

City typologies of Europe: a tool to support urban sustainability studies and practices

E. Aksoy¹, M. Gregor², J. Fons³, C. Garzillo⁴, M. Cugny-Seguin⁵,
M. Löhnertz² & C. Schröder¹

¹*European Topic Center of EEA – Urban, Land, Soil (ETC-ULS),
University of Malaga, Spain*

²*ETC-ULS, space4environment, Luxembourg*

³*ETC-ULS, Universitat Autònoma de Barcelona, Spain*

⁴*ETC-ULS, ICLEI European Secretariat, Germany*

⁵*European Environment Agency (EEA), Denmark*

Abstract

Cities and urban areas in Europe are very heterogeneous because of their geographic situations, climate, heritage and trajectory, activities, governance and urban management, population, etc. In order to better understand European urban systems despite the diversity of urban characteristics across Europe and taking into account its complexity, the European Environment Agency (EEA) and its European Topic Centre on Urban, Land and Soil systems (ETC-ULS) are developing a methodology to implement a comprehensive and integrated analysis of sustainability of a large number of very diverse cities. This paper reveals the first results of the on-going project on “City Typologies in Europe”.

With this aim, the concept of a city typology was elaborated and tested using existing and relevant data on the urban system at European level. 385 European cities were classified to identify different clusters which are derived from 59 indicators under different domains (economic, social and environmental) that reflect the major city characteristics.

The objective of this typology is to show the differences and similarities of the huge diversity of European cities by synthesizing available and comparable information. By reclassifying cities into groups, representative cities (role-models) will be identified for each cluster. Since cities are the key players in minimizing the use of resources and developing the circular economic model and are also crucial for the quality of life and well-being of cities’ residents, analyzing those

differences and similarities is very important to understand the real situations of the European cities, to make foresight scenarios, to develop effective tools of analysis of urban sustainability and to suggest a core set of urban sustainability indicators. Moreover, to achieve resource and energy efficient cities and sustainability, these typologies might be very helpful for the decision and policy makers to overcome the limitations of policy instruments that are insufficient to deal with the complexity of urban challenges. This study also might be helpful for local authorities to find the appropriate solutions for an effective, sustainable pathway which needs to take into account the similar characteristics of the cities as a group instead as individuals.

Keywords: city typologies, European cities, clustering, urban sustainability, city indicators.

1 Introduction

Cities and, more generally urban areas, are growing very fast. The urban areas are supposed to absorb all the population growth expected over the next decades. Europe already became predominantly urban in the beginning of the 1950s. Today, approximately 359 million of Europeans – 72% of the total EU population (Eurostat [1]) – live in cities, towns and suburbs. And this share will continue to increase.

Cities and urban areas in Europe are very heterogeneous because they are located in different geographic situations (littoral, mountain, island, fluvial corridor, etc.), have different climate, heritage (morphology, size, age of housing, spatial segregation, etc.) and trajectory (shrinking cities, sprawl, etc.), activities (industry, tourism, etc.), governance, urban management, population (demography, ageing), etc. These differences make it more complex both to define problems and to respond to them. Therefore, sustainability goals have to be defined and related indicators have to be assessed at a relevant level to take into account the ‘real city’. By consequence, it cannot be at the core city level, which is too small and too administratively defined, but more at the level of the functional urban area. The city typology gives a simplified image of the huge diversity of cities (size, development, demography, environmental profile, etc.) that eases the analysis.

Each city has a unique typology profile and its sustainable development should therefore follow an individual development pathway. Nevertheless, groups of cities can be discerned to characterize their common but differentiated policy challenges. The main objective of the city typology is to provide information on groups of cities that share the same or very similar characteristics; it is not conceived as a benchmarking tool. Furthermore, it can also be helpful identifying common policy needs for a specific group of cities, regional, national and EU level.

There is a need to find group characteristics that are responsible for the similarities and differences; and possible ways to influence the different development paths of cities towards sustainability. An important factor in understanding such differences is a city typology. A city typology characterizes a



generally recognizable sustainability feature of a group of cities that has far-reaching consequences for the scores on a number of sustainability indicators. So benchmarking of cities would be more meaningful when the benchmark is chosen within a similar type of cities than within the total group of cities studied (Zoeteman *et al.* [2]). A classic benchmarking is based on the assumption that all cities are equally comparable what does not reflect the reality; some cities get a low score that is rather related to the history and the localisation of the city than to the current urban management. For example, Northern European cities cannot be assessed in the same group than Balkan cities because there is a huge gap between the trajectories of these cities. Therefore, a clustering of the cities in the framework of our city typology according to their similarities (in all terms: ecological, economic, environmental and social characteristics) will make a comparison more feasible.

Some environmental data concerning the urban system exist, such as noise, air quality, waste water management or soil sealing. For certain others cases, proxy data can be used or computed. A set of indicators on urban sustainability can be achieved by developing specific indicators (green infrastructure, urban sprawl, etc.), by re-using data (parameters) that already exist (in particular for air quality, noise, water and vulnerability), and by gathering scattered data (including from research project) to fill the open gaps. The work consists more in giving sense to all the available but dispersed pieces of information than create new information (Gregor *et al.* [3]).

The aim is to provide synthesized information in an easy-to-understand way in order to facilitate the communication of important key messages. A typology of cities will be developed that is based on an integrated analysis of a number of parameters and takes into account the huge diversity of cities (size, development, demography, environmental profile, etc.). It is foreseen that each class of the typology, composed of a group of cities, will be followed in the future in order to analyse the trajectory of the group and not the individual trajectory of each urban area. The expected final outcome should be a typology to support the assessment of urban sustainability from a primarily environmental perspective, but also taking into account socio-economic and governance parameters.

2 Material and method

2.1 Material

Finding and gathering useful and compatible data on the city level by covering the highest possible number of cities in Europe was a really hard task. It has been achieved to build a database from different data sources for 385 European cities with 59 parameters under different domains (economic, social and environmental) which reflect the major city characteristics in Europe. This database was used as the main dataset in this study.

All of the data and references used for the study are listed in Table 1. The sources of the database were mainly Eurostat and EEA, as well as National statistics for filling the gaps in the database.



Table 1: Indicator list.

Indicator Name	Unit	Ref. Year	Data Source	
			Static indicators	
S01_01: Administrative area	km ²	2004		LAU layer
S01_02: Area covered with buildings and infrastructure	%	2006		Urban Atlas
S01_03: Green urban areas	%	2006		Urban Atlas
S01_06: Degree of soil sealing	%	2009		HRL Soil Sealing
S01_15_1: Built-up area inside UMZ (PBA)	%	2009		ETC-UMA
S01_15_2: Built-up area outside UMZ (PBA)	%	2009		ETC-UMA
S01_15_3: Difference of Built-up area (PBA)	%	2009		ETC-UMA
S01_16_1: Dispersion inside UMZ (DIS)	UPU/m ²	2009		ETC-UMA
S01_16_2: Dispersion outside UMZ (DIS)	UPU/m ²	2009		ETC-UMA
S01_16_3: Difference of dispersion (DIS)	%	2009		ETC-UMA
S01_17_1: Urban permeation inside UMZ (UP)	UPU/m ²	2009		ETC-UMA
S01_17_2: Urban permeation outside (UP)	UPU/m ²	2009		ETC-UMA
S01_17_3: Difference of urban permeation (UP)	%	2009		ETC-UMA
S01_18_1: Land uptake per person inside UMZ (LUP)	m ²	2009		ETC-UMA
S01_18_2: Land uptake per person outside UMZ (LUP)	m ²	2009		ETC-UMA
S01_18_3: Difference of land uptake per person (LUP)	%	2009		ETC-UMA
S01_19: Distribution of green urban areas	m/ha	2006		Urban Atlas
S01_20: Effective green infrastructure (urban hinterland)	%	2006		ETC-UMA (GI Task, based on Urban Atlas data)
S01_21: Hotspot ratio (hinterland)	%	2006		ETC-UMA (GI Task, based on Urban Atlas data)

Table 1: Continued.

Indicator Name	Unit	Ref. Year	Data Source
Static indicators			
S01_22: Urban blue areas	%	2006	Urban Atlas
S02_01: Compactness	%	2006	ETC-SIA (based on Urban Audit)
S02_02: Highly sealed areas	%	2006	ETC-SIA
S02_03: Lowly sealed areas	%	2006	ETC-SIA
S03_01: Air temperature	°C	2011	Urban Audit
S03_03: Precipitation	mm	2011	Urban Audit
S05_01: No. of inhabitants	persons	2011	Urban Audit
S05_02: Population density	persons/km ²	2011	ETC-SIA (based on Urban Audit)
S05_03: Age structure	%		Urban Audit
S05_05: Old-age dependency ratio	%	2012	Urban Audit
S05_07: Employment: rate of unemployed	rate		Urban Audit
S05_08: Tourism: Nights spent	%	2011	Urban Audit (+ local/national data)
S05_09: Industrial facilities	# of the ind. fac.	2007	Urban Audit
S05_10: Transnational firms	no. of firms	2006-2009	Urban Audit
S05_11: National population	%	2013	Urban Audit
S05_15: Students in higher education	no. of students	2004-2012	Urban Audit (+ local/national data)
S07_01: Lights in the night			ETC-SIA
S08_01: Waste generation per capita	T/person	2007-2009	Urban Audit
S10_01: Length of transport network	km ²	2006	Eurostat: Networks
S10_02: City-pair contactability	# of the cities	2009	ESPON
S12_04: WGI (Worldwide Governance Indicators)	Index	2012	Worldbank
S13_02: Altitude	m	2004	ETC-SIA

Table 1: Continued.

Indicator Name	Unit	Ref. Year	Data Source
Dynamic indicators			
D01_04: Changes of S1-6		2006	ETC-UMA
D01_15_1: Change of built-up area inside UMZ (PBA-DIF)	%	2006–2009	ETC-UMA
D01_15_2: Change of built-up area outside UMZ (PBA-DIF)	%	2006–2010	ETC-UMA
D01_16_1: Change of dispersion inside UMZ (DIS-DIF)	UPU/m ²	2006–2011	ETC-UMA
D01_16_2: Change of dispersion outside UMZ (DIS-DIF)	UPU/m ²	2006–2012	ETC-UMA
D01_17_1: Change of urban permeation inside UMZ (UP-DIF)	UPU/m ²	2006–2013	ETC-UMA
D01_17_2: Change of urban permeation outside UMZ (UP-DIF)	UPU/m ²	2006–2014	ETC-UMA
D01_18_1: Change of land uptake per person inside UMZ (LUP-DIF)	m ²	2006–2015	ETC-UMA
D01_18_2: Change of land uptake per person outside UMZ (LUP-DIF)	m ²	2006–2016	ETC-UMA
D02_01: Changes in S2-1		2000–2006	ETC-SIA
D05_01: No. of inhabitants, growth and shrinkage	%		Urban Audit
D05_02: Population density	%		Urban Audit
D05_03: Age structure	Absolute change		Urban Audit
D12_04: Changes in S12-4 (2002–2012)	Index	2002–2012	Worldbank

2.2 Data preparation and processing

It is important to pre-process the data because of data-inherent noise, errors, inconsistencies, outliers and lack of variable values. Different data pre-processing techniques like cleaning method, outlier detection, data integration and transformation can be carried out before the clustering process to achieve a successful analysis. It is also known that outlier detection and removal by means of a normalization approach improve the effectiveness and performance of the K-Means clustering algorithm (Patel and Mehta [4]). Therefore, the steps listed below were followed before executing the models.

- Dealing with missing values/gap filling;
- Detecting outliers;
- Normalization process; and
- Finding the optimum number of clusters.

As soon as data existed for the majority of the cities, gap filling procedures making use of either external sources or mathematical approaches were put in place. Several different methods have been applied to fill the gaps in the database for certain indicators, such as:

- Looking for a local/country statistics and websites;
- Looking for literature if there is data or for the methodology to calculate;
- Extracting the missing data for the cities from NUTS3 or NUTS0 levels;
- Taking the proportional calculation with the population and the numbers of the certain indicator for the city;
- Taking the country averages (if the data exist for more than half of the cities of the country, etc.).

Most of the studies from the literature (Patel and Mehta [4]; Visalakshi and Thangavel [5]; Pham *et al.* [6]; Ding [7]) show that normalizing is an essential pre-processing step to standardize values of all attributes or features from different dynamic range into a specified range and leads to obtain better quality clusters. Data normalization is the linear transformation of data to a specific range. Therefore, it is worthwhile to enhance the clustering quality by normalizing the dynamic range of input data objects into a specific range (Visalakshi and Thangavel [5]). Z-score normalization technique was applied by using standard deviation and the mean for the better comparability of the indicators in K-Means clustering.

The K-means algorithm requires the number of clusters in the data to be pre-specified. Finding the appropriate number of clusters for a given data set is generally a trial-and-error process made more difficult by the subjective nature of deciding what constitutes 'correct' clustering. Several techniques, as listed below, have been tried to decide on the optimum number of clusters:

- Rule of thumb;
- The elbow method;
- Silhouette plot;
- Principal component analysis (PCA).



2.3 Clustering method

K-means clustering was used to find the cities typologies in this study. K-means is an iterative algorithm. The clusters are first formed arbitrarily by choosing the first K cases, assigning each subsequent case to the nearest of the K, then calculating the centroids of each cluster. Subsequently, each case is tested to see whether the centre of another cluster is closer than the centre of its own cluster; if so, the case is reassigned. If cases are reassigned, the centroids are recalculated and the algorithm repeats. Since the algorithm is significantly sensitive to the initial randomly selected cluster centres; working iteratively (running the algorithm multiple times) helps reduce this effect (Aksoy [8]).

3 Results

Since several techniques suggested choosing between 7 and 14 clusters as the most appropriate number of clusters, it was decided to make the analysis for 10 classes. Each of these 10 clusters has the following number of memberships:

- Cluster 1: 35 cities
- Cluster 2: 12 cities
- Cluster 3: 90 cities
- Cluster 4: 5 cities
- Cluster 5: 56 cities
- Cluster 6: 55 cities
- Cluster 7: 83 cities
- Cluster 8: 1 city (Paris)
- Cluster 9: 1 city (London)
- Cluster 10: 47 cities

All of the 59 indicators given in Table 1 were clustered by applying K-Means clustering method to find the different typologies of the 358 European cities. The output of the K-Means clustering analysis and the distribution of the different cluster memberships can be seen in Figure 1. Significant characteristics of the each of the clusters can be seen from the spider diagrams given in Figure 2. Each of those 59 indicators is represented with one of the value axis in the spider diagrams for each of the cluster. Therefore, highest positive and the lowest negative values can be interpreted as the most significant characteristics each of the clusters and they can easily be observed from Figure 2.

According to the results, 35 cities in cluster 1 mainly were shaped by the high rate of unemployment and high rates of dispersion outside and inside of the urban morphological zones (UMZ). 33 cities out of 35 are Spanish cities.

Cluster 2 was formatted mainly the influences by the large administrative areas; high share of green urban areas; low values in almost all other indicators/parameters of the urban dimension domain (Area covered with

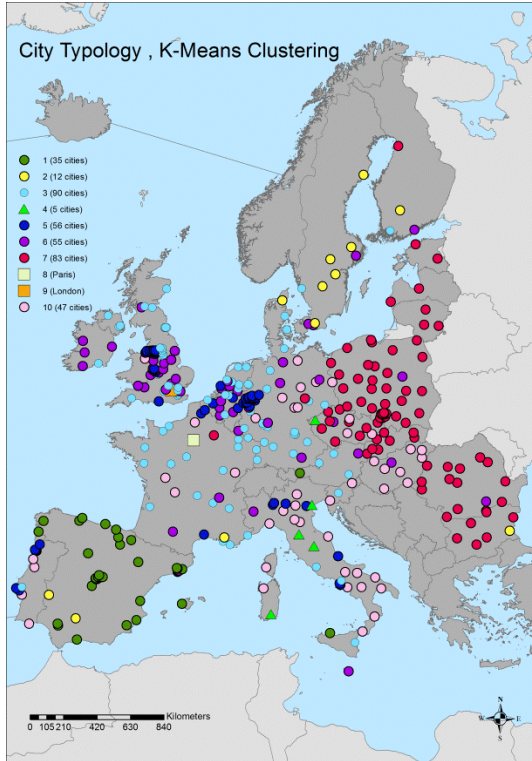


Figure 1: K-Means clustering result map.

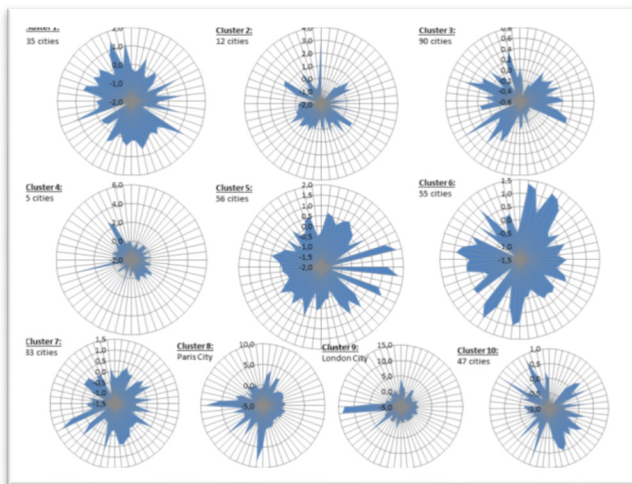


Figure 2: Spider diagrams of the cluster characteristics.

buildings and infrastructure; degree of soil sealing; changes of soil sealing degree; compactness and high density areas); Low values for most of the urban sprawl indicators inside and outside of the UMZ (“urban permeation”, “built-up area” and “dispersion”). 8 out of 12 cities are Scandinavian in cluster 2.

The largest cluster was found as cluster 3 therefore it is very heterogeneous. There is no outstanding significant indicator for this cluster. Many of the cities are medium-sized cities with a tendency of residential development towards the outside of the city.

Cluster 4 is the smallest cluster and mainly characterized by high number of nights spent on tourism; high share of urban blue areas and high rate for old age dependency. 4 of 5 cities (Firenze, Venezia, Perugia, Cagliari) are in Italy except Karlovy Vary.

Cluster 5 was mostly shaped by the urban sprawl indicators; “Urban permeation”, “built-up area” and “dispersion” for inside and outside of UMZ is high. Both, lots of old typically industrial cities/regions (German Ruhrgebiet, northern Italy, UK) from many different countries, and a high number of Portuguese cities were found all in this cluster together.

Cluster 6 contains a substantial number of the European capitals, such as Wien, Brussels, Praha, Berlin, Copenhagen, Budapest, Dublin, Luxemburg, Amsterdam, Warszawa, Bucharest, and Stockholm, with only 4 Eastern and 5 Mediterranean cities. Often the cities are relatively wealthy and business centres of regional and national importance. This also includes a higher percentage of foreign population. The most significant variables of this cluster were found as high values in the urban dimension (degree of soil sealing, changes of soil sealing degree, compactness, high density areas, etc.) and very low rate of national population.

Cluster 7 mainly consists of small and medium sized eastern European cities. Most indicative parameters were found as a high proportion of national population; low WGI, but very high change rate. In general, those cities have a low share of foreigners, are presumably structurally weak with a relatively high unemployment rate, and have a low attraction for tourists and investors.

Cluster 8 only consists of Paris, so there are no real cluster specificities, but the spider diagram describes the characteristics of Paris. Similarly, cluster 9 represents only London, as one the two megacities in Europe besides Paris. So again the spider diagram describes the characteristics of one city.

Cluster 10 consists of mostly green, small sized, old-age citizen cities with the high share of green urban areas; and high rate of old-age dependency.

4 Discussion

The major difficulty of this approach is to find comparable and relevant data for the same time period. We must admit that this approach is more driven by data availability rather than by strict application of the urban sustainability concept. However, given the large number of data and covered domains, we can assume that this approach can be assimilated to a coarse analysis of urban sustainability. Since this paper presented the first results of the on-going project on “City Typologies in Europe”, we will be able improve our database by adding more data



(water abstraction of cities, air quality indicators, risk of poverty, etc.) with the aim of filling the missing domains and dynamics of the cities.

The number of cities per cluster varies substantially and ranges from 1 city in two clusters (Paris and London) up to 83 cities in the second-largest and 90 cities in the largest cluster. The cluster analysis confirms the outstanding characteristics of the two agglomerations Paris and London, falling in separate one-city clusters. This also confirms the validity of the analytical approach since these two global cities do not fit in any of the other clusters in our data-driven approach. It is however interesting to observe that they do not fall into the same cluster, but in two separate ones.

Robust comparisons of groups of urban areas across Europe are lacking and the poor knowledge on urban dynamics has important consequences on policy-making. The distinct identity of the city typology concept is linked to the creation and development of selected clusters of cities. We are not seeking to define individual trajectories of each urban area. Our task is to examine the trajectory of the group/cluster of cities. In addition, by reclassifying cities into groups, we are interested in identifying representative cities (role-models) for each cluster, defining its characteristics as well as the ranges of the cluster.

Finally, the scope of the city typology is to use its broad-base data brokerage capability, and convert it into useful information that builds knowledge and understanding of where cities stand today, and is crucial for recalibrating sustainability policies and improving decision making. We should not ignore the dynamic and time-evolving characteristics of cities and remind ourselves that patterns of urbanisation in different parts of Europe follow different forces. Cities change over time in a manner that cannot be simply labelled; for example, some can be positioned in ‘transition zones’ between two clusters. Transformation of urban areas and their independent cores, evolution over time, emergence of unexpected new issues and naturally occurring patterns open up new governance challenges and permit shifting from one typology to another. Also urban policies have an impact on the form and speed of urban development. By this we mean that a city that successfully tackles a challenge can raise or change its profile accordingly – such changes can consequently be monitored. To an extent, the city typology offers this additional opportunity as quite often rankings and benchmarks say little about the efforts actually made by the cities.

References

- [1] Eurostat, Union Européenne and Commission Européenne, Eurostat regional yearbook 2013, Publications Office of the European Union, Luxembourg, 2013.
- [2] Zoeteman, K., Zande, M., & Smeets, R., Integrated Sustainability Monitoring of 58 EU Cities. Draft Report of Telos Project, Tilburg University, The Netherlands, March 2015.
- [3] Gregor, M., Schröder, C., Aksoy, E., Fons, J., Sainz, M., Storch, A., Thielen, P., & Schieder, W., City typology. ETC-SIA Technical Report, November 2014.



- [4] Patel, V.R. & Mehta, R.G., Impact of Outlier removal and normalization approach in modified k-means clustering algorithm. *IJCSI International Journal of Computer Science Issues*, Vol. 8, Issue 5, No. 2, September 2011.
- [5] Visalakshi, N.K. & Thangavel, K., Impact of normalization in distributed K-Means clustering. *International Journal of Soft Computing* 4 (4): 168-172, 2009. ISSN:1816-9503, Medwell Journals, 2009.
- [6] Pham, D.T., Dimov, S.S., & Nguyen, C.D. Selection of K in K-means clustering. *Proc. IMechE Vol. 219 Part C: J. Mechanical Engineering Science*. DOI: 10.1243/095440605X8298, 2004.
- [7] Ding, C. K-means clustering via Principal Component Analysis. *Proceedings of the 21st International Conference on Machine Learning, Banff, Canada, 2004*.
- [8] Aksoy, E., *Classifying Turkish District Data with K-Means and SOM Algorithms Book*, VDM Publication, ISBN: 978-3-639-13560-2 Germany, 2009.

