Graz Economics Papers – GEP

GEP 2020–02

Measuring Urban Sprawl using Land Use Data

Miriam Steurer and Caroline Bayr

February 2020

Department of Economics
Department of Public Economics
University of Graz

Measuring Urban Sprawl using Land Use Data

Miriam Steurer*, Caroline Bayr°

December 31, 2019

* Institute of Public Economics, University of Graz, Austria, miriam.steurer@uni-graz.at

° Joanneum Research, Policies, Graz, Austria, caroline.bayr@joanneum.at

Abstract: Digital land use data, generally derived by remote sensing and often organized in grid form, have become widely available for even the most remote areas of the globe. Here we investigate how to use land use data to measure three of the most characteristic aspects of urban sprawl: low density, low continuity of land use type (scatteredness), and low compactness of the shape of the city. For each of these categories we present multiple urban sprawl indicators. Some of these indicators have been used in the literature before, others we developed ourselves. For density measurements we illustrate how simple changes to common density indicators can improve their meaningfulness. With respect to scatteredness we show that the interpretation of entropy measures can be ambiguous. Using a variant on Moran's I index does a better job at measuring scatteredness. When it comes to measuring compactness, the grid structure of land use data can inflate the boundary of the measured area. We introduce new compactness indices that correct for this problem. To illustrate the discussed indices, we apply them to Graz, the second largest city in Austria, using data from the CORINE Land Cover (CLC) Project [1].

JEL: R11, R52

Keywords: surban sprawl, density, entropy, GIS, remote sensing, urban dynamics, spatial analysis, compactness

Note: An earlier version of this paper previously appeared as Graz Economics Paper **GEP 2015-06.**

1. Introduction

Measuring urban sprawl is important because it is associated with a variety of major ecological, social, and health effects and developments. The impacts associated with urban sprawl range from the lack of scale economies [2], ecological problems such as air pollution, traffic congestion, and water shortages [3-5], overburdening of amenities [6], fragmentation of the eco-system, loss of agricultural land, social problems of increased segregation [7], increased isolation [8], health problems such as obesity [e.g. 9-11], to increased public spending by local governments [12,13].

Urban sprawl is however a complex phenomenon and it is therefore not surprising that many different definitions of urban sprawl exist in the literature. Frenkel and Ashkenazi [2] summed up the situation well when they wrote: "we know that sprawl is significant, but we are not yet sure what it is exactly or how to measure it." The attributes that are used to describe urban sprawl range from: low density housing (especially at the urban fringe), low diversity, sprawling housing developments, poor accessibility, to aesthetic criteria such as boring architecture. However, the three categories that are most often used to describe urban form are density, diversity, and spatial-structural pattern [14]. In this paper we focus on how to measure these three characteristics of urban sprawl using land use data.

Land use data are typically available in digital code that refers to individual cells. Each cell is allocated one special land use type that indicates the dominant use of this plot of land. Our land use data comes from the Corine Land Cover Project [1]. This dataset is freely available for all European countries (and also Turkey, Madeira, and the Canary Islands). The grid size in the Corine Land Cover 1Project is 100 by 100 meters and there are 42 land use classes present in the CLC Project. Population is mainly found in CLC class 1 (continuous urban fabric) and 2 (discontinuous urban fabric).

The most prominent indicator used in the literature to describe urban sprawl is density. In section 2 we illustrate how traditional density measures can be improved by making use of the classifications of land use data. Rather than just using population density per area, land use data can be used to fine-tune the area definition by "cutting out" areas that are (for legal or other reasons) not used for housing. The resulting density measure is a more meaningful indicator of urban sprawl and also allows for comparisons across cities.

Section 3 discusses diversity-based urban sprawl indices. Entropy indices are the main measure of this type, but other indices like the Moran's I index provide alternatives to capture diversity or continuity across space. Entropy indices are often used in the literature to measure the diversity of population and land use patterns across an area [15-21]. We explore two different variants of Shannon's entropy index and find them difficult to interpret. The difficulty arises because entropy indices measure two components within one index: On the one hand the prevalence of one cell type compared to another, on the other hand how scattered the cell types are within the chosen area. To get around this problem we therefore turn to an alternative measure of diversity, the (global) Moran's I index. The Moran's I index is a spatial autocorrelation index and measures the similarity of neighboring cells. We argue that it is better suited to capture the discontinuity aspect of urban sprawl than entropy measures.

In section 4 we investigate which spatial-structural indices are well suited to measure the compactness of an urban settlement. We introduce two new indices into the urban sprawl literature that take account of the grid structure of land use data. Using conventional geometrical compactness measures with digital land use data would lead to overestimation of circumference measures and thus to an overestimation of sprawl. Our versions of Milne's [22] shape index and Bribiesca's [23] compactness index correct for this problem.

In section 5 we illustrate the workings of these indices by applying each of them to the city of Graz, the second biggest city in Austria using data from the Corine Landcover Project [1]. The paper concludes with a discussion of the relative advantages and disadvantages of each urban sprawl indicator.

2. Density Measures

2.1. Density Measure 1: number of people divided by total area of district/municipality

The most popular methods to measure urban sprawl focus on population density. Low population density numbers are suggestive of high degrees of urban sprawl. We can see large differences in the population density for major international cities. The 6,299 square kilometer city area of Los Angeles has a population of 14,900,000, which corresponds to a population density of 2,400 people per square kilometer. For the 2,163 square kilometer area of Seoul-Incheon in South Korea with a population of 22,547,000 people, the population density is 10,400 people per square kilometer [24]. The standard version of the density measure is to divide the number of people by the total area of a district/municipality ($\boldsymbol{D_1}$), where the size of the district is generally given by the political boundaries of the settlement.

$$D_1 = \frac{\sum_i \sum_j pop_{ij}}{\sum_i \sum_j x_{ij}}$$

where i indexes the cells in district j, pop_{ij} is the population of cell i in district j, and x_{ij} is the square kilometer area of cell i in district j. For the case of 100 by 100 meter grids, for all cells ij, x_{ij} equals 0.01 square kilometers.

2.2. Density Measure 2: number of people divided by built-up cells in district/municipality

One problem with $\boldsymbol{D_1}$ is that it is very sensitive to the area included in the measurement. Most often this measurement is performed using the political (administrative) boundary of a city. However, this is an artificial concept and can differ substantially from one town to another. Sometimes the outer suburbs and rural areas are included within a political city boundary, while other times they are not. Consistency of what constitutes an urban area is needed. Sadly, lack of consistency in where urban boundaries are drawn is the norm. The consequence is over- or underestimation of urban land use and a lack of comparability across or within nations [25, 26]. Remote sensing data allows for a more uniform definition of functional urban areas instead of arbitrary political boundaries.

Another related problem with the $\boldsymbol{D_1}$ measure is that it does not take account of different land uses within a city area. For example, a city that consists of dense urban spaces and a large park can have the same density measure as one that consists of a sprawling urban landscape without any large green spaces. To counteract these shortcomings, we propose a density measure which considers the population density in built-up areas only ($\boldsymbol{D_2}$).

The formula for this density measure reads as follows:

$$D_2 = \frac{\sum_i \sum_j pop_{ij}}{\sum_i \sum_j \sum_k x_{ijk}}$$

where i indexes the cells, j the district, and k indexes the land use types that are counted as residential.

As a population density measure, $\boldsymbol{D_2}$ is more meaningful than $\boldsymbol{D_1}$. Excluding those areas where no-one can live (for example rivers, lakes, high mountains) or where no-one is allowed to live (for example forests, parks, or other areas excluded by zoning restrictions) before calculating population density makes sense and improves the meaningfulness and comparability of the urban density measurements.

3. Diversity measures

3.1. Entropy measures to indicate urban sprawl

The concept of entropy comes from the field of thermodynamics and is generally considered to be a measure of disorder. Entropy indices can be used to measure "disorder" in various areas – apart from thermodynamics; it is used in inequality measurement, diversity measurement, and also for measuring urban sprawl. Entropy is highest when disorder in a system is greatest. As such it seems ideal for the concept of urban sprawl. No one would say that either the Gobi Desert or down-town Manhattan is affected by urban sprawl. So urban sprawl is something that happens in between these extremes. Figure 1 graphically illustrates this point. Because they are steadily increasing as more people move into an area, density measures cannot deal with this aspect but entropy measures can.

Entropy measures of urban sprawl attempt to capture the expansion of urban areas into the countryside. In contrast to density measures, entropy measures focus on differences in density across districts rather than the density level itself. In this way entropy is an indicator of the continuity of land use type, or if there is lack of continuity, the "scatteredness" of land use types throughout the area.

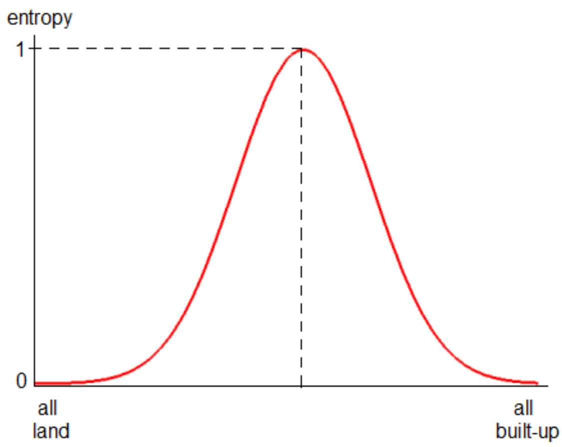

Figure 1: The curve of the entropy measure.

According to [27], continuity refers to the degree to which developable land has been built upon at urban densities in an unbroken fashion. Here we use Shannon's measure of entropy to measure the degree of continuity (or discontinuity) of urban development. A high entropy score indicates that population density varies a lot throughout the area which in turn is interpreted as a high level of urban sprawl. A low entropy score conversely implies low variability in population density across districts and a low level of urban sprawl.

Entropy measures have been used to measure urban sprawl in many locations; e.g. by [15, 16] for areas in China, [18, 19] for India, [17] for Canada. [28] provide a good overview of the use of entropy measures in the urban sprawl literature. One of the main advantages of entropy measures in the urban sprawl context is their simplicity and the ease of integration with GIS based data [15]. These advantages have made it "the most widely used technique to measure the extent of urban sprawl with the integration of remote sensing and GIS"[29]. Another benefit of the entropy measure is its invariance to the size, shape, or number of regions of the area under consideration and it is therefore a quite robust spatial statistic [15, 29].

Most entropy applications to urban sprawl use Shannon's entropy measure (or the Theil index, which is closely related to it). However, any variation of the general entropy index could be used to measure urban sprawl. Other indices of this type – for example the Atkinson index or the Gini coefficient – could be used instead. In this paper we consider two variations of Shannon's entropy measure – one that includes population density measures, and one that does not.

3.2. Measuring entropy at the district/municipality level with population data

Urban sprawl according to Shannon's entropy measure is calculated as follows:

$$E = -\sum_j p_j \cdot \log(p_j),$$

where p_j is the weight for district j (j = 1,… n) and $\sum_j p_j = 1$.

Using the two different concepts of density introduced in the chapter above (***D₁*** and ***D₂***), we can derive two different measures for p_j. Using the density concept of $\boldsymbol{D_1}$ leads to weights equal to:

$$p_j = \frac{\sum_i pop_{ij} \,/\, \sum_j x_{ij}}{\sum_j\left(\sum_i pop_{ij} \,/\, \sum_j x_{ij}\right)}$$

While defining density in the same way as $\boldsymbol{D_2}$ leads to weights equal to:

$$p_j = \frac{\sum_i pop_{ij} \,/\, \sum_j \sum_k x_{ijk}}{\sum_j\left(\sum_i pop_{ij} \,/\, \sum_j \sum_k x_{ijk}\right)}$$

Again, the subscript *i* indexes cells, *j* indexes the districts, and *k* indexes land use types and x_{ijk} denotes the land area of cells *i* of type *k* in district *j*, while pop_{ij} denotes population of cell *i* in district *j*.

The upper bound for *E* is $\log(n)$, where n is the number of districts. E takes the value $\log(n)$ when the population density is the same in all districts. Dividing E by $\log(n)$ provides us with an index that lies between zero and one. This normalized entropy measure makes it easier to compare entropy results between regions.

Using the two different weights for p_j above gives us two different measures of Shannon's entropy ($\boldsymbol{E_1}$and $\boldsymbol{E_2}$).

$$E_1 = \frac{-\sum_j p_j \cdot \log(p_j)}{\log(n)} \qquad \text{with} \quad p_j = \frac{\sum_i pop_{ij}/\sum_j x_{ij}}{\sum_j(\sum_i pop_{ij}/\sum_j x_{ij})}$$

$$E_2 = \frac{-\sum_j p_j \cdot \log(p_j)}{\log(n)} \qquad \text{with} \quad p_j = \frac{\sum_i pop_{ij}/\sum_j \sum_k x_{ijk}}{\sum_j(\sum_i pop_{ij}/\sum_j \sum_k x_{ijk})}$$

3.3. Entropy measures that do not use population data

The Shannon entropy measure can also be applied to density measures that are not based on population data. An advantage of an approach that does not use population data is that it is readily applicable in areas for which no reliable population data exist. Gathering detailed population data is cost intensive, in contrast gathering land use data is generally very cost effective. Also, as land use data can be gathered remotely they are available for even the remotest parts of the world and areas for which no reliable population data exist.

One example of such a density measure for a district is the number of built-up cells divided by the total number of cells in the district.[1] Applying this definition to the above index we get an entropy measure of the following form:

$E_3 = -\sum_j p_j \cdot \log(p_j)/\log(n)$,

where

$$p_j = \frac{\sum_i \sum_k x_{ijk}/\sum_i x_{ij}}{\sum_j(\sum_i \sum_k x_{ijk}/\sum_i x_{ij})}$$

are the weights summing to one.

$\sum_i \sum_k x_{ijk}$ denotes the total built up area of cells in district j with land use type k (where k denotes the land use types within which residential building occurs). On the other hand, $\sum_i x_{ij}$ is the total built up area of all cells in district j. For E_3 urban sprawl depends only on whether land is used or not for residential purposes, and not on how many people live on it.

3.4. Critique on the use of entropy indices

Even though entropy indices are very popular in the urban sprawl literature, we find that their interpretation is not straightforward. This difficulty comes from the fact that entropy indices combine two components within one index: on the one hand the prevalence of one cell type compared to another, on the other hand how scattered these cell types are within the chosen area. This means that the same level of entropy can exist in areas that are very different in their properties. For example, an area with a high prevalence of one particular cell type and a high degree of scatteredness will get a similar entropy score to another area that is characterized by a low prevalence of this cell type and a low degree of scatteredness. This makes entropy comparisons between different areas potentially problematic. For this reason, we employ an alternative measure of urban continuity (scatteredness) which does not suffer from this problem. The (global) Moran I index, a measure of spatial autocorrelation, does just that.

3.5. Spatial autocorrelation as an alternative for entropy

The Moran I index is a measure of spatial autocorrelation that focuses on the degree of clustering within an area. The Moran I index was introduced to the urban sprawl literature by [14] to distinguish between compactness and sprawl, and also in order to characterize metropolitan form. Similar to entropy measures, Moran's I index captures unequal distribution in an area. However, while entropy

---

[1] Because we are interested in urban sprawl, which is a phenomenon primarily linked to housing sprawl, we define a cell as built-up if there is residential housing in that cell.

refers to the distribution and composition of cells over the entire area, spatial autocorrelation measures the similarity of neighboring cells. Thus, a gradual decline in intensity of settlement from the inner-city outwards would give us high neighborhood similarity values. On the other hand, a city with multiple urban centers and agricultural areas in between these centers would give us low neighborhood similarity values. Entropy also measures cell similarity, but it captures the cell similarity average over the entire area, with the consequence that those two scenarios get similar entropy values. For this reason, we believe that the Moran I index is better suited to measure the scatteredness aspect of urban sprawl than entropy indices. Low autocorrelation values will generally mean a high degree of scatteredness and thus a high degree of urban sprawl. The interpretation of high autocorrelation numbers remains context specific as high autocorrelation can be due to the cells of the chosen area showing either similarly high population numbers or similarly low population numbers.

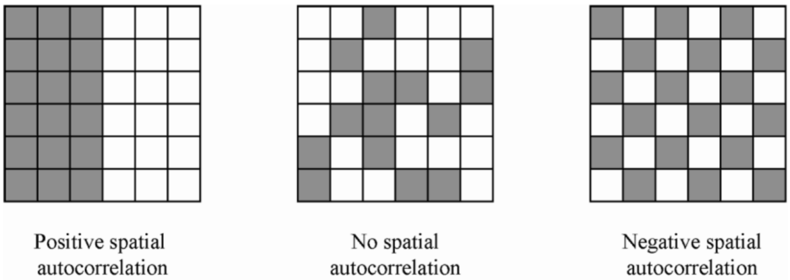

Figure 2: llustration of spatial autocorrelation

The Moran I index is defined as follows:

$$Moran = \frac{n}{w} \cdot \frac{[\sum_i \sum_j w_{ij}\ (pop_i - \overline{pop})]\ \cdot [pop_j - \overline{pop}]}{\overline{pop})^2}$$

where pop_i now denotes the population in cell i, while pop_j denotes the population in cell j, $\overline{pop}$ denotes average cell population (over the entire area), $w = \sum_i \sum_j w_{ij}$ with the distance weight w_{ij} set equal to 1 if the cells are neighbors, and to zero if they are not.

The Moran I index has a range from -1 to +1. To be consistent with the entropy measures discussed above – where a value close to 1 indicates a high degree of urban sprawl – we change the direction of the Moran's I index so that a higher value implies more urban sprawl. Also, we rescale it to lie between 0 and 1. Our modified version of the Moran I index is calculated as follows:

$$M_1 = \frac{1 - Moran}{2}$$

The $\boldsymbol{M_1}$ index will be below 0.5 whenever the values in the dataset tend to cluster spatially – high values cluster near other high values and/or low values cluster near other low values. Thus, low values of $\boldsymbol{M_1}$ mean that neighboring areas are similar to each other in terms of development intensity. When the value of M_1 becomes close to 0.5 the cell characteristics show less spatial autocorrelation. Values above 0.5 indicate negative spatial autocorrelation – something that we do not expect to find in the empirical urban sprawl context. The larger the $\boldsymbol{M_1}$ index, the more scattered the housing development over the area and thus the higher is the level of urban sprawl.

There is one problem with the interpretation of $\boldsymbol{M_1}$ however: without knowing the closer context we cannot say whether a low urban sprawl result (low $\boldsymbol{M_1}$) comes from an environment in which there is very little urban development (and low sprawl because most of the area is uninhabitated), or because of very high and dense urban development, or even if this result is based on low density housing throughout the area (often considered one of the main characteristics of urban sprawl). Tsai [14] introduced the Moran I index into the urban sprawl literature in order to characterize metropolitan form, and he sees it as the ideal candidate to replace a mix of other urban sprawl indices (like density indices, inequality indices, and metropolitan size measures). But because of the difficulty in interpreting the Moran I index results without looking at the underlying data structure we do not completely agree with [14] on this point. We believe that the Moran I index (and our $\boldsymbol{M_1}$ variant on it) is a valuable addition to the tool kit for measuring urban sprawl, but that it is not suitable to replace all other measures on this topic.

4. Spatial Compactness Measures

With each of the measures considered so far, we moved a step further away from the notion of urban sprawl as a population density concept and more towards urban sprawl as a land use issue. In section 4.2 we already introduced an urban sprawl measure that is independent of population data. Now we take this concept a step further and consider a class of methods for measuring urban sprawl that do not require any population data. We have already mentioned above that such an approach is useful in situations in which no reliable population data exist. Obtaining population density measures can also be problematic when the political boundary of a town does not coincide with its functional boundary and no cell level population data exist. The class of urban sprawl measures in this section focus on the shape and relative size of a town's boundary. The Corine Land Cover Project [1] provides an ideal framework for illustrating these types of methods. The question of "where does a town end?" has various possible answers. For the compactness measures we define the "end of town" as the end of cells with Corine Land Cover classification 1 or 2 (continuous or discontinuous urban fabric) and use the shape of the urban structure that evolves when we take all urban cells that are less than 200 meters apart. This follows the Urban Morphological Zone (UMZ) definition by the [1].

One of the often-mentioned aspects of urban sprawl is "strip development" along public roads and highways. This is the result of an externality problem. "Poor land use regulation may allow developers to strip-develop along state-funded roads, rather than absorb the full cost of construction of local road systems deep into residential developments. The social benefits of higher densities at the fringe, such as enhanced social interactions or preservation of green spaces, may not be incorporated into land use and development actions" [30]. Compactness measures will capture this aspect of sprawl which cannot be addressed with the other indices considered so far.

The most compact boundary is a circle. Arranging all populated cells within a circle would minimize the urban sprawl according to this type of measure. The more the shape departs from a circle the higher is the resulting measure of urban sprawl.

A simple compactness measure consists of the scalar-adjusted ratio between area and circumference of a settlement. This measure is dimensionless and the simplest measure of shape complexity.

$$C_0 = \frac{p}{\sum_i x_i}\frac{2}{r} - 1,$$

where p denotes the perimeter of the town, $\sum_i x_i$ denotes the area of the included cells, and r is the average radius of the settlement. A circle is the most compact two-dimensional form, and for a circular settlement C_0 reduces to zero. $\boldsymbol{C_0}$ can therefore be used to compare the shape of an actual town with the ideal of a circle.

However, there are some problems with this simple compactness measure C_0. First, it is not clear how to measure the radius of an irregular shape. Second, when using data based on a grid, the grid structure of the data produces "noisy contours" and consequently larger perimeters [23] which potentially inflates the compactness measure ($\boldsymbol{C_0}$ or any other traditional compactness measure). This problem is similar to Mandelbrot's classic example of how the length of the coastline of Britain can be increased by reducing the size of the measuring stick with which it is measured [31].

To avoid the problem of overstating contours, we now turn to compactness measures that are defined on grids in the first place. Compactness index $\boldsymbol{C_1}$ treats a square as the most compact shape, while compactness measure $\boldsymbol{C_2}$ treats a grid approximation to a circle as the most compact shape.

4.1. Compactness Index 1

$\boldsymbol{C_1}$ is based on an index introduced by Milne [22] and asks how much a particular area differs from a square. Milne's index is a shape index that calculates patch perimeter (given in number of cell surfaces) divided by the minimum perimeter (given in number of cell surfaces) possible for a maximally compact patch (in a square raster format) of the corresponding patch area [32]. We change the index slightly by letting $\boldsymbol{C_1}$ take the form of one minus the reciprocal of Milne's original index. We do this for two reasons: first we want all our parameters be bounded between 0 and 1 (which Milne's original index is not), second, we want all sprawl measurements to be pointing in the same direction with higher numbers indicating a higher level of urban sprawl.

$$C_1 = 1 - \frac{min\,p}{p}$$

where:

p = perimeter of the town in terms of number of cell surfaces.

$min\ p$ = minimum perimeter of the town in terms of number of cell surfaces. $min\ p$ is calculated as follows:

$min\ p = 4s$, when $m = 0$,

$min\ p = 4s + 2$, if $s^2 < a \leq s(1+s)$, or

$min\ p = 4s + 4$, if $a > s(1+s)$

where $s = \sqrt{a}$, $m = a - s^2$, a denotes the area of the shape under consideration (in terms of number of cells), and s is the side of the largest integer square (i.e., a square formed from the individual cells) smaller than a.

$\boldsymbol{C_1}$ takes on the value of 0 when the settlement area is maximally compact (i.e., square or almost square) and increases towards 1 as patch shape becomes more irregular. Given that GIS based land use data is presented in square grid form, this index is easy to apply in this context. A drawback of this index is however, that a square is not a natural form for settlements to take. Cities that grow over time generally are more likely to resemble circles than squares. Our next compactness measure provides a measure that compares settlements with the ideal compact form of a circle while at the same time correcting for grid-cell data problems.

4.2. Compactness Index 2 – Bribiesca's Contact Perimeter approach

Here we introduce an approach for measuring the compactness of objects composed of a finite number of cells that was developed by [23]. Bribiesca's paper appeared in the Mathematics and Computer Science literature and has so far not been applied in an urban context. Bribiesca's index basically allows us to do with grid data what the $\boldsymbol{C_0}$ measure can do in continuous space – it can compare the shape of an area with the ideal of a circle while adjusting for the grid structure. It is also size invariant. Size invariance is important not only to compare the compactness of settlements of different sizes, but also to compare measurements of the same settlement that are using different size grid structures. The Bribiesca Contact Perimeter approach is a discrete measure of compactness using the individual grid cell as its measurement instrument. Using the individual cell as the unit of measurement has the advantage that the accuracy of the shape contour as well as the shape area are measured using the same element – the cell [23].

To understand the method, it is first necessary to introduce the concept of the contact perimeter. The perimeter of a shape that is composed of cells consists of the sum of the lengths of the sides of the closed shape. In contrast, the "contact perimeter" P_C of a shape that is composed of cells corresponds to the sum of the lengths of segments which are common to the two cells. The notion of a contact perimeter and how it differs from the "normal" perimeter are illustrated in Figure 3 below.

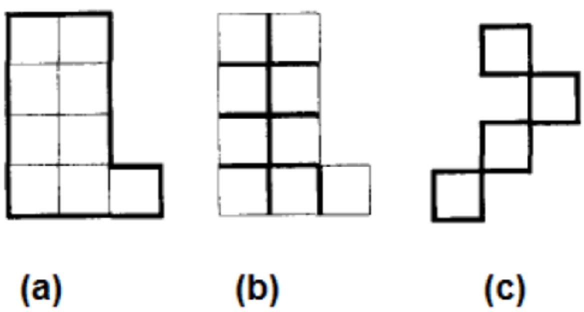

Figure 3: The perimeters and their relations: (a) the perimeter of an L-shape; (b) the contact perimeter of this same L-shape; (c) a shape without contact perimeter. Source: modified from [23].

The measure of normalized discrete compactness according to [23] is defined as follows:

$$Bribiesca's\ index = \frac{C_D - C_{D_{min}}}{C_{D_{max}} - C_{D_{min}}}$$

where

$C_D = P_C$ (Contact perimeter)

$C_{D_{min}} = n - 1$

$C_{D_{max}} = 2(n - \sqrt{n})$

n = number of cells in shape

The value of this index varies continuously from 0 (minimum compactness) to 1 (maximum compactness) [23]. Figure 4 illustrates digital circles and their values for this index.

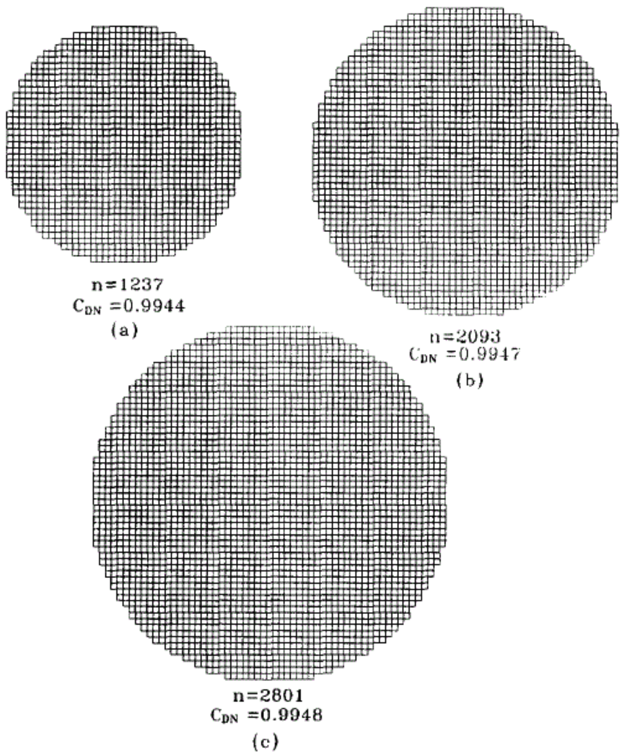

Figure 4: Examples of three different levels of resolution for a digital circle: (a) the digital circle composed of 1237 pixels; (b) the digital circle composed of 2093 pixels; (c) the digital circle composed of 2801 pixels; source [23].

This measure of discrete compactness is invariant under translation, rotation, and scaling [23]. Especially its invariance under scaling is important for our purposes: it allows the index to remain constant when the same shape is presented using different scale grid representations (and therefore a different number of pixels). Thus, a settlement's compactness value will remain (approximately) constant if grid size changes from 100m x 100m to 25m x 25m.

Again, we will change the index slightly to make it comparable with the other urban sprawl indices we consider. Our $\boldsymbol{C_2}$ index, which is based on Bribiesca's [23] index, is defined as follows:

$$C_2 = \frac{C_{D_{max}} - C_D}{C_{D_{max}} - C_{D_{min}}}$$

$\boldsymbol{C_2}$ will be bounded between 0 and 1, with larger numbers indicating a less compact form and thus a higher degree of urban sprawl. In our trials this index performs well and we believe that it is a useful addition to the literature on the measurement of urban sprawl.

5. Discussion

The Corine Land Cover dataset is provided by the European Environment Agency and provides land-use data in polygon as well as in grid structure. For the grid-based data, the Corine Land Cover Project classifies individual cells according to their dominant usage into 42 possible land use classes.

In the case of Graz, Austria there are 12,762 cells within the official city limits, with each cell measuring 100 by 100 meters. It follows that x_{ij} equals 0.01 square kilometers for all i and j.

Table 1 illustrates the results of all presented urban sprawl measures for our test case. For all but the population density measures, results lie between zero and one and higher numbers indicate a higher degree of urban sprawl. Each family of indices measures a different aspect of urban sprawl. These aspects are related but not identical – therefore we cannot say that one type of sprawl measure is categorically better than another. However, within each family we can generalize somewhat.

The simple $\boldsymbol{D_1}$ population density measure within the official city limits of Graz is 1,944 people per square kilometer which drops to 311 people per square kilometer when the 57 surrounding communities within the "Greater Graz Area" are included. For the population density measure $\boldsymbol{D_2}$ we need to establish in which of these land-use classes built-up area exists.[2] We find that for Graz, significant housing is only found in cells classified as CLC1 and CLC2. $\boldsymbol{D_2}$ is then calculated as follows: we divide the city population by the area of all CLC class 1 and 2 cells. Population density for the greater Graz area (including 57 municipalities outside of Graz) is calculated in the same way.

We apply these population density indices $\boldsymbol{D_1}$ and $\boldsymbol{D_2}$ to derive two versions of Entropy measures. Using $\boldsymbol{D_1}$ to calculate the entropy weights p_j we get entropy measure $\boldsymbol{E_1}$, while using $\boldsymbol{D_2}$ will give us $\boldsymbol{E_2}$. $\boldsymbol{E_3}$ does not need any population data.

Sprawl Measures	**Density Index D_1**	**Density Index D_2**	**Entropy Index E_1**	**Entropy Index E_2**	**Entropy Index E_3**	**Moran I Index M_1**	**Compact sprawl Index C_1**	**Compact sprawl Index C_2**
Graz	1944/km²	3635/km²	0.93	0.94	0.99	0.12	0.88	0.19
Graz including surrounding area	311/km²	2546/km²	0.92	0.83	0.92	0.19	0.95	0.31

Table 1: Results of Sprawl Indices

Next, we measure $\boldsymbol{M_1}$ with respect to population per cell for the area within the city limit of Graz and for the wider-Graz area. The $\boldsymbol{M_1}$ index which is based on the (global) Moran's I index also attempts to measure the scatteredness of an area. We find that the $\boldsymbol{M_1}$ values for of the test areas are less than 0.5, which indicates a positive spatial autocorrelation. Furthermore, our results show that the city of Graz has a higher degree of spatial autocorrelation (lower $\boldsymbol{M_1}$) and thus less urban sprawl than the greater Graz region.

While entropy measures both the relative prevalence of cell types and the overall distribution (scatteredness) of relative cell types, $\boldsymbol{M_1}$ focuses exclusively on the distribution of cell types. This is a positive feature of the $\boldsymbol{M_1}$ index that avoids the interpretation difficulties of the entropy indices. This is probably why Tsai (2005) sees the Moran's I index as the ideal candidate to replace a mix of other urban sprawl indices (like density indices, inequality indices, and metropolitan size measures). Nevertheless, we find that our $\boldsymbol{M_1}$ index (and thus the related Moran's I index as well) has other difficulities with respect to its interpretation. Low scatteredness (i.e., a low value of $\boldsymbol{M_1}$) means low variability between cells. This can come from extremely dense settlement throughout the area (such as in Dhakar in Bangladesh), particularly low settlement throughout the entire area (e.g., the Gobi desert), or because of a vast area of low density housing development. All three situations are characterized by low variability. The first two we would not want to classify as urban sprawl, but the third situation is one of the key characteristics of urban sprawl. To know which of these situations is the right interpretation one needs to know the underlying data. Therefore, the interpretation of a single out-of context $\boldsymbol{M_1}$ number is difficult.

With respect to the compactness indices, the results for C_1 indicate that the city of Graz shows lower urban sprawl than the greater Graz region – which coincides with our regional knowledge. However, the C_1 index is not ideal as its base shape is a square, which is not a typical shape for organically grown cities. This drawback is resolved in the C_2 index, which performs well with our sample data and is in our opinion the best index from the third category. Being invariant to size changes, C_2 provides a meaningful comparison of the compactness of the city of Graz and greater Graz regions. Our results for C_2 show that the city of Graz has a quite low shape-based sprawl value (0.12) which rises as we include the surrounding area of Graz. It performs exactly as it should.

Figure 5: CLC class 1 and 2 cells within city of Graz and for greater Graz region

6. Conclusions

Urban sprawl has many interpretations and different measures capture different aspects of it. We have concentrated here on three main characteristics of urban sprawl: density, scatteredness, and shape of the urban development. We have developed and applied density, entropy, spatial autocorrelation, and compactness indices that measure these three components of urban sprawl. As traditional methods of measurement are not always appropriate to use on GIS data, we presented indices that are optimally suited to this type of data. Urban sprawl is a complex phenomenon and it is therefore useful to apply a variety of measurements to capture what is going on. Focusing on a single measurement of urban sprawl can be misleading.

Summing up, we can say the following: While we believe that urban sprawl has many different components – each of which needs a different approach – we find that within each group some indicators are better suited for this task than others. For the density indices we favor D_2 as it corrects for areas that do not contain population. Contrary to its popularity in the literature, we find entropy indices are not ideal for capturing the scatteredness of urban sprawl. Instead we find that the Moran's I index (and our variant of it M_1) performs better. Compactness indices attempt to formalize the "intuition" we get when looking at a map to decide whether a town has a compact form or not. The grid structure of the data imposes some restrictions that were not considered in the literature so far. We find the C_2 index is well suited to deal with any problems arising from the grid structure and is able to give meaningful, size-invariant results on the compactness of settled areas. We hope it will be used in further studies on urban sprawl.

References

1. European Environment Agency (2010), *Corine Land Cover project 2006*, Version 13, http://www.eea.europa.eu/data-and-maps/data/corine-land-cover-2006-raster.
2. Frenkel A., and Ashkenazi M. (2008), Measuring Urban Sprawl; How can we deal with it? *Environment and Planning B: Planning and Design* 35(1), 56-79.
3. Brueckner J. K. (2000), Urban Sprawl: Diagnosis and Remedies, *International Regional Science Review* 23(2), 160-171.
4. Nechyba T. J. and Walsh R. P. (2004), Urban Sprawl, *Journal of Economic Perspectives* 18(4), 177–200.
5. Morote A.-F. and Hernández M. (2016), Urban sprawl and its effects on water demand: A case study of Alicante, Spain, *Land Use Policy* 50, 352-362
6. Al Jarah S. H., Zhou B., Abdullah R. J., Lu Y. and Yu W. (2019), Urbanization and Urban Sprawl Issues in City Structure: A Case of the Sulaymaniah Iraqi Kurdistan Region, *Sustainability* 11, 485, 1-21.
7. Glaeser E. L. and Kahn M. E. (2003), Sprawl and Urban Growth, Working Paper 9733, *National Bureau of Economic Research*, Cambridge, MA 02138.
8. Frumkin H. (2002), Urban Sprawl and Public Health, *Public Health Reports* 117(3), 201-217.
9. Ewing R., Schmid T., Killingsworth R., Zlot A., Raudenbush, S. (2003), Relationship Between Urban Sprawl and Physical Activity, Obesity, and Morbidity, *American Journal of Health Promotion*, 18(1), 47-57.
10. Bray R, Vakil C., Elliott D. (2005), Report on Public Health and Urban Sprawl in Ontario: a review of the Pertinent Literature, *Energy Technology Data Exchange World Energy Base* (ETDEWEB).
11. Coughenour C., de la Fuente-Mella H. and Paz A. (2019), Analysis of Self-Reported Walking for Transit in a Sprawling Urban Metropolitan Area in the Western U.S., *Sustainability* 11. 852, 1-16.
12. Carruthers J. (2008), Does Smart Growth Matter to Public Finance? Urban Studies 45(9), 1791-1823.
13. Fernandez Milan B. and Creutzig F. (2016), Municipal policies accelerated urban sprawl and public debts in Spain, *Land Use Policy* 54, 103-115
14. Tsai, Y.-H. (2005), Quantifying Urban Form: Compactness versus ‘Sprawl’, *Urban Studies* 42(1), 141-161.
15. Yeh A. G. O. and Li X. (2001), Measuring and Monitoring of Urban Sprawl in a Rapidly Growing Region Using Entropy, *Photogrammetric Engineering and Remote Sensing* 67(1), 83-90.
16. Li X. and Yeh A. G. O. (2004), Analyzing spatial restructuring of land use patterns in a fast-growing region using remote sensing and GIS, *Landscape and Urban Planning* 69(4), 335–354.
17. Sun H., Forsythe W., and Waters N. (2007), Modelling Urban Land Use Change and Urban Sprawl: Calgary, Alberta, Canada, *Networks and Spatial Economics* 7, 353-376.
18. Jat M. K., Garg P. K. and Khare D. (2008), Monitoring and modelling of urban sprawl using remote sensing and GIS techniques, *International Journal of Applied Earth Observation and Geoinformation* 10, 26-43.
19. Bhatta B. (2009), Analysis of Urban Growth Pattern Using Remote Sensing and GIS: a Case Study of Kolkata, India, *International Journal of Remote Sensing* 30(18), 4733-4746.
20. Yue W., Liu Y. and Fan P. (2013), Measuring urban sprawl and its drivers in large Chinese cities: The case of Hangzhou, *Land Use Policy* 31, 358–370.
21. Nazarnia N., Harding C. and Jaeger J. A. G. (2019), How Suitable Is Entropy as a Measure of Urban Sprawl? *Landscape and Urban Planning* 184, 32-43.
22. Milne, B. T. (1991), *Lessons from applying fractal models to landscape patterns, Quantitative Methods in Landscape Ecology*, edited by Turner M. G. and Gardner R. H., Springer Verlag, New York, 199-235.
23. Bribiesca E. (1997), Measuring 2-D Shape Compactness Using the Contact Perimeter, *Computers Mathematical Applications* 33(11), 1-9.

24. New Geography: *World Urban Areas Population and Density*: a 2012 Update, http://www.newgeography.com/content/002808-world-urban-areas-population-and-density-a-2012-update.
25. Cohen B. (2004), Urban Growth in Developing Countries: a Review of Current Trends and a Caution Regarding Existing Forecasts, *World Development* 32, 23-51.
26. Schneider A. and Woodcock C. E. (2008), Compact, Dispersed, Fragmented, Extensive? A Comparison of Urban Growth in Twenty-Five Global Cities using Remotely Sensed Data, Pattern Metrics and Census Information, *Urban Studies* 45(3), 659-692.
27. Galster G., Hanson R., Ratcliffe M. R., Wolman H., Coleman S. and Freihage J. (2001), Wrestling Sprawl to the Ground: Defining and measuring an elusive concept, *Housing Policy Debate* 12(4), 681-717.
28. Cabral P., Augusto G., Tewolde M. and Y. Araya (2013), Entropy in Urban Systems, Entropy 15, 5223- 5236.
29. Bhatta B., Saraswati S. and Bandyopadhyay D. (2010), Urban sprawl measurement from remote sensing data, *Applied Geography* 30, 731-740.
30. Henderson J. V. (2009), Comment on “Urban Sprawl in Europe” by Patacchini E and Zenou Y, *Brookings-Wharton Papers on Urban Affairs*, 125-149.
31. Mandelbrot B. (1967), How Long is the Coast of Britain? Statistical Self-Similarity and Fractional Dimension, *Science* 156, 636–638.
32. Department of Environmental Conservation, University of Massachusetts, website http://www.umass.edu/landeco/research/fragstats/documents/Metrics/Shape%20Metrics/Metrics/P8%20-%20SHAPE.htm.

Graz Economics Papers

For full list see:
http://ideas.repec.org/s/grz/wpaper.html
Address: Department of Economics, University of Graz,
Universitätsstraße 15/F4, A-8010 Graz

02–2020	**Miriam Steurer and Caroline Bayr**: Measuring Urban Sprawl using Land Use Data
01–2020	**Thomas Aronsson, Sugata Ghosh and Ronald Wendner**: Positional Preferences and Efficiency in a Dynamic Economy
14–2019	**Nicole Palan, Nadia Simoes, and Nuno Crespo**: Measuring Fifty Years of Trade Globalization
13–2019	**Alejandro Caparrós and Michael Finus**: Public Good Agreements under the Weakest-link Technology
12–2019	**Michael Finus, Raoul Schneider and Pedro Pintassilgo**: The Role of Social and Technical Excludability for the Success of Impure Public Good and Common Pool Agreements: The Case of International Fisheries
11–2019	**Thomas Aronsson, Olof Johansson-Stenman and Ronald Wendner**: Charity as Income Redistribution: A Model with Optimal Taxation, Status, and Social Stigma
10–2019	**Yuval Heller and Christoph Kuzmics**: Renegotiation and Coordination with Private Values
09–2019	**Philipp Külpmann and Christoph Kuzmics**: On the Predictive Power of Theories of One-Shot Play
08–2019	**Enno Mammen, Jens Perch Nielsen, Michael Scholz and Stefan Sperlich**: Conditional variance forecasts for long-term stock returns
07–2019	**Christoph Kuzmics, Brian W. Rogers and Xiannong Zhang**: Is Ellsberg behavior evidence of ambiguity aversion?

06–2019 **Ioannis Kyriakou, Parastoo Mousavi, Jens Perch Nielsen and Michael Scholz**: Machine Learning for Forecasting Excess Stock Returns The Five-Year-View

05–2019 **Robert J. Hill, Miriam Steurer and Sofie R. Waltl**: Owner-Occupied Housing, Inflation, and Monetary Policy

04–2019 **Thomas Aronsson, Olof Johansson-Stenman and Ronald Wendner**: Charity, Status, and Optimal Taxation: Welfarist and Paternalist Approaches

03–2019 **Michael Greinecker and Christoph Kuzmics**: Limit Orders under Knightian Uncertainty

02–2019 **Miriam Steurer and Robert J. Hill**: Metrics for Measuring the Performance of Machine Learning Prediction Models: An Application to the Housing Market

01–2019 **Katja Kalkschmied and Jörn Kleinert**: (Mis)Matches of Institutions: The EU and Varieties of Capitalism

21–2018 **Nathalie Mathieu-Bolh and Ronald Wendner**: We Are What We Eat: Obesity, Income, and Social Comparisons

20–2018 **Nina Knittel, Martin W. Jury, Birgit Bednar-Friedl, Gabriel Bachner and Andrea Steiner**: The implications of climate change on Germanys foreign trade: A global analysis of heat-related labour productivity losses

19–2018 **Yadira Mori-Clement, Stefan Nabernegg and Birgit Bednar-Friedl**: Can preferential trade agreements enhance renewable electricity generation in emerging economies? A model-based policy analysis for Brazil and the European Union

18–2018 **Stefan Borsky and Katja Kalkschmied**: Corruption in Space: A closer look at the world's subnations

17–2018 **Gabriel Bachner, Birgit Bednar-Friedl and Nina Knittel**: How public adaptation to climate change affects the government budget: A model-based analysis for Austria in 2050

16–2018 **Michael Günther, Christoph Kuzmics and Antoine Salomon**: A Note on Renegotiation in Repeated Games [Games Econ. Behav. 1 (1989) 327360]

15–2018 **Meng-Wei Chen, Yu Chen, Zhen-Hua Wu and Ningru Zhao**: Government Intervention, Innovation, and Entrepreneurship

14–2018 **Yu Chen, Shaobin Shi and Yugang Tang**: Valuing the Urban Hukou in China: Evidence from a Regression Discontinuity Design in Housing Price

13–2018 **Stefan Borsky and Christian Unterberger**: Bad Weather and Flight Delays: The Impact of Sudden and Slow Onset Weather Events

12–2018 **David Rietzke and Yu Chen**: Push or Pull? Performance-Pay, Incentives, and Information

11–2018 **Xi Chen, Yu Chen and Xuhu Wan**: Delegated Project Search

10–2018 **Stefan Nabernegg, Birgit Bednar-Friedl, Pablo Muñoz, Michaela Titz and Johanna Vogel**: National policies for global emission reductions: Effectiveness of carbon emission reductions in international supply chains

09–2018 **Jonas Dovern and Hans Manner**: Order Invariant Tests for Proper Calibration of Multivariate Density Forecasts

08–2018 **Ioannis Kyriakou, Parastoo Mousavi, Jens Perch Nielsen and Michael Scholz**: Choice of Benchmark When Forecasting Long-term Stock Returns

07–2018 **Joern Kleinert**: Globalization Effects on the Distribution of Income

06–2018 **Nian Yang, Jun Yang and Yu Chen**: Contracting in a Continuous-Time Model with Three-Sided Moral Hazard and Cost Synergies

05–2018 **Christoph Kuzmics and Daniel Rodenburger**: A case of evolutionary stable attainable equilibrium in the lab

04–2018 **Robert J. Hill, Alicia N. Rambaldi, and Michael Scholz**: Higher Frequency Hedonic Property Price Indices: A State Space Approach

03–2018 **Reza Hajargasht, Robert J. Hill, D. S. Prasada Rao, and Sriram Shankar**: Spatial Chaining in International Comparisons of Prices and Real Incomes

02–2018 **Christoph Zwick**: On the origin of current account deficits in the Euro area periphery: A DSGE perspective

01–2018 **Michael Greinecker and Christopher Kah**: Pairwise stable matching in large economies