University of Naples Federico II

Department of Civil, Architectural and Environmental Engineering

PhD in Civil Systems Engineering

XXIX cycle

Cities and energy consumption: how to reduce CO_2 emissions and address climate change

Candidate: Ing. Laura Russo

Tutor: prof. arch. Carmela Gargiulo

February 2017

A mio padre

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

SECTION 1. EXECUTIVE SUMMARY

1.1 Executive summary

According to IEA (2016), urban areas consume about two-thirds of primary energy demand and produce over 70 per cent of global carbon dioxide emissions (CO_2). Consequently "cities are the heart of the decarbonisation effort" (IEA, 2016) and can be the solution to climate change.

In order to support local policy makers' decisions and foster the transition towards a low-carbon future, a growing body of international researchers has been studying the complex and multidimensional relationship between cities and energy consumption. Urban planning policies, indeed, can effectively improve energy saving in cities and reduce urban emissions only if the interactions between urban factors and energy use are investigated and are found to be significant. However, despite the great interest of the literature for this topic, a consistent number of interactions between urban features and energy use at urban scale still lacks consensus.

Therefore, this research aimed to investigate the relationship between cities and energy consumption to identify the urban factors that significantly affect a city's energy and carbon footprint, thus supporting policy-makers in the definition of effective strategies and policies that can be implemented at an urban scale to reduce energy consumption and resulting CO_2 emissions.

By using a holistic approach rather than a sectoral one, this work considered together a comprehensive set of urban variables – grouped into four categories according to the general system theory (i.e. physical, functional, geographical, and socio-economic variables) – and energy variables. The selection of variables was based on a critical review of the recent interdisciplinary scientific literature on the relationship between cities and energy consumption. The review allowed the definition of a theoretical framework that presented the main urban factors influencing the energy and carbon footprint of a city according to the scientific community and described the key relationships between these features and energy consumption. Furthermore, the theoretical framework also illustrated those relationships amongst the different urban features, which may significantly affect energy consumption but are often ignored by the scientific literature.

Based on this framework, a set of eighteen urban variables and five energy variables was selected and included in the model, which was developed and calculated for a sample of seventy-three Italian capital cities, uniformly distributed across the country.

After an intensive data collection procedure, the dataset was explored and analyzed using different statistical methods each of which provided useful insights into the complex relationship between cities and their carbon footprint. In particular, an exploratory data analysis (EDA) was performed in order to identify potential outliers and evaluate the distribution of data, thus gaining a better knowledge of the research dataset and sample. After EDA, the data were analyzed using a correlation analysis, which provided two main results: (1) it enabled the measurement of the association between the eighteen urban variables in order to identify redundant information and, most importantly, significant interconnections amongst these factors; (2) it showed the significance of the linear relationship between each individual urban variable and CO_2 emissions by sector. Later, three regression models (OLS) were estimated in order to measure the direct relationships between urban and energy features. In these three models, the dependent variables are three of the five categories of CO_2 emissions – residential, transport and total – and the eleven independent variables are housing density, house material, green areas, concentration of manufacturing activities, concentration of commercial activities, concentration of touristic activities, degree-days, topography, income, car ownership and household composition. Finally, a multivariate statistical analysis was performed in order to identify groups of cities with similar urban characteristics and compare their energy behaviors.

When considering both direct and indirect effects (i.e. the results of regression and correlation analyses respectively), all the four groups of urban variables affect total CO_2 emissions per capita. More specifically, three physical variables (i.e. housing density, house material and green areas), one functional variable (i.e. the concentration of commercial activities), and two geographical variables (i.e. degree-days and topography) have a direct effect on CO_2 emissions: lower density of dwelling units and lower air temperatures, as well as valley topography and higher concentrations of masonry buildings, green areas and commercial activities increase CO_2 emissions per capita. On the other hand, three socio economic variables (i.e. income, education and ethnicity) and one functional variable (i.e. land-use mix) indirectly affect CO_2 emissions through the mediators of other urban features. In particular, a higher level of education and a higher share of foreign residents are both associated with higher income that, in turn, is associated with a higher land-use mix that corresponds to higher housing density, which reduces CO_2 emissions per capita.

Therefore, two main policy implications are drawn from the results of the correlation and regression analysis; one at the building scale and one at the urban scale. (a) At the building scale, interventions should focus on buildings materials, especially for reducing the energy use of masonry buildings. (2) At the urban scale, planning strategies should encourage compact developments in order to reduce energy consumption and total CO_2 emissions. Moreover, besides the lower energy footprint of compact cities, in Italy, higher densities of housing units correspond to higher densities of jobs, which in turn are characterized by higher incomes, and therefore strategies for promoting urban compactness can also have positive economic effects.

Furthermore, the results of the cluster analysis corroborate the findings of the correlation and regression analysis and provide additional insights about the sample of Italian cities considered in this research. The cluster analysis, indeed, shows that Italian colder-valley-inland-wealthier cities – such as Torino, Bolzano, Padova, Mantova, etc. – produce higher level of both residential and total CO_2 emissions, and are mainly located in the northern part of the country. On the contrary, Italian cities by the sea, with warmer climate and densely urbanized – such as Genova, Salerno, Bari, etc. – emit less CO_2 per capita, thus being more energy efficient than the others.

The results of this research substantiate the complexity and multidimensionality of the relationship between cities and energy consumption and the strategic role of both building and urban interventions for energy saving (Zanon & Verones, 2013). Furthermore, these results, which only partially support previous findings, suggest that important trade-offs exist between the different urban characteristics and cities' energy consumption and CO_2 emissions (Doherty et al., 2009; Lee & Lee, 2014; Papa et al., 2016). Measuring all of the trade-offs is a very challenging task, and this research proposed a first step in this direction.

1.2 Structure of the thesis

This thesis includes five sections. Section 2 provides a comprehensive critical review of recent empirical and modeling peer-reviewed studies on the relationship between cities and energy consumption. The review highlights the knowledge gap between what is known and what is still under debate and, based on that, it proposes a theoretical framework to guide future research on this topic.

Section 3 focuses on the methodology for developing the statistical models performed to investigate the relationship between urban features and energy consumption. In particular, it first describes the set of physical, functional, geographical, socio-economic and energy variables selected for the model, as well as the data collection procedure.

Then, for each variable, it shows the performance of the Italian cities included in the research sample, thus providing useful information about similarities and differences in urban and energy characteristics amongst main Italian urban areas. Section 3 concludes by presenting the statistical methods used for the analysis of data, i.e. correlation analysis, multiple regression analysis and cluster analysis.

Section 4 shows the results of both correlation and regression analysis, as well as cluster analysis. The results are carefully interpreted and discussed considering previous findings found in the scientific literature in order to highlight two types of significant relationships: (1) the relationships between the different urban factors, which may indirectly affect CO_2 emissions; (2) the direct relationships between urban factors and CO_2 emissions. This section ends by proposing an updated version of the theoretical framework presented in Section 2, based on the obtained new results.

Section 5, lastly, provides some concluding remarks that could be helpful for supporting policy makers in the definition of effective strategies to be implemented at an urban scale to reduce energy consumption and resulting CO_2 emissions. Furthermore, this Section highlights the main limitation of this work and outlines potential directions for future research on this topic.

SECTION 2. CITIES AND ENERGY CONSUMPTION: A CRITICAL REVIEW

2.1 Introduction

Adopting the Paris Agreement in 2015, for the first time governments from all over the world agreed to "hold the increase in the global average temperature well below 2 °C above pre-industrial levels and to pursue efforts to limit the temperature increase to 1.5 °C above pre-industrial levels, recognizing that this would significantly reduce the risks and impacts of climate change" (FCCC, 2015). Local governments play a key role in the implementation of actions aimed at decarbonisation (OECD, 2014). According to IEA (2016), urban areas consume about two-thirds of primary energy demand and produce over 70 per cent of global carbon dioxide emissions (CO_2). Consequently "cities are the heart of the decarbonisation effort" (IEA, 2016) and can be the solution to climate change. However, urban growth shows no sign of slowing, and the energy and carbon footprint of cities does not seem to decrease. Therefore, energy efficiency improvements in urban areas are urgently needed to meet national and global ambitious sustainable goals.

To support local policy makers' decisions and foster the transition towards a low-carbon future, a growing body of international researchers has been studying the complex and multidimensional relationship between cities and energy consumption. These studies differ from each other in a wide variety of ways. First of all, they take into account different types of urban characteristics (e.g. density, household size, income, etc.) and consider different types of energy consumption (e.g. total, transport, or residential energy consumption). Additionally, the samples of cities analyzed differ in scale, size and geographical location. Therefore, it is no surprise that this heterogeneity in approaches and methodologies leads to a variety in results. Literature does not provide a comprehensive critical review highlighting the gap between what we know – and we all agree about – and what we need to know about how cities affect energy consumption and CO_2 emissions (Jabareen, 2006). So the aim of this Section is to critically categorize and compare recent interdisciplinary scientific literature on the relationship between cities and energy consumption to develop a theoretical model to guide future research based on the resultant new knowledge.

Section 2 is structured as follows. In paragraph 2.2 I present the approach used for this review and sets the temporal and contextual limitations of this work. In paragraph 2.3 I describe the critical review of the relevant literature on the relationship between urban areas and energy use, comparing the approach, methodology and results of the different contributions. Finally, in paragraph 2.4 I propose a conceptual framework that provides new understanding based on the integration of the results previously described, and helps stimulating the debate on this topic. This framework aims to offer a direction for future research and support local policy makers in the definition of strategies, policies and actions that can effectively reduce urban energy use and carbon dioxide emissions at city scale.

2.2 Approach

A good review on the relationship between urban form and travel patterns can be found in Stead & Marshall (2001), while a detailed review on the relationship between urban structure (construction, maintenance and use of residential dwellings) and residential and transport related energy use can be found in Rickwood et al. (2008). However, urban form and structure are just two aspects of a bigger picture. In both reviews an integrated approach is missing, which takes into account the variety of urban factors affecting energy consumption and CO2 emissions at city level. The relationship between cities and energy consumption is multidimensional, especially because cities are complex and dynamic systems (Batty, 2008; Papa, 2009); therefore, a comprehensive review about this topic calls for a holistic approach that considers a wider range of urban factors – physical, functional, geographical, social, economic – influencing the energy and carbon footprint of cities. Moreover, an integrated approach also allows the identification of the existing trade-off between different urban features and energy saving (Doherty et al., 2009; Lee & Lee, 2014; Papa et al., 2016), providing a broader and more complete framework on such a complex topic.

Based on these considerations, this review combines interdisciplinary researches that investigate the multidimensional relationship between cities (in their complexity) and energy consumption. Using a holistic perspective, the critical review of these contributions revealed that different studies have considered different categories of urban features influencing energy consumption and CO_2 emissions. I have classified and summarized these features into four groups, each including a different number of variables: (1) physical features; (2) functional features; (3) geographical features; (4) socio-economic features. Giving that there is no single way of identifying different categories (Stead & Marshall, 2001), this classification is based on the General System Theory (von Bertalanffy, 1969) applied to the urban phenomenon (Gargiulo, Papa, 1993). In particular, according to the systemic principles, cities can be defined "as sets of elements or components tied together through sets of interactions" (Batty, 2008) and an urban system can be represented as a set of four subsystems: physical subsystem;

functional subsystem; geomorphological subsystem; anthropic subsystem (Papa et al., 1995). The four categories of urban features previously introduced reflect the aforementioned four urban subsystems.

The first group of urban features – physical features – includes those variables measuring the physical subsystem of a city, which consists of the spaces/areas of an urban system that have been transformed in order to accommodate all different types of human activities. This set of variables describes the so-called urban form of a city. There is a little doubt that urban form – typically measured in terms of density – has been given a brighter spotlight within the overall scientific debate. Nevertheless, there are other physical factors whose influence on energy consumption and CO_2 emissions has been investigated by the reviewed studies, including those measuring polycentricity (Bereitschaft & Debbage, 2013; Chen et al., 2011; Lee & Lee, 2014) and fragmentation (Chen et al. 2011) as well as green areas (Banister et al., 1997; Gargiulo et al. 2016; Holden & Norland, 2005; Ye et al., 2015).

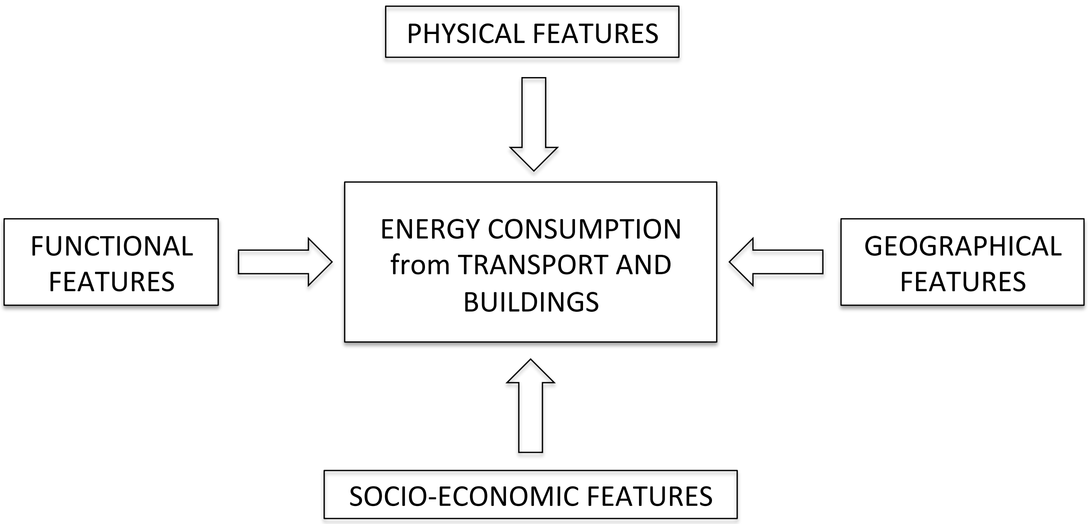
The second group of urban features – functional features – includes those variables describing the type and scale of activities carried out in a given city and, therefore, it reflects the urban functional subsystem. Some examples of functional factors include the proportion of jobs in the city center (Camagni et al., 2002; Mindali et al., 2004; Newman & Kenworthy, 1989) or the mix of housing, business and services (Holden & Norland, 2005; Jabareen, 2006) within a specific area.

The third group of urban features – geographical features – comprises those factors that refer to the specific context of reference and describe the differences in geographic aspects such as topography – e.g. percentage of coastal area (Creutzig et al. 2015; Ewing & Rong, 2008) – and climate – e.g. heating/cooling degree days (Baur et al., 2013; Creutzig et al. 2015; Ewing & Rong, 2008; Kennedy et al., 2009). This group provides a characterization of the whole urban territory, so reflecting the city's geomorphological subsystem.

Finally, the fourth and last group of urban features – socio-economic features – reflects the urban anthropic subsystem, which consists of all of the city's inhabitants as well as those people conducting activities for a limited amount of time within the urban perimeter. These urban features describe both social and economic aspects: examples of social variables analyzed by the reviewed studies include the level of education (Brownstone & Golob, 2008; Holden & Norland, 2005) and the proportion of young population (Banister et al., 1997), while examples of economic indicators are the income (Baur et al., 2013; Clark, 2013; Creutzig et al., 2015; Ewing & Rong, 2008; Holden & Norland, 2005; Kennedy et al., 2009; Makido, 2012) and the number of vehicles per inhabitant (Banister et al., 1997; Brownstone & Golob, 2009; Mindali et al., 2004).

In addition to this first categorization, the review also allowed the identification of different categories of energy consumption and/or CO2 emissions. Therefore, I have distinguished between: (a) energy consumption/CO2 emissions from the transport sector; (b) energy consumption/CO2 emissions from the residential sector; (c) total energy consumption/CO2 emissions. Based on this structure (Figure 1), I have developed a conceptual framework that integrates the different connections between urban features and energy consumption/CO2 emissions that have been empirically evaluated by published studies.

Figure 1 Structure of the review



In particular, this review includes empirical and modeling peer-reviewed studies that encompass a variety of cities samples, many of which located in Western Europe, in the United States and East Asia. Although some studies up to 2000 are reviewed, greater attention is given to those studies published after 2000. As to the scale of analysis considered in this review, I limited our analysis to those studies that evaluate the connections between urban areas and energy use at urban scale. Table 1 presents a synthesis of the review. In particular, each article has been categorized based on the urban feature/s (axis y) and the type of energy consumption/CO_2 emissions (axis x) considered. This table helps identifying on what researchers' attention has mainly focused and where critical knowledge gaps concentrate.

Table 1 Scientific studies categorized by urban feature and type of energy consumption / CO_2 emissions

Categories of energy consumption/CO_2 emissions **Class of urban feature**	Energy consumption / CO_2 emissions from TRANSPORT	Energy consumption / CO_2 emissions from BUILDINGS	TOTAL Energy consumption / CO_2 emissions
PHYSICAL features	Banister et al. (1997) Baur et al. (2013) Bereitschaft & Debbage (2013) Brownstone & Golob (2009) Camagni et al. (2002) Clark (2013) Creutzig et al. (2015) *Echenique et al. (2012)* Holden & Norland (2005) Kennedy et al. (2009) *Lee & Lee (2014)* Makido et al. (2012) *Marshal (2008)* Mindali et al. (2004) Newman & Kenworthy (1989) *Nuzzolo et al. (2014)*	Chen et al. (2008) Chen et al. (2011) *Echenique et al. (2012)* Ewing & Rong (2008) Holden & Norland (2005) Kennedy et al. (2009) *Lee & Lee (2014)* Makido et al. (2012) Ye et al. (2015)	Baur et al. (2013) Creutzig et al. (2015) *Echenique et al. (2012)* Kennedy et al. (2009)
FUNCTIONAL features	Banister et al. (1997) Camagni et al. (2002) Creutzig et al. (2015) Holden & Norland (2005) Mindali et al. (2004) Newman & Kenworthy (1989)	Holden & Norland (2005)	Creutzig et al. (2015)
GEOGRAPHICAL features	Bereitschaft & Debbage (2013)	Ewing & Rong (2008) Kennedy et al. (2009) Makido et al. (2012)	Baur et al. (2013) Creutzig et al. (2015)
SOCIO-ECONOMIC features	Banister et al. (1997) Baur et al. (2013) Brownstone & Golob (2009) Camagni et al. (2002) Clark (2013) Creutzig et al. (2015) Holden & Norland (2005) Kennedy et al. (2009) Makido et al. (2012) Mindali et al. (2004) Newman & Kenworthy (1989)	Ewing & Rong (2008) Holden & Norland (2005) Kennedy et al. (2009) Makido et al. (2012)	Baur et al. (2013) Creutzig et al. (2015) Kennedy et al. (2009)

2.3 Relationships

2.3.1 Physical features and energy consumption

Despite numerous efforts to define urban form, a shared approach for measuring the physical component of a city is still missing (Jabareen, 2006; Levy, 1999; Marshall, 2005; Newton, 2000). The complexity of connections between the city and both natural and anthropic activities makes the definition of urban form a challenging task that depends on multiple factors, which are often underestimated or even unrecognized (Lynch 1981). Nevertheless, there is a wide consensus of opinions that urban form – in all its definitions – can have an influence on energy consumption and CO_2 emissions, and consequently a great number of studies have investigated this relationship. In this context, the dichotomy between compact and dispersed city appears to be a key factor in the identification of a sustainable urban form. However, although it has long been argued that sprawling cities tend to consume higher amounts of energy than compact ones (Banister et al., 1997; Clark, 2013; Ewing and Rong, 2008; Marshal, 2008; Newman and Kenworthy, 1989), there has also been some criticism (Baur et al., 2013; Brownstone & Golob, 2008; Echenique et al., 2012; Mindali et al. 2004). Therefore, the relationship between urban compactness and environmental sustainability is not straightforward, yet (Chen et al., 2008, Williams et al., 2000).

If considering the general system theory and the physical subsystem, urban form should be measured in terms of housing density (i.e. the number of dwelling units in a given area) rather than population density (i.e. the number of inhabitants in a given area). Housing density, indeed, specifically refers to the built-up area of a city and provides a more precise idea of the physical urban development. However, most studies have considered population density a reliable and effective variable for the measurement of urban compactness (Breheny, 2001). Among these studies – both empirical and modeling – many agree that population density is negatively correlated with energy consumption and CO_2 emissions from transport and buildings. In particular, as far as the transportation sector is considered, Newman and Kenworthy (1989) find a strong negative correlation between population density and annual gasoline use per capita for a global sample of 32 cities, using an analysis of correlation. Similar results are shown by Camagni et al. (2002) for the case study of Milan, that find a significant inverse relationship between population density and the index of mobility impact (which refers to the mobility demand generated in each municipality within the city's perimeter), using an analysis of regression. Same results are found by Banister et al. (1997) for five cities in the UK, that argue that "higher density urban areas may help reduce the need to travel", and by Kennedy et al. (2009), whose analysis of ten big cities in the world shows that GHG emissions from ground transportation fuels are negatively correlated with population density.

If the residential sector is considered, supporters of compactness are Holden & Norland (2005), who compare eight residential areas within the Oslo region and show that "in densely developed areas, residents use less energy than do residents in areas with lower-density housing. This is mainly the result of more efficient energy supply systems – such as remote heating systems based on heat pumps – than can be introduced in areas with a large number of housing units per area unit". In line with this argument, the study carried out by Chen et al. (2008) for a sample of 45 Chinese cities evaluates the relationship between population density and a set or urban environmental variables, including domestic electricity and natural gas consumption. Through an analysis of correlation, the authors find a weak inverse relationship between urban compactness and domestic energy consumption.

More recently, new support to the theory that compact developments are more energy efficient than dispersed ones came from Makido et al. (2012), Clark (2013), Bereitschaft and Debbage (2013), and Creutzig et al. (2015): Makido et al. (2012) use a correlation analysis and a multiple linear regression analysis to investigate the relationship between urban form and CO_2 emissions in 50 Japanese cities and find that higher population density is associated with less CO_2 emissions from the passenger transport sector; according to Clark (2013), "higher population density – particularly in core areas – correlates with lower levels of per capita travel, and transport-related energy consumption and carbon emissions in the United States", but it is also associated with diminished housing affordability and increased congestion; same geographical context – the U.S. – for the study carried out by Bereitschaft and Debbage (2013), that find for every standard deviation increase in residential density, CO_2 emissions from on-road vehicles decreases of approximately 1.9 million t. On the other hand, Creutzig et al. (2015) find a strong negative correlation between population density and both transport energy use and GHG emissions for a sample of 274 global cities, using either a correlation and a regression analysis.

Along the same line of thoughts, however using a modeling approach rather than an empirical one, Marshal (2008), Lee & Lee (2014) and Nuzzolo et al. (2014) support the greater sustainability of denser urban areas, and quantify the impact of density on transport energy consumption and emissions. In particular, by comparing five U.S. urban growth scenario – high sprawl, business as usual (BAU), reduced sprawl, no sprawl, infill – Marshal finds that the reduced sprawl, no sprawl and infill scenarios decrease on-road gasoline CO_2 emissions compared to BAU, between 2005 and 2054, by 41%, 53% and 60% of a wedge respectively. Weaker but similar results are estimated by Nuzzolo et al., who compare five different scenarios – compact, transit oriented development, sprawl, trend, and BAU – for the city of Rome, and find that the compact scenario reduces CO_2 emissions and energy consumption deriving from car use by 24%. Analogously, Lee & Lee estimate for 125 urbanized areas in the U.S. that a 10% increase in population-weighted density – "*estimated as the weighted mean of census block group level densities, with each block group's population being used as the weight*" – decreases CO_2 emissions from travel and residential energy consumption by 4.8% and 3.5% respectively.

Criticizing all findings previously described, a smaller but consistent body of literature doubts the inverse correlation between population density and energy consumption/CO_2 emissions from transport and buildings. In particular, Mindali et al. (2004) highlight the inconsistency of the data collection method used by Newman and Kenworthy in the 1989 study and find very different results using the same sample and data set but a multivariate statistical approach: when cities are divided into clusters – one of North American and Australian cities and one of European cities – urban density has no effect on energy consumption from transport for both groups. Similarly, Baur et al. (2013) critic the robustness of the sample used by Newman and Kenworthy, in terms of geographical heterogeneity and numerosity. Also for a group of 62 European cities of different size they find that "population density is not, per se, a strong determinant of greenhouse gas emissions (neither for transportation GHG emissions, nor for total urban GHG emissions)". Similar results, but limited to California, are shown by Brownstone & Golob (2009), who argue that higher housing density decreases household vehicle use and resulting CO_2 emissions, but the impacts are too modest in magnitude to be considered significant – i.e. a 40% increase in housing density corresponds to a 5.5% fuel use reduction. In line with these findings, Echenique et al. (2012) use different models to estimate the sustainability of four spatial options – compaction, sprawl, edge expansion, and new town – for three different English city regions. They find that compaction decreases vehicle distance travel, but only by 5% compared to the trend, and the associated CO_2 reduction benefits are too small compared to "the potential socioeconomic consequences of less housing choice, crowding, and congestion".

In addition to the studies just described, which measure urban form in terms of population density, other researchers considered more complex indicators for assessing urban compactness and the way it affects energy consumption. Ewing and Rong (2008) measure urban form using Ewing et al.'s (2003) county sprawl index, which is calculated based on population density as well as street accessibility and clustering of development. For a sample of 266 U.S. counties, the authors indirectly estimate that urban sprawl positively affects residential energy use and, therefore encourage compact development. Similarly, Ye et al. (2015) analyze the case study of Xiamen and propose a normalized compactness index (NCI) based on Thinh et al.'s (2002) metric, which measures urban compactness in terms of gravity or attraction of a specific urban area. They find a positive correlation between the NCI and residential energy consumption, and interpret these results suggesting "that a compact city with heat and energy conservation from less-exposed wall and roof areas per capita, and more multifamily houses sharing foundations and resources, has residential energy savings".

A plurality of indicators is used by Chen et al. (2011) and Makido et al. (2012), who describe urban form using five and four different variables respectively. In particular, Chen et al. (2011) adopt a panel data analysis to study the relationship between five landscapes metrics – total urban class area, number of urban patches, mean perimeter-area ratio, Euclidean nearest neighbor distance, largest patch index – and energy intensity in production and living, in five Chinese cities. They find that (1) bigger cities consume more energy; (2) fragmentation in urban development increases energy consumption; (3) connectivity between patches is negatively correlated with energy consumption; (4) the largest patches index is negatively correlated with energy consumption, which suggests that concentration of urban activities should be encouraged, supporting the environmental sustainability of compact development. A similar approach is that employed by Makido et al. (2012), who consider three spatial metrics – the buffer compactness index (BCI), the compactness index (CI), and the area weighted mean patch fractal dimension (AWMPFD) – in addition to population density (measured in terms of urban area per capita and previously discussed), to estimate the relationship between urban form and CO_2 emissions from transport and buildings in Japan. Using a multiple linear regression analysis, the authors find that the BCI in the only spatial metric significantly correlated with energy consumption; in particular, increased BCI (i.e. increased compactness and monocentricity) decreases emissions from the passenger transport sector, but increases residential CO_2 emissions.

Although studies on the relationship between urban form and energy consumption mostly focus on the dichotomy between compact and sprawl development, some researchers include other physical urban variables in their analysis, such as house size, house typology, house age and availability of green spaces. In this context, it is shared opinion that bigger house size is associated with higher CO_2 emissions from transport (Lee & Lee, 2014) and buildings (Baur et al., 2013; Ewing & Rong, 2008; Holden & Norland, 2005), and that attached new houses are more energy efficient than detached old ones (Ewing & Rong, 2008; Holden & Norland, 2005). As far as green areas are concerned, results are not unanimous. In particular, Banister et al. (1997) find that the amount of open space in positively correlated with transport energy use in the case of Banbury and negatively correlated in the case of Oxford, while Ye et al. (2015) find that a greater connectivity and a weaker accessibility of green spaces is associated with higher CO_2 residential energy use.

To summarize, two main groups can be recognized in the debate on the relationship between urban form and energy consumption: those who support the compact city and those who question the magnitude of its environmental benefits. While compact development advocates support the idea that people living in dense urban settlements are less automobile dependent, tend to live in multifamily houses, and thus consume less energy than do residents in sprawl areas, critics suggest that the energy savings associated with the intensification of land use are too small to be considered significant, and they may be associated with negative externalities such as congestion, higher housing price, and less availability of green areas.

2.3.2 Functional features and energy consumption

Some of the studies on the relationship between urban form and energy consumption (described in the previous paragraph) also evaluate the energy and carbon footprint of a number of urban features that measure the functional organization of an urban system. It is of interest to note that the scientific literature does not offer any research that is exclusively focused on the relationship between urban functional features and energy consumption, but functional and physical features are always considered together. This may be because these two types of urban characteristics are very much connected to each other, and are both associated to the aforementioned compact city concept: in general, high-density and mixed-use development are typical of what can be defined a compact urban settlement (Burton, 2000), while the segregation of different land uses is typical of urban sprawl (Anderson, 1996).

In this context, the study carried out by Holden & Norland (2005) – earlier described for its results in terms of physical features and energy consumption – finds that the mix of housing, business and services does not have any significant effect on energy consumption from transport. Furthermore, they find a similar result for housing density, and suggest that "high density and high local mix must be combined with proximity to a center offering everyday services to bring about a reduction in energy use for everyday travel". However, stronger results are those found by Camagni et al. (2002), which use the ratio of jobs to resident population to measure the functional mix of a specific urban area, and find that this indicator is significantly inversely correlated with mobility, thus showing that higher mobility impact is associated with residential areas rather than with mixed ones. Similar results are those of Banister et al. (1997), that also use the ratio of jobs to population as a measure of functional mix, and find that mixed developments consume less energy from transport if local jobs and facilities are appropriate for local residents.

The proportion of jobs in the city center is one more indicator that describes the functional characteristics of different urban development and that has been considered by the scientific literature for its impact on energy consumption. In particular, Mindali et al. (2004) divide Newman and Kenworth's (1989) sample of 32 global cities in two groups (i.e. North American and Australian cities; European cities) and find a strong negative correlation between this variable and gasoline consumption for both groups. This result confirms Newman and Kenworthy results from 1989. However, Newman and Kenworthy also find no correlation between the absolute number of jobs in the city center and gasoline use for their sample of 32 global cities. The two results together suggest that the effect of the strength of the city center on gasoline consumption is not straightforward and that it may be that "it is largely the transportation policies applied to central cities that determine whether or not a significantly centralized work force is going to have a positive or negative effect on gasoline use" (Newman and Kenworthy, 1989).

Finally, it is of interest to also look at the indicator employed by Creutzig et al. (2015) for measuring the economic activity of the world cities included in their sample. The authors use the "center of commerce index" (Worldwide Mastercard, 2008), which classifies 75 leading urban centers based on their role in enabling commerce worldwide, and find a positive correlation between this proxy and the total final energy use. This finding highlights the role of production activities as key factors affecting the carbon footprint of urban areas.

In summary, there are relatively few studies that investigate the impacts of urban functional features on energy consumption. Although some results may appear contradictory, the general argument that emerges is that the positive effect of mixed-use development on energy saving from transport is not significant by it self, but becomes significant when combined with high density and supply of transit services.

2.3.3 Geographical features and energy consumption

Ewing and Rong (2008) are the first to consider topographic and climatic variables in their analysis on the relationship between cities and residential energy consumption. In particular, they find a positive correlation between heating degree days (HDDs) and energy use for heating, as well as between cooling degree days (CDDs) and energy use for cooling. Furthermore, they include data describing the topographic configuration of the 266 U.S. counties in their sample, but employ these two dummy variables – coast and valley – only to evaluate their relationship with climate. Thus, the authors don't provide any information about the way territorial geography may affect energy consumption. In this context, Creutzig et al. (2015) conduct a similar analysis by including HDDs, CDDs and coastal city location in their study of 274 global cities. Their analysis of regression shows that HDDs are positively correlated with both final energy and GHG emissions and "explain an important fraction of the energy use variability of cities", while CDDs and coastal city location do not significantly affect either energy use or GHG emissions.

The positive effect of HDDs on residential energy use found by both Ewing and Rong (2008) and Creutzig et al. (2015) is further confirmed by Kennedy et al. (2009), who analyze 10 global cities and find that the amount of fuel used for heating and industrial activities increases with HDDs. On the contrary, Baur et al. (2013) don't find any significant influence of HDDs on total GHG emissions for 62 European cities, possibly because their data on GHG emissions were previously corrected for seasonal variations, as specified by the authors. Similarly, in their analysis on urban form, air pollution and CO_2 emissions in 86 U.S. metropolitan areas, Bereitschaft and Debbage (2013) show that the two climate factors considered – temperature and moisture – are not associated with total CO_2 emissions, but only with O_3 concentrations and $PM_{2.5}$, VOC_s, and NO_x respectively. More controversial are the results of Makido et al. (2012), who use cities' average temperature instead of HDDs, and find a negative effect on residential CO_2 emissions. In this case, the authors admit the difficulties in interpreting such results and suggest the inclusion of HDDs rather than the average temperature in a future research.

To summarize, the relationship between geographical features and energy consumption has been interpreted by the literature as that between climate – specifically HDDs – and energy consumption from buildings. In this context, it is widely argued that an increase in HDDs is associated with an increase in CO_2 emissions from heating. As far as the geographical location of cities is concerned, only one research finds that the proximity to the ocean does not affect energy consumption. Future research should further investigate the importance of these aspects as well as that of urban topography with respect to energy consumption.

2.3.4 Socio-economic features and energy consumption

Researchers have extensively studied the impacts of economic and social factors on energy use. As far as the economic features are concerned, most of the attention has been focused on the effects of three main variables – income, fuel price and car ownership – on transportation first, and on residential and total energy consumption later. In particular, Newman and Kenworthy (1989) find that these three indicators are responsible for about 60% of gasoline use, while the remaining 40% depends on urban form and land use factors.

With respect to income, it is widely recognized that higher standard of living results in higher emissions from both transport (Brownstone & Golob, 2009; Clark, 2013; Holden & Norland, 2005; Newton & Kenworthy, 1989) and buildings (Ewing & Rong, 2008; Kennedy et al. 2009). In this regard, the results by Creutzig et al. (2015) are of particular interest. When considering the whole sample of 274 global cities, the authors find that final energy consumption is strongly positively associated with economic activity, but in the moment that they divide the sample in eight groups based on gross domestic product (GDP) per capita, density, fuel price, and HDDs, they find that "energy consumption for urban transport increases with GDP at low GDP levels, but decreases with GDP at high GDP levels". These findings give new insight into the question, and open up new avenues for future research. With regard to fuel price, Newman & Kenworthy (1989) argue that this economic factor is inversely correlated with transport energy consumption, and Ewing & Rong (2008) find a similar negative relationship between energy price and residential energy demand. More recently, Creutzig et al. (2015) find a negative relationship between fuel price and total energy use and emissions, thus supporting both previous results. Finally, considering car ownership, as reasonably expected, studies find that higher levels of car ownership are associated with higher energy use from transport (Banister et al., 1997; Mindali et al., 2004).

As far as the social features of urban areas are concerned, the impacts of different social aspects on energy consumption have been investigated by the scientific literature, but weak consensus exists among researchers. According to Camagni et al. (2002), for example, population growth rate positively affects mobility, while on the contrary, Baur et al. (2013) find that this indicator doesn't significantly influence total GHG emissions. Similar contradictory results are found when household composition is investigated: while Brownstone and Golob (2009) show that in California fuel use increases with the number of children, Ewing and Rong (2008) don't find any significant relationship between residential energy consumption and either the number of children or the number of adults, in the U.S.. There is the same debate when the level of education is considered, because those who find that education positively affects transport energy use – "households headed by a respondent with a college degree tend to have a vehicle fleet with greater overall lower fuel economy than their less educated counterparts. This effect is accentuated if the household is headed by a respondent with a postgraduate degree" (Brownstone & Golob, 2008) – are criticized by those who don't find any significant correlation (Holden & Norland, 2005). One last social aspect considered for its potential impacts on energy consumption is ethnicity; in particular, both Ewing and Rong (2008) and Brownstone and Golob (2009) find that energy consumption varies by race, but this relationship needs more specific research to be fully understood.

To summarize, it is widely recognized that social and economic factors affect energy consumption. However, while there is great consensus about the relationships between economic variables – income, fuel price, and car ownership – and energy consumption, there is far less agreement about the way social characteristics, such as demographic growth, household composition, education, and race may influence energy use.

2.4 A conceptual framework to guide future research

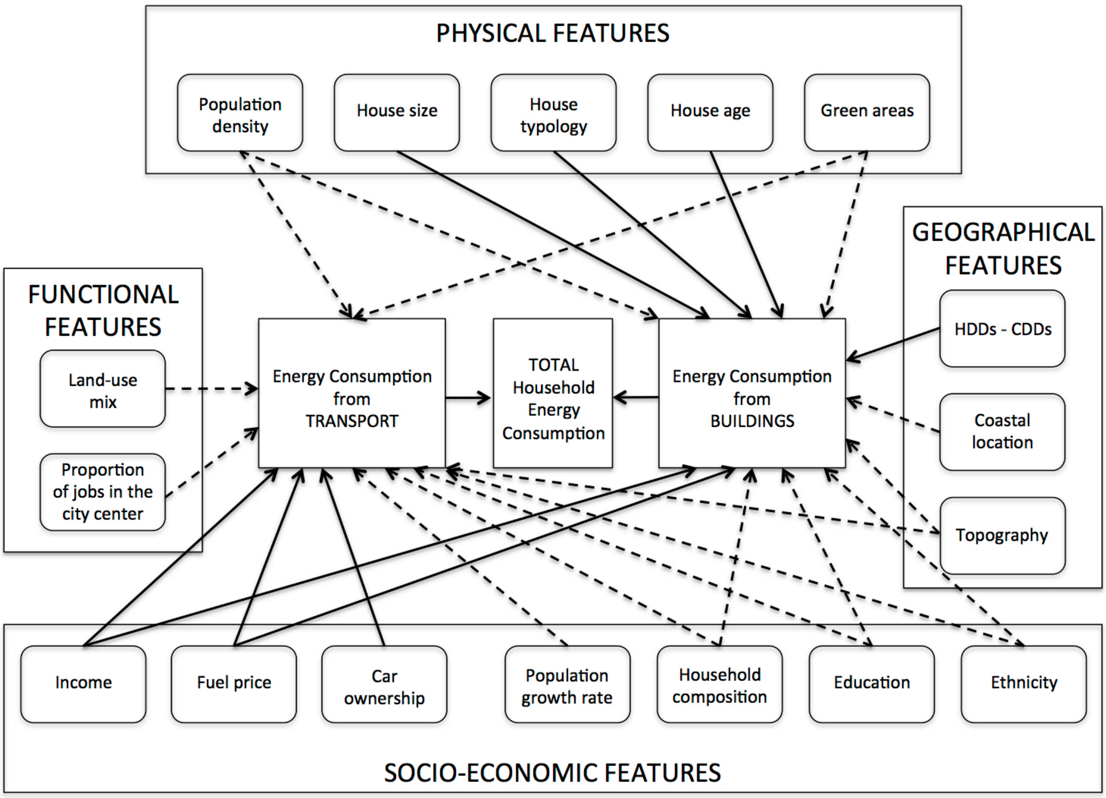
The review of the scientific literature on the relationship between cities and energy consumption allows the construction of a conceptual framework (Figure 2), which has two main goals: (1) to provide a state of the art summary on this topic, and (2) to suggest some directions for future research. The conceptual framework is built based on the integration of the findings previously described, and it takes into account the four categories of urban features that have been used to represent the urban system (according to the general system theory applied to the urban phenomenon). In particular, for each group of features, the main variables are specified and the relationships between these variables and the two types of energy consumptions – from transport and from buildings – are identified. Two different types of arrows are used: solid arrows represent those relationships for which there is a wide consensus within the scientific community, both in terms of "sign" (i.e. positive or negative relationship) and significance; on the contrary, dashed arrows indicate those relationships that require further investigation because of the conflicting results found in the literature so far.

At the top of the figure are the five physical features – population density, house size, house typology, house age, and green areas – that emerge from the literature review as key factors significantly affecting energy consumption at city scale. As far as population density is concerned, two dashed arrows connect this variable with both types of energy consumption; this is because, although there are numerous studies on the relationship between urban form and energy use, and the majority agree that population density is negatively correlated with both transport and building energy use, there is still a lack of consensus among researchers about the size of this correlation, and thus its significance. Similarly, further research is needed to explore the way green spaces affect energy consumption. On the contrary, the scientific findings about the relationships between the other three physical features – house size, house typology, and house age – and residential energy consumption are sufficiently reliable and widely shared in the literature, thus these arrows are solid.

At the left of the figure are the two functional variables – land use mix and the proportion of jobs in the city center – influencing energy consumption from transport, but in both cases the relationship is not straightforward, either because of the relatively small number of studies on this issue or because of the strength of these two relationships depend on other external variables (e.g. urban density and transit service), as previously described in paragraph 2.3.2. Therefore, embracing the complexity of the urban system, additional effort should be made to investigate the influence of the urban functional subsystem on energy consumption.

At the right of the figure are the three geographical features – heating and cooling degree days, coastal location and urban topography – that affect household energy consumption. In particular, a solid arrow connects HDDs/CDDs and residential energy use, because it is widely argued that climate conditions significantly influence fuel consumption for heating and cooling. On the other hand, with regard to the other two geographical features, too little research has been done in order to assess the impacts of coastal location on residential energy use and of topography on either residential energy use or transport energy use. Thus, three dashed arrows associate these two variables and the two types of energy consumption.

Figure 2 Conceptual model and key relationships between the four groups of urban features and energy consumption



At the bottom of the figure are the seven socio-economic features – income, fuel price, car ownership, population growth rate, household composition, education, and ethnicity – that are in part responsible of both transport and residential energy use, according to the reviewed literature. While there is wide consensus on the relationship between economic variables and energy consumption, there is less of a consensus on the impacts of social factors on energy use. In particular, it is widely demonstrated that income and fuel price are correlated – positively and negatively respectively – with energy consumption, from both transport and buildings, and that an increase in car ownership results in higher transport energy use. On the contrary, more complex are the influences of the four considered social features on energy use, which may explain the dissimilarity in findings among studies. Future research, indeed, should focus more on the influence of household composition, education and ethnicity on energy consumption. Furthermore, more scientific attention should be paid to measure the consequences of demographic growth on energy consumption, especially today that urbanization processes are extremely pervasive.

2.4.1 Relationships amongst different urban features

Using a holistic approach (as previously described in paragraph 2.2), the conceptual model proposed above does not provide a comprehensive picture of the complexity of the relationship between cities and energy consumption. Indeed, another group of interaction exists and significantly contributes to such complex relationship. This group includes the interactions amongst the four different types of urban features (physical, functional, geographical, and socio-economic). Differently from the relationships described in the previous paragraphs, these interactions indirectly affect energy consumption. Nevertheless, these indirect effects can be significant and should not be ignored.

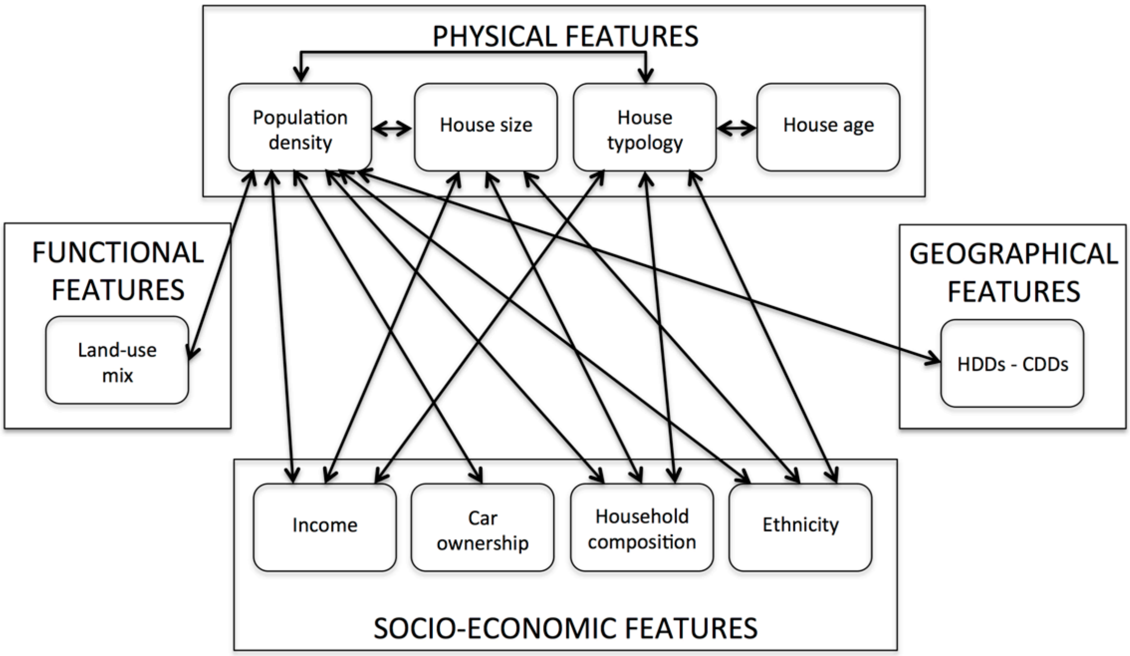
However, only a small part of the literature reviewed here considers these secondary interactions, which are summarized in Figure 3. In particular, Holden and Norland (2005) are the first to find a significant interaction between two physical features, i.e. house typology and house age. They find that the difference in energy consumption between single-family housing, row houses and multifamily housing is lower when considering housing units built after 1980. In other words, the energy efficiency of multifamily housing compared to single-family housing has decreased in recent years. This means that the direct effect of house typology on residential energy consumption becomes weaker when the indirect effect of house age is considered.

Similarly, Chen et al. (2008) find a positive interaction between population density and density of facilities (land use mix), which means that densely populated cities in China also have higher concentrations of activities. On the same page, Brownstone and Golob (2008) find that population density is negatively associated with car ownership, income and the number of family components, and that "households which are solely Black, solely Asian, solely Hispanic, or mixed White and Hispanic, all tend to reside in higher-density areas". Population (weighted) density is also found to be inversely association with housing type (calculated as an ordinal variable: 0 = multi-family, 1 = single attached, and 2 = single detached) and housing size (using the number of rooms as proxy), according to the results obtained by Lee and Lee (2014) using a multilevel structural equation model, which means that in denser populated areas there is a higher concentration of multi-family houses with a lower number of rooms.

Finally, Ewing & Rong (2008) devote much effort to analyze the way urban form can indirectly affect residential energy consumption through the housing stock and the formations of urban heat islands (UHIs). By using a hierarchical modeling, the authors find that house typology and house size are significantly associated with several socio-economic features. In particular, as the number of family members and income increase, both house size and the odds that the household will choose a single-family detached house increase. Analogously, also ethnicity is found to significantly affect the choice of both house typology and house size: White households are more likely to choose bigger single-detached homes than Black, Hispanic and Asian ones. Furthermore, Ewing and Rong also find that multifamily houses are associated with denser urban areas and that houses are significantly larger in sprawling counties than in compact ones. In addition to these results, the study shows that the effect of the urban heat island (UHI) is greater in compact developments, which implies that in denser areas "temperatures are higher that they would be otherwise".

Considered together, these results suggest that the indirect effects of these secondary interactions between physical, functional, geographical and socio-economic factors can significantly contribute to the increase and/or decrease of transport and residential energy consumption at urban scale. In other words, the correlations between different urban features and energy consumption found by the literature so far (and described in paragraph 2.3) cannot prove a causal relationship. Indeed, they may partially be the effect of secondary interactions between other variables. For example, a strong positive correlation between housing size and residential energy consumption may not be exclusively due to a direct link between these two variables, but it may also include the indirect effects of other physical (e.g. population density) and socio-economic (e.g. income and ethnicity) variables. However, it is very difficult to identify and untangle all the direct and indirect effects from different urban features on transport and residential energy consumption. Therefore, the task of establishing independent links between cities' characteristics and their energy and carbon footprint remains very challenging (Rickwood et al., 2008) and requires further investigation.

Figure 3 Key relationships amongst the four groups of urban features



2.5 Conclusions

Section 2 put together and compared the relevant literature on the relationship between cities and energy consumption over the last twenty years. Two main energy sectors have attracted the interest of the scientific community – transportation and residential sectors – and a large number of urban features have been analyzed. In particular, as I distinguished between four different categories of urban features (physical, functional, geographical, and socio-economic), the review showed that a great body of the literature has focused on the relationship between urban form (i.e. physical features) and energy consumption, while fewer researches have also investigated the effects on energy use and CO_2 emissions of other urban characteristics, such as those describing the functional, geographical and socio-economic aspects of a city.

Despite the great interest of the literature on this topic, a consistent number of interactions between urban features and energy use at urban scale still lacks consensus. One of the main open questions is about the relationship between population density and energy consumption. While it is widely argued that density is negatively correlated with both transport and residential energy use, there is less agreement about the scale (and significance) of this correlation and whether this inverse association can be generalized or whether it exists only for precise density ranges and specific clusters of cities. In addition to this open debate, the impact of social factors on energy use still requires further investigation. In particular, the effect of some social factors such as the level of education or the ethnicity on households' travel behavior and residential energy use.

Overall, three main limitations to the studies included in this review have emerged. Both first and second limitations concerns the approach used to analyze the relationship between cities and energy consumption. Many studies employ a sectorial approach rather than a holistic one. Consequently, they investigate a limited number of urban factors (physical factors prevalently) without accounting for other urban features that might confound previous findings. Furthermore, most research only considers direct effects of a number of urban factors on energy consumption or CO_2 emissions, without taking into consideration the possible indirect effects associated with the interactions that may exist amongst the different urban factors. As previously mentioned, these indirect effects may be significant and cannot be ignored if we want to explore the relationship between cities and energy consumption in its complexity and multidimensionality.

Finally, the third issue concerns the limited data availability. As highlighted in many of the reviewed studies, the lack of a comprehensive dataset about cities' energy consumption and CO_2 emissions by sector represents a significant limitation, which has been overcome by merging different data sources or by collecting data using questionnaires, whose reliability could be questionable. Similarly, once again because of data limitations, many of the described researches report as a limit that they have considered just a restricted number of urban variables while others, which may be equally important, could not been captured.

Given the findings of the studies presented above and taking into consideration the limitations previously described, I proposed a conceptual framework to guide future research on the relationship between cities and energy consumption. The proposed framework presents the main urban factors influencing the energy and carbon footprint of a city and illustrates clearly the key relationships between these features and both transport and residential energy consumption, highlighting those relationships that are not straightforward and require therefore further research (Figure 2). Most importantly, this framework also illustrates a second group of relationships – i.e. those amongst the four categories of urban features (Figure 3) – which may significantly affect energy consumption but are often ignored by the scientific literature, thus providing a more comprehensive picture of the complex and interconnected interactions between cities and energy consumption.

This wider picture represents the starting point of this research, whose aim is to evaluate the extent to which urban characteristics influence transportation and residential energy consumption on a city scale. Based on this framework, the next Section presents the methodology developed for measuring these impacts, which have to be clearly understood in order to enable urban planning policies to effectively improve energy saving in cities and reduce urban emissions.

SECTION 3. METHODOLOGY

3.1 Introduction

The goal of this research is to identify the urban factors that significantly affect a city's energy and carbon footprint, thus supporting policy-makers in the definition of effective strategies and policies that can be implemented at urban scale to reduce energy consumption and resulting CO_2 emissions.

This Section, in particular, presents the methodology for developing the statistical model performed to investigate the relationship between urban features and energy consumption on a city scale.

The theoretical framework proposed in Section 2 represents the starting point for the selection of the variables to be included in the statistical model. In particular, I collected data on a sample of 73 Italian cities for the chosen set of urban characteristics, categorized into four groups according to the general system theory applied to the urban phenomenon (physical, functional, geographical and socio-economic features). In addition, I collected data on CO_2 emissions per capita from residential, transport, tertiary, and municipal energy use. The complete set of data, including both urban and energy indicators, was analyzed using correlation and multiple regression analyses, as well as cluster analysis.

Section 3 is structured as follows. Paragraph 3.2 describes the data sample used for this research and the criteria adopted for this selection. Paragraph 3.3 illustrates the data collection procedure and the sources used for collecting the necessary information. Moreover, this paragraph also presents the set of physical, functional, geographical, socio-economic and energy variables selected for the statistical model. For each variable, I outline and discuss the performance of the Italian cities included in the research sample, thus providing useful information about similarities and differences in urban and energy characteristics amongst main Italian urban areas. Section 3 concludes by presenting the statistical methods used for the analysis of data, i.e. correlation, multiple regression and cluster analysis.

3.2 Data sample

The data sample for this research includes 73 Italian provincial capitals uniformly distributed across the country (Table 2 and Figure 4), which account for 12.340.751 inhabitants and correspond to the 21% of the total Italian population (recorded in 2011). This choice is the result of three main considerations: (1) sample homogeneity, (2) direct knowledge of the context, and (3) data availability.

(1) Several studies previously reviewed (Baur et al., 2013; Creutzig et al., 2014; Mindali et al., 2004) show the importance of sample clustering when different cities from around the world are considered together, in order to have more homogeneity and so more reliable results. Thus, the choice of selecting a group of cities within the same national context, with similar historical background and socio-economic development, is in line with this thinking. (2) The choice of the Italian context comes, then, from the direct knowledge of its main physical, functional, geographical and socio-economic features. This better understanding of the context of analysis enables a better interpretation of results, as well as an easier identification of potential outliers. (3) Why the Italian capital cities? Because, in Italy, considerable open access data are provided by the Italian National Institute of Statistics (ISTAT) for the 116 Italian provincial capitals, which represented the initial sample for this thesis. Because of the limited availability of data on energy consumption and CO_2 emissions, I had to reduce the original sample from 116 to 73 capital cities. This will be further discussed in paragraph 3.3.2.

Table 2 List of the 73 selected Italian cities, their geographic location and their population

City	Population (2011)	City	Population (2011)	City	Population (2011)
Alessandria	89,411	La Spezia	92,659	Potenza	66,777
Ancona	100,497	Latina	117,892	Ragusa	69,794
Andria	100,052	Lecce	89,916	Ravenna	153,740
Ascoli Piceno	49,958	Livorno	157,052	Reggio nell'Emilia	162,082
Bari	315,933	Lodi	43,332	Rimini	139,601
Barletta	94,239	Lucca	87,200	Roma	2,617,175
Belluno	35,591	Macerata	42,019	Salerno	132,608
Bergamo	115,349	Mantova	46,649	Sassari	123,782
Bologna	371,337	Matera	59,796	Savona	60,661
Bolzano	102,575	Messina	243,262	Tempio Pausania	13,946
Brindisi	88,812	Modena	179,149	Teramo	54,294
Cagliari	149,883	Monza	119,856	Torino	148,028
Campobasso	48,747	Napoli	962,003	Tortolì	10,743
Catania	293,902	Novara	101,952	Trani	55,842
Chieti	51,484	Nuoro	36,674	Trento	114,198
Cosenza	69,484	Olbia	53,307	Treviso	81,014
Cremona	69,589	Oristano	31,155	Trieste	202,123
Fermo	37,016	Padova	206,192	Udine	98,287
Ferrara	132,545	Palermo	657,561	Venezia	261,362
Firenze	358,079	Parma	175,895	Verbania	30,332
Forlì	116,434	Pavia	68,280	Verona	252,520
Genova	586,180	Pesaro	94,237	Vicenza	111,500
Grosseto	78,630	Pescara	117,166	Villacidro	14,281
Isernia	22,025	Piacenza	100,311		
L'Aquila	66,964	Pisa	85,858		

Figure 4 Map of the 73 Italian provincial capitals included in the sample

3.3 Data collection and description

Based on the conceptual framework described in Section 2 and on the availability of data for the sample of cities considered in this research, I selected eighteen urban variables describing the urban system and five variables measuring CO_2 emissions from energy use (Table 3). These variables were later included in the statistical model developed to estimate the relationship between urban features and energy consumption.

Table 3 Selected urban and energy variables

Category	Variable	Description	Source
PHYSICAL features	Housing density	Housing units per square kilometer	ISTAT (2011)
	House size	Average floor area per capita	ISTAT (2011)
	House age	% of buildings built after 1980	ISTAT (2011)
	House material	% of masonry buildings	ISTAT (2011)
	Green areas	% of green areas	ISTAT (2011)
FUNCTIONAL features	Land use mix	Jobs per square kilometer	ISTAT (2011)
	Functional specialization	Concentration of manufacturing activities	ISTAT (2011)
		Concentration of commercial activities	ISTAT (2011)
		Concentration of public activities	ISTAT (2011)
		Concentration of touristic activities	ISTAT (2011)
GEOGRAPHICAL features	Degree days	°C per day / year[1]	DPR 412/09
	Coastal location	Binary variable (0=coastal; 1=inland)	ISTAT (2009)
	Topography	Binary variable (0=non mountain; 1=mountain)	ISTAT (2009)
SOCIO-ECONOMIC features	Income	Average income per capita	Ministry of Economy and Finance (2012)
	Car ownership	Number of cars per 1.000 inhabitants	ISTAT (2011)
	Household composition	% of children (<15 years old)	ISTAT (2011)
	Education	Graduates per 1.000 inhabitants	ISTAT (2011)
	Ethnicity	% of foreign residents	ISTAT (2011)
ENERGY consumption	CO_2 emissions	Residential CO_2 emissions per capita	Sustainable Energy Action Plans (2005)
		Transport CO_2 emissions per capita	Sustainable Energy Action Plans (2005)
		Tertiary CO_2 emissions per capita	Sustainable Energy Action Plans (2005)
		Municipal CO_2 emissions per capita	Sustainable Energy Action Plans (2005)
		Total CO_2 emissions per capita	Sustainable Energy Action Plans (2005)

Detailed variable-by-variable data description is specified later herein (paragraph 3.3.1). When possible, the sample's data for each variable are compared to data at Italian and European level[2], in order to contextualize the information into a wider framework.

[1] Calculated as the sum of daily positive differences between a temperature of reference of 20°C (for Italy) and the average daily outdoor temperature, extended to the whole year.

[2] All data at Italian and European level were collected using Eurostat as data source.

Furthermore, Figures 5-27 provide a visual distribution of data for each parameter[3]. These maps were created using the natural breaks[4] function available in ArcGIS (Version 10.2).

I used two main data sources to collect the necessary information, i.e. the ISTAT for urban data and the Sustainable Energy Action Plans (SEAP) developed by each city for energy/emission data. Using data from only two data sources provides greater reliability and consistency. However, because of limited data availability on energy use at urban scale, urban and energy data refer to two different time period, 2011 and 2005 respectively. This aspect may represent a limitation in the accuracy of these research findings.

3.3.1 Urban variables

Following the general system theory (Gargiulo, Papa, 1993), I classified the 18 urban variables into four categories[5], which reflect the four subsystems of a city: (1) physical features; (2) functional features; (3) geographical features; (4) socio-economic features. Table 4 shows some descriptive statistics for the sample of this analysis.

The Italian National Institute of Statistics (ISTAT) provided open access data for sixteen out of eighteen urban variables. In particular, the majority of this information was collected from the 15th Population and Housing Census dated 9 October 2011, whose official results were published in 2014.

3.3.1.1 Physical features

This category of urban features describes the physical component of a city, i.e. the spaces that people live in (for more details see paragraph 2.2). The five physical features considered in this research are: (1) housing density, calculated as the ratio of the number of housing units to total land area; (2) house size, calculated as the average floor area per capita; (3) house age, calculated as the percentage of buildings built after 1980; (4) house material, calculated as the percentage of buildings made of masonry material; (5) green areas, calculated as the proportion of green areas to total land area. With respect to the theoretical framework previously proposed, one physical feature was not included (i.e.

---

[3] Although all data were collected at municipal level, I chose to represent data spread over the provincial territory. The representation at municipal level, indeed, would have been too small to be visible.

[4] *"Natural breaks classes are based on natural groupings inherent in the data. Class breaks are identified that best group similar values and that maximize the differences between classes. The features are divided into classes whose boundaries are set where there are relatively big differences in the data values"* (*http://pro.arcgis.com/en/pro-app/help/mapping/symbols-and-styles/data-classification-methods.htm*)

[5] This classification is described in more detail in Section 2, paragraph 2.2.

house typology) because of limited data availability, and one feature was added (i.e. house material) because different building materials may influence residential energy use, and thus CO_2 emissions (Balaras et al., 2007; Gerundo et al., 2016; Reddy & Jagadish, 2003).

Table 4 Descriptive statistics on urban characteristics for the sample of 73 Italian capital cities

Variable	Minimum	Maximum	Average	Standard deviation	Unit of measurement
Housing density	28.36	3214.38	579.96	623.96	housing units/km^2
House size	31.67	49.71	42.05	4.01	m^2/ inhabitants
House age	2.61	53.00	22.64	10.49	%
House material	23.43	91.19	53.40	14.13	%
Green areas	0.09	30.70	3.97	5.52	%
Land use mix	15.39	2496.70	422.98	467.74	jobs/km^2
Concentration of manufacturing activities	0.21	2.11	0.96	0.44	Location quotient
Concentration of commercial activities	0.48	1.65	1.02	0.21	Location quotient
Concentration of public activities	0.33	1.97	1.03	0.39	Location quotient
Concentration of touristic activities	0.09	8.67	1.08	1.12	Location quotient
Degree days	707	3001	1860.33	599.69	°C per day/year
Coastal location	0.00	1.00	0.44	0.50	binary variable
Topography	0.00	1.00	0.32	0.47	binary variable
Income	12183	26744	20411	2895	€
Car ownership	412.19	745.12	614.21	62.06	car/1000 person
Household composition	9.92	17.50	12.87	1.39	%
Education	55.50	242.21	150.51	38.64	grad./ 1000 person
Ethnicity	0.50	15.43	7.23	4.18	%

Housing density

As previously specified (paragraph 2.3.1), population density is widely used by the scientific community for measuring the compactness of urban settlements. However, according to the general theory system, the number of housing units in a given area (i.e. housing density) provides a more precise measure of urban form than the number of inhabitants to total land area. The latter one, indeed, refers more specifically to the anthropic subsystem rather than to the physical one. Therefore, after having collected data on both housing and population density and verified that the two variables are highly correlated (Pearson's coefficient = 0.99), I chose to use housing density as a physical variable for studying the relationship between cities and energy consumption. On the other hand, population density proved to be useful for contextualizing the cities included in the research sample within the Italian and European panorama, as described later herein.

Comparative statistics on housing density show that the cities in the sample are characterized by very different urban patterns (Figure 5). In particular, few major cities have very high level of density. For example, the density of Turin (3.214 dwellings per km^2) – that is the densest city in the sample as well as in the country – is over one hundred times bigger than housing density in Tempio Pausania (28 dwellings per km^2), which is at the bottom in the sample.

In order to have a better idea of the levels of density in Italian capital cities, it is useful to compare these values with those at both national and European scale. For this purpose, population density represents a more suitable indicator than housing density, for at least two reasons: it is more used at international level; and it considers population rather than housing units, thus being a more unbiased measurement because dwellings are likely to differ in typology and dimension when different contexts of urban development are considered. Therefore, in terms of population density, Italian capital cities are characterized by very high density (sample's average is 1.177 inhabitants per km^2) compared to the whole country (201 inhabitants per km^2) and Europe (116 inhabitants per km^2). Only two Italian capital cities – Villacidro and Tempio Pausania – have a population density lower than the European average, while Naples has a population density (8.203 inhabitants per km^2) about forty times bigger than the Italian one and seventy times bigger than the European one. These few comparative statistics show how compact main Italian cities are within both the national and European context.

House size

Small differences amongst Italian capital cities occur when house size is considered, resulting in a low standard deviation value (Table 4). Furthermore, very differently from what previously found in the case of housing density, the full sample's average floor area per capita (42 m^2) is very similar to the average value for Italy (40.68 m^2). Another interesting finding is the geographical distribution of the variable (Figure 6). The map clearly shows that the average floor area is greater in northern Italian capital cities and smaller in Southern territories. However, the city of Lecce (in Apulia) represents an exception: with almost 50 square meter of floor area per capita is the city with the highest house size within the research sample.

House age

The first Italian regulation for reducing building energy consumption was approved in 1976 (Legislative Decree 373/76), when the cost of oil unexpectedly rose because of the international oil crisis in 1973. It thus seems reasonable to assume that in Italy different building standards have been used since then. Based on these considerations and on the fact that data on buildings provided by the ISTAT are aggregated by decades, I chose the percentage of buildings built after 1980 as a proxy for house age.

In relation to the research sample, the average percentage of buildings built after 1980 is 23%. This value confirms that of Italy, where only 26% of residential buildings were built after 1980. As shown in Figure 7, most of these buildings are located in Sardinia, Basilicata and Apulia, all regions in the South of Italy. In particular, in two Sardinian cities – Tortolì and Olbia – more than half of the buildings were constructed after 1980.

House material

Two main types of buildings characterize the Italian context: masonry and concrete buildings. Masonry buildings represent almost 60% of total buildings in Italy, while concrete buildings are about 30%. The two different construction materials may have an influence on buildings energy use, especially on heating and cooling loads (Balaras et al., 2007; Reddy & Jagadish, 2003). For this reason I chose to include the "percentage of masonry buildings" in the dataset, and later investigate its relationship with CO_2 emissions.

Figure 8 shows that the highest percentage of masonry buildings is concentrated in the central-northern part of the country and in Sardinia, while in the South there is a greater use of concrete. In particular, over 80% of buildings in Tortolì and Olbia are masonry structures while, on the opposite side, Nuoro and Lodi have less than 30% of masonry buildings.

Green areas

This data on the concentration of green spaces comprises all different typologies of green areas located within the urban perimeter, including forests, municipal parks, green areas used as residential and recreational facilities, school and vegetable gardens.

The percentage of green areas varies a lot between the different cities, similarly to housing density. In both cases, indeed, the standard deviation is greater than the average (Table 4). In particular, the concentration of green areas varies from 0.09 % of Villacidro to 30.70% of Trento. The average percentage of green spaces for the research sample (3.8%) is slightly greater than that for whole country (2.7%). In the sample, only six capital cities have at least 10% of green areas, while forty-three cities (about 60% of the full sample) have less than 3% of green spaces (Figure 9).

Figure 5 Housing density

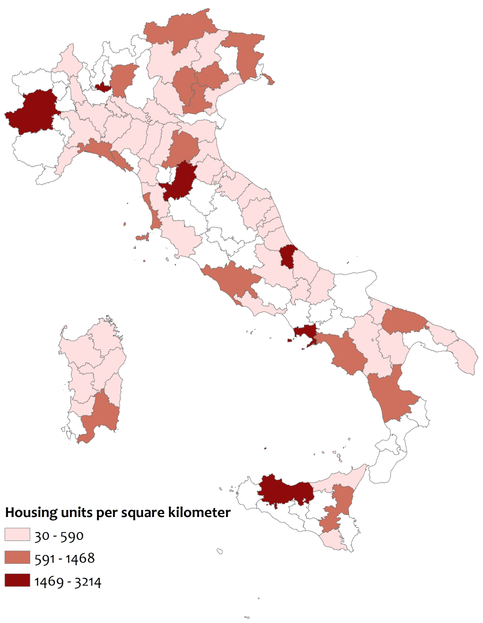


Figure 6 House size

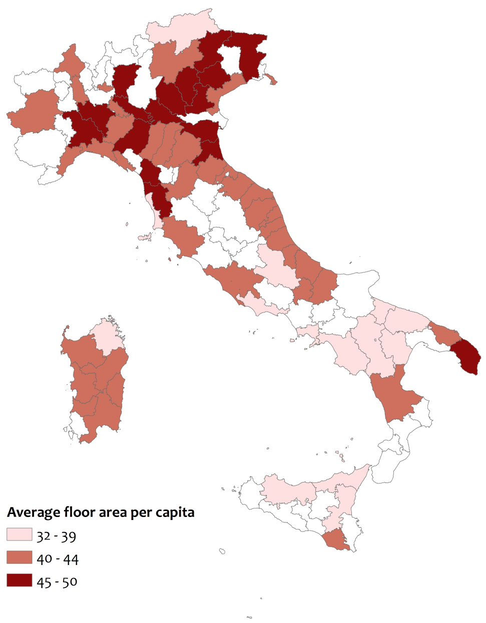


Figure 7 House age

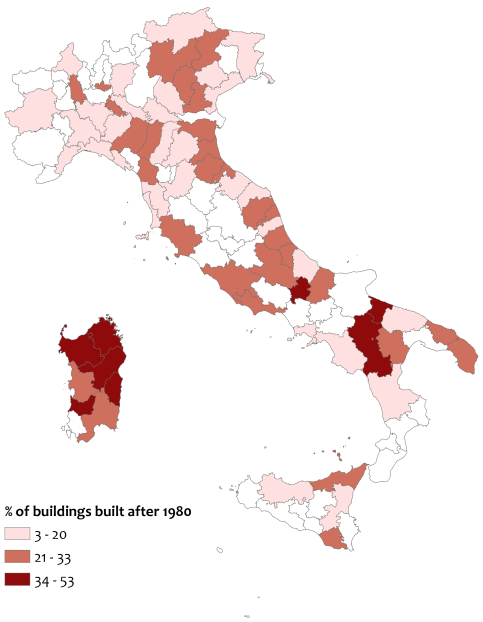


Figure 8 House material

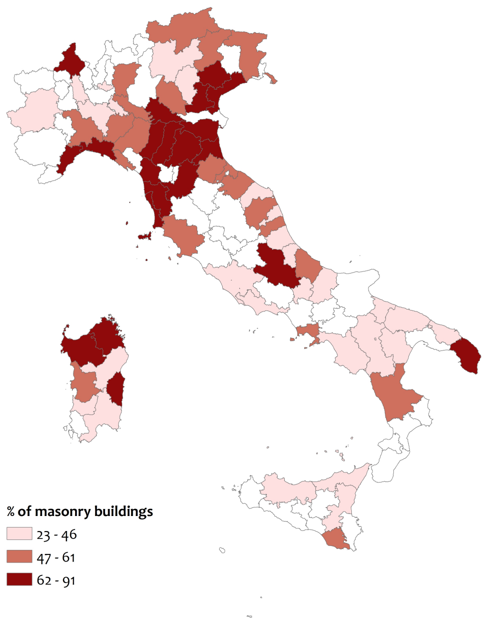
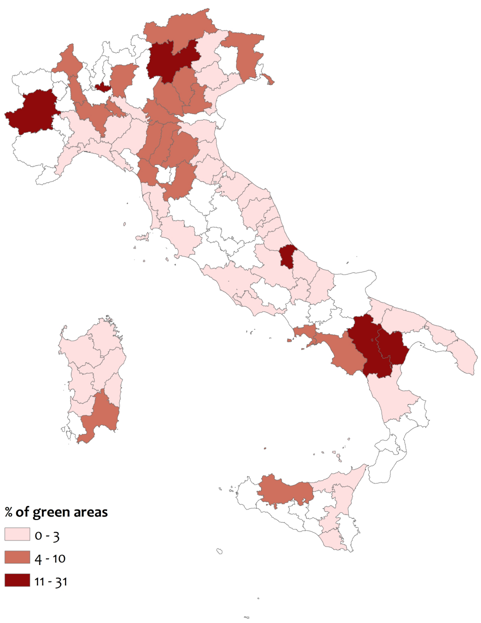


Figure 9 Green areas



3.3.1.2 FUNCTIONAL FEATURES

The functional features describe the functional subsystem of a city, which refers to the frequency and variety of activities occurring within the urban perimeter. For this research, I selected the following five functional variables: (1) land use mix, measured as the number of jobs per square kilometer; (2) concentration of manufacturing activities, (3) concentration of commercial activities, (4) concentration of public activities, and (5) concentration of touristic activities, all measured by the location quotient (LQ).

The choice of including in the dataset four variables identifying the economic base of a city comes from the consideration that urban economics may be a factor influencing the energy and carbon footprint of an urban settlement because different activities produce and consume different amount of energy (Habitat – UN, 2011). Nevertheless, the relationship between the economic base of a territory and CO_2 emissions have not been explored yet. Thus, including these variables in this analysis represents one small step forward in the scientific debate on the relationship between cities and energy consumption.

Manufacturing, commercial, public and touristic activities were selected based on the “economic base theory” (Andrews, 1953; Haig, 1928) that distinguishes between basic and non-basic activities: “a city's basic activity links the settlement with other portions of the earth's surface; non-basic endeavors link the settlement with itself” (Alexander, 1954). In particular, manufacturing and touristic activities can be considered basic o primary activities in the sense that they produce goods and services that brings new money into the local economy from outside the city, while commercial and public activities are non-basic or auxiliary activities because they serve the local demands within the population itself.

In order to identify the economic base of the 73 cities, I calculated the location quotient for each of the four economic activities. The location quotient, indeed, measures how concentrated a specific activity is in each city as compared to the full sample. Considering a city (*j*) and a specific economic activity (i), the mathematical expression of the location quotient is:

$$LQ_i = \frac{\frac{E_{ij}}{E_j}}{\frac{E_{is}}{E_s}}$$

where:

E_{ij} = city's employment in activity *i;*

E_j = city's total employment;

E_{is} = full sample employment in activity *i;*

E_s = full sample total employment.

The more the LQ is above one, the more the city is specialized in that economic activity. The closer the LQ drops towards zero, the more likely that sector does not represent the economic base of that city. In order to reflect this interpretation of the LQ in the corresponding maps (Figures 11-14), I did not use the natural break function as for the other variables. More specifically, after analyzing the data, I distinguished between specialized cities (LQ>1) and not-specialized cities (LQ<1) in the case of manufacturing, commercial and public activities, while I included a third class, that of very specialized cities (LQ>3), in the case of touristic activities, because of the very high value of one city.

Land use mix

Similarly to housing density and green areas, the standard deviation of this variable is higher than the average, which indicates that the concentration of jobs per square kilometer considerably differs among Italian capital cities. Moreover, remarkable correspondences between housing density and jobs concentration are evident when looking at the maps (Figure 5 and 10). Indeed, the seven cities with the highest concentration of jobs – Turin, Napoli, Firenze, Bergamo, Monza, Pescara and Bologna – are also the densest ones. This similarity also occurs between cities with fewer jobs and fewer dwellings per square meter (e.g. Tempio Pausania, Villacidro, etc.), and is statistically proven (Pearson's correlation coefficient = 0.96).

Concentration of manufacturing activities

The concentration of manufacturing activities as measured by the LQ shows that this sector represents the economic base of thirty Italian capital cities (41% of the full sample). The map (Figure 11), moreover, illustrates that these cities are mainly located in the northern part of the country, more precisely in Lombardy and Emilia Romagna.

Concentration of commercial activities

Commercial activities include retailing of every type of good.

Slightly higher than the number of cities specialized in manufacturing, the number of cities specialized in commercial activities (thirty-four) represent almost 50% of the full sample. They are concentrated in the northern part of Italy and in Sardinia (Figure 12).

Concentration of public activities

The concentration of public activities was measured considering employment in all different public institutions, such as governments, hospitals and universities. Figure 13 shows that these types of activities are concentrated in the southern part of the country. By comparing this map with that of manufacturing (Figure 11), the overlap is evident: cities specialized in manufacturing are not specialized in public activities and vice versa. This negative association is also statistically demonstrated (Pearson's correlation coefficient = -0.78).

Concentration of touristic activities

About 35% of Italian capital cities are specialized in tourism. Given the touristic attractiveness of cities like Venice, Rome, Firenze, Pisa and Naples, we would expect these to have the highest concentration of touristic activities. Nevertheless, these cities do have a location quotient higher than 1, but the city with the greatest specialization in this sector is Rimini (LQ=8.67), as shown in Figure 14. The city, indeed, can be considered the biggest beach destination of the Adriatic Riviera and every summer it attracts over two millions of tourist presences.

Figure 10 Land use mix

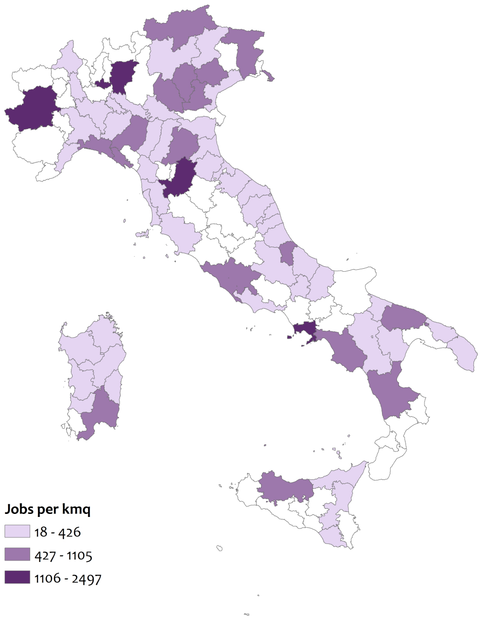


Figure 11 Concentration of manufacturing activities

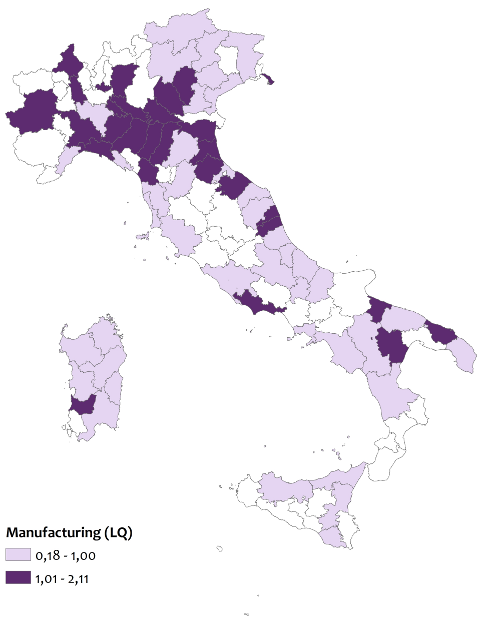


Figure 12 concentration of commercial activities

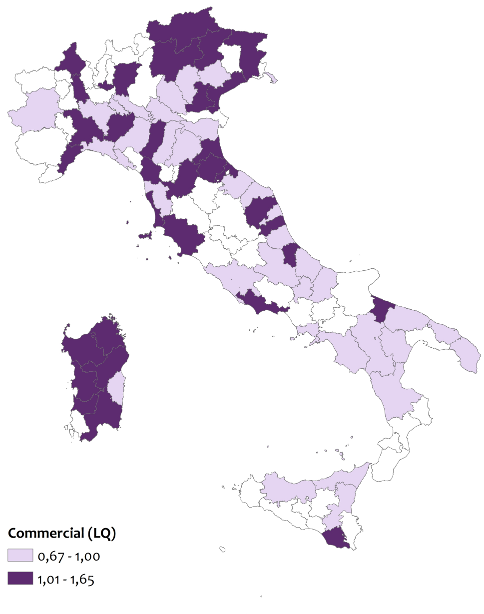


Figure 13 Concentration of public activities

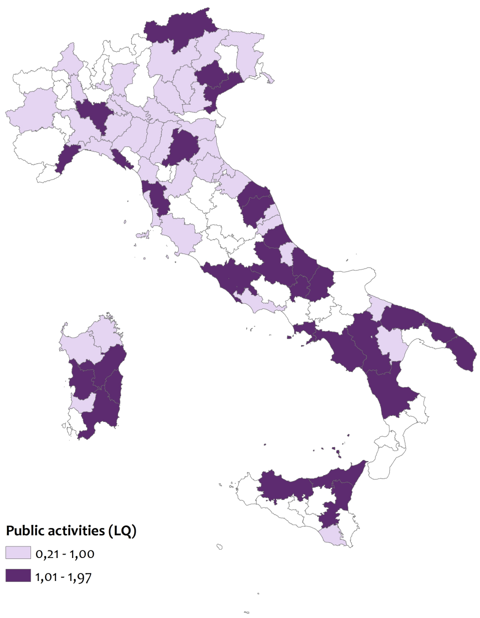
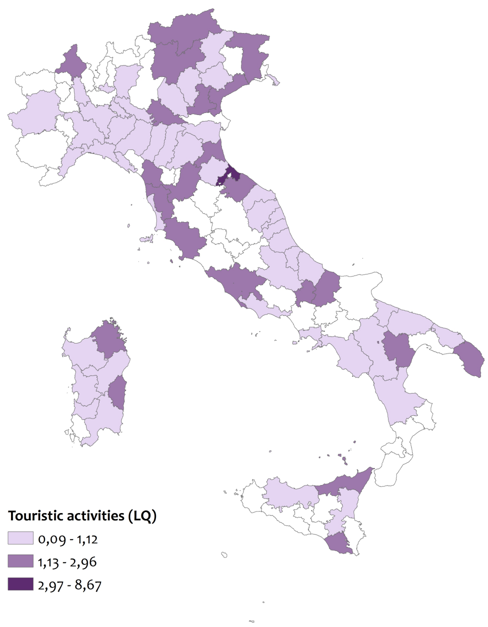


Figure 14 Concentration of touristic activities



3.3.1.3 GEOGRAPHICAL FEATURES

The geographical features describe the climatic and topographic characteristics of Italian capital cities, and in particular they comprise: (1) degree days, calculated as the sum of daily positive differences between a temperature of reference of 20°C (for Italy) and the average daily outdoor temperature, extended to the whole year; (2) coastal location, measured as a binary variable with the value "0" for coastal cities and the value "1" for inland cities; (3) topography, measured as a binary variable with the value "0" for valley cities and the value "1" for mountain cities.

Degree days

Degree-days were calculated using the following formula provided by UNI EN ISO 15927-6:2008, on a base temperature set to 20° (UNI 9019:1987):

$$DD = \sum_{e=1}^{n} (20° - T_e)$$

where:

n = number of days of the heating season, set by law for each Italian city;

T_e = daily mean external air temperature.

This indicator, therefore, measures the thermal requirement of a specific city. More specifically, higher values of DD correspond to a higher number of heating days (i.e. colder cities), whereas warmer territories are characterized by lower values of DD.

Looking at the map (Figure 15), not surprisingly, the highest values of degree-days are concentrated in the northern part of the country, while middle values are concentrated in the central part and the lowest ones in the South. The only exceptions are three southern cities (L'Aquila, Potenza and Campobasso) that have high values of DD because of their high elevation.

Coastal location

The research sample includes thirty-two coastal cities (43.8% of the full sample) and forty-one inland cities (Figure 16). Not surprisingly, coastal cities are concentrated in the two main islands: five in Sardinia and four in Sicily. Furthermore, about 47% of coastal cities are located on the Tyrrhenian coast, and about 41% on the Adriatic side.

It is important to remind the reader that the different areas colored in the maps do not correspond to the city's perimeter, but to the province. Hence, when the province is by the ocean but its color provides the opposite information, it means that the respective capital city is inland and not by the coast.

Topography

Mountain and valley cities were classified based on Law n. 991/1952, which defines as mountain those cities having more than 80% of the land area located more than 600 meters above sea level and where the difference between the highest and lowest elevations is not lower than 600 meters.

This definition allowed the construction of the map (Figure 17) where the fifty mountain cities (68.5%) and the twenty-three valley cities (31.5%) are represented in different colors. Differently from the map on degree-days, the distribution of mountain cities does not follow a geographical criterion. Moreover, comparing the maps of the three geographical variables, no evident overlaps comes out. Therefore, the general idea that mountain cities are colder and inland does not find correspondence in the data. Some mountain cities (e.g. Olbia, in Sardinia), indeed, are coastal ones, and some of the coldest cities (e.g. Pavia and Lodi, in Lombardy) are not mountain cities.

Figure 15 Degree days

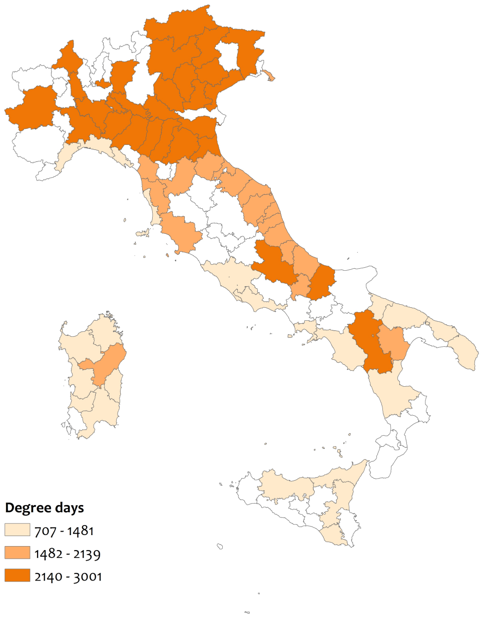


Figure 16 Coastal location

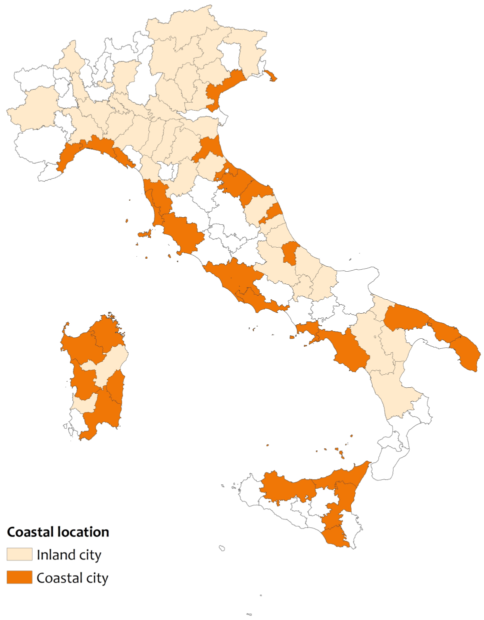
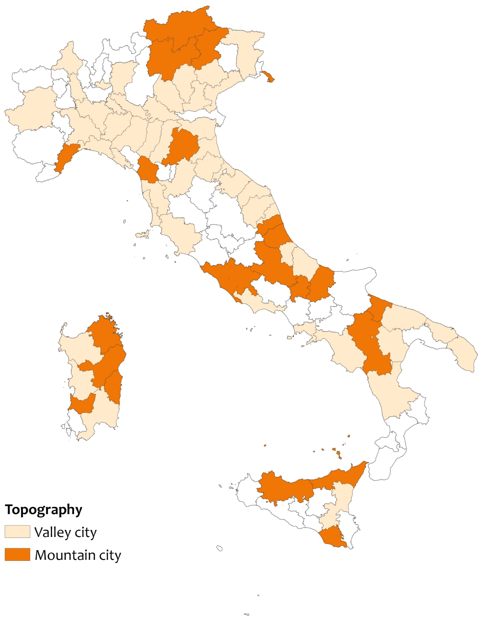


Figure 17 Topography



3.3.1.4 SOCIO-ECONOMIC FEATURES

The five socio-economic variables included in the dataset describe the main characteristics of the population living in the 73 Italian capital cities that may affect energy consumption and CO_2 emissions on an urban scale. In particular, these include: (1) income, measured as the average income per capita; (2) car ownership, calculated as the number of private cars per 1.000 person; (3) household composition, measured as the percentage of population younger than 15 years old; (4) education, measured as the number of graduates per 1.000 person; (5) ethnicity, measured as the percentage of foreign citizens.

Income

Data on average income per capita were calculated using the individual income tax return data at 2012, provided by the Ministry of Economy and Finance. In 2012, the average income in Italy was 19.660€, similar to the sample's one that is 20.411€.

In order to compare Italian data with European data and see how the local economic standard of living fits into a wider picture, it is possible to look at the at-risk-of-poverty rate[6] by Eurostat. In 2012, while in Europe less than one fourth of the population (24.7%) was viewed as being at-risk-of-poverty, in Italy almost one third (29.9%), corresponding to about eighteen millions of citizens.

The map (Figure 18) representing the geographical distribution of this variable provides interesting findings. It shows, indeed, a clear divide between a leader North and a laggard South. With only two exceptions (Rome and Cagliari), northern cities perform much better than southern ones in economic terms. Some cities in the North have even an average income double than those located in the South (e.g. Monza in Lombardy (26.744€) and Andria in Apulia (12.183€)).

Car ownership

Considering the full sample, the average number of cars per 1.000 inhabitants at 2011 is 614, slightly lower than the data for the whole country (625 cars per 1.000 person). In 2011, Italy was in the top three European countries for car ownership, following Liechtenstein (756 cars per 1.000 person) and Luxembourg (675 cars per 1.000 person).

[6] *"The at-risk-of-poverty threshold is set at 60 % of national median equivalised disposable income. It is often expressed in purchasing power standards (PPS) in order to take account of the differences in the cost of living across countries."*(http://ec.europa.eu/eurostat/statistics-explained/index.php/Income_distribution_statistics)

Considering the research sample and the geographical distribution of this variable (Figure 19), car ownership is higher in central cities and in Sardinia (from 0.65 to 0.75 car per person). The city with the lowest number of cars per capita is Venice (0.41).

Household composition

I used the percentage of population younger than 15 years old as a proxy of household composition. These data provide information about the population structure of each city included in the research sample. In particular, with 12.9% of children up to 14 years old, the population of the 73 Italian capital cities is slightly older than that of the whole country (14.1% of citizens with less than 15 years) and that of Europe (15.7%).

Figure 20 shows that the spatial distribution of this variable does not follow geographical criteria, thus meaning that cities with the highest concentration of children are uniformly distributed across the country, from the North to the South. However, three Apulian cities, namely Andria (17.5%), Barletta (16.6) and Trani (15.8%) occupy the podium with the highest concentration of young population. On the other hand, Cagliari (9.9%), Ferrara (10.5%) and Pavia (10.8%) score the lowest share of children.

Education

The number of graduates per 1.000 inhabitants corresponds to the population with at least either a university or a tertiary education. The average value of this indicator for the research sample is 151 graduates per 1.000 person, and it is higher than the number of graduates for Italy, which is about 106 graduates per 1.000 person. This finding suggests that people living in major cities have a higher education level than those living in smaller towns. From a geographical point of view, values are uniformly distributed across Italy (Figure 21).

Ethnicity

Ethnicity is measured as the share of foreign residents in each of the 73 Italian capital cities. On average, a city included in the sample has 7.23% of the population being foreign born. The average value for Italy is 9.70%, thus meaning that foreigners find capital cities less attractive than the other ones. Another interesting finding comes from a look at Figure 22, where it is evident that the country is divided into two blocks: North and South, where northern capital cities show the highest concentration of foreign residents (e.g. Piacenza 15.4%, Reggio nell'Emilia 14.8% and Vicenza 14.5%).

Figure 18 Income

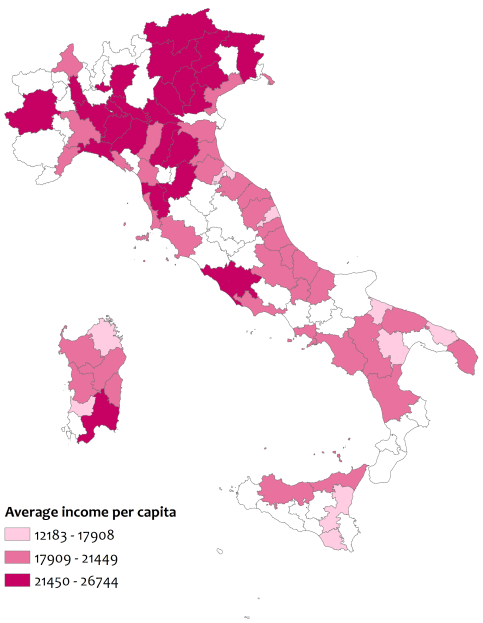


Figure 19 Car ownership

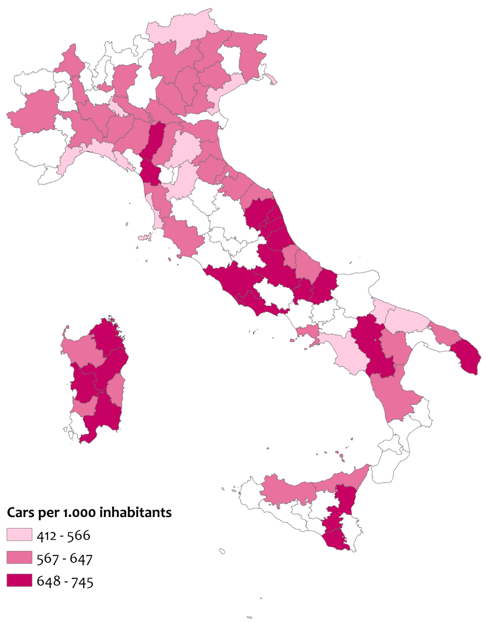


Figure 20 Household composition

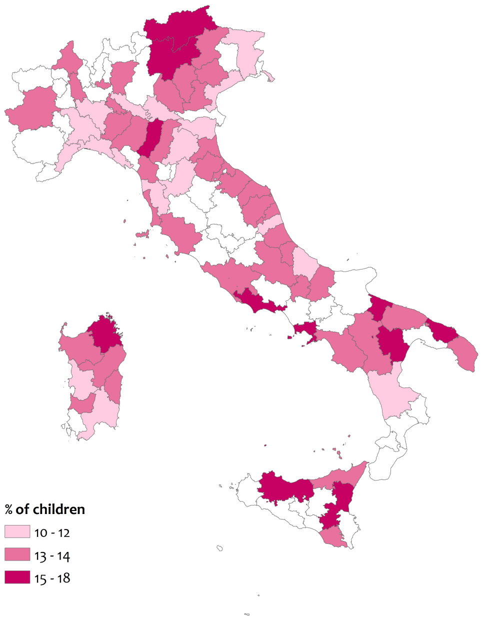


Figure 21 Education

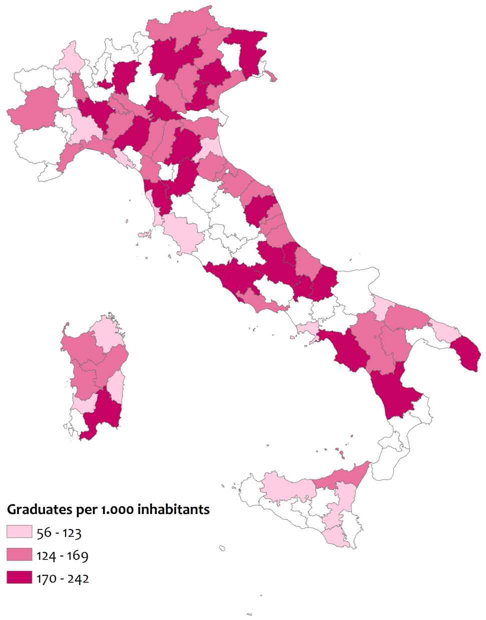
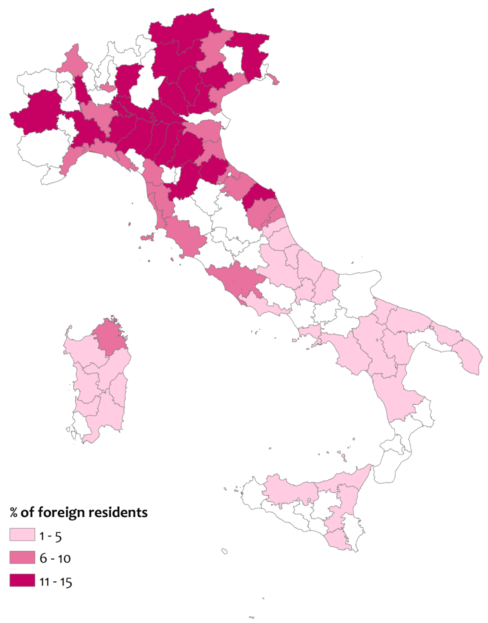


Figure 22 Ethnicity



3.3.2 Energy variables

In order to investigate the relationship between cities and energy consumption, after having collected data on urban variables I proceeded to gather information on the energy and carbon footprint of the Italian capital cities included in the sample.

According to Eurostat, in 2015 Italy final energy consumption was 116 Mtoe, equivalent to 10.7% of total EU-28 final energy consumption, which was 1.082 Mtoe. In particular, a more detailed analysis of the final end use of energy in Italy in 2015 shows that three main sectors prevail (Figure 23): transport (34%), buildings (28%) and industry (23). Similarly, the three dominant categories in the EU-28 final energy use in 2015 are (Figure 24): transport (33%), industry (25%) and buildings (25%).

Initially, for the collection of energy data, I used data on gas and electricity consumption for residential use provided by the ISTAT for the entire sample of 116 Italian capital cities. However, such data were not sufficient for the purpose of this research because they did not include any information about energy consumption from transport, which is the other main sector responsible for energy consumption at urban level together with buildings. Therefore, I decided to use a different data source – the Sustainable Energy Action Plans (SEAP) – that provided more detailed information, but for a smaller number of capital cities. Indeed, only 73 out of 116 Italian capital cities had approved a SEAP by the time I started the data collection (April, 2015).

Figure 23 Final energy consumption by sector, Italy (2015)
Source: author elaboration on Eurostat data (online code: nrg_100a)

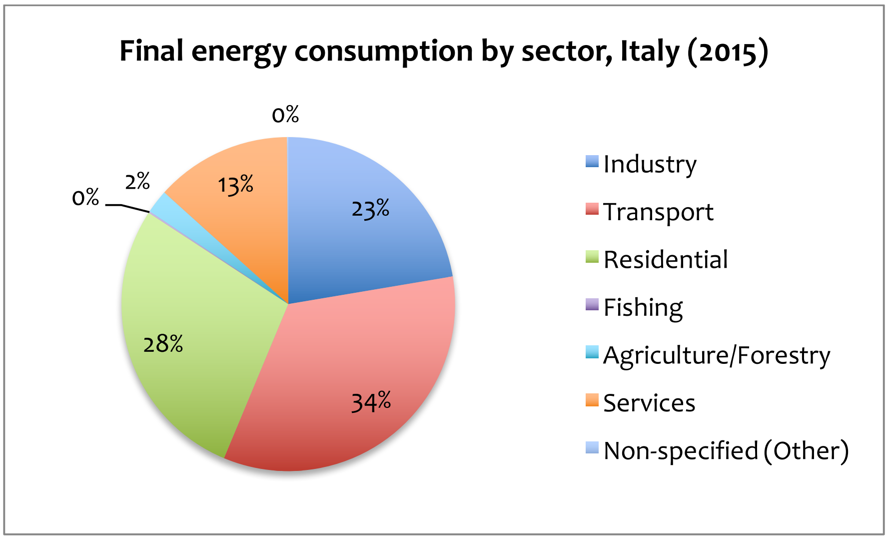
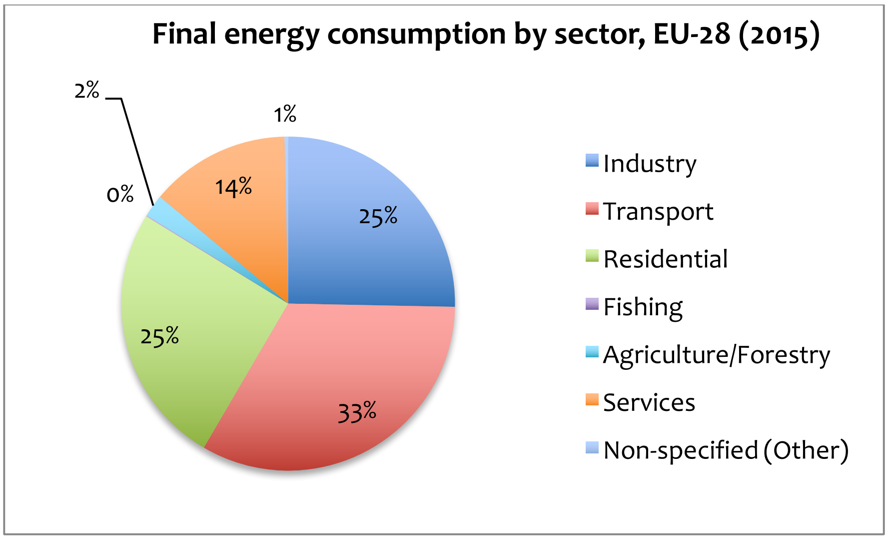


Figure 24 Final energy consumption by sector, EU-28 (2015)
Source: author elaboration on Eurostat data (online code: nrg_100a)



The data collection procedure required considerable time, because data were manually collected using the SEAP approved by each capital city. In particular, each SEAP provided the information on final energy consumption (MWh) and CO_2 emissions (t) by sector. The sectors considered were the following two: transport and residential buildings, tertiary buildings and municipal buildings. A more detailed description of SEAPs' objectives and contents follows.

3.3.2.1 THE SUSTAINABLE ENERGY ACTION PLAN

The European Commission has launched the Covenant of Mayors initiative in 2008, after the adoption of the 2020 EU Climate and Energy Package. This initiative aims to *"endorse and support the efforts deployed by local authorities in the implementation of sustainable energy policies"*. The Covenant of Mayors successfully managed to involve a great number of local and regional authorities, which committed to (1) prepare a Baseline Emission Inventory (BEI) and (2) develop and implement a Sustainable Energy Action Plan (SEAP) within the year following their formal adhesion to the initiative.

The BEI quantifies the amount of CO_2 emissions due to energy consumption in the territory of the Covenant signatory. Based on this information, the SEAP identifies the appropriate actions aimed at reducing CO_2 emissions and final energy consumption by end users, including both the public and private sectors (European Commission, 2010).

Up to now (January 2017), 3.014 Italian cities have submitted a SEAP and 2.559 of these plans have been accepted. When I finished collecting data for this research (April 2015), 2.433 Italian cities had submitted a SEAP and 1.440 of these plans had been accepted. In particular, 73 Italian capital cities had submitted a SEAP by April 2015, and therefore represent the final sample of this research. I downloaded the 73 action plans and used the BEIs for collecting data on the amount of CO_2 emitted due to energy consumption.

Elaborating a BEI can be a challenging task, especially because a large amount of data referred to different sectors is needed, which is not always easy to access and elaborate. Therefore, detailed information for compiling a BEI can be found in the SEAP guidebook – "How to develop a Sustainable Energy Action Plan" – by the European Commission (2010). These guidelines include a large number of recommendations but do not provide a single universally agreed methodology for the preparation of the BEI, on the contrary they give the local authority *"the flexibility to use any approach or tool that it considers appropriate for the purpose"*. However, they also encourage local authority *"to use emission factors that are in line with those of the Intergovernmental Panel on Climate Change (IPCC) or European Reference Life Cycle Data base (ELCD)"*.

Together with the SEAP guidebook, the European Commission provides the SEAP template tables related to the Baseline Emission Inventory for reporting the results of the BEI (Figure 25-26). These tables represented the main data source used for collecting energy data from the SEAPs of the 73 capital cities initially included in the research sample.

Figure 25 SEAP template table for BEI inventory – final energy consumption
Source: Annex II, "*How to develop a Sustainable Energy Action Plan*", European Commission (2010)

A. FINAL ENERGY CONSUMPTION (MWh)																
CATEGORY	ELEC-TRICITY	HEAT/COLD	FOSSIL FUELS								RENEWABLE ENERGIES					TOTAL
			Natural gas	Liquid gas	Heating oil	Diesel	Gaso-line	Lignite	Coal	Other fossil fuels	Plant oil	Biofuel	Other biomass	Solar thermal	Geo-thermal	
BUILDINGS, EQUIPMENT/FACILITIES AND INDUSTRIES																
Municipal buildings, equipment/facilities																
Tertiary (non municipal) buildings, equipment/facilities																
Residential buildings																
Municipal public lighting																
Industries (excluding industries involved in the EU Emission trading scheme – ETS)																
Subtotal buildings, equipments/facilities and industries																
TRANSPORT																
Municipal fleet																
Public transport																
Private commercial transport																
Subtotal transport																
TOTAL																
MUNICIPAL PURCHASES OF CERTIFIED GREEN ELECTRICITY (IF ANY) (MWh)																
CO_2 EMISSION FACTOR FOR CERTIFIED GREEN ELECTRICITY PURCHASES (FOR LCA APPROACH)																

Figure 26 SEAP template table for BEI inventory – CO_2 emissions
Source: Annex II, "*How to develop a Sustainable Energy Action Plan*", European Commission (2010)

B. CO_2 OR CO_2 EQUIVALENT EMISSIONS (t)															
CATEGORY	ELECTRICITY	HEAT/COLD	FOSSIL FUELS								RENEWABLE ENERGIES				
			Natural gas	Liquid gas	Heating oil	Diesel	Gasoline	Lignite	Coal	Other fossil fuels	Biofuel	Plant oil	Other biomass	Solar thermal	Geothermal
BUILDINGS, EQUIPMENT/FACILITIES AND INDUSTRIES															
Municipal buildings, equipment/facilities															
Tertiary (non municipal) buildings, equipment/facilities															
Residential buildings															
Municipal public lighting															
Industries (excluding industries involved in the EU Emission trading scheme – ETS)															
Subtotal buildings, equipments/facilities and industries															
TRANSPORT															
Municipal fleet															
Public transport															
Private commercial transport															
Subtotal transport															
OTHER															
Waste management															
Waste water management															
Please specify here ohter missions															
TOTAL															
CORRESPONDING CO_2-EMISSION FACTORS IN [t/MWh]															
CO_2 EMISSION FACTOR FOR ELECTRICITY NOT PRODUCED LOCALLY [t/MWh]															

In particular, while all 73 SEAPs provided data on CO_2 emissions, only 61 SEAPs provided data on energy consumption. For this reason, I considered CO_2 emissions by sector as energy variables, and I discarded data on final energy consumption. Moreover, the BEI quantifies the emissions that occurred in the baseline year, which is the year against which the achievements of the emission reductions in 2020 shall be compared. Different cities chose different baseline years, mainly depending on data availability. Therefore, I identified the most frequent baseline year in the sample, which is 2005 (corresponding to 35 out of 73 cities, equal to 48% of the sample), and I transformed the other data.

In order to report all emissions data to the same baseline year, I used time series data on greenhouse gas (GHG) emissions provided by the ISTAT for Italian regions. More specifically, for each region, the ISTAT offers data on GHG emissions for the following years: 1990, 1995, 2000, 2005, 2010. Based on these data, I calculated the annual change in emissions and I used this value for estimating the CO_2 emissions at 2005 for the 52% of the sample with a different baseline year. The following example better illustrates the procedure here adopted. Novara is a city in Piedmont region and the baseline year of its BEI was 1998; data by the ISTAT for Piedmont show that GHG emissions were 8.8 t per capita in 1995 and 9.8 t per capita in 2005, corresponding to a 10.4% increase during the ten years of reference, i.e. 1.04% annual increase, which corresponds to a 7.32% growth between 1998 and 2005.

At the end of the data collection procedure I had the following information for 73 Italian capital cities: (1) residential CO_2 emissions per capita; (2) transport CO_2 emissions per capita; (3) tertiary CO_2 emissions per capita; (4) municipal CO_2 emissions per capita; (5) total CO_2 emissions per capita, calculated as the sum of all previous emissions. Table 5 provides some descriptive statistics on CO_2 emissions by sector for the final sample.

Residential emissions

Comparative statistics show that the building sector is the most emitting one, among the four here considered. Residential CO_2 emissions per capita, indeed, represent almost 38% of total CO_2 emissions. Moreover, the geographical distribution of these emissions as reported in Figure 27 highlights that cities in the North of the country have higher average values of residential emissions per capita compared to those in the South, with the exception of Andria, which is located in the South (i.e. in Apulia) and is the second of the 73 cities emitting the most (3.3 t per capita) following Ravenna (3.6 t per capita). On the contrary, Campobasso (0.6 t per capita), Bari (0.6 t per capita), and Salerno (0.7 t per capita), are the cities with the lowest CO_2 emissions from buildings.

Transport emissions

With an average value of 1.78 t of CO_2 emissions per capita, transportation is the second dominant sector for emissions (36.5%). Differently from residential emissions, transport emissions do not seem to follow a geographical logic (Figure 28). Ravenna (4.6 t per capita), Trento (3.7 t per capita) and Potenza (3.6 t per capita) are the cities with the highest emissions per person, while Villacidro (0.1 t per capita), Tortolì (0.1 t per capita) and Messina (0.5 t per capita) are the cities with the lowest emissions per person.

Tertiary emissions

Tertiary emissions account for about 23% of total CO_2 emissions per capita. Similarly to residential emissions, they appear to be higher in the northern part of Italy and lower in the South (Figure 29). However, when looking at the performance of each city in the sample, the city with the highest tertiary CO_2 emissions per capita is Sassari (3.4 t per capita), which is located in Sardinia, followed by Cremona (2.2 t per capita) and Firenze (2.1 t per capita). On the other hand, two Apulia cities emit less than 0.1 t per capita, namely Barletta, Andria and Trani.

Municipal emissions

Municipal emissions account for only about 2.4% of total CO_2 emissions per capita, uniformly distributed across the country (Figure 30). The top three cities for municipal emissions are: Novara (0.4 t per capita), Barletta (0.4 t per capita) and Trani (0.4 t per capita). At the bottom of the ranking, there are Napoli (0.02 t per capita), Alessandria (0.03 t per capita) and Lecce (0.03 t per capita).

Table 5 Descriptive statistics on CO_2 emissions by sector for the sample of 73 Italian capital cities

Variable	Minimum	Maximum	Average	Standard deviation	Unit of measurement
Residential CO_2 emissions	0.59	3.57	1.85	0.70	t per capita
Transport CO_2 emissions	0.07	4.58	1.78	0.89	t per capita
Tertiary CO_2 emissions	0.00	2.35	1.13	0.53	t per capita
Municipal CO_2 emissions	0.00	0.40	0.11	0.08	t per capita
Total CO_2 emissions	1.95	8.77	4.87	1.47	t per capita

Figure 27 Residential emissions

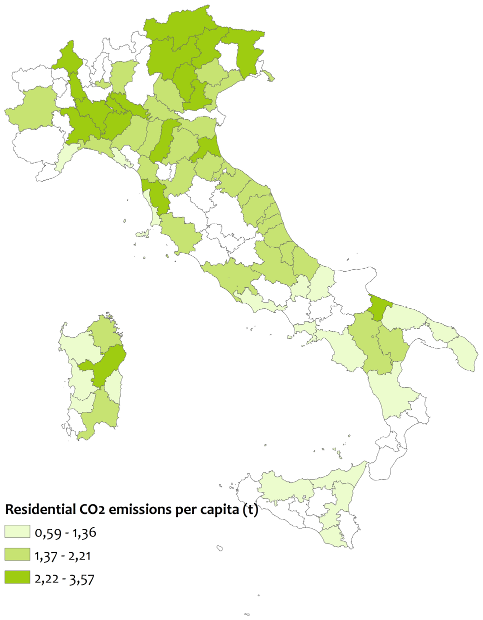


Figure 28 Transport emissions

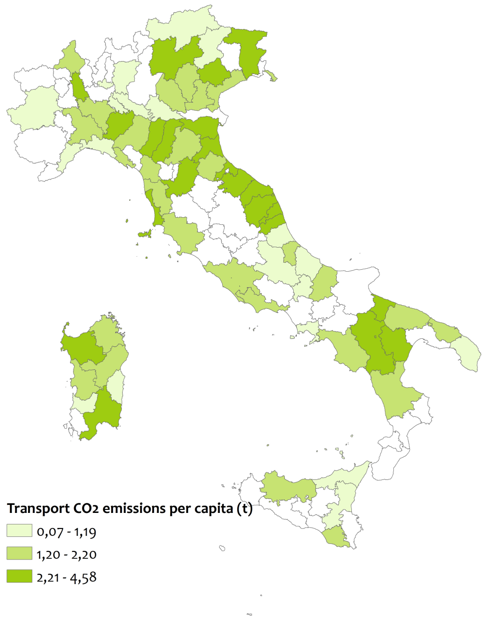


Figure 29 Tertiary emissions

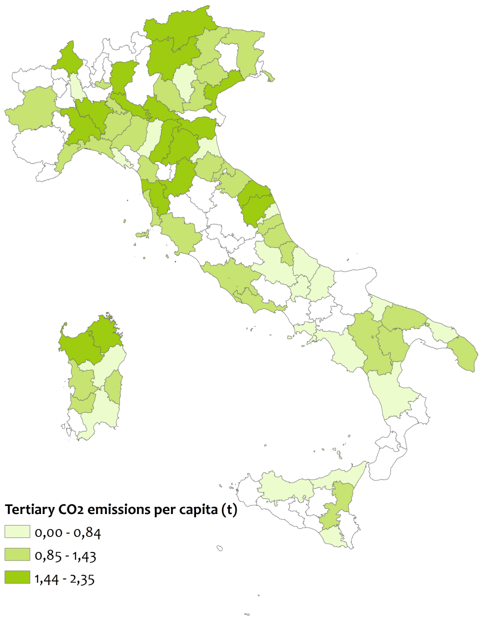


Figure 30 Municipal emissions

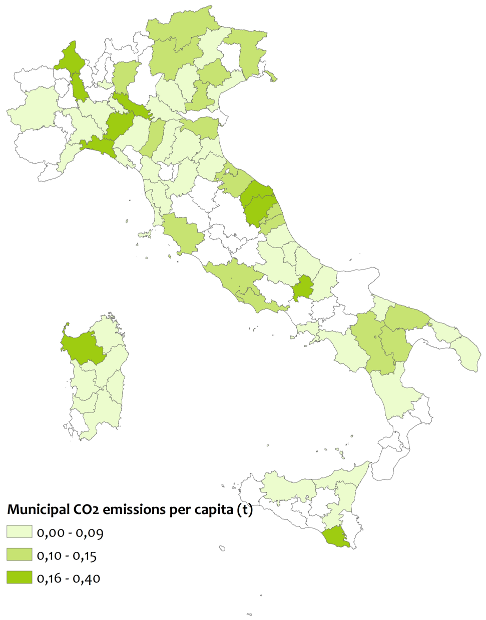
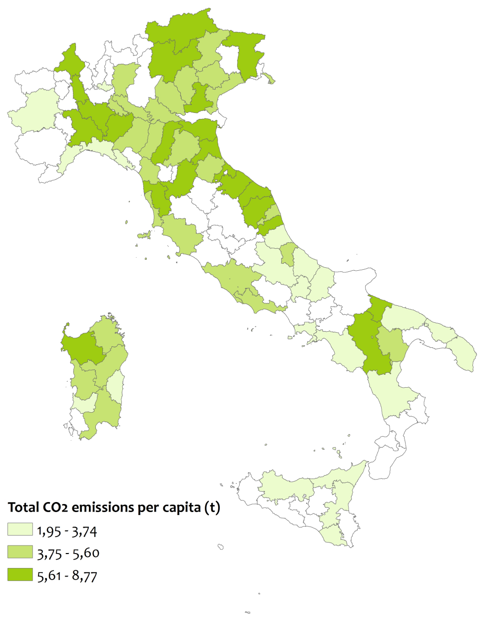


Figure 31 Total emissions



Total emissions

Total CO_2 emissions per capita were calculated as the simple arithmetic sum of residential, transport, tertiary and municipal CO_2 emissions per capita. This variable integrates together all previous information and provides a synthetic data on the carbon footprint of each of the 73 cities included in the final sample. More specifically, on average Italian capital cities emit 4.9 t per capita of CO_2 emissions; Trento, Ravenna and Ferrara rank in the top three with 8.8, 8.4 and 7.4 t per capita respectively. On the other side of the ranking, the most sustainable cities are Messina, Villacidro and Tortolì with 1.9, 2.1 and 2.4 t per capita respectively. These results confirm what Figure 31 shows, i.e. northern cities appear to have higher emissions per person than Southern ones.

3.4 Statistical tools for data analysis

Once the data collection procedure was complete, the data were analyzed according to the following steps: (1) exploratory data analysis; (2) correlation analysis; (3) regression analysis; (4) cluster analysis. Each of these statistical tools is described later herein, thus providing non-expert readers with the statistical knowledge necessary to better understand and interpret the results presented in Section 4.

3.4.1 Exploratory data analysis

Exploratory data analysis (EDA) was first introduced by Tukey's book "Exploratory Data Analysis" in 1977. This analysis approach aims to identify possible outliers, trends and patterns in the data, that might not have been anticipated. EDA is an important first step in any data analysis. Understanding where outliers occurred and how urban and energy variables were distributed helped to design statistical models that yield meaningful results.

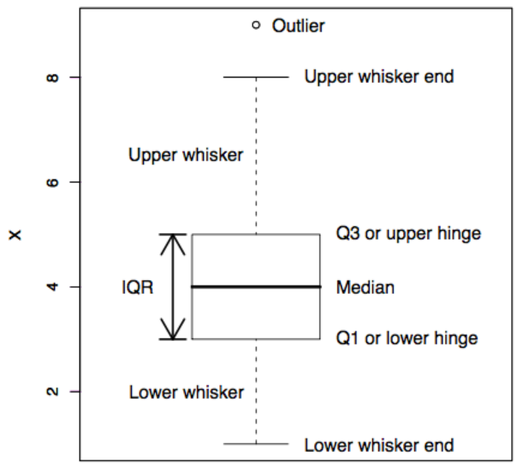
There are a number of graphical techniques used in EDA. For this research three graphical tools were used: (1) box plot; (2) histogram; (3) scatter plot. Therefore, the explanation that follows is limited to these three tools.

3.4.1.1 BOX-PLOT

The following description of box plots comes from Seltman (2012).

Boxplots are very good at presenting information about the central tendency, symmetry and skew, as well as outliers. The boxplot consists of a rectangular box bounded above and below by "hinges" that represent the quartiles Q3 and Q1 respectively, and with a horizontal "median" line through it. You can also see the upper and lower "whiskers", and a point marking an "outlier". The vertical axis is in the units of the quantitative variable.

Figure 32 Example of a box plot
Source: "Experimental design and analysis", Seltman (2012)



Any data value more than 1.5 IQRs beyond its corresponding hinge in either direction is considered an "outlier" and is individually plotted (Figure 32). Sometimes values beyond 3.0 IQRs are considered "extreme outliers" and are plotted with a different symbol. Each whisker is drawn out to the most extreme data point that is less than 1.5 IQRs beyond the *corresponding hinge. Therefore, the whisker ends correspond to the minimum and maximum values of the data excluding the "outliers". One additional thing you should notice on the plot is the symmetry of the distribution. Symmetry is appreciated by noticing if the median is in the center of the box and if the whiskers are the same length as each other.*

3.4.1.2 HISTOGRAM

The following description of histograms comes from Epstein (2014) and Seltman (2012).

The distribution of a variable tells us what values the variable takes and how often it takes these values. A histogram is a graph of the distribution of outcomes for a single numerical variable. The height of each bar is the number of observations in the class of outcomes covered by the base of the bar. All classes should have the same width and each observation must fall into exactly one class.

A graph is symmetric if the right and left sides of the histogram are approximately mirror images of each other. A graph is skewed to the right if the longer tail is on the right side (Figure 33). This is also called positively skewed. A graph is skewed to the left if the longer tail is on the left side. This is also called negatively skewed. An outlier is an individual data value that falls outside the overall pattern (Figure 34).

Many natural and social phenomena produce a continuous distribution with a bell-shaped (also called Gaussian or normal) distribution. The central limit theorem (CLT) explains why many real-world variables follow a Gaussian distribution.

Figure 33 Example of a positively skewed distribution
Source: "Exploring data distribution", Epstein (2014)

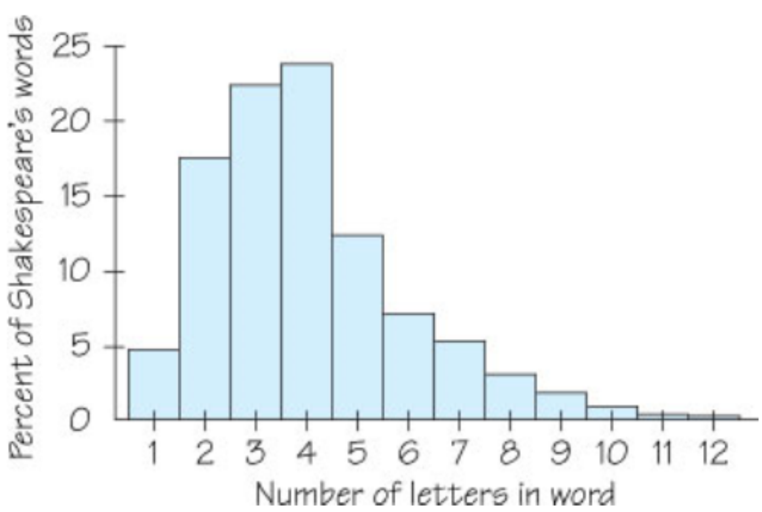


Figure 34 Example of a negatively skewed distribution and of an outlier
Source: "Exploring data distribution", Epstein (2014)

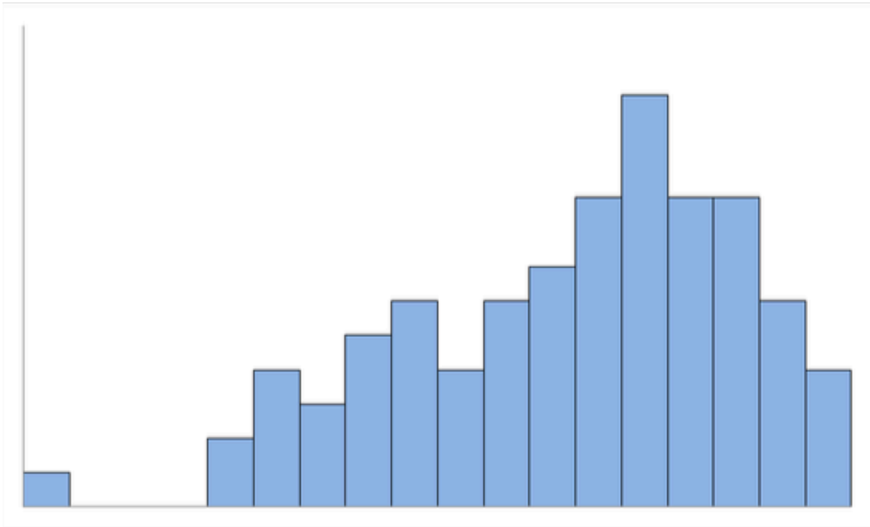
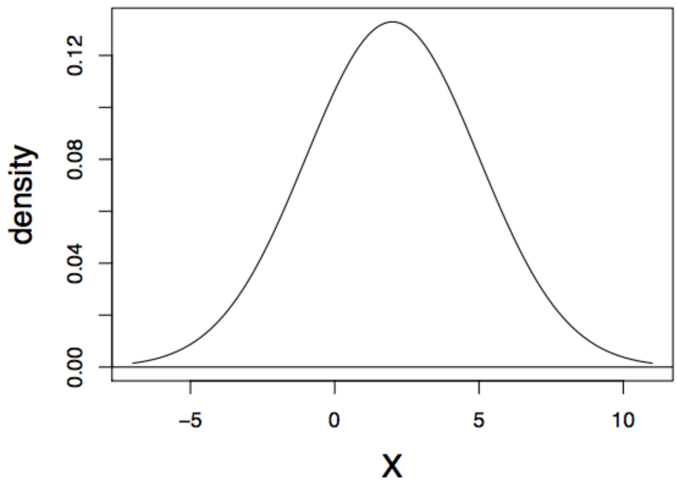


Figure 35 Example of a normal distribution
Source: "Experimental design and analysis", Seltman (2012)



In non-mathematical language, the "CLT" says that if we randomly sample a "large" number (say k) of independent values from that random variable, the sum or mean of those k values, if collected repeatedly, will have a Normal distribution. It takes some extra thought to understand what is going on here. The process I am describing here takes a sample of (independent) outcomes, e.g., the weights of all of the rats chosen for an experiment, and calculates the mean weight (or sum of weights). Then we consider the less practical process of repeating the whole experiment many, many times (taking a new sample of rats each time). If we would do this, the CLT says that a histogram of all of these mean weights across *all of these experiments would show a Gaussian shape, even if the histogram of the individual weights of any one experiment were not following a Gaussian distribution.*

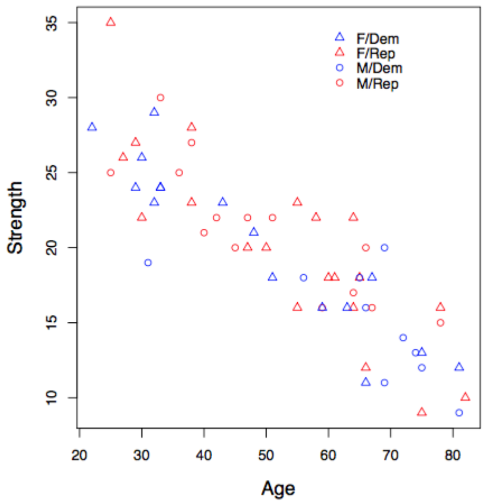
3.4.1.3 Scatterplots

The following description of scatterplots comes from Seltman (2012).

For two quantitative variables, the basic graphical EDA technique is the scatterplot which has one variable on the x-axis, one on the y-axis and a point for each case in your dataset. If one variable is explanatory and the other is outcome, it is a very, very strong convention to put the outcome on the y (vertical) axis.

One or two additional categorical variables can be accommodated on the scatterplot by encoding the additional information in the symbol type and/or color. An example is shown in Figure 35. Age vs. strength is shown, and different colors and symbols are used to code political party and gender.

Figure 36 Example of a scatterplot with two additional variables
Source: "Experimental design and analysis", Seltman (2012)



3.4.2 Correlation analysis

The following description of correlation analysis comes from Mukaka (2012).

The term correlation is sometimes used loosely in verbal communication. Among scientific colleagues, the term correlation is used to refer to an association, connection, or any form of relationship, link or correspondence. Webster's Online Dictionary defines correlation as a reciprocal relation between two or more things; a statistic representing how closely two variables co-vary; it can vary from -1 (perfect negative correlation) through 0 (no correlation) to +1 (perfect positive correlation).

In statistical terms, correlation is a method of assessing a possible two-way linear association between two continuous variables. Correlation is measured by a statistic called the correlation coefficient, which represents the strength of the putative linear association between the variables in question. It is a dimensionless quantity that takes a value in the range -1 to +1. A correlation coefficient of zero indicates that no linear relationship exists between two continuous variables, and a correlation coefficient of -1 or +1 indicates a perfect linear relationship. The strength of relationship can be anywhere between -1 and +1. The stronger the correlation, the closer the correlation coefficient comes to $\pm$ 1. If the coefficient is a positive number, the variables are directly related (i.e., as the value of one variable goes up, the value of the other also tends to do so). If, on the other hand, the coefficient is a negative number, the variables are inversely related (i.e., as the value of one variable goes up, the value of the other tends to go down).

There are two main types of correlation coefficients: Pearson's product moment correlation coefficient and Spearman's rank correlation coefficient. Pearson's product moment correlation is used when both variables being studied are normally distributed. This coefficient is affected by extreme values, which may exaggerate or dampen the strength of relationship, and is therefore inappropriate when either or both variables are not normally distributed. For a correlation between variables x and y, the formula for calculating the sample Pearson's correlation coefficient is given by:

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})\,(y_i - \bar{y})}{\sqrt{[\sum_{i=1}^{n}(x_i - \bar{x})^2][\sum_{i=1}^{n}(y_i - \bar{y})^2]}}$$

where x_i and y_i are the values of x and y for the ith individual.

In summary, correlation coefficients are used to assess the strength and direction of the linear relationships between pairs of variables. There is no attempt to establish one variable as dependent and the other as independent. Thus, relationships identified using correlation coefficients should be interpreted for what they are: associations, not causal relationships.

3.4.3 Regression analysis

The following description of regression analysis comes from Wilson et al. (2017).

Regression analysis is a statistical tool that allows us to describe the way in which one variable is related to another. This description may be a simple one involving just two variables in a single equation, or it may be very complex, having many variables and even many equations. From the simplest relationships to the most complex, regression analysis is useful in determining the way in which one variable is affected by one or more other variables.

In the simplest form of regression analysis, you have only the variable you want to model (or predict) and one other variable that you hypothesize to have an influence on the variable you are modeling. The variable you are is called the dependent variable (often represented as Y). The other variable is called the independent variable (often represented as X). The relationship or model you seek to find could then be expressed as:

$$Y = a + bX$$

In the expression above, a represents the intercept or constant term for the regression equation. The intercept is where the regression line crosses the vertical axis. Conceptually, it is the value that the dependent variable (Y) would have if the independent variable (X) had a value of zero.

The value of b tells you the slope of the regression line. The slope is the rate of change in the dependent variable for each unit change in the independent variable. If b has a positive value, Y increases when X increases and Y decreases when X decreases. On the other hand, if b is negative, Y changes in the opposite direction of changes in X.

How Can You Determine the Best Regression Line for Your Data? The most commonly used criterion for the "best" regression line is that the sum of the squared vertical differences between the observed values and the estimated regression line be as small as possible. This criterion is called "ordinary least squares" (OLS) regression.

3.4.3.1 Multiple Linear Regression

The following description of multiple linear regression analysis comes from Wilson et al. (2017).

In many (perhaps most) applications, the dependent variable of interest is a function of more than one independent variable. That is, there is more than one independent variable that can be used to explain variation in the dependent variable. In such cases, a form of OLS regression called multiple linear regression is appropriate. This is a straightforward extension of simple linear regression and is built upon the same basic set of assumptions. The general form of the multiple linear regression model is:

$$Y = f(X_1, X_2, \dots, X_k)$$

and the regression equation becomes:

$$Y = a + b_1X_1 + b_2X_2 + \cdots + b_kX_k$$

where Y represents the dependent variable and X_i represent each of the k different independent variables. The intercept, or constant, term in the regression is a and the b_i

terms represent the slope, or rate of change, associated with each of the k independent variables in the model.

The dependent variable (Y) used in a regression analysis will have some variability. Otherwise, there would be no reason to try to model Y. It would be convenient to have a measure of how much of that variation in Y is explained by the regression model. This is where the coefficient of determination (R^2) comes in handy. The coefficient of determination (R^2) gives the percentage of the variation in the dependent variable (Y) that is explained by the regression model. The worst possible explanatory power a model could have is to explain none of the variation in the dependent variable (R^2 = 0), and the best possible model would be one that explains all of the variations in the dependent variable (R^2 = 1.0). Hence, the coefficient of determination (R^2) will always be a value between 0 and 1. The closer it is to 0 the lower the explanatory power of the model, while the closer it is to 1 the greater is the explanatory power of the model.

3.4.3.2 INTERPRETING THE LOGLINEAR MODEL

The following description of how to interpret the loglinear model comes from Verbeek (2008).

An elasticity measures the relative change in the dependent variable due to a relative change in one of the x_i variables. Often, elasticities are estimated directly from a linear regression model involving the logarithms of most explanatory variables (excluding dummy variables), that is:

$$log\ y_i = (log x_i)'\gamma + \upsilon_i$$

We shall call this a loglinear model.

If x_{ik} is a dummy variable (or another variable that may take nonpositive values) we cannot take its logarithm and we include the original variable in the model. Thus we estimate:

$$log\ y_i = x_i'\beta + \varepsilon_j \qquad (*)$$

Of course, it is possible to include some explanatory variables in logs and some in levels. In (*) the interpretation of a coefficient β_k is the relative change in y_i due to an absolute change of one unit in x_{ik}.

3.4.4 Cluster analysis

The following description of cluster analysis comes from Friedman et al. (2001).

The goal of cluster analysis is to partition the observations into groups ("clusters") so that the pairwise dissimilarities between those assigned to the same cluster tend to be smaller than those in different clusters.

Hierarchical clustering methods require the user to specify a measure of dissimilarity between (disjoint) groups of observations, based on the pairwise dissimilarities among the observations in the two groups. As the name suggests, they produce hierarchical representations in which the clusters at each level of the hierarchy are created by merging clusters at the next lower level. At the lowest level, each cluster contains a single observation. At the highest level there is only one cluster containing all of the data.

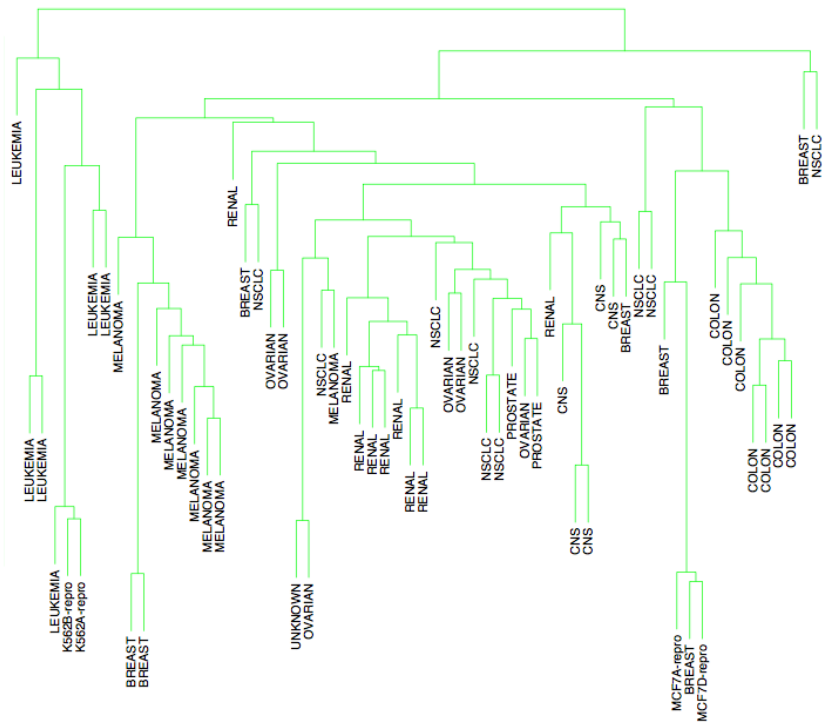
Strategies for hierarchical clustering divide into two basic paradigms: agglomerative (bottom-up) and divisive (top-down). Agglomerative strategies start at the bottom and at each level recursively merge a selected pair of clusters into a single cluster. This produces a grouping at the next higher level with one less cluster. The pair chosen for merging consist of the two groups with the smallest intergroup dissimilarity. Divisive methods start at the top and at each level recursively split one of the existing clusters at that level into two new clusters. The split is chosen to produce two new groups with the largest between-group dissimilarity. With both paradigms there are N – 1 levels in the hierarchy.

Each level of the hierarchy represents a particular grouping of the data into disjoint clusters of observations. The entire hierarchy represents an ordered sequence of such groupings. It is up to the user to decide which level (if any) actually represents a "natural" clustering in the sense that observations within each of its groups are sufficiently more similar to each other than to observations assigned to different groups at that level.

A dendrogram provides a highly interpretable complete description of the hierarchical clustering in a graphical format. This is one of the main reasons for the popularity of hierarchical clustering methods.

Cutting the dendrogram horizontally at a particular height partitions the data into disjoint clusters represented by the vertical lines that intersect it. These are the clusters that would be produced by terminating the procedure when the optimal intergroup dissimilarity exceeds that threshold cut value. Groups that merge at high values, relative to the merger values of the subgroups contained within them lower in the tree, are candidates for natural clusters. Note that this may occur at several different levels, indicating a clustering hierarchy: that is, clusters nested within clusters.

Figure 37 Example of a dendrogram
Source: "The Elements of Statistical Learning", Friedman et al. (2001)



3.5 Conclusions

The sample of cities selected for this research includes seventy-three Italian capital cities, which account for 21% of the total Italian population. Based on this sample and on the theoretical framework presented in Section 2, I selected a set of eighteen urban variables and five energy variables. Urban data were mainly collected from the ISTAT statistics, while I used the SEAPs as data source for energy variables. At the end of the data collection/elaboration procedure, I obtained a dataset consisting of twenty-three records for each of the seventy-three cities included in the sample. Four main statistical analysis were performed on this data seta: (1) exploratory data analysis; (2) correlation analysis; (3) multiple regression analysis; (4) cluster analysis. The following Section describes the main findings of these analyses and compares them with previous results found in the literature, in light of the theoretical framework developed in Section 2.

SECTION 4. RESULTS AND DISCUSSION

4.1 Introduction

Section 4 presents the main findings of this research, identifying those urban features that significantly affect energy consumption and resulting CO_2 emissions at urban scale. The dataset selected and described in the previous Section was explored and analyzed using different statistical methods (i.e. exploratory data analysis, correlation analysis, regression analysis, and cluster analysis), each of which provided useful insights into the complex relationship between cities and their carbon footprint. The cluster analysis was conducted using the SPAD software, while the statistical software SPSS 20 was used for the other analyses.

Exploratory data analysis (EDA) was performed in order to identify potential outliers and evaluate the distribution of the data, thus gaining a better knowledge of the research dataset and sample. After EDA, the data were analyzed using a correlation analysis, which provided two main results: (1) it enabled the measurement of the association between the eighteen urban variables in order to identify redundant information and, most importantly, significant interconnections amongst these factors; (2) it showed the significance of the linear relationship between each individual urban variable and CO_2 emissions by sector. Later, multiple linear regression analysis was performed to estimate the relationship between CO_2 emissions per capita and a number of determinants selected from the initial set of eighteen urban variables. Models were estimated separately for three categories of CO_2 emissions (i.e. residential, transport, and total CO_2 emissions per capita). CO_2 emissions by sector were regressed on eleven out of eighteen urban variables, i.e. housing density, house material, green areas, concentration of manufacturing activities, concentration of commercial activities, concentration of touristic activities, degree-days, topography, income, car ownership and household composition. Finally, a multivariate statistical analysis was performed in order to identify groups of cities with similar urban characteristics and compare their energy behaviors.

The results provided by these statistical analyses are discussed here considering previous findings found in the scientific literature, and in light of the theoretical framework proposed in Section 2. In particular, in order to address the research goal, the results are described highlighting the two main types of relationships affecting energy consumption and CO_2 emissions on a city scale, introduced and described in Section 2: namely, (1) the primary relationships between the urban features and CO_2 emissions, which directly affect a city's energy and carbon footprint; and (2) the secondary relationships among the different urban features, which indirectly (but significantly) affect a city's energy and carbon footprint.

Section 4 is structured as follows. Paragraph 4.1 presents the main results of the EDA. Paragraph 4.2 illustrates the correlation analysis and provides interesting findings on the associations among the urban variables included in the dataset. The results of the multiple regression analysis are later presented in paragraph 4.3 for each of the five regression models calculated. Section 4 ends with the description of the results of the cluster analysis.

4.2 Exploratory data analysis

4.2.1 Box plot

Visual data exploration using the box plot was very useful as a first step after completing the data collection procedure. This technique, indeed, allowed the identification of potential outliers. In particular, all the five box plots of CO_2 emissions per capita revealed one potential outlier: Torino. Therefore, the data for Torino were rechecked and a mistake in the unit of measurement indicated by the city's Sustainable Energy Action Plan was found. Figure 36 and 37 show the box plots of total CO_2 emissions per capita before and after fixing the data for Torino.

Figure 38 Box plot with Torino

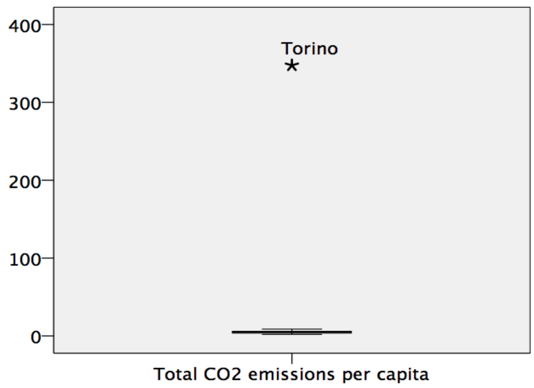
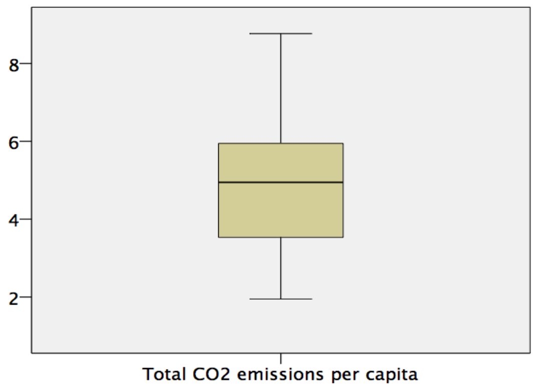


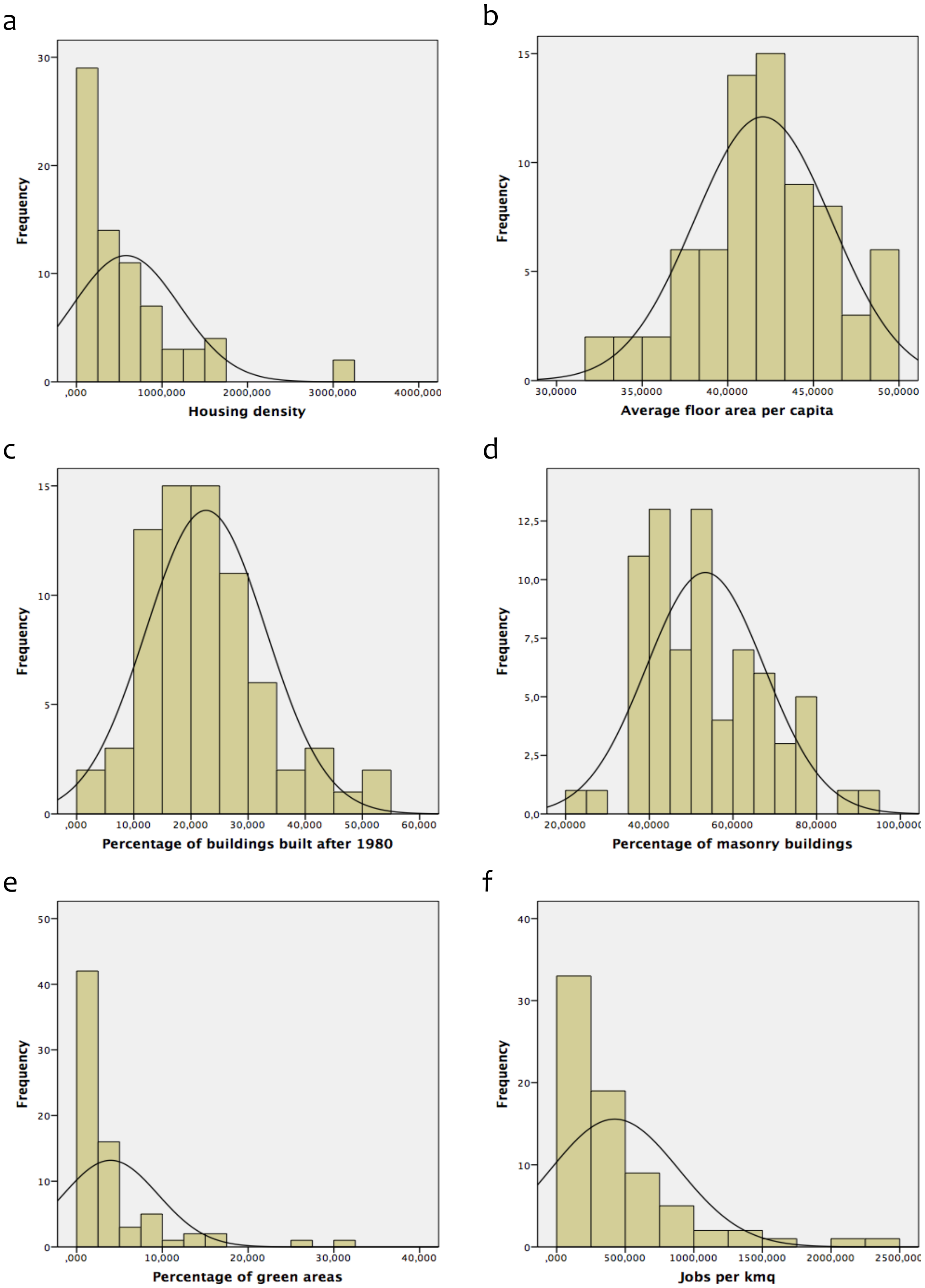
Figure 39 Box plot without Torino



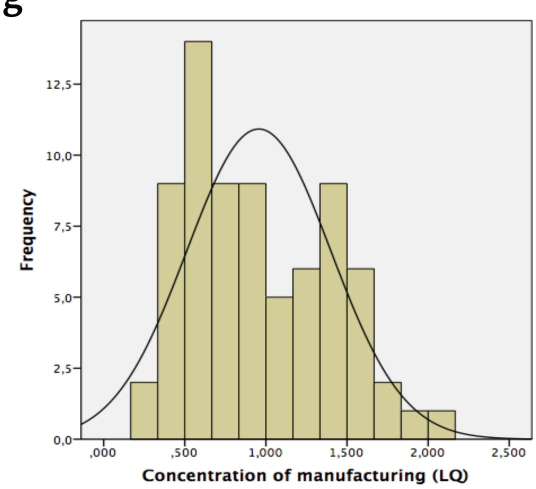
4.2.2 Histogram

Histograms were used to assess the normality of the distribution of variables (Figures 38 a-u). Four urban variables were found to be positively skewed: housing density (Figure 38 a), the percentage of green areas (Figure 38 e), jobs per square kilometer (Figure 38 f), and the concentration of touristic activities (Figure 38 j). These results confirm what was anticipated in paragraph 3.3.1, where the urban variables were described using descriptive statistics and the maps.

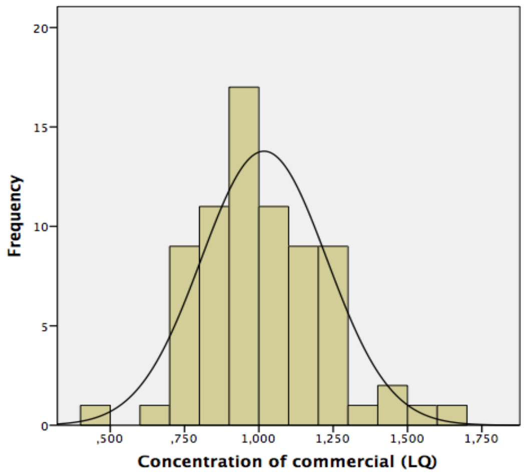
Figure 40 frequency distribution of variables



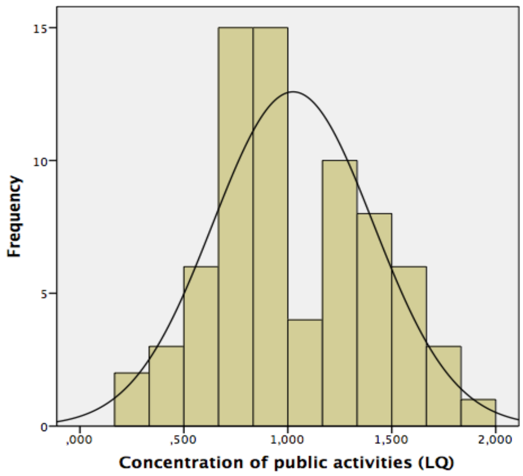
g


Frequency
0,0
2,5
5,0
7,5
10,0
12,5
,000
,500
1,000
1,500
2,000
2,500
Concentration of manufacturing (LQ)

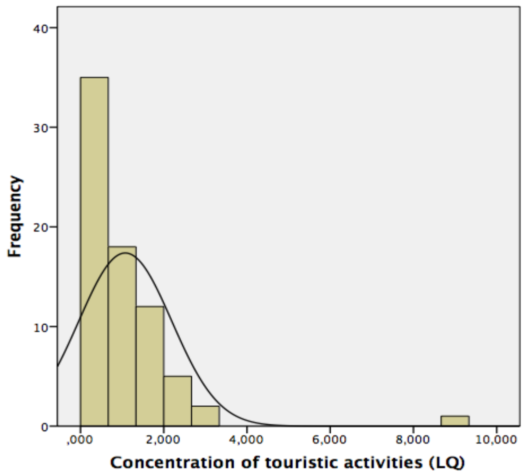
h


Frequency
0
5
10
15
20
,500
,750
1,000
1,250
1,500
1,750
Concentration of commercial (LQ)

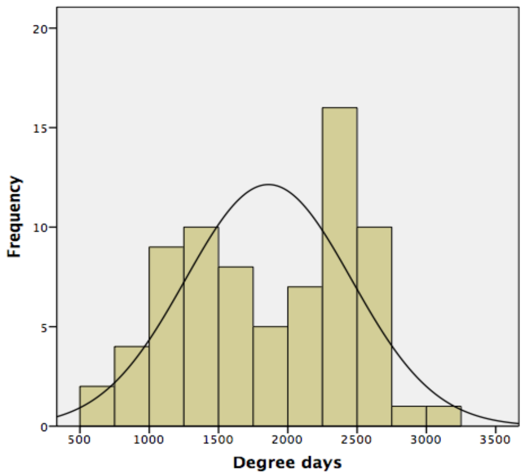
i


Frequency
0
5
10
15
,000
,500
1,000
1,500
2,000
Concentration of public activities (LQ)

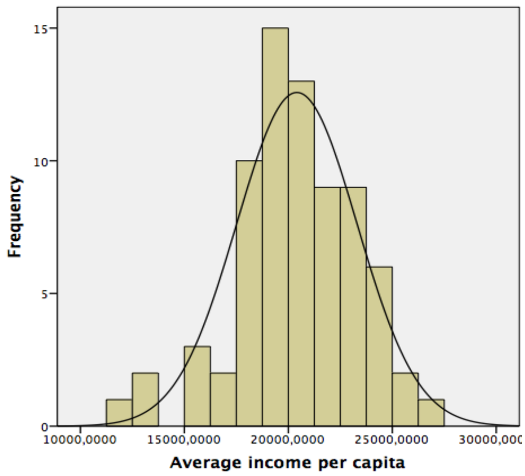
j


Frequency
0
10
20
30
40
,000
2,000
4,000
6,000
8,000
10,000
Concentration of touristic activities (LQ)

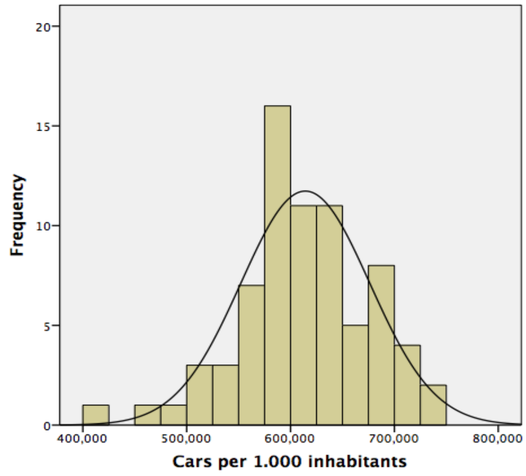
k


Frequency
0
5
10
15
20
500
1000
1500
2000
2500
3000
3500
Degree days

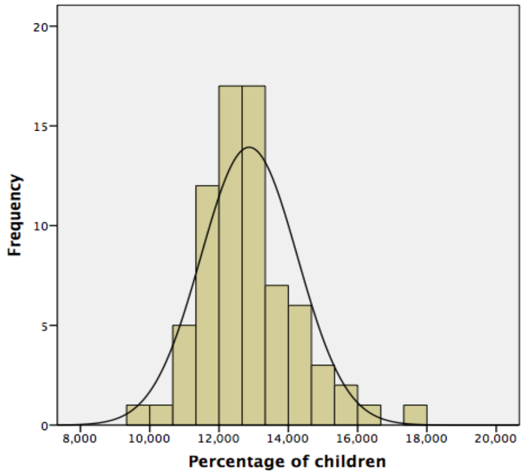
l


Frequency
0
5
10
15
10000,0000
15000,0000
20000,0000
25000,0000
Average income per capita

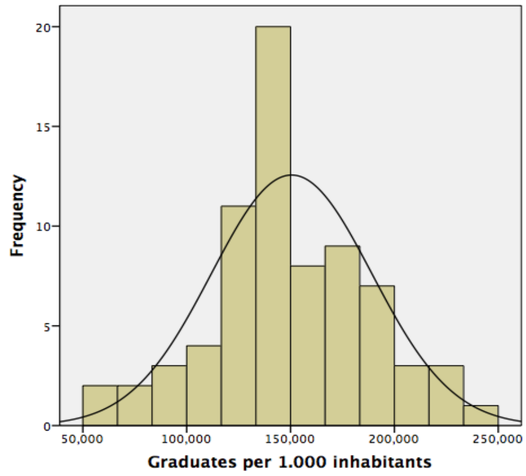
m


Frequency
0
5
10
15
20
400,000
500,000
600,000
700,000
800,000
Cars per 1.000 inhabitants

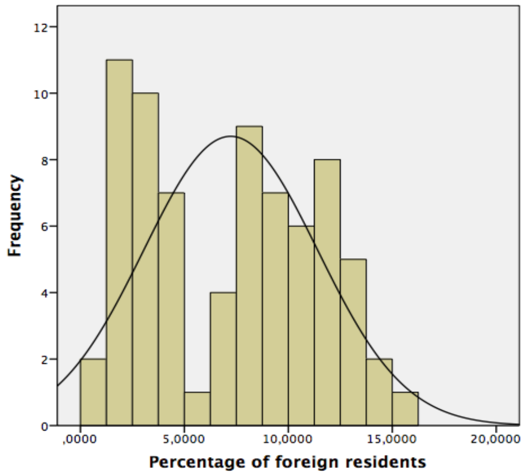
n


Frequency
0
5
10
15
20
8,000
10,000
12,000
14,000
16,000
18,000
20,000
Percentage of children

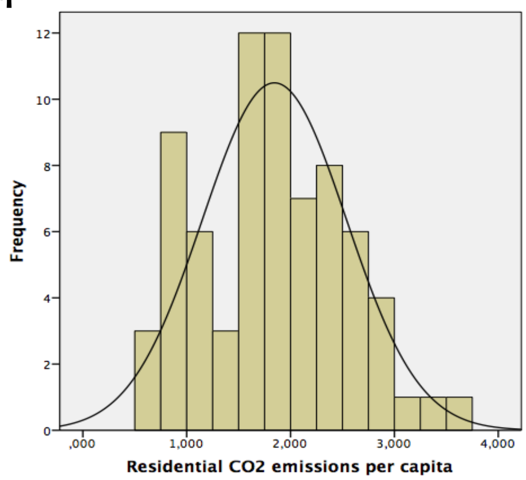
o


Frequency
0
5
10
15
20
50,000
100,000
150,000
200,000
250,000
Graduates per 1.000 inhabitants

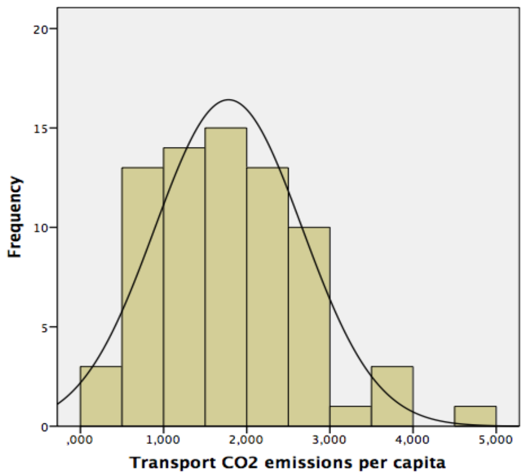
p


Frequency
0
2
4
6
8
10
12
,0000
5,0000
10,0000
15,0000
20,0000
Percentage of foreign residents

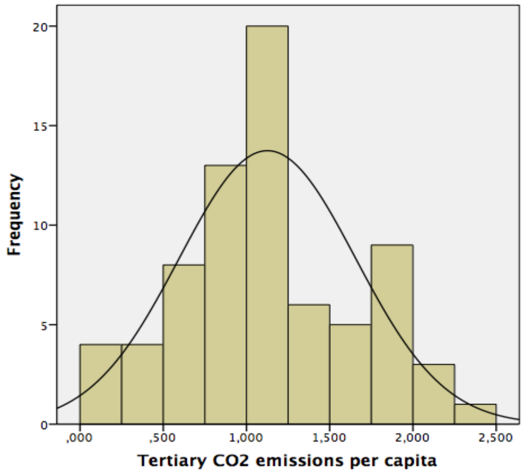
q


Frequency
0
2
4
6
8
10
12
,000
1,000
2,000
3,000
4,000
Residential CO2 emissions per capita

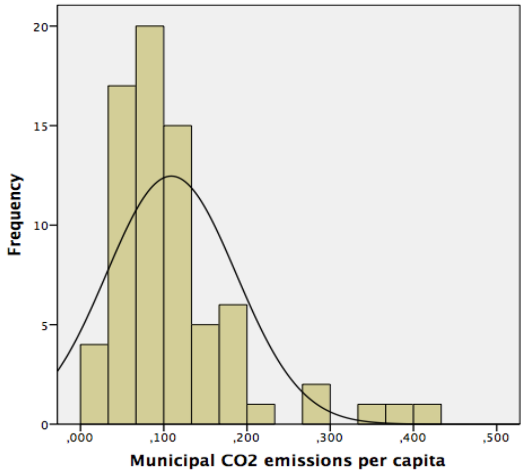
r


Frequency
0
5
10
15
20
,000
1,000
2,000
3,000
4,000
5,000
Transport CO2 emissions per capita

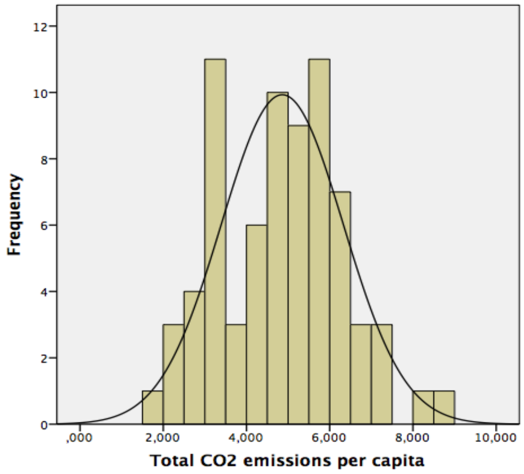
s



t



u



4.3 Correlation analysis

Correlation analysis was used to determine the strength and direction of the linear relationships between each pair of both urban and energy variables. Before performing the correlation analysis, the four variables positively skewed were log transformed to normalize their distribution and calculate Pearson's correlation coefficients. Table 6 reports the correlation coefficients for the twenty-three variables, and the values above 0.65 are in bold and red color. The threshold of 0.65 was later used to identify the best subset of urban variables to be included in the regression model to avoid multicollinearity issues.

Table 6 Correlation analysis, Pearson's correlation coefficients

	Housing density (ln)	Average floor area per capita	Percentage of buildings built after 1980	Percentage of masonry buildings	Percentage of green areas (ln)	Jobs per kmq (ln)	Concentration of manufacturing	Concentration of commercial	Concentration of public activities	Concentration of touristic activities (ln)	Degree days	Coastal location	Topography	Average income per capita	Cars per 1.000 inhabitants	Percentage of children	Graduates per 1.000 inhabitants	Percentage of foreign residents	Residential CO2 emissions	Transport CO2 emissions	Tertiary CO2 emissions	Municipal CO2 emissions	Total CO2 emissions
nsity (ln)	1.00																						
oor area per capita	0.00	1.00																					
e of buildings built after 1980	-0.66**	-0.12	1.00																				
e of masonry buildings	-0.12	0.35**	-0.11	1.00																			
e of green areas (ln)	0.61**	0.09	-0.45**	-0.14	1.00																		
mq (ln)	0.98**	0.07	-0.66**	-0.10	0.63**	1.00																	
ion of manufacturing	0.05	0.18	-0.18	0.02	0.13	0.11	1.00																
ion of commercial	-0.16	0.03	0.22	-0.06	-0.08	-0.16	-0.11	1.00															
ion of public activities	0.07	-0.18	0.01	-0.05	-0.06	0.01	-0.79**	-0.46**	1.00														
ion of touristic activities (ln)	-0.04	0.04	0.08	0.22	0.01	-0.04	-0.22	0.17	-0.07	1.00													
/s	0.09	0.46**	-0.21	0.08	0.36**	0.18	0.42**	-0.04	-0.38**	0.03	1.00												
ation	0.06	-0.36**	0.01	0.01	-0.18	-0.01	-0.30*	0.05	0.22	0.21	-0.69**	1.00											
y	-0.32**	-0.17	0.39**	-0.04	-0.31**	-0.36**	-0.29*	0.05	0.23	0.05	-0.08	-0.12	1.00										
come per capita	0.59**	0.52**	-0.42**	0.12	0.61**	0.67**	0.08	-0.13	0.00	-0.06	0.53**	-0.29*	-0.24*	1.00									
00 inhabitants	-0.41**	0.10	0.50**	-0.10	-0.24*	-0.39**	-0.19	0.04	0.13	0.05	-0.08	-0.12	0.24*	-0.14	1.00								
e of children	-0.15	-0.67**	0.31**	-0.28*	-0.15	-0.17	0.11	0.13	-0.18	-0.01	-0.23*	0.12	0.13	-0.46**	0.05	1.00							
per 1.000 inhabitants	0.40**	0.48**	-0.30*	0.00	0.42**	0.46**	-0.23*	-0.19	0.31**	0.04	0.31**	-0.24*	-0.16	0.73**	0.12	-0.58**	1.00						
e of foreign residents	0.42**	0.50**	-0.46**	0.23*	0.45**	0.50**	0.47**	-0.05	-0.43**	0.07	0.69**	-0.37**	-0.33**	0.69**	-0.25*	-0.19	0.31**	1.00					
l CO2 emissions	-0.06	0.23	-0.10	0.09	0.17	-0.01	0.34**	0.02	-0.34**	-0.01	0.61**	-0.42**	-0.08	0.22	-0.15	0.06	0.05	0.49**	1.00				
CO2 emissions	-0.17	0.06	0.04	0.07	0.17	-0.14	-0.02	0.18	-0.13	0.18	0.16	0.02	-0.19	0.00	0.19	-0.01	0.07	0.13	0.30**	1.00			
2 emissions	0.10	0.42**	-0.20	0.49**	0.23*	0.16	0.03	0.11	-0.10	0.17	0.33**	-0.17	-0.09	0.49**	-0.16	-0.41**	0.37**	0.43**	0.15	0.07	1.00		
CO2 emissions	-0.08	-0.13	-0.01	-0.12	-0.01	-0.09	0.19	0.09	-0.22	-0.09	0.15	-0.08	-0.13	-0.15	-0.06	0.21	-0.14	0.07	0.29*	0.11	-0.01	1.00	
missions	-0.10	0.29*	-0.09	0.25*	0.27*	-0.04	0.17	0.16	-0.29*	0.16	0.51**	-0.25*	-0.19	0.27*	-0.02	-0.12	0.19	0.47**	0.73**	0.78**	0.47**	0.25*	1.00

tion is significant at the 0.01 level (2-tailed).

on is significant at the 0.05 level (2-tailed).

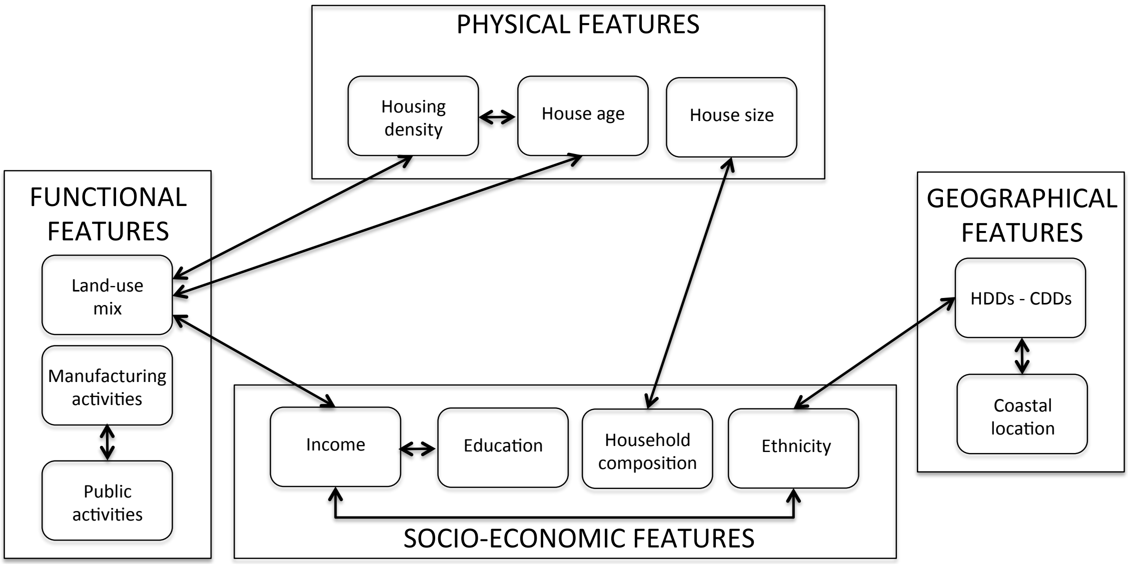
The correlation analysis allowed the assessment of two different types of relationships: (1) the relationships among the different physical, functional, geographical and socio-economic characteristics, which may indirectly affect energy consumption and CO_2 emissions on a city scale, and therefore represent a crucial result of this research; (2) the relationships between each pair of urban and energy variables, which provide interesting, but preliminary, findings that will be further explored through multiple regression analysis later herein.

Relationships among the urban variables

With respect to the first type of relationships, a number of significant associations ($r > 0.65$) emerge. Housing density is highly positively correlated with the concentration of jobs per square kilometer ($r = 0.98$) and is negatively correlated with the percentage of buildings built after 1980 ($r = -0.66$); subsequently the concentration of jobs per square kilometer is also negatively correlated with the percentage of buildings built after 1980 ($r = -0.66$). In other words, densely built cities have also higher land use mix, as well as a higher concentration of older buildings. Another significant association is that between house size (i.e. average floor area per capita) and household composition (i.e. percentage of children), which are negatively correlated ($r = -0.63$), meaning that a greater concentration of children corresponds to a lower average floor area per person. Moreover, income is positively correlated with the share of graduates ($r = 0.73$) and foreigners ($r = 0.69$), as well as with the number of jobs per square kilometer ($r = 0.67$): richer cities have higher levels of education, higher concentration of foreign residents and higher concentration of jobs. With respect to the geographical characteristics or urban areas, not surprisingly, degree-days are negatively correlated to coastal location ($r = -0.69$), i.e. colder cities tend to be inland cities and vice versa, and, less predictably, are positively correlated with the percentage of foreign residents ($r = 0.69$), i.e. warmer cities attract less foreigners than colder ones. The last strong correlation is that between manufacturing cities and those with a high concentration of public activities ($r = -0.79$), thus meaning that cities are likely to be either specialized in manufacturing or in public activities, not both. Figure 39 summarizes these relationships.

A number of similarities and differences emerge when these results are compared with those found in the scientific literature (Figure 3). More specifically, given the equivalence of housing and population density (see paragraph 3.3.1.1), three types of differences in results can be identified: (1) some relationships found in the literature have not been found here; (2) some relationships found here have not been found in the literature; (3) some relationships found in the literature have been also found here, but were not strong enough to be considered significant ($r < 0.65$).

Figure 41 Strongest correlations (r > 0.65) amongst the four groups of urban features



The first group includes the relationship between population density and heating degree days (Ewing & Rong, 2008) that does not come out as a significant result here ($r = 0.09$), and that between population density and house size (Ewing & Rong, 2008; Lee & Lee, 2014) that also turns out to be not significant here ($r = 0.00$). Even opposite are the results on the relationship between density and income: while Brownstone and Golob (2008) find a negative association for a sample of California households, the correlation found for the seventy-three Italian cities is positive ($r = 0.59$). This difference does not have to surprise if we consider the substantial differences of the two contexts in both socio-economic development and historical background.

The second group includes all the relationships that involve the coastal location and the concentration of manufacturing and touristic activities, as well as the relationships between density and house age, house age and land-use mix, land-use mix and income and income and ethnicity. These relationships cannot be found in the literature because they have not been investigated yet.

Finally, the third group includes three relationships found in the scientific literature as well as here, but that have not been included in Figure 39 because the Pearson's correlation coefficients are lower than 0.65, i.e. the positive relationship between income and house size ($r = 0.52$) and between density and ethnicity ($r = 0.42$), and the negative association between density and car ownership ($r = -0.41$).

Only two similarities can be found comparing the results presented above and those found in the literature and described in paragraph 2.4.1; namely, the relationship between density and land-use mix (Chen et al., 2008) and the relationship between house size and household composition (Ewing & Rong, 2008).

The few similarities and many differences in the results substantiate the argument that researches on the relationship between cities and energy consumption should pay more attention on these aspects rather than mainly focus on the direct relationships between urban features and energy use. As previously described in paragraph 2.4, indeed, the relationship between cities and energy consumption is complex and multidimensional because of the complexity and multidimensionality of cities, therefore in order to better understand this complexity is necessary to consider both the direct relationships between the different urban features and energy consumption and the relationships among the urban features, which may indirectly but significantly affect energy consumption and CO_2 emissions.

Relationships between urban and energy variables

With respect to the second type of relationships, the matrix of correlation coefficient shows few moderate associations. In particular, differently from the relationships among the urban variables, the Pearson's coefficients are never greater than 0.65. Nevertheless, the most interesting associations are those between the degree-days and both residential ($r = 0.61$) and total CO_2 emissions ($r = 0.51$), and that between income and tertiary CO_2 emissions ($r = 0.49$): not surprisingly, an increase in degree-days (i.e. colder cities) corresponds to an increase in residential/total CO_2 emissions (Kennedy et al. 2009), and an increase in income corresponds to an increase in tertiary CO_2 emissions. The scatterplots in Figure 40, 41 and 42 display these associations, pair by pair.

When comparing these results with those found in the scientific literature, three other relationships should be considered, two of which are still under debate: (1) the relationship between density and transport energy consumption/CO_2 emissions; (2) the relationship between density and residential energy consumption/CO_2 emissions; (3) the relationship between income and energy consumption/CO_2 emissions. Both relationships involving density still lack of consensus. In line with Lee and Lee (2014) and Makido et al. (2012) I found a negative association between density and residential CO_2 emissions (Figure 43) but this relationship cannot be considered significant ($r = -0.06$; $R^2 = 0.03$). Similarly, I found a negative association between density and transport CO_2 emissions (Figure 44) as found by Baur et al. (2014), Kennedy et al. (2009), Makido et al. (2012) and Newman and Kenworthy (1989) but, again, the magnitude of this association is very limited (r = -0.17; R^2 = 0.03). Slightly stronger, instead, is the positive relationship between income and total CO_2 emissions (r = 0.27; R^2 =0.07) shown in Figure 45, which supports what found by Baur et al. (2014) and Creutzig et al. (2015), among others.

Figure 42 Scatterplot of residential CO_2 emissions per capita versus degree-days

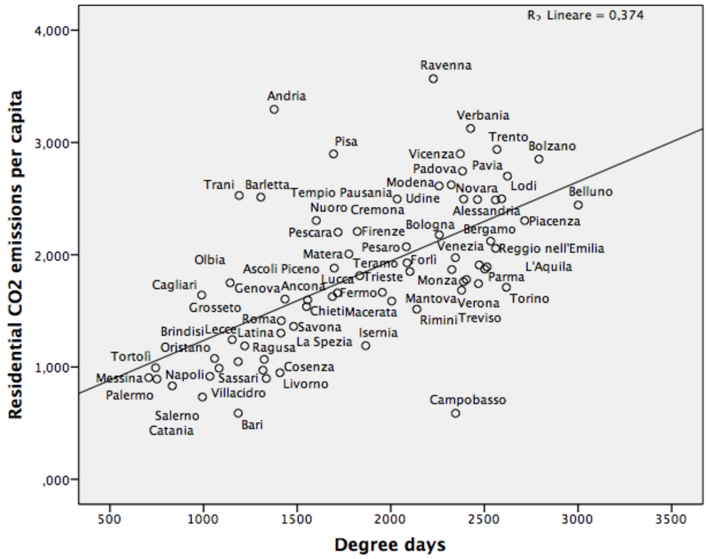


Figure 43 Scatterplot of total CO_2 emissions per capita versus degree-days

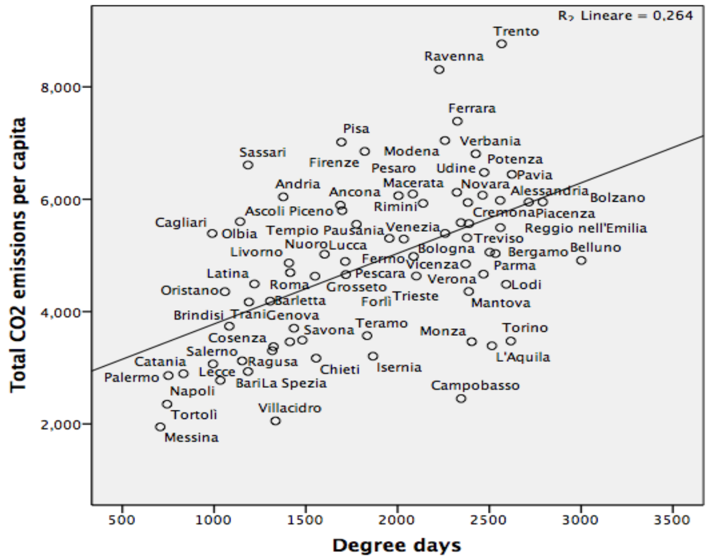


Figure 44 Scatterplot of tertiary CO_2 emissions per capita versus income

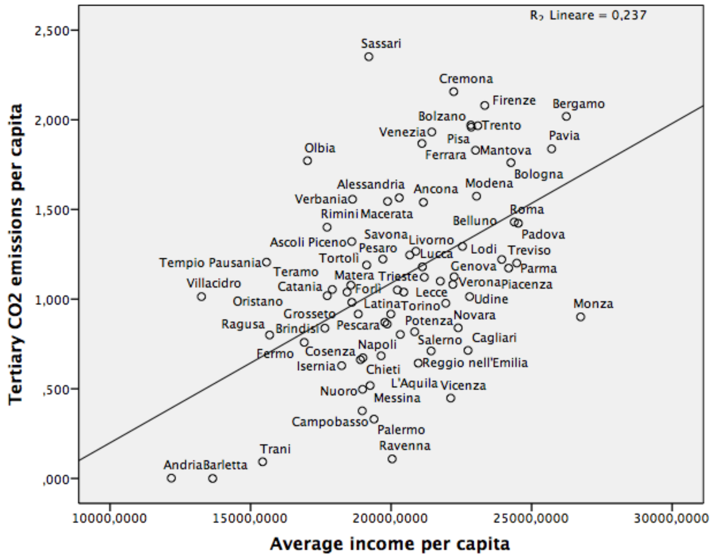


Figure 45 Scatterplot of residential CO_2 emissions per capita versus housing density

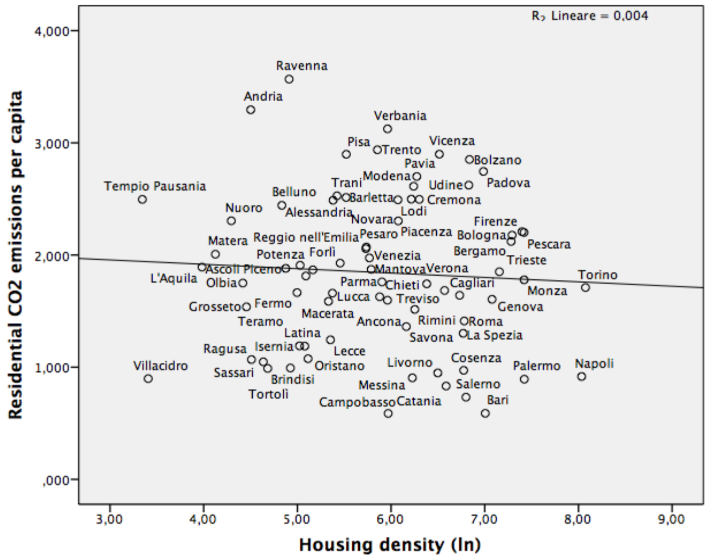


Figure 46 Scatterplot of transport CO_2 emissions per capita versus housing density

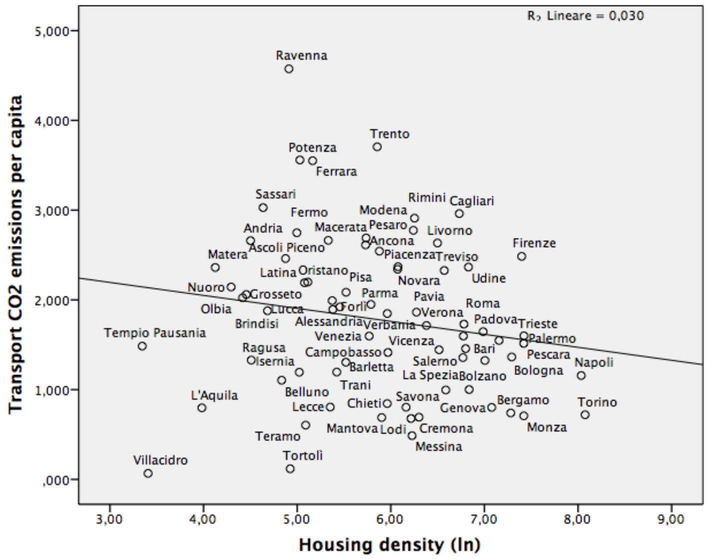
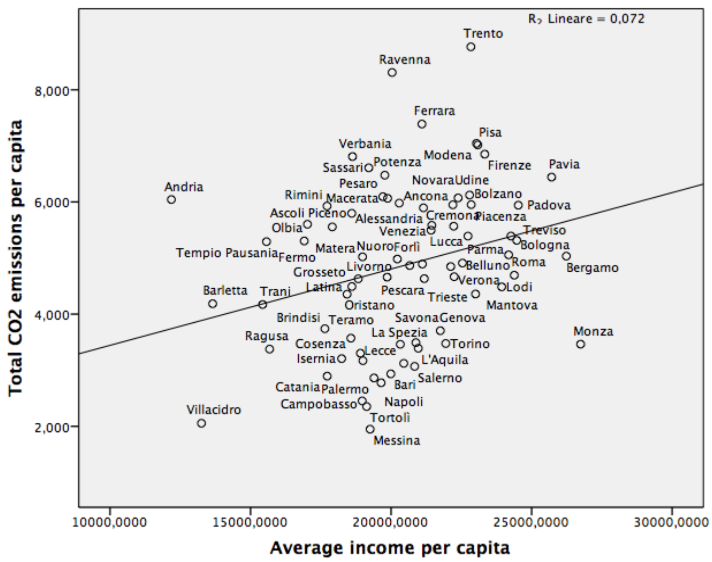


Figure 47 Scatterplot of total CO_2 emissions per capita versus income



4.4 Regression analysis

Three regression models (OLS) were estimated in order to measure the direct relationships between urban and energy features. In these three models, the dependent variables are three of the five categories of CO_2 emissions – residential, transport and total – and the eleven independent variables are housing density, house material, green areas, concentration of manufacturing activities, concentration of commercial activities, concentration of touristic activities, degree-days, topography, income, car ownership and household composition.

The selection of the eleven independent variables was based on the results of the correlation analysis: variables with a Pearson's correlation coefficient greater than 0.65 were discarded (i.e. average floor area per capita; percentage of buildings built after 1980; jobs per square kilometer; concentration of public activities; coastal location; graduates per 1.000 inhabitants; percentage of foreign residents).

Both dependent and independent variables included in the three multiple regression analyses are in natural log form[7], with the exclusion of the dummy variable "*Topography*". Therefore, the regression coefficients can be interpreted *"as the average percentage change in the dependent variable corresponding to a percentage change in the independent variable"* (Creutzig et al. 2015). The predictors initially included in each of the three regression models were then reduced by applying the backward stepwise method[8].

The results for the regression model with residential CO_2 emissions per capita as dependent variable (Table 7) show that emissions from buildings change considerably with changes in climate conditions: every 1% increase in degree-days corresponds to a statistically significant 0.75% increase of residential emissions. This result is in line with previous findings (Creutzig et al., 2015; Ewing & Rong, 2008). Furthermore, the model also shows that only one determinant explains two-fifth of the variance in residential emissions per capita. The other independent variables, including those describing the physical subsystem of a city, are not statistically significant, and thus removed from the final model (backward elimination procedure).

---

[7] See paragraph 3.4.3.2 for more details on the loglinear model.

[8] "*Backward elimination of variables chooses the subset models by starting with the full model and then eliminating at each step the one variable whose deletion will cause the residual sum of squares to increase the least. This will be the variable in the current subset model that has the smallest partial sum of squares*" (Rawlings et al. 2001).

The results for the regression model with transport CO_2 emissions per capita as dependent variable (Table 8) show that emissions from transportation moderately depend on the geographical and functional characteristics of urban settlements. Emissions are higher in colder cities with a concentration of touristic activities. Moreover, valley cities (topography = 0) emit 41% more for transportation than mountain cities (topography = 1).

Table 7 OLS results for residential CO_2 emissions

	Residential CO_2 emissions per capita		
	Coefficient	p-Value	VIF
Degree-days	0.753	0.000	1.000
Adjusted R^2		0.420	

Table 8 OLS results for transport CO_2 emissions

	Transport CO_2 emissions per capita		
	Coefficient	p-Value	VIF
Degree days	0.391	0.066	1.012
Concentration of touristic activities	0.228	0.015	1.030
Topography	-0.410	0.027	1.019
Adjusted R^2		0.140	

When residential and transportation emissions are considered together with emissions from municipal and tertiary buildings, results become even more interesting. Table 9 shows that total CO_2 emissions per capita decrease with increasing housing density, increase with increasing house age, degree days, the concentration of green areas and that of commercial activities, and that mountain cities emit less than valley ones.

The geographical features are the most important factors: every 1% increase in degree days corresponds to a statistically significant 0.39% increase of total emissions; valley cities emit 0.13% more CO_2 emissions per capita than mountain ones.

With respect to the physical variables, every 1% increase in housing density corresponds to a statistically significant 0.13% decrease of total emissions. This result substantiates the argument that compact cities consume less energy and emit less CO_2 than sprawl ones (Bereitschaft and Debbage, 2013; Clark, 2013; Creutzig et al., 2015; Ewing and Rong, 2008; Makido et al., 2012). Similarly, every 1% increase in the percentage of masonry buildings corresponds to a 0.23% increase of total emissions, thus meaning that house material significantly affect the energy performance of buildings when residential, tertiary and municipal buildings are considered together. If tertiary and municipal buildings are excluded, indeed, this correlation is not found significant (Table 7). Furthermore, if we look at the matrix of correlation coefficient (Table 6), tertiary CO_2 emissions and the percentage of masonry buildings have a correlation coefficient of 0.49, which suggests that the different construction materials have a significant influence on the energy use of tertiary buildings.

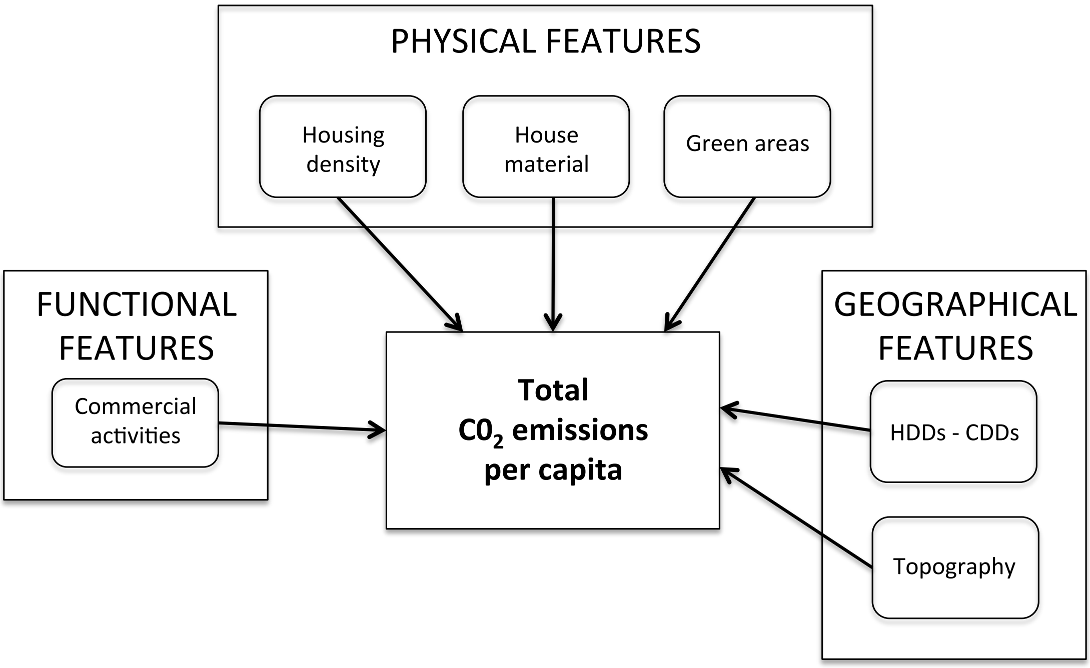
Another interesting finding is that every 1% increase in the percentage of green areas corresponds to a 0.09% increase of total CO_2 emissions. This result is particularly interesting because, up to now, green areas have always been considered for their microclimatic benefits. Vegetation, indeed, can effectively contribute to mitigate the urban heat island phenomenon (Dimoudi & Nikolopoulou, 2003; Gargiulo et al., 2016; Oliviera et al., 2011; Zoulia et al., 2009), i.e. the increase in urban temperatures compared to the surrounding rural areas. However, the studies on the positive effects of green spaces are limited to the cooling effect during summer time. Therefore, they did not consider the potential negative impacts of vegetation during winter time, which may increase energy use for space heating by increasing winter air temperatures. This negative effect can explain the positive correlation between the density of green areas and total CO_2 emissions that is shown in Table 9.

Table 9 OLS results for total CO_2 emissions

	Total CO_2 emissions per capita		
	Coefficient	p-Value	VIF
Housing density	-0.128	0.002	2.265
% of masonry buildings	0.234	0.037	1.061
% of green areas	0.092	0.006	2.493
Concentration of commercial activities	0.243	0.079	1.020
Degree-days	0.391	0.000	1.236
Topography	-0.132	0.046	1.151
Adjusted R^2		0.458	

These results are summarized in Figure 47, which visually outlines the direct relationships between the different urban features and CO_2 emissions. This schema shows that two categories of urban features mostly affect CO_2 emissions on a city scale: namely, physical and geographical features. Although the geographical characteristics of urban areas cannot be modified by human actions, the physical attributes can be transformed through land-use planning decisions, and thus are of main interest for planners and policy makers.

Figure 48 Significant relationships between the urban features and total CO_2 emissions



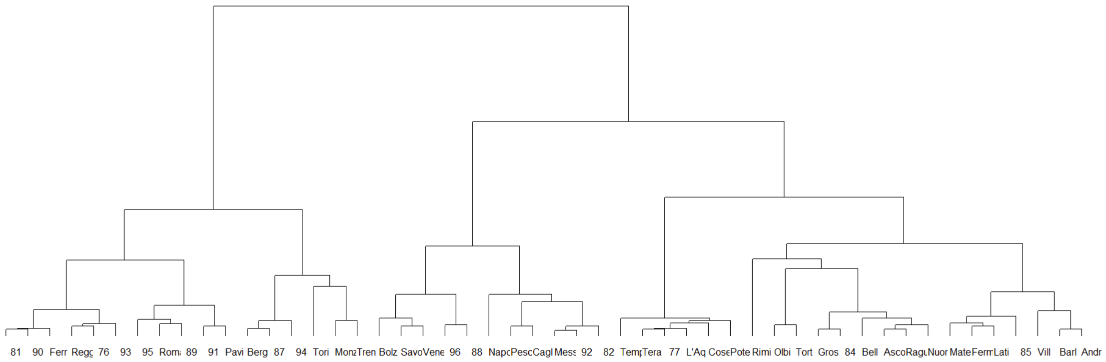
4.5 Cluster analysis

The last step of this research was to perform a cluster analysis and produce a dendrogram grouping the 73 Italian capital cities based on their urban characteristics. The energy variables were then projected onto these clusters in order to verify the extent to which groups of cities with similar physical, functional, geographical and socio-economic features produce similar CO_2 emissions.

The multivariate statistical analysis was conducted using the SPAD 55 software, which uses the Ward algorithm when applying the hierarchical classification. For this type of statistical analysis, only continuous variables can be used, therefore the two binary indicators included in the data set were previously replaced: the percentage of valley territory replaced the variable *topography* ($r = -0.71$), while the length of coastline divided by total area replaced the variable *coastal location* ($r = 0.69$).

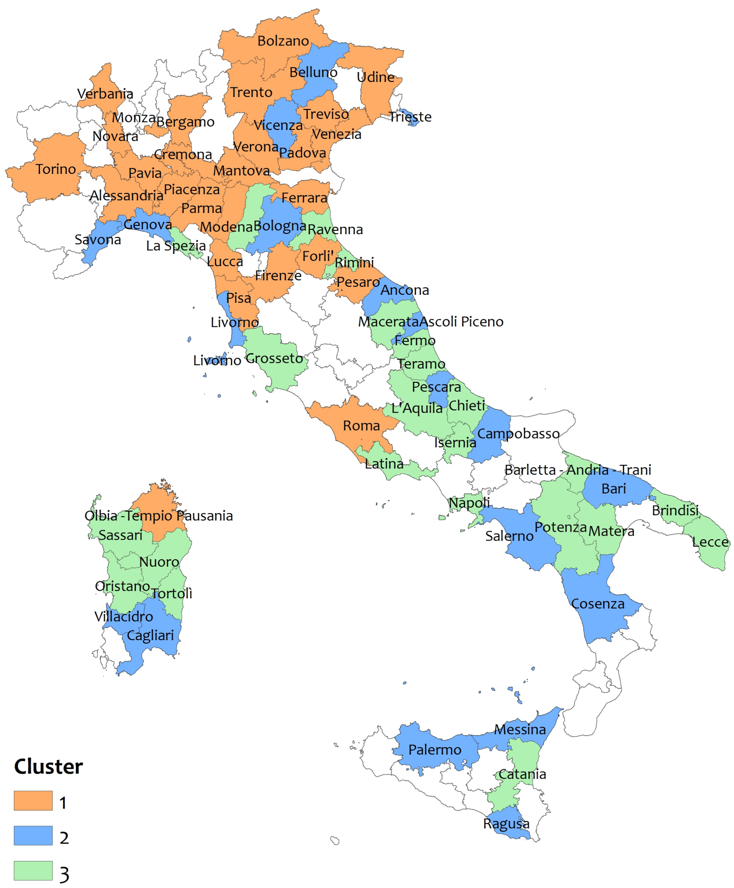
Based on this new data set, the analysis produced a hierarchical clustering dendrogram of the cities included in the research sample (Fig. 49). Given that there is no one objectively correct method to determine the best cutting point of a dendrogram, I cut the clustering tree where there is a large jump between two merged clusters, so obtaining three independent clusters.

Figure 49 Hierarchical clustering dendrogram of the cities dataset (using SPAD)



The first cluster includes 33 cities (45% of the sample), the second cluster includes 14 cities (19%) and the third cluster comprises 26 cities (36%). Figure 50 shows the geographical distribution of the three clusters. In particular, the elements of cluster 1 appear to be concentrated in the northern part of the country, differently from cities included in cluster 2 and 3, which are more uniformly distributed across Italy.

Figure 50 Map of the 73 Italian provincial capitals grouped into three clusters



4.5.1 Cluster 1

From a geographical point of view, cluster 1 includes mainly inland cities with high degree days and mostly located at the sea-level. The socio-economic characteristics of this cluster are: high income level, high concentration of foreign residents and graduates. Furthermore, these urban areas are mostly specialized in manufacturing, with a low concentration of jobs in public activities. Finally, from a physical perspective, these cities have a low concentration of "new" buildings and their average floor area per inhabitants is higher compared to the entire sample. To summarized, these are the wealthier cities of the sample, which not surprisingly are concentrated in the North of the country.

When projecting the energy variables on this cluster, the results show that these urban areas are characterized by residential, tertiary and total CO_2 emissions per capita higher than that of the full sample (Table 10). This result substantiates some of the results of the regression analysis, described in detail in the previous paragraph; in particular, it confirms those findings about the direct and significant effect of geographical features on CO_2 emissions at city scale: valley cities with higher degree days produce higher CO_2 emissions.

Table 10 Cluster 1 – characteristic variables

Characteristic variables	**Mean of the class**	**Mean of the sample**	**Test Values**	**Probab.**
Percentage of foreign residents	10.92	7.228	6.85	0.000***
Degree days	2342	1860	6.23	0.000***
Average income per capita	22573	20411	5.80	0.000***
Residential CO_2 emissions per capita	2.28	1.85	4.89	0.000***
Average floor area per capita	44.47	42.05	4.68	0.000***
Total CO_2 emissions per capita	5.73	4.86	4.59	0.000***
Tertiary CO_2 emissions per capita	*1.40*	*1.13*	*3.94*	*0.000****
Concentration of manufacturing (LQ)	1.15	0.96	3.33	0.000***
Graduates per 1.000 inhabitants	164.94	150.51	2.90	0.002**
Percentage of valley territory	90.77	79.42	2.80	0.003**
Percentage of buildings built after 1980	18.51	22.64	-3.06	0.001**
Concentration of public activities (LQ)	0.87	1.03	-3.17	0.001**
Coastline / total area	14.45	120.89	-4.14	0.000***

,,*** indicates the level of significance at the 10, 5, 1 percent level respectively*
Energy variables are in italics because they have been projected on the cluster

4.5.2 Cluster 2

Cluster 2 comprises mainly coastal cities with warmer climate (i.e. lower degree days). These urban areas have very high levels of housing density and a low concentration of buildings built after 1980. Moreover, they are characterized by a number of cars per capita lower than that of the full sample.

Similarly to cluster 1, the projection of the energy variables on this cluster corroborates parts of the results of the regression analysis. In particular, this cluster is characterized by lower residential and total CO_2 emissions per capita (Table 11). Therefore, higher housing density and lower degree days correspond to lower CO_2 emissions, as previously estimated using the regression models.

Table 11 Cluster 2 – characteristic variables

Characteristic variables	**Mean of the class**	**Mean of the sample**	**Test Values**	**Probab.**
Coastline / total area	437.69	120.89	6.61	0.000***
Housing density	1105.76	579.96	3.51	0.000***
Concentration of public activities (LQ)	1.32	1.03	3.15	0.001**
Jobs per kmq	698.54	422.98	2.45	0.007*
Concentration of manufacturing (LQ)	0.700	0.96	-2.40	0.008*
Percentage of buildings built after 1980	16.03	22.64	-2.62	0.004**
Cars per 1.000 inhabitants	570.72	614.22	-2.92	0.002**
Total CO_2 emissions per capita	3.79	4.86	-3.05	0.001**
Residential CO_2 emissions per capita	*1.25*	*1.85*	-3.55	*0.000****
Degree days	1302	1860	-3.88	0.000***

,,*** indicates the level of significance at the 10, 5, 1 percent level respectively*
Energy variables are in italics because they have been projected on the cluster

4.5.3 Cluster 3

The 26 urban areas included in cluster 3, in opposition to those included in cluster 1, are poorer cities, both in economic and social terms. Indeed, they are characterized by low income per capita, low number of foreigners and graduates and low job density. At the same time, these cities have a high car ownership level and a high percentage of children. From a physical point of view, cluster 3 corresponds to "newer" and sprawl urban settlements, with a low concentration of green areas. Furthermore, once more in contrast with cluster 1, they have warmer climate and are mostly located at higher elevation.

Differently from both cluster 1 and 2, the projection of the energy variables on cluster 3 does not provide substantial results. Only one category of CO_2 emissions (i.e. tertiary), indeed, significantly characterize this last cluster. However, tertiary emissions only account for about 23% of total CO_2 emissions per capita.

Table 12 Cluster 3 – characteristic variables

Characteristic variables	**Mean of the class**	**Mean of the sample**	**Test Values**	**Probab.**
Percentage of buildings built after 1980	31.43	22.64	5.34	0.000***
Cars per 1.000 inhabitants	650.78	614.22	3.74	0.000***
Percentage of children	13.45	12.87	2.66	0.004**
Percentage of green areas	1.75	3.97	-2.55	0.005**
Tertiary CO_2 emissions per capita	*0.91*	*1.13*	*-2.63*	*0.004***
Percentage of valley territory	65.53	79.42	-2.80	0.003**
Average floor area per capita	40.17	42.05	-2.98	0.001**
Graduates per 1.000 inhabitants	130.67	150.51	-3.26	0.001**
Degree days	1550	1860	-3.29	0.000***
Housing density	182.00	579.96	-4.05	0.000***
Jobs per kmq	114.51	422.98	-4.19	0.000***
Percentage of foreign residents	3.51	7.23	-5.65	0.000***
Average income per capita	17634	20411	-6.10	0.000***

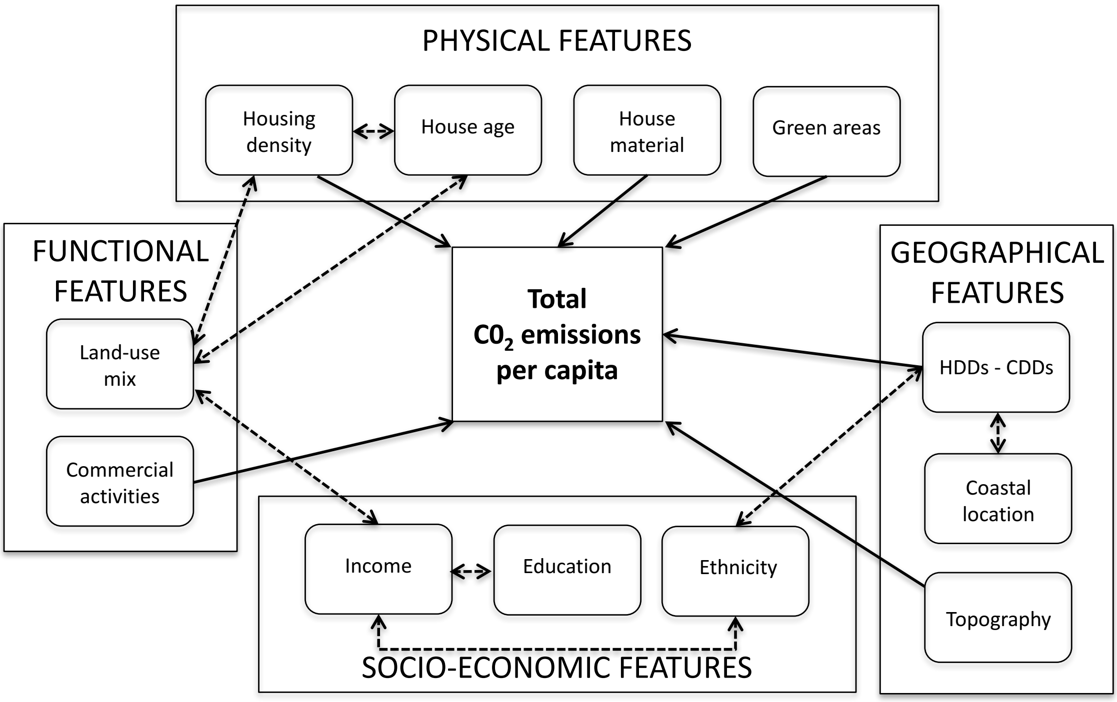
,,*** indicates the level of significance at the 10, 5, 1 percent level respectively*
Energy variables are in italics because they have been projected on the cluster

4.6 Conclusions

As demonstrated in Section 2, the complex relationship between cities and energy consumption/CO_2 emissions involves two main types of interactions: (1) the direct relationships between urban and energy variables; and (2) the relationships among the different urban features, which indirectly but significantly affect the carbon and energy footprint of cities. Based on this conceptual framework, the main results of this research are summarized in Figure 48, which includes both types of interactions: solid arrows represent the relationships between the urban features and CO_2 emissions, as in Figure 47; dashed arrows represent the relationships amongst the four groups of urban features, as in Figure 39.

When considering both direct and indirect effects, all the four groups of urban variables affect total CO_2 emissions per capita. More specifically, three physical variables (i.e. housing density, house material and green areas), one functional variable (i.e. the concentration of commercial activities), and two geographical variables (i.e. degree-days and topography) have a direct effect on CO_2 emissions: lower density of dwelling units and lower air temperatures, as well as valley topography and higher concentrations of masonry buildings, green areas and commercial activities increase CO_2 emissions per capita.

Figure 51 Direct and indirect relationships between the urban features and total CO_2 emissions



On the other hand, three socio economic variables (i.e. income, education and ethnicity) and one functional variable (i.e. land-use mix) indirectly affect CO_2 emissions through the mediators of other urban features. In particular, a higher level of education and a higher share of foreign residents are both associated with higher income that, in turn, is associated with higher land-use mix that corresponds to higher housing density, which reduces CO_2 emissions per capita. However, at the same time, a higher concentration of foreigners also corresponds to higher degree-days, which increases CO_2 emissions per capita. Figure 48, therefore, clearly highlights the complex trade-offs existing between the different urban factors and energy saving (Doherty et al., 2009; Lee & Lee, 2014; Papa et al., 2016), which were partially estimated in this research, but need further investigation to be fully understood and measured.

Furthermore, the results of the cluster analysis not only substantiate the findings of the regression analysis, but also provide additional insight about the sample of Italian cities considered in this research. The cluster analysis, indeed, shows that Italian colder-valley-inland-wealthier cities – such as Torino, Bolzano, Padova, Mantova, etc. – produce higher level of both residential and total CO_2 emissions, and are exclusively located in the northern part of the country (except for Roma and Tempio Pausania). On the contrary, Italian cities by the sea, with warmer climate and densely urbanized – such as Genova, Salerno, Bari, etc. – emit less CO_2 per capita, thus being more energy efficient than the others.

SECTION 5. CONCLUSIONS

5.1 Conclusions

As a factor of economic growth, technological progress and social change, energy can be considered the driving force behind urban development. Nevertheless, using and producing energy is responsible for about 65% of global greenhouse gas (GHG) emissions (Marrero, 2010), which, in turn, are responsible for climate change. The dramatic consequences of climate change – such as the warming of the atmosphere and the oceans, the intensification of extreme weather events, and the issue of food security, to name a few – are expected to increase in the coming years, especially if no action is taken. Therefore, in order to prevent climate change and preserve our planet for future generations, the European Commission (EC) has recently proposed new ambitious energy goals to cut its GHG emissions substantially and turn Europe into a highly energy efficient and low-carbon economy. In particular, key EU targets for 2030 require EU countries to increase energy efficiency by 27%, and cut GHG emissions by 40% compared with 1990. Thus, European countries have to consider energy inside the urban planning process (Papa & Fistola, 2016), in order to pursue together social, economic and environmental goals.

In this context, urban areas should be at the center of these sustainable policies, because "urban energy systems provide significant opportunities for increased efficiency in delivering transport and building services" (IEA, 2016). Cities, indeed, consume up to 75% of global energy and account for 78% of carbon emissions (CO_2) produced by human activities (Habitat-UN, 2011; IEA 2008). Thus, urban areas play a key role in addressing climate change.

A growing body of international researchers has been studying the complex and multidimensional relationship between cities and energy consumption so to support local policy makers' decisions and foster the transition towards a low-carbon future. If the interactions between urban factors and energy use are investigated and are found to be significant, indeed, urban planning policies can effectively improve energy saving in cities and reduce urban emissions. However, despite the great interest of the literature for this topic, a consistent number of interactions between urban features and energy use at urban scale still lacks consensus.

This research aimed to identify the urban factors that significantly affect a city's energy and carbon footprint, thus supporting policy-makers in the definition of effective strategies and policies that can be implemented at urban scale to reduce energy consumption and resulting CO_2 emissions.

Two main innovations were introduced in this research. The first innovation concerns the approach: I used a holistic approach rather than a sectorial one, thus considering at the same time a comprehensive set of urban factors – physical, functional, geographical, and socio-economic – describing the complexity and multidimensionality of cities for measuring their impacts on CO_2 emissions. Second, I didn't limit the analysis to the direct relationships between the urban features and energy consumption, but I also investigated the relationships among the different urban features, which indirectly but significantly affect energy consumption on an urban scale. This integrated approach allowed the identification of the existing trade-offs between different urban features and energy saving, providing a broader and more complete picture on such a complex topic.

The first step in this research was to review the scientific literature on the relationship between cities and energy consumption over the past twenty years. This review allowed the identification of the urban and energy variables to be included in the statistical models later developed to investigate this relationship. In particular, a set of eighteen urban variables and five energy variables was collected for a sample of seventy-three Italian capital cities, uniformly distributed across the country. The eighteen urban variables comprise five physical variables (i.e. housing density, house size, house age, house material, and green areas), five functional variables (i.e. jobs per square kilometer, concentration of manufacturing activities, concentration of commercial activities, concentration of public activities, and concentration of touristic activities), three geographical variables (i.e. degree-days, coastal location, and topography), and five socio-economic variables (i.e. income, car ownership, household composition, concentration of graduates, and concentration of foreign residents). The five energy variables include residential, transport, tertiary, municipal, and total CO_2 emissions per capita; total CO_2 emissions per capita was calculated as the sum of all other emissions.

After an intensive data collection procedure, the complete dataset was explored and analyzed using different statistical methods, each of which provided useful insights into the complex relationship between cities and their carbon footprint. Correlation analysis enabled the measurement of the association between the eighteen urban variables and showed the linear relationship between each individual urban variable and the five categories of CO_2 emissions. Then, multiple regression analysis (OLS) allowed the estimation of the relationships between CO_2 emissions per capita and a subset of the initial eighteen urban variables, which was identified by excluding those variable with a Pearson's correlation coefficient higher than 0.65. Lastly, cluster analysis was used to explore the urban characteristics of the capital cities included in the sample to verify whether homogeneous classes of urban areas have similar energy behaviours.

The results of the regressions analysis show that three physical features – housing density, house material, green areas – and two geographical features – degree-days and topography – significantly affect total CO_2 emissions per capita (R^2=0.46, p value < 0.05): with each 1% increase in housing density, total CO_2 emissions decrease by 0.13%; every 1% increase in the percentage of masonry building respect to concrete ones corresponds to a 0.23% increase in emissions; every 1% increase in the density of green spaces corresponds to a 0.09% increase in emissions; every 1% increase in degree-days corresponds to a 0.39% increase in emissions; and total CO_2 emissions decrease by 0.13% when passing from valley to mountain cities. The significant effects of housing density and degree-days on CO_2 emissions substantiate previous findings (Bereitschaft and Debbage, 2013; Clark, 2013; Creutzig et al., 2015; Ewing and Rong, 2008; Makido et al., 2012). On the other hand, results on the influence on CO_2 emissions of construction materials, green areas and topography have not been found in the literature so far, and therefore require further investigation for being validated. Green areas, in particular, have always been considered a positive element within the urban context because of their capability of reducing air temperature during summer time; the potential negative effects of green spaces on urban temperature during the winter have not been investigated so far.

The results of the correlation analysis show that more compact cities have a higher density of jobs that, in turn, corresponds to a higher average income, a higher number of graduates and a higher share of foreign residents. These findings partially contradict previous results found in the literature. More specifically, Brownstone and Golob (2008) as well as Ewing & Rong (2008) find that density and income are negatively correlated in the US: richer people are more likely to live in sprawl counties. Within the Italian context, in contrast, the association between density and income is positive: higher levels of income are concentrated in urban settlements with a higher share of jobs and dwelling units. This difference can be explained considering the dissimilarities between North American and Italian cities in terms of urban development due to the very different historical backgrounds and economic growth paths. These considerations highlight the importance of sample homogeneity for the investigation of the relationship between cities and energy consumption, because the physical, functional and socio-economic characteristics of urban areas may significantly differ among cities of different countries. Therefore, when the relationship between cities and energy consumption is investigated at global scale (i.e. considering a sample of cities of different countries), cities should be clustered in order to account for historical and socio-economic differences which might confound the final results.

Integrating both correlation and regression analysis results, this research shows that the two main categories of urban factors directly affecting CO_2 emissions per capita are the geographical and physical features, whereas the functional and socio-economic characteristics of urban areas have an indirect effect on CO_2 emissions. In other words, the climate condition of a city and its physical structure (both in terms of urban density and buildings characteristics) are in large part responsible for the use of energy and the resulting CO_2 emissions within the urban perimeter. Given that the geographical factors of cities cannot be changed by human intervention, the key role of urban policies and spatial planning in addressing energy and environmental issues becomes of strategical importance for addressing climate change.

In this regard, two main policy implications are drawn from the results of both correlation and regression analysis; one at the building scale and one at the urban scale. (a) At the building scale, interventions should focus on buildings materials, especially for reducing the energy use of masonry buildings. (2) At the urban scale, planning strategies should encourage compact developments in order to reduce energy consumption and total CO_2 emissions. Furthermore, besides the lower energy footprint of compact cities, in Italy higher densities of housing units correspond to higher densities of jobs, which in turn are characterized by higher incomes, and therefore strategies for promoting urban compactness can also have positive economic effects.

Beyond the aggregate results provided by both correlation and regression analysis, the cluster analysis gave significant information about the sample of 73 Italian capital cities considered in this research. In particular, it enabled the identification of three groups of cities with similar physical, functional, geographical and socio-economic characteristics, which have very different carbon footprints. Corroborating the findings of the analyses of correlation and regression outlined above, the results of the cluster analysis shows that northern Italian cities, mainly located at sea-level, with colder climate and a higher income, as well as a higher concentration of foreign residents and graduate students have a higher carbon footprint; on the other hand, warmer cities located on the coast and characterized by a compact urban development have a lower carbon footprint.

This work confirms the complexity and multidimensionality of the relationship between cities and energy consumption and the importance of both building and urban interventions to increase energy saving and decrease CO_2 emissions on a city scale (Zanon & Verones, 2013). Furthermore, the results of this research, which only partially support previous findings, suggest that important trade-offs exist between the different urban characteristics and the energy and carbon footprint of cities (Doherty et al., 2009; Lee & Lee, 2014; Papa et al., 2016). Measuring all of the trade-offs is a very challenging task, and this research proposed a first step in this direction.

5.2 Limitations and future developments

This research has several limitations. The first limitation is data availability. Because of data limitations, indeed, (1) data on urban areas and energy consumption/CO_2 emissions refer to two different time periods, and moreover, (2) some urban sectors that significantly affect energy consumption and resulting CO_2 emissions, such as industry, could not been considered. Furthermore, if more data were available, a more numerous sample of cities would have allowed the construction of different regression models for each group of cities obtained with the cluster analysis, thus providing more detailed information about the energy behavior of different typologies of urban areas.

The second limitation concerns the statistical methods used to estimate the relationship between cities and energy consumption: the correlation analysis and the multiple regression analysis do not allow the identification of a causal link between the variables considered. In other words, a strong correlation between two variables does not imply a direct link between these variables but it could be the results of an indirect interaction that involves other variables. Therefore, future research should focus on using different statistical model to study the complex relationship between cities and energy consumption, such as, for example, a multilevel structural equation model (SEM), which simultaneously tests multiple causal relations (Lee & Lee, 2014).

The third limitation concerns the research's sample: the results of the analyses previously described refer to a sample of Italian cities, therefore they may not apply to different geographical contexts. As previously highlighted, indeed, urban and energy features significantly differ from one country to another, and these differences may translate into different results. An interesting future development of this work would be to apply the same methodology to different countries and compare the results in order to identify similarities and differences and better support policy makers in the definition of effective urban strategies for reducing energy consumption and CO_2 emissions on a city scale.

REFERENCES

Alexander, J. W. (1954). The basic-nonbasic concept of urban economic functions. *Economic Geography*, 30(3), 246-261.

Anderson, W. P., Kanaroglou, P. S., & Miller, E. J. (1996). Urban form, energy and the environment: a review of issues, evidence and policy. *Urban studies*, 33(1), 7-35. DOI: 10.1080/00420989650012095

Andrews, R. B. (1953). Mechanics of the urban economic base: historical development of the base concept. *Land Economics*, 29(2), 161-167.

Balaras, C. A., Gaglia, A. G., Georgopoulou, E., Mirasgedis, S., Sarafidis, Y., & Lalas, D. P. (2007). European residential buildings and empirical assessment of the Hellenic building stock, energy consumption, emissions and potential energy savings. *Building and Environment*, 42(3), 1298-1314.

Banister, D., Watson, S., & Wood, C. (1997). Sustainable cities: transport, energy, and urban form. *Environment and Planning B: planning and design*, 24(1), 125-143.

Batty, M. (2008). Cities as complex systems: scaling, interactions, networks, dynamics and urban morphologies (*UCL Working paper 131*). Centre for Advanced Spatial Analysis, University College London.

Baur, A. H., Thess, M., Kleinschmit, B., & Creutzig, F. (2013). Urban climate change mitigation in Europe: looking at and beyond the role of population density. *Journal of Urban Planning and Development*, 140(1), 04013003.

Bereitschaft, B., & Debbage, K. (2013). Urban form, air pollution, and CO2 emissions in large US metropolitan areas. *The Professional Geographer*, 65(4), 612-635. DOI: 10.1080/00330124.2013.799991

Breheny, M. J. (2001). Densities and sustainable cities: the UK experience. In M. Eschenique, & A. Saint (Eds.), *Cities for the new millennium*. Spon Press.

Brownstone, D., & Golob, T. F. (2009). The impact of residential density on vehicle usage and energy consumption. *Journal of Urban Economics*, 65(1), 91-98.

Burton, E. (2000). The compact city: just or just compact? A preliminary analysis. *Urban studies*, 37(11), 1969-2006. DOI: 10.1080/00420980050162184

Camagni, R., Gibelli, M. C., & Rigamonti, P. (2002). Urban mobility and urban form: the social and environmental costs of different patterns of urban expansion. *Ecological economics*, 40(2), 199-216.

Chen, Y., Li, X., Zheng, Y., Guan, Y., & Liu, X. (2011). Estimating the relationship between urban forms and energy consumption: a case study in the Pearl River Delta, 2005–2008. *Landscape and urban planning*, 102(1), 33-42.

Chen, H., Jia, B., & Lau, S. S. Y. (2008). Sustainable urban form for Chinese compact cities: Challenges of a rapid urbanized economy. *Habitat international*, 32(1), 28-40.

Clark, T. A. (2013). Metropolitan density, energy efficiency and carbon emissions: Multi-attribute tradeoffs and their policy implications. *Energy Policy*, 53, 413-428.

Creutzig, F., Baiocchi, G., Bierkandt, R., Pichler, P. P., & Seto, K. C. (2015). Global typology of urban energy use and potentials for an urbanization mitigation wedge. *Proceedings of the National Academy of Sciences*, 112(20), 6283-6288.

Dimoudi, A., & Nikolopoulou, M. (2003). Vegetation in the urban environment: microclimatic analysis and benefits. *Energy and buildings*, 35(1), 69-76.

Doherty, M., Nakanishi, H., Bai, X., & Meyers, J. (2009). Relationships between form, morphology, density and energy in urban environments. *GEA Background Paper*.

Echenique, M. H., Hargreaves, A. J., Mitchell, G., & Namdeo, A. (2012). Growing cities sustainably: does urban form really matter?. *Journal of the American Planning Association*, 78(2), 121-137. DOI: 10.1080/01944363.2012.666731

Epstein, J. (2014). *Exploring data distribution*. College Station: Texas A&M University, Chapter 5.

European Commission (2010). *How to develop a sustainable energy action plan (SEAP) - Guidebook*. Publications Office of the European Union, Luxembourg.

Ewing, R., & Rong, F. (2008). The impact of urban form on US residential energy use. *Housing policy debate*, 19(1), 1-30.

FCCC – Framework Convention on Climate Change. (2015). *Adoption of the Paris agreement.*

Friedman, J., Hastie, T., & Tibshirani, R. (2001). *The elements of statistical learning (Vol. 1).* Springer, Berlin: Springer series in statistics.

Gargiulo, C., & Papa, R. (1993). Caos e caos: la città come fenomeno complesso. Per il *XXI Secolo: una enciclopedia e un progetto*, 297-306. ISBN 978-88-97110-15-6

Gargiulo, C., Tulisi, A., & Zucaro, F. (2016). Small green areas for energy saving: effects on different urban settlements. *Arquitectura, Ciudad y Entorno*, 11(32), 81-94.

Gerundo, R., Fasolino, I., & Grimaldi, M. (2016). ISUT Model. A Composite Index to Measure the Sustainability of the Urban Transformation. In *Smart Energy in the Smart City*, pp. 117-130. Springer International Publishing.

Habitat, U. N. (2011). *Cities and climate change: Global report on human settlements 2011*. London: Earthscan.

Haig M. Robert, *Major Economic Factors in Metropolitan Growth and Arrangement*; Vol. I of Regional Survey of New York and Environs (Extract of a letter from Mr. Frederick Law Olmsted to Mr. John M. Glenn dated February 21, 1921) (New York: 1928), p. 43.

Holden, E., & Norland, I. T. (2005). Three challenges for the compact city as a sustainable urban form: household consumption of energy and transport in eight residential areas in the greater Oslo region. *Urban studies*, 42(12), 2145-2166.

IEA – International Energy Agency. (2016). *Energy Technology Perspectives 2016. Towards Sustainable Urban Energy Systems*.

IEA – International Energy Agency. (2008). *World Energy Outlook 2008*, International Energy Agency, OECD, Paris (2008).

Jabareen, Y. R. (2006). Sustainable urban forms their typologies, models, and concepts. *Journal of planning education and research*, 26(1), 38-52. DOI: 10.1177/0739456X05285119

Kennedy, C., Steinberger, J., Gasson, B., Hansen, Y., Hillman, T., Havranek, M., ... & Mendez, G. V. (2009). Greenhouse gas emissions from global cities. *Environmental science & technology*, 43(19), 7297-7302.

Lee, S., & Lee, B. (2014). The influence of urban form on GHG emissions in the US household sector. *Energy Policy*, 68, 534-549.

Levy, A. (1999). Urban morphology and the problem of the modern urban fabric: some questions for research. *Urban Morphology*, 3(2), 79-85.

Lynch, K. (1981). *Good city form*. Cambridge: MIT Press.

Makido, Y., Dhakal, S., & Yamagata, Y. (2012). Relationship between urban form and CO 2 emissions: evidence from fifty Japanese cities. *Urban Climate*, 2, 55-67.

Marrero, G. A. (2010). Greenhouse gases emissions, growth and the energy mix in Europe. *Energy Economics*, 32(6), 1356-1363.

Marshall, J. D. (2008). Energy-efficient urban form. *Environmental science & technology*, 42(9), 3133-3137.

Mindali, O., Raveh, A., & Salomon, I. (2004). Urban density and energy consumption: a new look at old statistics. *Transportation Research Part A: Policy and Practice*, 38(2), 143-162.

Mukaka, M. M. (2012). A guide to appropriate use of correlation coefficient in medical research. *Malawi Medical Journal*, 24(3), 69-71.

Newman, P. W., & Kenworthy, J. R. (1989). Gasoline consumption and cities. *Journal of the American planning association*, 55(1), 24-37. DOI: 10.1080/01944368908975398

Newton, P., Tucker, S., Ambrose, M. (2000). Housing form, energy use and greenhouse gas emissions, in: Williams, K., E. Burton and M. Jenks. (2000). *Achieving Sustainable Urban Form*. Spon Press, Taylor and Francis Group, London, pp. 74 – 84.

Nuzzolo, A., Coppola, P., & Papa, E. (2014). Urban from and sustainability: modelling evidences from the empirical case study of RomeE. In *ETC European Transport Conference 2014* (pp. 1-20). AET 2014 and contributors.

OECD – Organization for Economic Co-operation and Development. (2014). *Cities and climate change*. National governments enabling local action.

Oliveira, S., Andrade, H., & Vaz, T. (2011). The cooling effect of green spaces as a contribution to the mitigation of urban heat: A case study in Lisbon. *Building and Environment*, 46(11), 2186-2194.

Papa, R. (2009). *Il governo delle trasformazioni urbane e territoriali. Metodi, tecniche e strumenti*. Franco Angeli, Milan, Italy.

Papa, R., Battarra, R., Fistola, R. & Gargiulo, C. (1995). La città come sistema complesso in crisi strutturale, in: Bertuglia, C.S., Fuccella, R., Sartorio, G.L., *La città come sistema complesso in crisi strutturale - strumenti e tecniche per il governo metropolitano*, Giuffré, Roma, 1995. ISBN 88-14-05263-8

Papa, R., & Fistola, R. (2016). Preface. In *Smart Energy in the Smart City*, 151-175. Springer International Publishing.

Papa, R., Gargiulo, C., & Zucaro, F. (2016). Towards the Definition of the Urban Saving Energy Model (UrbanSEM). In *Smart Energy in the Smart City*, v-vi. Springer International Publishing.

Papa, R., Gargiulo, C., & Zucaro, F. (2014). Climate Change and Energy Sustainability. Which Innovations in European Strategies and Plans. *Tema. Journal of Land Use, Mobility and Environment*, 0. DOI: http://dx.doi.org/10.6092/1970-9870/2554

Rawlings, J. O., Pantula, S. G., & Dickey, D. A. (2001). *Applied regression analysis: a research tool.* Springer Science & Business Media.

Reddy, B. V., & Jagadish, K. S. (2003). Embodied energy of common and alternative building materials and technologies. *Energy and buildings*, 35(2), 129-137.

Rickwood, P., Glazebrook, G., & Searle, G. (2008). Urban structure and energy—a review. *Urban policy and research*, 26(1), 57-81.

Seltman, H. J. (2012). *Experimental design and analysis*. Pittsburgh: Carnegie Mellon University, 428.

Stead, D., & Marshall, S. (2001). The relationships between urban form and travel patterns. An international review and evaluation. *European Journal of Transport and Infrastructure Research*, 1(2), 113-141.

Thinh, N. X., Arlt, G., Heber, B., Hennersdorf, J., & Lehmann, I. (2002). Evaluation of urban land-use structures with a view to sustainable development. *Environmental Impact Assessment Review*, 22(5), 475-492.

Verbeek, M. (2008). *A guide to modern econometrics*. John Wiley & Sons.

Von Bertalanffy, L. (1969). *General system theory: foundations, development, applications* (Revised Edition).

Williams, K., Jenks, M., & Burton, E. (2000). *Achieving sustainable urban form*. Taylor & Francis.

Wilson, J. H., Keating, B. P., & Beal, M. (2017). *Regression Analysis: Understanding and Building Business and Economic Models Using Excel.* Second Edition. Business Expert Press.

Worldwide, M. (2008). *Worldwide centers of commerce index*. Retrieved April, 5, 2013.

Ye, H., He, X., Song, Y., Li, X., Zhang, G., Lin, T., & Xiao, L. (2015). A sustainable urban form: The challenges of compactness from the viewpoint of energy consumption and carbon emission. *Energy and Buildings*, 93, 90-98.

Zanon, B., & Verones, S. (2013). Climate change, urban energy and planning practices: Italian experiences of innovation in land management tools. *Land Use Policy*, 32, 343-355.

Zoulia, I., Santamouris, M., & Dimoudi, A. (2009). Monitoring the effect of urban green areas on the heat island in Athens. *Environmental monitoring and assessment*, 156(1), 275-292.