

Document downloaded from:

<http://hdl.handle.net/10251/94472>

This paper must be cited as:

Gielen, E.; Riutort-Mayol, G.; Palencia Jiménez, JS.; Cantarino-Martí, I. (2017). An urban sprawl index based on multivariate and Bayesian factor analysis with application at the municipality level in Valencia. *Environment and Planning B Planning and Design*. 1-27. doi:10.1177/2399808317690148



The final publication is available at

<http://doi.org/10.1177/2399808317690148>

Copyright SAGE Publications

Additional Information

Environment and Planning B: Urban Analytics and City Science

An Urban Sprawl Index Based on Multivariate and Bayesian Factor Analysis with Application at the Municipality Level in Valencia

Journal:	<i>Environment and Planning B: Urban Analytics and City Science</i>
Manuscript ID	EPB-2015-0045.R3
Manuscript Type:	Manuscript
Keywords:	spatial planning, urban planning, cities, land use, multicriteria analysis

SCHOLARONE™
Manuscripts

Pre-Only

An Urban Sprawl Index Based on Multivariate and Bayesian Factor Analysis with Application at the Municipality Level in Valencia

ABSTRACT

Urban sprawl is now a common and threatening phenomenon in Europe, severely affecting environmental and economic sustainability. An analytical characterization and measurement of urban sprawl is required to gain a better understanding of the phenomenon and to proposed possible solutions. Traditional factor analysis techniques, especially Principal Component Analysis and Factor Analysis, have been commonly used. In this paper, we additionally test Independent Component Analysis with the aim of obtaining a multidimensional characterization of the sprawl phenomenon. We also use Bayesian Factor Analysis to obtain a single (unidimensional) measuring index of sprawl, which also allows us to obtain the uncertainty of the inferred index, in contrast to traditional approaches. All these techniques have been applied to study the phenomenon of urban sprawl at the municipality level in Valencia, Spain using a wide set of variables related to the characteristics and patterns of urban land use.

1. Introduction

Urban sprawl is the spread of human populations from central urban areas to outlying residential areas with lower population densities and more open spaces, where single family homes and low-rises predominate, and to which residents tend to travel by car. This urban model is characterized by low-density urban expansion patterns, producing a patchy, scattered, discontinuous and strung-out urban development, often under a lax planning control. This kind of urban land development is typical of the United States but is more

1
2
3 recent in Europe, where cities have traditionally grown around historical settlements, and
4
5 even more so in the Mediterranean region, which has been characterized by compact cities
6
7 until only a few years ago. Urban sprawl is a global challenge, and now a common
8
9 phenomenon in Europe, where there has recently been great interest in studying it
10
11 (European Environment Agency (EEA), 2016). This is also the case in Spain, where growth
12
13 in urban areas has created a situation that is becoming critical because of its impact on the
14
15 environmental and economic sustainability of cities (EEA, 2006). Therefore, it is necessary
16
17 to improve the framework and tools to control urban sprawl, through urban policies, which
18
19 requires a more extensive knowledge of the phenomenon under study.
20
21

22
23 The sprawl pattern is a complex construct, difficult to conceptualize and measure, with
24
25 many different conditions involved related to the density, uses, form and structure, as well
26
27 as to economic and social aspects, of urban areas. Sprawl is not directly observable, but
28
29 there are several indicators related to it that can be defined and measured, so that their
30
31 correlation structure and the latent dimensions (underlying factors not directly observable or
32
33 measureable) can be extracted by factorial statistical techniques (Johnson and Wichern,
34
35 2007). Latent dimensions can be a useful multi-dimensional measure of sprawl, and a
36
37 better understanding of the problem can be achieved by interpreting their correlation
38
39 structure.
40
41
42

43
44 Traditional factor analysis techniques, such as Principal Component Analysis (PCA) and
45
46 Factor Analysis (FA), are commonly used for the purpose of extracting the correlation
47
48 structure and the latent dimension from a set of indicators. FA is well suited for a
49
50 multidimensional characterization, since it identifies the underlying latent factors in the input
51
52 variables by minimizing the linear correlation among latent factors (Frenkel and Ashkenazi,
53
54 2008). The first component of PCA, which results from the first eigenvector and eigenvalue
55
56 of the covariance matrix of the input set of variables, represents the common factor that
57
58
59
60

1
2
3 maximizes the variability; therefore, it is commonly used for unidimensional
4 characterizations. Both techniques are based on linear model structures and assume
5 normality of the input variables. Many environmental variables may not be normally
6 distributed, which can be handled through transformation, so the results must be interpreted
7 in terms of the transformed variables. In contrast, Independent Component Analysis (ICA)
8 (Stone, 2004) does not require normally distributed variables and results in statistical
9 independence between the latent variables where the joint probability of the latent factors
10 can be factorized (Shlens, 2005). As far as we know, there is no application of this
11 technique in urban studies.
12
13
14
15
16
17
18
19
20
21
22
23

24 FA performed by means of Bayesian hierarchical models is a very powerful method for
25 computing the unidimensional common factor that maximizes variability (Mezzetti and
26 Billari, 2005; Conti et al., 2014) and does not assume any specific probability distribution for
27 the variables. Furthermore, the non-linear covariance structure over the common factor,
28 such as spatial and/or temporal covariance, can be easily specified (Mari-dell'Olmo et al.,
29 2011), and credible intervals (uncertainty) can be inferred from the posterior distributions of
30 the common factor parameter estimated in each municipality. Finally, the treatment of
31 missing values is handled more appropriately, which are estimated during the Markov Chain
32 Monte Carlo (MCMC) simulation method.
33
34
35
36
37
38
39
40
41
42
43
44

45 Thus, the first objective of this study is to extract the latent dimensions that underlie the
46 input set of variables by means of the application of the ICA method, thus obtaining a
47 multidimensional analytical characterization at the municipality for Valencia in Spain. A
48 comparison between the ICA, FA and PCA solutions is also performed. The second
49 objective is to compute a unidimensional measurement index as the common factor to all
50 input variables using Bayesian hierarchical models (Ntzoufras, 2009). Although the
51 underlying multidimensional information is lost, the phenomenon is simplified and facilitates
52
53
54
55
56
57
58
59
60

1
2
3 the measurement and interpretation of the spatial distribution of sprawl throughout the
4 territory, especially for professionals from different disciplines who may find it difficult to deal
5 with multiple dimensions.
6
7
8

9 10 **2. Brief background**

11
12 Even though there is abundant literature, it is difficult to find a definitive consensus.
13 Following Galster et al. (2001), urban sprawl is one name for many conditions. Most
14 definitions are based on the assumption of a traditional compact city understood as the
15 ideal model and the starting point to define the opposite model (Chin, 2002; EEA, 2006).
16 For the EEA (2006), sprawling cities are characterized by a physical pattern of low-density
17 expansion of large urban areas growing into the surrounding agricultural areas, which
18 produces a patchy, scattered and strung-out urban development, with a tendency to
19 discontinuity. According to Fulton et al. (2001), sprawl is associated with a land
20 consumption higher than the population growth, and its commonly used indicator is the per
21 capita consumption of land.
22
23
24
25
26
27
28
29
30
31
32
33
34

35
36 Some studies are conducted from a descriptive point of view, while others take a more
37 morphological perspective, and others focus on the dynamics of changes. Therefore, it is
38 clear that measuring sprawl is a complex task, because of its multiple characteristics or
39 dimensions. Some authors choose socio-economic data for analysis because their
40 definition of sprawl demands this kind of data support; other authors, using a more
41 morphological definition, prefer spatial measures. Table 1 shows the different characteristics
42 considered to measure sprawl.
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

	Fulton et al., 2001	Galster et al., 2001	Ewing et al., 2002	Cutsinger and Galster, 2006	EEA, 2006	Frenkel and Ashkenazi, 2008	Torrens, 2008	Patacchini and Zenou, 2009	Arribas-Bel et al., 2011	Colannino et al., 2011	Kew et al., 2013	Jaeger and Lee, 2014	Paulsen, 2014	Zeng et al., 2014
Density	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Changes-Dynamics	X				X		X	X			X	X		X
Land consumption-Footprint	X	X			X		X	X		X	X		X	X
Fragmentation				X		X	X		X	X	X	X	X	X
Concentration		X	X	X									X	
Nuclearity		X		X										
Mixed use			X	X		X	X		X					
Centrality		X		X			X		X				X	X
Proximity		X		X										X
Open space									X					
Accessibility-Connectivity			X				X		X					X
Complexity						X				X				X
Social							X							

Table 1. Factors used to measure sprawl in the literature

There are important differences in these characteristics yet all authors consider density as an essential feature to measure sprawl. Sprawl has static as well as dynamic components associated with population growth or expansion of urban land. Another essential characteristic is the land consumption or the footprint caused by sprawl development. Furthermore, factors linked to how dispersed urban land is are present in almost all studies with various forms (such as fragmentation, concentration and nuclearity) and different metrics, most of them derived from landscape ecology, such as the Shannon entropy index (Torrens, 2008; Colannino et al, 2011; Zeng et al., 2014), the degree of landscape division or the Gini index of concentration (Colannino et al., 2011). Many studies also considered a composition dimension and the consequent possible relations between uses: proximity, mixed use, accessibility, open space or connectivity. Finally, some authors considered other characteristics, like complexity, with the use of other landscape metrics, such as Fractal

dimension or Shape Index (Frenkel and Ashkenazi, 2008; Colannino et al., 2011; Zeng et al., 2014), or social factors (Torrens, 2008).

Moreover, the observed variables may differ according to the scale of the study (Torrens, 2008; Zeng et al, 2014). Sprawl is a multi-scale phenomenon, with different characteristics, so measuring dispersion requires a multiple-level approach in accordance with the function of the scale (Zeng et al., 2014). The macro-scale measurements, when analysis units correspond to regional or metropolitan areas, normally use socio-economic census indicators (Galster et al., 2001; Ewing et al., 2002; Cutsinger and Galster, 2006; Arribas-Bel et al., 2011). Measuring sprawl at a micro or meso-scale, corresponding to a local characterization, requires a more accurate analysis of the spatial characteristics of the urban land use pattern, as spatial or morphological metrics (Frenkel and Ashkenazi, 2008; Colannino et al., 2011).

In Table 2, different approaches are summarized according to the type of variables used in the studies, the type of output index produced, and the methodology used.

	Type of variables	Output-index	Methodology
Fulton et al., 2001	- Socio-economic census indicators	Multidimensional	-
Galster et al., 2001	- Socio-economic census indicators - Spatial metrics	Unidimensional	Aggregate standardization
Ewing et al., 2002	- Socio-economic census indicators	Unidimensional	Aggregate standardization Principal Component Analysis
Cutsinger and Galster, 2006	- Socio-economic census indicators	Multidimensional	-
EEA, 2006	- Socio-economic census indicators	Multidimensional	-
Frenkel and Ashkenazi, 2008	- Spatial metrics	Unidimensional	Factor analysis
Torrens, 2008	- Spatial metrics	Multidimensional	-
Patacchini and Zenou, 2009	- Socio-economic census indicators	Multidimensional	Principal Component Analysis
Arribas-Bel et al., 2011	- Socio-economic census indicators	Multidimensional	Artificial neural network
Colannino et al., 2011	- Spatial metrics	Multidimensional	Factor analysis
Kew et al., 2013	- Socio-economic census indicators	Unidimensional	Standardization procedure

Jaeger and Schwick, 2014	- Spatial metrics	Unidimensional	Aggregate weighted standardization
Paulsen, 2014	- Socio-economic census indicators	Multidimensional	-
Zeng et al., 2014	- Socio-economic census indicators - Spatial metrics	Multidimensional	-

Table 2. Methods used to measure urban sprawl

Some researchers have obtained a multidimensional output index without using any statistical procedure, only using the original variables simultaneously to conceptually categorize the sprawl (Fulton et al., 2001; Cutsinger and Galster, 2006; EEA, 2006; Torrens, 2008; Paulsen, 2014; Zeng et al., 2014). Others have obtained a unidimensional index only by means of an aggregate standardization of various input variables.

PCA, FA and Neuronal Networks are the statistical models commonly used in the literature either for unidimensional or multidimensional characterizations of sprawl. The work of Frenkel and Ashkenazi (2008) obtains a unidimensional index from FA with only one factor, which is equivalent to the first component of PCA. No work has been found in the literature in which ICA or a Bayesian factor approach have been used for characterizing sprawl.

Regarding its impacts, urban sprawl development has left a very large footprint in Spain (EEA, 2006). The Corine Land Cover project and the Spanish Land Use Information System (SIOSE) reaffirmed these trends, even after the end of the housing boom in 2010 (Zornoza, 2013). Also, the European Parliament, by passing the *Auken* resolution, points at the impact of extensive urbanization in Spain.

There are interesting reviews in the literature of the effects of sprawl in Johnson (2001), Ewing et al. (2002), Burchell et al. (2003), Ewing (2008) or Wilson and Chakraborty (2013). Urban sprawl threatens sustainability, transforming natural and rural environments, raising noise pollution over the safety limits, and creating more greenhouse gases, which affect

1
2
3 climate change (EEA, 2006; Wilson and Chakraborty, 2013). There are also public and
4
5 private costs associated with sprawl as outlined in Burchell et al. (2003).
6
7

8
9 Despite the commonly known negative effects of sprawl (e.g. the private cost associated
10
11 with congested roads), urban sprawl also generates some benefits to people individually. In
12
13 fact, these benefits are generally the causes of sprawl: people look for a better quality of
14
15 life, more open space, wish to live in houses with gardens, etc. However, the balance
16
17 seems to be negative on the community (Chin, 2002; Pichler, 2007; Garbiñe, 2007; Miralles
18
19 et al., 2012).
20
21

22
23 From an economic perspective, urban sprawl is an expensive model of urban development
24
25 (EEA, 2006) and is clearly inefficient (Garbiñe, 2007). The sprawled city has economic
26
27 costs with a significant impact on public local finance because of the public services
28
29 assumed by municipalities, as demonstrated in many studies in the United States (RERC,
30
31 1974; Speir, 2002; Burchell et al., 2003; Carruthers and Ulfarsson, 2003). In Spain, some
32
33 authors have also addressed this issue (Garbiñe, 2007; Solé-Ollé and Hortas-Rico, 2008;
34
35 Hortas-Rico, 2010; Benito et al., 2010). In any case, all of them provide very different
36
37 results, depending on which expenditure is considered and where it has been studied.
38
39 Therefore, the discussion on the compact or dispersed city requires more studies on
40
41 measuring and evaluating the cost of sprawl in order to be able to give the correct
42
43 responses and make the correct decisions on which urban model to choose. It is important,
44
45 therefore, to continue increasing the knowledge about how to measure urban sprawl,
46
47 especially in the Spanish context, where important economic effects may arise due to the
48
49 urban development occurred in recent decades.
50
51
52

53 54 **3. Scope of the study and information sources** 55 56 57 58 59 60

1
2
3 The study area is the autonomous community of Valencia, located on the Mediterranean
4 coast in southeast Spain. More than 5 million people currently live there, in a total of 542
5 municipalities and an area of 23,255 km². The territory is administratively divided into three
6 provinces: Castellon, Valencia and Alicante. From a geomorphological point of view, the
7 Valencian Community can also be divided into three different geographical areas: coastal,
8 intermediate and rural areas (Figure 1).
9
10
11
12
13
14
15

16
17 80% of the population is concentrated in the coastal area, in an area equivalent to 25% of
18 the total. The territory is flat with less than 100 meters above sea level, and a densely
19 populated area (696.48 people per hectare), where 85% of the GDP of the region is
20 concentrated. It is characterized by coastal urban areas and metropolitan areas, with a
21 predominance of irrigated agriculture and a greater economic weight of services, especially
22 tourism. The intermediate area represents 17% of the population of Valencia on 37% of its
23 surface with a lower density (101.80 people per hectare) and only 13% of the region's GDP.
24 It corresponds to an intermediate zone between the coast and the mountains, with medium-
25 sized cities and predominance of industry and rain-fed agriculture. Thirdly, the rural area
26 (38% of the territory) is a mountain area, with small towns in which only 2% of the
27 population lives. This area only represents 2% of the region's GDP, with forested areas and
28 an economy based on agriculture and services.
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43

44
45 In regard to urban occupation, there are large differences between areas: in the coastal
46 area, 70% of the total urban development is concentrated in Valencia with 37% and 3.3% in
47 the intermediate and rural areas, respectively. The type of settlement also varies: 40% of
48 the urban development in the coastal area is contiguous, while 60% is not. In the
49 intermediate area, more than 70% of the urban development is not contiguous. Lastly, 67%
50 of the urban development in the rural area is characterized by a contiguous urban
51 development.
52
53
54
55
56
57
58
59
60

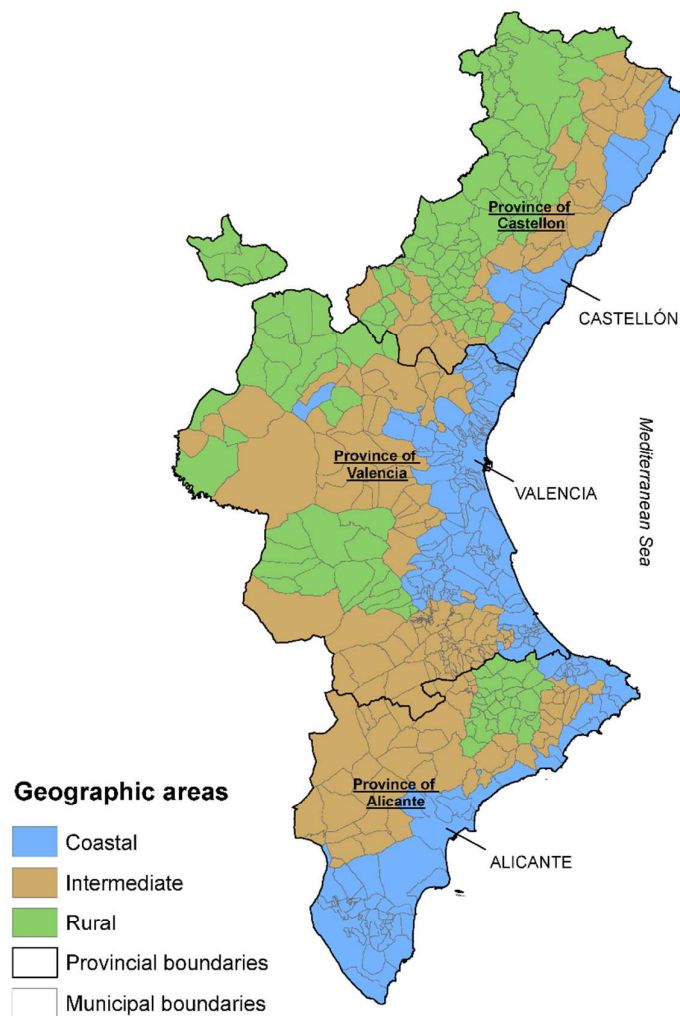


Figure 1. Geographical areas of the Valencian Community

Mediterranean cities were examples of compact cities for a long time, “historically characterized by the archetypal image of density, urban complexity and social diversity” (Munoz, 2003). However, recently, this paradigm has changed, due to urban development, especially in coastal areas. This is also the case for Valencia.

In Spain, this phenomenon is the result of three expansive economic cycles associated with corresponding housing bubbles (Miralles, 2014): a first cycle, in 1959-1972, which created the first tourist developments on the coast; a second cycle, from 1985 to 1990, which generated a major urban development on the Mediterranean coast; and a third cycle, from

1
2
3 1997 to 2006, which resulted in new investments in the real estate construction sector. In a
4
5 context marked by greater access to private transport, helped by falling transport costs
6
7 (Muniz, 2006), at least during the first two cycles, as well as a change in the lifestyle of
8
9 people, asking for more space (Jaeger, 2014), and poor control mechanisms in the
10
11 legislation of land use planning, cities became more dispersed. Fortunately, the passing of
12
13 the Valencian Community Territorial Strategy in 2011 introduced some rationality in urban
14
15 growth, for a sustainable and compact city.
16
17

18
19 The main information sources used were: the Population and Housing Census, the
20
21 boundary limits, both from the National Statistics Institute, and the SIOSE (the Spanish
22
23 Land Use and Cover Information System), a high-resolution GIS database on land cover
24
25 and use, at a 25.000 scale, from the Valencian Cartographic Institute. We focused on
26
27 residential buildings included in a composite coverage known as Mixed Urban. Each cover
28
29 is defined according to the sum of different types of artificial simple uses (houses, streets,
30
31 gardens, pools, etc.). There are three Mixed Urban covers: the old area, the expansion area
32
33 (new parts), and the discontinuous area or suburbs. In addition, these covers are
34
35 characterized by attributes about the residential building type: isolated block of apartments
36
37 or non-isolated block of apartments, single-family detached home, houses in rows or semi-
38
39 detached houses.
40
41
42
43

44 45 **4. Methodology**

46
47 The first step was to obtain as many characteristics of urban sprawl as possible. Using the
48
49 datasets outlined above, a set of urban sprawl indicators were calculated as outlined in the
50
51 next section.
52
53

54 55 **4.1 Indicators**

1
2
3 Based on our morphological definition of sprawl, we define a set of 12 representative
4 indicators concerning density, land use, form and structure of the urban patches, which are
5 based on an extensive literature review and the data available for Valencia (Table 3).
6
7
8

9
10 Density is the most popular measure of sprawl. The *Net housing density* indicator (NetDen)
11 is the area-weighted average of housing density of all the urban patches in a municipality.
12 However, sprawl is characterized by low housing density in single-family homes and
13 discontinuous urban areas, so we also define an *Urban form* indicator (CCont) that
14 assesses the proportion of continuous urban typology from the total urban surface.
15 Likewise, the *Dwelling type* indicator (CEdif) evaluates the proportion of the multiple family
16 homes from the total amount of homes. The *Open space density* indicator (FSpace) refers
17 to green areas and street density in a municipality. Sprawled settlements are usually
18 characterized by private gardens, swimming pools, bigger parks, parking areas and streets.
19 The *Shape index* (Shape) and *Fractal dimension* (Fractal) are indicators of shape
20 complexity based on the relationships between the perimeter and the area of the urban
21 patch (Colannino et al., 2011). Sprawl models tend to have irregular shapes in urban areas.
22 The *Fragmentation degree* (Frag) and *Index of concentration* (Gini) are indicators related to
23 fragmentation (Marmolejo and Stallbohm, 2008) and concentration of urban land
24 (Colannino et al., 2011). The *Distance to center* (Dist) indicator refers to the area-weighted
25 average of distances from the urban patches to the city center. The *Standard deviation of*
26 *the distance to center* (sdDist) measures the variability of the Dist indicator. In addition, we
27 have computed the *coefficient of variation* of the Dist indicator (cvDist), which represents
28 the variability independently of the magnitude of the mean. Finally, to account for the
29 importance of the sprawl model in a municipality, we have computed the *Discontinuous*
30 *area* indicator (Disc), which is the proportion of discontinuous urban surface in the
31 municipality.
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Indicator	Formula	Indices
Net built-up density (unit: m^2/m^2)	$NetDen = \frac{\sum_{i=1}^n \left(\frac{\sum_{a=1}^p (U_i^a * K_a)}{S_i} \right)}{n}$	U_i^a built-up area of type a in urban patch i and K_a number of floors of building type a (Cantarino et al., 2014) S_i area of urban patch i n number of urban patches D discontinuous urban area T area of municipality C continuous urban area M multi-family housing area F single-family dwelling area Z green area R roads area H the number of houses in municipality p_i perimeter of urban patch Q_i is the cumulative percentage of area of patch i over total area U_i built-up area in urban patch i d_i distance from i to center city i
Discontinuous surface (unit: %)	$Disc = \frac{D}{T}$	
Urban form (unit: %)	$CCont = \frac{C}{C + D}$	
Dwelling type (unit: %)	$CEdif = \frac{M}{M + F}$	
Open space density (unit: $m^2/dwelling$)	$FSpace = \frac{Z + R}{H}$	
Shape Index (unit: none)	$Shape = \frac{\sum_{i=1}^n \left(\frac{p_i}{\sqrt{\pi S_i}} * S_i \right)}{\sum_{i=1}^n (S_i)}$	
Fractal dimension (unit: none)	$Fractal = \frac{\sum_{i=1}^n \left[\frac{2 * \ln(0,25 * p_i)}{\ln(S_i)} * S_i \right]}{\sum_{i=1}^n (S_i)}$	
Fragmentation degree (unit: none)	$Frag = -1 * \sum_{i=1}^n \left\{ \frac{S_i}{\sum_{i=1}^n S_i} * \left[\ln \left(\frac{S_i}{\sum_{i=1}^n S_i} \right) \right] \right\}$	
Index of concentration (unit: none)	$Gini = \frac{\sum_{i=0}^n (1 - Q_i)}{n}$	
Distance to center (unit: meters)	$Dist = \frac{\sum_{i=0}^n \left(d_i * \frac{U_i}{\sum_{i=1}^n (U_i)} \right)}{n}$	
Standard deviation of the distance to center (unit: meters)	$sdDist = \sqrt{\frac{\sum_{i=0}^n \left(d_i - \frac{\sum_{i=0}^n \left(d_i * \frac{U_i}{\sum_{i=1}^n (U_i)} \right) \right)^2}{n - 1}}$	
Coefficient of variation of the distance to center (unit: meters)	$cvDist = \frac{sdDist}{\frac{\sum_{i=0}^n \left(d_i * \frac{U_i}{\sum_{i=1}^n (U_i)} \right)}{n}}$	

Table 3: Selected sprawl indicators

4.2 Multivariate Factor Techniques

Three factor multivariate techniques were then used to explore the correlation structure of the set of indicators and extract their underlying dimensions: ICA, which handles non-

1
2
3 normally distributed variables (Hyvärinen and Oja, 2000), FA (Bartholomew, 2007) and PCA
4 (Abdi and Williams, 2010), where the latter two approaches require normally-distributed
5 variables. As some of the indicators showed non-normal distributions, they were
6 transformed prior to the application of FA and PCA; more details are provided in the
7 Supplementary Material.
8
9

10
11
12
13
14
15 ICA aims to extract the latent factors that are statistically independent of each other. This is
16 achieved by a *mutual-based information method*, which has a form of nonlinear optimization
17 minimizing the information mutually shared among the extracted latent factors (Hyvärinen
18 and Oja, 2000). The ICA method models the set of original indicators as a linear
19 combination of new statistically independent components, as
20
21
22
23
24

$$X_{(I*J)} = S_{(I*K)} \cdot A_{(K*J)} \quad (1)$$

25
26
27
28
29
30 Where X is a matrix $(I * J)$ containing each one of the original indicators j measured for
31 each one of the municipalities i , S is a matrix $(I * K)$ containing each one of the
32 independent components k , and A is a matrix $(K * J)$ containing the contributions or
33 loadings of the linear combination. The estimation of the matrix S and A by the mutual
34 information-based method is carried out so that S is composed of independent components.
35
36 In this work, the package *PearsonICA* of the statistical software R (R-Development Core
37 Team, 2010) has been used to perform this method.
38
39
40
41
42
43
44
45

46
47 FA resolves the linear combination of the new components, following the same expression
48 in (1), by formulating a linear model that relates the original set of indicators X and the
49 latent factors S , taking into account the correlation structure of the variables in the
50 formulation of the model and the method of maximum likelihood to estimate the model
51 parameters (Graffelman, 2013). The R package *Robustfa* has been used to apply this
52 method.
53
54
55
56
57
58
59
60

1
2
3 PCA performs *Singular Value Decomposition* of the correlation matrix of the set of
4
5 variables. Using the same expression (1), this procedure leads to new latent factors
6
7 determined by the eigenvectors and eigenvalues of the correlation matrix (Shlens, 2005).
8
9 The R package *FactoMineR* has been used to apply this method. The first latent factor (first
10
11 eigenvector and eigenvalue) will therefore be a common factor among all the variables that
12
13 explain the maximum variability. The others (the other eigenvectors and eigenvalues) will
14
15 then be consecutively orthogonal to each other. Consequently, as the first latent factor is
16
17 the maximum-variance consensus among all the variables, this can include potentially
18
19 different dimensions.
20
21
22
23

24 **4.3 Bayesian Factor Analysis**

25
26 This section aims to define a single index for urban sprawl, which is the common factor
27
28 maximizing the variability contained in the set of variables. Although the multidimensional
29
30 information is lost, the phenomenon is simplified and facilitates the interpretation of the
31
32 spatial distribution of sprawl throughout the territory.
33
34
35

36
37 In this study Gaussian models were used for all the variables, which were previously
38
39 transformed into normal distributions (see Supplementary Materials). The Bayesian factor
40
41 model for inferring only one single latent factor is based on finding the linear combination
42
43 between scores and loadings that maximize the likelihood of the model (Mari-Dell'Olmo et
44
45 al., 2011). If the likelihood is considered as Gaussian, then the procedure is equivalent to
46
47 traditional FA. Furthermore, the Bayesian approach allows us to estimate the uncertainty of
48
49 the inferred latent factors and to incorporate the process of imputing the missing-values of
50
51 the set of variables at the same time as the resolution of the Bayesian factor model and the
52
53 estimation of its parameters.
54
55
56
57
58
59
60

The Bayesian factor model is performed using hierarchical models. Let Y_{ij} be the value of the indicator j , $j = 1, \dots, 12$, in municipality i , $i = 1, \dots, 542$.

$$Y_{ij} \sim \text{Normal}(\mu_{ij}, \sigma_j^2)$$

$$\mu_{ij} = \alpha_j + S_i \cdot L_j$$

$$\alpha_j \sim \text{Normal}(0, 1000)$$

$$S_i \sim \text{Normal}(0, \sigma_S^2)$$

$$\sigma_S^2 = 1$$

$$L_j \sim \text{Normal}(0, 1000)$$

$$\sigma_j \sim \text{IGamma}(0.001, 0.001)$$

α_j is the general mean for each indicator j , S_i is the common factor of all the indicators in municipality i , L_j is the loading which quantifies the contribution of the indicator j to the common factor, and finally σ_j^2 is the residual variance of the model for each indicator j . As Marí-Dell'Olmo (2011) suggests, to avoid identifiability problems between the scales of S_i and L_j , the variance of S is fixed to one ($\sigma_S^2 = 1$). Also, the product $S_i \cdot L_j$ needs to be restricted (1) to a unique solution, since different and equivalent solutions may be obtained by simply changing their sign (Marí-Dell'Olmo et al., 2011),

$$L_j = \begin{cases} L_j & \text{if } L_1 \geq 0 \\ -L_j & \text{if } L_1 < 0 \end{cases} \quad (1)$$

$$S_i = \begin{cases} S_i & \text{if } L_1 \geq 0 \\ -S_i & \text{if } L_1 < 0 \end{cases}$$

For the general means α_j and loadings L_j , prior non-informative normal distributions (normal distributions with a very large variance) are defined. For the variance parameters

σ_j^2 , prior non-informative Inverse Gamma distributions (with very low Gamma parameters) are defined (Kass and Wasserman, 1995).

Once the complete Bayesian hierarchical model has been formulated, a learning process based on MCMC can be applied to estimate the posterior distributions of the parameters of the model. We have used the WinBUGS software (Lunn et al., 2000) to run the MCMC on the proposed model (Ntzoufras, 2009). From the posterior distributions of the estimated parameters, their mean and variance can be inferred. The variance represents the uncertainty of the estimated parameter.

5. Results

5.1 Results of the Multivariate Factor Analysis

Table 4 shows the loadings of each extracted dimension to each of the original variables for the ICA solution. With the analysis of the loadings matrix, the variables can be discriminated, with higher weights over each extracted dimension, so that the greater the contribution, in absolute values, the greater the presence in the dimension. In this way, it is possible to know which variables make up the different dimensions or extracted latent factors.

	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
<i>Disc</i>	0.15	0.16	-0.08	0.13	0.94
<i>CCont</i>	-0.77	-0.24	-0.12	-0.16	-0.36
<i>CEdif</i>	-0.65	-0.22	0.02	-0.03	-0.49
<i>NetDen</i>	-0.65	-0.21	-0.02	-0.11	-0.05
<i>FSpace</i>	0.90	-0.06	0.07	0.14	0.20
<i>Frag</i>	0.38	0.69	-0.17	0.19	-0.01
<i>Shape</i>	0.12	0.30	-0.03	0.94	0.19
<i>Fractal</i>	0.14	0.14	0.01	0.91	0.06
<i>Dist</i>	-0.07	-0.25	0.95	-0.07	-0.11
<i>sdDist</i>	0.03	0.05	1.00	-0.10	0.09
<i>cvDist</i>	0.24	0.69	0.24	0.03	0.48
<i>Gini</i>	0.25	0.77	-0.11	0.06	0.14

Table 4: Results of ICA

With the aim of checking and contrasting the dimensions revealed by ICA, FA has also been tested on the data, expecting both solutions to be congruent. Table 5 shows the resulting loadings for the FA. Both generate similar resulting latent factors and are only different from Disc. In the case of FA, Disc contributes to Dimensions 1, 2 and 5; however, in ICA, this is only present with a high weight in Dimension 5.

The ICA method seems to produce the best results and the clearest interpretation of the latent factors. It should be noted that the individual weights (loadings) of the dominant variables in each of the latent factors are more balanced in ICA than in FA solution. Although in ICA solution the explained variances of each of the latent factors are not shown, they are expected to be quite similar to the FA.

	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
<i>Disc</i>	0.49	0.57	-0.16	0.13	0.41
<i>CCont</i>	-0.77	-0.28	-0.12	-0.19	-0.41
<i>CEdif</i>	-0.65	-0.34	-0.02	-0.07	-0.07
<i>NetDen</i>	-0.95	-0.18	0.03	-0.13	0.22
<i>FSpace</i>	0.71	0.19	0.04	0.15	0.12
<i>Frag</i>	0.40	0.52	-0.29	0.22	0.07
<i>Shape</i>	0.16	0.24	-0.12	0.88	0.05
<i>Fractal</i>	0.16	0.06	-0.05	0.98	0.02
<i>Dist</i>	-0.03	-0.35	0.92	-0.13	0.01
<i>sdDist</i>	0.10	0.22	0.97	-0.05	-0.02
<i>cvDist</i>	0.23	0.87	0.23	0.13	0.03
<i>Gini</i>	0.34	0.68	-0.19	0.14	0.03

	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
SS-loadings	3.36	2.40	1.91	1.96	0.34
Variance %	28.0	20.0	15.9	16.3	2.8
Cumulative Variance %	28.0	44.0	64.0	80.3	83.1

Table 5: Results of FA

In summary, five clear dimensions were extracted. The weakest dimension is probably Dimension 5, with only around 6% of the explained variance. Dimension 1 is composed of

CCont, CEdif, NetDen and FSpace, all of which are variables related to intensity of use, i.e. density and concentration of housing and population. Dimension 2 is composed of Frag, cvDist, and Gini, all related to spatial fragmentation, dispersion and concentration of urban patches. Dimension 3 is composed of Dist and sdDist, related to the mean and variance of distance from the urban patch to the center of town. Dimension 4 is composed of the Shape and Fractal shape complexity indicators. Dimension 5 is related to the Disc variable, which measures the relative importance of the discontinuous land.

Finally, we compare the results with PCA (Table 6), which shows that Dimension 1 is a consensus of all the variables. The fact that Dimensions 2 and 3 are clearly dominated by Dist and sdDist and by Shape and Fractal, respectively, means that these two dimensions have different information from the common factor of all the variables, especially Dimension 2, which is composed of Dist and sdDist.

	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
<i>Disc</i>	0.84	-0.11	-0.24	0.07	-0.17
<i>CCont</i>	-0.87	-0.22	0.06	0.20	0.12
<i>CEdif</i>	-0.76	-0.15	0.20	0.27	-0.45
<i>NetDen</i>	-0.82	-0.15	0.08	0.37	-0.18
<i>FSpace</i>	0.74	0.20	-0.06	-0.40	-0.41
<i>Frag</i>	0.80	-0.25	-0.13	0.18	0.06
<i>Shape</i>	0.64	-0.29	0.68	0.08	0.02
<i>Fractal</i>	0.54	-0.22	0.79	-0.06	0.03
<i>Dist</i>	-0.24	0.91	0.25	-0.02	0.04
<i>sdDist</i>	0.25	0.90	0.15	0.29	0.00
<i>cvDist</i>	0.73	0.22	-0.09	0.52	-0.04
<i>Gini</i>	0.79	-0.18	-0.20	0.38	0.04

	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
Eigenvalue	5.84	2.05	1.34	0.95	0.46
Variance %	48.69	17.08	11.21	7.95	3.83
Cumulative Variance %	48.70	65.77	76.98	84.93	88.76

Table 6: Results of PCA

5.2 Results of the Bayesian Factor Analysis

Table 7 shows the loading (as the mean of the estimated parameter L_j) of each indicator j on the estimated common factor S_i , as well as its uncertainty (as the standard deviation of the estimated parameter L_j). The variables with the smallest contribution (weight) are *sdDist* and *Dist*, which form the distance to center dimension (Dimension 3), followed by *Shape* and *Fractal*, which form the complexity dimension (Dimension 4). All the other variables associated with Dimensions 1 (intensity of use), 2 (fragmentation and dispersion) and 5 (magnitude) have high weights in the common factor; they are the ones which best characterize the sprawl phenomenon. Figure 2 shows, for each municipality i , spatial representation of sprawl index (common factor) and Figure 3 its uncertainty. His histograms are shown in Figure 4 and Figure 5, respectively.

The uncertainty is around 10% of the index, which seems to be admissible for the phenomenon under study. Furthermore, its variation is very small. It varies from 0.25 to 0.33, which shows the good consistency of the model. Only two groups of municipalities with anomalous values can be seen, showing a lack of fit in the model. The values higher than 0.33 correspond to municipalities in inland and rural areas, whose behavior in their urban models is quite different to most of the other municipalities, with a much smaller or non-existent sprawl model. In most cases, they are small municipalities with a single urban patch. The municipalities with an uncertainty higher than 0.45 have additional missing values in their indicators, which is the reason for the high uncertainty.

Indicator	Loading	Error
<i>Disc</i>	0.90	± 0.042
<i>CCont</i>	-0.88	± 0.043
<i>CEdif</i>	-0.74	± 0.046
<i>NetDen</i>	-0.78	± 0.037
<i>FSpace</i>	0.70	± 0.039
<i>Frag</i>	0.79	± 0.040
<i>Shape</i>	0.57	± 0.042

<i>Fractal</i>	0.46	± 0.042
<i>Dist</i>	-0.23	± 0.044
<i>sdDist</i>	0.22	± 0.049
<i>cvDist</i>	0.70	± 0.039
<i>Gini</i>	0.78	± 0.042

Table 7: Loadings of latent factor (common factor) from Bayesian Factor Analysis

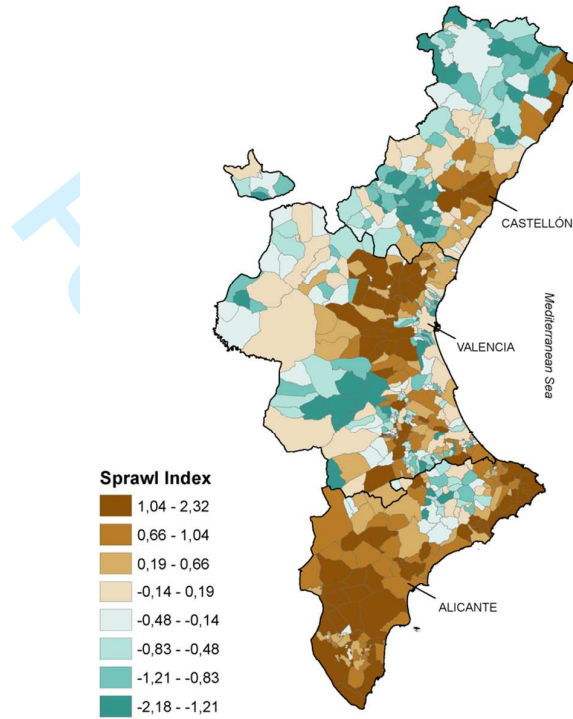


Figure 2. Map of sprawl index

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

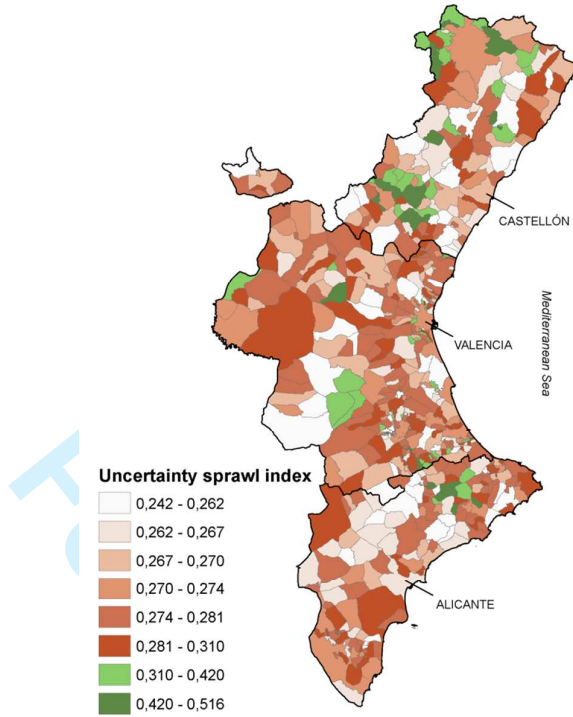


Figure 3. Map of uncertainty of sprawl index

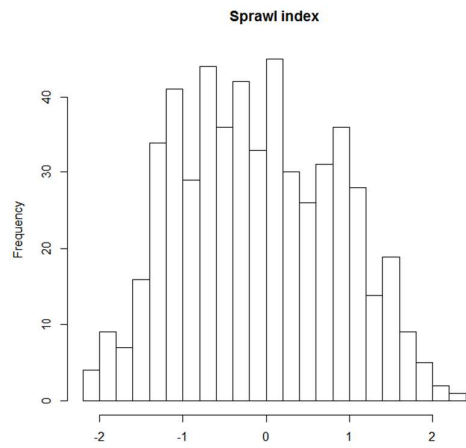


Figure 4. Histograms of sprawl index

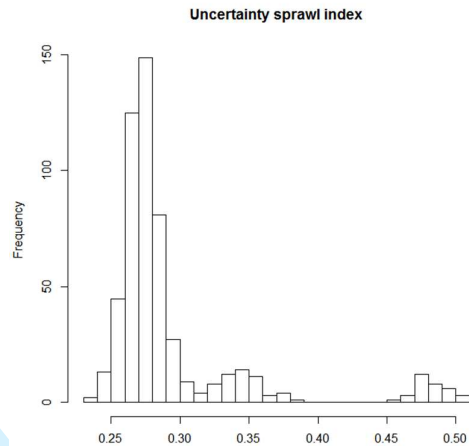


Figure 5. Histograms of the uncertainty of the sprawl index

5.3 Understanding Urban Sprawl in Valencia

Using multivariate factor techniques, we extracted five different dimensions from the set of original indicators to achieve a better understanding of urban sprawl in Valencia.

First underlying dimension (Figure 6) is composed of the following variables: net housing density (NetDen), proportion of continuous urban typology (CCont), proportion of multiple family homes (CEdif) and density of net open space (FSpace), associated with the intensity of residential use and load capacity of the territory, which are clearly dependent on urban typology and density, both of which are planning policy parameters. The development of this dimension appears especially in western municipalities in the metropolitan area of Valencia and Castellon, favored by the expansion of the cities and the development of bedroom communities in middle-class towns. It also appears intensively throughout the province of Alicante and in the north coast of the province of Castellon due to a low density urban model promoted by tourism. However, this dimension is not developed in most of the coastal municipalities due to the high concentration and densification of buildings and population, caused by the appeal of the beach. In the more rural and mountainous areas, it is clear that this dimension does not have a significant effect.

1
2
3 Second dimension is associated with the dispersion and fragmentation of urban land. The
4 variables making up this dimension are Frag, Gini and cvDist, i.e. measures of division,
5 concentration and dispersion of urban land. The presence of this second dimension shows
6 how the sprawl phenomenon creates fragmented and dispersed urban patterns. Basically,
7 this dimension adds coastal municipalities to the areas affected by the first dimension
8 (Figure 7), due to the growing demand to occupy coastal areas, producing highly
9 fragmented patterns in these areas. In addition, some large rural or intermediate
10 municipalities seem to have high fragmentation values. In any case, these first two
11 dimensions are linked to the morphological aspects of the urban model.
12
13
14
15
16
17
18
19
20
21
22
23

24 Third dimension reflects distance or proximity to the center of town and is measured by the
25 distance to the center (Dist) plus the variability of these distances (sdDist). Mapping this
26 dimension (Figure 8), a quite different behavior from the first two dimensions may be noted.
27 The combination of both high distances and a variability in these distances appears mainly
28 in the medium and large rural areas, due to the fact that they probably have few and distant
29 urban patches. On the coast, the larger distances between urban patches are located in the
30 northern municipalities of the provinces of Alicante and Castellon.
31
32
33
34
35
36
37
38
39

40 Fourth dimension deals with the Shape and Fractal shape complexity indicators, which
41 measure the complexity of the geometric shape of urban patches. Normally, high
42 complexity corresponds to low density, dispersed and fragmented urban areas, while
43 compact areas tend to have simpler and more regular shapes. Nevertheless, modern
44 residential areas tend to be regularly shaped, whereas older and usually denser areas may
45 present complex shapes. Although this dimension is usually associated with sprawl, it can
46 also be a characteristic of compact areas. This ambiguous behavior is reflected in the
47 results of PCA analysis, since these variables, apart from contributing to the common factor
48 (Dimension 1 in PCA), are also present in dimension 3. Mapping this dimension (Figure 9),
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 this ambiguous behavior may also be noted, showing high complexity values in many
4 coastal municipalities, which are areas with sprawl models, and in many rural and mountain
5 municipalities characterized by irregular older urban areas.
6
7
8

9
10 Finally, fifth dimension is associated with the Disc variable, which is the relative importance
11 of discontinuous land in the municipality. This dimension is related to the magnitude of the
12 phenomenon, which explains why it appears in a separate dimension. However, as can be
13 seen in PCA and the Bayesian common factors, this variable is part of the common factor.
14 As shown in Figure 10, the relative high importance of discontinuous urban occupation
15 appears in the whole coast of the province of Alicante, in the western municipalities of the
16 metropolitan area of the city of Valencia and in the entire metropolitan area of the city of
17 Castellon. Furthermore, high values may appear in some inland and rural municipalities of
18 the province of Castellon, due to the fact that the land use database characterizes urban
19 centers of small rural settlements as discontinuous.
20
21
22
23
24
25
26
27
28
29
30
31

32
33 Dimensions 1 (Intensity of use), 2 (Fragmentation) and 4 (Complexity) seem to be
34 morphological characteristics of sprawl, while Dimensions 3 (Distance to center) and 5
35 (Magnitude) are rather related to the expansion and magnitude of sprawl. Regarding the
36 importance of each dimension, Dimensions 1 (Intensity of use), 2 (Fragmentation) and 5
37 (Magnitude) contribute substantially to the common factor (Dimension 1 in PCA or Bayesian
38 common factor), and therefore seem to be better dimensions to characterize sprawl. The
39 complexity dimension (Dimension 4) contributes less to common factor, as shown in
40 loadings from Bayesian Factor Analysis (Table 7), and seems to have a more ambiguous
41 behavior (Dimension 3 in PCA). However, Dimension 3 (Distance to center) does not seem
42 to characterize sprawl, at least for our study area and with the variables formulated: it
43 appears in a separate PCA dimension and contributes almost nothing to the common factor.
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

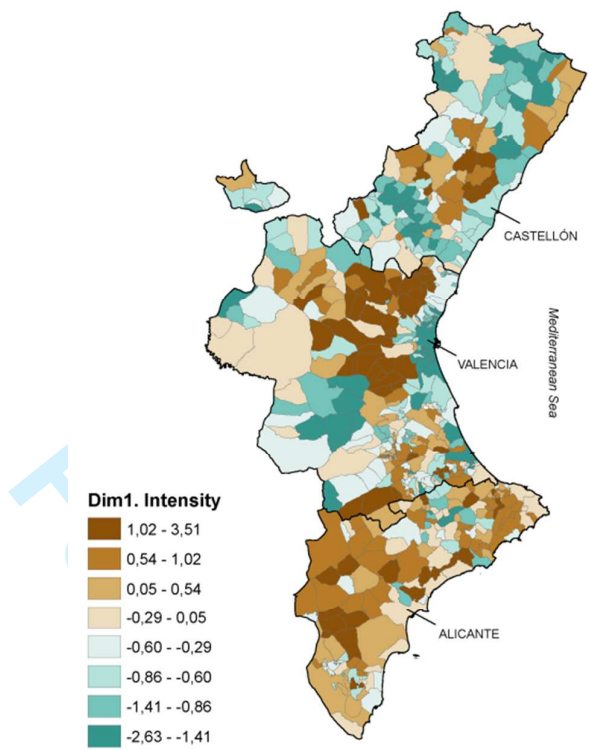


Figure 6. Map of Dimension 1, Intensity

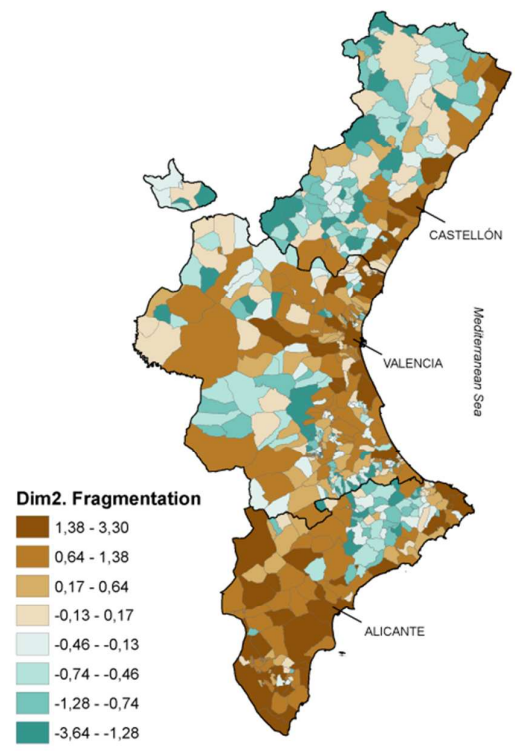


Figure 7. Map of Dimension 2, Fragmentation

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

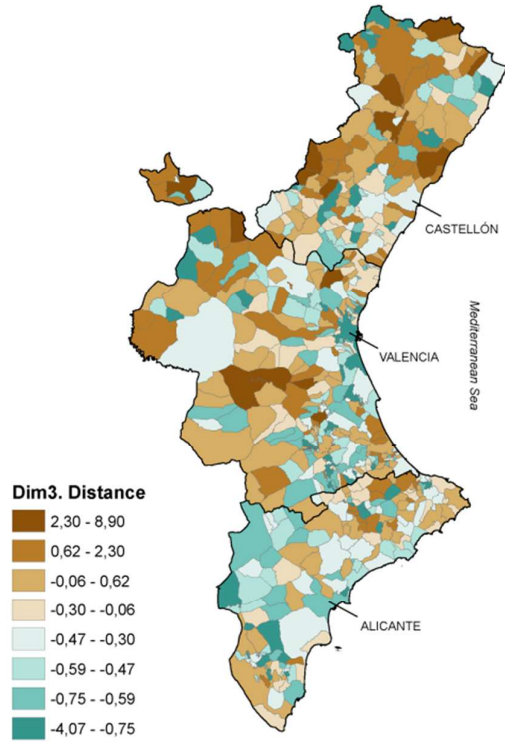


Figure 8. Map of Dimension 3, Distance

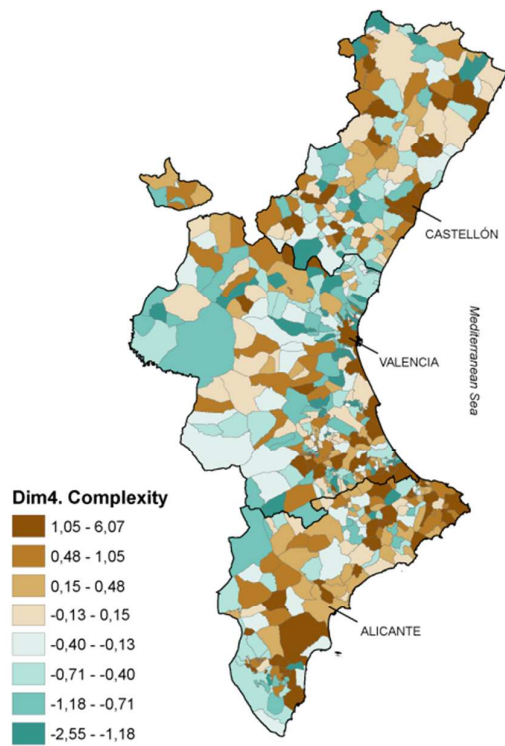


Figure 9. Map of Dimension 4, Complexity

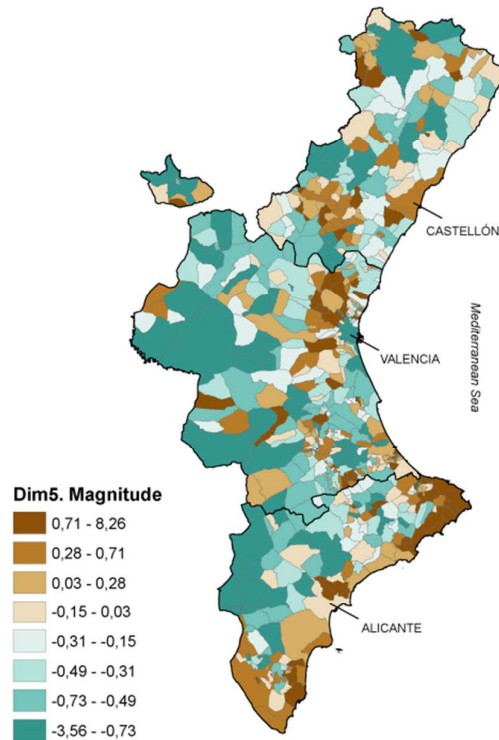


Figure 10. Map of Dimension 5, Magnitude

In addition, a single unidimensional measurement of each municipality has been obtained by computing the common factor of all the indicators, which quantifies the degree of sprawl in each municipality as shown in Figure 2. Significant values of the Sprawl Index appear from south to north, on the coast and in the intermediate area in the province of Alicante, especially in the north of Costa Blanca, in the western metropolitan area of the city of Valencia, in the metropolitan area of the city of Castellon and in some parts of the coast in the province of Castellon. These results reflect the consequences of urban planning in recent decades, in which the Valencian Community has experienced significant changes, from a dense and compact urban area to large low-density urban areas, probably as a consequence of higher living standards and the increased demand for second homes and tourist accommodation, especially in the coastal areas. At the other end, low sprawl values appear mostly in inland and rural areas where there is no demand for urban growth, and in

1
2
3 some coastal locations which mainly belong to small municipalities with severe space
4
5 restrictions for urban development.
6
7

8 **6. Discussion and Conclusions**

9

10
11 The purpose of this study was to characterize and understand urban sprawl in Valencia,
12
13 Spain and to obtain a unidimensional global index variable using a combination of ICA and
14
15 Bayesian Factor Analysis.
16

17
18 By applying ICA, we identified five dimensions of sprawl: density, dispersion and
19
20 fragmentation, distance to center, complexity and magnitude. On the basis of these
21
22 dimensions, urban sprawl can be understood as a complex model of occupation of a
23
24 significant part of the territory, characterized by a large low-density urban area and divided
25
26 into many widely separated and complex parts.
27
28

29
30 To study the spatial distribution of sprawl, we defined a simpler model using Bayesian
31
32 Factor Analysis to reduce the identified dimensions to a single index, easy to represent and
33
34 analyze, which has validated the research on urban development in Valencia. Over the past
35
36 20 years, many municipalities have experienced a paradigm shift and have gone from being
37
38 traditional compact areas to diffuse areas, e.g. the Sprawl Index shows higher values in
39
40 most of the coastal areas, in the province of Alicante and around the three major cities in
41
42 the region: Alicante, Valencia and Castellon de la Plana.
43
44
45

46
47 It is also important to point out some limitations of the research. Firstly, the lack of more
48
49 detailed information on the location of the population within the administrative boundaries
50
51 prevents us from attributing exact data on population or housing to each urban patch,
52
53 forcing us to work at the municipal level. The second problem arises from an over-
54
55 generalization of the database used (SIOSE) (Temes and Moya, 2016), which may give
56
57
58
59
60

1
2
3 anomalous information on land uses in some small municipalities. Despite these limitations,
4
5 which have mostly been corrected using statistical techniques, the results are satisfactory.
6
7 Another advantage of this approach is that the 12 indicators were built with simple
8
9 formulations from public and online databases, ensuring that the study can be repeated.
10

11
12 Secondly, the results from this study can be used to determine the costs of sprawl and
13
14 therefore support decision-making and planning processes through comparative
15
16 assessment of different development alternatives. Sprawl could be studied in relation to
17
18 sustainability or used to complement existing studies on how sprawl patterns affect
19
20 municipal costs and thus design a decision-making tool for municipal authorities.
21
22

23 24 25 **References**

26
27 Abdi H, Williams J, 2010, "Principal component analysis" *WIREs Computational Statistics*
28
29 **2**(4) 433–459, doi:10.1002/wics.101
30

31
32 Arribas-Bel D, Nijkampy P, Scholten H, 2011, "Multidimensional urban sprawl in Europe: a
33
34 self-organizing map approach" *Computers, Environment and Urban Systems* **35**(4) 263–
35
36 275
37
38

39
40 Bartholomew DJ, 2007, "Three faces of factor analysis", in *Factor analysis at 100. Historical*
41
42 *development and future directions* Eds Cudeck R, Maccallum RC pp 9-21
43
44

45
46 Benito B, Bastida F, Guillamón MD, 2010, "Urban Sprawl and the Cost of Public Services:
47
48 An Evaluation of Spanish Local Governments", *Lex Localis – Journal of local self-*
49
50 *government* **8**(3) 245-264, doi:10.4335/8.3.245-264(2010)
51

52
53 Burchell RW, Lowenstein G, Dolphin WR, Galley CC, Downs A, Seskin S, Still KG, Moore T,
54
55 2003, "Costs of Sprawl -- 2000. Transit Cooperative Research Program (TCRP)" Report 74,
56
57 Transportation Research Board, Washington
58
59
60

1
2
3 Cantarino I, Torrijo F, Palencia S, Gielen E, 2014, "Assessing residential building values in
4 Spain for risk analyses – application to the landslide hazard in the Autonomous Community
5 of Valencia" *Nat. Hazards Earth Syst. Sci.* **14** 3015-3030, doi:10.5194/nhess-14-3015-2014
6
7
8

9
10 Carruthers J, Ulfarsson G, 2003, "Urban sprawl and the cost of public services"
11 *Environment and Planning B: Planning and Design* **30**(4) 503 – 522
12
13

14
15 Chin N, 2002, "Unearthing the roots of urban sprawl: a critical analysis of form, function and
16 methodology" *WP 47, Centre for Advanced Spatial Analysis, University College London*
17
18

19
20 Colaninno N, Cerda J, Roca J, 2011, "Spatial patterns of land use: morphology and
21 demography, in a dynamic evaluation of urban sprawl phenomena along the Spanish
22 Mediterranean coast" *ERSA conference papers. Ersa* 11 p.776
23
24
25

26
27 Conti G, Frühwirth-Schnatter S, Heckman J.J, Piatek R, 2014, "Bayesian exploratory factor
28 analysis" *Journal of Econometrics* **183**(1) 31–57
29
30
31

32
33 Cutsinger J, Galster G, 2006, "There is No Sprawl Syndrome. A New Typology of
34 Metropolitan Land Use Patterns" *Urban Geography* **27**(3) 228-252
35
36

37
38 European Environment Agency (EEA), 2006, "Urban sprawl in Europe. The ignored
39 challenge" EEA. Report N° 10/2006
40
41
42

43
44 EEA, 2016, "Urban sprawl in Europe. Joint EEA-FOEN report" EEA. Report N° 11/2016
45
46

47 Ewing R, Pendall R, Chen D, 2002, *Measuring Sprawl and Its Impact*. Smart Growth
48 America
49
50

51
52 Ewing RH, 2008, "Characteristics, causes, and effects of sprawl: a literature review" In
53 *Urban Ecology*, Marzluff JM et al. (eds), pp 519–535, doi:10.1007/978-0-387-73412-5_34
54
55
56

1
2
3 Frenkel A, Ashkenazi M, 2008, "Measuring urban sprawl: how can we deal with it?"
4
5 *Environment and Planning B: Planning and Design* **35** 56-79 doi:10.1068/b32155
6
7

8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
Fulton W, Pendall R, Nguyen M, Harrison A, 2001, *Who Sprawls Most? How Growth
Patterns Differ Across the U.S.* Washington D.C. Center on Urban & Metropolitan Policy,
The Brookings Institution, Survey Series, 24p.

Garbiñe H, 2007, "Los costes económicos y sociales de la ciudad de baja densidad", in *La
ciudad de baja densidad: lógicas, gestión y contención* Indovina F, 203-240.

Galster G, Hanson R, Ratcliffe M, et al., 2001, "Wrestling sprawl to the ground: Defining and
measuring an elusive concept" *Housing Policy Debate* **12** 681-718

Graffelman J, 2013, "Factor Analysis" in *Encyclopedia of Environmetrics* 2nd Edition. Wiley

Hyvärinen A, Oja E, 2000, "Independent component analysis: algorithms and applications"
Neural Networks **13**(4-5) 411-430

Hortas-Rico M, 2010, "Urban sprawl and municipal budgets in Spain: a dynamic panel data
analysis" *Documents de Treball de l'IEB* 2010/43

Jaeger J, Schwick C, 2014, "Improving the measurement of urban sprawl: Weighted Urban
Proliferation (WUP) and its application to Switzerland" *Ecological Indicators* **38** 294-308.

Johnson M, 2001, "Environmental Impacts of Urban Sprawl: A Survey of the Literature and
Proposed Research Agenda" *Environment and Planning A* **33** 717-735. doi:10.1068/a3327

Johnson RA, Wichern DW, 2007, *Applied Multivariate Statistical Analysis*. Upper Saddle
River, NJ: Prentice Hall

1
2
3 Kass R, Wasserman L, 1995, "A reference Bayesian test for nested hypotheses and its
4 relationship to the Schwarz criterion" *Journal of the American Statistical Association*
5
6
7 **90**(431) 928-934
8

9
10 Kew B, Lee BD, 2013, "Measuring Sprawl across the Urban Rural Continuum Using an
11 Amalgamated Sprawl Index" *Sustainability* **5** 1806-1828. doi:10.3390/su5051806
12
13

14
15 Lunn DJ, Thomas A, Best N, Spiegelhalter D, 2000, "WinBUGS - A Bayesian modelling
16 framework: Concepts, structure, and extensibility" *Statistics and Computing* **10** 325-337
17
18

19
20 Marmolejo C, Stallbohm M, 2008, "En contra de la ciudad fragmentada: ¿hacia un cambio
21 de paradigma urbanístico en la Región Metropolitana de Barcelona?" *Scripta Nova, Revista*
22 *Electrónica de Geografía y Ciencias Sociales* Vol XII **270**(65)
23
24
25

26
27
28 Marí-Dell'Olmo M, Martínez-Beneito MA, Borrell C, Zurriaga O, Nolasco A, Domínguez-
29 Berjón MF, 2011, "Bayesian factor analysis to calculate a deprivation index and its
30 uncertainty" *Epidemiology* **22** 356-364
31
32
33

34
35
36 Mezzetti M, Billari FC, 2005, "Bayesian correlated factor analysis of socio-demographic
37 indicators" *Stat Methods Appt.* **14** 223-241
38
39

40
41 Miralles JL, Diaz S, Altur VJ, 2012, "Environmental impact on the Mediterranean Spanish
42 coast produced by the latest process of urban developments" *Proc of The Sustainable City*
43 *VII, WIT Press: Boston*
44
45
46

47
48 Miralles JL, 2014, *El darrer cicle immobiliari al País Valencià. O el progrés de la misèria.*
49 Fundació Nexè, Demos
50
51

52
53
54 Munoz F, 2003 "Lock living: Urban sprawl in Mediterranean cities" *Cities* **20**(6) 381-385,
55 doi:10.1016/j.cities.2003.08.003
56
57
58
59
60

1
2
3 Nechyba T, Walsh R, 2004, "Urban Sprawl" *Journal of Economic Perspectives* **18**(4) 177-
4
5 200

6
7
8 Ntzoufras I, 2009, *Bayesian Modeling using Winbugs*. John Wiley & Sons

9
10
11 Patacchini E, Zenou Y, 2009, "Urban Sprawl in Europe" *Papers on Urban Affairs*, 2009(1)
12
13 125-149. Project MUSE, muse.jhu.edu/article/271142

14
15
16 Paulsen K, 2014, "Geography, policy or market? New evidence on the measurement and
17
18 causes of sprawl (and infill) in US metropolitan regions" *Urban Studies* **51** 2629-2645
19
20 doi:10.1177/0042098013512874

21
22
23 Pichler N, 2007, "European Urban Sprawl: Sustainability, Cultures of (Anti) Urbanism and
24
25 "Hybrid Cityscapes". *Dela* **27** 101-133

26
27
28
29 R-Development Core Team, 2010, *R: A Language and Environment for Statistical*
30
31 *Computing*. Vienna, Austria: R Foundation for Statistical Computing. Online: [http://www.R-](http://www.R-project.org)
32
33 [project/org](http://www.R-project.org)

34
35
36 RERC. Real Estate Research Corporation, 1974, *The Cost of Sprawl, Environment and*
37
38 *Economic Costs of Alternative Residential Development Patterns at the Urban, Fringe*.
39
40 Washington D.C., U.S. Government Printing Office, 278p.

41
42
43
44 Shlens J, 2005, *A tutorial on principal component analysis*. Online:
45
46 <http://www.cs.cmu.edu/helaw/papers/pca.pdf>

47
48
49 Solé-Ollé A, Hortas-Rico M, 2008, "Does urban sprawl increase the costs of providing local
50
51 public services? Evidence from Spanish municipalities" *Documents de treball IEB* 6/2008

52
53
54 Speir C, Stephenson K, 2002, "Does Sprawl Cost Us All?" *Journal of the American Planning*
55
56 *Association* **68**(1) 56-70

1
2
3 Stone JV, 2004, *Independent Component Analysis: A Tutorial Introduction* MIT Press,
4
5 Boston

6
7
8
9 Temes R, Moya A, 2016, "Method Of Identifying Urban Morphologies From Cadastral
10 Attributes. Evidence From The Littoral Municipalities Of The Valencian Community (Spain)"
11
12 *International Journal of Sustainable Development and Planning* **11**(3) 316-324
13

14
15
16 Torrens PM, 2008, "A Toolkit for Measuring Sprawl" *Applied Spatial Analysis and Policy* **1**(1)
17
18 5-36, doi:10.1007/s12061-008-9000-x.
19

20
21
22 Wilson B, Chakraborty A, 2013, "The Environmental Impacts of Sprawl: Emergent Themes
23 from the Past Decade of Planning Research" *Sustainability* **5** 3302-3327,
24
25 doi:10.3390/su5083302
26
27

28
29 Zeng C, He S, Cui J, 2014, "A Multi-Level and Multi-Dimensional Measuring on Urban
30 Sprawl: A Case Study in Wuhan Metropolitan Area, Central China" *Sustainability* **6** 3571-
31
32 3598, doi:10.3390/su6063571
33
34
35

36
37 Zornoza C, 2013, *Crecimiento urbanístico en la zona costera de la Comunidad Valenciana*
38
39 *(1987-2009): Análisis y perspectivas de futuro*, Universitat de València, 152p.
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

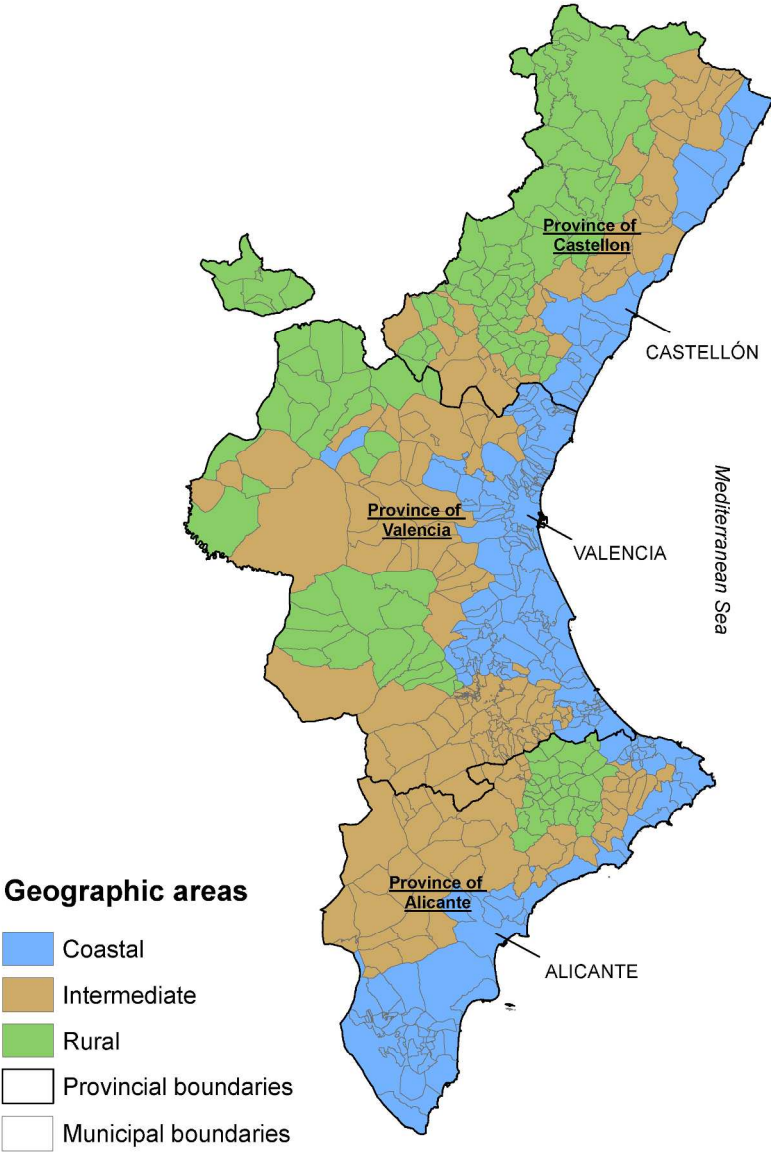


Figure 1. Geographical areas of the Valencian Community
 Figure 1
 210x296mm (300 x 300 DPI)

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

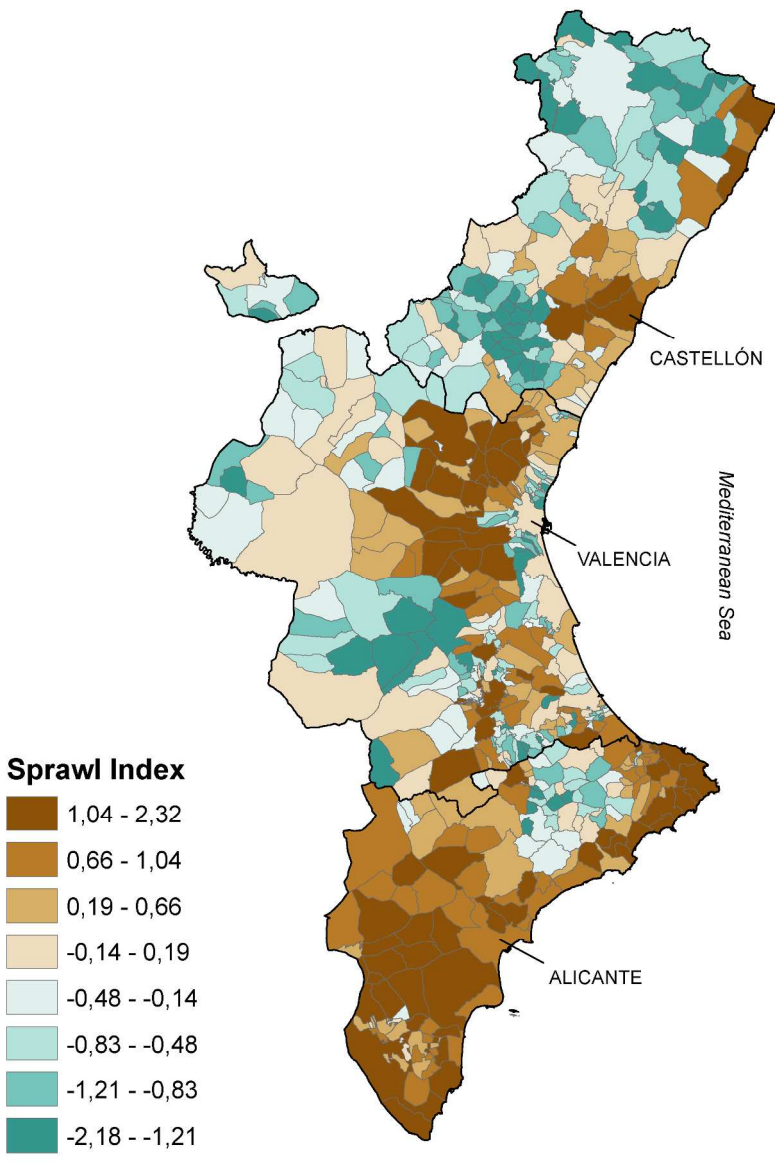


Figure 2. Map of sprawl index
Figure 2
210x296mm (300 x 300 DPI)

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

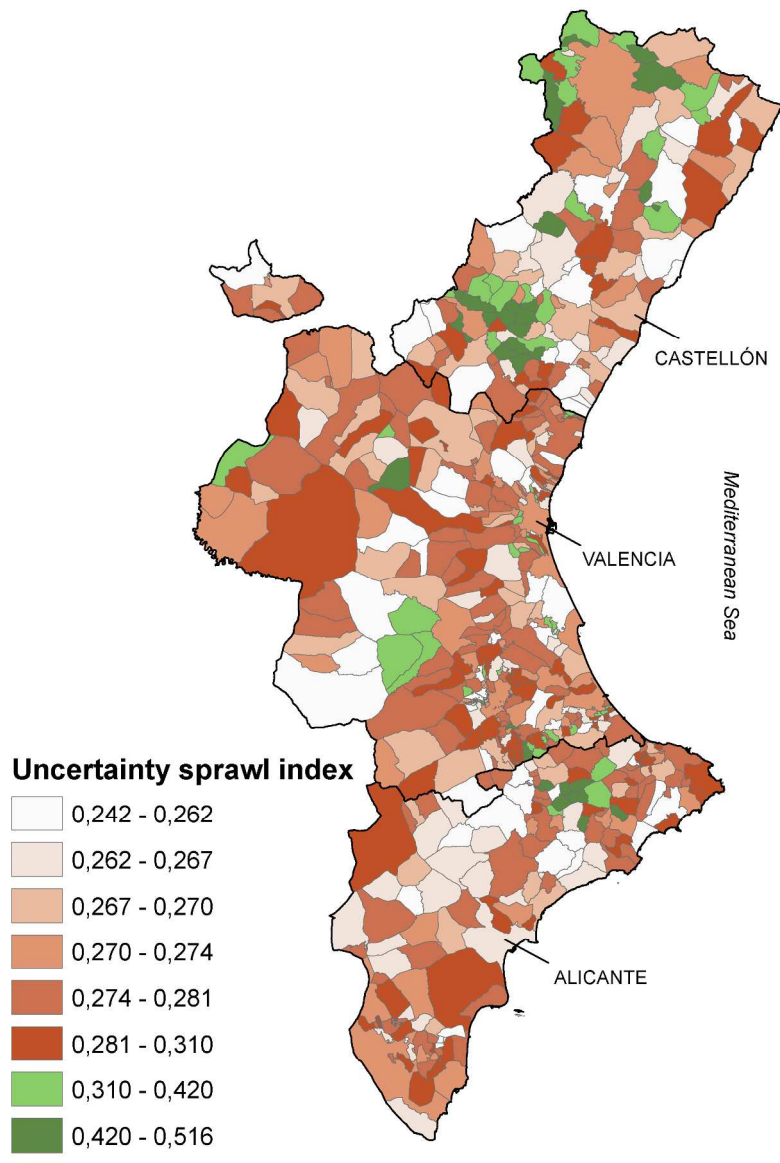


Figure 3. Map of uncertainty of sprawl index
 Figure 3
 210x296mm (300 x 300 DPI)

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

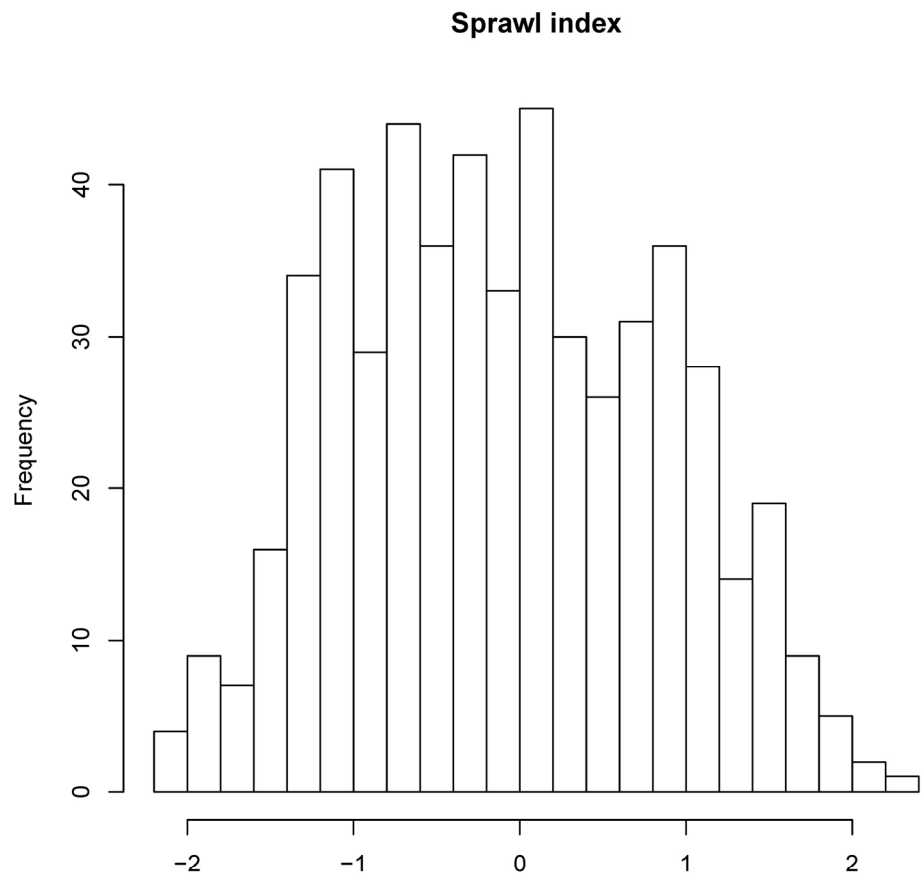


Figure 4. Histograms of sprawl index
Figure 4
177x177mm (300 x 300 DPI)

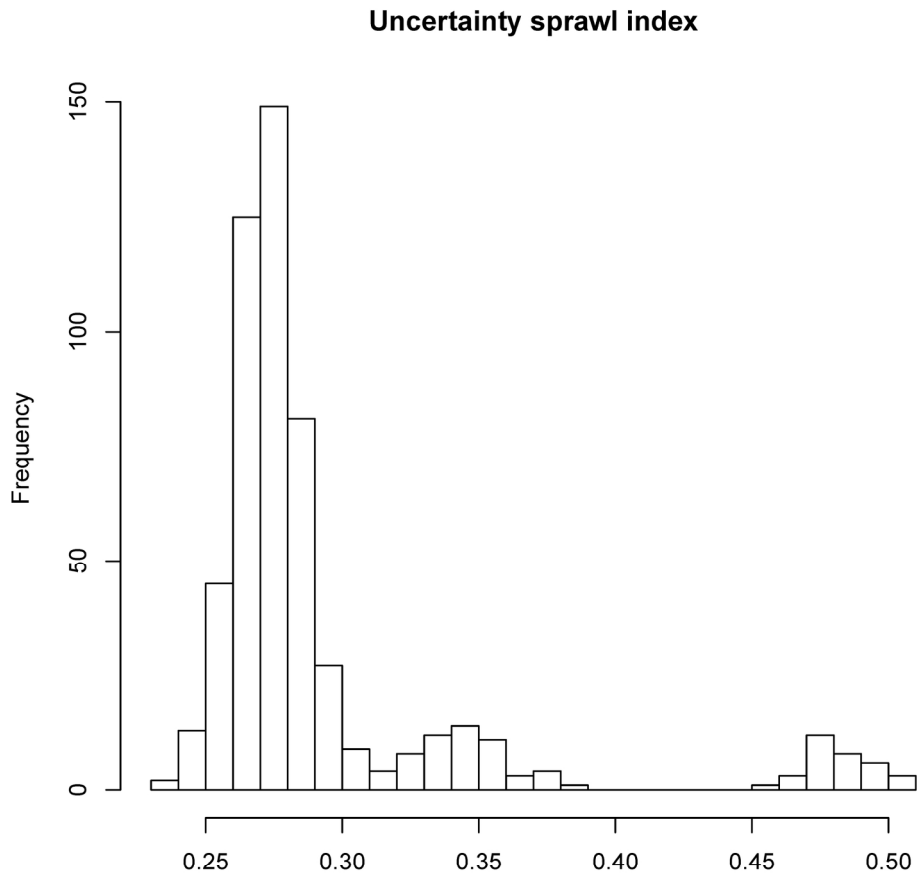


Figure 5. Histograms of the uncertainty of the sprawl index
 Figure 5
 177x177mm (300 x 300 DPI)

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

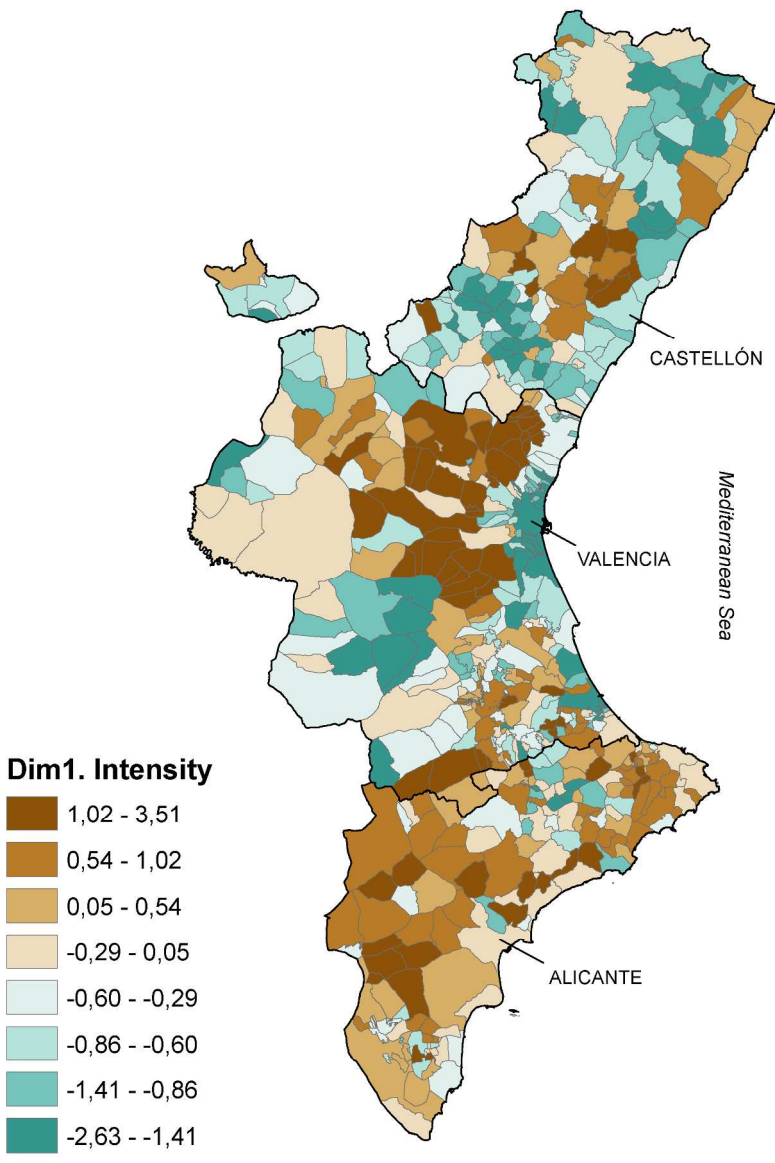


Figure 6. Map of Dimension 1, Intensity
Figure 6
210x296mm (300 x 300 DPI)

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

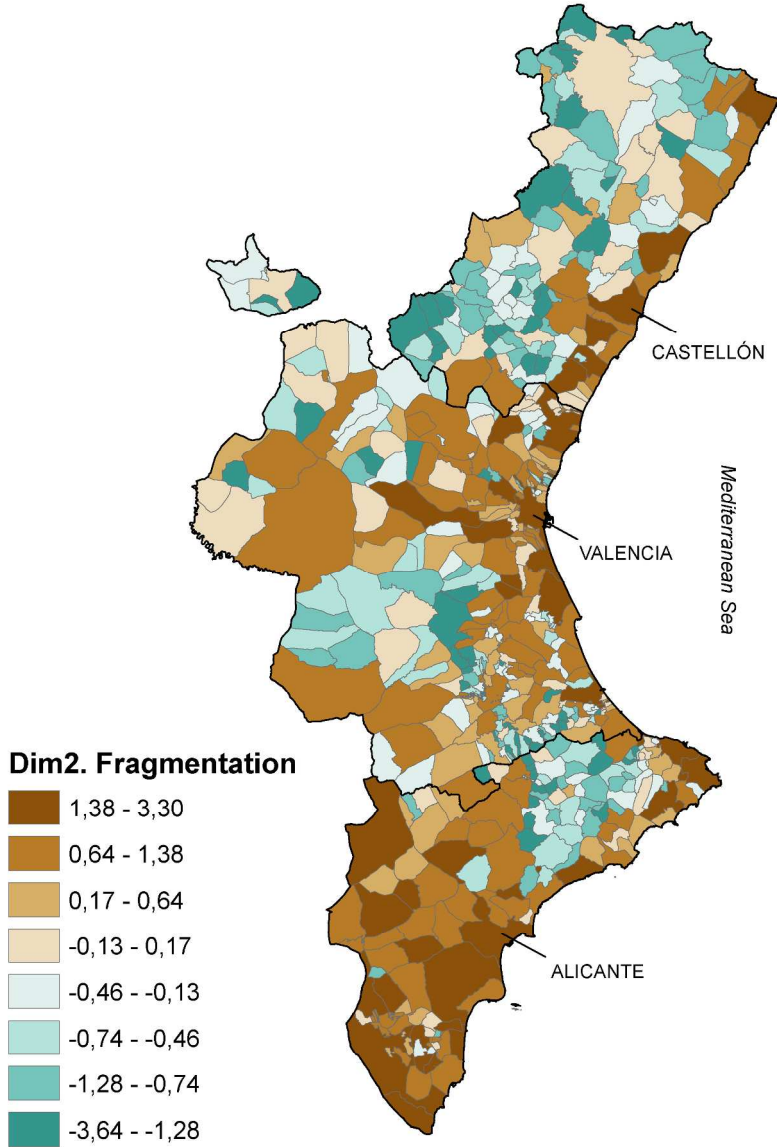


Figure 7. Map of Dimension 2, Fragmentation
Figure 7
210x296mm (300 x 300 DPI)

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

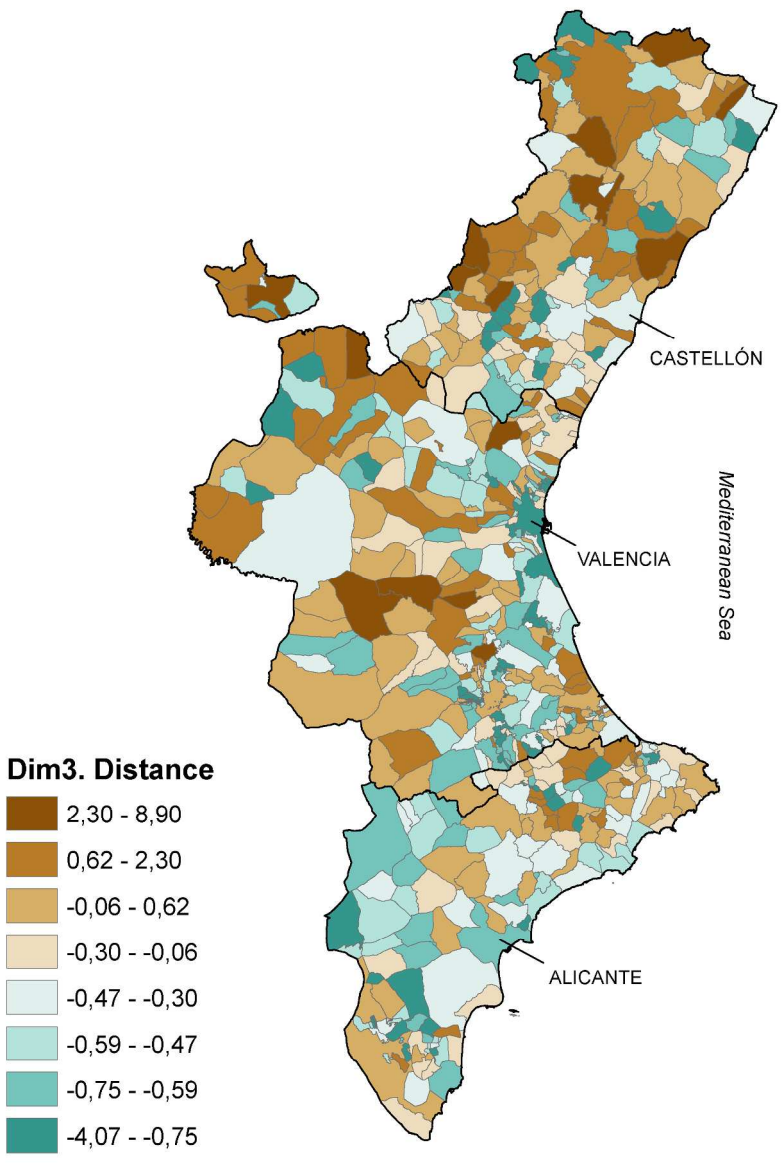


Figure 8. Map of Dimension 3, Distance
Figure 8
210x296mm (300 x 300 DPI)

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

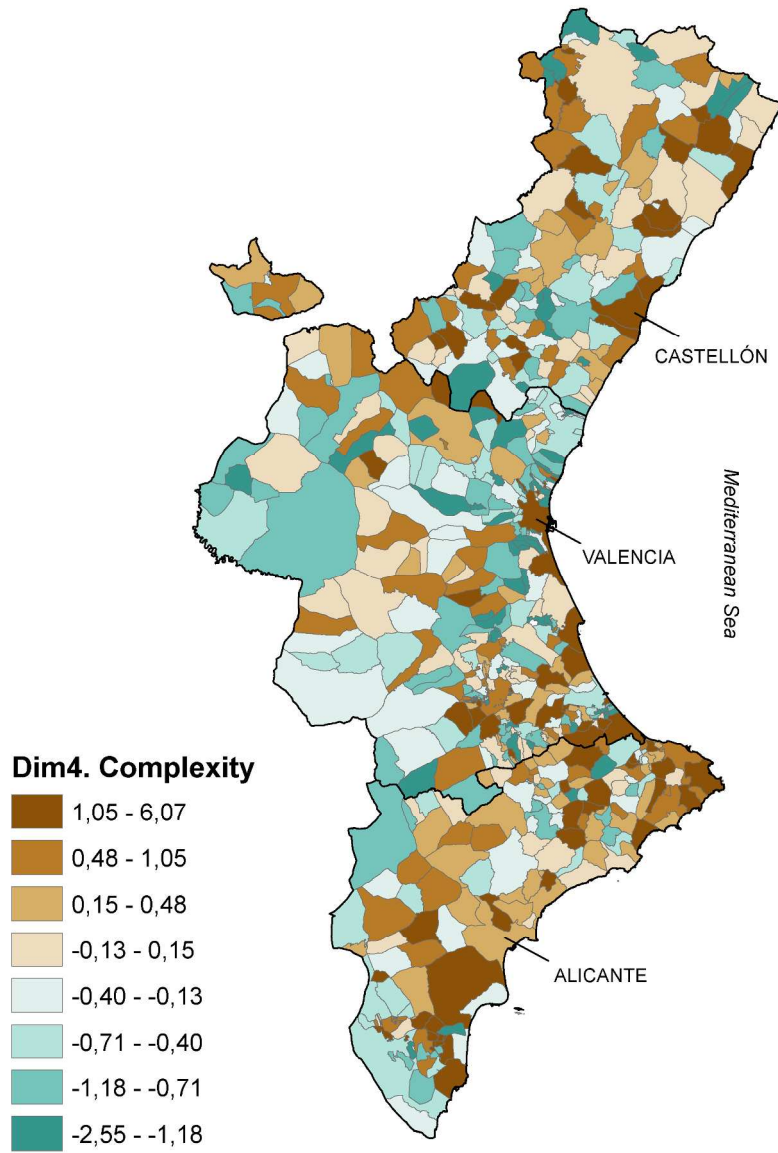


Figure 9. Map of Dimension 4, Complexity
Figure 9
210x296mm (300 x 300 DPI)

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

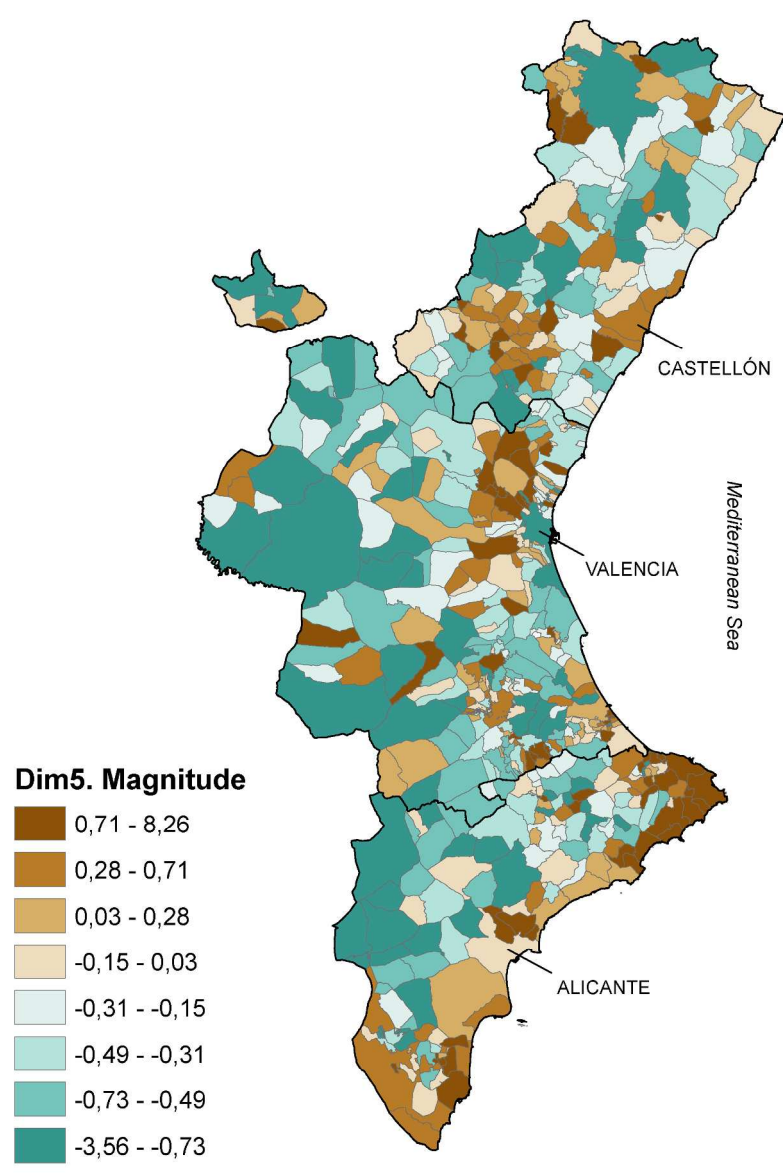


Figure 10. Map of Dimension 5, Magnitude
Figure 10
210x296mm (300 x 300 DPI)

Supplementary Material: Transformation of Variables

Preprocessing on the data to detect anomalous values as well as histogram transformations to adapt some variables to the needs of normality for Principal Component Analysis and Factor Analysis is done in this section.

Due to the limitations of the information sources some municipalities have a zero value in some of the indicators. The reason for that is the precision of the data-bases, which assigns a zero value when it actually should be a little more. These are, in general, small municipalities in inland and rural areas with very small populations, in which there is scarcely any discontinuous urban development, and they may even be represented by only one single urban patch, usually pertaining to a compact, historic center. Moreover, when the indicators are derivative indices from the data, they can also accumulate values of zero and/or one.

To avoid accumulating values of a certain quantity, we proceeded to remove these “anomalous” values, and treated them as missing values, which were then imputed from the Principal Component Analysis (PCA) in the case of (FA), Independent Component Analysis (ICA) and PCA. In the case of the Bayesian Factor Analysis, the imputation is implemented in the model and estimated at the same time as its resolution by the MCMC (Markov Chain Monte Carlo) technique (Hastings, 1970; Ntzoufras, 2009). 395 out of 6492 indicator values belonging to 114 out of 541 different municipalities were anomalous values, and were removed and treated as missing values.

What is more, most of observed indicators are not normal distributed; FA and PCA techniques assume normality in their input variables. Then, an appropriate

transformation over the observed indicators is needed before being used in the FA and PCA analysis. In Figure 1 non-normal variable distributions can be noted, as well as accumulated values of zero and one (“anomalous” values).

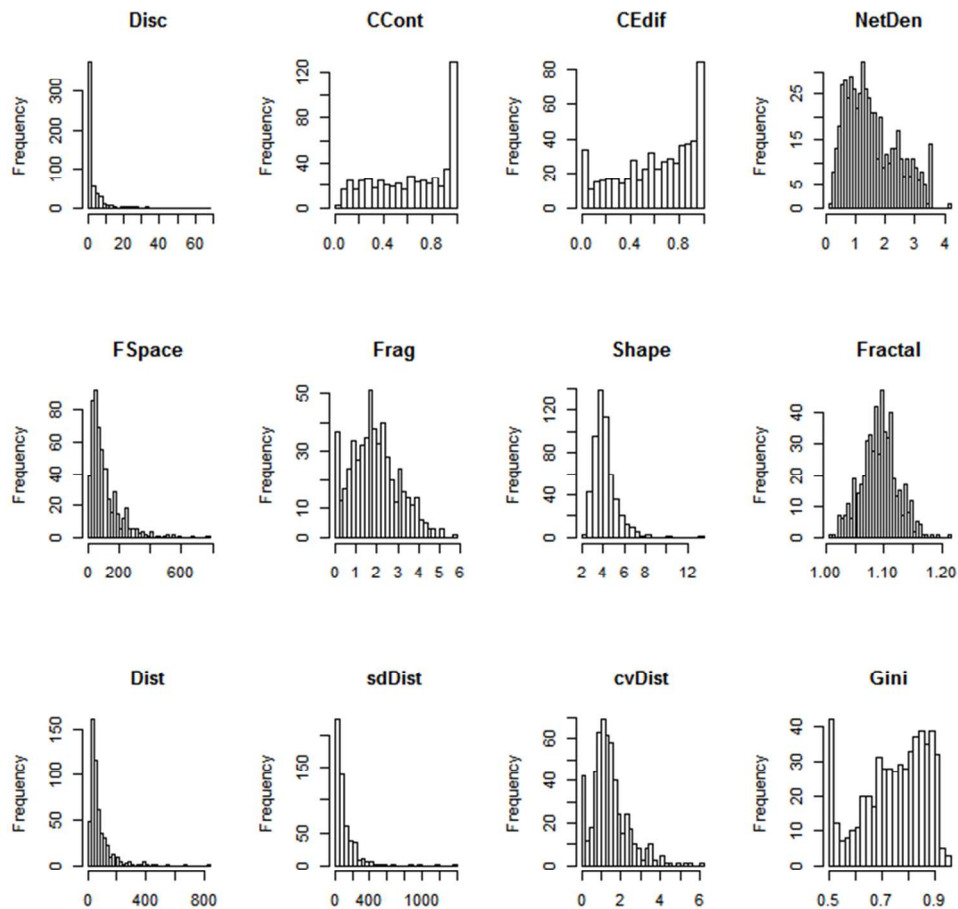


Figure 1. Original variable histograms.

The Disc, CCont, and CEdif observed indicators are proportions, so the logit transformation ($Z_{ij} = \log(Y_{ij}) - \log(1 - Y_{ij})$) has been applied to obtain normal distributions. Gini is a left-skewed variable to which the square transformation has been applied. FSpace is a density with right skewness to which logarithmic transformation has been applied. Dist and sdDist are continuous measures with right skewness and the root

square has been applied to them. The other indicators, Frag, cvDist, Shape and Fractal, can be considered as normal distributions. Finally, the resulting transformed variables are showed in Figure 2.

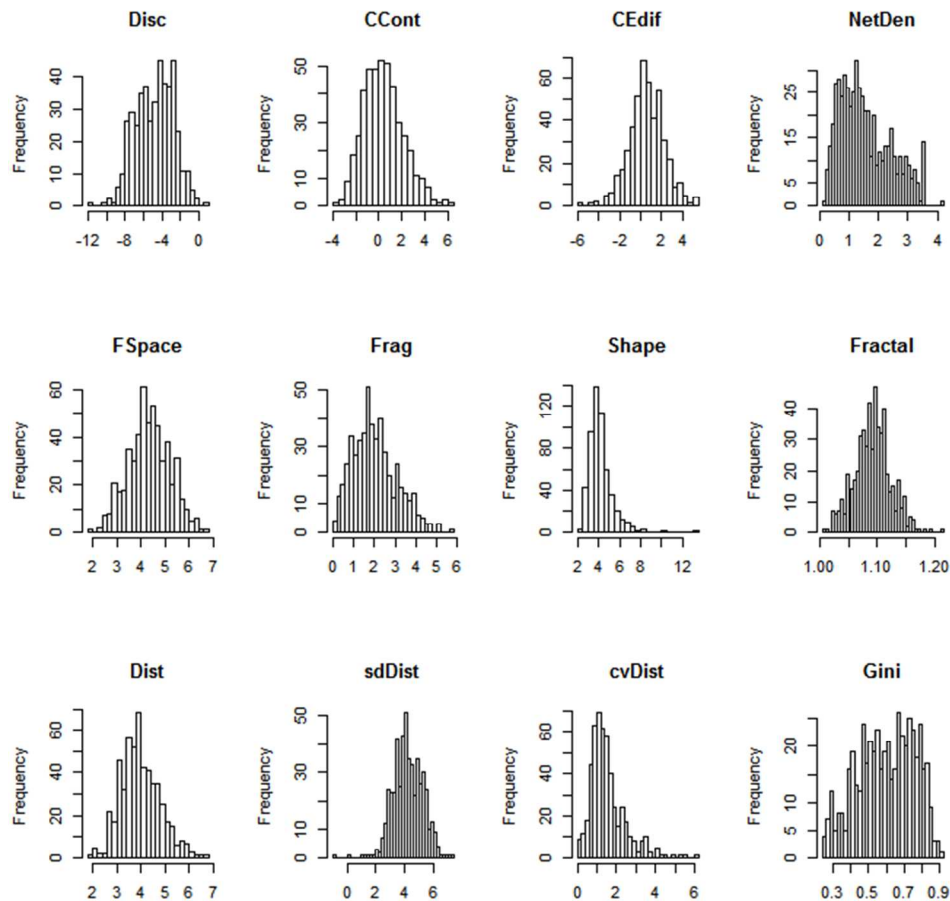


Figure 2. Transformed variables histograms

References:

- Hastings WK, 1970, "Monte Carlo Sampling Methods Using Markov Chains and Their Applications" *Biometrika* **57** 97–109
- Ntzoufras I, 2009, *Bayesian Modeling using Winbugs*. John Wiley & Sons