be able to separate the heterogeneity of land use data due to the lack of spatial compactness from the heterogeneity due to other components. Such indices are also suitable for delivering results across different areas of expertise.

The main aim of this work is to employ entropy for measuring urban sprawl in terms of spatial compactness or dispersion. If sprawl is considered a negative condition and is measured by means of spatial entropy, a low level of entropy is desirable, corresponding to a non-chaotic (compact) urban configuration. We present a thorough assessment of the advantages and disadvantages of a set of spatial entropy measures, which have not been employed in the context of urban sprawl measurement yet, both with a comparative study on simulated data and via a case study on three European cities. The simulation study compares spatial entropy values across the representative urban configurations currently proposed in the literature (Tsai, 2005): the monocentric, the polycentric and the decentralized city. In addition, the results of the simulation study provide reliable ranges of entropy values that may be used as reference intervals for assessing an indication of urban sprawl in real case studies, as in the application, where we propose an example of comparison over space and time.

The motivating case study comes from official European land use data. Such data are available for environmental aims, and a binary raster dataset is specifically produced, for a few time points, for the evaluation of urban expansion (EEA, 2011); the dichotomization process divides the land cover categories into urban and non-urban classes and provides a pixel grid, with pixel size of 250×250 metres. The data production process ensures the appropriate resolution to analyse the phenomenon of (inefficient) urban expansion over Europe, and also that the dichotomization prevents from misinterpretation of the urban phenomenon. For example, urban parks are classified within urban areas, in order to avoid underestimation of city compactness. Data possess the proper quality and the right scale for the study at hand, and provide homogeneous information on a very vast territory. Details on such classification are given in Section 4. We selected two time points, 1990 and 2012, for the commuting belts of three cities in Europe: Eindhoven, Lublin and Bologna (also studied in Altieri et al., 2014). We chose these cities since they have similar geographical extension and population amounts. In addition, they belong to countries with different levels of urban sprawl (EEA, 2006), giving us the opportunity of exploring if different levels of sprawl are also detectable at the city level.

The focus of this work is on the spatial aspect of inefficient urban development. Should this be of interest, our results can be combined with measures integrating relevant demographic, social or economical variables affecting urban sprawl, e.g. following the idea of EEA (2006), to grasp the global environmental footprint of a city. Though spatial entropy is applied to the specific issue of urban sprawl, the techniques illustrated in the present paper may be used for any environmental phenomenon whose spatial distribution and heterogeneity is of interest. The evaluation of the abilities of such techniques is relevant for climate and meteorology studies, e.g. the spatial distribution of metereological phenomena, for ecological purposes, e.g. species distribution (Altieri et al., 2018a), for landscape and geographical studies, for the assessment of environmental risks, e.g. earthquakes and wildfires, for atmospheric studies, e.g. polluting substances, and for disease mapping.

In the present paper, in Section 2 we revisit the works by Batty (1974), Karlström and Ceccato (2002) under a unified statistical framework. We also illustrate the approach of Altieri et al. (2018a) with a special focus on its use in urban sprawl studies. In Section 3, we build a simulation study, which helps in comparing and discussing the performances of the two approaches for spatial entropy measures under different urban scenarios. This is useful for further applications, since the study covers the main urban configurations explored by the theory (Tsai, 2005), and also as a contribution to the statistical theory of spatial entropy measures. In Section 4, the measures are applied to the case study; this constitutes a further practical contribution to the discussion on urban sprawl. Some concluding remarks can be found in Section 5.

This work is implemented in `R` (R Core Team, 2017). It makes use of the packages `sp` (Bivand et al., 2013), `spatstat` (Baddeley et al., 2015) and dependencies, and of the recent package `SpatEntropy` (Altieri et al., 2018b), now available on CRAN.

2 Spatial entropy measures for urban sprawl

In many environmental and urban studies, the definition of entropy coincides with Shannon's formula: given a categorical variable X with I possible outcomes, the entropy is

$$H(X) = \sum_{i=1}^{I} p(x_i) \log \left(\frac{1}{p(x_i)} \right) \tag{1}$$

where $p(x_i)$ is the probability of the ith outcome, $i = 1, \ldots, I$, and $\log(1/p(x_i))$ is the information function, measuring the information brought by outcome x_i (Cover and Thomas, 2006). Entropy is a non-negative quantity, which quantifies the average 'information' or 'surprise' concerning X. The more the categories of X are equally likely, the higher the entropy; if a category of X is far more likely than others, the entropy is low, as one can predict the behaviour of X and data do not carry much information. Thus, entropy synthesizes the heterogeneity of X in a single number; data with very different spatial configurations but the same probability mass function for X share the same entropy, which is not desirable in the context of urban sprawl. For example, an area which is partly urbanized and partly rural may be either compact, with an urban nucleous and rural surroundings, or dispersed, i.e. sprawled, with many small scattered urban areas. Shannon's entropy does not detect the difference in the two patterns and returns the same value when the proportion of urbanized and non-urbanized territory is the same across the two configurations.

For this reason, an extension to spatial entropy is needed. The seminal attempt to extend (1) into a spatial entropy measure, developed by Batty (1974), is presented in Section 2.1; its most relevant extension, proposed by Karlström and Ceccato (2002), is sketched in Section 2.2. A recent approach to spatial entropy, proposed by Altieri et al. (2018a), is in Section 2.3. The remainder of this work shows that these measures are very suitable in distinguishing between urban compactness and urban sprawl, though with theoretical differences and peculiar properties.

The spatial entropy measures presented in this Section make use of a few concepts. 'Space', in this work, is intended as the two-dimensional space, as everything is applied to geographical maps. The 'observation window' is a fixed, limited spatial region with known size and shape; the spatial phenomenon under study potentially exists everywhere, but is only detected over the observation window. The window is partitioned into 'spatial units', which may be pixels, each carrying only one realization of X, or 'areas', i.e. polygonal regions that may contain several pixels; spatial units are defined via representative coordinate pairs, such as the unit centroids, used to measure distances. 'Distances' are always intended as the Euclidean distance on the two-dimensional space. Moreover, the ideas of spatial adjacency and neighbourhood are fundamental in Section 2.2 and 2.3. The concept of neighbourhood means that occurrences at certain spatial units are influenced, in a positive or negative sense, by what happens at surrounding units, i.e. their neighbours. The system can be represented by a graph (Bondy and Murty, 2008), where each spatial unit corresponds to a vertex and neighbouring units are graphically connected by edges. The simplest way of representing a neighbourhood system is via an adjacency matrix. In the graph representation, for G vertices, $A = \{a_{gg'}\}_{g,g'=1,\ldots,G}$ is a square $G \times G$ matrix such that $a_{gg'} = 1$ when there is an edge from vertex g to vertex g', and $a_{gg'} = 0$ otherwise. In an observation window with G spatial units, the element of the square matrix $a_{gg'} = 1$ if the unit $g' \in \mathcal{N}(g)$, the neighbourhood of unit g. In the standard definition of A, the diagonal elements are all zero (Anselin, 1995). In the remainder of the paper, the word 'adjacent' means 'neighbouring', i.e. connected in the graph, while the word 'contiguous' is used for pixels or polygons sharing a border on the map, i.e. for a topological contact.

The definition of entropy as a discrete random variable relies on categories' probabilities. In the literature and in the present work, spatial entropy measures are proposed as descriptive indices, substituting the unknown probabilities with the observed relative frequencies, obtaining the non parametric as well as the maximum likelihood entropy estimator (Paninski, 2003).

2.1 Batty's spatial entropy

A very appreciable attempt to include spatial information into Shannon's entropy starts from a reformulation of (1). This approach is proposed by Batty (1974; 1976) to define a spatial entropy which extends Theil's work (1972), and does not introduce the adjacency matrix yet. In a spatial context, a phenomenon F (the binary case of the variable X, i.e. presence/absence) is detected over an observation window. The window has size T and is partitioned into G areas of size T_g, $g = 1, \dots, G$. The partition into G areas defines G dummy variables identifying the occurrence of F over a generic area g. The occurrence of F in area g has probability p_g, where $\sum_g p_g = 1$. The intensity of F over the area g is defined by $\lambda_g = p_g/T_g$, where T_g is the area size, and is assumed constant within each area. Shannon's entropy of F follows, where 'categories' are replaced by 'spatial units':

$$H(F) = \sum_{g=1}^{G} p_g \log\left(\frac{1}{p_g}\right) = \sum_{g=1}^{G} \lambda_g T_g \log\left(\frac{1}{\lambda_g}\right) + \sum_{g=1}^{G} \lambda_g T_g \log\left(\frac{1}{T_g}\right). \tag{2}$$

Batty (1976) shows that the first term on the right hand side of the formula converges to the continuous version of Shannon's entropy (Rényi, 1961), namely the differential entropy, as the area size T_g tends to zero, and rewrites it in terms of p_g, giving Batty's spatial entropy

$$H_B(F) = \sum_{g=1}^{G} p_g \log\left(\frac{T_g}{p_g}\right). \tag{3}$$

It expresses the average amount of information brought by the occurrence of F over the areas, and includes T_g, that accounts for unequal space partition. Analogously to Shannon's entropy, which is high when the I categories of X are equally likely in a non-spatial context, Batty's entropy is high when the phenomenon of interest F tends to be equally intense over the G areas partitioning the observation window; its maximum value is $\log(T)$, reached when $\lambda_g = \lambda = 1/T$ for all g. The maximum value depends neither on the area partition, nor on the discrete or continuous nature of F, but only on the size of the observation window. Batty's entropy $H_B(F)$ reaches a minimum value equal to $\log(T_{g^*})$ when $p_{g^*} = 1$ and $p_g = 0$ for all $g \neq g^*$, with g^* denoting the area with the smallest size.

In conclusion, Batty's entropy evaluates the heterogeneity in the intensity of a spatial phenomenon; it reaches a maximum when the intensity is constant over the areas, irrespective of the value of the intensity itself. When the target is to measure urban sprawl, F denotes the presence of urbanization, and, for a reliable use of Batty's entropy, the index must be computed on a city as a whole, with a monocentric scenario (Tsai, 2005) as the archetype of a positive development. With this assumption, a high level for Batty's entropy is not desirable, as it indicates constant urban intensity, i.e. scattering of urban patches across regions, denoting sprawl. A low level, on the contrary, indicates that some areas in the window have a very high urban density (usually, the city centre) while others tend not to present urbanization (i.e. the outside areas). Therefore, when Batty's entropy is low the city is compact and a scarce level of sprawl is present. The measure does not produce reliable results if applied to a sub-area of the city: for example, if only applied to the city centre (divided into G districts), a very high degree of urbanization with constant intensity over the selected area would be reflected by a high value for the entropy, wrongly detecting sprawl.

2.2 Karlström and Ceccato's spatial entropy

A challenging attempt to introduce additive properties, and to include the idea of neighbourhood in Batty's entropy index (3) via an adjacency matrix, is due to Karlström and Ceccato (2002), following the theory of Local Indices of Spatial Association (LISA, Anselin, 1995). Karlström and Ceccato's entropy index $H_{KC}(F)$ starts by weighting the probability of occurrence of F in a given spatial area g, p_g, with its neighbouring values:

$$\widetilde{p}_g = \sum_{g'=1}^{G} a_{gg'} p_{g'}. \tag{4}$$

In this proposal, the adjacency matrix A is row-standardized, and the elements on the diagonal a_{gg} are greater than zero for all g, i.e. each area neighbours itself and enters the computation of $I(\widetilde{p}_g)$. Then, an information function is defined, where T_g is discarded, as $I(\widetilde{p}_g) = \log(1/\widetilde{p}_g)$. Karlström and Ceccato's entropy index is

$$H_{KC}(F) = \sum_{g=1}^{G} p_g \log\left(\frac{1}{\widetilde{p}_g}\right), \tag{5}$$

which weights the new information function with the original probabilities, in contrast to the traditional formulation of Shannon's entropy. The maximum of $H_{KC}(F)$ does not depend on the choice of the neighbourhood and is $\log(G)$. As the neighbourhood reduces, i.e. as A tends to the identity matrix, $H_{KC}(F)$ coincides with Batty's spatial entropy (3) in the case of $T_g = 1$ for all g (where the unit size depends on the measurement unit of the observation window). The sum of local measures $p_g I(\widetilde{p}_g)$ constitutes the global index (5), preserving the LISA property of additivity.

One major disadvantage of (3) and (5) is that it is unable to consider a categorical variable X with $I > 2$ outcomes, since only one category enters the measure. In other words, F may be a specific category of X, labelled as X_i^*, and $H_{KC}(X_i^*)$ is computed to assess the spatial configuration of the realizations of X_i^*. Thus, for a X variable with $I > 2$, I different $H_{KC}(X_i^*)$ are computed, but no way is proposed to synthesize them into a single spatial entropy measure for X. Moreover, the local components are not entropy measures themselves. Lastly, conclusions are affected by the choice of the area partition. Nevertheless, Batty's and Karlström and Ceccato's approach is expected to be helpful in the context of urban sprawl with binary variables, and is assessed in Sections 3 and 4.

2.3 Spatial mutual information and residual entropy

A second way to build a spatial entropy measure consists in defining a new categorical variable Z, with $R = (I^2 + I)/2$ categories for I categories of X. Each outcome z_r, $r = 1, \ldots, R$, identifies unordered pairs of occurrences of X over space: $z_r = \{x_i, x_j\}$ with $i, j = 1, \ldots, I$ and $j \neq i$ (O'Neill et al., 1988; Li and Reynolds, 1993; Leibovici, 2009). Such change of variable is crucial in a spatial context, since observations must now be linked to spatial units (e.g. pixels), so that space is now taken into account via the distance between each pair of observations. The attention moves to the computation of (1) as Shannon's entropy of Z, $H(Z)$, instead of $H(X)$.

Altieri et al. (2018a) follow the approach based on Z and introduce a second discrete variable W, that represents space by classifying the distance at which two observations (also called occurrences) take place. A set of distances $d_0, \ldots, d_K$, with known K, is fixed, where $d_0 = 0$ and d_K is the maximum distance between any two spatial units inside the observation window. The choice of the distances $d_1, \ldots, d_{K-1}$ is exogenous and depends on the study at hand (Altieri et al., 2018a). Then, each category of W is a class $w_k =]d_{k-1}, d_k]$, with $k = 1, \ldots, K$. The classes w_k cover all possible distances within the observation window. Each distance category w_k implies the choice of a corresponding adjacency matrix A_k, which identifies pairs where the two observations lie at a distance belonging to the range $]d_{k-1}, d_k]$.

Thanks to the introduction of W, the entropy of Z may be decomposed as

$$H(Z) = MI(Z, W) + H(Z)_W \tag{6}$$

following the fundamentals of Information Theory (Cover and Thomas, 2006): the first term $MI(Z,W)$ is known as mutual information and measures the amount of the entropy of Z which is explained by its relationship with W, while the second term $H(Z)_W$ is the conditional, or residual, entropy, quantifying the remaining amount of entropy of Z once the effect of W is removed. In a spatial context, the two terms acquire a new meaning: $MI(Z,W)$ is the quantity of interest and is called spatial mutual information, because Z identifies pairs of categories of spatial observations and W collects categories of distances at which pairs occur. Spatial mutual information quantifies the part of entropy of Z due to the spatial configuration W; for the same reason, $H(Z)_W$ is the spatial global residual entropy, quantifying the information brought by Z after space has been taken into account. When Z depends on W, i.e. when the realizations of X are spatially associated, the spatial mutual information is high. Conversely, when the spatial association among the realizations of X is weak, the entropy of Z is mainly due to spatial global residual entropy.

When sprawl is under study and the variable of interest X is binary with categories urban and non-urban, Z identifies pairs with the three possible unordered combinations of urban/non-urban areas, i.e. $\{$urban, urban$\}$, $\{$urban, non-urban$\}$, $\{$non-urban, non-urban$\}$. A compact city represents the situation where the X outcomes should be highly positively correlated. In such case, spatial mutual information tends to be high, because urban areas generally have urban neighbours, while non-urban areas have non-urban neighbours; space plays a relevant role in determining the entropy of Z. The overall value of $MI(Z,W)$, however, can be negatively influenced by what happens at large distance ranges, where usually scarce correlation is present. Hence, spatial mutual information for the whole dataset may approach zero even when a compact pattern occurs, which is seemingly a flaw.

The variable W helps in overcoming this drawback, since the two terms forming $H(Z)$ can be further decomposed. Indeed, K subsets of realizations of Z are available, formed by pairs of observations belonging to each distance range, denoted by $Z|w_k$; a set of K conditional distributions is obtained, that sum up to the two components of (6). When measuring urban sprawl, this means that the degree of compactness of a city may be quantified at different distance ranges, which can help in understanding the extent and seriousness of the sprawl phenomenon.

From Information Theory, spatial mutual information

$$MI(Z,W) = \sum_{k=1}^{K} p(w_k) PI(Z|w_k) = \sum_{k=1}^{K} p(w_k) \sum_{r=1}^{R} p(z_r|w_k) \log\left(\frac{p(z_r|w_k)}{p(z_r)}\right) \quad (7)$$

is a weighted sum of partial terms $PI(Z|w_k)$, each quantifying the contribution of the kth distance range to the relationship between Z and W. In other words, each partial term measures the degree of association (compactness) in the city pattern at each distance range. The focus is expected to be on short distance ranges, where the difference between a compact city and a dispersed one is evident. By exploring these terms, an indication of the degree of sprawl can be provided.

Analogously, spatial residual entropy is

$$H(Z)_W = \sum_{k=1}^{K} p(w_k) H(Z|w_k) = \sum_{k=1}^{K} p(w_k) \sum_{r=1}^{R} p(z_r|w_k) \log\left(\frac{1}{p(z_r|w_k)}\right), \quad (8)$$

where the partial residual entropy terms measure the partial contributions to the entropy of Z due to sources other than the spatial configuration. As regards sprawl, a high value for $H(Z|w_k)$, especially at short distance ranges, is a hint for urban dispersion.

The additive terms in (7) and (8), together with their sums, constitute a rich set of spatial entropy measures. In particular, spatial mutual information has theoretical support to be considered a reliable method for measuring urban heterogeneity. It is able to maintain the information about the categories of X by exploiting the trasformed variable Z, to consider different distance ranges simultaneously, to quantify the overall role of space, and to be easily interpretable. A comparative study for different urban configurations is developed in what follows, in order to verify the ability of the set of indices to detect sprawl.

3 Spatial entropy measures on simulated urban settings

The flexibility and informativity of the spatial entropy indices discussed in Section 2 are assessed with a comparative study, which helps in understanding the differences between the two approaches over three urban configuration archetypes. Following Tsai (2005), they are identified as 'monocentric city', 'polycentric city' and 'decentralized city'. The monocentric city, with one compact centre and surrounding peripheral areas, is considered the most positive situation as regards the urban pattern; the polycentric city, occurring when different city centres merge into one greater urban area, is an intermediate, less compact, situation which may suffer from sprawl; the decentralized configuration, denoted by urban scattering, is concerned by the sprawl issue. An example of the three settings is shown in Figure 1. The three archetypes are theoretical reference scenarios in the literature and correspond to a low, medium and high indication of sprawl, respectively; they can be used for comparison to real situations, for a better understanding of the type of urban expansion under study.

Insert Figure 1 about here

In this simulation study, the observation window is a square of size 100 km^2 and represents a city's metropolitan area. In order to simulate cities, the window is firstly partitioned into districts. Since the measures of Section 2.1 and 2.2 may be affected by the partition, we check two partition options. Firstly, the window is partitioned into $G_1 = 20$ districts of different size, by randomly generating 20 centroids over the area following a homogeneous Poisson process, which is known to produce uniformly distributed points over an area (Baddeley et al., 2015), and then performing a Dirichlet tessellation, i.e. assigning each part of the window to the district with the closest centroid. The second option is to partition the window into concentric districts, which can give a better idea of a city expansion into surrounding areas. We choose $G_2 = 5$ annuli, defined by concentric rings, with the same width, i.e. the same difference between the radius of the outer ring and the one of the inner ring. The annuli center is the observation window centroid, and their width is chosen so that they cover the whole window. The two options are shown in Figure 2 for a monocentric dataset.

Insert Figure 2 about here

The second step consists in generating urban patches over the districts. The window is gridded by 40×40 pixels, so that each pixel is 250 metres wide, the same resolution as the data used in Section 4. The binary variable X occurs over the pixels, with categories $x_1 =$ urban and $x_0 =$ non-urban. Consequently, for the measures of Section 2.1 and 2.2 the phenomenon F regards the presence of x_1, while for the approach of Section 2.3 the transformed variable Z has 3 categories: $z_1 = \{$urban, urban$\}$, $z_2 = \{$urban, non-urban$\}$, $z_3 = \{$non-urban, non-urban$\}$. The three urban configurations are generated exploiting the theory of point processes. The monocentric and polycentric scenarios are generated from the intensity function of a Thomas process (Baddeley et al., 2015), i.e. a Poisson cluster point process, with one cluster for the monocentric case and four clusters for the polycentric case. Such process is known to produce spatially correlated events, and is analogous to a spatial autoregressive model with a smoothly decreasing intensity function (Baddeley et al., 2015). The decentralized pattern is generated following the intensity function of a homogeneous Poisson process. For the three urban scenarios, 1000 datasets are simulated. Then, the point patterns are turned into raster data, where a pixel gets value x_1 (urban) if it contains at least one point. For each of the 1000 realizations, the number of urban and non-urban pixels is the same across the three scenarios. This way, Shannon's entropy would not be able to distinguish among the configurations, while we check how the measures of Section 2 succeed in detecting sprawl.

The objective of the simulation study is to verify that all measures presented in this work, though based on different distance concepts, incorporate spatial information and are able to distinguish among the three theoretical urban configurations. In particular, the distinction should be neat between the first two scenarios, which denote an acceptable urban development, and the decentralized pattern, which corresponds to a serious degree of sprawl. The study aims at showing that they are a substantial improvement with regard to non-spatial entropy measures, still commonly used as landscape and urban expansion metrics. Results are concerned with the type of urban development, rather than with its spatial extension. Therefore, they hold for any city, irrespective of its size, as long as its configuration can be compared to one of the archetypical scenarios.

3.1 Batty's and Karlström and Ceccato's entropies

For both partition options, probabilities p_g of (3) and (5) are estimated in each of the 1000 simulations as the proportions of urban pixels over the districts. Batty's entropy for the three scenarios and the two partition options is shown in the boxplots of Figure 3. Results are presented in relative terms, i.e. all entropies are divided by their maximum $\log(100)$ in order to vary between 0 and 1 and enable comparison.

Insert Figure 3 about here

The index performance with the concentric partition is outstanding: the measure is able to make a neat distinction among the three urban configurations as regards spatial entropy. Results are nearly as good for the random partition, with only a partial overlapping between the values of the monocentric and the polycentric scenario. The most important distinction is between the first two configurations and the decentralized pattern, which denotes sprawl; under this perspective, Batty's entropy proves to be effective in detecting a seriously sprawled development, irrespective of the district partition.

For Karlström and Ceccato's entropy, different possibilities for the neighbourhood distances between the districts' centroids are considered, in order to quantify $I(\widetilde{p}_g)$. For the random partition option, three neighbourhoods are set, up to the 5th percentile, first quartile and median of the distribution of distances among the $G_1 = 20$ centroids; they are equal to $nd_{11} = 1473$, $nd_{12} = 3654$ and $nd_{13} = 5335$ metres. For the concentric option, four neighbourhoods are possible over the 5 annuli, i.e. up to the jth farthest district, $j = 1, \ldots, 4$. We name them $nd_{21} = 1$Ann, $nd_{22} = 2$Ann, $nd_{23} = 3$Ann and $nd_{24} = 4$Ann, where jAnn means 'up to the jth farthest annulus'. The estimates of $\widetilde{p}_g$ in (5) are computed as averages of the neighbouring estimated probabilities. Results for Karlström and Ceccato's entropy are shown in Figure 4, in relative terms, for the three urban configurations, the two partition options and all neighbourhoods.

Insert Figure 4 about here

The inclusion of a neighbourhood system does not look helpful in the context of urban sprawl. Neighbourhood nd_3 of option 1 leads to a general overlap in the results, showing that it is not possible to distinguish among the three configurations when the neighbourhood is too wide. As for the other panels of Figure 4, the measure is still able to achieve the most important result, i.e. distinguishing the first two urban patterns from the decentralized one. The extreme values obtained from a decentralized, sprawled scenarios are never reached by the other two configurations, irrespective of the partition and neighbourhood extent.

While Karlström and Ceccato's extension to Batty's entropy is interesting from a theoretical point of view because of the LISA-type properties, it does not seem to provide major advantages in practical situations, especially due to the risk of overlapping between the values resulting from the polycentric and the decentralized scenario, when the partition is not concentric. The original formulation of Batty's entropy is well performing on simulated data, and is also used for the case study of Section 4. The neat distinction among positive and negative scenarios in Figure 3 allows to provide intervals for an indication of the degree of sprawl in cities; intervals must not be intended as absolute benchmarks, rather they help in evaluating the seriousness of the sprawl phenomenon in a city according to its spatial distribution. The interval $[0, 0.847]$ refers to a monocentric scenario and suggests a positive expansion; the interval $[0.848, 0.985]$ indicates a polycentric configuration and possibly an intermediate level of sprawl. The interval $[0.985, 1]$ suggests a high level of sprawl.

3.2 Spatial mutual information and residual entropy

For the computation of the entropy set of Section 2.3, breaks for the distance ranges must be chosen, where the distance concerns pairs of pixels, not districts as in Section 2.2, and is measured between pixel centroids. Two options are considered in the simulation study, where it should be remembered that the global values of spatial mutual information and residual entropy are not affected by the choice of the w_k. The first one is motivated by the tradition of spatial statistics, where the so called '4 nearest neighbour system' (i.e. pixels sharing a border) and the analogous '12 nearest neighbours system' are of standard use (Anselin, 1995). Accordingly, the first two distance classes chosen for option 1 (in metres) are $w_{11} = [0, 250]$ and $w_{12} =]250, 500]$, where

250 metres is the distance between contiguous pixels' centroids; the remaining classes are $w_{13} =]500, 1250]$, i.e. up to 5 pixels along the cardinal directions, and $w_{14} =]1250, d_{max}]$, $d_{max} = 13789$ metres being the maximum distance between pixels within the observation window. In the measurement of urban sprawl, the focus is on what happens at small distance ranges, where a lack of spatial association, i.e. a high presence of pairs of type {urban, non-urban}, indicates dispersion, thus sprawl. Therefore, detailed results are needed for small distances, while aggregate results are enough at large distances. The second option follows the same criterion as the neighbourhood distance choice in Section 3.1: the empirical distribution of pixel distances (in metres) is computed, and the breaks are chosen as the 5th, 25th and 50th percentile, resulting in classes $w_{21} = [0, 1346]$, $w_{22} =]1346, 3260]$, $w_{23} =]3260, 5130]$, $w_{24} =]5130, 13789]$. Each specific adjacency matrix A_k identifies pairs of pixels at a distance that belongs to class w_k. The rule of moving rightward and downward is adopted along the observation window in order to identify neighbouring pairs avoiding double counting. The number of pixel pairs contained within each distance class does not depend on the simulated data or configuration, but only on the choice of the grid and on w_k, and is shown in Table 1: the great amount of data involved in the computation einforces the validity of the results. Then, each $p_{Z|w_k}$ is estimated using proportions for the three categories of Z at the corresponding distance range.

Insert Table 1 about here

Since the main focus of this work is on the contribution of the partial terms, rather than on the global value, spatial partial information terms are shown, in relative terms, in Figure 5 for the two distance class options.

Insert Figure 5 about here

Spatial mutual information can be interpreted as a sprawl detector: a high mutual information value implies positive association among urban pixels and positive association among non-urban ones, and indicates a compact urban expansion. The measure successfully distinguishes the first two spatial patterns (mono- and polycentric) from the decentralized one in all cases, except at very large distances (w_{14}) where the lack of distinction among patterns is expected and is of scarce interest in sprawl studies. No mutual information is detected at any distance over the decentralized patterns, where no spatial structure is present and space does not help in explaining the data behaviour, irrespective of the choice of the distance breaks. Results are not shown for spatial partial residual entropy terms, as the interpretation is symmetrical to that of spatial mutual information.

As for Batty's entropy, the boxplots in Figure 5 can be used as a reference for assessing real case studies, since no overlap occurs between an acceptable and a sprawled situation. According to the simulation study, relative values of spatial mutual information in $[0, 0.001]$ are an indication of a decentralized pattern.

4 Measuring urban sprawl in Europe via spatial entropy

The case study comes from official European sources. Land use data for the European territory are provided by CORINE project (COoRdination of INformation on the Environment, EEA, 2011), which integrates remote sensing images and photo interpretation to produce a pixel grid with a 250×250 metres resolution; pixels are classified according to 44 land use classes. Guidelines are then provided to dichotomize the dataset into urban and non-urban pixels, transforming land use data in Urban Morphological Zone (UMZ) data. An Urban Morphological Zone can be defined as ‘a set of urban areas laying less than 200m apart’[2] (EEA, 2011). Urban classes are selected in a way, e.g. by including green urban areas, that avoids misinterpretation in the phenomenon of urban scattering, which might otherwise be overestimated. The nature of UMZ data makes them appropriate to identify shapes and patterns of urban areas, and thus to detect urban sprawl. Polygonal maps with administrative boundaries are superimposed over the European dataset for selecting the areas of interest. We select data for years 1990 and 2012, the first and last release of CORINE’s dataset, for three cities in different European countries with their commuting belts, i.e. an extension of the urban centre beyond the administrative city boundaries, which includes the municipalities surrounding the main city. The three cities have a similar population and spatial extension, and are chosen based on results in EEA and FOEN (2016): we focus on the *DIS* index, ‘DISpersion of built-up areas’, which characterises the settlement pattern according to a geometric perspective. The first city is Eindhoven, the Netherlands, chosen because the country is classified among the highly sprawled ones. The second city is Lublin, in Poland, one of the countries below the average European sprawl level. The third one is Bologna, Italy, a country with an average level of sprawl. From 1990 to 2012, all cities increase the fraction of urbanized pixel: Eindhoven from 18% to 25%, Lublin from 9% to 16% and Bologna from 16% to 18%. A total of six binary raster datasets (3 cities at 2 time points) is considered and displayed in Figure 6.

Insert Figure 6 about here

The objective of the present study is to show how the proposed entropy measures can be used in an absolute way to obtain an indication of the level of uban sprawl in the three cities, or in a relative

[2]The Corine Land Cover classes used to build the Urban Morphological Zone dataset are ‘Continuous urban fabric’, ‘Discontinuous urban fabric’, ‘Industrial or commercial units’, ‘Green urban areas’. ‘Port areas’, ‘Airports’, ‘Road and rail networks’ and ‘Sport and leisure facilities’ are also considered if they are neighbours to the core classes.

way for comparison across cities and for evaluating the phenomenon evolution over time. This is done with both approaches presented in Section 2 and with the support of the results of Section 3. The focus of the results is on the ability of the proposed measures to capture the phenomenon of urban sprawl in an interpretable and comparable way: even though the two approaches to spatial entropy are based on different concepts of neighbourhood and even if distance is measured between different types of spatial units, the advantages of both can be highlighted. For the three cities, we expect to find consistent results with regard to the country level analysis of EEA and FOEN (2016). We follow the usual approach in the literature to treat the indices as descriptive measures, without relying on their unknown probability distribution for inference and testing. The remainder of the present Section outlines the main findings; a more comprehensive discussion can be found in Section 5.

4.1 Batty's spatial entropy

Each city with its commuting belt is partitioned according to the administrative boundaries of the municipalities, the G districts for Batty's entropy. In order to compare results, values are divided by their maxima, i.e. the logarithm of each city window's size.

Insert Table 2 about here

Results in Table 2 show that Batty's entropy values confirm the EEA country level sprawl ranking: Lublin is the less sprawled one, the highest indication of sprawl is detected for Eindhoven and Bologna constitutes an intermediate case. In addition, the result for Eindhoven suggests that the city may suffer from a high level of sprawl, following the reference set of intervals of Section 3.1: Eindhoven's entropy values are greater than 0.985, that was found as the lowest value of the decentralized configuration range in the simulation study. The comparison of results over time is also consistent with the EEA and FOEN report: the issue of urban sprawl tends to increase for all cities, especially for Lublin.

The second partition option introduced for the simulation study in Section 3.1, with concentric areas defined by annuli centered in the centroid of each main city, has been also checked and leads to the same results.

4.2 Spatial mutual information and residual entropy

Spatial partial mutual information and partial residual entropy terms are computed following the first distance option of Section 3.2: w_1 and w_2 are the 4 and 12 nearest neighbour systems respectively, w_3 considers up to 5 pixels along the cardinal directions, w_4 captures all greater distances. The number of observation pairs contained within each distance class is shown, for each city, in Table 3; the huge amount of data used for computations enhances the reliablity of the results. All distances refer to pairs of pixels and are measured, in metres, between pixel centroids.

Insert Table 3 about here

Insert Figure 7 about here

Results are summarized in Figure 7, which displays the values of partial spatial mutual information $PI(Z|w_k)$ and partial residual entropies $H(Z|w_k)$. To allow for space and time comparisons, their proportional versions are computed by setting the sum $PI(Z|w_k) + H(Z|w_k)$ to 1 at each distance class w_k. The increase of sprawl over time is evident in these results, while the ranking of cities in terms of urban sprawl is more evident in 1990 than in 2012, again aligning with the EEA country results: Eindhoven has a low proportion of spatial information at all distances, identifying a high sprawl level; Lublin is the least sprawled one, with the highest values of partial spatial information terms. By considering the reference set of values identified in Section 3.2, none of the cities is classified, in an absolute way, as decentralized; this suggests a less extreme evaluation than Batty's entropy results, and may be more sensible in this context. Indeed, Eindhoven, though more sprawled than the other two, looks like a polycentric city according to Figure 6. Extreme decentralized values are more likely to be found in really chaotic cities, e.g. Calcutta (see Bhatta et al., 2010).

The most informative distance classes for detecting urban sprawl are the small ones. A second option, employing the 5th, 25th and 50th percentile of the empirical distribution of distances for each city to choose the breaks d_1 to d_3 of the distance classes, proves to be useless for detecting and comparing the urban sprawl of the three cities over space and time. The distance classes are too broad, and the partial terms of spatial mutual information are all very low.

5 Discussion and concluding remarks

In this work, the approaches proposed by Batty (1976), Karlström and Ceccato (2002) and Altieri et al. (2018a) are employed to quantify the level of urban sprawl, i.e. the chaotic expansion of cities, and their properties are assessed with a comparative study. The first objective is to show that they constitute a major step forward in the use of entropy measures as landscape metrics; another objective is to make a comparative evaluation of the theoretical and practical properties of such indices; lastly, we aim at providing indications about the level of sprawl of three European cities.

The first objective is fully met by the results from both simulated and real data: while Shannon's entropy would not be able to distinguish among scenarios with similar proportions of urban areas, both approaches successfully separate a clustered situation from a sprawled one.

The second objective may be summarized by a few main comments. Batty's entropy is a fundamental step toward spatial entropy measures, but it suffers from a few theoretical drawbacks: it requires a dichotomous variable and is affected by the choice of the window partition. Nevertheless, in the context of urban sprawl such measure proves to be very efficient, since it neatly distinguishes among differently sprawled scenarios and returns coherent results with respect to EEA and FOEN (2016) as regards European cities. Karlström and Ceccato's approach represents an interesting proposal from a theoretical point of view, exploiting the idea of neighbourhood and enjoying LISA-type properties; however, it suffers from the same drawbacks as Batty's entropy, moreover it focuses on a single, predefined neighbourhood system. In the simulation study, the addition of a neighbourhood system generates overlapping among the results from different urban scenarios, therefore this approach is discarded in the case study. Spatial partial mutual information is the key component of Altieri et al.'s entropy for drawing conclusions on sprawl. Its advantages lie in the possibility of managing variables with any number of categories, e.g. more than two land use classes, decomposing the entropy due to space from the one due to other sources of heterogeneity, investigating the global values and the partial terms jointly, to identify the role of space at different distance ranges. Beyond enjoying such theoretical properties, spatial mutual information proves to be effective in measuring urban sprawl and distinguishing among clustered and decentralized scenarios. The set of measures proposed by Altieri et al. is more exhaustive than Batty's and Karlström and Ceccato's indices, since it allows to quantify the contribution of the partial terms. The choice of the distance classes in this proposal does not affect the global result, unlike the choice of Batty's window partition. The last advantage is that results in the case study of Section 4 are less extreme than Batty's results, and more realistic for a set of European cities.

As for the last objective, some conclusive points can be stated about the case study of Eindhoven, Lublin and Bologna. The EEA country ranking in terms of dispersion of built up areas is reproduced here at city level: Lublin is the least sprawled one, Bologna has an intermediate level of sprawl and Eindhoven is the most sprawled one. The situation of Eindhoven seems very critical: according to Batty's entropy, its values belong to the range of the decentralized pattern identified by the simulation study, while spatial mutual information classifies it as the worst situation, though not extreme. All cities become more affected by sprawl over time, denoting an inefficient urban expansion from 1990 to 2012.

Results from both simulation and application suggest some general recommendations for working with real data. In the study of urban sprawl, the most interesting distances are the small ones. At this regard, spatial mutual information and spatial residual entropy are very flexible, as they can focus on the most informative distance range to interpret the phenomenon under study. Distance classes must be suitably proposed according to the context. The focus on small distances is not an issue for the set of spatial entropy measures proposed by Altieri et al., as the choice of the classes does not affect the global result; the theoretical framework illustrated in this paper shows that, when the distance classes system is modified, the set of measures can be easily, rapidly and intuitively adapted. As a further point, the data finest available resolution should be used, i.e. points if data are a point pattern, or the finest grid provided if data are lattice. Pixel aggregation is not recommended unless motivated, as it may reduce precision in the results and requires expertise in classifying the new pixel according to land use classes.

The well performing spatial entropy measures considered in this work capture the spatial aspect of the complex phenomenon of dispersed urbanization. As a further development, they can be suitably integrated with other indices in order to obtain a comprehensive quantification of the environmental footprint of a city. This helps in focusing on the worst developed areas and contributes to solve environmental issues such as dangers to ecosystems, forest destruction, pollution and climate change. The present work is also useful in highlighting the abilities of spatial entropy measures, which can be employed in other environmental contexts as well: meteorology, species abundance studies, pollution, earthquakes and fires are just a few examples.

Acknowledgements

This work is developed under the PRIN2015 supported project 'Environmental processes and human activities: capturing their interactions via statistical methods (EPHASTAT)' [grant number 20154X8K23] funded by MIUR (Italian Ministry of Education, University and Scientific Research).

References

Altieri, L., D. Cocchi, G. Pezzi, E. Scott, and M. Ventrucci (2014). Urban sprawl scatterplots for Urban Morphological Zones data. *Ecological Indicators 36*, 315–323.

Altieri, L., D. Cocchi, and G. Roli (2018a). A new approach to spatial entropy measures. *Environmental and Ecological Statistics 25*, 95–110.

Altieri, L., D. Cocchi, and G. Roli (2018b). *SpatEntropy: Spatial Entropy Measures*. R package version 0.1.0.

Anselin, L. (1995). Local indicators of spatial association - LISA. *Geographical Analysis 27*, 94–115.

Baddeley, A., E. Rubak, and R. Turner (2015). *Spatial Point Patterns: Methodology and Applications with R*. London: Chapman and Hall/CRC Press.

Bart, I. (2010). Urban sprawl and climate change: a statistical exploration of cause and effect, with policy options for the EU. *Land Use Policy 27*, 283–292.

Batty, M. (1974). Spatial entropy. *Geographical Analysis 6*, 1–31.

Batty, M. (1976). Entropy in spatial aggregation. *Geographical Analysis 8*, 1–21.

Batty, M. (2010). Space, scale, and scaling in entropy maximizing. *Geographical Analysis 42*, 395–421.

Benfield, F., M. Raimi, and D. Chen (1999). Once there were greenfields: how urban sprawl is undermining Americas environment, economy and social fabric. Technical report. Natural Resources Defense Council.

Bhatta, B., S. Saraswati, and D. Bandyopadhyay (2010). Urban sprawl measurement from remote sensing data. *Applied Geography 30*, 731–740.

Bivand, R. S., E. Pebesma, and V. Gomez-Rubio (2013). *Applied spatial data analysis with R. Second edition.* New York: Springer.

Bondy, J. A. and U. S. R. Murty (2008). *Graph Theory*. Springer.

Cabral, P., G. Augusto, M. Tewolde, and Y. Araya (2013). Entropy in urban systems. *Entropy 15*, 5223–5236.

Chong, C. H.-S. (2017). *Comparison of Spatial Data Types for Urban Sprawl Analysis Using Shannon's Entropy*. University of Southern California: Dissertation.

Couch, C., L. Leontidu, and G. Petschel-Held (2007). *Urban Sprawl in Europe. Landscapes, Land Use Change & Policy*. Oxford, Malden, MA: Wiley-Blackwell.

Cover, T. M. and J. A. Thomas (2006). *Elements of Information Theory. Second Edition*. Hoboken, New Jersey: John Wiley & Sons, Inc.

EEA (2006). Urban sprawl in Europe - the ignored challenge. Technical report. EEA Report No 10/2006.

EEA (2011). Corine land cover 2000 raster data. Technical report. Downloadable at http://www.eea.europa.eu/data-and-maps/ data/corine-land-cover-2000-raster-1.

EEA and FOEN (2016). Urban sprawl in Europe - joint EEA-FOEN report. Technical report. EEA Report No 11/2016.

Ewing, R. (2008). *Characteristics, Causes, and Effects of Sprawl: A Literature Review*. Springer, Boston: In J.M. Marzluff, E. Shulenberger, W. Endlicher, et al. (eds) Urban ecology: an international perspective on the interaction of humans and nature, 519-535.

Ewing, R. and S. Hamidi (2015). Compactness versus sprawl. A review of recent evidence from the United States. *Journal of Planning Literature 30*, 413–432.

Jaeger, J. A. G., R. Bertiller, C. Schwick, and F. Kienast (2010). Suitability criteria for measures of urban sprawl. *Ecological Indicators 28*, 427–441.

Jiang, F., S. Liu, H. Yuan, and Q. Zhang (2007). Measuring urban sprawl in Beijing with geospatial indices. *Journal of Geographical Sciences 17*, 469–478.

Johnson, M. P. (2001). Environmental impacts of urban sprawl: a survey of the literature and a proposed research agenda. *Environmental and Planning A 33*, 717–735.

Karlström, A. and V. Ceccato (2002). A new information theoretical measure of global and local spatial association. *The Review of Regional Research (Jahrbuch Für Regionalwissenschaft) 22*, 13–40.

Leibovici, D. G. (2009). *Defining spatial entropy from multivariate distributions of co-occurrences*. Berlin, Springer: In K. S. Hornsby et al. (eds.): COSIT 2009, Lecture Notes in Computer Science 5756, 392-404.

Leibovici, D. G., C. Claramunt, D. LeGuyader, and D. Brosset (2014). Local and global spatio-temporal entropy indices based on distance ratios and co-occurrences distributions. *International Journal of Geographical Information Science 28*, 1061–1084.

Li, H. and J. F. Reynolds (1993). A new contagion index to quantify spatial patterns of landscapes. *Landscape Ecology 8*, 155–162.

Liu, Y. and K. Chen (2018). An information entropy-based sensitivity analysis of radar sensing of rough surface. *Remote Sensing 10*.

O'Neill, R. V., J. R. Krummel, R. H. Gardner, G. Sugihara, B. Jackson, D. L. DeAngelis, B. T. Milne, M. G. Turner, B. Zygmunt, S. W. Christensen, V. H. Dale, and R. L. Graham (1988). Indices of landscape pattern. *Landscape Ecology 1*, 153–162.

Oueslati, W., S. Alvanides, and G. Garrod (2015). Determinants of urban sprawl in European cities. *Urban studies 52*, 1594–1614.

Paninski, L. (2003). Estimation of entropy and mutual information. *Journal of Neural Computation 15*, 1191–1253.

R Core Team (2017). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Rényi, A. (1961). *On Measures of Entropy and Information.* University of California Press, pp 547-561: Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability.

Rosni, N. and N. Noor (2016). A review of literature on urban sprawl: assessment of factors and causes. *Journal of Architecture, Planning & Construction Management 6*, 12–35.

Stone, B. J. (2012). *The city and the changing climate: Climate change in the places we live*. Cambridge, MA, USA: Cambridge University Press.

Theil, H. (1972). *Statistical Decomposition Analysis*. Amsterdam: North Holland.

Torrens, P. (2008). A toolkit for measuring sprawl. *Applied Spatial Analysis and Policy 1*, 5–36.

Tsai, Y. (2005). Quantifying urban form: compactness versus sprawl. *Urban Studies 42*, 141–161.

Yeh, A. and X. Li (2001). Measurement and monitoring of urban sprawl in a rapidly growing region using entropy. *Photogrammetric Engineering and Remote Sensing 67*, 83–90.

Figure 1: Examples of the 3 archetypical urban scenarios: monocentric, polycentric, decentralized.

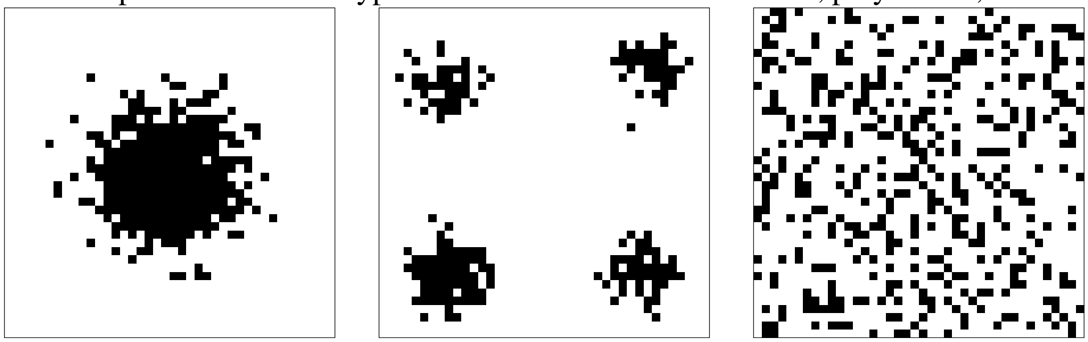


Figure 2: Two options for area partition in Batty's and Karlström and Ceccato's entropy over an example of monocentric dataset. Left panel: 20 random areas; right panel: 5 concentric rings.

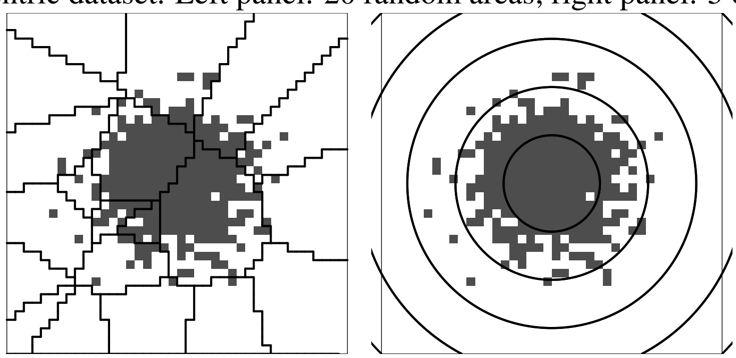


Table 1: Number of pairs in each distance class for the two distance options on simulated data

Distance option 1				Distance option 2			
w_1	w_2	w_3	w_4	w_1	w_2	w_3	w_4
3120	6082	48006	1221992	67988	253026	319624	638562

Figure 3: Results for Batty's entropy over the three urban scenarios, 1000 simulations, with the two partition options: 20 random areas (left panel), 5 concentric rings (right panel).

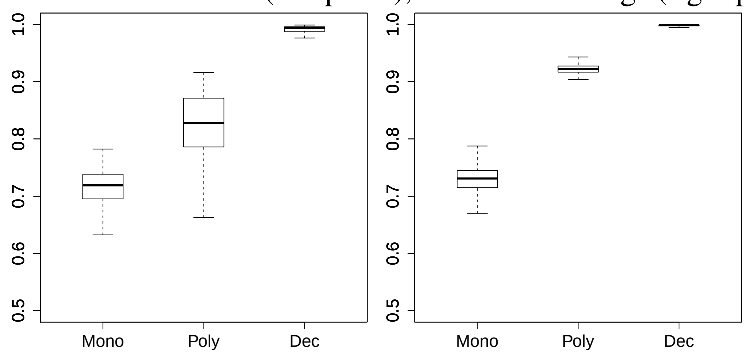


Figure 4: Results for Karlström and Ceccato's entropy over the three urban scenarios, 1000 simulations, with the two partition options: 20 random areas (higher panels), 5 concentric rings (lower panels), at different neighbourhood distances.

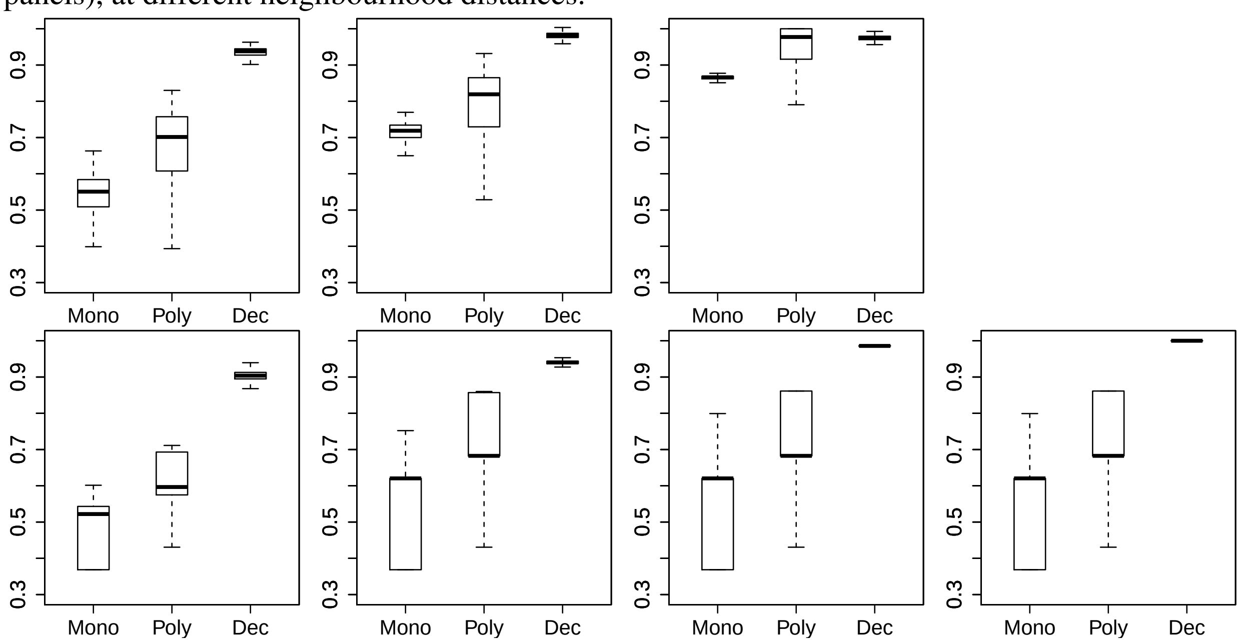


Figure 5: Spatial partial information for the three urban scenarios, 1000 simulations. First option for the distance ranges in the higher panels, second option in the lower panels.

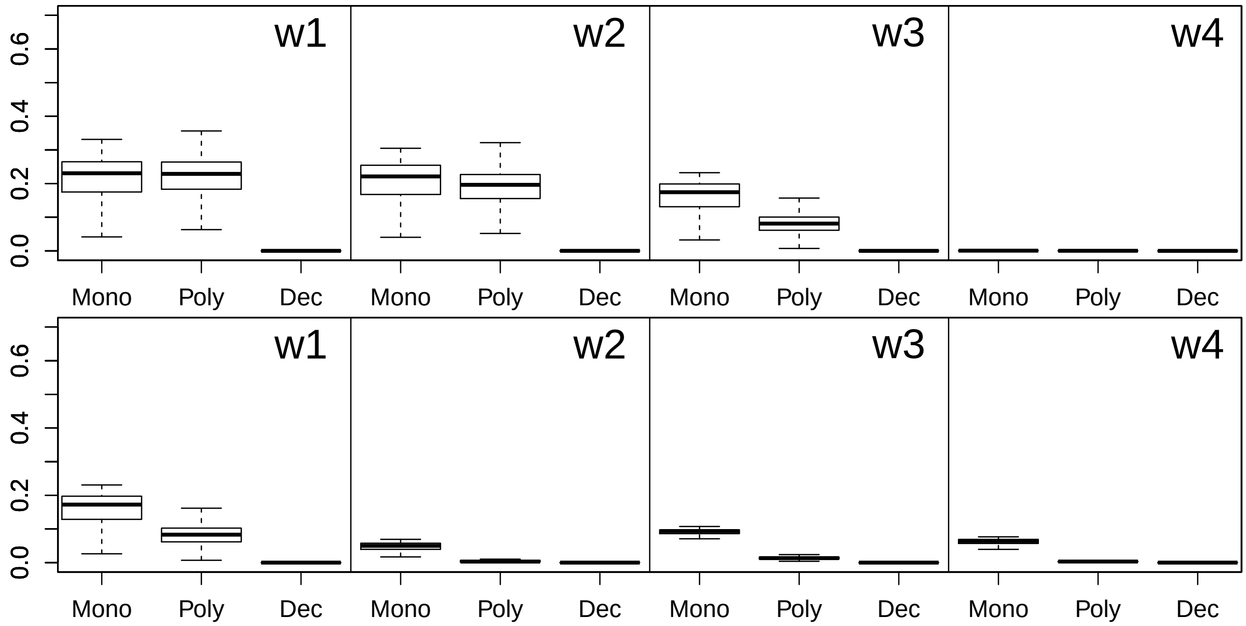


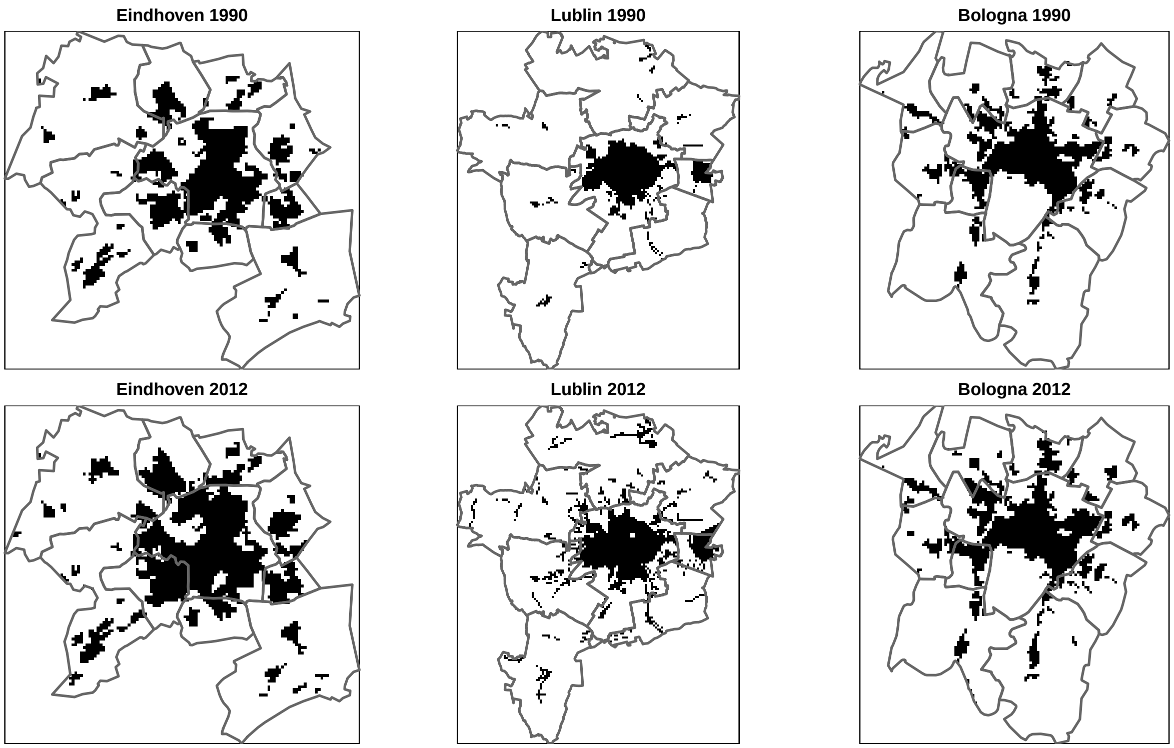
Table 2: Results for Batty's and Karlström and Ceccato's (KC) entropy for the three cities

	1990	2012
Eindhoven	0.987	0.990
Lublin	0.955	0.978
Bologna	0.980	0.983

Table 3: Number of pairs contained within each distance class for Eindhoven, Lublin and Bologna

	w_1	w_2	w_3	w_4
Eindhoven	17076	33657	274784	37732933
Lublin	23839	47118	387726	73249328
Bologna	19634	38780	318491	49618095

Figure 6: From left to right, Eindhoven, Lublin and Bologna together with their commuting belts, in 1990 (higher panels) and 2012 (lower panels).



Commuting belt for Eindhoven: Best, Eersel, Geldrop, Heeze-Leende, Nuenen, Oirschot, Son en Breugel, Veldhoven and Waalre.

Commuting belt for Lublin: Głusk, Jastków, Konopnica, Niedrzwica Duża, Niemce, Świdnik and Wólka.

Commuting belt for Bologna: Anzola dell'Emilia, Calderara di Reno, Casalecchio di Reno, Castel Maggiore, Castenaso, Granarolo dell'Emilia, Pianoro, San Lazzaro di Savena, Sasso Marconi, Zola Predosa.

Figure 7: Proportional partial spatial mutual information (grey area) and residual entropy (white area). From left to right, Eindhoven, Lublin and Bologna in 1990 (higher panels) and 2012 (lower panels).