I.S.S.N.: 0213-7585

REVISTA DE ESTUDIOS REGIONALES

2ª EPOCA Enero-Abril 2022

123

SUMARIO

UNIVERSIDADES DE ANDALUCÍA

REVISTA DE ESTUDIOS REGIONALES Nº 123, I.S.S.N.: 0213-7585 (2022), PP. 161-196

The industrial-urban relative index (IURI) in spanish urban areas: The productive relevance of medium-sized cities

El índice relativo industria-ciudad (IRIC) en las áreas urbanas españolas: La relevancia productiva de las ciudades medias

Gonzalo Andrés López
Francisco Javier González Moya
Universidad de Burgos

Recibido, Marzo de 2020; Versión final aceptada, Septiembre de 2020.

KEYWORDS: Medium-size City; Industrial City; IURI; Chebyshev; PCA; Clustering.

PALABRAS CLAVE: Ciudad Media; Ciudad Industrial; IRIC; Chebyshev; PCA; Clustering.

Clasificación JEL: C3, O18.

ABSTRACT:

The Industrial-Urban Relative Index (IURI) is a formulation that allows the classification of the degree of intensity of the productive activity in the Spanish urban areas by means of the analytical crossing of diverse variables referred to the industrial employment, the surface of industrial land and the percentage of workforce. The application of this index has enabled the definition and classification of typologies of medium-sized industrial cities in Spain in a previous study. Based on this analysis, this paper considers the need to consolidate this analytical research, reinforcing the definition of the profile of the medium-sized, industrial Spanish city and its characterization with a discrete number of variables. This article analyzes the structure of the IURI obtained, suggests a new model using the PCA algorithm and compares both versions. A small sensitivity analysis of the index is also carried out and the degree of alteration of the initial variables is established, obtaining changes in the position of the industrial cities in the Spanish urban network. Finally, by means of clustering techniques, the profile of the industrial profile of medium-sized cities is reinforced, contributing to highlighting the productive relevance of this type of urban areas.

RESUMEN:

El Índice Relativo Industria-Ciudad (IRIC) es una formulación que permite clasificar el grado de intensidad de la actividad productiva en las áreas urbanas españolas mediante el cruce analítico de diversas variables referidas al empleo industrial, la superficie de suelo industrial y el porcentaje de población activa. La aplicación de este índice nos ha permitido definir y clasificar tipologías de ciudades medias industriales en España en un estudio previo. Partiendo de este análisis, en este trabajo se plantea la necesidad de consolidar esta investigación analítica, reforzando la definición del perfil de la ciudad media industrial española y su caracterización con un número discreto de variables. En este artículo se analiza la estructura del IRIC obtenido, se plantea un nuevo modelo mediante el algoritmo PCA y se comparan ambas versiones. Se realiza asimismo un pequeño análisis de sensibilidad del índice y se establece el grado de alteración de las variables iniciales, obteniendo cambios de posición de las ciudades industriales en la red urbana española. Finalmente, mediante técnicas de clustering se refuerza el perfil del carácter industrial de las ciudades medias, contribuyendo a destacar la relevancia productiva que tienen este tipo de áreas urbanas.

1. INTRODUCTION AND PREVIOUS APPROACH

In Spain, there are currently 48 medium-sized urban areas with approximately 50 and 300,000 inhabitants, which make up the central segment of the national urban network. This group holds over 20% of the total population of the country, but if we look closely at the meaning of the productive activities, we can observe that these cities concentrate about 25% of the employment rate and more than 30% of the Spanish industrial land. The ratio of industrial land per inhabitant in these urban areas is much higher than that of large cities and urban agglomerations –65 m^2/inhabitant versus 45 m^2/inhabitant. In more than half of these 48 medium-sized urban areas, there are more than 5,000 industrial jobs and in approximately the same number –22 of the 48 cities– the percentage of the industrial population is higher than the national average. These figures clearly indicate the great significance that industry has in these cities, and the notable and outstanding role it has played in their urban configuration and in their socio-economic structure. These statements are the result of several previous research works in which, by studying the significance of industrial activity in this type of cities, the role of the industry in each phase of the urbanization process has been identified, and typologies of medium-sized industrial cities in Spain have been established and classified Andrés López, 2019 y 2020).

One of the main conclusions of this line of research has been to determine that, actually, the size of cities does not define their true industrial specialization, since if we look at the socio-economic impact that industry has in relative terms, medium-sized cities are always listed in hegemonic order. More particularly, Spain only has six urban areas with more than 20% of the industrial workforce, and all of them are medium-sized cities (Avilés, Alcoy, Burgos, Manresa, Sagunto and Gijón); only ten cities that exceed 60 industrial jobs per every 1,000 inhabitants and seven of them are also medium-sized cities (Avilés, Palencia, Burgos, Vitoria, Manresa, Alcoy and Gerona); and only fifteen cities where industrial land occupies more than 30% of the total urbanized area, twelve of which are also medium-sized cities. The data are undoubtedly revealing: the special and socio-economic impact of the industry in these cities is highly noticeable. Briefly, this situation can be seen in the model of dispersion of relative variables that is attached as Figure 1. This representation summarizes in a graphic and concise way the diagnosis of the territorial analysis carried out in previous research.

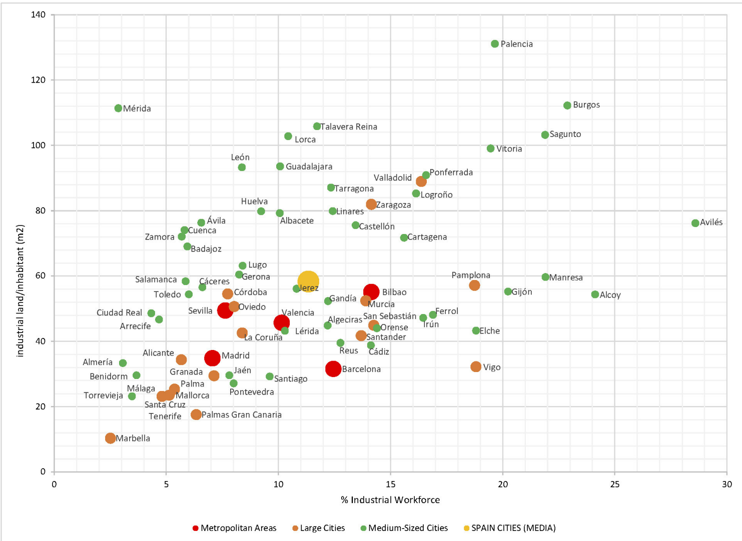
This diagnosis was conducted by applying a deductive methodological model, studying the existence of industrial specialization in medium-sized cities. Firstly, we began with the study of urban areas, as urban groups that gather a central municipality and some areas of influence in which industry is significantly important, as it extends to the adjacent municipalities. Of the many proposals for delimitation and classification of existing Spanish urban areas (Miramontes and Viera, 2016), the Functional Urban Areas (FUA) scheme included in the Urban Audit program has been used, as it is considered to be the most appropriate for a large number of medium-sized Spanish cities with areas of economic activity and for its possibilities of comparison with other European areas for further studies[1]. Thus, the research carried out focuses on the statistical study of the information available at the level of these urban areas, equating the concept of the medium-sized city to that of

1 In the studies carried out and in the review presented in this contribution, the model has been applied to the case of the Spanish urban network, in order to validate the operation of the index and its usefulness. However, its formulation is possible in groups of cities in any other country and will allow comparisons to be made in the future on the intensity of productive activity in cities in different areas. The results of its application will only depend on the reality of industrial activity in each set of cities and on the availability of the information necessary for the construction of the index (easily obtainable in Europe through the data of the Urban Audit program and in most countries of the world through the state statistics centers).

a medium-sized urban area. In fact, the terms "medium-sized urban areas" and "medium-sized cities" are used indistinctly in this paper, assuming the previous classifications existing on this segment of the Spanish urban scale (Bellet and Llop, 2000, 2004; Ganau y Vilagrasa, 2003; Andrés López, 2008; Sánchez Moral, 2009; Roca Cladera, 2012; Cebrián and Panadero, 2013; García Martín, 2014; Campos Sánchez, 2017; García, Martínez y Escudero, 2018). Therefore, all the available information refers to the whole of the urban area of each case, analyzing the reality of industrial cities as complex spaces that affect not only the central city but also its areas of influence in various surrounding municipalities.

FIGURE 1

DISPERSION OF THE RELATIVE VARIABLES RELATED TO INDUSTRIAL LAND AND INDUSTRIAL EMPLOYMENT (% INDUSTRIAL WORKFORCE AND INDUSTRIAL LAND/INHABITANT (M²) IN SPANISH URBAN AREAS



SOURCE: Author's elaboration from data obtained from the Spanish National Geographic Institute (Map of Land Occupation in Spain. CORINE Land Cover, 2012), National Statistics Institute (Municipal Register, 2017 and Urban Audit, 2017) and Social Security Treasury General Office (affiliation by activity sectors, 2017)

Likewise, it shall be pointed out that the space occupied in a generic sense by industrial spaces has been studied, taking into account that in many cases –although industry constitutes the main activity– it usually coexists with other tertiary activities. The information available in the sources used and at the level of detail of the urban scale handled does not allow for the disaggregation of such uses, so the impact that industrial estates (areas of economic activity) generate in cities is measured. The aim is to understand their comparative significance in urban areas as a whole from a general perspective (Andrés López, 2019).

In the light of these considerations, the information in this paper has been obtained from various sources and tabulated using different statistical tools and data mapping techniques. On the one hand, statistical information on the number of industrial companies and the volume of industrial employment provided by the National Statistics Institute and the Social Security Treasury General Office has been used. On the other, the industrial areas have been calculated from the measurement of the data obtained through technical work in a Geographic Information System (GIS). In this case, the information resulting from the CORINE Land Cover Project has been handled in order to obtain the measurement of the existing industrial spaces in each city. Avoiding the problems and exceptions we already know this source (Olazabal and Bellet, 2017, 2018; Bossard et al., 2000; Catalá Mateo et al., 2008; Díaz-Pacheco and Gutiérrez, 2014); lays out, a valid reference of the industrial surface in each urban area has been obtained. This information on the coverage of industrial land use has been crossed with the rest of the data on industrial activity. In this way, a significant number of variables have been analyzed and tabulated with different statistical techniques in order to obtain absolute and relative ratios and indicators that could illustrate this phenomenon in medium-sized cities.

Of all these variables, four indicators have finally been selected for further analysis to advance the following conclusions: percentage of the industrial workforce, industrial employment per 1,000 inhabitants, urbanized industrial land per inhabitant and industrial land with respect to the urbanized total. Based on the use of these indicators, categories and typologies of industrial cities have been established in the previous works referred, although the individual analysis of these rates did not allow for completion of a totally defined diagnosis on this issue. Due to the fact that not all the variables showed similar results, it was difficult to specify an overall view of the degree of industrial specialization of all the cities. To solve this problem, the last published work used the multivariate technique, combining these four indicators in a single index. The objective was not to remain only in the approximation allowed by the dispersion ranges or the classification of typologies according to the relative positions in the variables, but to achieve a classification of the impact of industry in these cities according to the combination of the four data together. Although the variables ordered individually already show quite clearly the results of relative concentration of industry in the medium-sized cities, the construction of a multivariate ranking on a composite index turned out to be more complete and interesting.

With this in mind, the Industrial-Urban Relative Index (IURI) has been drawn up. This is a formulation that allows the classification of the degree of relative intensity of the productive activity in the Spanish urban areas by means of the analytical crossing of the four indicated variables. This index has been obtained not only for medium-sized cities, but also for all Spanish urban areas, and it generates a ranking that classifies all the cities according to the intensity of the four variables. The index is based on a simple practical application of t theorem or inequality. With this as a reference, we have displayed a distance proportional to the standard deviation of each variable over its value, obtaining a high agglomeration tetra dimensional interval of points (the urban areas). In this way, each city has been positioned according to the four variables, and the relative index summarizes the joint effect of these variables in the characterization of industrial cities. Although by design it is an artificial and non-dimensional variable (it does not represent a real characteristic but the conjunction of four variables), we consider its usefulness not so much in the range of values it provides, but in the possibility of ordering the cities according to their value. As a result, those cities that are more typically industrial obtain a better index, and those that show a less defined industrial profile are placed at the bottom of the ranking and show lower values. The fact of having a variable with a high ranking is not what defines the city as industrial, but rather the obtaining of a high position in the set of all the variables. On the contrary, a low ranking in one or more of them relegates that urban area to the classification, even though it presents particularly high values in some of the other indicators.

TABLE 1
IURI, ORIGINAL VERSION

Ranking	City	IURI Index	Urban Area Tipology	Ranking	City	IURI Index	Urban Area Tipology
1	Palencia	100,00	MEDIUM-SIZED	36	Valencia	19,01	METROPOLITAN
2	Burgos	91,95	MEDIUM-SIZED	37	Zamora	18,57	MEDIUM-SIZED
3	Avilés	84,52	MEDIUM-SIZED	38	Vigo	18,53	LARGE
4	Vitoria/Gasteiz	73,56	MEDIUM-SIZED	39	Barcelona	18,17	METROPOLITAN
5	Logroño	55,06	MEDIUM-SIZED	40	Ávila	18,11	MEDIUM-SIZED
6	Sagunto	48,62	MEDIUM-SIZED	41	Reus	18,08	MEDIUM-SIZED
7	Manresa	45,84	MEDIUM-SIZED	42	Mérida	17,35	MEDIUM-SIZED
8	Alcoy/Alcoi	44,06	MEDIUM-SIZED	43	Jerez de la Frontera	17,19	MEDIUM-SIZED
9	Pamplona/Iruña	42,47	LARGE	44	Cuenca	16,72	MEDIUM-SIZED
10	Valladolid	40,52	LARGE	45	Santander	16,47	LARGE
11	Zaragoza	40,40	LARGE	46	Algeciras	15,82	MEDIUM-SIZED
12	Tarragona	40,22	MEDIUM-SIZED	47	A Coruña	15,51	LARGE
13	Ponferrada	38,64	MEDIUM-SIZED	48	Lleida	14,47	MEDIUM-SIZED
14	Lorca	38,23	MEDIUM-SIZED	49	Sevilla	13,20	METROPOLITAN
15	Talavera de la Reina	34,82	MEDIUM-SIZED	50	Córdoba	13,06	LARGE
16	Gijón	34,78	MEDIUM-SIZED	51	Salamanca	12,49	MEDIUM-SIZED
17	Bilbao	33,33	METROPOLITAN	52	Cáceres	11,52	MEDIUM-SIZED
18	Castellón de la Plana	31,73	MEDIUM-SIZED	53	Badajoz	11,12	MEDIUM-SIZED
19	Donostia / San Sebastián	28,28	LARGE	54	Toledo	10,91	MEDIUM-SIZED
20	Cartagena	27,36	MEDIUM-SIZED	55	Madrid	9,86	METROPOLITAN
21	Guadalajara	26,15	MEDIUM-SIZED	56	Santiago de Compostela	9,68	MEDIUM-SIZED
22	Irun	25,90	MEDIUM-SIZED	57	Jaén	8,62	MEDIUM-SIZED
23	León	25,64	MEDIUM-SIZED	58	Pontevedra	7,62	MEDIUM-SIZED
24	Linares	25,36	MEDIUM-SIZED	59	Alicante	6,86	LARGE
25	Huelva	23,89	MEDIUM-SIZED	60	Ciudad Real	6,79	MEDIUM-SIZED
26	Albacete	23,40	MEDIUM-SIZED	61	Granada	6,67	LARGE
27	Ourense	23,22	MEDIUM-SIZED	62	Arrecife	6,17	MEDIUM-SIZED
28	Girona	22,82	MEDIUM-SIZED	63	Almería	5,92	MEDIUM-SIZED
29	Ferrol	22,56	MEDIUM-SIZED	64	Santa Cruz de Tenerife	5,26	LARGE
30	Murcia	22,49	LARGE	65	Palma	4,89	LARGE
31	Elche/Elx	21,95	MEDIUM-SIZED	66	Las Palmas de Gran Canaria	4,72	LARGE
32	Gandía	20,96	MEDIUM-SIZED	67	Málaga	4,64	LARGE
33	Cádiz	20,70	MEDIUM-SIZED	68	Benidorm	2,69	MEDIUM-SIZED
34	Lugo	19,77	MEDIUM-SIZED	69	Torrevieja	2,28	MEDIUM-SIZED
35	Oviedo	19,26	LARGE	70	Marbella	0,00	LARGE

SOURCE: Author's elaboration based on data obtained from the Spanish National Geographic Institute (Land Occupancy Map in Spain. CORINE Land Cover, 2012), National Statistics Institute (Municipal Register, 2017 and Urban Audit, 2017) and the Social Security Treasury General Office (affiliation by activity sectors, 2017). Variables used: % of industrial workforce, industrial employment per 1,000 inhabitants, urbanized industrial land per inhabitant, and industrial land with regard to the urbanized total.

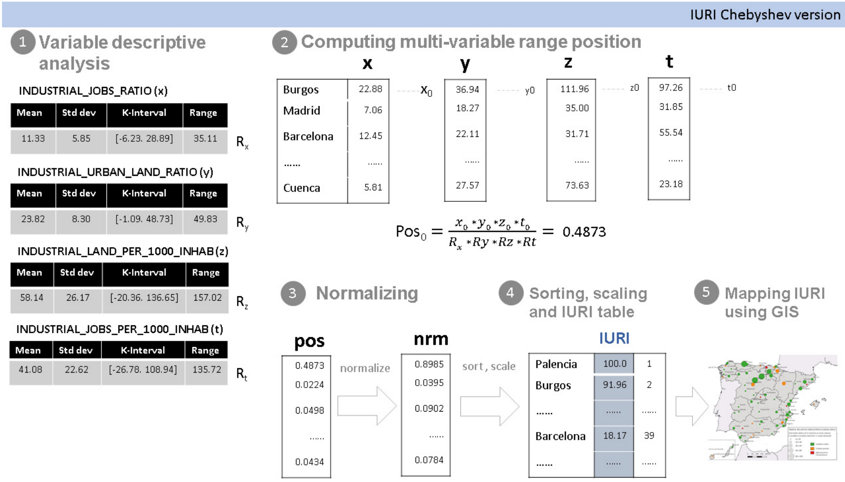
The original definition of this index and its results are shown in Table 1. From the highest to the lowest value, the index shows the cities with the highest industrial profile on a scale from 100 to 0. It can be observed how the upper part of the table is mostly occupied by medium-sized cities. The industrial profile of many of these cities in Spain is clearly defined, and the initial diagnosis posed by the previous results obtained with the dispersion diagrams, the graphs or the different maps that graphically represent the distribution of these cities according to their industrial intensity is confirmed in a ranking.

2. OBJECTIVES AND WORKING METHODOLOGY

The first version of IURI or Chebyshev version is based on the known Chebyshev inequality (Dragomir, 2019), which shows –in a very simple way– an interval containing a high percentage of the values of a random variable, at a certain confidence level. It uses the mean and standard deviation to be constructed. A striking aspect about this technique is that it does not require any previous condition of the distribution of the variable, achieving an interval that is relatively coarse depending on the selected confidence level. It contains at least a certain percentage of the observations. The confidence level "k" is proportional to the percentage of values by the expression $(1\text{-}1/k^2)$. This technique can be applied to a single variable or extended jointly to a group of variables, as in this case. A simple methodology has been developed to achieve a "relative position" of the values provided by a city in relation to the four-dimensional interval of the variables considered at the same time. Once normalized and put on a scale of zero to one hundred, it offers the first IURI index; as shown schematically in Figure 2A.

Based on this approach and the previous results shown, this article reviews the formulation of the IURI index, analyzes its quality, strengths and weaknesses, and establishes a new method of obtaining it, based on the PCA (Principal Component Analysis) statistical model. Once the new version of IURI is obtained, a comparison with the previous version is carried out, checking the improvements that take place in the generation of the index, without losing the fundamental objective of highlighting the fully industrial feature of these cities.

FIGURE 2A
DIAGRAM SHOWING THE METHODOLOGY FOR OBTAINING THE CHEBYSHEV VERSION OF THE IURI

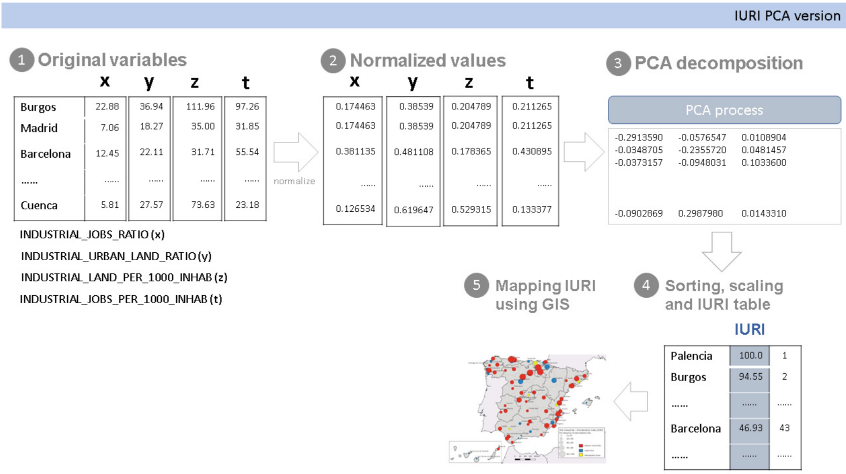


Source: Author's elaboration.

The technique used to generate the second version of the IURI is based on the well-known statistical model PCA (Smith, 2002; Abdi and Willimas, 2010). Given a set of numerical variables that characterize an entity (in this case, cities), and a large number of variables, PCA finds a set of artificial variables, derived from the initial ones. These are lower in number (fewer variables) and explain the variability in the original data (Andrecut, 2009). We therefore move from manipulating a high set of N numerical variables to a set of K numerical variables, so that K is less than N, managing to explain to a high degree the same information by reducing the dimension. Moreover, it is possible to know, after the application, the weight or relative importance of each of the new variables (known as main components). In our case, reducing the number of variables is not the problem, since we work with four or five.

As we will see, our interest is focused on the relative weights, since this is what will allow us to generate the new IURI as a linear combination of the weights (percentage of variability explained) applied to the main components. We present an abbreviated scheme of this methodology in Figure 2B.

FIGURE 2B
DIAGRAM DETAILING THE METHODOLOGY USED TO OBTAIN THE PCA VERSION OF THE IURI



Source: Author's elaboration.

A sensitivity analysis of both versions of IURI was carried out by performing unit tests for each variable and city and observing the degree of variation of the indicator. This was done by assuming a gain or loss of a position in the indicator, and by observing what degree of loss or gain is recorded in the variable. In this manner, we establish in a unitary way what will be the behavior of this indicator in the face of changes in the model's input variables.

When dealing with sensitivity analyses for the Chebyshev version of IURI, we have employed spreadsheets and used the "What-if Analysis" tool, which in practice allows us to pose and solve small linear programming problems in the search for a target value for a variable, altering another one with which it is related. By gathering the results obtained and summarizing the information, the tables provided are obtained.

As for the sensitivity analysis on the PCA version of the IURI, we have proposed variation ratios for each variable, up to a range from 50% loss in value to 50% gain. This has allowed a high set of test cases (70 * 70 * 5 * 9 = 220,500 cases) where with the help of a Python script (Ayyadevara, 2018; VVAA, 2020) the value of each variable has been modified for each city, regenerating the calculated IURI. Once again, we have used spreadsheets to provide a summary of the information that will be detailed in the accompanying tables.

Finally, by means of a well-known clustering technique (K-means), the conclusions previously obtained on the basis of the IURI are reinforced in both versions. On the industrial profile of the medium-sized city, without taking into account the indicator, we study the behavior of the numerical variables that define it and with a clustering technique we identify the profile of the cities with the greatest industrial specialization in the set of medium-sized urban areas.

When we considered letting the cities group “by themselves”, we carried out a test based on the well-known classification algorithm (clustering) called K-means. Given a set of numerical variables that characterize an entity (cities), it is possible to create groups so that, according to the data, the cities most closely related to each other are concentrated with respect to a center (Jain and Dubes, 1988; Huang, 1998; Hernández et al., 2004; Ding, 2004; Yingyu et al., 2013; Adnan et al., 2013). The aim is to achieve a double objective when grouping; on the one hand, to minimize the distance of each center (centroid) between the most similar cities and to maximize the distance with those that are less similar (Wu et al., 2008). This technique allows us to define and specify more precisely the real profile of the medium-sized industrial city in Spain. The centers or centroids are artificial points whose coordinates are calculated by the algorithm itself. In fact, we must first specify how many centroids (groups) we wish to group our cities into. Iteratively, the algorithm creates the centroids, calculates the distance between each centroid and each city, and assigns a class or group label to each city –depending on the closest centroid. This would be closest to choosing a number of leaders in a group of people and grouping them around those who are most like-minded, while also ensuring that those who are present are the least like-minded. In other words, this algorithm allows the identification of medium-sized cities with a clearer industrial profile by virtue of the variables used (industrial employment, industrial land and industrial workforce), as well as the clustering of their counterparts with more similar characteristics.

This grouping occurs a controlled number of times (iterations), so that the algorithm changes both the position of centroids and the type or label assigned to each city because of the new distances calculated. The central concept of distance must be understood here as Euclidean distance, whose mathematical definition is the square root of the sum of squares of differences from the mean value. As mentioned, the choice of the number of groups (clusters) is a decision made before applying the algorithm, and we have performed the corresponding tests with different number of clusters. We used a Python script (Ayyadevara, 2018) and summarized the results with the help of spreadsheets, since they compare the position in the group (cluster) with the ranking that each city obtains in the two versions of IURI.

3. RESULTS AND DISCUSSION

3.1 Review of the industrial-urban relative index (iuri) in Spain

As previously stated, in the relative industrial specialization ranking established by the IURI index, medium-sized cities always appear with a very striking productive identity, i.e. with a clear influence of the industrial activity in their territory. Most of the first cities classified by the index are medium-sized cities, and a clear differentiation can be observed in the spatial influence that industry has according to the size of the cities in the Spanish urban system.

More particularly, depending on the calculation variations of the first version, these cities occupy 9 of the first 10 places, or even all 10, as the case may be, within this classification. As observed, 8 of the medium-sixed cities out of the 48 existing ones (17%) occupy the first 10 positions of the ranking and 28 (58%) the upper half of the table, with IURI values ranging from 19.77 to 100 for the first position.

An additional concern arises when analyzing the structure of the index and its distribution. It is true that this is an artificial and dimensionless index that gives each city a value on a scale of 0 to 100 points, so the first classification objective is achieved. However, we also wonder if the distribution in the score is appropriate, so that it offers possible values in all quartiles in a reasonable way for those cities where this is the case, as shown in Table 2.

The distribution of the index values offers 3 cases in Q4 (4%), 2 values in the Q3 (3%), 17 values in Q2 (24%) and finally groups the less industrial cities in Q1 with 48 cases (69%).

The distribution by city typology is also presented in Table 2. Without altering the conclusions on the fully industrial profile of the cities that appear classified as such, the distribution offers some zeros in quartiles Q3 and Q4.

Actually constitutes the exception to the result generally observed. For this reason, we deem it necessary to provide an IURI definition that improves these “architectural” aspects of the index itself, without losing its true nature. The aim is to advance in the definition of the profile of the medium-sized industrial city in order to clarify the spatial influence that productive activity has on this type of city.

TABLE 2
DISTRIBUTION OF THE NUMBER OF CITIES ALONG THE QUARTILES OF THE IURI VALUE. DISTRIBUTION BY TYPE

Quartiles	Typology							
	Overall		Medium – sized		Large		Metropolitan	
	No	%	No	%	No	%	No	%
Q1. From 0 to 25	48	69%	29	60%	13	76%	4	80%
Q2. From 25 to 50	17	24%	14	29%	4	24%	1	20%
Q3. From 50 to 75	2	3%	2	4%	0	0%	0	00%
Q4. From 75 to 100	3	4%	3	6%	0	0%	0	0%
	70		**48**		**17**		**5**	

Source: Author’s elaboration.

3.2. Towards a new formulation

Following this reasoning and without losing sight of the objective already achieved of classifying and characterizing medium-sized cities as urban areas with a clear industrial profile, we have developed a second method for generating the IURI.

In this case, a well-known statistical model called PCA is used, (Smith, 2002; Ordóñez et al., 2014) which allows working with a large set of variables by reducing their number. Faced with a large set of data in terms of characteristics (columns), the model generates a smaller number of artificial variables, but is capable of explaining a high percentage of the variability of the initial variables. In some way, it takes advantage of the correlations between the initial variables to express the same information with fewer characteristics or columns. In addition to this, the model allows us to know the degree of variability expressed by each new column, called “main component”, which also appear in order of their importance.

For our particular case, the objective is not so much to obtain a reduction in dimensions (given that the number of variables is very low, and the level of significance we are looking for is high), but rather to know the relative importance of each main component.

In the construction of the IURI, according to this version, we decided to work with the variables for 2016 published by EuroStat [25] in particular in the section "Functional Urban Areas". In this way, we have referred all the variables derived from the population to that published for 2016.

In some of the tests carried out, in addition to contemplating the population variable updated to 2016, we have located an estimate of the Activity Rate. This is defined as the quotient between the employed population and the working-age population. This is a general variable (it does not speak specifically of industry), but it offers an idea of the level of economic activity of the population of a city.

For this reason, we have carried out tests to obtain an IURI based on PCA using sets of four (PCA 4) and five variables (PCA 5), the fifth variable being the activity rate referred to. In addition to obtaining the index in both scenarios, we compared both solutions and observed whether this fifth variable allows us to reinforce the IURI obtained or, on the contrary, contributes a negligible amount to what has already been expressed. Our hypothesis regarding this fifth variable maintains that it could reinforce the other variables, particularly the industrial employment rate.

Although this was our initial hypothesis, the results summarized in Table 2A do not present a very clear improvement over the IURI with four variables, shown in Table 2B below.

For the sake of certainty, Table 3A shows the PCA result with five variables (PCA 5), using the variables updated to 2016 and the activity rate as the fifth variable.

TABLE 2A
DISTRIBUTION OF THE NUMBER OF CITIES ALONG THE QUARTILES OF THE IURI VALUE WITH 5 VARIABLES EUR 2016 (PCA 5). DISTRIBUTION BY TYPOLOGY

Quartiles	Typology							
	Overall		Medium - sized		Large		Metropolitan	
	No	%	No	%	No	%	No	%
Q1. From 0 to 25	14	20%	7	15%	7	41%	0	0%
Q2. From 25 to 50	29	41%	19	40%	6	35%	4	80%
Q3. From 50 to 75	22	31%	17	35%	4	24%	1	20%
Q4. From 75 to 100	5	7%	5	10%	0	0%	0	0%
	70		**48**		**17**		**5**	

Source: Author's elaboration.

TABLE 2B
DISTRIBUTION OF THE NUMBER OF CITIES ALONG THE QUARTILES OF THE IURI VALUE WITH 4 VARIABLES EUR 2016 (PCA 4). DISTRIBUTION BY TYPOLOGY

Quartiles	Typology							
	Overall		Medium - sized		Great		Metropolitan	
	No	%	No	%	No	%	No	%
Q1. From 0 to 25	13	19%	6	13%	7	41%	0	0%
Q2. From 25 to 50	29	41%	19	40%	6	35%	4	80%
Q3. From 50 to 75	23	33%	18	38%	4	24%	1	20%
Q4. From 75 to 100	5	7%	5	10%	0	0%	0	0%
	70		**48**		**17**		**5**	

Source: Author's elaboration.

TABLE 3A
IURI CLASSIFICATION IMPLEMENTED WITH PCA (PCA 5, EUR 2016)

Ranking	City	IURI Index	Urban Area Tipology	Ranking	City	IURI Index	Urban Area Tipology
1	Palencia	100,00	MEDIUM-SIZED	36	Ferrol	44,65	MEDIUM-SIZED
2	Burgos	95,98	MEDIUM-SIZED	37	Ávila	44,31	MEDIUM-SIZED
3	Avilés	88,86	MEDIUM-SIZED	38	Valencia	44,01	METROPOLITAN
4	Vitoria/Gasteiz	86,84	MEDIUM-SIZED	39	Cuenca	43,79	MEDIUM-SIZED
5	Logroño	78,28	MEDIUM-SIZED	40	Oviedo	43,05	LARGE
6	Sagunto	70,95	MEDIUM-SIZED	41	Reus	42,13	MEDIUM-SIZED
7	Lorca	70,89	MEDIUM-SIZED	42	Barcelona	42,09	METROPOLITAN
8	Manresa	68,79	MEDIUM-SIZED	43	Jerez de la Frontera	41,68	MEDIUM-SIZED
9	Tarragona	67,98	MEDIUM-SIZED	44	Vigo	41,64	LARGE
10	Alcoy	67,13	MEDIUM-SIZED	45	Algeciras	38,89	MEDIUM-SIZED
11	Pamplona/Iruña	66,29	LARGE	46	Santander	37,47	LARGE
12	Zaragoza	66,26	LARGE	47	Coruña (A)	37,16	LARGE
13	Valladolid	65,84	LARGE	48	Lleida	37,12	MEDIUM-SIZED
14	Talavera de la Reina	65,83	MEDIUM-SIZED	49	Sevilla	35,46	METROPOLITAN
15	Ponferrada	63,07	MEDIUM-SIZED	50	Córdoba	34,73	LARGE
16	Bilbao	58,95	METROPOLITAN	51	Badajoz	33,57	MEDIUM-SIZED
17	Gijón	58,87	MEDIUM-SIZED	52	Salamanca	33,53	MEDIUM-SIZED
18	Castellón de la Plana/Castelló	57,90	MEDIUM-SIZED	53	Cáceres	32,65	MEDIUM-SIZED
19	Guadalajara	56,78	MEDIUM-SIZED	54	Toledo	32,43	MEDIUM-SIZED
20	Cartagena	54,13	MEDIUM-SIZED	55	Madrid	30,05	METROPOLITAN
21	Donostia-San Sebastián	53,05	LARGE	56	Santiago de Compostela	26,95	MEDIUM-SIZED
22	León	53,00	MEDIUM-SIZED	57	Arrecife	24,74	MEDIUM-SIZED
23	Huelva	52,68	MEDIUM-SIZED	58	Jaén	24,25	MEDIUM-SIZED
24	Linares	52,12	MEDIUM-SIZED	59	Ciudad Real	22,81	MEDIUM-SIZED
25	Albacete	51,23	MEDIUM-SIZED	60	Alicante/Alacant	22,79	LARGE
26	Irun	50,79	MEDIUM-SIZED	61	Almería	21,94	MEDIUM-SIZED
27	Girona	50,57	MEDIUM-SIZED	62	Pontevedra	21,54	MEDIUM-SIZED
28	Mérida	48,61	MEDIUM-SIZED	63	Granada	21,29	LARGE
29	Murcia	48,11	LARGE	64	Palma de Mallorca	20,01	LARGE
30	Elche/Elx	47,98	MEDIUM-SIZED	65	Santa Cruz de Tenerife	19,46	LARGE
31	Gandia	46,56	MEDIUM-SIZED	66	Las Palmas	16,41	LARGE
32	Cádiz	45,81	MEDIUM-SIZED	67	Málaga	16,39	LARGE
33	Zamora	45,29	MEDIUM-SIZED	68	Benidorm	10,83	MEDIUM-SIZED
34	Ourense	45,19	MEDIUM-SIZED	69	Torrevieja	8,08	MEDIUM-SIZED
35	Lugo	44,68	MEDIUM-SIZED	70	Marbella	0,00	LARGE

Source: Author's elaboration.

Once implemented, the new reference IURI is shown in Table 3B below.

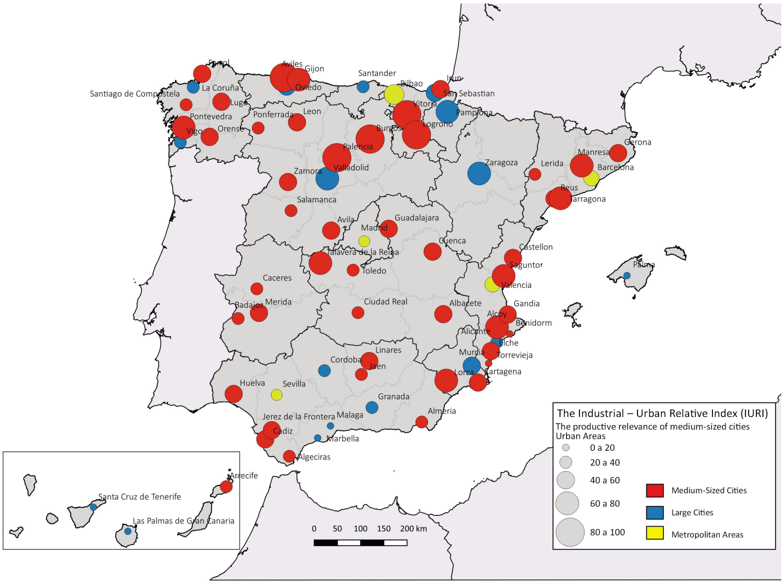
TABLE 3B
IURI CLASSIFICATION IMPLEMENTED WITH PCA (PCA 4, EUR 2016)

Ranking	City	IURI Index	Urban Area Tipology	Ranking	City	IURI Index	Urban Area Tipology
1	Palencia	100,00	MEDIUM-SIZED	36	Ávila	46,71	MEDIUM-SIZED
2	Burgos	94,55	MEDIUM-SIZED	37	Oviedo	46,69	LARGE
3	Avilés	87,04	MEDIUM-SIZED	38	Ferrol	46,59	MEDIUM-SIZED
4	Vitoria/Gasteiz	86,17	MEDIUM-SIZED	39	Valencia	45,06	METROPOLITAN
5	Logroño	80,20	MEDIUM-SIZED	40	Elche/Elx	44,53	MEDIUM-SIZED
6	Lorca	74,73	MEDIUM-SIZED	41	Jerez de la Frontera	44,30	MEDIUM-SIZED
7	Sagunto	71,67	MEDIUM-SIZED	42	Reus	42,37	MEDIUM-SIZED
8	Talavera de la Reina	69,80	MEDIUM-SIZED	43	Barcelona	40,93	METROPOLITAN
9	Tarragona	67,93	MEDIUM-SIZED	44	Vigo	39,69	LARGE
10	Valladolid	66,67	LARGE	45	Algeciras	39,66	MEDIUM-SIZED
11	Ponferrada	66,32	MEDIUM-SIZED	46	Coruña (A)	39,12	LARGE
12	Zaragoza	66,22	LARGE	47	Santander	37,76	LARGE
13	Manresa	66,10	MEDIUM-SIZED	48	Lleida	37,27	MEDIUM-SIZED
14	Alcoy	65,85	MEDIUM-SIZED	49	Sevilla	37,19	METROPOLITAN
15	Pamplona/Iruña	64,18	LARGE	50	Salamanca	37,06	MEDIUM-SIZED
16	Gijón	63,29	MEDIUM-SIZED	51	Córdoba	36,66	LARGE
17	Bilbao	61,69	METROPOLITAN	52	Badajoz	36,43	MEDIUM-SIZED
18	Castellón de la Plana/Castelló	58,29	MEDIUM-SIZED	53	Cáceres	35,34	MEDIUM-SIZED
19	León	57,80	MEDIUM-SIZED	54	Toledo	31,37	MEDIUM-SIZED
20	Guadalajara	56,81	MEDIUM-SIZED	55	Madrid	29,46	METROPOLITAN
21	Linares	56,59	MEDIUM-SIZED	56	Santiago de Compostela	26,76	MEDIUM-SIZED
22	Huelva	55,64	MEDIUM-SIZED	57	Jaén	26,12	MEDIUM-SIZED
23	Donostia-San Sebastián	55,33	LARGE	58	Ciudad Real	24,83	MEDIUM-SIZED
24	Cartagena	55,27	MEDIUM-SIZED	59	Almería	24,49	MEDIUM-SIZED
25	Irun	52,79	MEDIUM-SIZED	60	Arrecife	24,04	MEDIUM-SIZED
26	Mérida	52,71	MEDIUM-SIZED	61	Alicante/Alacant	23,30	LARGE
27	Albacete	52,29	MEDIUM-SIZED	62	Pontevedra	23,12	MEDIUM-SIZED
28	Zamora	50,38	MEDIUM-SIZED	63	Granada	22,13	LARGE
29	Cádiz	49,38	MEDIUM-SIZED	64	Santa Cruz de Tenerife	19,44	LARGE
30	Girona	48,19	MEDIUM-SIZED	65	Palma de Mallorca	17,68	LARGE
31	Lugo	47,91	MEDIUM-SIZED	66	Málaga	17,05	LARGE
32	Gandia	47,61	MEDIUM-SIZED	67	Las Palmas	16,70	LARGE
33	Murcia	47,53	LARGE	68	Benidorm	12,05	MEDIUM-SIZED
34	Ourense	47,12	MEDIUM-SIZED	69	Torrevieja	9,88	MEDIUM-SIZED
35	Cuenca	47,01	MEDIUM-SIZED	70	Marbella	0,00	LARGE

Source: Author’s elaboration.

FIGURE 3

THE NEW INDUSTRIAL-URBAN RELATIVE INDEX (IURI) ON URBAN SPANISH AREAS. THE RELEVANCE OF MEDIUM-SIZED CITIES. ICRI PCA 4 RANKING EUR 2016 VARIABLES VERSION



Source: author's elaboration. Graph elaborated with data from the Spanish National Geographic Institute (Map of land occupation in Spain. CORINE Land Cover, 2012), National Statistics Institute (Municipal Register 2017 and Urban Audit 2017) and Social Security Treasury General Office (affiliation by activity sectors, 2017). Variables used: % industrial workforce, industrial employment per 1.000 inhabitants, industrial land per inhabitant, and industrial land over urbanized land

In order to make homogeneous comparisons with the first version of IURI, we chose to keep the method with the four original variables (PCA 4), updated to the 2016 version of EuroStat, as the PCA version (Eurostat, 2020). At this point, it is possible to obtain for each city a scalar value as a linear combination of the main components and their degree of importance. The rearrangement of this scalar value brings us a new ranking, and the value normalized from 0 to 100 offers a new IURI value according to this technique. The objective is achieved, and the final result for each of the cities can be compared. In general terms, the guideline is the stability in the ranking of each of the cities, or in positions immediately above or below those they occupied in the previous ranking.

Exceptionally, there are some cities that present a significant rise, such as Lorca, which progresses from position 14 to the current position 6. As opposed to Manresa or Alcoy, which have significantly dropped from positions 7 and 8 to positions 13 and 14, respectively. These specific movements identify cases of cities that qualify their industrial position in the ranking as a consequence of their reformulation, when considering the degree of general activity and weighing up the variability. In any case, apart from these nuances, stability in the ranking is observed as a general trend, with the clear industrial specialization of medium-sized cities maintained above large urban areas.

From the point of view of absolute values of the index, our concern for adequate distribution persists. We can observe that, in general, a greater diversity of values is achieved in each quartile, shown in Table 2B. It is shown how zero values disappear at one of the intersections. In addition, the percentages in Q1 are smoothed out, distributing their weight towards Q2 and Q3 (going from 60%, 76% and 80% in Q1 to the current values of 13%, 41% and 0%). The quartiles Q2 and Q3 now provide more balanced values than in the previous case.

All in all, we can affirm that we have achieved an index that classifies the medium-sized industrial city effectively. In this new version, 9 of the 10 first cities (19%) are medium-sized urban areas and 23 of them in the first half of the table also belong to this category (48%).

3.3 IURI sensitivity considerations

An additional concern arises with regard to the sensitivity of the index insofar as it may vary according to the possibilities of progression or regression in the index by the cities. This means that, if in the future the variables that characterize each city change drastically, that urban area may rise or fall in the ranking by obtaining a score very different from the current one.

In the first version of the IURI (Chebyshev version), the statistical process based on Chebyshev inequality (Dragomir, 2019) provides, *a priori*, and based on its construction. It offers great stability to the classification in the face of discrete changes in the variables, given that at a confidence level k=5 the confidence intervals of the variables are very generous. As a result, a not very intense change in the value of a variable has a very moderate impact. This can be interpreted positively as strong point of the method of construction of this objective indicator and its validity over time, but also as a signal of insensitivity to discrete changes in one or more starting variables.

However, we have carried out a sensitivity analysis and studied each variable separately in each of the cities by observing what is the expected variation in the value of the variable to make the city gain or lose a position in the ranking. It is very likely that a multivariate sensitivity analysis would be much more complete but, still, the partial perspective of each variable provides us with relevant information.

The sensitivity analysis is reduced to an optimization problem for each particular case. It can offer three situations: no optimal solution, a mathematically valid optimal solution but in unacceptable reality (in case of suggesting negative values for these variables), or a positive numerical consistent value (and relatively realistic) with possible values of the variable.

After having synthesized the results of the test carried out, the following tables provide the conclusions in each case for each variable. We now show the previous table for medium-sized, large and metropolitan cities. Each of this tables show the percentage value of each variable that allows either gaining or losing a position in the ranking is detailed. Frequently, we obtain a series of values that form the range that is shown.

Tables 4, 5 and 6 provide these data. They reflect the reduction (below 100%) or increase in value (above 100%) that each individual variable must contribute to achieve a gain or loss of a position in the ranking:

TABLE 4
SENSITIVITY FOR MEDIUM-SIZED CITIES. % OF THE ORIGINAL VALUE OF THE VARIABLE

Medium – sized	Lose a position	Gain a position
Industrial employment (%)	Between 36% and 68%	Over 108%
Industrial land vs urban land	Between 36% and 84%	Over 104%
Industrial land per-inhabitant (m^2)	Between 36% and 612%	Over 108%
Industrial employment per 1,000 inhab.	Between 36% and 68%	Over 108%

Source: Author's elaboration.

TABLE 5
SENSITIVITY FOR LARGE CITIES. % OF THE ORIGINAL VALUE OF THE VARIABLE

Large	Lose a position	Gain a position
Industrial employment (%)	36%	Over 228%
Industrial land vs urban land	36%	Between 164% and 612%
Industrial land per inhabitant (m^2)	Between 36% and 612%	Between 228% and 612%
Industrial employment per 1,000 inhab.	No value	Between 228% and 612%

Source: Author's elaboration.

TABLE 6
SENSITIVITY FOR METROPOLITAN CITIES. % OF THE ORIGINAL VALUE OF THE VARIABLE

Metropolitan	Lose a position	Gain a position
Industrial employment (%)	Between 100% and 128%	92%
Industrial land vs urban land	Between 36% and 100%	80%
Industrial land per inhabitant (m^2)	Between 36% and 124%	92%
Industrial employment per 1,000 inhab.	128%	92%

Source: Author's elaboration.

In short, and as a general trend, in 3 of the 4 variables, medium-sized cities must more than double their value to gain a position in the IURI (PCA), while a sharp fall in their value (of 32% or more, or 16% or more) allows them to lose a position in the index. For practical purposes, it is a reasonable stability.

In the case of large cities, gaining a position becomes almost impossible, while losing a position is unlikely. Finally, urban agglomerations can lose a position relatively easily, while gaining a position is also feasible (although this only occurs in the case of Madrid).

Due to the sensitivity of the model with regard to the version of the IURI under the new formulation (PCA), we have carried out a test in which the values of the variables for each of the cities were intentionally altered.

For this particular version we have altered, with unit tests, the value of each variable in each city, from losses of 50%, 25%, 10% and 5% to gains in the same ratios. For each case (each tuple city - variable - ratio), the IURI has been recalculated and the relative position of each of the cities has been noted down. In the interest of simplicity, the cases of loss of value are gathered separately from the cases of gain of value, in order to provide a clearer picture, without descending to what occurs particularly for each case.

TABLE 7

SENSITIVITY FOR MEDIUM-SIZED CITIES. % OF CASES THAT LOSE, MAINTAIN OR GAIN A POSITION IN IURI

	Decreasing value [-50%, 0%)				Increasing value (0, +50%]			
	-1 pos	Neutral	+1 pos	% explained	-1 pos	Neutral	+1 pos	% explained
Industrial employment (%)	23%	46%	6%	75%	8%	81%	10%	99%
Industrial land vs urban land	15%	38%	13%	66%	0%	94%	6%	100%
Industrial lad per inhabitant (m^2)	8%	38%	13%	59%	8%	83%	8%	99%
Industrial employment per 1,000 inhab.	15%	58%	13%	86%	6%	90%	4%	100%

Source: Author's elaboration.

Once the results have been obtained, they are summarized in Tables 7, 8 and 9, and appear separated by city typology. We report the percentage of cases that lose, maintain, or gain a position in the IURI due to the change in the value of the variable in ranges of loss of value [-50%, 0%] or gain of value (0%, 50%).

TABLE 8

SENSITIVITY FOR LARGE CITIES. % OF CASES THAT LOSE, MAINTAIN OR GAIN A POSITION IN IURI

	Decreasing value [-50%, 0%)				Increasing value (0, +50%]			
	-1 pos	Neutral	+1 pos	% explained	-1 pos	Neutral	+1 pos	% explained
Industrial employment (%)	6%	88%	0%	94%		100%		100%
Industrial land vs urban land	12%	76%	12%	100%		100%		100%
Industrial land per inhabitant (m^2)	24%	65%	0%	89%		100%		100%
Industrial employment per 1,000 inhab.	0%	82%	12%	94%		100%		100%

Source: Author's elaboration.

TABLE 9

SENSITIVITY FOR METROPOLITAN CITIES. % OF CASES THAT LOSE, MAINTAIN OR GAIN A POSITION IN IURI

	Increasing value [-50%, 0%)				Decreasing value (0, +50%]			
	-1 pos	Neutral	+1 pos	% explained	-1 pos	Neutral	+1 pos	% explained
Industrial employment (%)		100%		100%		100%		100%
Industrial land vs urban land		100%		100%		100%		100%
Industrial ground per inhabitant (m^2)		80%	20%	100%		100%		100%
Industrial employment per 1,000 inhab..		100%		100%		100%		100%

Source: Author's elaboration.

In the case of medium-sized cities there is greater dispersion, although as shown, changes in loss of value lead to the continuity or alteration of a position in more than 66% of cases. While those of gain explain this situation in most of the cases, in at least 99% of cases, according to each variable.

Regarding large cities, neutrality or the loss of a position in the case of a loss in value of the variable is highlighted, while neutrality predominates in the case of a gain in value. In each of them, the percentage of cases explained exceeds 89%. Finally, neutrality to these changes is the keynote of metropolitan areas, in 3 of the 4 variables.

It may be concluded that the PCA version of the IURI (PCA 4) incorporates some sensitivity to changes in the case of medium-sized cities, although it is true that position stability is offered at least 59% in cases of loss of value, and at least 99% in cases of gain in value. For practical purposes, a much more discrete scale is expected in the behavior of the variables, which means that this version of IURI enjoys a reasonable stability and allows us to consolidate the identification of the medium-sized Spanish industrial cities.

3.4. Affinity and clustering: the grouping of medium-sized cities with a more defined industrial profile

So far, we have focused our efforts on observing the industrial profile of medium-sized cities through IURI, which allows a ranking as a tool. This is an artificial –albeit useful– construction to highlight our hypothesis of this profile in this type of city in a major way.

We will eventually steer our efforts towards characterizing these cities by themselves, letting them gather by affinity according to the data of the variables provided, but without creating any artificial variable, using a data mining technique called clustering.

We have managed to group together cities of a purely industrial profile. In fact, we have let them group themselves together. To this effect, we have used the well-known K-means algorithm through the concept of Euclidean distance [21] to ensure that the cities that are most similar to each other are grouped around a centroid or mean point.

This algorithm takes as inputs a set of numerical data of the number of city groups to be gathered. If we think of the available numerical variables as the coordinates of a point, the algorithm calculates as many centroids as desired, and it will try to position the most similar cities very close to that centroid and very far from the rest of the centroids. Once this is done, it will measure the distance from each city to each centroid (mean) and will assign it a label by proximity indicating the group it belongs to. In this way, each city is classified and grouped by its closest centroid.

The following steps iteratively recalculate the coordinates of new centroids and successively both the belonging of each city to each group and the position of each centroid, with the double objective of minimizing the distance to the centroid of one's group and at the same time maximizing the distance to the centroids of the rest of the groups. The algorithm ends up offering the belonging of each city to each cluster or group and the distance of each city to its centroid. This allows the selection of the medium-sized cities with the highest industrial profile in Spain, clearly identifying them according to their productive specialization.

In fact, the result obtained in each of the tests is a set of groups or clusters, and the relationship of each city to one of the existing groups by the assignment of a group label. The way in which this grouping is achieved is based on measuring the Euclidean distance between each city and the successive centroids. This is a simple calculation that is reduced to the square root of the sum of the squares of the differences in each variable of the value of the same in each city with respect to the value of the given centroid. This provides a numerical scalar value that gives an overview of the distance between the city and its centroid, as if they were two points.

The number of groups is a parameter of the algorithm, and is provided as an input to it. We have carried out different tests, achieving aggregations using from 2 to 8 different groups. Establishing the optimal number of groups is a question that goes beyond our scope, so we have applied common sense and performed different tests with a fixed number of groups between 2 and 8 (theoretically, this provides a minimum of 8 to 9 cities per group).

By observing the results, we can see that the medium-sized cities of an industrial profile often appear together in the same group, characterized as medium-sized productive cities. The results vary from an initial classification into 2 groups (which would separate industrial and non-industrial cities) to a classification into 8 groups, where more than one of the groups could actually be considered as a group of cities of an industrial profile. Although the IURI has not been used as a variable in the model, we have taken its conclusions into account to focus on the cluster or group considered to be purely industrial in each case.

The following tables, from 12 to 16, summarize for each number of established clusters (from 2 to 8), the result of the K-means algorithm with the cities that appear within the "most industrial" cluster. All the tables provided below have the same structure: each city's ranking or position within the cluster, their corresponding ranking within the IURI in the Chebyshev version, and within the IURI in the PCA version. The reader may wonder how a ranking within each cluster can be provided, but the answer is simple: we can use the information of the distance to the centroid and generate this ranking internally within the cluster. Although IURI has not been used as a model variable, its application has helped us select the most appropriate industrial cluster in each test.

TABLE 10

CLUSTER OF INDUSTRIAL CITIES (K-MEANS WITH 2 CLUSTERS)

Ranking	City	IURI Ranking (Chebychev)	IURI Ranking (PCA)
1	Avilés	3	3
2	Palencia	1	1
3	Vigo	37	44
4	Burgos	2	2
5	Lorca	14	6
6	León	23	19
7	Talavera de la Reina	16	8
8	Gijón	15	16
9	Ourense	27	35
10	Cádiz	33	29
11	Ferrol	30	38
12	Vitoria	4	4
13	Tarragona	12	9
14	Manresa	7	13
15	Linares	24	20
16	Alcoy	8	14

Source: Author's elaboration.

TABLE 11
CLUSTER OF INDUSTRIAL CITIES (K-MEANS WITH 3 CLUSTERS).

Ranking	City	IURI Ranking (Chebychev)	IURI Ranking (PCA)
1	Avilés	3	3
2	Palencia	1	1
3	Logroño	5	5
4	Sagunto	6	7
5	Alcoy	8	14
6	Pamplona	9	15
7	Manresa	7	13
8	Burgos	2	2
9	Vitoria	4	4

Source: Author's elaboration.

TABLE 12
CLUSTER OF INDUSTRIAL CITIES (K-MEANS WITH 4 CLUSTERS)

Ranking	City	IURI Ranking (Chebychev)	IURI Ranking (PCA)
1	Avilés	3	3
2	Palencia	1	1
3	Sagunto	6	7
4	Alcoy	8	14
5	Pamplona	9	15
6	Burgos	2	2
7	Manresa	7	13
8	Vitoria	4	4

Source: Author's elaboration.

TABLE 13
CLUSTER OF INDUSTRIAL CITIES (K-MEANS WITH 5 CLUSTERS)

Ranking	City	IURI Ranking (Chebychev)	IURI Ranking (PCA)
1	Avilés	3	3
2	Palencia	1	1
3	Logroño	5	5
4	Sagunto	6	7
5	Alcoy	8	14
6	Valladolid	10	10
7	Burgos	2	2
8	Pamplona	9	15
9	Manresa	7	13
10	Vitoria	4	4

Source: Author's elaboration.

TABLE 14
CLUSTER OF INDUSTRIAL CITIES (K-MEANS WITH 6 CLUSTERS)

Ranking	City	IURI Ranking (Chebychev)	IURI Ranking (PCA)
1	Avilés	3	3
2	Palencia	1	1
3	Alcoy	8	14
4	Pamplona	9	15
5	Manresa	7	13
6	Burgos	2	2
7	Vitoria	4	4

Source: Author's elaboration.

TABLE 15
CLUSTER OF INDUSTRIAL CITIES (K-MEANS WITH 7 CLUSTERS)

Ranking	City	IURI Ranking (Chebychev)	IURI Ranking (PCA)
1	Avilés	3	3
2	Palencia	1	1
3	Vitoria	4	4
4	Burgos	2	2

Source: Author's elaboration.

TABLE 16
CLUSTER OF INDUSTRIAL CITIES (K-MEANS WITH 8 CLUSTERS)

Ranking	City	IURI Ranking (Chebychev)	IURI Ranking (PCA)
1	Avilés	3	3
2	Palencia	1	1
3	Vitoria	4	4
4	Burgos	2	2

Source: Author's elaboration.

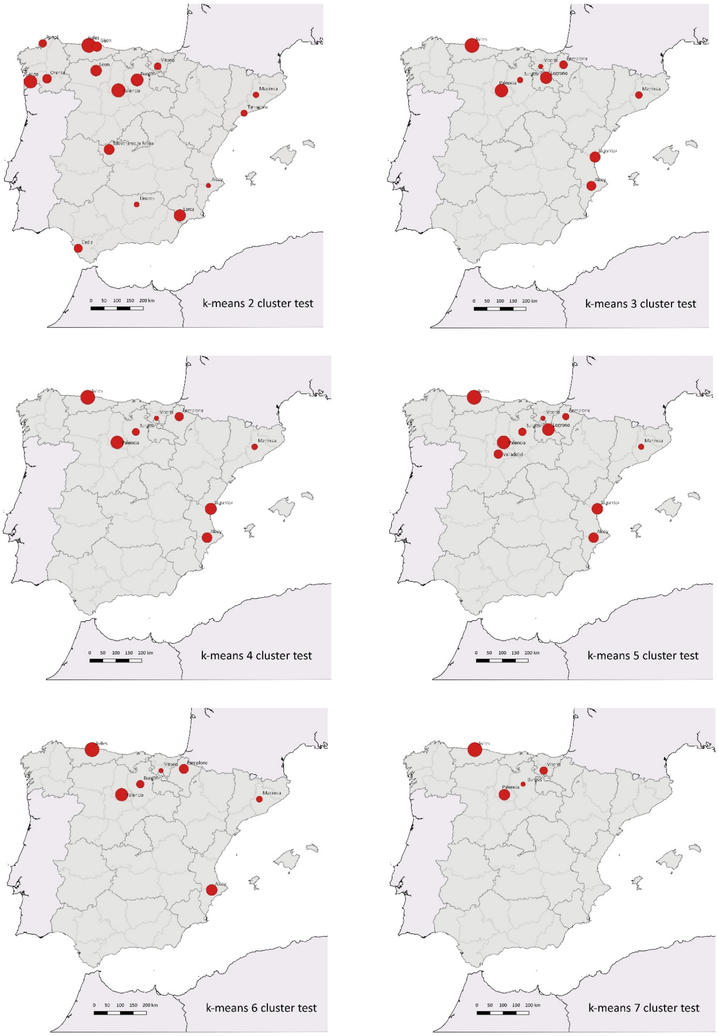
Table 10 summarizes the results for two clusters. We have 6 of the 10 most industrial cities of the IURI in the Chebyshev version and 7 of 10 in the PCA version, although in this first grouping there are cities with a high ranking (Vigo, Cádiz or Ferrol) and others that are moderately industrial (Tarragona, Lorca, Gijón or Talavera de la Reina).

Tables 11, 12 and 13 report the test results for 3, 4 and 5 clusters respectively. We can observe that the results are very similar in both cases, with a cluster of between 8 and 10 cities incorporating the first 8, 9 or 10 industrial cities of the IURI in the Chebyshev version. With regard to the PCA version, there are always 3 or 4 moderately industrial cities.

Tables 14, 15 and 16 show a higher demand in the execution of the algorithm by forcing the grouping as much as possible, since we demand to generate 6, 7 or 8 clusters. This requirement forces the recalculation of the distances and the result is always a small industrial cluster of between 4 to

FIGURE 4

A GRAPHICAL BRIEFING USING K-MEANS TESTS, FROM 2 TO 7 CLUSTER. AS CLUSTER AMOUNT IS INCREASED, GENUINE INDUSTRIAL MEDIUM-SIZE CITIES STAND ALL TOGETHER



Source: Author's elaboration.

7 cities. Particularly in the last two tests, the cluster offers the cities that we understand to be purely industrial, consolidating Avilés, Palencia, Vitoria and Burgos as the 4 main medium-sized industrial cities in the country.

As the number of clusters grows, the dispersion of cities increases, with the most industrial group gradually decreasing in number, so the analysis becomes somewhat distorted. In any case, we see that as the number of groups grows, the cities with the most noticeable industrial profile remain grouped in a strong nucleus and the aforementioned urban areas are shown to be the medium-sized spaces of a more defined productive nature. In other words, these are the four cities in the Spanish urban system with the greatest industrial specialization and in which industrial land, industrial employment and industrial workforce present a greater intensity in the territory. Finally, Figure 4 provides an industrial profile from Tables 10 to 15: the more groups used, the more reliably those cities with a more genuine industrial profile are obtained.

4. CONCLUSIONS

The territorial picture shown by the processing of the data indicates that the Spanish city system has an average industrial city profile with the following basic characteristics:

- A population of around 200,000 inhabitants.
- A relevant volume of industrial employment (between 10,000 and 25,000 employees).
- A very high relative value of the industrial workforce, clearly above the national average and that of the large cities (between 15 and 30% of the total).
- A high degree of land occupied by productive facilities (between 1,000 ha and 2,700 ha), with values that reach a third of the urbanized total and even exceed this percentage –industry occupies between 25% and 45% of the current urban land.

This profile associated with productive activity can be seen in almost 30 of the 50 medium-sized urban areas in Spain, an aspect that shows an intense spatial and socio-economic significance of the industry in this group of cities (two thirds of the cities present relevant industrial activity in their territory). With regard to this, the main conclusion that can be drawn from the treatment of this statistical information is therefore linked to the clarification and identification of the industrial profile of many of these cities. The improvement of the IURI index made in this paper shows that the size of the cities in Spain does not define their industrial specialization. The fact is that medium-sized urban areas occupy the highest positions in the relative ranking on industrial-urban intensity and the grouping of clusters also generates a profile in which medium-sized cities stand out over large urban areas.

In this sense, the revision and improvement of the quality of the IURI as a tool to characterize cities of an industrial profile has proved to be efficient, providing a new method for its calculation. This method does not offer important variations in the ranking, but it does offer more quality in the definition of the model. Both the original version (Chebyshev) and the current version (PCA) allow, by relative comparison, to establish the cities with the most industrial profile (Avilés, Palencia, Vitoria and Burgos). The quality of the index, its reasonable distribution, the strength of the classification method and the degree of sensitivity it can offer to slight or important changes to the variables handled are all confirmed.

After the sensitivity tests carried out in both versions, we can conclude that both ways of characterizing and classifying the medium-sized industrial cities will be fully valid in the future, and we will be able to use them to maintain their conclusions in the face of changes in the values of the variables, given that the possible changes in some of their variables will not alter reality as much as in the cases of the tests carried out.

Finally, in order to reinforce our thesis, and beyond the IURI indicator, we have conducted some clustering tests using the K-means algorithm. This algorithm classifies our cities into groups, using the concept of Euclidean distance, which is calculated with the variables used. It allows us to check if the variables considered help us acquire the profile of the purely industrial cities, observing the groups of cities obtained. We have carried out tests to obtain between 2 and 8 city clusters, and we have compared the results provided by this classification with respect to the IURI ranking in its two versions. The result also identifies the group of 4 or 5 main cities in the scale of industrial intensity in the country, all of them medium-sized urban areas.

The comparison between the versions of IURI and the result of the clustering has enabled us to establish the similarities in the results of both techniques, confirming that in both cases a different approach is taken but aimed at defining the industrial profile of the medium-sized Spanish city. This also corroborates the significant results within these groupings.

This technique makes it possible to analyze the relative position of each urban industrial area and to compare the intensity of productive activity in all the cities of the same urban system. The index, as formulated, can be applied not only to Spain, but also to the study of the urban reality in other countries. This will provide interesting comparative dynamics to consider whether this same behavior of medium-sized cities is occurring in other cases.

ACKNOWLEDGEMENT:

This research has been financed by the Ministry of Science, Innovation and Universities, with the support of European Union ERDF funds, within the 2018 Call for Proposals of the State Program for Research, Development and Innovation Projects Oriented to the Challenges of Society. This is part of the coordinated research project titled “Ciudades medias españolas: urbanización y políticas urbanísticas. 40 años de Ayuntamientos democráticos”, sub-project “Procesos de cambio en la urbanización de las ciudades medias y sus áreas urbanas, 1979-2019” (RTI-2018-096435-B-C22). The opinions, findings, conclusions, recommendations or omissions reflected in the text are the sole responsibility of the authors and do not necessarily reflect the views of the Ministry.

5. REFERENCES

ABDI, H. and WILLIMAS, L. J. (2010): “Principal component analysis”, *Interdisciplinary Reviews*. Computational Statistics, vol. 2, issue 4, 433-459.

ADNAN ALRABEA, A. V. SENTHILKUMAR, HASAN AL-SHALABI, and AHMAD BADER (2013): “Enhancing K-Means Algorithm with Initial Cluster Centers Derived from Data Partitioning along the Data Axis with PCA”, *Journal of Advances in Computer Networks*, vol. 1, no. 2, 137-142.

ANDRECUT, M. (2009). Parallel GPU implementation of iterative PCA algorithms, *Journal of Computational Biology*, 16 (11), 1593-1599.

ANDRÉS LÓPEZ, G. (2008): “Geografía y ciudades medias en España: ¿a la búsqueda de una definición innecesaria?”, *Scripta Nova*, no. 263. Ed. Digital.

ANDRÉS LÓPEZ, G. (2019): “El significado de los espacios de actividad económica en la estructura urbana de las ciudades medias españolas”, *Ciudades. Revista del Instituto Universitario de Urbanística de la Universidad de Valladolid*, no. 22, Universidad de Valladolid, 1-22.

ANDRÉS LÓPEZ, G. (2019): “Las ciudades medias industriales en España. Evolución histórica, proceso de urbanización y estructura urbana”, *Ería. Revista Cuatrimestral de Geografía*, vol. 2019-1, year XXXVIII, Universidad de Oviedo, 25-49.

ANDRÉS LÓPEZ, G. (2020): “Las ciudades medias industriales en España. Caracterización geográfica, clasificación y tipologías”, *Cuadernos Geográficos*, vol. 59, no. 1 (2020), Universidad de Granada, 99-125.

AYYADEVARA, V. (2018): *Pro Machine Learning Algorithms*, Apress, 372 p.

BELLET, C. and LLOP, J. M. (Ed.) (2000): *Ciudades intermedias: urbanización y sostenibilidad*, Editorial Milenio, Lérida, 560 p.

BELLET, C. and LLOP, J.M. (2004): “Ciudades intermedias: entre territorios concretos y ciudades y espacios globales”, *Ciudad y Territorio. Estudios Territoriales*, XXXVI (141-142), 2004, 569-582.

BOSSARD, M., FERANEC, J. and OTAHEL, J. (2000). *CORINE Land Cover technical guide. Addendum 2000*. Technical Report no. 40, EuroEnvironment Agency, Copenhague, 105 p.

CAIMING ZHONG, XIAODONG YUE, ZEHUA ZHANG and JINGSHENG LEI. (2015): “A clustering ensemble: Two-level-refined co-association matrix withpath-based transformation”, *Pattern Recognition*, vol. 48, issue 8, 2699-2709.

CAMPOS SÁNCHEZ, F. S. (2017): “Ciudades medias españolas siglo XXI. Revisión y análisis bibliométrico de enfoques y temáticas predominantes”, *Cuadernos Geográficos*, no. 56(1), 217-241.

CAPO VICEDO, J. et al. (2010): “Análisis de las publicaciones sobre distritos industriales y clusters en las revistas españolas de geografía“,*Revista de Estudios Regionales*, no. 89, 91-114.

CATALÁ MATEO, R., BOSQUE SENDRA, J. and PLATA ROCHAS, W. (2008): “Análisis de posibles errores en la base de datos Corine Land Cover (1990-2000) en la Comunidad de Madrid”, *Estudios Geográficos*, no. LXIX, 81-104.

CEBRIÁN ABELLÁN, F. and PANADERO MOYA, M. (Coord.) (2013): *Ciudades medias. Formas de expansión urbana*, Ed. Biblioteca Nueva, 246 p.

DÍAZ-PACHECO, J. and GUTIÉRREZ PUEBLA, J. (2014): “Exploring the limitations of Corine Land Cover for monitoring urban land-use dynamics in metropolitan areas”, *Journal of Land Use Science*, no. 9(3), 243-259.

DING, H., HE, X. (2004): "K-means Clustering via Principal Component Analysis", ICML '04: *Proceedings of the twenty-first international conference on Machine learning*.

DRAGOMIR, S.S. (2019): "On some Chebyshev type inequalities for the complex integral", *Integración - UIS* [online]. 2019, vol. 37, no.2, 307-317.

ESPEJO BENITEZ, J.M. (2011): "Un indicador de competitividad para las provincias españolas ", *Revista de Estudios Regionales*, no. 92, 43-84.

ESPINO (del) HIDALGO, B. (2017): "Las ciudades medias del centro de Andalucía. Aproximación a un fenómeno territorial y urbano", *Revista de Estudios Regionales*, no. 108, 165-191.

EUROSTAT DATABASE (2020): *Electronic source*. Available at: (https://ec.europa.eu/eurostat/data/database, "Functional Urban Areas").

FERNÁNDEZ CUESTA, G. and FERNÁNDEZ PRIETO, J.R. (1999a): *Atlas industrial de España. Desequilibrios territoriales y localización de la industria*, Ed. Nobel, 205 p.

FERNÁNDEZ CUESTA, G. and FERNÁNDEZ PRIETO, J.R. (1999b): "La distribución de la industria en España. Pautas regionales y cambios recientes", *Ería*, no. 49, 129-158.

GANAU CASAS, J. and VILAGRASA IBARZ, J. (2003): "Ciudades medias en España: posición en la red urbana y procesos urbanos recientes", *Mediterráneo Económico*, no. 3, 37-73.

GARCÍA MARTÍN, F. M. (2014): "¿Ciudades intermedias? La conformación de un nuevo tipo de ciudades en España a lo largo del siglo XX", *VII Congreso Internacional de Ordenación del Territorio*, Fundicot, Madrid, 1090-1104.

GARCÍA MARTÍN, F. M. (2016): "Compacidad y densidad de las ciudades españolas", Eure, vol. 42, no. 127, 5-27.

GARCÍA, J.A., MARTÍNEZ, J.M. and ESCUDERO, L.A. (2018): "La importancia de las ciudades medias en el Sistema urbano de España", Cebrián Abellán, F. (Coord.): *Ciudades medias y áreas metropolitanas. De la dispersión a la regeneración*, Servicio de Publicaciones Universidad de Castilla La Mancha, Cuenca, 25-41.

GOERLICH GISBERT, F. J. and MAS IVARS, M. (2008): *Sobre el tamaño de las ciudades en España. Dos reflexiones y una regularidad empírica*, BBVA, Documentos, no. 6, 68 p.

GOODALL, T., GIBSON, S. and SMITH, M. C. (2012): "Parallelizing principal component analysis for robust facial recognition using cuda," *Application Accelerators in High Performance Computing (SAAHPC)*, 2012 Symposium, 121–124.

HERNÁNDEZ ORALLO, J., FERRI RAMIREZ, C. y RAMIREZ QUINTANA, M. J. (2004): *Introducción a la Minería de Datos,* Ed. Pearson, 680 p.

HOLANDA FILHO, R. and EVERARDO BESSA MAIA, J. (2010): "Network Traffic Prediction using PCA and K-means", *IEEE Network Operations and Management Symposium*.

HUANG, Z. (1998): "Extensions to the k-Means Algorithm for Clustering Large Data Sets with Categorical Values", *Data Mining and Knowledge Discovery*, 2, 283–304.

JAIN, AK. and DUBES, RC. (1988): *Algorithms for clustering data*, Prentice-Hall, Englewood Cliffs, 320 p.

MÉNDEZ GUTIÉRREZ DEL VALLE, R. and PASCUAL RUIZ VALDEPEÑAS, H. (2006): *Industria y ciudad en España. Nuevas realidades, nuevos retos*, Ed. Thomson-Civitas, Navarra, 618 p.

MIRAMONTES CARBALLADA, A. and VIERA DE SÁ MARQUES, T. M. (2016): "Las áreas urbanas en la península ibérica. Un ejercicio de delimitación", *Papeles de Geografía*, 62, 47-63.

MIRET PASTOR, L. et al. (2011): "Identificación de sectores de servicios y de alta tecnología en la Comunidad Valenciana: ¿Un nuevo cluster mapping?", *Revista de Estudios Regionales*, 90, 71-96.

MOLINA MORALES, F.X. et al. (2012): “La heterogeneidad dimensional de los distritos industriales. Un estudio longitudinal del caso español”, *Revista de Estudios Regionales*, 93, 43-63.

NADAL, J. (Dir.) (2003): *Atlas de la industrialización de España*, 1750-2000. Crítica. Fundación BBVA, 664 p.

OLAZABAL, E. and BELLET, C. (2017): “Análisis de las nuevas dinámicas de urbanización en España. Su estudio a través del uso de Corine Land Cover y SIOSE”, *Actas del XXV Congreso de la Asociación de Geógrafos Españoles*, Madrid, AGE, 2639-2648.

OLAZABAL, E. and BELLET, C. (2018): “Procesos de urbanización y artificialización del suelo en las aglomeraciones urbanas españolas (1987-2011)”, *Cuadernos Geográficos*, 57, (2), 189-210.

ORDONEZ, C., MOHANAM, N. and GARCIA-ALVARADO, C. (2014): “PCA for large data sets with parallel data summarization”, *Distrib Parallel Databases*, 32, 377–403.

PRECEDO LEDO, A., and MÍGUEZ IGLESIAS, A. (2018): “Los efectos de la crisis en el posicionamiento de las ciudades españolas”, *Boletín de la Asociación de Geógrafos Españoles*, AGE, Madrid, 76, 79-101.

ROCA CLADERA, J., MOIX BERGADÁ, M. and ARELLANO RAMOS, B. (2012): “El sistema urbano en España”, *Scripta Nova. Revista Electrónica de Geografía y Ciencias Sociales*, Vol. XVI, no. 396, Ed. Digital.

SALA RÍOS, M. et al. (2011): “Fluctuaciones cíclicas, shocks y asimetrías. Un análisis desagregado para las regiones e industrias españolas “, *Revista de Estudios Regionales*, 91, 97-123

SÁNCHEZ MORAL, S. et al. (2009): “Dinámicas de las ciudades de tamaño intermedio en el sistema urbano español: entre el declive y la recuperación”, Pillet Capdepon, F. y otros (Dir): *Geografía, territorio y paisaje: el estado de la cuestión*, Univ. Castilla La Mancha, Ciudad Real, 655-670.

SMITH, L. I. (2002) *A tutorial on principal components analysis. Electronic source*. Available at: http://www.iro.umontreal.ca/~pift6080/H09/documents/papers/pca_tutorial.pdf

VVAA. (2020): *Scikit learn decomposition PCA. Electronic source*. Available at: https://scikit-learn.org/stable/modules/generated/sklearn.decomposition.PCA.html

WU, X., KUMAR, V., QUINLAN, J. R., GHOSH, J., YANG, Q., MOTODA, H. and ZHOU, Z. H. (2008). “Top 10 algorithms in data mining”, *Knowledge and information systems*, 14(1), 1-37.

YINGYU LIANG, MARIA-FLORINA BALCAN and VANDANA KANCHANAPALLY (2013): “Distributed PCA and k-Means Clustering”, *Computer Science*, 1-8.