Land Use Policy 105 (2021) 105425

Contents lists available at ScienceDirect

Land Use Policy

journal homepage: www.elsevier.com/locate/landusepol

Where the city lights shine? Measuring the effect of sprawl on electricity consumption in Spain

Alfredo Cartone [a], Alberto Díaz-Dapena [b], Raquel Langarita [c], Fernando Rubiera-Morollón [b,*]

[a] Department of economic studies, University G. d'Annunzio of Chieti-Pescara, Pescara, Italy
[b] Regional Economics Laboratory of the University of Oviedo, Oviedo, Spain
[c] University of Zaragoza, Zaragoza, Spain

ARTICLE INFO

Keywords:
Electricity
Entropy econometrics
Disaggregation
Urban sprawl
Urban sustainability
Spatial heterogeneity

ABSTRACT

Urban sprawl is a phenomenon that is generally growing across all continents. As a result, modern city structures need larger areas for similar populations. Few studies have evaluated the effects of sprawl on an important aspect in terms of sustainable development: energy consumption. The aim of this paper is to analyse whether urban sprawl has a significant effect on electricity consumption in Spanish municipalities. The increase in sprawl in Spanish cities is heterogeneous, and the growth of household income during recent decades has allowed households to move to scattered residential areas. This situation makes this country especially interesting as a case study to evaluate the impacts of urban sprawl. In this paper, by disaggregating the electricity consumption of households at the local level using entropy, we measure the effect of sprawl to evaluate whether there is an effect on household energy consumption. The joint consideration of disaggregated data and spatial heterogeneity allows us to assess the effect that sprawl has for certain urban configurations on electricity consumption, which points to the need for policies that involve national, regional and local land use policies.

1. Introduction

Over the last few years, the phenomenon of cities covering large areas of available land, known as urban sprawl, has occurred in Europe at a high rate (see, among others, European Comision, 2006; Couch et al., 2007; Christiansen and Loftsgarden, 2011; Oueslati et al., 2015; Hennin et al., 2015; Milan and Creutzig, 2016). In fact, the size of many European cities is increasing faster than their population, evidencing agglomeration in large urban contexts, which tends to take up an increasing rate of land. Nevertheless, although this is not a general phenomenon across Europe, many cities continue to grow compactly, remaining faithful to the traditional European urban model (Patacchini et al., 2009; Travisi et al., 2010). The case of Spain is especially illustrative of this heterogeneity of behaviour with respect to sprawl that is observed in Europe in general. As Rubiera et al. (2016) show, many Spanish cities have drifted towards intense sprawled growth, while others remain faithful to a compact urban model.

Global warming and the great environmental challenges of our time have led researchers to study energy efficiency from several perspectives. This includes, of course, the energy efficiency of homes. Much emphasis has been placed on the study of construction materials that achieve greater energy efficiency; the development of electricity-efficient appliances is also worth investigating. Additionally, an analysis of the impact of the urban form is also part of the research lines reinforced by the search for maximum energy efficiency and lower environmental impact (Itsván, 2010; Wilson and Chakraborty, 2013). The idea of how to design and develop sustainable cities is now at the centre of academia as well as the political debate. Policy makers very often refer to the concept of "smart cities", so energy consumption emerges as a relevant aspect to be addressed. This topic has been recently debated in terms of the effects of urban sprawl. Stiri (2014), Wiesmann et al. (2011), Heionen y Junnila (2014), and Huang (2015), among others, have studied residential energy consumption for a set of housing types, namely, apartment buildings, row-terraced houses, and detached houses. They consider degrees of urbanization (cities, semi-urban and rural), finding that sprawled houses show significant increments in different types of energy consumption for different countries (see, for a review, Rubiera-Morollón and Garrido-Yserte (2020)). Lasarte et al. (2018) extend a study on 17 regions of Spain to determine the effects of urban sprawl on electricity demand via quantile regression, reaching similar conclusions for the Spanish case.

The objective of this paper is to add to the current literature on the

* Corresponding author.
E-mail address: frubiera@uniovi.es (F. Rubiera-Morollón).

https://doi.org/10.1016/j.landusepol.2021.105425
Received 25 September 2020; Received in revised form 18 February 2021; Accepted 19 March 2021
Available online 1 April 2021

effects of urban sprawl on household electricity consumption. We model electricity consumption at the local level, including measured urban sprawl, using georeferenced aerial photos as a control variable. In the model, sprawl is introduced as a factor directly influencing electricity consumption, assuming that the direct effect of rapid sprawl is likely to affect demand in electricity due to the lower efficiency of detached/sprawled houses. Our analysis is specific to the case study of Spain (specifically, we apply the model for 248 cities in Spain). The strong changes in income per capita, social customs, and land use have caused Spain to be rapidly affected by urban sprawl, which justifies investigating more of the consequences expected from sprawl (Rubiera et al., 2016). According to statistical data from the Household Budget Survey (HBS) of the National Statistical Institute (INE in its Spanish acronym), in 2014, approximately 35% of the population of Spain lived in houses, with 11% living in detached houses and 24.2% in semi-detached houses (INE, 2016). The main contribution is to estimate the effect that urban sprawl has on energy consumption by exploiting a robust spatial approach. In this direction, focusing on local models instead of a global one should allow much more accurate urban policies that influence sprawl.

Two major issues arise in our evaluation of the effect of sprawl on electricity consumption in Spain. First, special consideration is paid to the spatial scale, disaggregating data at the municipality level of the analysis to support land use policies at a granular level. To this aim, we adopt a novel technique in the analysis of electricity consumption. Following Tarozzi and Deaton's (2009) small area estimation (SAE), local estimates are made consistent by applying generalized maximum entropy (GME). Second, different urban patterns may cause variations in the impacts of climate, culture, and demographic structures. For this reason, we tackle the potential heterogeneity by applying a regression tree algorithm (Breiman et al., 1984). According to this methodology, we consider groups of municipalities that show homogenous behaviours in terms of electricity consumption. A comparison to another widespread technique is offered to treat spatial heterogeneity, geographically weighted regression (Fotheringham et al., 2002).

The paper is structured as follows. Section 2 includes a brief review on the environmental effects of urban sprawl, with a focus on energy issues. Section 3 describes datasets explaining the methodologies adopted to obtain information about electricity consumption at the local level, as well as how the sprawl index is calculated. Section 4 presents the basics of the estimation strategy used to control the spatial heterogeneity: the regression trees approach. In Section 5, the results are reported, while Section 6 concludes.

2. Effects of the sprawling process on the energy efficiency of cities

The first studies on urban sprawl were more attentive to describing the phenomenon and understanding its causes than to analysing its consequences. However, the debate about the effects of the dispersion of cities soon began. Jenks et al. (1996) presented a comprehensive review of this first stage of the international literature on urban sprawl in which sprawl was basically understood to imply soil predation with consequences on the natural environment of cities and to generate greater distances in the city, which can make mobility more difficult. However, there was no consensus among social scientists about the ultimate impact of the phenomenon. Some researchers argued that urban sprawl was an inevitable consequence of economic growth and the development of societies, which could have positive effects. For example, Ewing (1997) argued that the dispersed city allowed lifestyles that reconnect human beings with nature without having to abandon the advantages of urban life, and Breheny (1996) positively valued the creation of large and polycentric cities, which achieved a more efficient distribution of economic and social activity, while avoiding congestion around a single centre.

The more cities sprawled or did so with greater intensity, the more evidence accumulated on the effects of this urban model we had. In parallel, we have witnessed an extraordinary improvement in the methodologies for capturing information. All of this has given rise to much more precise empirical studies that have clarified the effects of dispersion on various fields. A greater consensus among the scientific community exists after this on the lower social and environmental sustainability of dispersed cities. Two good reviews of the literature on this second phase can be found in Camagni et al. (2002) or in Wilson and Chakraborty (2013).

Already from a position of consensus against the dispersed urban model, although with nuances according to each author, the literature has continued to advance in the last decade paying attention to four main fronts: (i) immediate impact of the sprawl on the environment or the landscape, (ii) sprawl, mobility, and sustainability, (iii) social or economic effects of sprawl, and (iv) sprawl, climate change, energy efficiency, and urban sustainability. These topics are studied in cities around the world, although there has been a growing interest in studying the phenomenon in the emerging economies of Latin America and, mainly, Asia. A complete revision of these more recent studies can be found in Rubiera-Morollón and Garrido-Yserte (2020).

On the direct impact of the extension of the city and the soil predation produced by its dispersion, we can see the works of Hasse and Lathrop (2003), Robinson et al. (2005), Skog and Stennes (2016), Abu Hatab et al. (2019), and Vicenzotti and Qvistrom (2018) that show how dispersion invades landscapes and damages natural environments with great aggressiveness. Slemp et al. (2012) paid attention to the damage to the traditional rural culture that, as they pointed out, is exterminated in environments around the dispersed city. Yan (1995) applies this type of impact analysis in the environment for Chinese cities with similar conclusions. More recently, Yang et al. (2019) update their conclusions using advanced GIS models that allow us to see the growth of urban slick and the simultaneous disappearance of natural spaces.

The second most commonly studied aspect has been related to mobility. Sprawl implies greater distances and lower population density, which hinders the success of collective means of transport, either due to the cost of the infrastructures in large urban areas, in the case of the subway or commuter rail, or due to the number of necessary stops, which makes urban bus systems or the like very inefficient (Camagni et al., 1998; Johnson, 2001). Several studies have analysed how dispersion generates a deterioration in public transport, making mobility dependant on access to private vehicles (Camagni et al., 2002). Going beyond this idea, some works have explored how this greater intensity of private transport in dispersed cities affects the traditional centre (Lee, 2005), generating cumulative processes: greater dispersion implies greater use of private vehicles and deterioration of public transport, with negative effects on the increase in CO_2 emissions. This damages the centre, which, when weakened, loses compacting power, accelerating the dispersion process (Schneider and Woodcock, 2008). Most recent research has also paid attention to the complexity of mobility in dispersed cities in emerging countries in Latin America (see, among others, Coq and Asian, 2019) or Asia (see, among others, Xu et al., 2019).

Another part of the international literature has focused on exploring the relationship between urban sprawl and other socioeconomic aspects, reaching very interesting conclusions. For example, Briggs (2005), among others, has identified that urban dispersion increases social conflict in cities. The coexistence of different social classes, which, in most cases, coincide with ethnic or religious groups, in dispersed cities with damaged main centres facilitates the generation of enormously dangerous urban ghettos. With an analysis from a historical perspective, Axelrod (2007) shows how dispersed cities have greater difficulties integrating immigrants and facilitating the assimilation of new cultures. From another perspective, Florida and Mellander (2015) paid attention to how less integration of urban life generated by dispersed cities with weak centres translates into less capacity to generate the positive effects of agglomeration on creativity, talent, or cultural development. Other

Table 1
Municipalities and population in Spain in EUZ, year 2011.

Number Inhabitants (in thousands)	Number Municipalities	% total in number of municipalities	Population (2011)	% in total population	Population/ Municipalities
> 1000	2	0.02	4,880,486	10.38	2,440,243
500–1000	4	0.05	2,743,809	5.83	685,952.25
100–500	52	0.64	9,921,453	21.10	190,797.17
50–100	74	0.91	5,327,459	11.33	71,992.69
20–50	111	1.37	3,404,624	7.24	30,672.29
5–20	240	2.96	2,619,183	5.57	10,913.26
< 5	174	2.14	472,581	1.00	2715.98
Total	657	8.10	29,369,595	62.45	44,702.58

Note: [(a)] with the exception of Ceuta and Melilla.
Source: Own elaboration derived from the Ministry of development database (2011).

authors, such as Frumkin (2002), explore the effect of dispersion on health. The dispersed city model encourages the intensive use of the automobile, reducing physical activity and facilitating the increase in obesity; see, among others, Ewing et al. (2003). Finally, there are several works that relate urban dispersion with taxation. The key conclusion is that compact urban growth generates a better fiscal position for local governments, whereas, when development occurs with high levels of urban dispersion, the fiscal situation of local government worsens. Kotchen and Schutle (2009) can be seen for a general review.

However, the aspect around which the greatest scientific production has been carried out in recent years is the environmental impact of urban sprawl. Obviously, most of the studies have explored aspects related to mobility, as commented above. However, there are also several works dedicated to exploring the relationships between urban sprawl and energy and electricity consumption, which can also affect the environment since the share of renewables in the energy mix is still low. As an example, see, among others, Stiri (2014), Wiesmann et al. (2011), Huang (2015), and Lasarte et al. (2018). All these works coincide in offering evidence on the lower energy efficiency of typical single-family homes in scattered urban settings. This residential model increases the energy costs borne by families and significantly reduces energy efficiency. This occurs because, in the first place, it is more complex and expensive to provide primary sources, whether gas, electricity or others, to homes located in dispersed urban areas. One has to travel more km of pipelines or power lines that imply higher costs and greater losses of resources in transport. All of this is needed just to reach a very low volume of families because dispersed growth implies a low density. However, second, and more importantly, single-family homes are a mode of construction that is more energy inefficient than apartments/-flats in buildings. They are more exposed to cold or heat and are more expensive to maintain the temperature of the house.

Most of the previous literature in this emerging line of research is conducted on a household scale. Therefore, most authors measure the energy efficiency of families/houses. However, there are hardly any works – to our knowledge – that have attempted to extend these conclusions at the city level, where the effect of their shape should be evaluated. This is the specific contribution that we try to make with our analysis.

3. Methodology and data

In our aim to evaluate the effect of urban sprawl on electricity consumption for Spanish cities, we need to include some variables. First, the spatial unit used should be defined, which is the objective of sub-Section 3.1. Then, the methods to obtain the dependent variable (electricity consumption) and the main independent variable (urban sprawl index) are presented, sub-Sections 3.2 and 3.3. Finally, sub-Section 3.4 summarizes the variables that we have to incorporate as controls in the estimation.

3.1. The spatial level of the analysis: Spanish municipalities

The first step for our analysis, especially relevant when we consider the complexity of the local/regional institutional structure in Spain, is to properly delimit the level of spatial disaggregation and define the spatial unit of the analysis.

The highest level of spatial disaggregation of the public administration in Spain is the municipality. Spain is made up of 8114 municipalities. Nevertheless, most of these municipalities are rural areas for which the studies of urban phenomena, such as urban sprawl, are not as interesting they would be in cities. According to the Ministry of Development, approximately 84% of Spanish municipalities have less than 5000 inhabitants and account for only around 13% of the total Spanish population. For comparison, all the smallest municipalities with less than 5000 inhabitants have almost the same total population as the two largest municipalities, the Madrid and Barcelona municipal areas, with more than 1,000,000 inhabitants which account for 10.38% of the population. If we only selected those urban areas with more than a specific population size (individually considered), we would be missing the fact that most of the large urban areas or metropolises in Spain cover larger areas than those within the municipal boundaries. For example, in Madrid or Barcelona, the real city includes more than 50 municipalities; some of them are very small in terms of population, but well integrated in the dynamics of the metropolitan area.

For the aforementioned reasons, we select municipalities that are clear urban areas based on one of the following factors: have more than 50,000 inhabitants, belong to a metropolitan area, or be located in its influential area. To determine the metropolitan areas and their influential areas, we use the official delimitation of the Spanish government for 2011 (Ministry of development, 2011), known as Extended Urban Zones (EUZ). This delimitation implies a total amount of 657 Spanish municipalities according to the official information. The main data are summarized in Table 1.

3.2. The dependent variable: disaggregating electricity consumption

The lack of disaggregated information on energy consumption at the local level represents a problem in identifying the effect of urban sprawl on energy consumption. Moreover, the aggregation of heterogeneous spatial units within the same region makes this problem even larger. Estimating any relationship in these conditions could easily be spurious due to the modifiable areal unit problem, as shown in Openshaw and Taylor (1979) or Openshaw (1983). To solve this problem, small area estimation (SAE) provides a proxy at the local level. A basic explanation of this approach can be found in Rao (2003) and an extended summary of different methodologies in Pfeffermann (2002, 2013), and Guadarrama et al. (2014). Through this approach, we disaggregate this variable at the local level, which constitutes another relevant contribution of the paper.

This research adapts the Fernandez-Vazquez et al. (2017) methodology to disaggregate figures of electricity consumption. To the best of

our knowledge, there are no previous studies applying this methodology to disaggregate figures of electricity consumption. This methodology is based on the well-known framework of Elbers et al. (2003), Lanjouw (2003), and Tarozzi and Deaton (2009), where disaggregated information of a key variable in the consumption survey can be obtained using a group of auxiliary variables that are common in microdata from the HBS and census. This procedure has been used by the World Bank to estimate economic information in different countries. Some recent examples of this perspective can be found in Bramley and Watkins (2013) for Scotland, Melo et al. (2014) for the United Kingdom, Modrego and Berdegué (2015) for several countries of Latin America, and Morales et al. (2017) for the region of Valencia (Spain).

Fernandez-Vazquez et al. (2017) enhances this methodology through entropy (see Golan et al., 1996 or Golan, 2017), adding information about the observable aggregates in the survey, similar to Bernadini--Papalia and Fernández-Vázquez (2018). This information improves the accuracy of the estimates due to the lack of assumptions and makes them consistent with the aggregates. The procedure starts with the Tarozzi and Deaton (2009) specification:

$$\widehat{y}_i = \frac{1}{N_i}\sum\nolimits_{h\epsilon H(i)} \widehat{E}\left(y_h|x_h;\widehat{\gamma}_r\right) \tag{1}$$

where x_h stands for the additional characteristics of household, h, in the census and $\widehat{\gamma}_r$ is a set of parameters in each region, r, that indicates the relationship between the key variable in the household, in our case, electricity and additional characteristics. The value of the variable in city i, $\widehat{y}_i$, can be obtained through aggregation of the estimated values in each location.

According to this formulation, it is necessary to follow a sequence of steps. First, the relationship between electricity in logs[1] and relevant information about the households (such as sex, age, working status of the household head, or number of members) must be estimated.[2] These variables must provide the same definitions in the HBS and census.

It is possible to predict the electricity consumption of the households in the census by applying the estimated relationship to their observable characteristics. Households in the census can be located at a municipal level, so the estimated electricity consumption in the municipality can be calculated by using the mean of the predictions.

Despite the advantage of this methodology to study spatial problems, one problem remains: ordinary least squares (OLS) extrapolation of HBS relations over census observations can easily differ from observable regional aggregates. Therefore, according to IMAJINE's approach, this issue can be solved through optimization with restrictions. In this optimization, the known aggregates are used as a restriction for the estimates. The resulting estimates should be the most fitting ones that also fulfil the condition of aggregation given by the HBS.

The fitness of the predictions is judged according to the Shannon (1948) entropy indicator, which indicates the 'uncertainty' value. This measure depends on the probability of each value and gives a score that is the maximum for a homogeneous distribution (see Golan (2017) for a more extended review of this approach). As a result, this approach tries to modify the original values as little as possible, given a set of restrictions. This value can be calculated through a supporting vector with M possible values, $b_h^{'} = [b_{h1},\ldots,b_h^*,\ldots,b_{hM}]$. In this case, the supporting vector is designed using three standard deviations of distance from the mean as extreme values and the Tarozzi and Deaton (2009) estimation in the middle. These values are chosen to allow enough variability in the estimates, avoiding the fact that the estimates depend on an ad hoc value of the interval; see Golan (2017) for more information. Consequently, each value in (1) could be expressed as a multiplication of the supporting vector and their probabilities p_{hm}.

$$\widetilde{y}_h = \sum\nolimits_{m=1}^{M} p_{hm} b_{hm} \tag{2}$$

This is a weighted mean of the values in the supporting vectors using probabilities as weights. These probabilities will be unknowns in the optimization problem. By means of the probabilities of the supporting vectors, the Shannon (1948) entropy measure can be expressed with the following equation:

$$Ent(p) = -\sum\nolimits_{h=1}^{H}\sum\nolimits_{m=1}^{M} p_{hm}\ln(p_{hm}) \tag{3}$$

This measure has the optimum value in the homogeneous distribution of probabilities. Therefore, it deviates as little as possible from this distribution. With this property in mind, restrictions and errors can be introduced in the optimization as in (4):

$$\underset{\boldsymbol{P},\boldsymbol{U}}{Max}Ent(\boldsymbol{P},\boldsymbol{U}) = -\sum\nolimits_{h=1}^{H}\sum\nolimits_{m=1}^{M} p_{km} ln(p_{km}) - \sum\nolimits_{j=1}^{J} u_j ln(u_j)$$

subject to:

$$\boldsymbol{y_r} = \sum\nolimits_{h=1}^{H}\sum\nolimits_{m=1}^{M} \boldsymbol{b_{hm}} p_{hm} \widehat{\boldsymbol{x}}_{\boldsymbol{h}} + \sum\nolimits_{j=1}^{J} v_j u_j \tag{4}$$

$$\sum\nolimits_{m=1}^{M} p_{hm} = 1; k = 1,\ldots,K$$

$$\sum\nolimits_{j=1}^{J} u_j = 1$$

This optimization problem maximizes the entropy measure expressed in (4) by adding a supporting vector for the errors v_j with their probabilities u_j. The three constraints indicate that the estimates have to be consistent with the observed regional aggregates, and at the same time, the sum of the probabilities in the supporting vector has to be equal to one.

This methodology will be used to decompose information about electricity consumption. To do so, two databases are combined: the Spanish HBS and the census. Both databases are provided by the Spanish National Institute.

The HBS database[3] provides information on 21,358 households with personal information about their members and their consumption. The location in this database is only provided at the NUTS 2 (Nomenclature des Unités Territoriales Statistiques) level to protect the confidentiality of the households in the sample. In the specific case of Spain, this level represents Spanish autonomous communities.

The census database, on the other hand, provides information for a sample of 1,619,806 anonymous households. It indicates the specific location of households in municipalities of more than 20,000 habitants, or the region with the population of the municipality (4 intervals) in areas of less than 20,000 habitants.

In this research, the key variable is household electricity consumption in the main residence, as defined by the National Institute.[4] Therefore, it indicates the level of electricity consumption in the main home as well as the electricity consumed in the garage or similar areas of the main house. Several variables, observable in both sources, can be used to explain the variability in energy consumption. The variables used in this analysis can be found in Table A1 in the appendix with their mean and standard deviation. These variables describe the personal characteristics of the head of household as well as the structure of the

[1] Given that the relation is obtained in a log equation, an additional parameter can also be obtained in order to obtain the expected value without logarithms, as in Wooldridge (2011): $\widehat{E}(y_h|x_h;\widehat{\gamma}_r) = \widehat{\alpha}_0 \exp[f_0](x_h\widehat{\gamma}_r)\forall h \in r$.

[2] Heterogeneity in the parameters is introduced with a set of parameters $\widehat{\gamma}_r$ for each region r available in the HBS.

[3] Year 2011 has been chosen to allow a perfect match with the population census of the same year.

[4] Category 4.5.1.1 of the HBS.

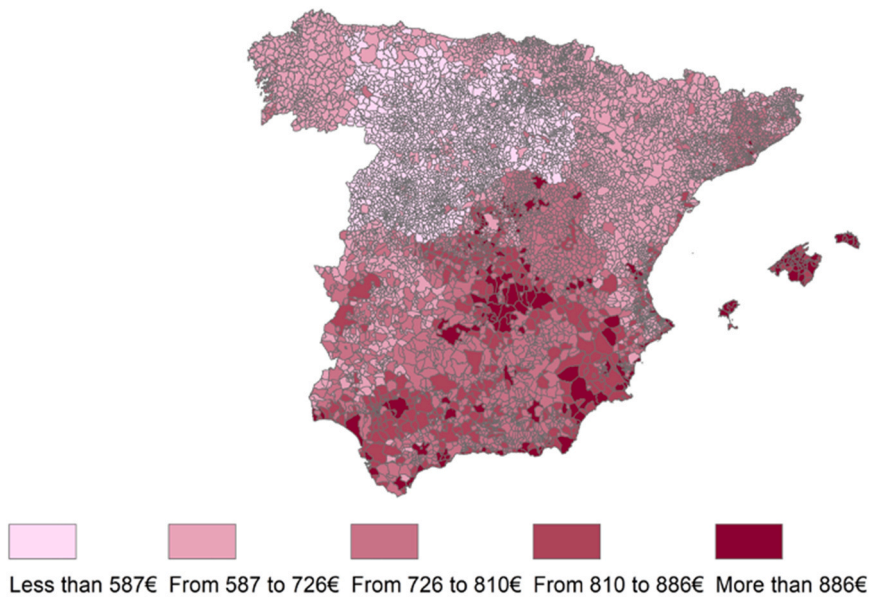

Fig. 1. Map of the mean estimates of electricity consumptions of the households in each municipality with the correction of GME, year 2011 (€/year). Source: Own elaboration.

household.[5] As a summary, explanatory variables of energy consumption (x_h) are presented in Table A1. These variables are used to estimate energy consumption for the households (y_h) in the census. However, given the information available on energy consumption in the HBS for the regions (y_r), the GME procedure in Eq. (4) is applied to the estimates to make them consistent with the observable aggregates. After the disaggregation process, we obtain the level of electricity consumption by municipality, which is represented in Fig. 1. Even from a national map, the extent of the differences between municipalities can be easily seen. In fact, 10% of the municipalities at the bottom are less than or equal to 36% of the mean, while 10% of the municipalities at the top are greater than or equal to 6.7% of the mean, indicating higher electricity consumption in the south than in the north of Spain. The highest observed values in electricity consumption are related to regions around Madrid, the capital of Spain. Greater values of electricity consumption concentrate on the Mediterranean Sea from north (i.e., Barcelona) to the south, with higher demand in the Valencia region. Conversely, average electricity consumption at the municipality level is lower in the internal regions in the north of the country, such as in Galicia.

3.3. The main independent variable: an urban sprawl index

The phenomenon of urban sprawl has been studied within different disciplines (geography, town planning, territorial planning, environmental science, economics, sociology, and even public health) from very different standpoints. This multifaceted nature leads to numerous definitions, and sometimes, one is inconsistent with another, thereby leading to confusion (Richardson and Chnag-Hee, 2004). One of the main goals over the past decade in terms of the analysis of urban sprawl has been to provide a precise definition of the concept that might also lead to quantitative research.

Galster et al. (2001) provided a definition that manages to encompass the complexity and multidimensionality of urban sprawl, defining it as "*a pattern of land use in an urban area that exhibits some level of combination of eight dimensions: density, continuity, concentration, clustering, centrality, nuclearity, mixed use, and proximity*". Squires (2002) defines sprawl in a similar way, including the *"vehicle dependency and exclusion of new developments in the outskirts of settled areas"*. Glaeser and Kahn (2004) compiled a complete review on urban sprawl, defining it similar to the authors mentioned previously and taking the different aspects that interact in a sprawled city into consideration. Dwyer and Childs (2004) connect sprawl with the decline in city centres and Sturm and Cohen (2004) focus more on its public health effects, a Davoudi (2003) and Johnson (2001) view it from an environmental perspective.

Although the way in which previous authors define the concept of urban sprawl is the most consistent with its multidisciplinary character, it presents the problems of the overlapping of the dimensions taken into consideration and the difficulty of combining them in a single measure, enabling its comparison over time and between different locations. Consequently, other authors have defined the phenomenon to be easily converted into an indicator. Peiser (2001) defines sprawl from a more spatial perspective. Jaeger and Schwick (2014) take a major step forward in this respect, focusing their definition on only three dimensions, namely, dispersion, ratio of built-up area, and density of use, and they propose an index pondering these three aspects. However, their index presents a serious problem of subjectivity in the weightings applied as well as in the choice of the dimensions considered.

Burchfield et al. (2005) go one step further in simplifying the concept and classify the phenomenon of urban sprawl as "*whether the residential development is scattered or compact,*" such that "*in the sprawling areas much of the land immediately surrounding the average house will not itself be developed*", bringing the definition of urban sprawl down to only one dimension, the degree to which building is dispersed, thus simplifying quantification. These authors propose an urban sprawl index (USI) consistent with their definition, which can be obtained via the possibilities offered by geographic information systems (GIS). These authors specifically use TM Landsat imagery at a resolution of 30×30 m, providing photointerpretation in a raster GIS scenario. This scenario indicates the delimitation of the pixels of the image as urban or rural and, for each pixel considered, urban counts the number of other urban pixels that fall within an area of 1 km^2 around it, applying the following formula:

$$\mathrm{USI} = 100\left[1 - \frac{\text{Urban pixel}}{18^2\pi}\right] \tag{5}$$

[5] In addition, control dummies for economic sector (1 digit) and occupation level in 5 categories (Manager/Technician or professional/Support worker or sales/craftsman machine operators or skilled agricultural worker/non-qualified workers) and the squared age of the head of the household were also included.

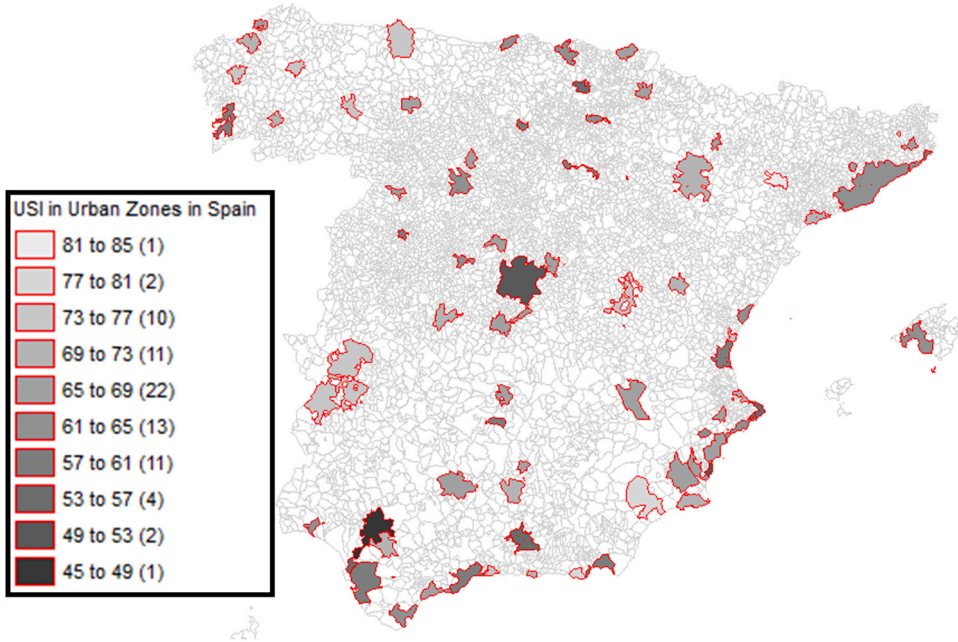

Fig. 2. Map of Urban Sprawl Index (USI) for the major Spanish urban and metropolitan areas, year 2001.
Source: Rubiera et al. (2016).

Thus, high values of USI (up to 100) indicate high levels of dispersion or sprawl, while low values indicate concentration (Burchfield et al., 2005). This paper has had a major impact, and this measure will be the one used in this analysis, making use of Rubiera et al. (2016), where the proceedings of Burchfield et al. (2005) are applied to the specific case of Spain by exploiting the potential of the databases available for this country.

This paper departs from the estimates of USI in Rubiera et al. (2016) for the studied 657 municipal areas in Spain. The USI estimates are mapped in Fig. 2. The average USI for all Spanish urban areas is 68.81. There is a strong dispersion in this index within the territory: from the metropolitan area of Seville, with the lowest value of 48.13, to Lleida, with the highest level of 81.12. Other important urban areas with a high level of sprawl include Madrid, Granada, and Vitoria. At the other extreme are cities such as Caceres, Lugo, and Santiago, with the lowest levels of sprawl. Barcelona presents the national average value.

Regarding the period of time, we are very limited by the data availability for the calculated sprawl index. Our assumption is that a phenomenon such as urban sprawl does not produce effects immediately. Consequently, it is necessary to have a significant time lag to observe the impact that urban sprawl may have on energy efficiency. We propose the use of sprawl data from the beginning of the decade, using 2001, which is the first year when the urban sprawl index was calculated, and the electricity consumption at the end of the same decade, 2011.

3.4. Other control variables

To obtain a reliable estimate of the relationship between the level of dispersion and electricity consumption, control variables are introduced in the model. Despite the limited available information for Spanish municipalities, estimations include controls for socioeconomic, geographical, and climatological factors, following the ideas of Rappaport and Sachs (2003). According to them, geography has an important role in the location of the population, productivity, and quality of life, creating a suitable environment where it is easier for cities to grow.

The Spanish Institute of Geography (ING) and Spanish Climatological Agency (AEMET) provide precise local information on geographical characteristics and climatological data at the local level. A description of all the variables incorporated in the model can be found in Table 2. All variables, including dependent variables and controls, are accounted for in logs and summaries are reported in Table A1.

Table 2
Control variables description (year considered for the control is 2011).

Other control variables	Description
Pop	Population resident
Pop 1970	Population resident in 1970
Prec Total	Level of precipitation (average liters/m^2)
Rat man	Share of manufacturer employees on the total employment
Rat_ SA	Share of advanced services employees on the total employment
Tmax Jul	Average of the maximum temperature in the month of July
Tmin Jan	Average of the minimum temperature in month of January

4. Econometric strategy

Different strategies may be adopted for treating spatial heterogeneity. Basically, we can distinguish between continuous and discrete heterogeneity (Anselin, 2010). While the former corresponds to the variation of the relationship in each unit (as in the case of geographically weighted regression; Fotheringham et al. 2002), the latter leads to the identification of clusters defined as spatial regimes (Anselin, 1988). In this sense, the identification of regimes may be conducted choosing both an exogenous and an endogenous criterion (Ramajo et al., 2008; Postiglione et al., 2013). This article focuses on the identification of spatial clusters in energy consumption. However, geographically weighted regression (GWR) results are also presented, allowing us to compare both methods and enhancing the robustness of the conclusions.

4.1. Geographically Weighted Regression

GWR has often been used in applied research. Estimated coefficients in GWR change in each locality so that nearby observations have more influence in determining local regression relationships. For example, parameters indicating sensitivity to climate conditions (as well as other variables) may be calculated at every spatial unit based on major effects

Table 3
A model for energy consumption for 248 Municipalities in Spain.

Variables	Coefficients
	5.215*** (0.284)
Tmax Jul	0.589*** (0.056)
Tmin Jan	0.029*** (0.004)
Prec	-0.021 (0.013)
Ratio Man	-0.029*** (0.008)
Ratio SA	0.052*** (0.016)
Pop	-0.018 (0.012)
Pop 1970	-0.046*** (0.008)
USI	0.040 (0.026)

Note: Model 3 includes Sprawl indicator (USI). *P-values* 0.10(*), 0.05(**), and 0.01 (***). Standard errors in brackets.
In order to avoid for the presence of multicollinearity in the estimation we verified the variance inflation factors (VIF) for different variables. Particularly, all the selected variables show VIFs below the level of the (Vittinghoff et al., 2011). A table including VIFs is reported in Table A2.
Source: Own elaboration.

observations on a dependent variable, y, and the p predictors, $\mathbf{X}$. The regression tree algorithm tries to find partitions so that the X space is split into k disjoint sets, $A_1, \ldots, A_k$. Each split is based on a single criterion, i.e., the value of a single variable X. The rule created at each split divides the data into two subsets with maximum homogeneity to minimize the sum of squared residuals, therefore minimizing misclassification costs (Breiman et al., 1984). Successively, each subset of a child node is split further until one of the following stopping criteria is reached:

(i) The division of a specified node does not improve the fit of the model (Therneau and Atkinson, 1997).
(ii) Further disaggregation of any current group generates subgroups whose size is less than a certain threshold value.
(iii) The constraint on the maximum number of groups k is active.

After terminal nodes are individuated, the chosen model will be estimated for each of the groups $A_1, \ldots, A_k$ by ordinary least squares (OLS). Hence, this technique avoids a large production of coefficients

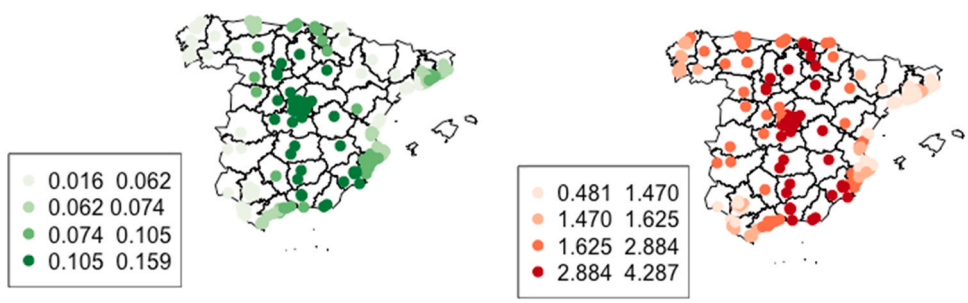

Fig. 3. Quantile map for estimated parameters related to USI using GWR (left) and significance levels (*T values* - right).
Source: Own elaboration.

from neighbours. The estimation is obtained by adopting a geographical version of the weighted least squares estimator (Fotheringham et al., 2002):

$$\widehat{\boldsymbol{\beta}}_i = (\mathbf{X}^t\mathbf{W}_i\mathbf{X})^{-1}\mathbf{X}^t\mathbf{W}_i\mathbf{y} \tag{6}$$

where $\boldsymbol{\beta}_i$ is the set of parameters corresponding to the regression covariates plus the intercept variance for each unit, $i = 1...n$, $\mathbf{X}$ is the set of p covariates plus a vector of ones, $\boldsymbol{y}$ is the vector including observations for the dependent variable, and $\mathbf{W}_i$ is a nxn diagonal matrix whose entries are specified on the basis of a kernel. In our application a Gaussian kernel is adopted:

$$w_{ij} = \exp\left(-d_{ij}^2/2\gamma^2\right) \tag{7}$$

where d_{ij} is the distance between location i and j and γ is the selected level of bandwidth. Following Fotheringham et al. (2002), the bandwidth is determined in this research by the AIC criterion. GWR has expanded the consideration of spatial heterogeneity, allowing the identification of geographical differences in the estimated parameters.

4.2. Structural differences and regression trees

Regression trees (Breiman et al., 1984; Hastie et al., 2001) have also been adopted in applied research to address spatial heterogeneity (among others, Tomczyk and Ewertowski, 2013; Postiglione et al., 2010).

A regression tree is developed through binary recursive partitioning that divides a data set of n observations into different subsets, imposing certain linear conditions on a set of covariates. The algorithm can be described as it follows. In step zero, we start considering a sample of n

and simplifies the interpretation for groups. In fact, regression trees could help policy makers localize spatial clusters that show homogenous effects of sprawl on electricity consumption.

5. Main results: effects of urban sprawl on electricity consumption

5.1. Results in a general estimation using OLS models

Table 3 presents the model estimated by ordinary least squares (OLS). Most of the control variables are significant with expected signs. Temperatures seem to adequately capture the influence that weather differences have on electricity consumption. The sign of the coefficient of the temperature variables is, for both the maximum and the minimum, positive. In fact, heating systems in Spain rely on various energy sources: gas, diesel, and, to a lesser extent, coal. Electric powered heating systems are also used, but this is less frequent. Therefore, in cold climate areas to the north of the peninsula, various energy sources are used for heating systems. In contrast, air conditioning works exclusively using electricity. Therefore, in warmer areas, the use of electrical energy increases significantly.

The model incorporates the urban sprawl level, which is not significant at the global level. This result is, at first glance, different from what was found by Lasarte et al. (2018) at the household level; however, OLS results do not take spatial heterogeneity into account. Therefore, we will focus our attention on the following estimation methodologies to be able to consider this effect.

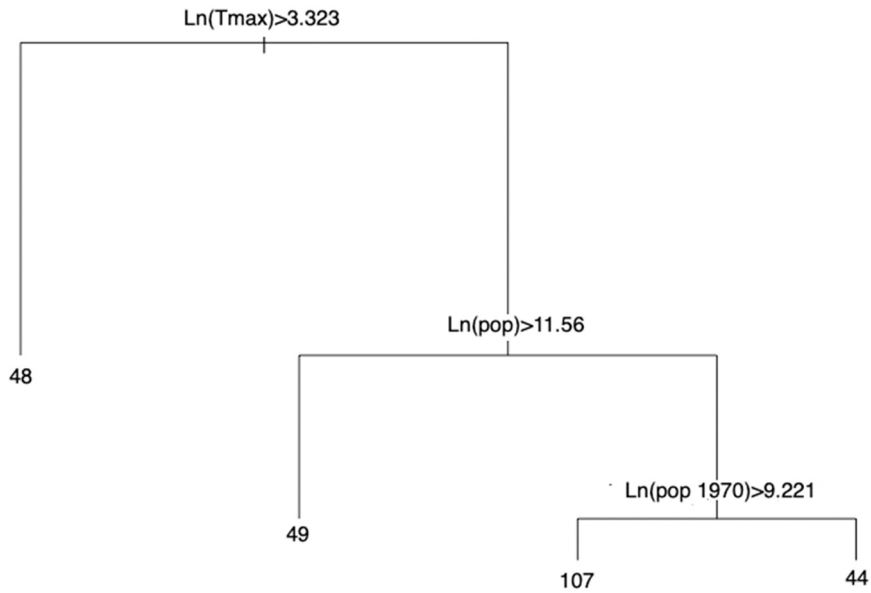

Fig. 4. Regression tree performed on the global model, including variables used for splitting and number of units in the spatial clusters.
Source: Own elaboration.

Table 4
Estimates of parameters for different spatial clusters obtained by regression trees.

	Atlantic	Inland	Mediter	Large
	5.786***	5.272***	5.231***	5.225***
Tmax Jul	0.594***	0.547***	0.313***	0.510***
Tmin Jan	0.052***	0.032***	0.023***	0.035**
Prec	-0.094**	-0.041	0.009	0.042
Ratio Man	-0.052**	-0.052***	0.002	-0.060***
Ratio SA	-0.074**	-0.074	0.019	0.097***
Pop	0.053	0.015	0.077***	-0.051*
Pop 1970	-0.093**	-0.031*	-0.073**	-0.033*
USI	-0.042	-0.008	0.069**	0.097*

Note: *P-values* 0.10(*), 0.05(**), and 0.01 (***). Standard errors in brackets.
Source: Own elaboration.

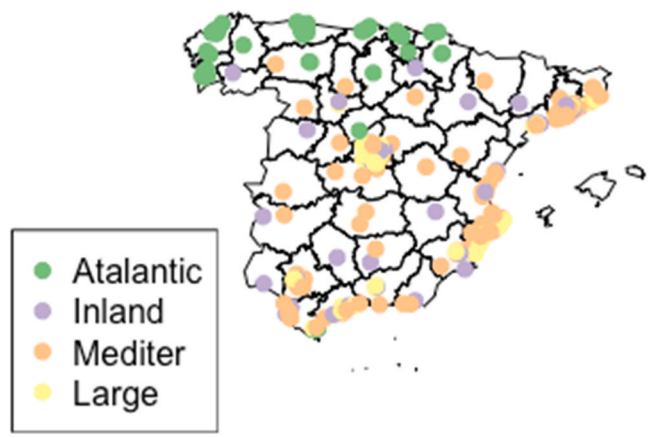

Fig. 5. Units (Municipalities) included into the four different clusters of the regression tree.
Source: Own elaboration.

5.2. *Energy consumption, urban sprawl, and spatial heterogeneity*

GWR is applied for the 248 cities of Spain by using a Gaussian adaptive kernel. The bandwidth is determined based on an AIC criterion.[6] GWR produces an extensive set of results, and a table summarizing variations in estimated parameters for each variable is reported in Table A3. In particular, variations in the minimum and maximum temperatures show higher coefficients in southern cities. For the special case of urban sprawl, our variable of interest, local variation of coefficients seems to justify a deeper analysis of spatial heterogeneity. In Fig. 3.a, we observe how the coefficients linked to USI have higher a magnitude especially around Madrid. On the one hand, few cities in the North present a high and significant estimated parameter for USI (see Fig. 3.b for the significance of estimated parameters at the local level).

On the other hand, discrete heterogeneity can also be studied through regression trees to identify spatial clusters. Fig. 4 summarizes the splitting, where a minimum of 40 units per group is set. The nodes follow a geographical logic (first split is determined by maximum temperature during summer – Tmax Jul). Urban areas in the north are grouped in the node that we have designated as "Atlantic". A split comes from urban characteristics, as we designated "inland" the node that collects urban areas of the peninsular interior with less than 80,000 inhabitants (Pop). Another split comes from urban population in the past (Pop 1970); we designated “Mediter” as the node collecting urban areas recently developed on the Mediterranean and “Large” as the node including the traditionally largest metropolitan areas (for example Madrid).

Table 4 and Fig. 5 report the results for the different regressions on each group. Considering heterogeneity allows us to derive accurate local models that change their coefficients over space. Major changes involve the variable of the urban sprawl index (USI). If this effect is not significant in the global model, we observe that it turns significant in two important cases by considering spatial heterogeneity. The effect of sprawl was significant for both the “Mediter” group and the “Large” group. Particularly, in the “Large” group, this coefficient is higher in magnitude, a circumstance that suggests how sprawl influences the energy consumption of households in greater cities. Additionally, in the “Mediter" group (where there are many vacation areas with second residences combined with some of the most important urban areas in the

[6] In this application an adaptive kernel is preferred to a fixed one, due to irregular distribution of municipalities considered. In an adaptive kernel the number of units is specified for the bandwidth. For this application optimal bandwidth corresponds to 29.

Table 5
Root of mean square errors (RMSE) for OLS, GWR, and regression tree (RT).

	OLS	GWR	RT
RMSE	0.8312	0.7612	0.7290

Source: Own elaboration.

country), the level of sprawl has a positive and significant effect, though slightly lower. In contrast, in "inland", which corresponds to *emptied Spain* (the area of Spain that loses population), the parameter is not statistically significant, which may be a result of the concentration process around the largest cities. Lastly, a relevant difference appears in the composition of the energy consumption model in the Atlantic node.

Comparing local results to those obtained by Lasarte et al. (2018), it can be observed that for "Large" and "Mediter", the impact at the municipal level is more relevant than that at the family level. One of the possible explanations is that we use the most accurate measurement of sprawl in this work and that dispersed areas also suffer energy inefficiencies in lighting and public services. This result is also consistent with the conclusions of other works that measure the impact of sprawl on other energy sources or for other economies (such as Stiri (2014), Wiesmann et al. (2011), Huang (2015)).

Looking at Table 5, we observe the root of the mean square error for specifications. Among the three options, OLS shows a higher level of RMSE, while both GWR and the regression tree option lead to a reduction in RMSE. Particularly, we see that by considering spatial clusters, we obtain an additional improvement in the specification.

Spatial heterogeneity should be carefully taken into account when modelling electricity consumption, especially in the case of Spain. The assessment of spatial heterogeneity can be supportive in terms of local policies and highlights relationships that are "hidden" at the global level. Moreover, while modelling heterogeneity, we should consider the difference between continuous heterogeneity and discrete heterogeneity. In our study, we see how both options can be valid alternatives for uncovering the effect of sprawl; however, due to a slightly better performance and easier interpretability, we gave broader consideration to the case of discrete spatial heterogeneity.

6. Concluding discussion

The main objective of the paper was to evaluate the degree to which the urban sprawl of cities can affect electricity consumption. Since the available information does not offer disaggregated information on electricity consumption at the municipal level, a first contribution of the paper is a disaggregation of the energy consumption, for which we use small area estimation. Exploiting micro-data from the census, estimations were obtained making use of generalized maximum entropy. This information is combined with several control variables available at the local level and combined in a model that was initially estimated using OLS. This preliminary approach does not seem to capture any effect of the sprawl on the energy consumption of the families. Nevertheless, we observe that behind this global result, there is important spatial heterogeneity. To control and consider this spatial heterogeneity, we used a regression tree approach. The joint consideration of working at the local level and considering the spatial heterogeneity with the regression tree offers a clearer picture.

In the larger cities of the country as well as those located along the Mediterranean coast, the cases in which sprawl is more intensive and grows more during the last decade, urban sprawl clearly affects electricity consumption. In these places, the higher the sprawl is, the higher the electricity consumption is, which indicates that more attention should be paid to the urban form and to the design of policies to reduce urban sprawl.

The *Spanish Energy Saving and Efficiency Action Plan* 2011–2020, elaborated by the Ministry of Industry, Tourism, and Commerce, and the IDEA (Institute for Energy Diversification and Saving; Spanish Government, 2011), includes recommendations on energy saving measures in the building sector, in accordance with the contents of Directive 2010/31/EU related to energy efficiency in buildings (European Commission, 2010) and the Resource Efficient Europe strategy (European Commission, 2011). Another strategy regards the introduction of small installations of renewable thermal energy and cogeneration in household electricity consumption. In that sense, the Spanish Government (2014) approved Law 413/2014, which establishes, for the first time in Spain, the conditions of electricity self-consumption, whose regulations were especially designed for households, since they were focused on small plants (less than 100 kV of installed capacity). The most recent Royal Decree in Spain related to self-consumption was approved on the 5th of April 2019 (RD 244/2019, see BOE (2019)) by the Spanish government, where more conditions and specifications are regulated, such as access to electricity production and distribution to the population. These measures are in line with European objectives, United Nations (2016), which try to encourage member countries to adopt more sustainable energy strategies by implementing policies, such as increasing the share of renewables in the electricity mix, trade, and efficiency. According to our results, major attention should be given to the urban form, especially in recently developed areas (i.e., urban areas growing faster than the Mediterranean) and in very large cities such as Madrid. Urban structure shows a heterogeneous impact in different geographical situations and climate conditions. In this direction, urban policies should be very aware of them, and above all, a "one size fits all" policy could be very far from reaching the expected results in terms of energy efficiency. Conversely, the definition of local policies, major attention to local planning, and collaboration between national, regional, and local authorities could lead to a reduction in urban sprawl and contribute to preserving the environment.

Finally, our analysis is restricted to immediate effects of electricity consumption associated with urban sprawl. As a next step in this analysis, we are interested in seeing the effects of sprawl in socioeconomic and environmental terms. The disaggregation of electricity consumption at the municipal level made in this paper, together with the econometric strategy developed, are suitable tools to be considered in the future. Through this methodology, it is possible to plan scenarios to change the sprawling processes of cities and determine the effect on electricity consumption, thus making use of a more global model, which will impact the rest of the variables of the economy. Another immediate future line of research is inclusion in the analysis of electricity prices. Although low, there can be differences among autonomous communities, which creates clear borders on the disaggregated maps of energy consumption. There was a contraction in the access tariff by capacity, the tax that consumers pay to cover some fixed costs, especially for transmission and distribution companies, which can include different taxes among regions. This reason can explain the clear differences among regions obtained in our results.

Acknowledgements

This rresearch has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 726950 IMAJINE.

Appendix

See Tables A1-A3.

Table A1
Summary statistics for other control variables. All variables are expressed in log.

Other control variables	Description	Min	1st Q	Median	3rd Q	Max	Mean
Pop	Population resident	9.879	10.315	10.914	11.462	14.999	11.024
Pop 1970	Population resident in 1970	5.442	9.309	10.003	10.934	14.954	10.170
Prec Tot	Level of precipitation (average liters/m^2)	2.015	3.618	3.846	3.953	4.979	3.853
Rat man	Share of manufacturer employees on the total employment	-4.241	-3.127	-2.644	-2.223	-1.370	-2.707
Rat_ SA	Share of advanced services employees on the total employment	-4.528	-3.406	-3.175	-2.978	-2.004	-3.187
Tmax Jul	Average of the maximum temperature in the month of July	3.082	3.367	3.405	3.470	3.592	3.395
Tmin Jan	Average of the minimum temperature in month of January	-1.240	-1.204	1.281	1.569	1.946	0.615

Table A2
Varianc Inflation Factors (VIFs) for control variable.

Control variables	Variance Inflation Factors (VIFs)
USI	1.991
Pop	5.880
Pop 1970	5.541
Prec Total	1.991
Rat man	1.209
Rat_ SA	2.050
Tmax Jul	2.205
Tmin Jan	1.381

Table A3
Summary of the estimated obtained by Geographically Weighted Regression.

	Min	1st Q	Median	3rd Q	Max
	3.155	4.427	4.728	5.344	6.917
Max Jul	0.008	0.461	0.614	0.728	1.002
Min Jan	-0.008	0.012	0.025	0.032	0.038
Prec Tot	-0.031	-0.009	0.012	0.053	0.135
Ratio Man	-0.064	-0.049	-0.038	-0.015	0.036
Ratio SA	-0.035	0.019	0.080	0.127	0.162
Pop	-0.083	-0.038	-0.007	0.005	0.020
Pop 1970	-0.061	-0.052	-0.049	-0.038	-0.015
USI	0.016	0.062	0.074	0.105	0.159

References

Abu Hatab, A., Cavinato, M.E.R., Lindermer, A., 2019. Urban sprawl, food security and agricultural systems in developing countries: a systematic review if the literature. Cities 94, 129–142.
Anselin, L., 1988. Spatial Econometric: Methods and Models. Kluwer Academic Publisher, Dordrecht.
Anselin, L., 2010. Thirty years of spatial econometrics. Pap. Reg. Sci. 89 (1), 3–25.
Axelrod, J.B.C., 2007. Keep the L out of Los Angeles: race, discourse and urban modernity in 1220s southern California. J. Urban History 34 (1), 3–37.
Bernadini-Papalia, R., Fernández-Vázquez, E., 2018. Information theoretic methods in small domain estimation. Econom. Rev. 37 (4), 347–359.
BOE, 2019. Royal Decree of 5th of April to regulate administrative, technical and economic conditions of the self-consumption of electrical energy by the Spanish Government. Spanish Government.
Bramley, G., Watkins, D., 2013. Developing Local and Small Area Estimates and Exploring Patterns of Income Distribution, Poverty and Deprivation. Report of Research for the Improvement Service on Behalf of Four Local Authorities and the Scottish Government.
Breheny, M., 1996. Centrist, decentriscts and compromisers: views on the future of urban form. In: Jenks, M., Burton, E., Willians, K. (Eds.), The Compact City: A Sustainable Urban Form? E&FN Spon.
Breiman, L., Friedman, J., Olshen, R., Stone, C., 1984. Classification and Regression Trees. Wadsworth, Belmont, California.
Briggs, X.S., 2005. More pluribus, less Unum? The changing geography of race and opportunity. Geography of Opportunity: Race and Housing Choice in Metropolitan America, pp. 17–41.
Burchfield, M., Overman, H.G., Puga, D., Turner, M.A., 2005. Causes of sprawl: a portrait from space. Q. J. Econ. 121 (2), 587–633.
Camagni, R., Capello, R., Nijkamp, P., 1998. Towards sustainable city policy: an economy-environment technology nexus. Ecol. Econ. 24 (1), 103–118.
Camagni, R., Gibelli, M.C., Rigamonti, P., 2002. Urban mobility and urban form: the social and environmental costs of different patterns of urban expansion. Ecol. Econ. 40, 199–216.
Christiansen, P., Loftsgarden, T., 2011. Drivers behind urban sprawl in Europe, p. 1136. TØI report.
Coq, D., Asian, R., 2019. Urban sprawl and sustainable urban policies. A review of the cases of Lima, Mexico City and Santiago de Chile. Sustainability 11 (20).
Couch, C., Ldontidou, L., Petschel-Held, G., 2007. Urban Sprawl in Europe. Landscapes. Land-Use Change & Policy. Blackwell Publishing, Oxford.
Davoudi, S., 2003. European briefing: polycentricity in European spatial planning: from an analytical tool to a normative agenda. Eur. Plan. Stud. 11 (8), 979–999.
Dwyer, J.F., Childs, G.M., 2004. Movement of people across the landscape: a blurring of distinctions between areas, interests, and issues affecting natural resource management. Landsc. Urban Plan. 69 (2–3), 153–164.
Elbers, C., Lanjouw, J.O., Lanjouw, P., 2003. Micro-level estimation of poverty and inequality. Econometrica 71 (1), 355–364.
European Comision, 2006. Environmental Aspects of Urban Spread in Europe. Joint Research Centre, European Commission, Brussels.
European Commission, 2010. Directive 2010/31/UE of the European Parliament and of the Council of 19 May 2010 on the Energy Performance of Building. European Commission, Brussels.
European Commission, 2011. Energy Roadmap 2050. European Commission, Brussels.
Ewing, R., 1997. Is Los Angeles-style sprawl desirable? J. Am. Plan. Assoc. 63 (1), 107–126.
Ewing, R., Pendall, R., Chen, D., 2003. Measuring Sprawl and its impact. Transp. Res. Rec. J. Transp. Res. Board 1831 (1), 175–183.
Fernandez-Vazquez, E., Plotnikova, M., Postiglione, P., Rubiera-Morollón, F. & Viñuela, A., 2017. Deliverable 2.2 Literature Review on Disaggregation Methodologies Vol. D2.2, IMAJINE WP2 Analysis of Territorial Inequalities in Europe Information.
Florida, R., Mellander, C., 2015. Exploring creativity and urban development with agent-based modeling. J. Artif. Soc. Soc. Simul. 18 (2), 12.
Fotheringham, A.S., Brunsdon, C., Charlton, M., 2002. Geographically Weighted Regression - The Analysis of Spatially Varying Relationships. Wiley, Chichester.
Frumkin, H., 2002. Urban sprawl and public health. Public Health Rep. 117 (3), 201–217.
Galster, G., Ratcliffe, M.R., Wolman, H., Coleman, S., Freihage, J., 2001. Wrestling sprawl to the ground: defining and measuring an elusive concept. Hous. Policy Debate 12 (4), 681–717.
Glaeser, E.L., Kahn, M.E., 2004. Sprawl and urban growth. In: Handbook of Regional and Urban Economics, Vol. 4. Elsevier, pp. 2481–2527.
Golan, A., 2017. Foundations of Info-metrics: Modeling and Inference with Imperfect Information. Oxford University Press.
Golan, A., Judge, G., Miller, D., 1996. A Maximum Entropy Econometrics: Robust Estimation with Limited Data. Wiley, New York.
Guadarrama, M., Molina, I., Rao, J.N.K., 2014. A comparison of small area estimation methods for poverty mapping. Stat. Trans. 17 (1), 41–66.
Hasse, J.E., Lathrop, R.G., 2003. Land resource impact indicators of urban sprawl. Appl. Geogr. 23 (2–3), 159–175.
Hastie, T., Tibshirani, R., Botstein, D., Brown, P., 2001. Supervised harvesting of expression trees. Genome Biol. 2 (1) research0003.1.
Hennin, E., Schwick, C., Soukup, T., Orlitová, E., Kienast, F., Jaeger, J., 2015. Multi-scale analysis of urban sprawl in Europe: towards a European de-sprawling strategy. Land Use Policy 49, 483–498.
Huang, W.H., 2015. The determinants of household electricity consumption in Taiwan: evidence from quantile regression. Energy 87, 120–133.
INE, 2016. Encuesta de condiciones de vida. Instituto Nacional de Estadística, Madrid.
Itsván, L.B., 2010. Urban sprawl and climate change: a statistical exploration of cause and effect, with policy options for the EU. Land Use Policy 27 (2), 287–293.
Jaeger, J.A., Schwick, C., 2014. Improving the measurement of urban sprawl: Weighted Urban Proliferation (WUP) and its application to Switzerland. Ecol. Indic. 38, 294–308.
Jenks, M., Burton, E., Willians, K., 1996. The Compact City: A Sustainable Urban Form? E&FN Spon.
Johnson, M.P., 2001. Environmental impacts of urban sprawl: a survey of the literature and proposed research agenda. Environ. Plan. A 33 (4), 717–735.
Kotchen, M., Schutle, S., 2009. A meta-analysis of cost community service studies. Int. Reg. Sci. Rev. 32, 376–399.
Lanjouw, P., 2003. Estimating geographically disaggregated welfare levels and changes. In: Silva, L.P.D., Bourguingnon, F. (Eds.), The Impact of Economic Policies on Poverty and Income Distribution: Evaluation Techniques and Tools. Oxford University Press, New York, pp. 1–17.
Lasarte, E., Rubiera, F., Cuartas, B.M., 2018. Energy consumption and urban sprawl: evidence for the Spanish case. J. Clean. Prod. 172, 3479–3486.
Lee, B., 2005. Edge or edgeless cities? Urban spatial structure in US metropolitan areas, 1980 to 2000. J. Reg. Sci. 47 (3), 479–515.
Melo, P.C., Copus, A., Coombes, M., 2014. Modelling small area at-risk-of-poverty rates for the UK using the world bank methodology and the EU-SILC. Appl. Spat. Anal. Policy 9 (1), 97–117.

Milan, B.F., Creutzig, F., 2016. Municipal policies accelerated urban sprawl and public debts in Spain. Land Use Policy 54, 103–115.
Ministry of development, 2011. Las Áreas Urbanas. Documento del Ministerio de Fomento.
Modrego, F., Berdegué, J.A., 2015. A large-scale mapping of territorial development dynamics in Latin America. World Dev. 73, 11–31.
Morales, D., Rueda, M., Esteban, D., 2017. Model-assisted estimation of small area poverty measures: an application within the valencia region. Soc. Indic. Res. 138 (3), 873–900.
Openshaw, S., 1983. The modifiable area unit problem. Concepts and Techniques in Modern Geography, pp. 1–41.
Openshaw, S., Taylor, P., 1979. A million or so correlation coefficients. In: Wrigley, N. (Ed.), Statistical Methods in the Spatial Sciences. Pion, London, pp. 127–144.
Oueslati, W., Alvanides, S., Garrod, G., 2015. Determinants of urban sprawl in European cities. Urban Stud. 52 (9), 1594–1614.
Patacchini, E., Zenou, Y., Henderson, J.V., Epple, D., 2009. Urban sprawl in Europe. Brookings-Wharton Papers on Urban Affairs, pp. 125–149.
Peiser, R., 2001. Decomposing urban sprawl. Town Plan. Rev. 72 (3), 275–298.
Pfeffermann, D., 2002. Small area estimation-new and directions. Int. Stat. Rev. 70 (1), 125–143.
Pfeffermann, D., 2013. New important developments in small area estimation. Stat. Sci. 28 (1), 40–68.
Postiglione, P., Andreano, M.S., Benedetti, R., 2013. Using constrained optimization for the identification of convergence clubs. Comput. Econ. 42, 151–174.
Postiglione, P., Benedetti, R., Lafratta, G., 2010. A regression tree algorithm for the identification of convergence clubs. Comput. Stat. Data Anal. 54 (11), 2776–2785.
Ramajo, J., Marquez, M.A., Hewings, G.J.D., Salinas, M.M., 2008. Spatial heterogeneity and interregional spillovers in the European Union: do cohesion policies encourage convergence across regions? Eur. Econ. Rev. 52, 551–567.
Rao, J.N.K., 2003. Small Area Estimation. Wiley, New Jersey.
Rappaport, J., Sachs, J.D., 2003. The United States as a coastal nation. J. Econ. Growth 8, 5–46.
Richardson, H.W., Chnag-Hee, C.B., 2004. Urban Sprawl in Western Europe and the United States. Ashgate, London.
Robinson, L., Newell, J.P., Marzluff, J.M., 2005. Twenty-five years of sprawl in the Seattle region: growth management responses and implications for conservation. Landsc. Urban Plan. 71 (1), 51–72.
Rubiera-Morollón, F., Garrido-Yserte, R., 2020. Recent literature about urban sprawl: a renewed relevance of the phenomenon from the perspective of environmental. Sustainability 12 (16), 6551.
Rubiera, F., González, V., Rivero, J.L., 2016. Urban sprawl in Spain: differences among cities and causes. Eur. Plan. Stud. 24 (1), 204–226.
Schneider, A., 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 Stud. 45 (3), 659–692.
Shannon, C.E., 1948. A mathematical theory of communications. Bell Syst. Tech. J. 27, 379–423.
Skog, K.L., Stennes, M., 2016. How do centrality, population growth and urban sprawl impact farmland conversion in Norway? Land Use Policy 59, 185–196.
Slemp, C., Davenport, M.A., Seekamp, E., 2012. Growing too fast: local stakeholders speak out about growth and its consequences for community well-being in the urban-rural interface. Landsc. Urban Plan. 106 (2), 139–148.
Spanish Government, 2011. Spanish Energy Saving and Efficiency Action Plan 2011-2020- Ministry of Industry. Tourism and Commerce and IDEA, Madrid.
Spanish Government, 2014. Real Decreto 413/2014, de 6 de junio por el que se regula la actividad de producción de energía eléctrica a partir de fuentes de energía renovables, cogeneración y residuos, BOE-A-2014-6123, Madrid.
Squires, G.D., 2002. Urban Sprawl: Causes, Consequences, & Policy Responses. The Urban Insitute.
Stiri, H., 2014. Building and household X-factors and energy consumption at the residential sector. A structural equation analysis of the effects of household and building characteristics on the annual energy consumption of US residential buildings. Energy Econ. 43, 178–184.
Sturm, R., Cohen, D.A., 2004. Suburban sprawl and physical and mental health. Public Health 118 (7), 488–496.
Tarozzi, A., Deaton, A., 2009. Using census and survey data to estimate poverty and inequality for small areas. Rev. Econ. Stat. 91 (4), 773–792.
Therneau, T.M., Atkinson, E.J., 1997. An Introduction to Recursive Partitioning Using the RPART Routines.
Tomczyk, A.M., Ewertowski, M., 2013. Planning of recreational trails in protected areas: application of regression tree analysis and geographic information systems. Appl. Geogr. 40, 129–139.
Travisi, C.M., Camagni, R., Nijkamp, P., 2010. Impacts of urban sprawl and commuting: a modelling study for Italy. J. Transp. Geogr. 18 (3), 382–392.
United Nations, 2016. United Nations Framework Convention on Climate Change (UNFCC). Conference of the parties. Report of the conference of the parties on its twenty-first session, Paris from 30 November to 13 December 2015.
Vicenzotti, V., Qvistrom, M., 2018. Zwischenstadt as travelling concept: towards a critical discussion of mobile ideas in transnational planning discourses on urban sprawl. Eur. Plan. Stud. 26 (1), 115–132.
Vittinghoff, E., Glidden, D.V., Shiboski, S.C., McCulloch, C.E., 2011. Regression Methods in Biostatistics: Linear, Logistic, Survival, and Repeated Measures Models, 2nd ed. Springer.
Wiesmann, D., Lima-Azevedo, I., Ferrão, P., Fernández, J.E., 2011. Residential electricity consumption in Portugal: findings from top-down and bottom-up models. Energy Policy 39, 2772–2779.
Wilson, B., Chakraborty, A., 2013. The environmental impacts of sprawl: emergent themes from the past decade of planning research. Sustainability 5, 3302–3327.
Wooldridge, J.M., 2011. Introductory Econometrics: A Modern Approach. Nelson education.
Xu, G., Jiao, L., Liu, J., 2019. Understanding urban expansion combining macro patterns and micro dynamics in three Southeast Asian megacities. Sci. Total Environ. 660, 365–383.
Yan, X.P., 1995. Chinese urban geography since the late 1970s. Urban Geogr. 16 (6), 469–492.
Yang, Y., Zhang, L., Ye, Y., 2019. Curbing sprawl with developing limiting boundaries in urban China: a review of literature. J. Plan. Lit.