IRIS AperTO

UNIVERSITÀ DEGLI STUDI DI TORINO

AperTO - Archivio Istituzionale Open Access dell'Università di Torino

A scalable method to quantify the relationship between urban form and socio-economic indexes

This is the author's manuscript

Original Citation:

Availability:

This version is available http://hdl.handle.net/2318/1730493 since 2020-02-24T16:23:50Z

Published version:

DOI:10.1140/epjds/s13688-018-0132-1

Terms of use:

Open Access

Anyone can freely access the full text of works made available as "Open Access". Works made available under a Creative Commons license can be used according to the terms and conditions of said license. Use of all other works requires consent of the right holder (author or publisher) if not exempted from copyright protection by the applicable law.

(Article begins on next page)

12 October 2024

RESEARCH

A scalable method to quantify the relationship between urban form and socio-economic indexes

Alessandro Venerandi[1*], Giovanni Quattrone[2] and Licia Capra[2]

[*]Correspondence: alessandro.venerandi.12@ucl.ac.uk
[1]Department of Civil, Environmental & Geomatic Engineering, UCL, Gower Street, London, UK
Full list of author information is available at the end of the article
[†]Equal contributor

Abstract

The world is undergoing a process of fast and unprecedented urbanisation. It is reported that by 2050 66% of the entire world population will live in cities. Although this phenomenon is generally considered beneficial, it is also causing housing crises and more inequality worldwide. In the past, the relationship between design features of cities and socio-economic levels of their residents has been investigated using both qualitative and quantitative methods. However, both sets of works had significant limitations as the former lacked generalizability and replicability, while the latter had a too narrow focus, since they tended to analyse single aspects of the urban environment rather than a more complex set of metrics. This might have been caused by the lack of data availability. Nowadays, though, larger and freely accessible repositories of data can be used for this purpose. In this paper, we propose a scalable method that delves deeper into the relationship between features of cities and socio-economics. The method uses openly accessible datasets to extract multiple metrics of urban form and then models the relationship between urban form and socio-economic levels through spatial regression analysis. We applied this method to the six major conurbations (i.e., London, Manchester, Birmingham, Liverpool, Leeds, and Newcastle) of the United Kingdom (UK) and found that urban form could explain up to 70% of the variance of the English official socio-economic index, the Index of Multiple Deprivation (IMD). In particular, results suggest that more deprived UK neighbourhoods are characterised by higher population density, larger portions of unbuilt land, more dead-end roads, and a more regular street pattern.

Keywords: urban form; socio-economics; spatial analysis; open data; OpenStreetMap

Introduction

Cities are growing faster than ever before. In 1950, only 30% of the total world population was living in cities. Today, this datum stands around 54%. By 2050, the estimates project that 66% of the total world population will be urban, with cities in developing countries attracting the greatest number of new city dwellers [1]. Urbanisation is regarded by institutions and governments as a positive phenomenon as it brings, for example, better and less costly public services and improved living standards due to the concentration of economic activities [2]. However, this very same phenomenon is also reported to bring more inequality worldwide, with some areas benefiting more from public investments and economic growth than others [3]. It is thus necessary to develop a better understanding of the relationship between physical features of the urban environment and socio-economic levels of city dwellers, to inform urbanists and city planners.

Urban form had been investigated extensively in the past, in relation to socio-economics and well-being. Jacobs, for example, observed different parts of her city (i.e., New York) and reached the conclusion that the traditional compact pedestrian friendly city form would have ensured the overall well-being of city dwellers [4]. Swiss architect Le Corbusier formulated a theory based on his personal perspective and reached an opposite conclusion, instead. In his view, the optimal city form was dispersed and more car-oriented [5]. The main limitation of these works lies in the use of qualitative methodologies which render studies difficult to repeat, and outcomes to generalise. More recently, researchers adopted quantitative methods to study the relationship between features of the urban environment and socio-economic aspects (see, for example, [6] and [7]). The main limitations of these works is that they analysed relatively small geographic areas (e.g., single neighbourhoods) and focused on single aspects of the urban environment (e.g., place accessibility) despite the fact that urban form is, by definition, the interplay of multiple elements and these should be thus studied together to best capture the socio-economics of city neighbourhoods.

The choice of using qualitative methods or analysing single metrics might have been dictated by the lack of available data. In the last decade, though, large data repositories and new techniques of data collection (e.g., crowd-sourcing) have become readily available (i.e., "open data revolution" [8, 9]). Researchers have recently started to take advantage of such phenomenon and study cities through quantitative methods. For example, they analysed crowd-sourced visual perceptions of different urban environments in relation to socio-demographic factors [10] and urban qualities, such as beauty [11].

Inspired by this set of works, which analysed urban form in a more comprehensive manner, and by taking advantage of the "open data revolution" too, we propose a quantitative method that uses openly accessible datasets and spatial regression analysis to study the relationship between multiple features of urban form and socio-economic indexes. Unlike previous works, our method (i) relies on multiple descriptors of the urban environment and (ii) can be applied to cities of different sizes and repeatedly over time, at almost no cost. To test this method, we applied it to the UK major urban areas as identified by an official document (i.e., London, Manchester, Birmingham, Liverpool, Leeds, and Newcastle) [12]. Both features of urban form and information on the socio-economic levels of their communities were extracted from openly accessible datasets. The former were extracted from Ordnance Survey (OS) VectorMap District and OpenStreetMap (OSM), while the latter was obtained from the English IMD database. Outcomes of the spatial models could explain between 27% and 70% of the variance of IMD, confirming the existence of a relationship between urban form and socio-economics across the six cities. Furthermore, we found some aspects of urban form to have similar behaviours across the case studies thus highlighting some common patterns. In particular, more deprived neighbourhoods of urban UK were found to be characterised by higher population density, more unbuilt land, a higher presence of dead-end roads, and a more regular street pattern. The method proposed in this paper and its outcomes can be helpful in the current urbanisation age as they constitute a data driven basis for reasoning on possible design schemes and urban policies.

The remainder of this paper is structured as follows. We firstly illustrate related works. We then present our method, starting from the metrics of urban form and the concept of socio-economic index, and by then following with the type of analysis conducted. We discuss the results of its application to the six UK cities under study, before offering our interpretations. We conclude with final remarks and limitations.

Related works

As we presented in the Introduction, the relationship between urban form and socio-economic outcomes has been investigated in a variety of ways. Several authors used qualitative methods based on observation or personal views. These authors can be subdivided in two schools of thought: those supporting the compact pedestrian city form and those favouring more spread out and car-oriented urban environments. Jacobs belongs to the former group as she supported the features of the traditional city, that is medium to high built density, perimeter blocks, walk-ability, and mixed-use [4]. Similarly, Whyte praised human scale streets, walk-ability, and argued that subtle urban details, such as shop widows, porticoes, steps, and doorways, were indispensable for city liveability [13]. Gehl is on the same page and favours pedestrian mixed-use streets, as well as active *city edges* (i.e., block frontages), which, in his view, can promote social interactions as well as stimulate commercial activities [14]. The other school of thought was openly against the traditional city form instead. Le Corbusier, for example, deemed it as disordered, chaotic, and unhealthy [5]. He proposed a city based on super blocks (i.e., urban blocks of big dimension) delimited by multi-lane highways and residential tower blocks retracted from the side-walks and laid out in open space (the so-called "towers in the park"). Similarly, Hilberseimer despised the traditional city form and proposed plans characterised by a grid of highways and repetitive, tall residential blocks aligned to them [15]. Although these works presented insightful perspectives, they lack generalizability as they were based on qualitative methods. Moreover, they are hardly replicable as mostly based on personal views and observations.

More recently, thanks to the diffusion of computers and Geographic Information Systems (GIS), the study of cities has become more quantitative. Vaughan *et al.*, for example, adopted a renowned technique for the analysis of street networks (i.e., Space Syntax [16]) to study the relationship between *integration* (i.e., a measure of spatial accessibility) and socio-economic levels of East London residents [6]. The researchers reported that more accessible places, such as main streets, were associated with more affluent residents while less accessible ones, such as back streets and interstitial spaces, were related to less advantaged citizens. Other scholars focused on urban form and criminal activity and studied the relationship between dwelling typologies and crime occurrences in a London borough [17]. These scholars found the flat to be the safest house type. Researchers also separately investigated density in relation to crime; however, they found discordant outcomes. Some reported absence of any relationship [18, 19], while a more recent work found that density was overall beneficial against crime [17]. Hillier studied whether a specific configuration of the street network, that of cul-de-sac, was associated with more or less crimes in a London neighbourhood. Results suggested that cul-de-sac did not attract more crimes than other spatial configurations when integrated in a street network with

significant through movement and many properties facing the streets [7]. Other researchers focused on the relationship between social aspects of urban life and the configuration of cities. Similar to density, outcomes were contrasting. Some scholars reported that higher densities improved social interactions [20] and that lower densities not only decreased social ties, but also favoured more car-oriented behaviours [21]. Contrary to this, other researchers found that higher densities diminished the will of people to socialize and increased stress [22, 23, 24]. Although these studies were based on quantitative methods rather than qualitative ones, findings are still hardly generalizable as the geographic contexts under study were limited (e.g., a single neighbourhood, a specific city). This was mainly due to a lack of data as most of it was proprietary and thus hardly accessible. Furthermore, these works mainly studied separate aspects of urban form in relation to socio-economics (e.g., spatial accessibility and social class in the work by Vaughan *et al.* [6]). However, cities are constituted by many interrelated components, which act in synergy rather than in isolation [25, 26].

In the last decade, given the increased availability of open data repositories and the rise of new data collection techniques (e.g. crowd-sourcing), the limitations stated above (i.e., few and hardly accessible datasets) have almost disappeared. Indeed, researchers have started to analyse cities in more comprehensive manners, for example, through the analysis of pictures, which, by default, comprise of multiple elements of the built environment. Quercia *et al.* used crowd-sourcing to ask more than 3,000 respondents whether photos of different urban environments transmitted beauty, quietness, and happiness [11]. The researchers then used this information to understand what visual features (i.e., colours, textures) best correlated with these qualities and found that the colour green was positively correlated, while wide roads and faceless buildings were inversely correlated. Selasses *et al.* used the same data collection technique (i.e., crowd-sourcing) to ask more than 7,000 people to rate their visual perceptions of street views in terms of safety, social status, and uniqueness [10]. The group of researchers then compared the responses to socio-demographic data and found that spatial dissimilarities in the perception of safety and social status better correlated with violent crimes than their absolute values. Furthermore, they found that, safety perception being equal, these crimes were more related to areas that looked more upper class. Although analysing multiple aspects of urban form at the same time, these works focused on "point-based" analysis (e.g., the characteristics of the urban space which can be seen in a single picture). What is still missing is a method that enables the analysis of multiple aspects of urban form at an "area-based" level. We present next the details of a method that permits this type of spatial analysis in a scalable manner.

Method

The methodology we propose mainly consists of two parts: (i) computation of metrics of urban form and socio-economics aggregated at areal level and (ii) quantitative analysis based on spatial linear regression to understand the relationship between the features of urban form and socio-economic levels of neighbourhoods. We present next the metrics and the procedural steps for carrying out the analysis, before describing how we applied it in practice.

Metrics

Urban form

Our method requires the computation of nine different metrics, capturing different aspects of the built environment. Five of these were derived from previous works, while four are being proposed in this paper. The five metrics derived from previous works are:

- Connected Node Ratio (CNR). This measures the level of connectivity and walk-ability of a street network and was derived from the work by Garrick and Marshall [27]. CNR is computed as the ratio between the number of street intersections that are not cul-de-sac (i.e., node degree equals to 0) and the total number of street intersections in an area. We include this metric as several different authors, including Jacobs [4] and Gehl [14], considered connectivity and walk-ability fundamental aspects for thriving neighbourhoods. These not only positively affected the health conditions of citizens, but also improved social interactions, commercial activities, and provided informal protection against crime.
- Intersection Density (ID). This metric quantifies the density of street intersections in city areas. As for CNR, also ID has been extracted from the work by Garrick and Marshall [27]. Intersection Density is computed as the ratio between the total number of intersections lying in an area and the extension (in square meters) of such area. The reason for including this metric is similar to the one stated above as ID and CNR are closely related. Generally, a dense street network tends also to be more connected and walk-able and thus be associated with the positive aspects mentioned above (i.e., better health conditions, more social and economic benefits, more control against crime [4, 14]).
- Percentage of Unbuilt Land (PUL). It measures the amount of land that is not covered by buildings and was derived from previous work by Banister *et al.* [28]. PUL is calculated dividing the extension (in square meters) of land that is left unbuilt in a city area, by the overall extension (in square meters) of such area (then multiplying this value by 100). PUL provides information on how buildings are distributed across an area. Smaller values of PUL mean that buildings are spread all over an area. Conversely, greater values of PUL mean that buildings are concentrated in fewer spots. We include this metric as the spatial distribution of buildings in city areas was considered a relevant aspect by different researchers. Modernist planners, for example, favoured few buildings surrounded by open space (i.e., "towers in the park") [5, 15]. Authors supportive of the compact city form favoured a more continuous urban fabric instead [4, 13, 14].
- Population Density (PD). It quantifies how densely populated is a city area. PD is a common statistical datum used by institutions and governments and is computed as the ratio between the number of residents living in an area and the extension (usually, in hectares) of such area. PD together with PUL provides information on how built density is distributed across an area. For example, a neighbourhood with a big portion of unbuilt land and a high population density is likely to be characterised by residential towers. Conversely,

a neighbourhood with small unbuilt surface and a high population density is likely to be characterised by perimeter blocks.

- Betweenness Centrality (BC). It measures the level of centrality of streets. To be more specific, BC is based on the concept that a street is central if it is included in many of the shortest paths linking pairs of nodes (street intersections) in a street network. This metric was derived from the work by Porta *et al.* [29]. We include BC as previous works showed it to be associated with positive aspects of cities such as employment density [30], agglomeration of economic activities [31], and street quality [32]. BC is usually computed for street segments. However, since our method analyses areal units, some sort of aggregation is necessary. We do so by calculating BC for the street segments of the urban region under study, and by then identifying the maximum value within each area of such region. We argue that the maximum value of BC can be representative of the level of accessibility of different areas with respect to the overall urban region.

We present next four metrics of urban form that we propose in this paper to complement the previous ones:

- Percentage of Green Areas (PGA). This metric quantifies the amount of public green space available in a city area. It is computed dividing the amount of green space (in square meters) in an area by the total extension (in square meters) of such area (then multiplying this value by 100). We include this metric as several authors deemed the presence of greenery an important aspect for city neighbourhoods. Jacobs, for example, argued that parks and garden positively affected city liveability. However, she also pointed out that they could potentially have negative effects, particularly in terms of safety, if these were relegated to peripheral areas with low densities [4].
- Irregularity of the Street Network (ISN). It measures to what extent the street network of a specific city area is irregular. ISN is computed by dividing the standard deviation of the node degrees associated with the intersections lying within an area by the average node degree relative to the same intersections. Intuitively, a small ISN value reflects an area with a small variation in node degrees, for example, an area characterised by a grid layout, where the majority of street intersections are four-way ones. Conversely, a great ISN value corresponds to an area with a greater variation in node degrees, for example, that of an area characterised by a mix of different street intersections (e.g., cul-de-sac, three-way intersections, four-way intersections, six-way intersections). We consider this metric as the configuration of the street network was another aspect deemed important for city liveability by several authors. For example, Jacobs generally favoured the grid layout [4]. However, she also argued that this had to be interrupted by squares or diagonal roads as, in her view, these urban elements would have offered "visual interruptions" that enhanced urban life.
- Dead-end Density (DD). It measures the density of dead-end roads (cul-de-sac) in a specific city area. It is calculated as the ratio between the number of cul-de-sac lying in an area and the extension (in square meters) of such area. We chose this metric as the presence of dead-end roads in neighbourhoods

attracted the attention of several authors. On the one hand, Jacobs argued that cul-de-sac negatively affected urban liveability as they diminished connectivity and thus the positive effects linked to it (e.g., better health and socio-economic outcomes, more safety against crime) [4]. On the other, Newman supported cul-de-sac as a reduced connectivity would have had positive outcomes, especially in terms of safety, as fewer strangers and more locals would have walked in neighbourhoods [33].

- Offering Advantage of Historic Properties ($OAHP$). It quantifies whether a city area offers more or less historic properties (i.e., built pre-1900) than the average city area. 1900 is selected as temporal threshold as, from roughly that point on, the modernist style started to unfold [34] and thus properties built in the period following this year cannot be considered historic. Note that this threshold is mainly valid for the European context. To compute this metric, we use a formula adopted in a previous work (i.e., Offering Advantage) [35] that showed to be effective in capturing variations in the offering of urban elements across a city. In this paper, Offering Advantage is adapted to reflect to what extent a city area a_k offers more historic properties h_i, compared to the average area. More specifically:

$$OA(h_i, a_k) = \frac{count(h_i, a_k)}{\sum_{j=1}^{N} count(h_j, a_k)} \cdot \frac{\sum_{j=1}^{N} count(h_j)}{count(h_i)}$$

where $OA(h_i, a_k)$ corresponds to the Offering Advantage of historic properties h_i in the area a_k; $count(h_i, a_k)$ represents the number of historic properties h_i in the area a_k; N is the total number of historic properties; $count(h_i)$ is the number of historic properties h_i across a whole city. $OAHP$ can be considered a proxy for the traditional urban form. The more an area offers historic properties, the more likely is that such area is characterised by features of the traditional compact city form (e.g., density, connectivity, perimeter blocks). We include $OAHP$ as compactness and distribution of built density were deemed fundamental aspects that affected urban life by several authors. Jacobs, for example, supported the traditional compact city form as, in her view, it enhanced social tights, commercial activities, and safety [4]. Modernist architects, such as Le Corbusier, on the other hand, despised such city form as they saw it as overly dense and unhealthy, and proposed more dispersed urban plans [5].

Socio-economic indexes

Apart from quantitatively capturing urban form by means of the nine metrics above mentioned, our method also requires access to an index that captures the socio-economic levels of the area under study. Such indexes are ready available for many countries around the world. Although they differ in how they are computed, most of them are based on the concept that wealth or poverty are not caused only by economic factors (e.g., income, employment) but also by other aspects of life (e.g., education, health) and are thus composite. Socio-economics indexes are usually computed at a fine level of spatial granularity in developed countries (though, they tend

to be coarser in developing ones). In England and Wales, there exists, for example, the Index of Multiple Deprivation (IMD), [36] which is computed by weighting seven different domains (i.e., income, employment, education, health, crime, barriers to housing and services, living environment), for small census areas of approximately 1,500 residents. In developing countries, there exists the Multidimensional Poverty Index (MPI), [37] which is calculated as a weighted mean of three macro domains (i.e., health, education, and standard of living) at household level.

Analytical approach

Having illustrated the metrics that our method requires, we now present the analytical approach. This is based on spatial linear regression, as it allows to directly compare and interpret regression coefficients, and thus to measure to what extent our metrics of urban form can explain of the variance of the socio-economic index relative to one another. The analytical approach consists of a four-step process: (i) selection of the areal unit of analysis and computation of the metrics of urban form and socio-economics for such unit; (ii) normalisation and standardisation of the metrics to meet the assumption of linear regression and obtain comparable regression coefficients; (iii) test for collinearity to avoid overinflated regression coefficients and unexpected signs; (iv) test for spatial autocorrelation to check for this issue and use of spatial regression in case the phenomenon is present. We follow with more details next.

Spatial unit of analysis

The spatial unit of analysis is the basic geographic entity for which the metrics presented above are computed. While there is no systematic method to select such unit, two considerations should be taken into account when choosing one. First, given that some of the metrics are based on the count of street intersections, the spatial unit should be big enough to contain some of these elements. For example, a unit of analysis comprising of a 500 meters by 500 meters block might not be suited for the approach proposed, as it might not have any intersection within its boundary. Second, official units that existed for a long period of time are generally better suited than, for example, more grid-shaped ones. Historical boundaries, in fact, tend to keep the morphological unity of neighbourhoods, for example, by not cutting buildings or blocks. This provides metrics that better reflect what exists in the real world. Once the spatial unit of analysis is selected, metrics of urban form and socio-economics should be computed and aggregated for such unit.

Normalisation and standardisation

Linear regression requires that candidate variables are normally distributed. Normalising the metrics is thus necessary to meet this assumption. This can be achieved in different ways, depending on how the values of the metrics are distributed. Common normalisation techniques are exponentiation and logarithmic transformation. Standardisation is required because the metrics have different magnitudes: some are measures of density (e.g., Population Density, Dead-end Density), some others are percentages (e.g., Percentage of Unbuilt Land, Percentage of Green Areas). If these were regressed untransformed against the socio-economic index, their relative

regression coefficients would be hard to compare and interpret. To avoid this, our method requires the computation of the standard scores (or z scores) associated with the normalised metrics. This can be achieved through the following formula:

$$z = \frac{X - \mu}{\sigma}$$

where X represents the metric raw value, μ is the mean value of such metric, and σ its standard deviation.

Test for collinearity and linear model

It is possible that two or more candidate variables (the metrics of urban form) show collinearity, that is they are strongly correlated. If strongly collinear variables are used in a regression, it is likely that their relative regression coefficients would be inflated or show unexpected signs. Since this approach is based on the interpretation of such coefficients, it is necessary to detect and discard strongly collinear variables. This can be achieved through the computation of the Variance Inflation Factors (VIFs) associated with each candidate variable. Let reg be a regression model with predictor variables v_1, v_i, ..., v_n. The VIF of the variable v_i is obtained by, first, performing linear regression with v_i as dependent variable and the other variables as independent ones v_1, $v_i - 1$, $v_i + 1$, ..., v_n, and, second, by using the overall model fit (i.e., R^2 value) obtained at the previous step in the following formula:

$$VIF = \frac{1}{1 - R^2}.$$

If a variable has a strong linear relation with at least another one, its correlation coefficient is likely to be close to 1 and the VIF related to that variable large. A VIF equal to or greater than 10 is a sign of a collinearity issue [38]. If the candidate variables show VIFs smaller than 10, they can be regressed against the socio-economic index. Conversely, if the candidate variables have VIFs equal to or greater than 10, it is necessary to implement a stepwise procedure that, first, excludes the candidate variable with the highest VIF and, second, repeats the same process until none of the variables has a VIF equal to or greater than 10. At the end of this procedure, the candidate variables should be devoid of collinearity and can be regressed against the socio-economic index.

Once obtained the linear model has been obtained, our method requires to check for spatial autocorrelation. This phenomenon occurs when observations located near one another are correlated or, as Tobler put it: 'everything is related to everything else, but near things are more related than distant things' [39]. Not considering this special dependency in linear models can cause over-inflated regression coefficients or unexpected signs. To check for this issue, our method relies on a renowned technique in spatial studies called Moran's test [40]. This checks whether the residuals of a regression analysis are spatially autocorrelated. The outputs of the Moran's test are an index I and a p-value. The former varies between -1 and 1, and can be interpreted similarly to a Pearson's correlation coefficient. The latter measures the statistical significance of the test. A negative Moran's I means that dissimilar values cluster together thus forming a dispersed pattern. Conversely, a positive Moran's

I means that similar values are located near one another thus forming a clustered pattern. If the Moran's test is not significant (i.e., there is no statistical evidence that the residuals are spatially autocorrelated), the linear model can be trusted and interpreted. Conversely, if the Moran's test were to be significant (i.e., there is statistical evidence of the presence of spatial autocorrelation in the residuals), our method requires the use of a spatial model that incorporates the overlooked spatial information. We propose the use of the Spatial Autoregressive (SAR) model, a type of spatial model that accounts for the proximity of observations in space by including a spatial weighting matrix in the equation [41]. To ascertain that the SAR model accounts for all the spatial autocorrelation present in the data, our method requires to perform again the Moran's test. If the outputs are negative, the SAR model can be accepted and interpreted. Conversely, if they are positive, it should be rejected.

We present next the application of the proposed method to six UK cities.

Application of the method

We applied the method presented above to UK urban areas. To do so, we first identified what areas are considered urban. We extracted such information from an official document called *Rural Urban Classification* [12]. In this document, areas were classified in ten classes depending on their level of "urbanity", with the most rural category being *Hamlets and Isolated Dwellings in a Sparse Setting* and the most urban being *Major conurbation*. We chose the areas classified as *Major conurbation* for this analysis. The resulting urban areas corresponded to the cities of London, Manchester, Liverpool, Birmingham, Newcastle, and Leeds. These vary quite substantially both in socio-economic, historic, and cultural terms and in size. London covers the vastest surface and is the most populated, while Newcastle is the smallest and least populated of the set. We provide more information on the six urban areas under study in Table 1 and a map with their locations in Figure 1.

Table 1 Population and extension of the six cities under study. Source: UK Census 2011 [42].

City	Population	Surface (ha)
London	8,173,941	229,546
Birmingham	2,736,460	94,661
Manchester	2,682,528	144,284
Leeds	2,226,058	94,315
Liverpool	1,381,189	52,267
Newcastle	1,104,825	57,127

The first step of our method consists in the computation of the nine metrics of urban form and the socio-economic index at a *suitable* spatial unit of analysis. We selected the *ward* as spatial unit of analysis for the present study, for two reasons. First, their spatial extension was never too small to cause issues in the computation of the metrics, yet it was small enough to offer city planners fine grained units of analysis and possible intervention. Second, wards are long standing administrative boundaries defined by the UK government, which have both electoral and ceremonial functions, and were first implemented in the Middle Ages [43]. We identified 847 wards for London, 238 for Manchester, 183 for Birmingham, 119 for Newcastle, 100 for Liverpool, and 93 for Leeds.

Having chosen the spatial unit at which to perform our analysis, we then needed access to openly accessible datasets from which to compute the metrics of urban form and socio-economics. We used five of such datasets: OS VectorMap District, Dwellings by Property Build Period and Type, the 2011 UK Census, OSM, and IMD. We present these datasets next.

Ordnance Survey (OS) VectorMap District. This is one of the official digital maps of the UK [44]. It contains information on various geographic objects such as roads, building footprints, and natural areas. The content of this dataset is generated and kept updated by Ordnance Survey, the UK official mapping agency, and was made freely accessible for the first time in 2010. The geographic information is provided in vectorial format for tiles of 100 km by 100 km. We selected the tiles corresponding to the six urban areas under study and extracted the information needed for the computation of the metrics. More specifically, we computed the degrees of each node (i.e., street intersection) in the street networks and the areas occupied by buildings in the six cities under study. This information was then used to compute six of the nine metrics of urban form at ward level: Connected Node Ratio (CNR), Intersection Density (ID), Dead-end Density (DD), Irregularity of the Street Network (ISN), Percentage of Unbuilt Land (PUL), and Betweenness Centrality (BC).

Dwellings by Property Build Period and Type. This database contains the count of properties in England and Wales for several build periods and housing types [45]. To be more specific, the information on build periods is subdivided in twelve classes of around ten years each, with the first being the class with properties built before 1900 and the last being the one with properties built between 2010 and 2015. This information is provided for official census areas of about 1,500 residents, the Lower-layer Super Output Areas (LSOAs). However, LSOAs are much smaller than wards (indeed too small for our analysis). The information on build periods was thus aggregated at the level of wards by summing the values associated with the LSOAs contained in each ward. This information was then used to compute the metric Offering Advantage of Historic Properties ($OAHP$).

2011 Census: Population and Household Estimates for Wards and Output Areas in England and Wales. This database contains information on the population density (i.e., persons per hectare) of each ward in England and Wales [42]. This data was used to compute the metric Population Density (PD).

OpenStreetMap. With more than three million users, OSM is probably the best known example of geographic crowd-sourcing [46]. OSM contributors are collectively building and keeping updated the first free and editable map of the world. Many studies have been carried out to ascertain the quality of its content in different parts of the world, for example in the UK [47], France [48], and Germany [49], and reported an overall good level of spatial accuracy, especially in urban areas. For the purpose of this analysis, we used the OMS dataset as source of information for public green areas. Once these were identified, we assigned them to each of the wards of the urban regions under study and computed their areas. This information was then used to compute the metric Percentage of Green Areas (PGA).

Table 2 Metrics of urban form and IMD with relative descriptions, means, and standard deviations.

Variable name	Description	Mean	SD
Connected Node Ratio (CNR)	Level of connectivity of the street network	0.743703	0.081798
Intersection Density (ID)	Density of street intersections	0.619632	0.328992
Percentage of Unbuilt Land (PUL)	Proportion of land left unbuilt	73.234177	12.303271
Population Density (PD)	Density of city dwellers	62.762976	43.150516
Betweenness Centrality (BC)	Level of accessibility	0.027086	0.040125
Percentage of Green Areas (PGA)	Amount of green areas	7.929976	8.185188
Irregularity of the Street Network (ISN)	Level of irregularity of the street network	0.365015	0.058786
Dead-end Density (DD)	Density of dead-end roads (cul-de-sac)	0.200358	0.111242
Offering Advantage of Historic Properties ($OAHP$)	Weighted offering of historic properties	0.904616	0.994179
Index of Multiple Deprivation (*IMD*)	Socio-economic deprivation	24.162269	13.067144

Index of Multiple Deprivation. This dataset includes information on the socio-economic deprivation of communities of England and Wales [36]. It is computed for LSOAs by weighting seven different domains: income deprivation, employment deprivation, health deprivation, education deprivation, barriers to housing and services, crime levels, and living environment deprivation. IMD is provided as an aggregate score: higher values correspond to more deprived areas, lower values to more advantaged ones. Since the spatial unit adopted in this analysis was the ward, we aggregated IMD values at the level of such unit by averaging the IMD values associated with the LSOAs contained in each ward. This was possible since the variation of IMD scores of the LSOAs contained in each ward was small (i.e., their standard deviation values were small and always smaller than their average values). This data constituted the metric of socio-economic deprivation (IMD).

We present a summary table with metrics, relative brief descriptions, means, and standard deviations in Table 2.

Results

In this section, we present first some preliminary results drawn from the observation of the density distribution plots of the metrics of urban form and IMD. Second, we illustrate the outcomes of the regression analyses performed for each city. Third, we offer interpretations for the behaviours of the regression coefficients.

Preliminary Results

We computed frequency distribution plots for each metric under study, to explore what data we were dealing with and what it meant in terms of urban form. We present these plots, for each city under study, in Figure 2. As shown, none of the metrics were normally distributed, with only Connected Node Ratio (CNR) showing a distribution close to the normal. The majority of the metrics (six out of nine) showed positive skews, having most of their values close to their respective first quartiles. These were: Percentage of Green Areas (PGA), Intersection Density

(ID), Dead-end Density (DD), Population Density (PD), Betweenness Centrality (BC), and Offering Advantage of Historic Properties ($OAHP$). Irregularity of the Street Network (ISN) and Percentage of Unbuilt Land (PUL) showed negative skews, instead, with most of their values concentrating around their respective third quartiles. IMD also presented a non-normal distribution (i.e., positive skew).

We drew some observations from these preliminary results. Most metrics of urban form had similar distributions across the six cities under study. This was not surprising since the metrics were computed for regions within the same country, which thus had been subject to similar historic, social, and cultural phenomena. However, there were also some unique variations. First, the urban form of London seemed to be denser in terms of built form and population compared to the other cities (i.e., more low values of PUL, more high values of PD). Moreover, it seemed to have a better connected street network (i.e., more high values of CNR, more low values of DD) and being less deprived (i.e., more low values of IMD). Second, Birmingham's urban features showed peaks of values rather than more varied distributions. In particular, most of its neighbourhoods seemed to have low values of ID (around 0.5), moderately high values of PUL (around 70%), and low values of $OAHP$ (close to 0). Third, Newcastle seemed to offer less green areas than the other cities. The majority of its neighbourhoods, in fact, showed values of PGA close to 0. Finally, Leeds seemed to be more sparsely built (i.e., more high values of PUL) and offer more historic properties (i.e., more high values of $OAHP$).

Linear models

After having normalised and standardised the variables of urban form, we checked whether they were collinear through the VIF test. Outcomes of this test indeed showed that some of the variables presented collinearity. In particular, CNR and ID were strongly collinear in all cities, with VIFs significantly greater than 10. PUL was found to be collinear only in Leeds. Such variables were thus discarded from the list of candidates for the regression analysis.

We then input the remaining variables in six regression models, one for each city, with IMD as dependent variable. Model outcomes suggested that multiple features of urban form were associated with deprivation. Models were all statistically significant, at 99% confidence level, and generally presented moderate fits, with four models out of six being able to explain around 50% of the variance of IMD. To be more specific, urban form could explain 50% of the variance of deprivation in Birmingham and Leeds; it could explain 49% of the variance in London and 48% in Manchester. The explanatory power of the models for Liverpool and Newcastle was lower, instead. Urban form could explain 25% of the variance of IMD in the first city and only 11% in the second.

To check whether spatial autocorrelation was not affecting these outcomes, we performed the Moran's test on the residuals. Outputs of such test showed statistical evidence of the presence of spatial autocorrelation in all models. Moran's I values were statistically significant, at 99% confidence level, with a minimum value of 0.16 (Manchester, Leeds, and Newcastle) to a maximum of 0.44 (London). Given the presence of spatial autocorrelation, which could have biased regression coefficients and overall model fit, we implemented the SAR technique to account for such

phenomenon. SAR models were all statistically significant, at 99% confidence level, and showed greater explanatory powers and smaller regression coefficients, meaning that part of IMD was indeed explained by the spatial proximity of observations. To be more specific, the model for Birmingham could explain 67% of the variance of IMD, the one for London could explain 70%, the one for Manchester 56%, the one for Leeds 59%, the one for Liverpool 49%, and the one for Newcastle 27%. We performed a second Moran's test to ascertain whether spatial autocorrelation did not affect the residuals of the SAR models. Outputs of such test confirmed that there was no statistical evidence of the presence of the issue in none of the models (i.e., p-values > 0.05). Full results of the linear regression (LR) and SAR models can be found in Table 3. Knowing that models were robust, we proceeded to investigate their relative regression coefficients. Common patterns are summarised below:

- Dead-end Density (DD) was statistically significant and positively associated with deprivation, in five cities out of six (i.e., Birmingham, London, Manchester, Liverpool, and Newcastle);
- Irregularity of the Street Network (ISN) was significant and negatively associated with deprivation, in four cities out of six (i.e., Birmingham, Manchester, Liverpool, and Newcastle);
- Percentage of Unbuilt Land (PUL) was significant and positively associated with deprivation, in four cities out of six (i.e., London, Manchester, Liverpool, and Newcastle);
- Population Density (PD) was significant and positively associated with deprivation, in four cities out of six (i.e., Birmingham, London, Manchester, and Leeds).

For what concerned the remaining coefficients, Betweenness Centrality (BC) was associated with deprivation in two cities out of six (i.e., London and Leeds), Offering Advantage of Historic Properties ($OAHP$) was related to advantaged areas in London only, while Percentage of Green Areas (PGA) was negatively associated with deprivation in Newcastle only. We elaborate more on these findings next.

Interpretations

As mentioned above, several regression coefficients (i.e., DD, ISN, PUL, and PD) showed similar patterns across the cities under study. This meant that we could identify an urban form that was associated with deprivation at country level. In particular, deprived English neighbourhoods appeared to be characterised by high population density, vast portions of unbuilt land, numerous cul-de-sac, and regular street patterns. This seemed to closely resemble the modernist "towers in the park" design scheme, which saw the concentration of residents in small portions of land (i.e., residential towers laid out in open space), conspicuous presence of dead-end roads, and regular street patterns of major roads [5, 15].

The link between these urban features and deprivation thus seemed to discredit the modernist theories and support the ones of the compact city form [4, 13, 14], for the UK context. Jacobs, for example, was in favour of perimeter blocks rather than isolated residential towers as the retraction of buildings from side walks diminished social interactions, as fewer points of exchange (e.g., doors, windows, porticoes) between buildings and streets were present [4]. Furthermore, they reduced commercial

Table 3 LR and SAR models for the six urban areas under study with outcomes of the Moran's test. P-value symbols correspond to: 0 '*' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1**

	London				Manchester			
	LR		SAR		LR		SAR	
	p-val.	β	p-val.	β	p-val.	β	p-val.	β
(intercept)		0.00		-0.01		0.00		-0.02
CNR		-		-		-		-
ID		-		-		-		-
PUL	***	0.36	***	0.27	***	0.91	***	0.78
PD	***	1.06	***	0.60	***	0.35	***	0.27
BC	**	0.10	.	0.04		-0.05		-0.07
PGA		0.01		-0.01		0.06		0.04
ISN	*	0.11		0.03	***	-0.53	***	-0.46
DD		0.02	*	0.08	***	0.96	***	0.85
$OAHP$	***	-0.17	***	-0.14		-0.06		-0.05
adj. R^2		**0.49**		**0.70**		**0.48**		**0.56**
p-value		0.00		0.00		0.00		0.00
Moran's		0.04		0.03		0.16		0.02
p-value		0.17		0.08		0.00		0.26

	Birmingham				Leeds			
	LR		SAR		LR		SAR	
	p-val.	β	p-val.	β	p-val.	β	p-val.	β
(intercept)		0.00		-0.05		0.00		-0.01
CNR		-		-		-		-
ID		-		-		-		-
PUL		0.12		0.13		-		-
PD	***	0.45	***	0.35	**	0.44	**	0.37
BC	**	0.19		0.06	**	0.21	*	0.17
PGA	*	0.14		0.07		-0.10		-0.07
ISN	*	-0.23	*	-0.17		-0.10		-0.10
DD	*	0.24	*	0.18	.	0.23		0.19
$OAHP$	.	0.11		0.06		-0.04		-0.05
adj. R^2		**0.50**		**0.67**		**0.50**		**0.59**
p-value		0.00		0.00		0.00		0.00
Moran's		0.30		0.03		0.16		-0.02
p-value		0.00		0.21		0.00		0.59

	Liverpool				Newcastle			
	LR		SAR		LR		SAR	
	p-val.	β	p-val.	β	p-val.	β	p-val.	β
(intercept)		0.00		-0.02		0.00		0.01
CNR		-		-		-		-
ID		-		-		-		-
PUL	.	0.40	*	0.36	*	0.60	*	0.50
PD	.	0.33		0.18		0.10		0.07
BC		0.14		0.07	.	0.16		0.13
PGA		-0.06		-0.02	*	-0.20	.	-0.14
ISN	.	-0.28	.	-0.26	.	-0.38	.	-0.34
DD	**	0.66	**	0.58	**	0.62	**	0.53
$OAHP$		-0.11		-0.12		0.07		0.05
adj. R^2		**0.25**		**0.49**		**0.11**		**0.27**
p-value		0.00		0.00		0.01		0.00
Moran's		0.25		0.01		0.16		-0.02
p-value		0.00		0.37		0.00		0.56

activity, as there was no physical space on the sides of streets for amenities. Finally, isolated tower blocks also reduced safety, as streets were not informally controlled by windows facing them (the so-called "eyes on the street" effect). Similarly, she also supported well-connected streets rather than cul-de-sac, as the latter diminished connectivity and thus also the ability of pedestrians to move in the urban space. This aspect, in her view, was not only fundamental for the vitality of urban spaces but also for their economic prosperity and safety. She argued that fewer pedestrians corresponded to fewer social interactions, smaller use of amenities, and fewer "eyes on the street" to prevent crime. Finally, she supported street networks with irregularities (e.g., diagonal roads and squares) rather than overly regular grids. This aspect, in her view, would have provided "visual interruptions", which enhanced the perception of space and, ultimately, urban life.

The link between presence of cul-de-sac and deprivation seemed also to discredit the theory proposed by Newman (i.e., cul-de-sac were beneficial against crime as they reduced passage of people thus making urban spaces more controllable [33]). Although we did not test a pure measure of crime, a domain associated with such topic was included in the computation of IMD.

Three more associations were found to be significant, although not across all cities. The relationship between BC and deprivation seemed to be linked to the detrimental effects of too much accessibility, which a recent study found to be associated with more road traffic [50], on people's well-being. These negative effects, in fact, could be associated with more air and noise pollution, more congestion, and higher levels of stress. The inverse relationship between $OAHP$, our proxy for the traditional urban form, and deprivation seemed to be backed up by theories of the compact city form [4, 13, 14]. As we mentioned earlier, aspects of the traditional urban form (e.g., density, connectivity, perimeter blocks) were deemed fundamental for social, economic, and safety reasons. Finally, the negative relationship between PGA and deprivation in Newcastle could be linked to the beneficial effects of urban greenery on the well-being of residents. This was supported, for example, by the study of Maas *et al.* who found that higher percentages of urban green were associated with higher scores of perceived health [51].

Limitations

We ought to acknowledge some limitations for this work. First, urban form is not the only factor influencing the socio-economic levels of city areas. Many other aspects are at stake, for example, specific housing policies, economic interventions, gentrification. While it would be impossible to account for all of these different factors in one model, research outcomes from other fields (e.g., demography, econometrics) can be used to contextualise and interpret the results provided by the application of our method. Second, the proposed approach and relative outcomes do not imply causation. For instance, this means that, although one may find that connectivity is associated with better socio-economic outcomes, increasing the connectivity of a neighbourhood might not necessarily bring actual improvements of the socio-economic conditions of the resident population. Third, the results found for the British cities cannot be generalised as they only hold for the specific geographic regions (i.e., the six cities under study) and time frame (i.e., 2015) investigated

in this work. Nonetheless, one can use the very same method to test larger areas and different time frames and thus extend generalisability. Fourth, the selection of the spatial unit of analysis inevitably comes with the issue of the Modifiable Areal Unit Problem [52]. This states that values of metrics can vary quite substantially if computed for different spatial units. This can clearly bias the outcomes of a spatial analysis. While there is no systematic method to address this issue, one should be aware of this problem and eventually test different units of analysis. Finally, the proposed method models the relation between aspects of urban form and socio-economics in a linear fashion. However, it is possible that such relationships are not linear. For example, interactions between metrics of urban form might be better suited to explain socio-economic levels than the metrics taken separately. This warrants a separate future investigation.

Conclusions

With the world undergoing a process of fast urbanisation, inequality is on the rise as some areas are benefiting more than others of public fundings and international investments. Analysing the relationship between urban form and socio-economics has thus become urgent as it can assist planners and policy makers when debating how to design cities and where to allocate resources. In this paper, we proposed a quantitative method to analyse such relationship at scale through spatial linear regression. More specifically, the method extracts metrics of urban form and socio-economics from openly accessible datasets. It then identifies, through regression analysis, what set of urban features are associated with socio-economic levels of city areas. When applied to the major UK cities, the method found that urban form could explain up to 70% of the variance of IMD, an official deprivation index. We also observed that some specific regression coefficients showed common patterns across the cities under study: high population density, vast portions of unbuilt land, presence of cul-de-sac, and regular street patterns were all related to deprivation. By connecting these findings to previous works, we argued that the relationship between this specific combination of urban features and deprivation discredited modernist theories and supported theories of the traditional city form, in the UK case.

img/6cities

Figure 1 Location of the six urban areas under study.

img/final

Figure 2 Density distribution plots for the metrics of urban form and IMD, for the six cities. BR: Birmingham. LN: London. MA: Manchester. LE: Leeds. LI: Liverpool. NC: Newcastle.

Declarations

Competing interests

The authors declare that they have no competing interests.

List of abbreviations
UK: United Kingdom IMD: Index of Multiple Deprivation
OSM: OpenStreetMap
OS: Ordnance Survey
GIS: Geographic Information System
CNR: Connected Node Ratio
ID: Intersection Density
PUL: Percentage of Unbuilt Land
BC: Betweenness Centrality
PGA: Percentage of Green Areas
DD: Dead-end Density
OAHP: Offering Advantage of Historic Properties
MPI: Multidimensional Poverty Index
LSOA: Lower-layer Super Output Area
VIF: Variance Inflation Factor
SAR: Spatial Autoregressive model
LR: Linear Regression

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Availability of data and materials
The normalised and standardised version of the data used in this research is provided in the supporting files.

Funding
The research presented in this paper has been developed thanks to the Engineering and Physical Sciences Research Council (EPSRC) grant number EP/G037698/1.

Author's contributions
Main author: Alessandro Venerandi. Co-authors: Giovanni Quattrone, Licia Capra.

Acknowledgements
Not applicable.

Endnotes
Not applicable.

Author details
[1]Department of Civil, Environmental & Geomatic Engineering, UCL, Gower Street, London, UK. [2]Department of Computer Science, UCL, Gower Street, London, UK.

References
1. Department of Economic and Social Affairs: World urbanization prospects, the 2014 revision: highlights. Technical report, Population Division United Nations (2014)
2. Overseas Development Institute: Briefing paper 44: Opportunity and exploitation in urban labour markets. Technical report (2008)
3. Moreno, E.L., Bazoglu, N., Mboup, G., Warah, R.: State of the world's cities 2008/2009 - harmonious cities. Technical report, UN-HABITAT (2008)
4. Jacobs, J.: The Life and Death of Great American Cities. New York: Random House, ??? (1961)
5. Corbusier, L.: The City of Tomorrow and Its Planning. London: Architectural Press, ??? (1947)
6. Vaughan, L., Clark, D.L.C., Sahbaz, O., Haklay, M.M.: Space and exclusion: does urban morphology play a part in social deprivation? Area **37**(4), 402–412 (2005)
7. Hillier, B.: Can streets be made safe? Urban design international **9**(1), 31–45 (2004)
8. Mayer-Schönberger, V., Cukier, K.: Big Data: A Revolution that Will Transform How We Live, Work, and Think. Houghton Mifflin Harcourt, ??? (2013)
9. Kitchin, R.: The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences. Sage, ??? (2014)
10. Salesses, P., Schechtner, K., Hidalgo, C.A.: The collaborative image of the city: mapping the inequality of urban perception. PloS one **8**(7), 68400 (2013)
11. Quercia, D., O'Hare, N.K., Cramer, H.: Aesthetic capital: what makes london look beautiful, quiet, and happy? In: Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing, pp. 945–955 (2014). ACM
12. Department for Environment, Food & Rural Affairs: The 2011 Rural Urban Classification For Small Area Geographies. Technical report, Office for National Statistics (2013)
13. Whyte, W.H.: City: Rediscovering Its Center. New York: Doubleday, ??? (1988)
14. Gehl, J.: Cities for People. Island press, ??? (2013)
15. Hilberseimer, L.: The New City: Principles of Planning. Theobald, ??? (1944)
16. Hillier, B.: Space Is the Machine: a Configurational Theory of Architecture. Space Syntax, ??? (2007)
17. Hillier, B., Sahbaz, O.: An evidence based approach to crime and urban design, or, can we have vitality, sustainability and security all at once. R. Cooper, G. Evans, C. Boyko, C.(eds.), Designing Sustainable Cities: Decision-making Tools and Resources for Design, Wiley Blackwell, 163–186 (2009)

18. Harries, K.: Property crimes and violence in United States: an analysis of the influence of population density. International Journal of Criminal Justice Sciences **1**(2), 24–34 (2006)
19. Haughey, R.M.: Higher-density Development: Myth and Fact. ULI–Urban Land Institute, ??? (2005)
20. Duany, A., Plater-Zyberk, E., Speck, J.: Suburban Nation: The Rise of Sprawl and the Decline of the American Dream. Macmillan, ??? (2001)
21. Burchell, R.W., Shad, N.A., Listokin, D., Phillips, H., Downs, A., Seskin, S., Davis, J.S., Moore, T., Helton, D., Gall, M.: The Costs of Sprawl-revisited vol. Project H-10 FY'95, (1998)
22. Bridge, G.: The neighbourhood and social networks. cnr paper **4**, 1–32 (2002)
23. Georg, S.: The metropolis and mental life. Metropolis: Centre and Symbol of Our Times. London: Macmillan (1995)
24. Freeman, L.: The effects of sprawl on neighborhood social ties: An explanatory analysis. Journal of the American Planning Association **67**(1), 69–77 (2001)
25. Muratori, S.: Studi per Una Operante Storia Urbana di Venezia vol. 1. Instituto poligrafico dello Stato, Libreria dello Stato, ??? (1959)
26. Conzen, M.R.G.: Alnwick, Northumberland: a study in town-plan analysis. Transactions and Papers (Institute of British Geographers) (27), 122 (1960)
27. Marshall, W., Garrick, N.: The shape of sustainable street networks for neighborhoods and cities. In: Congress for the New Urbanism XVII, Denver, CO (2009)
28. Banister, D., Watson, S., Wood, C.: Sustainable cities: transport, energy, and urban form. Environment and Planning B: planning and design **24**(1), 125–143 (1997)
29. Porta, S., Crucitti, P., Latora, V.: The network analysis of urban streets: a primal approach. Environment and Planning B: planning and design **33**(5), 705–725 (2006)
30. Wang, F., Antipova, A., Porta, S.: Street centrality and land use intensity in Baton Rouge, Louisiana. Journal of Transport Geography **19**(2), 285–293 (2011)
31. Timothée, P., Nicolas, L.-B., Emanuele, S., Sergio, P., Stéphane, J.: A network based kernel density estimator applied to Barcelona economic activities. In: International Conference on Computational Science and Its Applications, pp. 32–45 (2010). Springer
32. Remali, A.M., Porta, S., Romice, O.: Correlating street quality, street life and street centrality in Tripoli, Libya. The Past, Present and Future of High Streets (2014)
33. Newman, O.: Defensible Space. Macmillan New York, ??? (1972)
34. Curtis, W.J.: Modern Architecture Since 1900 vol. 2. Phaidon London, ??? (1996)
35. Venerandi, A., Quattrone, G., Capra, L., Quercia, D., Saez-Trumper, D.: Measuring urban deprivation from user generated content. In: Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing, pp. 254–264 (2015). ACM
36. Smith, T., Noble, M., Noble, S., Wright, G., McLennan, D., Plunkett, E.: The English indices of deprivation 2015. London: Department for Communities and Local Government (2015)
37. Alkire, S., Conconi, A., Roche, J.M.: Multidimensional Poverty Index 2012: Brief Methodological Note and Results. University of oxford, Department of International Development, oxford Poverty and Human Development Initiative, oxford, UK (2012)
38. Chatterjee, S., Hadi, A.S.: Regression Analysis by Example. John Wiley & Sons, ??? (2015)
39. Tobler, W.R.: A computer movie simulating urban growth in the Detroit region. Economic Geography **46**, 234–240 (1970)
40. Cliff, A.D., Ord, J.K.: Spatial Processes: Models & Applications. Taylor & Francis, ??? (1981)
41. Haining, R.P.: Spatial Data Analysis: Theory and Practice. Cambridge University Press, ??? (2003)
42. 2011 Census: Population and Household Estimates for Wards and Output Areas in England and Wales. http://tinyurl.com/jwwcz95. Accessed 4 May 2017.
43. Borer, M.I.C.: The City of London: A History. Constable & Company Limited, ??? (1977)
44. OS VectorMap District. http://tinyurl.com/jxrhcqw. Accessed 4 May 2017.
45. Dwellings by Property Build Period and Type, LSOA and MSOA. http://tinyurl.com/m3we5sn. Accessed 4 May 2017.
46. OpenStreetMap. www.openstreetmap.org. Accessed 4 May 2017.
47. Haklay, M.: How good is volunteered geographical information? A comparative study of OpenStreetMap and Ordnance Survey datasets. Environment and planning B: Planning and design **37**(4), 682–703 (2010)
48. Girres, J.-F., Touya, G.: Quality assessment of the French OpenStreetMap dataset. Transactions in GIS **14**(4), 435–459 (2010)
49. Ludwig, I., Voss, A., Krause-Traudes, M.: A Comparison of the Street Networks of Navteq and OSM in Germany. In: Advancing Geoinformation Science for a Changing World, pp. 65–84. Springer, ??? (2011)
50. Xiao, Y.: Urban morphology and housing market. PhD thesis, University College London (2012)
51. Maas, J., Verheij, R.A., Groenewegen, P.P., De Vries, S., Spreeuwenberg, P.: Green space, urbanity, and health: how strong is the relation? Journal of epidemiology and community health **60**(7), 587–592 (2006)
52. Openshaw, S., Taylor, P.J.: A million or so correlation coefficients: three experiments on the modifiable areal unit problem. Statistical applications in the spatial sciences **21**, 127–144 (1979)