Goerlich Gisbert, FJ.; Cantarino Martí, I.; Gielen, E. (2017). Clustering cities through urban metrics analysis. Journal of Urban Design. 22(5):689-708. https://doi.org/10.1080/13574809.2017.1305882

Clustering cities through urban metrics analysis

ABSTRACT

This paper describes a process for measuring and characterizing urban morphological zones. These urban zones are delineated for the entire area of Spain, independently of administrative boundaries and excluding demographic data, using a high resolution land use dataset. Given the rich information available on land cover and subsequently assigned population data, it is possible to calculate a set of urban spatial metrics to classify these urban zones into homogenous morphological groups. Four types of urban agglomerations are identified in Spain by working with these urban metrics and applying a final cluster analysis. Although these groups have a general complex monocentric typology, each has its own specific characteristics. Finally, a picture of patterns and trends of urbanization for the main urban agglomerations in Spain is provided, offering some perspectives on the urban sprawl phenomenon.

Keywords: Land use, Land Cover and Use Information System (*SIOSE*), urban morphological zones, urban areas, urban form, urban morphology, urban sprawl

Introduction

With the increasing acceptance of sustainable development as a guiding concept, researchers have focused renewed attention on matters of urban form that trace back to the start of modern planning and urban studies (Harris and Ullman, 1945; Conzen, 2001). The discussion on the form of contemporary city and its applications has therefore grown in importance in the last decades (Besussi et al., 2010, Schwarz, 2010, Lowry et al., 2013), especially as related to sustainable urban development, and urban and regional policies. The process of territorial planning is by nature complex and requires tools capable of analyzing a comprehensive set of information. There is an increasing demand to monitor and quantify urban dynamics to meet the challenges of regional planning.

In this paper, the target is to define urban typologies through the study of their general morphology or *urban forms*, composed mainly of the physical elements within a city. Working with spatial data, and some statistical data, a series of morphological quantitative indices (or "metrics") are calculated in order to analyze, and to understand, the urban forms. The definitions of urban form vary greatly in the literature. While some authors rely solely on land use/land cover to measure urban form in terms of the physical structure of a city (Herold et al., 2005 and Huang et al., 2007), others also include socio-economic aspects such as population number or density (Frenkel and Ashkenazi, 2008, and Tsai, 2005).

A specific high resolution Spanish land use dataset (*SIOSE*) is used to represent built-up areas. The most relevant feature of *SIOSE* is its object oriented data model (Valcárcel, 2011), since it gives more precision and detail to the polygon definition. Indeed, this Spanish model clearly improves on other hierarchical land use models such as, for example, the widely known Corine Land Cover (*CLC*) in Europe. The resolution of the *SIOSE* dataset has allowed to calculate some urban metrics in order to draw up a general characterization and identification of Spanish urban typologies.

In addition to the spatial limits defined with the land cover, demographic data and their spatial distribution should be considered to improve the study of urban forms. There are several methods to downscale population data, but the dasymetric-based methods are the most commonly used. Population reallocation can be based on different ancillary data, as land cover, cadastral information or even street network (Pavía and Cantarino, 2016). Beyond population data, some authors deal with the functional structure of the city as services, transport networks or economic structures.

For Weeks (2010), an urban area is *"a spatial concentration of people whose lives are organized around nonagricultural activities"* and is determined by the population size and density, and also by the land area and economic and social organization. In fact, the delimitation of urban areas has been a challenging problem for a long time, and there are no standard methods to solve it to date (see the review by Liang et al., 2010). This type of analysis generally emphasizes commuting, so that the ability to move from one municipality to another in the same day to work, study, shop, or for leisure activities is crucial to establishing the urban area range.

Urban areas are identified from a functional point of view as the place where residents carry out their daily activities, and are characterized by having one or several population nuclei. Due to the lack of available data and the low significance of this type of mobility until recent decades, these considerations were not taken into account in historical studies. However, they are becoming increasingly relevant in the contemporary world (Clifton et al., 2008; Parr, 2007), especially in Europe, where the higher levels of daily mobility that have appeared in the last decade have led to enlarged city size and to the need to consider many cities as part of a functional urban region (Parr, 2007).

Urban sprawl is a well-known topic in metropolitan studies, and is especially common in the United States (Brueckner, 2000). The phenomenon is directly related to increased commuter activity and enlarged city size. There is no general agreement about the exact definition of this term (EEA, 2016), but its effects are clearly perceived in the urban landscape as a consequence of an important increase of land occupation and enlarged city size, so that urban sprawl patterns make it difficult to establish urban boundaries and require a review of the definition of urban areas.

The classic papers of Galster et al. (2001) and Wolman (2005) give some dimensions of sprawl based on urban forms. Ewing et al. (2002) add the functional factors (activity centres, accessibility, for example). Burchfield et al. (2006) assign the causes of sprawl to factors such as dynamic processes based on population growth, although sprawl can also be estimated by evaluating indicators of dispersion and complexity with density and aggregation. In summary, urban sprawl comprises a combination of multiple aspects, among which form, density and land use patterns can be highlighted (Besussi et al., 2010).

According to Clifton et al. (2008), the first approach to characterizing cities is to define their urban form by analyzing and measuring their spatial pattern of land use distribution, or *urban structure*, formed mainly by a combination of physical elements, such as buildings, streets, etc. Nowadays, satellite images and high resolution land cover models offer an unprecedented opportunity to develop the more precise comparative indicators needed. By employing this data for the first time in a national comparative analysis of systematic indicators, this study aims to strengthen the understanding of the national variants in urban form, particularly between the different regions and even lifestyles in Spain (rural dispersion in the northern green belt). From the abundant information available on land cover and population distributions it is possible to draw an accurate picture of urban structures, and take it as the starting point.

The urban spatial structure is, firstly, a way of organizing named physical elements. In general terms, however, the urban structure has several dimensions, including functional, demographic and economic dimensions. Indeed, the urban form of a specific city is the result of a variety of influences, including site and topography, economic and demographic development and past planning efforts (Batty and Longley, 1994). Further dimensions should also be added to obtain the form of spatial structure manifested and uncover these urban patterns.

The analysis of spatial structures is central to geographic research. Spatial primitives such as location, distance, direction, orientation, linkage, and their patterns have been discussed as general spatial concepts in geography (Golledge, 1995). Here these basic spatial concepts and the analysis of spatial structure and its urban patterns will be approached from the perspective of spatial metrics. These tools are useful to objectively quantify and describe the underlying structures and patterns of the urban landscape from geospatial data (Pham and Yamaguchi, 2011).

Developed in the late 1980s, the analysis of spatial morphology in natural landscapes (by means of *landscape metrics*) is a technique used to quantify the shape and pattern of vegetation.The work of McGarigal et al. (2002) and their software *Spatial Pattern Analysis Program for Categorical Maps* (FRAGSTATS) is one prominent example. Applied to fields of research outside landscape ecology and across different kinds of environments (in particular urban areas), the approaches and assumptions of landscape metrics may be more generally referred to as ‘‘spatial metrics”. These metrics are essential to better understand the characteristics of a landscape and can be defined in general as “*numerical indices to describe the structures and patterns of a landscape*” (as cited in Bhatta, 2010, p 87). Herold et al. (2005, p. 288) also defined spatial metrics as, "*quantitative and aggregate measurements derived from digital analysis of thematic-categorical maps showing spatial heterogeneity at a specific scale and resolution*".

Scholars have presented several urban form indices to attempt to define their characteristics (size, extent, land use organization, demographic distribution, and so on). Huang (2007) analyses 77 cities worldwide by means of seven quantitative indices representing five physical characteristics of the urban form, namely, compactness, centrality, complexity, porosity and density. Colaninno et al. (2011a) identify nine useful indices for quantifying the morphology of the city, based on several features, and calculated according to the formal and relational characters of the buildings. Schwarz (2010) reviews the urban metrics and groups them into two types: landscape metrics (with 27 different indexes) and socio-economic indicators (18). Finally, Lowry (2013) compares 18 spatial metrics and organises them into four urban form categories: density, centrality, accessibility, and neighbourhood type.

Researchers and practitioners aiming to quantify the urban form of a single city or a whole range of cities can choose from numerous indicators. At least two strands of discussion with respect to measuring urban form are distinguished: landscape metrics and socio-economic indicators. Landscape metrics identify landscape forms through map analysis of land use or cover. Population-related indicators for measuring urban form are also discussed in the literature; some examples are population number, population density, or the administrative area of the city. Finally, the broadest definition possible of urban form is used for this paper. Accordingly, urban form here encompasses the physical structure and size of the urban fabric as well as the distribution of population within the area, following the studies of Huang (2007) and Schwarz (2010).

A wide variety of spatial metrics have been created to characterize and quantify urban form (Glaser et al., 2001, Glaser et al., 2001, Galster et al., 2001, Ewing et al., 2002, Weston, 2002, Frenkel and Ashkenazi, 2008 and Clifton et al., 2008). The increased use and growing demand for spatial metrics are due in part to the availability of commercial and open source computing tools, such as geographic information systems (GIS), which can store and analyze large amounts of spatial data, and the increased availability of spatial data in the public domain (Kerski and Clark, 2012).

Methods

To approach the study of forms of urban areas, firstly it is necessary to define their limits and the method to obtain them. This definition is not trivial because the limits of urban areas are not defined simply by their administrative or official limits, and a more precise definition is needed based on urban fabric, without including rural or uninhabited areas. Several methods are available to obtain the spatial scope of urban areas (see Liang and Mao, 2010). In this paper, and as explained above, land cover datasets are used to define the Spanish urban areas following the methodology applied by the EEA. This provides more possibilities to apply several urban metrics.

Having delimited the urban areas, their population is assigned in grid format by means of a dasymetric method described in Goerlich and Cantarino (2013). The demographic information was taken from the census tract of official National Institute of Statistics data (INE, 2006). This urban area demographics model allows analyzing both population and land use. Its structure and urban patterns can be described through numerical indices, which have been applied extensively in the last decades (Bracken, 1994).

The spatial metrics selected are derived from population and urban form. This is the most important information for analysing cities with functional data, specifically, density, population spatial correlation, and several indexes for defining form. These metrics are similar to those selected by Schwarz (2010) to characterize two hundred European cities.

The last step was to group the spatial metrics data set of the defined urban areas through a cluster analysis. This analysis identified some homogeneous groups that enabled the characterization of all the urban form patterns studied and to satisfactorily explain their meaning (see Figure 1 for an explanation of the process).

Insert Figure 1 here. Flow diagram of complete process.

Land cover models

The usual way of representing urban areas in a region is by selecting the built-up areas, normally by means of remote sensing techniques and then integrating them in a land use dataset. In the European region, the best known reference model is *CORINE* Land Cover (*CLC*), which was established by the European Commission and developed by the European Environment Agency (EEA) in 1985 in order to compile, coordinate, and homogenize the information on the status of the environment and natural resources in Europe. It is a hierarchical-type georeferenced model which divides the land into relatively homogeneous polygons assigning them a unique cover of up to 44 classes at the higher disaggregation level.

However, for the purpose of the present study, a high resolution product derived from remote sensing, recently introduced into Spain, was used. The Land Cover and Use Information System of Spain, (*SIOSE*) was developed by the Spanish National Geographical Institute (*IGN*) and aims to solve most of the problems arising from the poor resolution of the *CLC*. The *SIOSE* dataset has an object oriented structure and provides different and potentially infinite combinations of the possible land cover elements. The information is highly versatile and can be adapted to the researcher's needs, although it is also much more complicated to manipulate than the classical hierarchical land cover models of the *CLC* type. Its polygons are complex, and contain combinations of different covers and attributes. In these conditions, a hierarchical model with a single coverage per polygon, similar to the one in the *CLC* model, is much more versatile as a working tool, without losing the rich amount of information in the original database model.

For these reasons, a new model with hierarchical nomenclature based on *SIOSE* has been developed, called the *SIOSE* Hierarchical Model (*SHM*). This model respects the geometry of the initial database but each polygon can only be assigned to a class in the hierarchical nomenclature, and at the same time all polygons must be classified. The model production methodology, validation, application and results of the *SHM* model are thoroughly explained in Cantarino (2013).

Defining urban areas

Urban areas are complex objects that can be approached through different definitions and delineations. As seen before, urban areas can be characterized through several types of structural dimensions, of which the physical and demographic elements were selected. Specifically, to define the limits of the urban areas, only the data available in the SIOSE dataset have been used, following the "morphological" city delineation shown by the European Observation Network for Territorial Development and Cohesion (ESPON, 2015).

The main bases for work defining the urban areas for the whole European territory are Morphological Urban Areas (*MUA*) and Urban Morphological Zones (*UMZ*). The first, the MUA database, was created in 2007 and updated in 2011 by IGEAT (Institut de Gestion de l'Environnement et d'Aménagement du Territoire, Université Libre de Bruxelles) and is mainly based on the selection of the most densely populated municipalities (LAU2), specifically, those over 650 inhab/km^2 (IGEAT, 2007).

The second term, *Urban Morphological Zones*, is the most interesting European morphological urban database and the only one that is freely available. It is defined at the European scale with harmonised criteria, updated regularly by the EEA, and includes all cities with populations over 10,000. The name UMZ refers to the fact that they are determined exclusively by the polygons defined in land use and land cover (LULC) datasets. They were created by the EEA on the basis of the information on LULC provided by the CLC. A UMZ is defined as "*a set of urban areas lying less than 200 m apart*" in which urban areas are identified by the land covers that contribute to the urban fabric and typically urban functions (Simon et al., 2010, p. 4). This concept is the direct precedent of the approach chosen to delineate the urban areas in this study.

The method used to obtain the urban morphological zones is based on the *SHM* database. This one differs substantially from the EEA method, which uses a raster approach to carry out the proximity calculation and changes the original contours of the polygons in the contact areas and angular edges. As a consequence, the EEA UMZs have no information on the internal land coverage types (see Table 1 for a comparison of different urban area delineations).

Insert Table 1 here. Main characteristics of urban agglomeration definition models in Europe.

Indeed, the *SHM* UMZ*s* allow to examine their coverage *SIOSE*-based in such a way that it is always possible to refer to the initial database, which is necessary to generate urban indices (see Section 2.3). Thus, a total of 5,589 UMZs were obtained, by establishing a UMZ minimum surface area of 20 ha, slightly more than those offered by the EEA.

Finally, the Statistical Atlas of Spanish Urban Areas (Ministry of Public Works and Transport, 2006) is the only official data source for urban areas in Spain with which to validate the UMZs. To compare the two urban areas, the total of 83 LUA (Large Urban

Areas) defined in this atlas was used. This method takes the municipalities and their population size as its starting point, so that a LUA must have more than 50,000 inhabitants. In conclusion, the LUA zones, drawn by municipality boundaries, are too coarse for sensitive urban area definitions since they are over-dependent on their administrative limits. Figure 2 shows the comparison of LUA and UMZ for Madrid, and the UMZ cover distribution.

Insert Figure 2 here. Comparison of LUA and UMZ for Madrid. UMZ covers distribution.

Urban demographics

The *SHM* UMZ population was assigned through the population grid constructed by Cantarino (2013) with a resolution of 1 km^2 and applying dasymetric techniques. The reference date was 2006 for population and 2005 for the land use dataset. The demographic information was taken from the Census Tracts population of the INE (2006). It also uses the SIOSE residential land coverage as auxiliary information, since the population distribution criterion used is based on four types of built-up residential areas with their specific population densities, and finally adjusted by means of an iterative process. This is the same dasymetric criterion applied to allocate the population on a smaller cell grid.

This grid includes the entire Municipal Population Registry: 44,708,964 inhabitants for the year 2006. The full process is run in vector format and consists of the intersection of the layer that contains the 5,589 UMZs with the vector layer of the population grid.

After the intersection process, a minimum threshold population of 10,000 inhabitants was established to classify a UMZ as a genuine urban area; this is the same EEA UMZ limit. Following this criterion, the original number decreases to 415 UMZs, which takes up a surface area of 5,926 km^2, and are inhabited by 33,151,122 people, 74.1% of the total. Furthermore, 101 UMZs have more than 50,000 inhabitants and 56 more than 100,000, the latter containing 66.9% of the urban population. The surface area is distributed somewhat more homogeneously; in any case, population density, in terms of UMZ surface areas, increases monotonously with respect to population size. The high value of the largest urban areas stands out.

Urban form through spatial metrics

As explained in the Introduction section, urban form is considered to configure inhabitant distribution throughout the urban space. In general, urban form can be classified by four key metrics: density, land use, connectivity and accessibility. However, in this research, activities or a description of urban functionality are not included. This simplification is dictated by the description of the UMZ that will be used, which is based solely on population and urban morphology.

After obtaining the dataset that contains the urban form elements to be considered, it is necessary to find a way to quantify the morphology. A wide variety of spatial metrics have been created to characterize and quantify urban form (Altieri et al., 2014, Ewing et al., 2002, Frenkel and Ashkenazi, 2008, Galster et al., 2001, Glaser et al., 2001). However, it should be remembered that there is no defined set of specific indicators for use in urban geography, as the significance of spatial metrics varies with the objective of the study and the characteristics of the urban landscape under investigation (Clifton, et al., 2008).

The different urban forms can be catalogued and reflected according to their spatial metrics based on quantitative indices representing the physical characteristics of the landscape mosaic, population and its distribution (Schwarz, 2010). The principle behind this claim is that an urban area, as a self-organism, has a unique identity that is preserved through the years, despite the passage of time and the spatial and functional transformations that occur (Frenkel, 2004). The configuration of the physical elements, with their own functional dynamics, produces different "drawings" of cities. By disaggregating the urban texture into different components, the topological and geometrical indicators can be studied to discover the urban patterns and other systems that affect urban morphology. (Colannino et al, 2011b)

In this study, simple quantitative indices obtained from population, land use elements and UMZ geometry are initially applied. According to Tsai (2005), *"Metropolitan form can be analyzed as four distinguishable dimensions: size (total population), intensity (population density), the degree of inequality of distribution (concentration of population in a small proportion of the urban space) and the degree of clustering, which is the tendency for dense areas to be located next to each other".* This author bases his analysis on the decentralization of population or employment by means of Shannon's entropy or coefficients like those of Geary, Gini, and Moran, the latter two being the most useful.

Fortunately, both the UMZ definition and the population data distributed in cells on UMZ allowed making these calculations. A set of indices to distinguish the main basic urban forms (for example, the monocentric/polycentric scheme) can be applied, which are sufficiently explanatory for the first classification of the UMZ.

Furthermore, the analysis has also identified characteristics of urban sprawl. According to Tsai (2005), the Moran and population density indexes can distinguish sprawl from compact distribution. However, a complete assessment of urban sprawl is beyond the scope of this research. Indeed, although a wide variety of form indicators are used, functional factors like employment or commuting distance are not taken into account, which are strongly associated with sprawl (Besussi et al, 2010).

The urban models of territorial occupation are analyzed by applying quantitative indices concerning the form and structure of the land cover polygons and the intensity of urbanization in terms of population density and occupied areas. Following the works of Colaninno et al. (2011b) and Huang et al. (2007), the select indices are shown in Table 2. They were computed for the 101 UMZs with populations above 50,000 inhabitants.

Insert Table 2 here. List of indices used

- Gross Density (GD) is a measure of the relative magnitude of a UMZ and describes the intensity of the population pressure over the whole area.
- Net Density (ND) measures the population concentration (intensity) in the built-up area. It also assesses the real "densifying" of an urban area (human pressure on the housing area) and can estimate the mean residential building height.
- Ratio Open Space (ROS) quantifies the ratio of natural and artificial urban green areas to total UMZ surface. It is crucial both as an amenity for residents and for the sustainability of cities, and indicates the degree of development.
- Degree of Municipality Division (DMD) shows the degree of dispersion and complexity of a UMZ, according to the surface municipalities clipped by that UMZ. This index is defined as the probability that two randomly chosen places in the

municipality under study are not situated in the same undissected area (adapted from Jaeger, 2000). It takes the value zero when the UMZ is inside one municipality only.

- Compactness Index (CI) is the ratio of the surface–perimeter to a standard circle reference and indicates the irregularity of the UMZ boundary.
- The Gini coefficient (GC) is a measure of statistical dispersion and is the most commonly used measure of inequality. The coefficient varies between 0, which reflects complete equality and 1, which indicates complete inequality. This coefficient calculates the level of population concentration in the UMZ.
- Moran's index I (MI) is a measure of spatial autocorrelation, i.e., the degree of aggregation or clustering. It is characterized by a correlation in a signal among nearby locations in space. For urban agglomerations, this index decreases with decentralization and dispersion, and is normally used for assessing sprawl and characterizing landscape fragmentation (Tsai, 2005; Fan and Myint, 2013).

Results

Relations among the indicators

The correlation analysis of 101 UMZs shows an adequate relation among most of the spatial metrics indicators (see Table 3). It is noteworthy that the overall compactness (CI) correlates very strongly with the other indices. In general, the shape indices are well correlated with each other. The main exception is that MI does not present a significant correlation with the ROS and DMD densities, nor does the Gini coefficient with the ND (see p-values). However, MI is an important index in that it reflects the patterns of population distribution within the UMZs. Tsai (2005), in his study of 219 U.S. metropolitan areas, also found no statistical correlation between Moran, Gini and density.

Insert Table 3 here. Correlation matrix. UMZs with more than 50,000 inhabitants

The strong collinearity between the density indicators is noteworthy and suggests one of them could be eliminated to avoid redundancy. ND will be kept because it offers more information than GD on the occupation of buildings.

Another variable to highlight is the Moran index (MI), which was calculated by a Python routine in ArcGIS 10.2 (with the inverse weighted Euclidean distance option), on population data distributed in cells of 1 km^2, as described in the previous section. As Le

Néchet (2010) has pointed out, this index was founded to be affected by the quantity of data analyzed and to be quite low for small sizes. This is a manifestation of a scale problem which could exert unspecified influence on the results of spatial analysis (Bhatta, 2010).

The ArcGIS Python script thus calculates a default threshold distance that ensures each feature will have at least one neighbor. For the data processed, MI reaches values of 0.60 - 0.70 for large UMZs in the study, while it falls rapidly for the smaller ones. However, the statistical significance (p-values higher than 0.05) for the smaller ones is enough to prevent rejection of the null hypothesis of spatial dependence.

Previous analysis of most populated UMZ

With the traditional approaches, based on the density and distribution of municipalities, a first exploration of some data of the main UMZ allows to make an initial approximation of Spanish urban forms and then carry out a cluster analysis to obtain some basic classification.

The results for these UMZs can also be analyzed first with magnitude indicators (GD) and the intensity of population in the housing environment (ND), reported in Table 4. The large differences in certain indices that define these UMZs are explained below.

Insert Table 4 here. Top ten Spanish UMZ by population.

For the municipalities, the differences in number/total area shown in Table 4 are due to the high degree of heterogeneity in the size of municipalities in Spain, for historical, social and economic reasons, especially in Madrid. The capital of Spain is located in the center of the country, in an area that was originally agricultural, of low performance and value, which gave rise to large administrative units. This, together with a high ND, points to a monocentric general structure with population subcenters.

In contrast, the other UMZs (Barcelona and Valencia) are on the coast, where precisely the opposite occurs. Furthermore, the high number of small municipalities in Barcelona suggests a significant complexity (high DMD) and some degree of polycentrism in comparison with other regions.

Especially prominent are Bilbao and San Sebastian as extensive and complex cities, but associated with high urban density. Their structures show a concentrated urban core and a considerably dispersed peri-urban area, in accordance with their location in the wet temperate belt of northern Spain (oceanic climate).

On the other hand, Seville, Zaragoza and Malaga have high density and a low number of municipalities, which suggests a monocentric scheme (especially in the last two) but with some population subcenters (see Figure 3). Finally, Alicante (Mediterranean) and Tenerife (Canary Islands) are less extensive with a more dispersed population, partly as a result of being on the coast and having warm dry climates. In summary, these are the scattered or expanded cities.

Classification of UMZ by cluster analysis

The 101 UMZs with more than 50,000 inhabitants could now be re-classified by means of a cluster analysis. For this analysis the indices shown in Table 2 were used, except the GD index, as explained above. First, the preceding hierarchical analysis showed that all the UMZs could be classified into 4 or 5 groups. Building on this result, four types were designated in the subsequent K-Means cluster analysis, with data standardization. The K-Means method has the advantage of allowing the group centers to be adjusted iteratively.

After the first trials, and from the results given in the preceding section, four characteristic Spanish cities were selected as seeds in the cluster analysis: Madrid (largest UMZ), Zaragoza (monocentric), Alicante (expanded) and Córdoba (small traditional city). With a few qualifications, the resulting classifications reaffirmed the broad contrasts between these types of urban forms of Spanish UMZs (Table 5).

Insert Table 5 here. Cluster centroids.

The first cluster with Madrid as the seed (6% of the total) includes the biggest UMZs with low compactness, high density and complexity and more than 300,000 inhabitants. All are provincial capitals with high ROS, and the MI is higher, due to its high density. The main examples are shown in Table 4, in the first positions. The second cluster represented by Zaragoza (29%) presents median density, compactness and low dispersion: monocentrism is quite high, with some population subcenters.

The third cluster represented by Alicante (38%) includes UMZs with low density and compactness, moderately high ROS and dispersion. It is worth noting that some

UMZs in this group are Mediterranean tourist and leisure centers with high occupation in summer such as Malaga-Marbella.

The fourth cluster with Córdoba as the seed (28%) includes highly monocentric small and traditional urban areas: high density, compactness and low ROS, dispersion and clustering. However, the MI is lower, due to their small surface areas. These are mainly inland UMZ < 300,000 inhabitants in the center or south of Spain, little developed and with the classical structure of old European cities. Coastal exceptions are due to geographical constraints.

A new grid was constructed for the graphical output of the population with the purpose improving the detail of the maps. This grid was calculated following the same dasymetric distribution process and source data mentioned in the Methods section. Thus, a 200m cell size for the graphic representation of the four types of cities was obtained, represented by Madrid, Zaragoza, Alicante and Córdoba. The representation of the non-residential urban areas present in the definition of UMZs (activity areas, infraestructure areas and open areas) was added (see Figures 3 and 4).

Insert Figure 3 here. UMZ Zaragoza, Alicante and Córdoba with grid cells population.

Insert Figure 4 here. UMZ Madrid with grid cells population. Same legend as Fig. 3

Discussion

The cluster analysis provided some highly significant results following the successful classification of all the Spanish agglomerations of over 50,000 inhabitants. Four homogeneous groups were clearly identified that satisfactorily characterized all the urban forms under study.

The indices used to define the clusters proved to be suitable and sufficiently significant in analyzing the correlations. However, the formal definition of the UMZs can almost be said to necessarily involve a monocentric scheme, since the polycentric types require urban centers separated by open areas, which is not the case here. ESPON (2011) supports this approach and in the UMZs designed by the EEA the normal scheme is to have only one main center (or “strong core”, representing 94% of European UMZ). Polycentric forms (“several cores”) are recognized in the industrial conurbations, such as the case of France, Italy, United Kingdom, Germany, etc. The cited suitability of Moran’s

Index to distinguish monocentrism from polycentrism is therefore not applicable, although it has been shown to be suitable for separating urban concentration-dispersion, in conjunction with other indices.

It should also be pointed out that Gini's coefficient becomes saturated with the high population variance values shown by the UMZs, although it did not perform differently from the others in the two intermediate groups (see Table 5). However, in combination with Moran's Index of correlation it does offer interesting aspects: a high Gini value combined with a moderate Moran coefficient indicates a certain degree of attenuated decentralization, as in the cases of Donostia and Bilbao, which show indications of urban sprawl. On the other hand, low Gini values combined with high Moran values show a strong degree of monocentrism, as in the case of the Córdoba UMZ, an example of a compact medium-sized area that coincides with its municipality.

Therefore, as mentioned with regard to Table 4, t the choice of seeds was the right one, as the results obtained. For example, Zaragoza as a type of low-complexity, clearly monocentric (MI = 0.48) compact city with a highly-concentrated population. On the other hand, Alicante is seen as a more complex decentralized city with low population density and low MI (0.29) for its size. Finally, Córdoba is a low-complexity, compact, monocentric (MI = 0.34) city with medium-high density.

In general, this analysis did not find any purely monocentric urban systems or any that were significantly affected by sprawl. The predominant urban morphology is a moderately monocentric type and is coupled with smaller population subcenters, generally associated with the municipalities it includes. This type of more or less decentralized monocentrism suggests the term *acentrism*, as proposed by Le Néchet (2010), in which densely populated urban centers lose their importance.

The variation interval over the mean for the six indices (see the centroids value in Table 5) was calculated, assuming its normal distribution. The level of significance is 98%, applying the t-Student's distribution to distinguish different population samples. Results are shown in Table 6. The intervals overlapped in the same index have been shaded for optimum comparison.

Insert Table 6 here. Interval value indices (t-Student's 98%).

Cluster 1 shows a clearly monocentric morphology, the most important example of which is Barcelona, with the highest MI (0.58). Madrid and Valencia are somewhat more acentric (0.50 and 0.48, respectively), as is Bilbao (0.35), but the latter's high GC value (0.82) seems to indicate a certain degree of decentralization. In fact, in this case it can be seen that a large part of the population is widely dispersed following the traditional pattern in the north of Spain's wet belt and repeated in the country's coastal and frontier regions. Donosti is in a similar situation, with the lowest MI of the larger cities (0.20). According to Table 6, these Moran values are quite similar to those of Cluster 2. Two cases in Cluster 3 deserve special attention; the first is Elche (in Alicante, SE Spain), with the highest Gini value (0.884). Its open and complex urban structure gives it an MI of 0.19, although its population is widely dispersed around subcenters. Durango (Vizcaya, northern Spain) also has a high Gini (0.77) and a low MI (0.06) and a dispersed urban morphology in a series of secondary centers. These examples of sprawl have two different origins: the first one would appear to be a case of a recent residential spread, while the latter is the result of the typical dispersed settlements of Spain's northern wet belt mentioned above.

The study of centroids in Table 5 and their intervals in Table 6 provide some keys to apply to the knowledge of the urban sprawl phenomenon that appears in Cluster 3. The most significant indices to describe the sprawl and its values are low density (ND), high green areas (ROS), high administrative fragmentation (DMD) and low autocorrelation (Moran index). The Gini value (GC) is not sufficiently significant to separate Cluster 2 from Cluster 3 (or even Cluster 1). According to Tsai (2005), this index seems to signal an unequal population distribution, and may be conceived as a dimension of urban form, rather than sprawl.

The results of the cluster analysis for the Spanish provincial capitals are summarized in Table 7:

Insert Table 7 here. Results of the cluster analysis: groups of Spanish provincial capitals.

Oddly, there are no remarkable differences in the population density when Cluster 1 and 4 are contrasted (Table 6). Indeed, urban dispersion has advanced very rapidly and often uncontrollably in some big Spanish (and European) cities, where urbanization expands at much faster rates than population growth. These urban areas are experiencing a change toward more dispersed and horizontal rather than vertical growth at the expense of farming and forested areas, semi-natural environments, and wetlands. This trend serves to attenuate the levels of over-densification of central districts and to reduce mean density (Catalán et al., 2008).

Cluster 4 contains cities that have not lost their original historical hierarchical structure or undergone extensive urban development, mostly because they have small populations and are situated inland. This means they have non-complex monocentric morphologies associated with a small incidence of urban green areas and fairly high population densities, similar to those of Cluster 1 (Table 6). The cases involving peripheral coastal locations are due to restrictions on urban expansion and lead to concentrations of the population. These restrictions can be due to natural geographical features, such as the Cádiz isthmus, the Huelva wetlands, coastal mountain ranges, and artificial features such as national frontiers or linear infrastructures. Population growth in coastal areas was also observed, especially along the Mediterranean shore due to seasonal tourism and the acquisition of second homes, which gives rise to diffuse urbanizations close to the more consolidated, and often small, urban centers. Their low densities and monocentristic features accompanied by a certain degree of complexity logically situates them in Cluster 3, as in the case of the examples cited above. Indeed, the intense urban expansion observed in most Mediterranean areas was in part supported by the impotence of planning control, in many cases a consequence of high administrative fragmentation (DMD index) and the lack of a common policy for the metropolitan area.

Conclusions

Although most of the population lives in urban environments, there is no general consensus on the definition of urban areas or agreement on the methodology for their classification. Many definitions rely too heavily on demographics and the administrative boundaries that constitute the reference for the compilation of demographic statistics. Part of the reason for this is that population figures and surface areas in which populations are located are easy to obtain and work with.

The UMZs defined in this paper are based exclusively on morphological criteria from the Spanish High Resolution Land Use database (SIOSE), and contain, first, all the original cover information and, second, their population from a population grid in vector format. The calculated spatial metrics are based on this information.

According to Table 6, these spatial metrics seem a good choice since they are complementary and can help to explain the diversity of urban structures. The only index without a clear application may be the Gini coefficient. This metric did not show a good fit to define the four types of UMZs, as it coincided in three of the types obtained.

The functional aspects linked to commuting were not taken into account because they are not directly related to land coverage and information on them is lacking, and they depend on the availability of an urban transport network. The resident-workplace relationship is an aspect requiring further research in order to correctly delimit metropolitan areas or large urban zones (OECD, 2012).

Moreover, there are numerous advantages to classifying urban areas by analyzing quantitative UMZ-form metrics. The main advantage is that it provides a spatial description summary of urban areas. This analysis can also capture minor differences between complex urban forms. However, the quantitative metrics do not provide as clear an image as that provided by maps. It is therefore difficult to choose the correct urban model from metropolitan form indices only: two UMZs with the same index value can represent two very different forms. The process cannot therefore be fully automated since the results would require expert revision before reaching final conclusions.

Remote sensing data and GIS have opened up a set of possibilities to study urban form. The comparative analysis of these methods enables more comprehensive and more systematic results than were previously possible. Both the regional averages and the individual patterns in these spatial indicators confirm the effects of contemporary development and the legacy of most of Spanish historic cities linked to them. The compactness, density and regularity of inland urban areas generally exceed the levels of the larger coastal ones.

Thus, the most important conclusion of the UMZ cluster analysis is that there is not only one urban agglomeration model, and that the polycentric/monocentric dichotomy does not appear to be relevant in this research. Indeed, social preferences for a suburban and dispersed life style are not seen as clearly here as they may be in North American or European countries, and in Spain there is a certain preference for urban centrality and proximity (Catalán et al., 2008). However, other authors using methods based on the functional connection of urban centers have indeed described polycentric morphologies in Spanish urban areas (Cladera, 2012).

Spanish urban areas can be described as an amalgam of complex juxtapositions of various types. The most commonly found example of a UMZ is a monocentric system with ill-defined subcenters that partially maintain the original hierarchy of the historical city, but with a clear tendency toward complexity, decentralization and sprawl in the larger cities. Inland cities still maintain, in part, those limits with a markedly monocentric morphology, though with a rather dispersed population. The most open and complex cities tend to be found near the coast and frontier regions, especially in the mild and wet belt in northern Spain and on the Mediterranean coastline, where the seasonal fluctuations of the population, together with a complex set of municipalities, are largely responsible for these characteristics.

Through the results of the cluster analysis, an interesting application of this study is obtained. Urban sprawl is closely related to net density, open spaces, administrative fragmentation and the population autocorrelation signaled by the Moran index (values in Table 6 for Cluster 3). These values provide an objective and appropriate assessment of the complex phenomenon of urban sprawl and separate it from a compact urban distribution.

Given the schedule for *SIOSE* updating, the explained methods can be routinely applied to update the hierarchical model and the morphological zones, and then to study changes in urban land size and cover. Moreover, the analysis of dynamic changes in land cover or population is commonly used by scholars to evaluate sprawl (Colaninno, 2011b), and too for understanding the urban regional interactions and urban habitats (De Noronha and Vaz, 2015; Vaz, 2016)

In summary, urban areas are dynamic spatial entities that are highly complex due to many factors involved in their structure. This study presents one of the first results of the general study of Spanish cities through their spatial structure based on land cover data and proposes some research lines for the study of the population spread phenomenon. The procedure described in this paper can be applied to other countries, as long as they have detailed land cover datasets and georeferenced populations. This spatial characterization of urban areas offers urban planners objective support for organizing urban development, designing appropriate city policies and allocating public resources where they are really needed.

References

Altieri, L. Cocchi, D., Pezzi, G., Scott, M. and Ventrucciet, M. 2014. "Urban sprawl scatterplots for Urban Morphological Zones data". *Ecological Indicators*. Volume 36, 315–323.

Besussi, E., Chin, N., Batty, M., and Longley, P. 2010. The Structure and Form of Urban Settlements. In: *Remote Sensing of Urban and Suburban Areas,* 13. T. Rashed and C. Jürgens (eds.), Remote Sensing and Digital Image Processing 10, DOI 10.1007/978-1-4020-4385-7_2, Springer Science+Business Media

Bhatta, B. 2010. *Analysis of urban growth and sprawl from remote sensing data* (1st ed.): Springer Berlin Heidelberg.

Bracken, I. 1994. "A surface model approach to the representation of population-related social indicators". In Fotheringham S. and Rogerson P. (eds), *Spatial Analysis and GIS*, London, Taylor and Francis, 247-259

Brueckner, J. 2000. "Urban sprawl: diagnosis and remedies". *International Regional Science Review* 23, 2: 160–171.

Burchfield, M., Overman, H., Puga, D., and Turner, M. 2006. "Causes of sprawl: A portrait from space". *Quarterly Journal of Economics* 121, (2), May 2006: 587-633,

Cantarino, I. 2013 ……

Catalán, B., Saurí, D. and Serra, O. 2008. "Urban sprawl in the Mediterranean?: Patterns of growth and change in the Barcelona Metropolitan Region 1993–2000". *Landscape and Urban Planning*, 85, 174-184.

Cladera, J.R., Bergadá, M. and Arellano, B. 2012. "El sistema urbano en España". *Scripta Nova,* XVI (396).

Clifton, K., Ewing, R., Knaap, G. and Song, Y. 2008. "Quantitative analysis of urban form: a multidisciplinary review". *Journal of Urbanism: International Research on Placemaking and Urban Sustainability*. Vol. 1, Iss. 1.

Colaninno, N., Cladera, J.R., Pfeffer, K. 2011a. *An automatic classification of urban texture: form and compactness of morphological homogeneous structures*. In Barcelona, In: 51st European Congress of the Regional Science Association International ERSA 2011, Barcelona, Spain.

Colaninno, N., Cerda, J. and Roca, J. 2011b. *Spatial patterns of land use: morphology and demography, in a dynamic evaluation of urban sprawl phenomena along the Spanish Mediterranean coast*. In: 51st European Congress of the Regional Science Association International ERSA 2011, Barcelona, Spain.

Conzen, M.P. 2001. “The study of urban form in the United States”. *Urban Morphologhy*, 5 (1), 3–14.

De Noronha, T., and Vaz, E. (2015). Framing urban habitats: The small and medium towns in the peripheries. Habitat International, 45, 147-155.

ESPON. 2011. *Naming UMZ: a database now operational for urban studies.* Technical report. On ESPON 2013 Database Project.

European Environmental Agence (EEA), 2016. *Urban sprawl in Europe*. EEA report 11/2016

Ewing, R., Pendall, R., and Chen, D. 2002. *Measuring sprawl and its impact*. Smart Growth America, Washington DC. Avalaible on: http://www.smartgrowthamerica.org/sprawlindex/MeasuringSprawl.PDF.

Fan, C. and Myint, S. 2014. “A comparison of spatial autocorrelation indices and landscape metrics in measuring urban landscape fragmentation”. *Landscape and Urban Planning*, Vol. 121, 117-128.

Frenkel, A. 2004. “The potential effect of national growth-management policy on urban sprawl and the depletion of open spaces and farmland”. *Land Use Policy*. Vol. 21, 321-382

Frenkel, A. and Ashkenazi, M. 2008. “Measuring urban sprawl: How can we deal with it?”. *Environment and Planning B: Planning and Design*, 35. 56–79

Galster, G., Hanson, R., Wolman, H., Coleman, S., and Freihage, J. 2001. Wrestling Sprawl to the Ground: Defining and Measuring an Elusive Concept. Housing Policy Debate 12, 681-717. Available on: www.mi.vt.edu/data/files/hpd%2012(4)/hpd%2012(4)_galster.pdf (accessed on February 21, 2009)

Goerlich, F. and Cantarino, I. 2013, ……………..

Golledge, R. G. 1995. *Primitives of spatial knowledge*. In: Nyerges et al. (Eds.), Cognitive aspects of human-computer interaction for Geographic Information Systems (pp. 29–44)

Glaser, E.L., Kahn, M.E. and Chu, C. 2001. Job sprawl: employment location in U.S. metropolitan areas. The Brookings Institution, Center on Urban & Metropolitan Policy, Washington DC. Available on: http://www.brookings.edu/es/urban/publications/glaeserjobsprawl.pdf

Herold, M., Couclelis, H., Clarke, K. 2005. "The role of spatial metrics in the analysis and modelling of urban land use change". *Computers, Environment and Urban System* 29 (4), 369-399.

Harris, C.D. and Ullman, E.L. 1945. "The nature of cities". *Ann. Am. Acad. Polit. So. Sci.*, 242, 7–17

Huang, J., Lu, X., Sellers, J. 2007. "A global comparative analysis of urban form: Applying spatial metrics and remote sensing". *Landscape and Urban Planning*. Volume 18, Issue 4, 184-197.

IGEAT et al. 2007. *Study on Urban function*. Final Report, ESPON 1.4.3. Program 2006.

IGN (National Geographic Institute) 2011. *Land Cover and Use Information System (SIOSE)*, Technical document Version 2.0, Madrid. Available on: htpp://*inspire.ec.europa.eu/ref_ser.cfm?id=33399*

Jaeger J. A. G. 2000. "Landscape division, splitting index, and effective mesh size: new measures of landscape fragmentation". *Landscape Ecology*, 15: 115-130.

Kerski, J. and Clark, J. 2012. *The GIS guide to public domain data*. ESRI Press, Redlands, CA.

Le Néchet, F. 2010. *Quantifier l'eloignement au modele de Bussiere : monocentrisme contre "Acentrisme".* in: Foltete J.-C., Tannier C. (dir.), *Actes des Neuviemes Rencontres de Theo Quant, Besancon.*

Liang, J., Li F., and Mao, L. 2010. *Review of the methods of delimitation for the spatial scope of urban agglomeration*. In: The 18th International Conference on Geoinformatics: GIScience in Change, Geoinformatics 2010, Peking University, Beijing, China, June, 18-20.

Lowry, J. and Lowry, M. 2013. "Comparing spatial metrics that quantify urban form". *Computers, Environment and Urban Systems*, vol. 44, 56-97.

Huang, J, Lu, X.X. and Sellers, J.M. 2007. “A global comparative analysis of urban form: applying spatial metrics and remote sensing”. Landscape Urban Plan., 82 (4), pp. 184–197.

McGarigal, K., Cushman, S. A., Neel, M. C., and Ene, E. 2002. FRAGSTATS: Spatial Pattern Analysis Program for Categorical Maps Computer software program. University of Massachusetts, Amherst, www.umass.edu/landeco/research/fragstats/fragstats.html.

Ministerio de Fomento 2006. *Atlas Estadístico de las Areas Urbanas de España*. Madrid

OECD. 2012. *Redefining "Urban": A New Way to Measure Metropolitan Areas*. Paris: OECD Publishing.

Parr, J. 2007. Spatial definitions of the City: Four Perspectives. *Urban Studies*, 44, nº 2, 381-392.

Pavia, J.M. and Cantarino, I. 2016. ………………………….

Pham, H. M., and Yamaguchi, Y. 2011. Urban growth and change analysis using remote sensing and spatial metrics from 1975 to 2003 for Hanoi, Vietnam. *International Journal of Remote Sensing*, 32(7), 1901-1915.

Schwarz, N. 2010. “Urban form revisited - Selecting indicators for characterising European cities”. *Landscape and Urban Planning*, volume 96, issue 1

Simon, A., Fons, J. and Milego, R. 2010. *Urban Morphological Zones. Version F2v0. Definition and procedural steps*. Final Report (28.12.2010). European Topic Centre, Land Use and Spatial Information (ETC LUSI), Copenhagen: European Environment Agency (EEA), 2010. Available on: http://www.eea.europa.eu/data-and-maps/data/urban-morphological-zones-2006-umz2006-f3v0/.

Tsai, Y. 2005. “Quantifying Urban Form: Compactness versus ‘Sprawl’”. *Urban Studies* 42:141.

Valcárcel, N. 2011. *Comparison and parameterization between SIOSE and CLC*. 6th EAGLE meeting and Joint Meeting EAGLE-Geoland2, June 2011, Málaga, Spain [Avalaible in http://sia.eionet.europa.eu/EAGLE/EAGLE_6thMeeting_g2_Malaga/04a_CLCAttribution_NV_Homework.pdf]

Vaz, E. (2016). “The future of landscapes and habitats: The regional science contribution to the understanding of geographical space”. *Habitat International*, 51, 70-78.

Vidal, C., Gallego, J. and Kayadjanian, M. 2001. *Geographical use of statistical data. Methodological overview*. In: Towards agri-environmental indicators. Integrating statistical and administrative data with land cover information. Copenhagen: European Environment Agency, Topic Report No. 6, chapter 1: 11-24.

Weeks, J. 2010. *Defining urban areas*. In: Remote Sensing of Urban and Suburban Areas. Springer Netherlands.

Wolman, H., Galster, G., Hanson, R., Ratcliffe, M., Furdell, K. and Sarzynski, A. 2005. "The Fundamental Challenge in Measuring Sprawl: Which Land Should Be Considered?". *The Professional Geographer*, Vol. 57, 94-105.

LIST OF TABLES

Table 1. Main characteristics of urban agglomeration definition models in Europe. MF: Ministry of Public Works and Transport; JRC: Joint Research Centre (grid for population data).

Name	**Producer**	**Date**	**Criteria**	**Source**	**Units Number**	**Coverage**
MUA	IGEAT	2007	Population density	Administra-tive unit	1988 > 20000 inh.	29 European countries
LAU	Spanish MF	2000-2012	Population total	Administra-tive unit	82 > 50000 inh.	Spain
UMZ	EEA	2001	Urban tissue and function (raster)	CLC Grid JRC	4300 > 10000 inh.	29 European countries
SHM UMZ	Own production	2006	Urban tissue and function (vector)	SIOSE/SHM Own grid	415 > 10000 inh.	Spain

Table 2. List of indices used

Indices	Formula	Notes
Gross Density *GD*	$GD = P/S_t$	Ratio population (P) to the entire UMZ area (S_t)
Net Density *ND*	$ND = P/S_{bu}$	Ratio population (P) to residential built-up area (S_{bu})
Ratio Open Space *ROS*	$ROS = \frac{S_{os}}{S_t}$	Ratio green area (or open spaces S_{os}) to the entire UMZ area (St).
Degree of Municipality Division *DMD*	$DMD = 1 - \sum \left(\frac{S_{mi}}{S_t}\right)^2$	Ratio municipality surface in UMZ (S_{mi}) to total UMZ Area (S_t)
Compactness Index *CI*	$CI = 4\,\pi \frac{S_t}{{P_e}^2}$	Ratio surface UMZ (S_t) to perimeter (P_e)
Gini Coefficient *GC*	$GC = 1 - \sum_{k=1}^{N} (Y_k - Y_{k-1})(X_k + X_{k-1})$	N is the number of populated cells in UMZ areas; X_k is the accumulated proportion of the population; and Y_k is the accumulated proportion of UMZ area
Moran's Index *MI*	$MI =$ $= \frac{N}{\sum_i \sum_j w_{ij}} \frac{\sum_i \sum_j w_{ij} (X_i - \overline{X})(X_j - \overline{X})}{\sum_i (X_i - \overline{X})^2}$	N is the number of cells indexed by i and j; X is the population in cells; $\overline{X}$ the mean of X; and w_{ij} is an element of a matrix of spatial distance weights

Table 3. Correlation matrix. UMZs with more than 50,000 inhabitants (101 sample). p-values in red and italics.

	GD	ND	ROS	CI	DMD	GC	MI
GD		0.745	-0.399	0.4589	-0.478	-0.496	-0.134
		*0.000**	*0.000**	*0.000**	*0.000**	*0.000**	*0.180*
ND	0.745		-0.246	0.433	-0.354	-0.158	-0.054
	*0.000**		*0.013**	*0.000**	*0.000**	*0.115*	*0.589*
ROS	-0.399	-0.2457		-0.408	0.440	0.363	0.053
	*0.000**	*0.013**		*0.000**	*0.000**	*0.000**	*0.600*
CI	0.459	0.433	-0.408		-0.575	-0.452	-0.336
	*0.000**	*0.000**	*0.000**		*0.000**	*0.000**	*0.001**
DMD	-0.478	-0.354	0.440	-0.575		0.305	0.029
	*0.000**	*0.000**	*0.000**	*0.000**		*0.002**	*0.775*
GC	-0.496	-0.158	0.363	-0.452	0.305		0.507
	*0.000**	*0.115*	*0.000**	*0.000**	*0.002**		*0.000**
MI	-0.134	-0.054	0.053	-0.336	0.029	0.507	
	0.180	*0.589*	*0.600*	*0.001**	*0.775*	*0.000**	

*: indicates a significant correlation coefficient at the level of 0.05

Table 4. Top ten Spanish UMZ by population.

UMZ Name	Population (inhab.)	Area (km^2)	Number munic.	Gross density (GD)	Net density (ND)
Madrid	4 833 124	605.47	30	7982.4	26,094.5
Barcelona	3 802 184	377.29	67	10077.6	22,150.4
Valencia	1 515 755	180.09	49	8416.6	17,088.6
Bilbao	932 789	240.38	51	3880.4	32,839.2
Seville	925 214	115.76	16	7992.4	25,134.3
Zaragoza	648 293	65.24	4	9937.0	25,005.6
Malaga	530 023	50.11	2	10576.8	33,521.4
S. Sebastian	428 222	158.01	30	2710.1	27,326.2
Alicante	426 036	77.58	7	5491.7	13,506.9
Tenerife	399 091	54.77	10	7286.5	14,868.6

Table 5. Cluster centroids.

Cluster nº	***ND***	***ROS***	***CI***	***DMD***	***GC***	***MI***
1	25980	0.210	0.0030	0.882	0.728	0.405
2	18660	0.138	0.0145	0.215	0.692	0.368
3	11400	0.183	0.0060	0.679	0.691	0.216
4	20300	0.099	0.0404	0.177	0.554	0.128

Table 6. Interval value indices (t-Student's 98%). Overlaps are shaded

Cluster nº	***Number UMZs***	***ND***	***ROS***	***CI***	***DMD***	***GC***	***MI***
1	8	33100 18800	0.27 0.15	0.006 0.000	0.98 0.78	0.80 0.66	0.58 0.23
2	28	21800 15600	0.16 0.12	0.020 0.010	0.30 0.13	0.72 0.66	0.41 0.33
3	38	13000 9800	0.21 0.16	0.008 0.004	0.73 0.63	0.72 0.66	0.26 0.18
4	27	26200 24400	0.12 0.08	0.053 0.028	0.29 0.07	0.60 0.51	0.19 0.06

Table 7. Results of the cluster analysis: groups of Spanish provincial capitals.

Group	Characteristics	Provincial Capitals/ Autonomous Cities
1	LARGEST CITIES Large cities of > 300,000 inhabitants. High density with green zones. Very complex, compact and highly mono-centric structure with several urban subcenters.	**Madrid**, Barcelona, Valencia, Bilbao, Donostia, Pamplona
2	COMPACT CITIES Medium-sized cities. Medium density with few green zones. Usually inland location. Compact, non complex and monocentric structure with some urban subcenters.	**Zaragoza**, Seville, Las Palmas (1), Malaga (1), Palma (1), Valladolid, Murcia, Vitoria, Oviedo, Almería (1), Albacete, Logroño, Lleida, Ourense, Cáceres, Lugo, Ciudad Real, Toledo, Segovia
3	EXPANDED CITIES Medium-sized cities. Low density with green zones. Usually coastal situation. Dispersed, fairly complex and monocentric structure with urban subcenters.	**Alicante** (2), Tenerife (2), Granada, Coruña (2), Castellón (2), León, Salamanca, Santander (2), Tarragona (2), Girona, Pontevedra (2).
4	TRADITIONAL CITIES Small cities of < 300,000 inhabitants. Medium-high density with few green zones.. Usually located inland. Non complex and highly centralized compact monocentric structure.	**Córdoba**, Burgos, Huelva (1), Cádiz (1), Badajoz, Jaén, Palencia, Guadalajara, Ceuta (1), Melilla (1), Zamora, Ávila

(1) Coastal cities with severe orographic or geographic restrictions.
(2) Coastal cities with no important restrictions.
The seeds used in the cluster analysis are shown in **bold type**.

LIST OF FIGURES

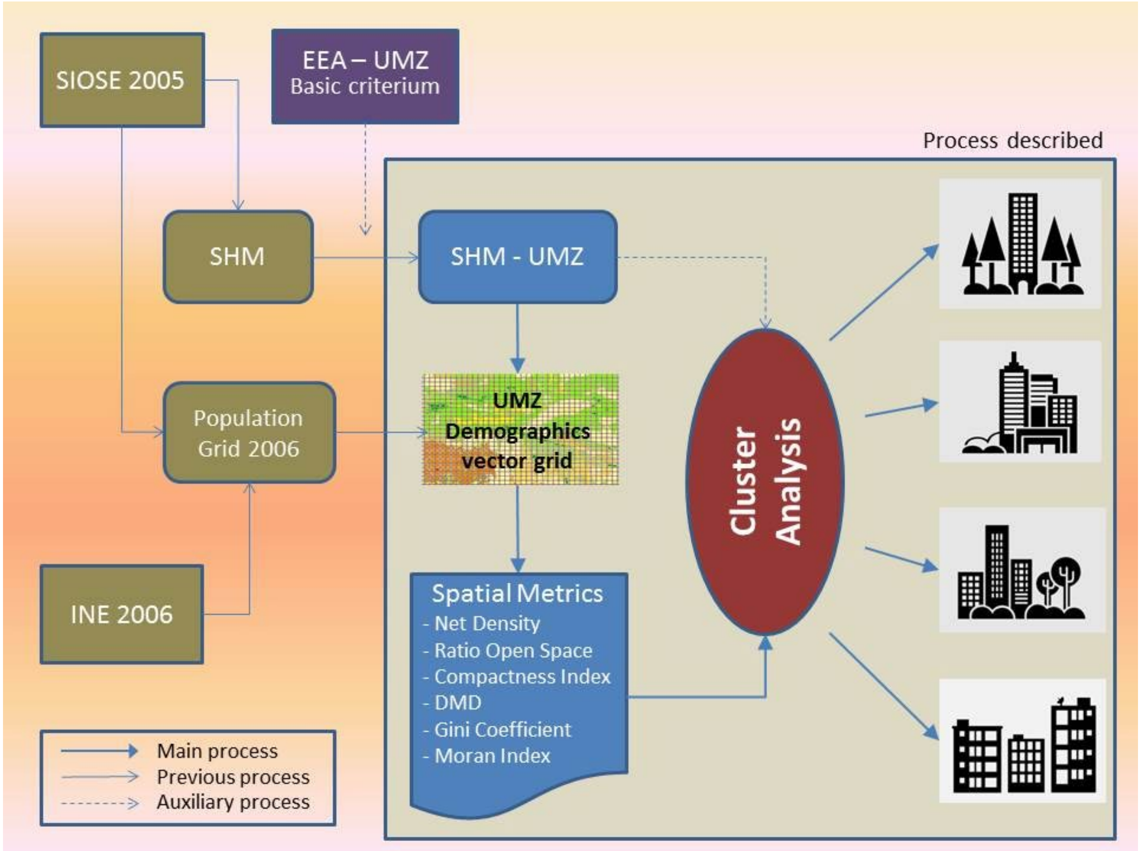

Figure 1. Flow diagram of complete process.

SIOSE: Land Cover and Use Information System of Spain; EEA: European Environment Agency; UMZ: Urban Morphological Zones; SHM: SIOSE Hierarchical Model; INE: Spanish Statistical Institute

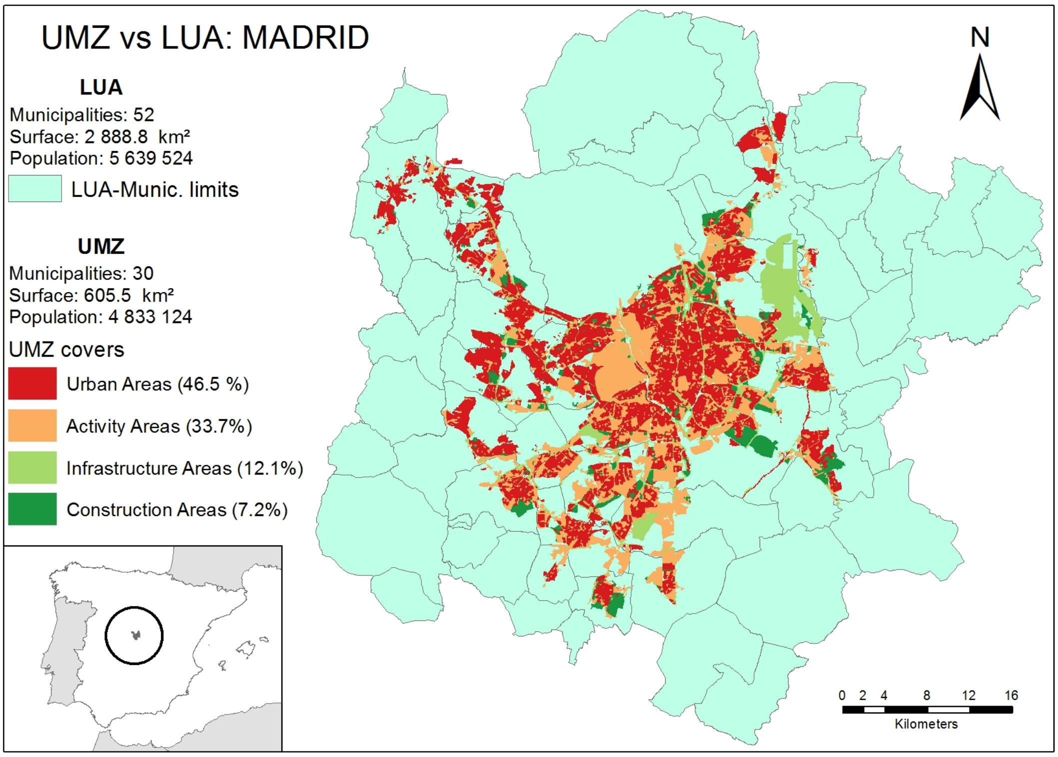

Figure 2. Comparison UMZ versus LUA for Madrid. UMZ covers distribution. Source: SIOSE (IGN 2005), Ministry of Public Works and Transport (2006) and own elaboration.

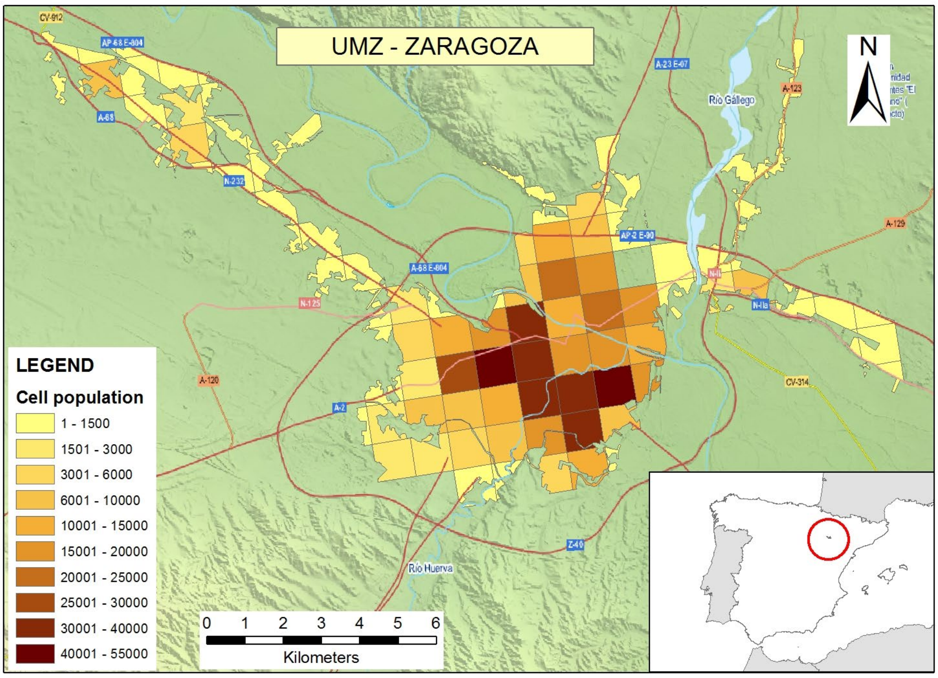

Figure 3. UMZ Zaragoza with grid cells population.

Source: SIOSE (IGN 2005), INE (2006), and own elaboration.

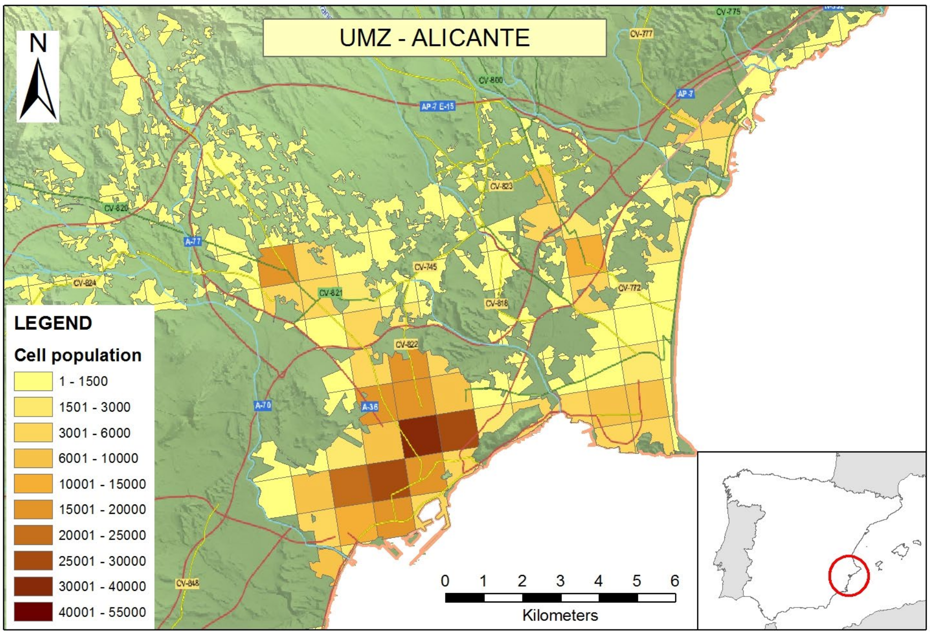

Figure 4. UMZ Alicante with grid cells population.

Source: SIOSE (IGN 2005), INE (2006), and own elaboration.