https://doi.org/10.5194/egusphere-2025-908
Preprint. Discussion started: 28 February 2025

Identifying urban settlement archetypes: clustering for enhanced multi-risk exposure and vulnerability analysis

Gabriella Tocchi[a], Massimiliano Pittore[b], Maria Polese[a]

[a] *University of Naples Federico II, Department of Structures for Engineering and Architecture, Naples, Italy*
[b]*EURAC Research, Center for Climate Change and Transformation, Bolzano, Italy*

Correspondence to: Gabriella Tocchi (gabriella.tocchi@unina.it)

Abstract. Identification of risks and vulnerabilities in urban areas is crucial for supporting city authorities in disaster risk reduction and climate change adaptation. Moreover, comparison of risk assessments across different cities may help effective allocation of adaptation funding towards more resilient and sustainable cities. The distinct physical, social, economic, and environmental characteristics of a city, along with the relevance of impending hazards, determine the level of risk and vulnerability faced by its residents. While the results of urban risk assessments will vary from one city to another, using general urban typologies (e.g. coastal cities, dryland cities, and inland or high-altitude cities) can effectively support in the understanding of risk in relation to its key drivers, helping to segmentate the complexity in otherwise too broad problem (Dickson et al., 2012).

This study aims to reduce complexity in urban risk assessment at regional and national scale, ensure a baseline for comparison and identify potential hotspots in multi-hazard and multi-risk assessment frameworks. We propose a clustering methodology that groups urban settlements based on open-source data, used as proxies of urban vulnerability and exposure. Applying two widely used clustering techniques, we define 18 urban archetypes for the Italian territory, incorporating geographic, demographic, and socio-economic characteristics. These archetypes satisfy multiple validity dimensions of archetype analysis (Piemontese et al., 2022) and can serve as a valuable tool for policymakers. By providing a structured understanding of urban vulnerability profiles, they support the design of targeted interventions and urban resilience strategies tailored to specific risk conditions.

1 Introduction

Over last few decades, natural disasters have caused devastation to many communities throughout the world, killing about 1.5 million of people and incurring losses exceeding 4.5 billion USD (Centre for Research on the Epidemiology of Disasters - CRED, 2024). Such disasters are the results of the interaction of hazards (natural or man-made) with vulnerable socio-ecological and socio-economical systems. Evidence shows that the level of disaster proneness of communities may vary greatly with their physical, demographic, socioeconomic and institutional characteristics (Cutter et al., 2003; Wang et al., 2022). For example, low-income and minority communities in New Orleans were disproportionately affected during Hurricane Katrina due to residing in flood-prone, lower-lying areas, and lacking personal transportation, which hindered evacuation (Flanagan et al., 2011). Similarly, aging communities with limited mobility face challenges in evacuating quickly during hazardous events, leading to higher mortality rates, as seen during the 2011 Tohoku Tsunami, Hurricane Katrina, and the 2017 and 2018 California wildfires (Brunkard et al., 2008; Hamideh et al., 2022; Miyazaki, 2022).

Climate change brings additional challenges to urban management and decision making for city governments and is associated with a growing variety of impacts on cities, the surrounding ecosystems, and livelihood of resident and temporary population (e.g., Dickson et al., 2012). As highlighted in the IPCC's 6th assessment report, in urban areas the risk to people and assets due to climate-related hazard has already increased, and climate impacts are felt disproportionately in urban communities, with the most economically and socially marginalized being most affected (Dodman et al., 2023). Such risks depend on the increase of intensity and frequency of extreme weather events (La Sorte et al., 2021; Mulholland & Feyen, 2021) as well as on the interplay with several non-climatic risk drivers including extent and features of the exposed systems and assets (e.g., European Environment Agency, 2024) and their vulnerability (e.g., Cutter & Finch, 2008; Dickson et al., 2012).

Exposure is intended as the presence of people; livelihoods; species or ecosystems; environmental functions, services, and resources; infrastructure; or economic, social, or cultural assets in places and settings that could be adversely affected, while vulnerability refers to the propensity or predisposition to be adversely affected. Vulnerability encompasses a variety of concepts and elements, including sensitivity or susceptibility to harm and lack of capacity to cope and adapt (Intergovernmental Panel on Climate Change - IPCC, 2022; Koren et al., 2017). It encompasses both the lack of coping

https://doi.org/10.5194/egusphere-2025-908
Preprint. Discussion started: 28 February 2025

capacity and adaptive capacity—factors that influence a community's ability to manage disasters effectively (Cardona et al., 2012; Marin Ferrer, 2017). The level to which urban settings are prone to the negative impacts of one or multiple hazards is also known as urban vulnerability (Thywissen, 2006), and its assessment is particularly challenging, as cities are intricate systems composed of interdependent networks of built environments, infrastructure, and social systems (Koren et al., 2017)

The concentration of assets and people may increase potential losses, while dynamic interactions between individual components that enable efficient system performance can lead to cascading failures. In addition, urban areas are often exposed to multiple hazards, such as earthquakes, floods, heatwaves, each interacting with the built environment and human activities in different ways. This complex interplay explains also why often non-extreme hazards can lead to severe consequences, while extreme events in other contexts may not result in disasters (Lavell et al., 2012).

Despite the high specificity of exposure and vulnerability of each urban environment, we assume that a relatively low number of representative archetypes could be found to decrease the level of complexity at regional and national scale, ensure a baseline for comparison and highlight potential hotspots in multi-hazard and multi-risk assessment frameworks.

We note that the concept of archetype is interpreted here as the optimal representation of a relatively homogeneous cluster of urban settlements. A disambiguation is necessary with respect to the statistical concept of archetype proposed by Cutler & Breiman (1994) which defines them as pure types whose mixture can be used to describe all elements in a dataset and can be considered as a form of extremal analysis. A different and broader interpretation of archetype in the context of sustainability research aims at finding recurrent patterns and similarities in the considered set of cases that are "crucial for describing the system dynamics or causal effect of interest" (Oberlack et al., 2019).

We therefore define urban settlement archetype as a representative instance (either real or ideal) of a cluster of urban settlements according to the selected indicators and the considered clustering approach and metrics.

2 A clustering proposal for decoding urban settlements archetypes

Archetypes may help manage the complexity of systems by highlighting essential patterns, facilitating better understanding and communication, and supporting more informed decision-making (Oberlack et al., 2023; Piemontese et al., 2022; Wicki et al., 2024). Archetypes have been extensively employed to classify cities based on socio-economic and socio-demographic parameters, to support policy decisions on fiscal interventions (Bruce, 1971; Dalton, 2015; Harris, 1943). An increasing amount of climate studies are dedicated to identifying recurring patterns and archetypes, in order to understand local climate vulnerabilities and to formulate specific adaption strategies (Rocha et al., 2020; Vidal Merino et al., 2019; Wicki et al., 2024). For instance, in Riach et al. (2023) recurring climate risk patterns at the municipal level in Baden-Wuerttemberg, Germany, are identified by analysing indicators for climatic hazards (e.g., annual mean temperature, hot/ice days, heavy precipitation) and exposure/vulnerability (e.g., proportion of elderly, energy production, population density). The nine urban archetypes derived represent municipalities with varying climate risk characteristics that require tailored adaptation measures.
Although several examples of city-scale archetypes analysis are available, they are often limited to the analysis of single-hazard risk (e.g., Awah et al., 2024; Carroll & Paveglio, 2016; Joshi et al., 2022; Riach et al., 2023).

Following the approach suggested in Piemontese et al. (2022), we perform the archetype analysis according to three phases of Design, Analysis and Application. In the *Design phase*, the problem framing and attributes selection is performed. In particular, this study seeks to address the challenge of assessing urban exposure and vulnerability by proposing a national-scale clustering of Italian urban settlements using open-source data. Municipality is selected as the primary geographical boundary for urban settlements since available authoritative open-source data is often referring to such administrative units. Municipalities are small, well-defined units, making them ideal for detailed spatial analysis and accurate identification of urban settlements. These boundaries often reflect historical settlements, preserving the cultural context that is essential for understanding contemporary urban dynamics. Additionally, municipalities are responsible for local governance and urban planning, making them relevant units for studying urban settlements, as local policies directly affect development and quality of life (actionability also for risk mitigation and climate change adaptation). The goal of this study is to group urban settings (municipalities) to define risk-oriented urban settlements archetype. To this end, the selected attributes are shared key drivers of urban vulnerability, as described in section 3. For the *Analysis phase*, methods of analysis should be defined, towards generalizability of results. The proposed methodology utilizes two widely-used clustering techniques—*agglomerative hierarchical clustering* and *partitioning clustering*—to analyse vulnerability-related data. By integrating a range of geographic, demographic, and socio-economic parameters, the study provides a robust assessment of the vulnerability of Italian urban settlements, identifying archetypes with varying levels of

https://doi.org/10.5194/egusphere-2025-908
Preprint. Discussion started: 28 February 2025

susceptibility to natural hazards. Using open-source data ensures the approach is both replicable and scalable, making it generalizable and applicable to other regions. Finally, the *Application phase* entails a real-world check of the archetypes identified towards their empirical validity, meaning they should correspond to variable levels of susceptibility to risk (according to the problem framing), and assessment of the impact, intended as the usefulness of results for application by final knowledge users. The identified archetypes offer a simplified framework for managing the complexity of diverse urban areas and their exposure to hazards. This risk-oriented classification offers valuable insights for urban resilience and disaster management professionals, enabling policymakers and urban planners to design targeted risk-reduction strategies tailored to the specific vulnerability profiles of each archetype, resulting in more efficient resource allocation.

Section 3 details the selection of geographic, demographic, and socio-economic factors considered representative of various dimensions of urban vulnerability. These factors, derived from open-source data, serve as the input for the clustering analysis. Section 4 explains the clustering process, including the metrics used to evaluate the performance of the clustering algorithms, and discusses the analysis results. Section 5 presents the urban settlement archetypes identified through the analysis. Section 6 describes the real-world validation of the identified archetypes, assessing their empirical validity. Finally, the conclusion highlights the potential applications and implications of the research findings, acknowledges the study's limitations, and provides recommendations for future research directions.

3 Key indicators of urban settlements vulnerability

To apply clustering techniques, it is essential to have a dataset containing meaningful features (attributes) that allow for clear differentiation between clusters. These attributes may include numerical, categorical, or mixed data types, depending on the clustering algorithm. Thus, the first step in clustering urban settlements at a national scale is to identify key drivers of vulnerability and assess data availability.

Vulnerability is multidimensional, defined by various physical, social, economic, environmental, and institutional factors that shape the susceptibility of systems to the impact of hazards (UNDRR., 2023; Van Westen & Woldai, 2012; Villagrán de León, 2006). Social vulnerability refers to the propensity of some social groups (e.g., poor, single-parent households, pregnant or lactating women, the handicapped, children, and elderly) to suffer negative consequences of hazards, due to their lack of capacity to react and manage the effect of hazard related processes (Cutter et al., 2003; Wisner et al., 2004). Economic vulnerability is the propensity of economic assets and processes to be harmed by exogenous shocks (Cardona et al., 2012), such as the potential impacts of natural and man-made hazards (i.e., business interruption, secondary effects such as increased poverty and job loss).

Table 1 – Vulnerability dimensions most common indicators.

Dimension	Indicator	Reference
Social	Dependency ratio	(Cutter et al., 2003; Eriksen & Kelly, 2007; Frigerio et al., 2018)
	Old age index	(Cutter et al., 2003; Frigerio et al., 2018; Marzi et al., 2019)
	Level of education	(Cutter et al., 2003; Frigerio et al., 2018; Marzi et al., 2019; Sibilia et al., 2024)
	Family structure	(Cutter et al., 2003; Frigerio et al., 2018; Marzi et al., 2019)
	Commuting rate	(Cutter et al., 2003; Frigerio et al., 2018; Marzi et al., 2019)
	Race/Ethnicity	(Cutter et al., 2003; Frigerio et al., 2018; Marzi et al., 2019)
	Access to medical services	(Cutter et al., 2003; Sibilia et al., 2024)
Economic	Employment rate	(Marzi et al., 2019; Opach et al., 2020; Sibilia et al., 2024)
	% women in the workforce	(Marzi et al., 2019; Opach et al., 2020)
	Household income	(Marzi et al., 2019; Sibilia et al., 2024)
	GPD per capita	(Eriksen & Kelly, 2007; Sibilia et al., 2024)
Physical	Housing conditions	(Marzi et al., 2019; Sibilia et al., 2024)
	Building typology/material/design	(FEMA, 2022; Kappes et al., 2012; Lagomarsino & Giovinazzi, 2006)
	Population density	(Marzi et al., 2019; Opach et al., 2020)
Institutional	Distance to services centres	(Marzi et al., 2019; Opach et al., 2020)
	Political stability	(Papathoma-Köhle et al., 2021; Sibilia et al., 2024)
	Risk awareness and perception	(Papathoma-Köhle et al., 2021)
	Transparency	(Papathoma-Köhle et al., 2021)

https://doi.org/10.5194/egusphere-2025-908
Preprint. Discussion started: 28 February 2025

Environmental	Vegetation cover/Land use	(Eriksen & Kelly, 2007; O'Brien et al., 2004; Sibilia et al., 2024)
	Water quality and availability	(Eriksen & Kelly, 2007; Rockstrom, 2013)
	Air pollution level	(Cohen et al., 2017; Eriksen & Kelly, 2007)

Physical vulnerability expresses the propensity of the built environment (e.g., buildings and infrastructure) and population to suffer the physical impact of hazardous events (Douglas, 2007). Institutional vulnerability may arise from weaknesses in governance and infrastructure, such as a lack of comprehensive disaster preparedness plans or poor coordination between agencies like emergency services and healthcare providers (Papathoma-Köhle et al., 2021). Environmental vulnerability is the susceptibility of ecosystems to sustain degradation (destruction of forest, farmland, or crops, lower yields) and loss of functionality following a hazardous event (Angeon & Bates, 2014; Marzi et al., 2019). Table 1 presents a list of key indicators used to assess each dimension of vulnerability mentioned.

In our work we focused on different categories of indicators expectedly linked with these vulnerability dimensions, and namely: geographical, demographic and socio-economic (Table 2). Geographic indicators include the altimetric zone and the urban centeredness degree. The altimetric zone of urban settlements, which refers to their elevation and topographical features, can be considered a proxy of access to the main services. As a matter of fact, accessibility of services of general interest can be particularly challenging in certain contexts (e.g. mountain regions, islands) due to their geomorphological and settlement structure conditions (Bertram et al., 2023). These accessibility issues can also complicate evacuation efforts and the delivery of emergency services during a disaster. Likewise, urban centeredness degree, that relates to the spatial characteristics and distribution of urban areas, is proxy for the availability of public services and the level of spatial connectedness. In classic urban geography, centrality is typically defined by attractiveness (Alonso, 1964; Isard W., 1956). This attractiveness can be assessed from various perspectives: historically, where centrality is viewed as historical heritage and the main place of memory for the urban community; architecturally, where it relates to the shape of urban blocks and the presence of monuments and significant buildings; and functionally, where it is linked to the presence, density, and types of activities and land uses (Cutini, 2001). Numerous studies also connect urban centrality to critical urban issues, such as the efficiency of transport systems and commuting patterns (Giuliano & Narayan, 2003; Schwanen et al., 2004). The degree of urban centeredness significantly influences the response and resilience of urban systems by affecting resource availability, infrastructure robustness, community networks, and emergency preparedness. Ensuring effective access to essential public services, such as healthcare and education, is challenging even under normal circumstances. However, it becomes even more crucial during crises like natural disasters, when the demands on these services and their operating conditions become significantly more complex (Fan et al., 2022; Loreti et al., 2022; Tariverdi et al., 2023). The level of peripherality of the areas with respect to the network of urban centres influence may determine not only difficulties of access to basic services but also lower quality of life of citizens and their level of social inclusion (Oppido et al., 2023).

Demographic indicators are linked to physical vulnerability. Residential population significantly influences the exposure to natural hazards, determining not only the higher presence of people and housing, but also of infrastructure, production capacities, species or ecosystems, and other tangible human assets in places and settings that could be adversely affected by one or multiple hazards. Higher population not only increases the potential for human and property losses, but also complicate evacuation efforts, and strain emergency response resources (Zhao et al., 2017). The other demographic parameter adopted is the degree of urbanization, often used to classify areas into cities, urban areas, and rural areas based on criteria such as population density, concentration of human activities, and built environment (Balk et al., 2018; United Nations, 2018). Indeed, highly urbanized densely populated areas are more likely to experience greater damage, congestion, and strain on resources during emergencies. It affects the capacity for evacuation and accessibility to essential services, due to dense infrastructure, complex urban layouts and the potential for cascading failures in infrastructure (Kendra et al., 2008; Lall & Deichmann, 2012). Finally, socio-economic factors taken into account are those parameters that influence both social and economic vulnerability, such as the presence of the elderly population, the employment rate and the educational level. For instance, past events highlight that elderly may be more vulnerable due to reduced mobility, poor health, and communication challenges (Ardalan et al., 2010; Carnelli & Frigerio, 2017; Cutter et al., 2003), while education levels can heighten vulnerability to natural hazards influencing risk perception and awareness, knowledge, and skills related to disaster preparedness (Alexander, 2012; Wachinger et al., 2013). Still, minority groups, including migrants, and ethnic communities, often face heightened social vulnerability, especially in high-risk areas, due to language barriers and communication challenges that can hinder access to critical emergency information (Carnelli & Frigerio, 2017; Walter Gillis et al., 2012).

Table 2 – Indicators selected for urban settlement clustering.

Indicator	Category	Vulnerability dimension
Altimetric zone	Geographical	Institutional

https://doi.org/10.5194/egusphere-2025-908
Preprint. Discussion started: 28 February 2025

Centeredness degree		
Urban degree Residential population	Demographic	Physical
Proportion of Children/ Elderly Aging index Dependency ratio Family structures Educational level Quality of buildings Commuting rate Employment/ Unemployment Female employment Crowding index Ethnicity /Foreign resident	Socio-economic	Social; Economic

It is worth mentioning that we only consider indicators for which publicly available data exist at the municipal level. For example, since GDP per capita is only available at national, regional, or provincial scales, it is not included in this study. Similarly, many building characteristics affecting physical vulnerability are either difficult to detect or unavailable at the municipal scale (e.g., structural system and earthquake-resistant design level; Tocchi et al., 2022). Moreover, building vulnerability indicators often vary depending on the type of hazard (Kappes et al., 2012), making it challenging to collect all relevant information for multiple hazards across Italy. For these reasons, only population density and general building quality are considered in this study. Indicators suggested for the environmental vulnerability dimension are not included due to data limitations as well. For instance, municipal-level air pollution data in Italy is limited, as such data is only available for major cities with monitoring stations.

Data on urban degree, urban centeredness degree, altimetric zone, social vulnerability factors used herein are primarily sourced from ISTAT (Italian National Institute of Statistics). All data are collected at the municipal level, aligning with the administrative boundaries adopted for the analysis. The dataset includes 7960 objects, representing the 7960 Italian municipalities, and 19 attributes (both numerical and categorical) related to the vulnerability factors outlined in sections 3.1 through 3.5.

3.1 Degree of urbanization

Eurostat (2021) proposed grid-based approach, implemented in geographic information systems (GIS), to determine the degree of urbanization based on a combination of geographical contiguity and population density. First, raster grid cells of 1 km² are categorized according to their total population and population density. Second, small statistical units are classified as urban centres (high-density units), urban clusters (moderate-density units) and rural cells (low density units) based on population thresholds and density criteria of groups of neighbouring cells. Finally, degree of urbanization of local administrative units is defined based on the share of population living in urban centres, urban clusters and rural grid cells: densely populated units (i.e., at least 50% of population living in urban centres) are categorized as *cities*, thinly populated units (i.e., at least 50% of population living in rural areas) as *rural areas* and intermediate density units (i.e., less than 50% of population living in rural areas and urban centres) as *towns and suburbs*. In Italy, classification of municipalities adopting to the above-mentioned Eurostat procedure is provided by ISTAT, the Italian National institute of statistics (https://www.istat.it/classificazione/principali-statistiche-geografiche-sui-comuni/) and reported in Figure 1.

It is found that only 3% of Italian municipalities are classified as cities, yet they account for 33% of the country's residential population. Conversely, rural areas make up 68% of municipalities but only represent 24% of the population. Towns and suburbs, which comprise 29% of municipalities, account for 43% of the population.

3.2 Urban centeredness degree

The urban centeredness degree is used herein as a measure of how centralized or decentralized an urban area is. Italian territory is a polycentric territory, i.e., a territory characterized by a network of municipalities or aggregations of municipalities (centres of offer of services) around which areas characterized by different levels of peripherality gravitate. These centres offer a wide range of essential services, capable of generating important catchment areas, even remotely, and of acting as "attractors" (in the gravitational sense). The methodology used to define the urban centeredness degree of municipalities is based on the approach proposed by the National Strategy for Internal Areas (SNAI, "Strategia

https://doi.org/10.5194/egusphere-2025-908
Preprint. Discussion started: 28 February 2025

Nazionale per le Aree Interne" in Italian). This territorial policy aims to enhance the quality of citizen services and economic opportunities in remote areas, which are characterized by significant distances from major service centres and are at risk of marginalization. The proposed methodology involves two main phases: i) identifying urban hubs based on their capacity to provide essential services and ii) classifying the remaining municipalities as peri-urban areas and inner areas based their distance from the hubs measured in travel time (DPS, 2013).

More specifically, the selection of hubs, which can also be defined as service offering centres, is based on service availability indicators for high school educational services (e.g., high schools, technical and professional institutes, and other higher education institutions), health services (e.g., presence of multiple health and emergency facilities, healthcare facilities with at least 250 beds), and rail transport services (e.g., train stations with an average of more than 6,000 travellers per day and a high number of daily trains). Some neighbouring municipalities are classified as intermunicipal hubs, meaning that several contiguous municipalities collectively provide the required level of services in a network system. The remaining municipalities are classified based on an accessibility indicator measured in minutes to reach the nearest hub. Peri-urban areas are less than 20 minutes away from the nearest hub, while inner areas are more than 20 min away. Further classification of the inner areas into four categories is also provided: it is possible to distinguish between intermediate areas, that are approximately 20 to 40 minutes away, peripheral areas, that are between 40 and 75 minutes away, and ultra-peripheral territories, that are more than 75 minutes away. Figure 1 (right) shows the classification of Italian municipalities based on their urban centeredness degree.

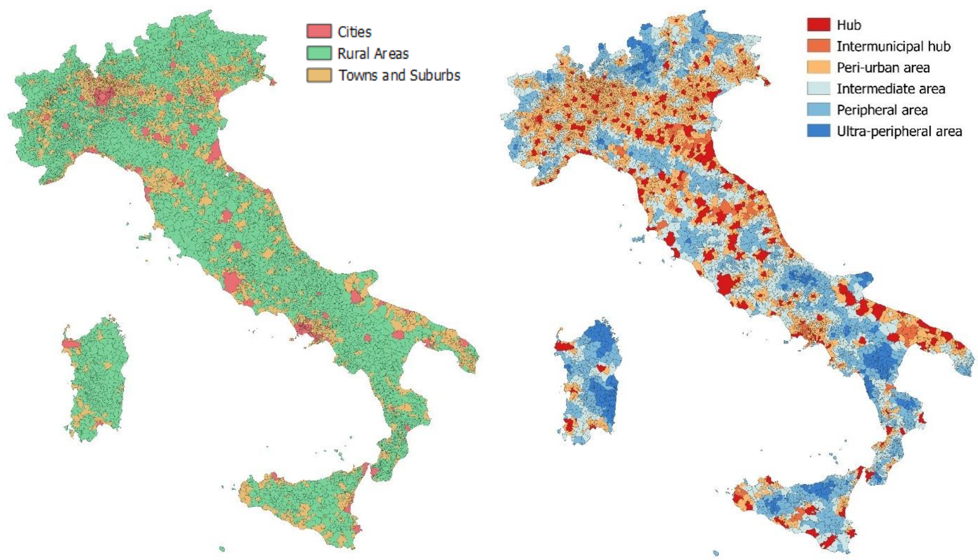


Figure 1 – Italian municipalities classified based on urban degree (left) and urban centeredness degree (right). Data used for the classification are derived from ISTAT (https://www.istat.it/classificazione/principali-statistiche-geografiche-sui-comuni/).

Most of the Italian population resides in peri-urban areas (37%) and urban hubs (35%), which account for 44% and 3% of municipalities, respectively. Intermediate, peripheral and ultra-peripheral areas account for 16%, 6% and 1% of population, respectively, and represent 28% (intermediate), 19% (peripheral), and 4% (ultra-peripheral) of Italian municipalities. Intermunicipal hubs represent only 2% of municipalities and house 5% of the population. Population density generally decreases from hubs to peripheral municipalities. High-density cities comprise 35% of hubs, while medium-density towns and suburbs make up 57%. Only 8% of hubs are low-density rural areas. Intermunicipal hubs exhibit medium-high population density, with 23% classified as cities, 50% as towns and suburbs and 27% as rural areas. Peri-urban municipalities exhibit medium-low population density, with 5%, 45% and 50% classified as cities, towns and suburbs and rural areas, respectively. Intermediate, peripheral, and ultra-peripheral municipalities are mostly low-density rural regions (83%, 91%, and 96%, respectively).

It is worth mentioning that only three classes are considered for the urban centeredness degree, namely urban hubs (represented by both hubs and intermunicipal hubs), peri-urban areas and inland areas (that includes intermediate, peripherical and ultra-peripherical areas), according to the main classification proposed by ISTAT. This simplification is

https://doi.org/10.5194/egusphere-2025-908
Preprint. Discussion started: 28 February 2025

adopted in order to: (i) minimizes noise and variability in the data, leading to more stable and reproducible clusters; (ii) prevents the model from overfitting to minor variations, improving generalizability; (iii) enhanced interpretability.

3.3 Altimetric zone

ISTAT classifies Italian municipalities into three altimetric zones based on elevation: mountain (>600 m.a.s.l.), hill (300-600 m.a.s.l.), and lowland (<300 m.a.s.l.) (ISTAT, 2020). Elevation data is derived from a Digital Elevation Model (DEM) developed by ISPRA (Italian Institute for Environmental Protection and Research) for a 20-meter grid. Using the DEM, statistics such as average, sum, minimum, and maximum elevation within the municipal boundaries are calculated using a zonal statistics tool in GIS software. The municipality's altimetric zone is then determined based on the surface prevalence criterion. Municipalities could be also subdivided to account for the moderating influence of the sea on the climate, as coastal or inland areas. However, only the main three altimetric classes are adopted in this study (i.e., mountain, hill and lowland), not considering the classification in coastal and inland zones. Indeed, coastal zones represent a minority of municipality (almost 90% of municipalities are located in inland areas, while only 10% in coastal areas) and this can lead the model becoming more likely to fit to noise, reducing its generalizability to new data. Furthermore, despite some correlation may exist between urban vulnerability and coastal/inland areas (e.g., a coastal city with strong infrastructure and preparedness may be less vulnerable than an inland town with weak governance and high poverty), generally the distinction between coastal and inland areas is primarily linked to the types of natural hazards affecting these regions rather than inherent differences in urban vulnerability. Figure 2 shows the classification of Italian municipalities by population classes and altimetric zones.

Geographically, 31% of municipalities are in mountainous areas (30% inland, 1% coastal), accounting for just 12% of the population (10% inland, 2% coastal). Hill areas encompass 43% of municipalities, with 33% in inland hills and 10% in coastal hills, representing 23% and 16% of the Italian population, respectively. Lowland areas include 26% of municipalities (24% inland, 2% coastal) and home to 49% of the population (34% inland, 15% coastal).

Many densely populated cities are located in lowland areas (75%), while only 22% are situated in hilly regions and 3% in mountainous areas. Low-density or rural municipalities are predominantly found in mountainous (39%) and hilly regions (42%), compared to only 19% in lowlands. Medium-density towns and suburbs are mostly located in lowlands (38%) and hills (45%), with only 17% in mountainous areas. Similarly, hubs (including intermunicipal hubs) and peri-urban municipalities are primarily situated in lowland (47% and 44%, respectively) and hilly areas (44% and 41%), with only 9% of hubs and 15% of peri-urban municipalities in mountainous regions. In contrast, most inner municipalities are in mountainous (47%) and hilly regions (44%), while only 9% are found in lowlands.

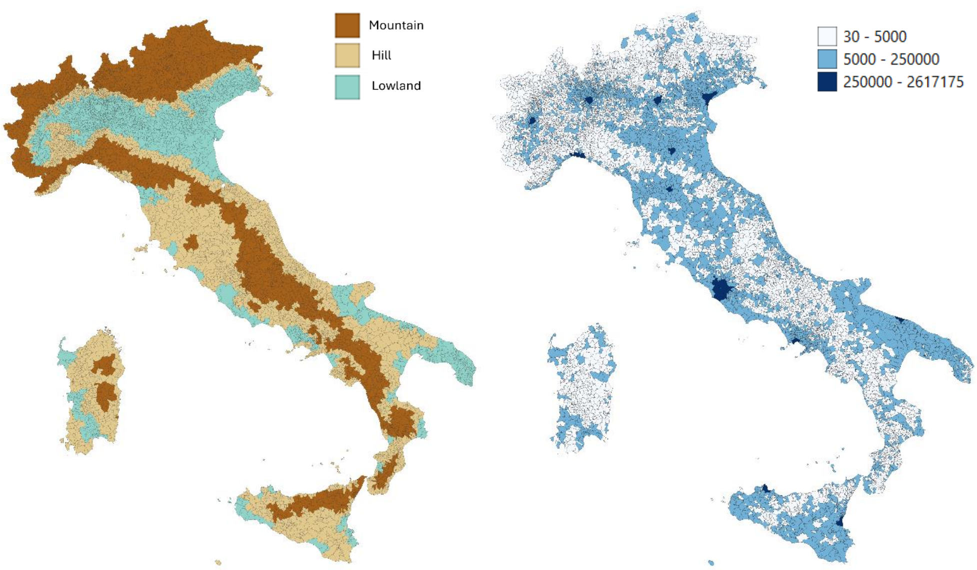


Figure 2 - Italian municipalities classified based on altimetric zone (left) and population (right). Data used for the classification are derived from ISTAT (https://www.istat.it/classificazione/principali-statistiche-geografiche-sui-comuni/).

https://doi.org/10.5194/egusphere-2025-908
Preprint. Discussion started: 28 February 2025

3.4 Residential Population

Three population classes (C_{pop}) were introduced by ISTAT to classify municipalities according to the number of inhabitants (ISTAT, 2020). The classes (Fig. 2) are defined using the following population thresholds: small municipalities (less than 5000 inhabitants, C_{pop}=1); medium municipalities (between 5001 to 250000 inhabitants, C_{pop}=2); big municipalities (more than 250000 inhabitants, C_{pop}=3). ISTAT provides updated statistics on the resident population per municipality every year. In this study information on population per municipality is updated to 2018, along with the most recent data on urbanization, centrality, and altitude zones, all referring to the same year.

A significant proportion of municipalities fall into the lowest population class, with 33% having between 501 and 2000 inhabitants, 26% having between 2001 and 5000 inhabitants and 11% being very sparsely populated, with fewer than 500 residents. The remain municipalities belong to the medium population class (30%), including 15% with population between 5001 and 10000 inhabitants, 13% between 10001 and 50000, and 2% between 50001 and 250000. Only 0.2% belong to the high population class, representing Italy's largest cities such as Rome and Milan.

As expected, 91% of hubs (including intermunicipal hubs) are found in higher population classes, with 87% in population class 2 and 4% in population class 3, highlighting their concentration in highly populated areas. In contrast, 84% of inner municipalities fall within the lowest population categories, with only 16% classified in class 2. Similarly, densely populated cities tend to have larger populations, with 73% in classes 2 and 3 (41% having between 10001 and 50000 inhabitants, and 32% exceeding 50001 inhabitants). Meanwhile, 88% of rural areas are also sparsely populated, with the vast majority (88%) having fewer than 5000 inhabitants.

3.5 Social vulnerability indicators

Parameters commonly used to assess social vulnerability, such as gender, age, education, socioeconomic status, public health condition, employment status, and access to resources, need to be tailored to the local context to accurately reflect place-specific dimensions (Chen et al., 2013; Cutter et al., 2003; Guillard-Gonçalves et al., 2015; Mesta et al., 2022). The variables representing social vulnerability adopted in this study are detailed in Table 3. These variables, which encompass seven demographic and socio-economic indicators pertinent to the Italian context, are inspired by the study conducted by Frigerio et al (2018).

Table 3 - Variables that characterize social vulnerability. The last column reports their impact on social vulnerability, i.e., increasing (+) or decreasing (-), according to Frigerio et al. (2018).

Variables	Description	Impact on social vulnerability
Children	Percentage of population aged under 15	+
Elderly people	Percentage of population aged over 65	+
Aging index	Ratio of elderly people compared to children	+
Dependency ratio	Ratio of nonworking-age people to those of working age	+
Families with more than 5 components	Proportion of families with more than five members	+
High educational index	People with at least a university degree compared to the total population aged over 15	-
Low educational index	People with at most a secondary school diploma compared to the total population aged over 15	+
Quality of buildings	Buildings with very bad or bad state of preservation	+
Commuting rate	Ratio of commuters to working-age population	+
Unemployed	Proportion of unemployed among working-age population	+
Employed	Proportion of employed among working-age population	-
Female employed	Proportion of employed women of working-age	-
Crowding index	Number of persons per dwelling	+
Foreign resident	Proportion of foreign population	+

Age indicators include the percentage of children (under 15) and elderly (over 65), the ageing index, calculated as ratio of elderly to children (Preedy & Watson, 2010), and the dependency ratio, i.e., ratio of nonworking-age people to working-age people (Simon et al., 2012), calculated as those under 15 and over 65. The family structure indicator measures the proportion of families with more than five members. Indeed, evidence shows that the larger the family the lower the

https://doi.org/10.5194/egusphere-2025-908
Preprint. Discussion started: 28 February 2025

income (ISTAT, 2024). The education indicator consists of the low educational index, calculated as number of people with at most a secondary school diploma compared to the total population aged over 15, and the high educational index, calculated as people with at least a university degree compared to the total population aged over 30. The commuting rate is the ratio of commuters to working-age people, while building quality indicator is calculated as the proportion of buildings in poor condition. In the census database, building quality is classified based on four categories of preservation: very good, good, bad, or very bad. For this study, the number of buildings in bad or very bad condition at the municipal level is used as a representative variable for building quality. The employment indicators cover unemployment, employment, and female employment rates among working-age people. The crowding index is calculated herein as the number of persons per dwelling. The foreign residents indicator is defined in terms of percentage of foreign population (i.e., not Italian citizens). Each variable is derived from last census (ISTAT, 2011) at census tract level and aggregated at municipal level.

It was observed that aging index, dependency ratio, low educational index and the percentage of buildings in poor conditions tends to increase from hubs to ultra-peripheral areas (average values of 1.7, 0.55, 0.51 and 14% for hubs, average values of 2.7, 0.61,0.60 and 20% for ultra-peripheral areas, respectively), while high educational index, percentage of employed and female employed as well as crowding index tend to decrease (average values of 0.17, 50%, 44% and 2.4 for hubs, average values of 0.08, 44%, 38% and 2.2 for ultra-peripheral areas, respectively). Rural municipalities exhibit a higher aging index (2.2) compared to towns (1.4) and cities (1.2), along with larger values of low educational index (0.61 vs. 0.58 in towns and 0.55 in cities) and dependency ratios (0.58 vs. 0.52 in towns and 0.50 in cities). Conversely, these rural areas show smaller values of high educational index (0.09 compared to 0.11 in towns and 0.15 in cities), employment rates (0.48 vs. 0.52 in towns and cities), and crowding index (2.3 compared to 2.5 in towns and 2.6 in cities).

Social vulnerability is often expressed through a composite index known as the Social Vulnerability Index (SoVI), which aggregates different metrics affecting it (e.g., Cutter et al., 2003; Frigerio et al., 2018). Using a unique index to represent social vulnerability provides a comprehensive and easily interpretable measure that encapsulates multiple dimensions of vulnerability, facilitating communication and policymaking. In this study the individual indicators affecting social vulnerability are considered in the cluster analysis, while the aggregated SoVI index is used to provide a synthetic description. This approach allows for a more nuanced understanding of the different dimensions of vulnerability. By analysing each indicator separately, cluster analysis can capture the unique contributions and relationships between factors like income, education, health, and housing quality, which may be masked in a single composite index. Additionally, considering individual indicators enables the identification of distinct patterns or subgroups within the data, leading to more effective archetype identification. In contrast, an aggregated index may oversimplify these dynamics and overlook important variations across clusters.

4 Cluster analysis

To identify archetypes by grouping entities based on shared characteristics, clustering analysis is widely used. Clustering refers to unsupervised learning techniques used to find subgroups (or clusters) within a data set by organizing elements according to their similarities. This method ensures that observations within the same group are highly similar, while those in different groups are distinctly different. Unlike supervised classification algorithms, which rely on labelled training data to categorize new information into predefined classes, clustering uncovers natural structures in the data by analysing similarities between data points without the need for predefined labels.

In this study clustering is adopted to group urban settlements based on their potential exposure and urban vulnerability-related factors to define risk-oriented urban archetypes, proposing an application for Italian municipalities as a case. We conducted a two-step clustering approach. A first cluster analysis is performed with a sub-set of attributes, specifically demographic and geographic parameters. The aim is to allow a first broad definition of urban archetypes, simple and highly interpretable. In the second step, a nested clustering approach is applied to further differentiate sub-clusters based on socio-economic attributes. Both hierarchical and partitioning clustering techniques are employed in each step to enable comparison of clustering outputs and to identify nuanced patterns that may not be captured by a single method. Most detailed information of algorithms used are reported in section 4.2 and 4.3. For each step, urban archetypes (presented in section 5) are defined based on the results of the most effective algorithm, selected according to widely used clustering performance metrics (see section 4.3 and 4.5).

https://doi.org/10.5194/egusphere-2025-908
Preprint. Discussion started: 28 February 2025

4.1 Data pre-processing

Data preprocessing is crucial for enhance quality of clustering. Specifically, we performed: (i) outlier detection; (ii) correlation analysis – to eliminate measurement redundancy; (iii) normalization of numerical data values.

The detection of outliers is necessary to ensure high quality of clustering. An outlier is an object in a data set that deviates significantly from the remaining data, Outlier detection is essential for ensuring high-quality clustering, as extreme values can distort normalization and affect cluster formation (Nowak-Brzezińska & Gaibei, 2022; OECD, 2008).Outliers are identified using the interquartile range (IQR) method, where data points beyond 1.5 times the IQR from the quartiles are considered outliers. In this study, residential population had the highest number of outliers. Instead of removing them, which would compromise the analysis, the population variable was transformed into categorical classes (i.e. the population classes presented in section 3.4) to reduce its impact on clustering. The 18 attributes selected for the clustering are thus divided into 4 categorical (i.e., urban degree, population class, urban centeredness degree and altimetric zone) and 14 numerical (the ones listed in table 3).

For the correlation analysis, the Pearson correlation coefficient (*r*) is used. This coefficient is a statistical measure used to assess the strength and direction of the linear relationship between two continuous variables (Cohen, 2013). It is one of the most used methods for correlation analysis. While it does not inherently assume normality, research indicates that Pearson's correlation is relatively robust to violations of normality, especially when sample sizes are large (Bishara & Hittner, 2017). The value of *r* ranges between -1 and 1, with values higher than 0 that indicate a positive correlation and values lower than 0 a negative correlation. Values close to 0 indicate no linear correlation between variables (Cohen, 2013). Figure 3 shows the correlation matrix obtained for the 14 numerical variables considered. The analysis shows that there is a very strong correlation ($| r | > 0.8$) between age indicators, namely: proportion of under 14 aged and over 65 aged ($r = -0.84$); aging index and proportion of people aged 65 and above ($r = +0.81$); dependency ratio and proportion of over 65 ($r = 0.91$). Strong correlation ($| r | > 0.5$) is also observed for: aging index and dependency ratio to proportion of people under 14 (r equal to -076 and –0.55, respectively); dependency ratio and aging index ($r = +0.73$); crowding index and dependency ratio ($r = -0.6$); crowding index and proportion of families with 5 or more components ($r = +0.78$), high and low educational index ($r = -0.78$), proportion of commuters and proportion of employed ($r = +0.61$); proportion of employed and proportion of over 65 aged ($r = -0.58$) and proportion of under 14 aged ($r = +0.54$).

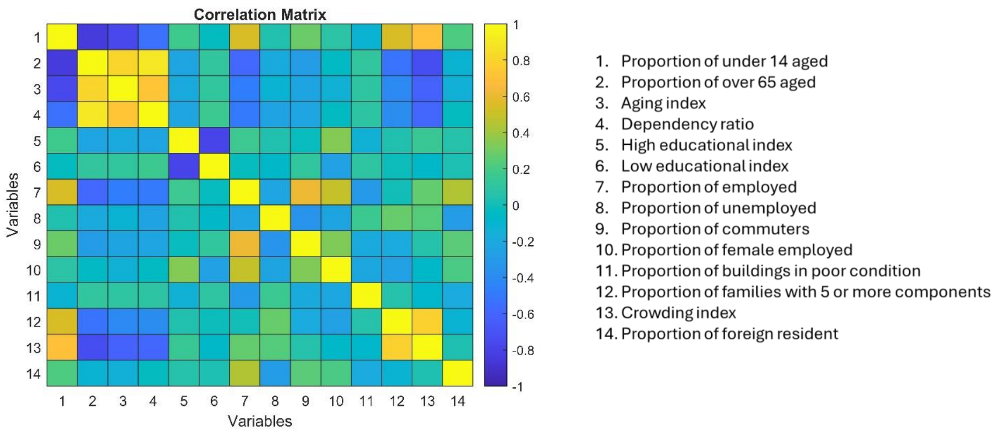


Figure 3 – Correlation matrix for numerical variables considered.

Based on the analysis results, the number of numerical attributes used for clustering was reduced from 14 to the following 8: aging index, low educational index, proportion of unemployed, proportion of commuters, proportion of female employed, proportion of buildings in poor condition, crowding index and proportion of foreign resident.

Finally, numerical data are normalized to enhance the quality of clustering by ensuring that all features contribute equally to the analysis, regardless of their original scales. Without normalization, features with larger ranges could dominate the clustering process, leading to biased results (Usman & Stores, 2020). As normalization method, the empirical cumulative distribution function (ECDF) is adopted. The empirical CDF approach ranks the data points by their cumulative probability, effectively distributing them between 0 and 1 based on their relative positions within the dataset. Compared to other normalization methods (e.g., min-max normalization, z-score), ECDF normalization offers several advantages: it

https://doi.org/10.5194/egusphere-2025-908
Preprint. Discussion started: 28 February 2025

effectively processes non-normally distributed data, minimizes the impact of outliers, and provides a clear, intuitive framework for interpreting data rankings relative to the overall distribution (Hoffman et al., 2017).

4.2 Hierarchical clustering

Hierarchical clustering organizes data into tree-structured clusters through either an agglomerative or divisive process (Han et al., 2011). In agglomerative clustering, each object is initially assigned to an individual cluster (that is, if there are n objects, the process will start with n clusters). Initial clusters are gradually merged into larger clusters based on their similarity (or dissimilarity), until a hierarchy of clusters is built and only one cluster remains, which contains all data points. The selection of an appropriate distance or dissimilarity measure crucially affects the clustering solution and depends on the nature of the considered variables. Most distance measures concern the analysis of either continuous only or categorical only data (e.g., Euclidean distance). The Gower distance (Gower, 1971) is a flexible dissimilarity measure that can work both with numerical and categorical variables. For numerical variables, the Gower distance uses the normalized absolute difference. If x_{ik} and x_{jk} are the values of the *l-th* numerical attribute for objects i and j, the distance between the two objects for attribute l is calculated as:

$$d_{ij}^{l} = \frac{|x_{il} - x_{jl}|}{\max(x_l) - \min(x_l)} \tag{1}$$

For categorical variables, the Gower distance assigns a value of 0 if the values are the same and 1 if they are different:

$$d_{ij}^{l} = \begin{cases} 0 & if\ x_{il} = x_{jl} \\ 1 & if\ x_{il} \neq x_{jl} \end{cases} \tag{2}$$

The Gower distance for a pair of objects i and j is calculated as the average value of the individual attribute distances, according to Eq. (3):

$$D_{ij} = \frac{\sum_{l=1}^{p} w_l d_{ij}^{l}}{\sum_{l=1}^{p} w_l} \tag{3}$$

Where p is the number of attributes, d_{ij}^{l} is the individual attribute distance and w_l is the weight for the *l-th* attribute, set to 1.

The main steps of agglomerative hierarchical clustering process can be outlined as follows:

1) For each pair of data points i and j in the dataset, the Gower distance is calculated. The result is an $n \times n$ distance matrix where each entry (i, j) represents the Gower distance between objects i and j across all attributes. This means that the values on the diagonal of this matrix will be equal to zero, since the distance between object i and itself is zero. It is important to note that the Gower distance already normalizes numerical variables, making additional normalization unnecessary.
2) Next, the two clusters with the smallest Gower distance are merged, reducing the total number of clusters by one.
3) The distance matrix is then updated to reflect the distance between the newly formed cluster and all other clusters. The recalculation of distances depends on the chosen linkage criterion (e.g., single, complete, average, or ward linkage). In this study, complete linkage is employed, where all pairwise dissimilarities between observations in cluster A and cluster B are computed, and the largest of these dissimilarities is recorded.
4) Repeat the merging process iteratively, continuing to merge the closest clusters and updating the distance matrix until all objects are grouped into a single cluster.

4.2.1 Optimal number of clusters

Throughout the merging process, a dendrogram is constructed—a tree-like diagram that visually represents the order and levels at which clusters are merged. The height of each node in the dendrogram corresponds to the Gower distance at which the clusters were merged. The dendrogram can be analysed to determine the optimal number of clusters, either visually by identifying the largest vertical distance (gap) between merges, known as the "cut" point (James et al., 2017), or by evaluating the inconsistency coefficient (Martin et al., 2022). The inconsistency coefficient measures the similarity of clusters connected by each link, comparing its length with the average length of other links at the same level of the dendrogram (Jatain et al., 2013). A higher coefficient indicates less similarity between clusters. The relationship between the inconsistency coefficient and the number of clusters indicates that a lower number of clusters corresponds to higher inconsistency, which suggests better clustering since distinct clusters tend to have high inconsistency. However, fewer clusters often lead to greater within-cluster variance. To strike a balance between distinct clusters and minimizing within-

https://doi.org/10.5194/egusphere-2025-908
Preprint. Discussion started: 28 February 2025

cluster variance, the variability of observations within each cluster is evaluated, and its trend is analysed as the number of clusters increases.

To evaluate the variability of the observations within each cluster, a coefficient representing the within cluster distance (WCD) is calculated as sum of the average values of distances between data points in a single cluster. Specifically, for the *l-th* numerical attribute WCD_l is calculated by taking the average of the squares of the differences between each pair of values i and j in the cluster C_k (Gordon, 1986):

$$WCD_l(C_k) = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n}(x_{il} - x_{jl})^2}{n^2 - n} \tag{4}$$

Where $x_1, x_2, \ldots, x_n$ are n observations within the *k-th* cluster on a quantitative variable, x.

For the *l-th* categorical attribute, the coefficient of unlikeability proposed by Perry & Kader (2005) is utilized as measure of WCD_k:

$$WCD_l(C_k) = \frac{\sum_{i \neq j} c(x_{il}, x_{jl})}{n^2 - n} \tag{5}$$

Where:

$$c(x_i, x_j) = \begin{cases} 1 & if\ x_{il} \neq x_{jl} \\ 0 & if\ x_{il} = x_{jl} \end{cases} \tag{6}$$

And $x_1, x_2, \ldots, x_n$ are n observations within the *k-th* cluster on a categorical variable, x.

The final value of WCD for the cluster C_k is given by the average values across all p attributes:

$$WCD\ (C_k) = \frac{\sum_{l=1}^{p} WCD_l(C_k)}{p} \tag{7}$$

While the overall value of WCD for the clustering – which accounts for all m clusters - can be defined as the average value:

$$WCD = \frac{\sum_{k=1}^{m} WCD(C_k)}{m} \tag{8}$$

The evaluation of variation of WCD with the increasing number of clusters together with the trend of inconsistency coefficient allows the definition of the best number of clusters for the specific case analyses.

4.3 Partitioning clustering

Partitioning clustering is a method that divides a dataset into a predefined number of clusters by assigning each data point to a single cluster based on similarity. The most used partitioning clustering technique is the k-means algorithm (MacQueen, 1967). To perform k-means clustering, the number k of clusters must be predefined, and k objects, representing the initial cluster centroids, are arbitrarily chosen. The remaining objects are then iteratively assigned to these clusters in a way that minimizes the distances of points to their respective centroids, thereby minimizing the within-cluster variance. The position of each centroid is updated each iteration by the mean value of the objects in a cluster. One of the main drawbacks of k-means algorithm is that it only works on numeric values, prohibiting its use to cluster data containing categorical values. The k-modes algorithm is an extension of k-means algorithm that employs a simple matching dissimilarity measure to handle categorical data, replacing cluster means with modes and using a frequency-based approach to update these modes during the clustering process (Huang, 1998). These modifications enable the k-modes algorithm to cluster categorical data in a manner similar to k-means. The k-prototypes algorithm combines elements of the K-means algorithm and the K-modes algorithm, allowing for the clustering of objects characterized by both numeric and categorical attributes (Huang, 1998). Like the k-means algorithm, this technique requires the user to set the number of clusters (k), while initial cluster centroids are chosen arbitrarily. Observations are iteratively assigned to the closest centroid in a way that minimizes within-cluster variance. To define the closeness between two objects, this method applies Euclidean distance to numeric attributes and uses a distance function simple matching dissimilarity (δ=0 if the values match, δ=1 if they do not) for categorical attributes. Thus, dissimilarity measure for a data point i and centroid j can be calculated as:

https://doi.org/10.5194/egusphere-2025-908
Preprint. Discussion started: 28 February 2025

$$D(i,j) = \sum_{numerical} (x_{il} - \mu_{kl})^2 + \gamma \sum_{categorical} \delta(x_{il}, \mu_{kl}) \tag{9}$$

Where x_{il} is the value of the *l-th* attribute of the *i*-th data point, μ_{kl} is the relative value for the *k-th* cluster centroid, and γ is a weighting factor to balance numerical and categorical distances (a value of 1 is adopted for γ). The centroids are updated after each iteration by taking the average values of numerical variables of the objects within a cluster and evaluating the modes for categorical attributes, i.e., the category with the highest frequency.

As one of the main drawbacks of this clustering techniques is that the clustering is very sensitive to the selection of initial centroid, the random selection of initial centroids and the clustering are repeated *t* times (*t*=10) and each iteration *t* the performance of the clustering algorithm is evaluated by calculating WCD. Among *t* different clustering obtained, the best clustering is determined based on WCD value (i.e., the clustering providing the lower WCD value).

4.4 Clustering based on demographic and geographic features

A preliminary cluster analysis is conducted using only demographic and geographic attributes, namely: urban degree, residential population, centeredness degree and altimetric zone. It is important to note that all these attributes are categorical. Consequently, hierarchical clustering is performed using Eq. (2) and (3), while partitioning clustering is applied using the dissimilarity measures for categorical variables presented in Eq. (9).

The optimal number of clusters is determined following the procedure outlined in Section 4.2.1. The inconsistency coefficient reaches its highest values (>4) when considering between 4 and 12 clusters. Conversely, the within-cluster dispersion (WCD) exhibits an inverse relationship with the inconsistency coefficient: as the number of clusters increases, WCD decreases. Specifically, WCD declines from 1.26 with 4 clusters to 0.70 with 10 clusters, remaining constant at 0.70 between 10 and 12 clusters. To achieve a balanced trade-off between the inconsistency coefficient and WCD, 10 clusters are selected as the optimal number for this dataset. Since partitioning clustering requires a predefined number of clusters, the optimal number identified through this methodology is also adopted for the partitioning clustering approach.

4.4.1 Results of hierarchical cluster analysis

Figure 4 shows the representativeness of clusters in terms of attributes considered. Clusters 1, 2, 3 and 6 represent rural municipalities with low (1, 2 and 3) and medium (6) population. Among them, cluster 1 identifies hubs (more specifically intermunicipal hubs), cluster 2 peripheral areas, cluster 3 peri-urban areas while cluster 6 includes both peripheral and peri-urban municipalities – with a very small portion of intermunicipal hubs.

Clusters 4, 5 and 7 identifies medium-density towns and suburbs with medium (4), medium-low (7) and low population (5). Cluster 4 is represented by hubs (specifically, intermunicipal hubs), cluster 7 by peripheral areas while cluster 5 in majority by peri-urban areas. Clusters 8 and 9 include high density cities, that are medium (8) and high (9) populated areas. Cluster 9 includes all the major Italian cities with more than 250’000 inhabitants (e.g., Rome, Milan and Naples), while cluster 8 includes cities located in peri-urban areas. Finally, cluster 10 includes all medium-populated municipalities not included in the other clusters, most of the medium-densely populated and located in peri-urban and peripheral areas.

Regarding the altimetric zone, most peripheral areas (e.g., clusters 2 and 7) are located in hilly and mountainous regions, whereas densely populated cities (e.g., clusters 8 and 9) are primarily situated in lowland areas. Cluster 4 distinctly represents towns and suburbs in hilly regions, while Cluster 6 includes rural municipalities in mountainous areas. Cluster 1 groups intermunicipal hubs found in both hilly and mountainous areas. However, the classification of municipalities in clusters 3, 5, and 10 is less clearly defined.

https://doi.org/10.5194/egusphere-2025-908
Preprint. Discussion started: 28 February 2025

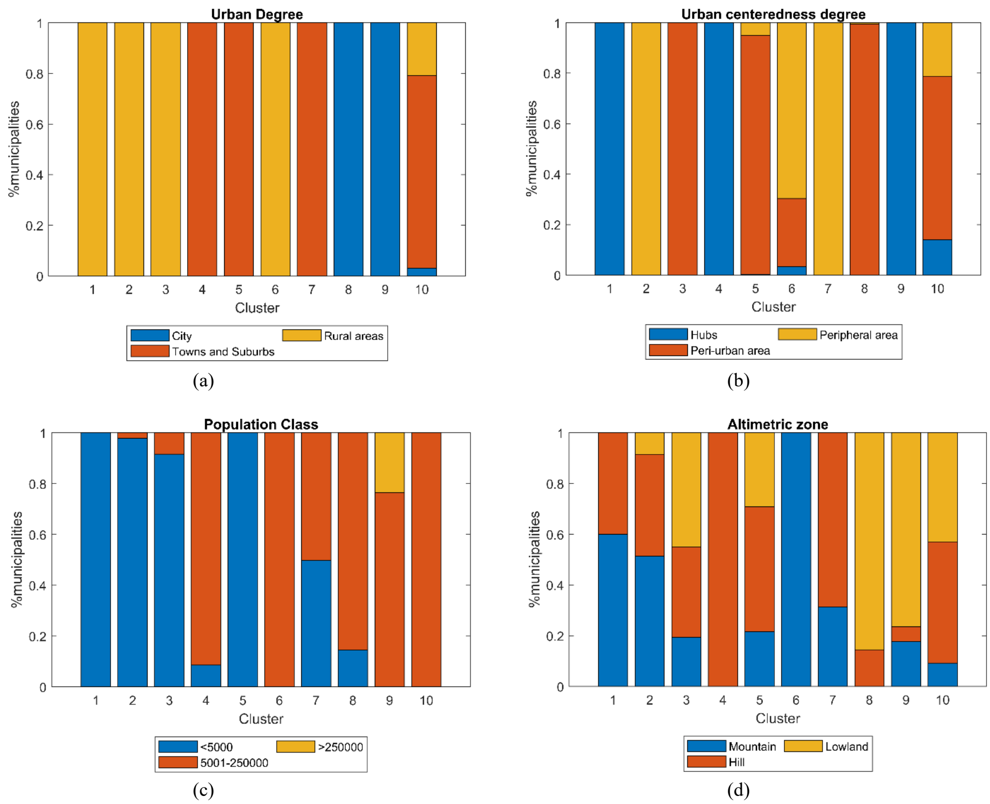
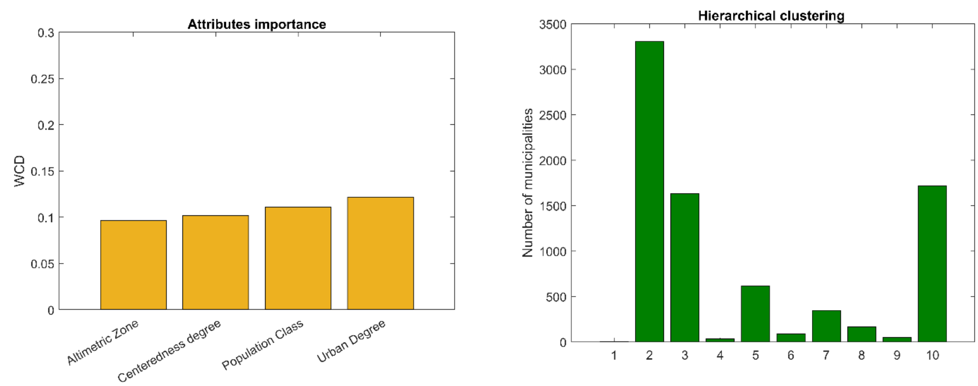


Figure 4 - Representativeness of clusters in terms of urban degree (a), urban centeredness degree (b), population class (c) and altimetric zone (d).

Attributes importance for clustering is evaluated adopting simplified procedure proposed in Fraiman et al. (Fraiman et al., 2008). The methodology is based on an iterative removal of variables, to assess its contribution to clustering based on the performance metric selected. In other words, variables are removed one at a time and the impact of the removal on the overall model is measured. The greater the impact of removing a variable, the more important it is considered. In this study we consider WCD as performance metric. Figure 5a shows that the most important attribute for the clustering is the urban degree, followed by the population class and the centeredness degree, while the less important attribute is the altimetric zone.

https://doi.org/10.5194/egusphere-2025-908
Preprint. Discussion started: 28 February 2025



(a) (b)

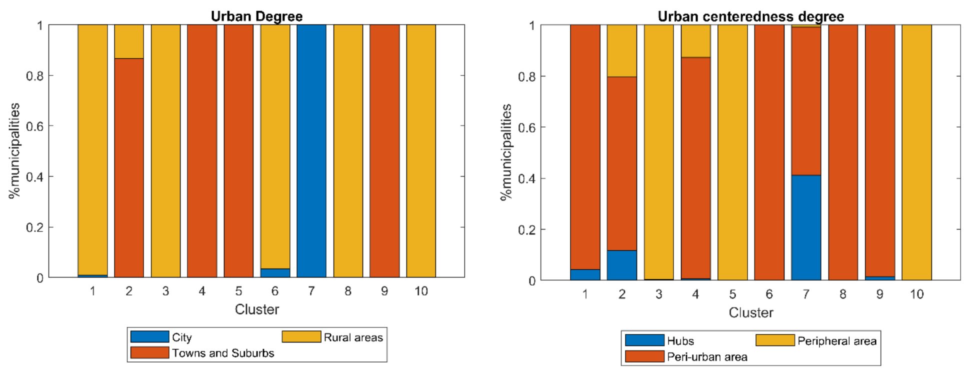
Figure 5 – Attribute importance in terms of variation of WCD (a); number of Italian municipalities belonging to each cluster (b).

A significant number of municipalities are classified within rural clusters, with Cluster 6 encompassing 3,305 municipalities and Cluster 3 including 1,632 municipalities (Fig. 5b). This aligns with the fact that 68% of Italian municipalities are categorized as rural areas (see also Section 3.4). The least populated cluster is Cluster 1, which contains only 5 municipalities, followed by Cluster 4 (35 municipalities), Cluster 9 (51 municipalities), and Cluster 6 (89 municipalities). Meanwhile, Clusters 8, 7, 5, and 10 include 166, 347, 615, and 1,715 municipalities, respectively.

4.4.2 Results of partitioning cluster analysis

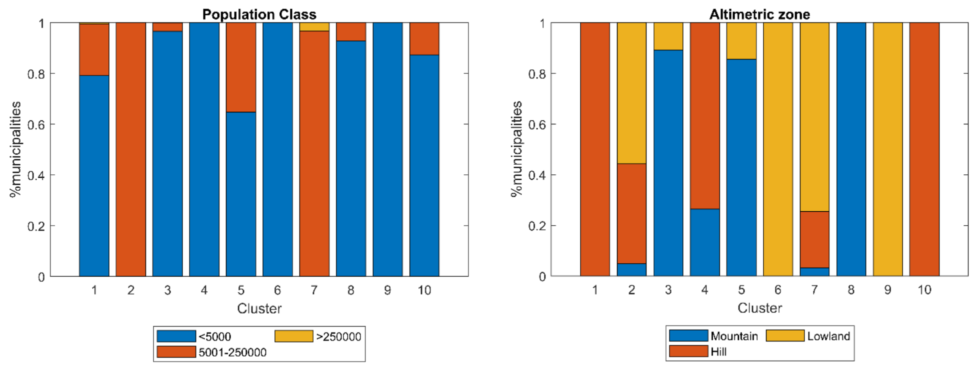
Results of partitioning clustering are presented in Fig. 6, highlighting the representativeness of different attributes. Clusters 3 and 10 represent low populated peripheral municipalities in rural areas, specifically located in mountainous (3) and hilly (10) regions. Clusters 1, 6 and 8 also include low populated municipalities in rural areas but classified as peri-urban, located in hilly (1), lowland (6) and mountainous areas (8). Clusters 4 and 9 identify low populated suburban municipalities in peri-urban areas, with cluster 4 representing those in mountainous and hilly regions, and cluster 9 those in lowland areas. Clusters 2 and 5 characterize medium-low (5) and medium populated (2) towns in peripheral and peri-urban areas. Municipalities in cluster 5 are predominantly located in mountainous regions, while those in cluster 2 are mainly found in hill and lowland areas. Finally, cluster 7 includes medium to high populated cities, encompassing both hubs and peri-urban municipalities, primarily located in lowland regions.

The attribute importance analysis (Fig. 7a) indicates that urban degree is the most influential factor in accurately distinguishing clusters, followed by altimetric zone, centeredness degree, and population classification.



(a) (b)

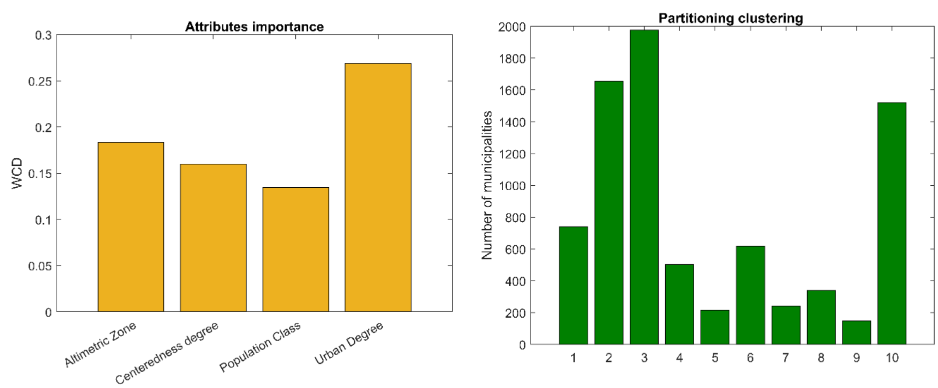
https://doi.org/10.5194/egusphere-2025-908
Preprint. Discussion started: 28 February 2025



(c) (d)

Figure 6 - Representativeness of clusters in terms of urban degree (a), urban centeredness degree (b), population class (c) and altimetric zone (d).

The distribution of municipalities across clusters varies significantly (Fig. 7b). Cluster 3 is the largest, comprising 1,976 municipalities, followed by Cluster 2 with 1,655 municipalities and Cluster 10 with 1,520 municipalities. Cluster 1 includes 740 municipalities, while Cluster 6 and Cluster 8 contain 618 and 341 municipalities, respectively. Cluster 4 and Cluster 5 represent smaller groups, with 503 and 216 municipalities. The smallest clusters are Cluster 9 with 148 municipalities and Cluster 7 with 243 municipalities, indicating distinct groupings within the dataset.



(a) (b)

Figure 7 - Attribute importance in terms of variation of WCD (a); number of Italian municipalities belonging to each cluster (b).

4.4.3 Comparison of clustering algorithms

In order to evaluate quality of clustering techniques, measure of intra-cluster distance (i.e., WCD presented in section 4.2.1) as well as inter-clusters distance are investigated. Specifically, the coefficient WCD is adopted to evaluate the performance of the clustering for each single attribute. The value of WCD_l for the attribute *l-th* across all clusters is calculated as follows:

$$WCD_l = \frac{\sum_{k=1}^{m} WCD_l(C_k)}{m} \tag{10}$$

Where $WCD_l(C_k)$ is the value of WCD for the *l-th* attribute and the *k-th* cluster, calculated according to Eq. (4) and Eq. (5), and m is the total number of clusters. The lower is the WCD_l value, the better the performance of the algorithm with respect to the considered attribute.

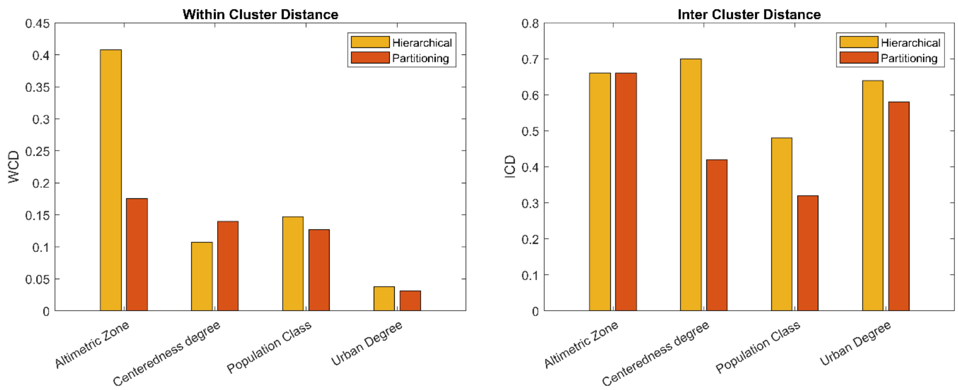
Inter-cluster distance (ICD) measures the separation between clusters in a clustering solution and is useful for evaluating how distinct the clusters are. To calculate inter-cluster distance, we adopt Centroid-to-Centroid Distance:

https://doi.org/10.5194/egusphere-2025-908
Preprint. Discussion started: 28 February 2025

$$ICD_l = \frac{\sum_{i \neq j} \left\| \mu_{i,l} - \mu_{j,l} \right\|}{m} \tag{11}$$

Where $\mu_{i,l}$ and $\mu_{j,l}$ are the centroids values (i.e., the mode of the objects within a cluster) of clusters *i* and *j* for the *l-th* attribute. Unlike WCD, a higher ICD indicates better algorithm performance, as it reflects greater differentiation between clusters.

From Fig. 8, it can be observed that the hierarchical clustering algorithm achieves better WCD performance for the centeredness degree attribute. However, it performs worse than partitioning clustering for the population class attribute and significantly worse for the altimetric zone attribute. Regarding urban degree, both clustering methods exhibit high and comparable WCD performance. Overall, considering the average WCD across all attributes, the hierarchical algorithm shows a higher WCD (0.18) compared to the partitioning algorithm (0.13). Despite this difference, both clustering approaches demonstrate relatively good performance. In terms of ICD, hierarchical clustering outperforms partitioning clustering across all attributes, except for the altimetric zone, where both methods yield the same ICD value. Overall, hierarchical clustering demonstrates superior performance in terms of ICD, with a value of 0.62, compared to 0.49 for partitioning clustering.



(a) (b)

Figure 8 – Comparison of clustering algorithms in terms of WCD_l (a) and ICD_l (b).

While WCD is slightly higher in hierarchical clustering compared to partitioning clustering, the overall clustering quality remains good in both methods. However, the significantly higher ICD (0.62 vs. 0.49) for hierarchical clustering suggests that it produces more distinct and well-separated clusters, making it the preferable choice for this analysis. Therefore, the results of hierarchical clustering are used for the initial broad definition of archetypes (presented in Section 5) and serve as the input for the second step of the analysis, namely the nested clustering, which is discussed in the following section.

4.5 Nested clustering

Nested clustering identifies clusters within clusters, unveiling data structures at multiple levels of granularity. This method is particularly valuable for detecting complex patterns in data, providing a more detailed understanding of the underlying relationships. In this study, for each first-level cluster (or broad archetype) identified in the previous section, we analyse nested clusters to capture the heterogeneity of urban settlements in terms of socio-economic vulnerability. To achieve this, we consider eight socio-economic attributes—aging index, low educational index, proportion of unemployed individuals, proportion of commuters, proportion of female employees, proportion of buildings in poor condition, crowding index, and proportion of foreign residents—selected based on the correlation analysis results presented in Section 4.1.

Both partitioning and hierarchical clustering algorithms are applied within each cluster to further refine the sub-groups. The optimal number of sub-clusters for each cluster is determined using WCD and the inconsistency coefficient, as detailed in Section 4.2.1. Based on the clustering performance metrics (i.e., WCD and ICD), partitioning clustering proves to be the most suitable approach for this second-level clustering. As results, we identified 3 sub-clusters for cluster 2 and cluster 3, 2 sub-clusters for cluster 4, 8, 9 and 10 and no sub-clusters for cluster 1, 5, 6 and 7 due to homogeneity of socio-economic data within the cluster. The list of clusters and sub-clusters with the average values of numerical attributes for each of the sub-clusters identified is reported in Table 4.

https://doi.org/10.5194/egusphere-2025-908
Preprint. Discussion started: 28 February 2025

Cluster names are derived from the geographical and demographic characteristics analysed (e.g., peri-urban settlements, peripheral rural areas), while sub-cluster names are assigned based on the mean value of the Social Vulnerability Index (SoVI) within each sub-cluster, which reflects demographic and socio-economic conditions of the settlements. Additionally, if necessary, sub-cluster names may also incorporate the specific social vulnerability factors that contribute most significantly to the SoVI value. For example, both sub-clusters 2a and 2b exhibit high social vulnerability; however, sub-cluster 2b has the highest aging index among all sub-clusters, leading to its classification as "*aged communities with high social vulnerability*." Similarly, sub-clusters 3a ("*aged communities with high social vulnerability*") and 3b ("*high household density settlements with high social vulnerability*") both exhibit high SoVI values, but the former is characterized by a high aging index, while the latter has a high crowding index. The SoVI indicator is calculated following the procedure proposed by Frigerio et al.(2018), utilizing the same socio-economic variables adopted in this study for clustering. The criteria used to identify the different socio-economic condition categories is based on SoVI values and specifically: a value lower than 1 corresponds to low social vulnerability, value between 1 and 1.20 to moderate social vulnerability, values between 1.20 and 1.40 to intermediate social vulnerability, values between 1.40 and 1.60 to high social vulnerability, values higher than 1.60 to very high social vulnerability. The average values of individual variables for each sub-cluster are provided in Table 4.

Table 4 – Cluster and sub-clusters identified with relative average value of numerical attributes within the cluster or sub-cluster.

Cluster	Sub-cluster	Aging Index	Low Educational Index	Unemployed	Comm uters	Female employed	Building bad state	Crowding index	Foreign	SoVI
1. Low populated intermunicipal hubs in rural areas	-	1.69	0.54	0.03	0.23	0.44	0.14	2.44	0.09	1.31
2. Low populated peripherical rural areas	a. High socially vulnerability settlements	2.04	0.61	0.04	0.22	0.39	0.19	2.40	0.04	1.50
	b. Aged communities with high social vulnerability	3.70	0.68	0.02	0.28	0.38	0.23	2.03	0.05	1.54
	c. Aged communities with moderate social vulnerability	3.31	0.59	0.03	0.19	0.43	0.19	2.12	0.04	1.15
3. Medium-Low populated peri-urban settlements in rural areas	a. Aged communities with high social vulnerability	2.38	0.61	0.03	0.34	0.44	0.22	2.21	0.05	1.46
	b. High household density settlements with high social vulnerability	1.51	0.56	0.05	0.26	0.39	0.21	2.59	0.03	1.56
	c. High socially vulnerable settlements	1.78	0.62	0.03	0.34	0.42	0.12	2.47	0.08	1.54
4. Medium-populated intermunicipal hubs	a. Settlements with intermediate social vulnerability	1.77	0.54	0.03	0.28	0.45	0.10	2.40	0.08	1.22
	b. High socially vulnerable settlements	1.22	0.55	0.05	0.15	0.40	0.18	2.74	0.03	1.40
5. Low Populated peri-urban suburbs	-	1.38	0.58	0.04	0.34	0.41	0.15	2.55	0.05	1.53
6. Medium-populated peri-urban and peripheric suburbs	-	1.71	0.54	0.03	0.14	0.41	0.19	2.50	0.04	1.17
7. Peripheric suburbs medium-low populated	-	1.43	0.59	0.04	0.24	0.40	0.19	2.50	0.06	1.51
8. Peri-urban cities	a. Very high socially vulnerable settlements	0.73	0.63	0.06	0.22	0.36	0.26	3.08	0.02	2.01
	b. Settlements with intermediate social vulnerability	1.17	0.55	0.03	0.40	0.45	0.11	2.44	0.07	1.38
9. Major urban hubs	a. Low socially vulnerable settlements	1.95	0.47	0.03	0.11	0.47	0.10	2.23	0.11	0.84
	b. Settlements with intermediate social vulnerability	1.50	0.51	0.05	0.08	0.42	0.24	2.60	0.03	1.20
10. Medium-populated towns	a. High socially vulnerable settlements	1.15	0.57	0.05	0.18	0.38	0.22	2.71	0.03	1.56
	b. Settlements with intermediate social vulnerability	1.54	0.57	0.03	0.28	0.43	0.13	2.48	0.08	1.39

https://doi.org/10.5194/egusphere-2025-908
Preprint. Discussion started: 28 February 2025

5 Urban settlements archetypes in Italy

The first-level clustering provides a “broad” definition of urban archetypes, considering only geographic and demographic attributes. Clusters 1, 2 and 3 (“*Low populated intermunicipal hubs in rural areas*”, “*Low populated peripherical rural areas*” and “*Low populated peri-urban settlements in rural areas*”, see table 4) represent archetypes of low populated, rural urban settlements in peripherical (2) and peri-urban (3) areas or close to urban hubs (1). Cluster 5 (“*Low Populated peri-urban suburbs*”) also represents archetypes for low populated settlements but characterized by higher population density. Clusters 4, 6, 7 and 10 (“*Medium-populated intermunicipal hubs*”, “*Medium-populated peri-urban and peripheric suburbs*”, “*Peripheric suburbs medium-low populated*” and “*Medium-populated towns*”) are archetypes for medium-populated peri-urban and peripheric suburbs (6, 7, 10) and towns that are intermunicipal hubs (4). Clusters 8 and 9 (“*Peri-urban cities*”, “*Major urban hubs*”) represent archetypes for densely populated peri-urban settlements (8) and hubs (9). These broad archetypes are mapped in Fig. 9. Notably, altimetric classification is not included in the archetype definition, as it exhibits high within-cluster variance and is the least significant attribute, making its contribution negligible in defining urban archetypes.

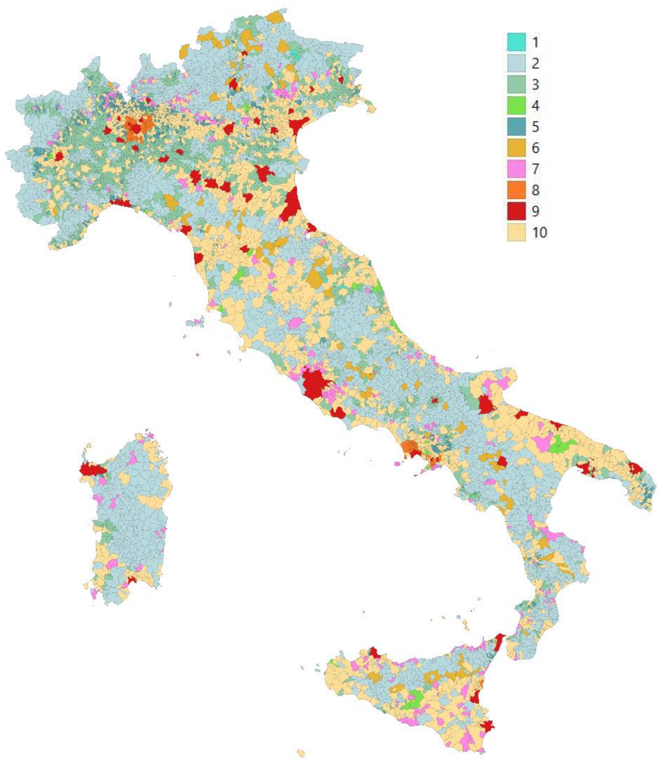


Figure 9 – Broad urban archetypes across Italian territory.

The proposed urban archetypes are obtained considering both first-level clusters and sub-clusters. Specifically, 18 urban archetypes are defined (listed in Table 5), characterized by geographic, demographic, and socio-economic features. Each archetype is identified by an alphanumeric code and a designation, derived from the combination of the codes and names of the clusters and sub-clusters from which they are obtained.

Table 5 – Urban archetypes identified. The number of municipalities belonging to each archetype and the related population share are provided in the last columns.

Urban Archetypes	n° municipalities	Population share (%)

https://doi.org/10.5194/egusphere-2025-908
Preprint. Discussion started: 28 February 2025

1	Low populated intermunicipal hubs in rural areas with intermediate social vulnerability	5	0.03
2a	Low populated peripherical rural settlements with high social vulnerability	2237	7.56
2b	Low populated peripherical rural settlements with socially vulnerable aged communities	525	0.54
2c	Low populated peripheral rural settlements with moderately socially vulnerable aged communities	543	1.16
3a	Medium-Low populated peri-urban settlements in rural areas with socially vulnerable aged communities	457	1.1
3b	Medium-Low populated, high household density peri-urban settlements in rural areas with high social vulnerability	220	1.47
3c	Medium-Low populated peri-urban settlements in rural areas with high social vulnerability	955	4.13
4a	Medium-populated intermunicipal hubs with intermediate social vulnerability	20	0.64
4b	Medium-populated intermunicipal hubs with high social vulnerability	15	0.64
5	Low Populated peri-urban suburbs with high social vulnerability	615	2.92
6	Medium-populated peri-urban and peripheric suburbs with moderate social vulnerability	89	1.17
7	Peripheric suburbs medium-low populated with high social vulnerability	347	5.14
8a	Peri-urban cities with very high social vulnerability	63	2.35
8b	Peri-urban cities with intermediate social vulnerability	103	2.22
9a	Major urban hubs with low social vulnerability	31	15.89
9b	Major urban hubs with intermediate social vulnerability	20	6.61
10a	Medium-populated towns with high social vulnerability	429	13.2
10b	Medium-populated towns with intermediate social vulnerability	1286	33.23

Urban *archetype 1* represents urban settlements characterized by low population, low population density and medium-low social vulnerability, functioning as intermunicipal hubs. Notably, only few municipalities fall into this archetype, because it represents a small minority of urban hubs in rural areas with very low population (< 500 inhabitants). The majority of urban archetypes in peripheral and rural areas (i.e., those with low population density) are characterized by low population and medium to high social vulnerability (e.g., *archetypes 2a*, *2b*, *3a*, *3b* and *3c*). *Archetype 2a* includes peripherical, low populated rural areas (< 2000 inhabitants) with high social vulnerability, mainly due to a medium-high aging index, high low educational index, high unemployment rate and high crowding index. *Archetype 2b* represents sparsely populated rural, aging communities, exhibiting the highest average aging index and low educational index, along with significantly high proportion of buildings in poor conditions and low percentage of female employed, all contributing to a high SoVI value. In contrast, *archetype 2c*, which also represents low populated peripheral settlements characterized by very high aging index, shows lower SoVI value thanks to the lower percentage of commuters, a lower low educational index and higher proportion of female employed.

Archetypes 3a, *3b* and *3c* represent low populated, socially vulnerable settlements in peri-urban areas, exhibiting high low educational index, high percentage of unemployed (*3b*), high percentage of buildings in bad state of preservation (3a and 3b), high crowding index (*3b* and *3c*) and high proportion of foreign resident (3c). *Archetypes 4a* and *4b* are medium-populated intermunicipal hubs with slightly different level of social vulnerability: medium-low social vulnerability the former, medium-high social vulnerability the latter. While both show quite high crowding index, *archetype 4b* shows higher crowding index, high unemployment rate and medium-high percentage of buildings in poor conditions.

Low populated peri-urban suburbs are represented by *archetype 5* and show high social vulnerability, primarily due to medium-high unemployment rate, high percentage of commuters and high crowding index. *Archetype 6* corresponds to medium-populated peri-urban and peripheric suburbs with relatively favourable demographic and socio-economic conditions, benefiting from medium-high educational level (i.e., medium-low value of low educational index), low percentage of commuters, high percentage of female employed, that compensate for the medium-high value of buildings in bad state of preservation and high crowding index. *Archetype 7*, representing medium-low populated peripheral suburbs, shares the same average values of *archetype 6* regarding buildings in poor conditions and crowding index. However, it is characterized by higher percentage of commuters and higher percentage of unemployed, leading to higher SoVI.

Archetypes 8a represents densely populated peri-urban settlements with the highest social vulnerability, driven by the highest percentage of unemployed, the highest percentage of buildings in poor conditions and the highest crowding index. In addition, this archetype also shows high low educational index and low percentage of female employed. *Archetype 8b* also represents densely populated peri-urban settlements but with lower social vulnerability (higher percentages of female

https://doi.org/10.5194/egusphere-2025-908
Preprint. Discussion started: 28 February 2025

employed, lower percentage of buildings in bad state of preservation), despite being characterized by the highest percentage of commuters.

Archetypes 9a and *9b* correspond to the largest densely populated cities, yet with contrasting degree of social vulnerability: *archetype 9a* has low SoVI (the lowest), attributed to the lowest value of low educational index, the highest percentage of female employed, together with low percentage of buildings in poor conditions; *archetype 9b* has intermediate social vulnerability, due to the higher percentage of unemployed, higher percentage of buildings in poor conditions and higher crowding index. *Archetypes 10a* and *10b* represent medium-populated towns (with medium-population density) that can function as hubs or be located in peri-urban and peripheral areas. These towns exhibit relatively high social vulnerability, influenced by high percentage of unemployed (*10a*), medium-high percentage of commuters (*10b*), relatively low percentage of female employed (*10a*), high percentage of buildings in bad state of preservation (*10b*) and high crowding index.

From a geographical perspective, we found that many low populated peripheric and peri-urban settlements in rural areas (*archetypes 2b*, *2c*, *3a*, *3b* and *3c*) are primarily located in northern region of Piemonte (20%) and Lombardy (15%). *Archetypes 2a*, which also represents low populated peripherical rural areas, includes several municipalities in Sardinia (12%), Calabria (9%), Lombardy (9%) and Piemonte (7%) regions. *Archetype 4a* ("*Medium-populated intermunicipal hubs with intermediate social vulnerability*") is exclusively found in the northern regions (i.e., Piemonte, Lombardy, Veneto, Liguria, Tuscany and Marche). Conversely, *Archetype 4b* ("*Medium-populated intermunicipal hubs with high social vulnerability*") is present only in the southern regions, specifically Abruzzo, Campania, Apulia and Sicily. Many municipalities (48%) represented by *archetype 5* ("*Low Populated peri-urban suburbs with high social vulnerability*") are in Lombardy region, while a notable percentage of municipalities represented by *archetype 6* ("*Medium-populated peri-urban and peripheric suburbs with moderate social vulnerability*") are in Trentino (20%), Tuscany (11%) and Basilicata (11%) regions. *Archetype 7* ("*Peripheric suburbs medium-low populated with high social vulnerability*") includes a high concentration of municipalities in Lombardy (27%) and Sicily (15%) regions.

"*Peri-urban cities with very high social vulnerability*" (*archetype 8a*) can be found exclusively in the southern regions, which almost all cases located in Campania (61 out of 63 municipalities). In contrast, "*Peri-urban cities with intermediate social vulnerability*" (*archetype 8b*) are concentrated in the northern regions, specifically in Lombardy (99%) and Piemonte (1%). Similarly, "*Major urban hubs with low social vulnerability*" (*archetype 9a*) are in the northern part of the country, covering Piemonte, Valle d'Aosta, Lombardy, Trentino, Veneto, Friuli-Venezia Giulia, Liguria, Emilia-Romagna, Tuscany and Lazio regions, while "*Major urban hubs with intermediate social vulnerability*" (*archetype 9b*) predominantly located in the southern regions, including Molise, Campania, Apulia, Sicily and Sardinia regions. *Archetype 10a* ("*Medium-populated towns with high social vulnerability*") is also primarily found in the southern regions (93%), while *archetype 10b* ("*Medium-populated towns with intermediate social vulnerability*") is distributed across the entire country, with a significant percentage in Lombardy (27%) and Veneto (16%) regions.

6 Discussion

The proposed study of urban archetypes leverages the framework and guidelines set forth by Piemontese et al. (2022) to ensure a robust and reliable archetype analysis, focusing on six dimensions of validity: conceptual validity, construct validity, internal validity, empirical validity, external validity, and application validity. The proposed urban archetypes conform to each of these dimensions as follows.

Conceptual Validity is achieved by ensuring the research problem and questions are scientifically sound and relevant to real-world issues. In this study we addressed the need to categorize urban areas based on geographic, demographic, and socio-economic factors to understand urban vulnerabilities better. By focusing on these pertinent aspects, this study aligns with the conceptual framework and reflects real-world challenges faced by urban settlements in Italy.

Construct Validity involves the careful selection of attributes that define the archetypes, ensuring their connection to the conceptual framework. We meticulously selected attributes such as unemployment rates, building conditions, crowding indices, educational levels, and percentages of female employment. These attributes are justified based on existing literature and their relevance to socio-economic conditions, thereby reinforcing the construct validity.

Internal Validity is maintained through the rigorous application of hierarchical and partitioning clustering methods. We employed WCD and ICD to select the most suitable clustering approach, ensuring the reliability and robustness of the clustering process. This methodological rigor ensures that the urban archetypes are well-defined and accurately represent the dataset.

https://doi.org/10.5194/egusphere-2025-908
Preprint. Discussion started: 28 February 2025

Empirical Validity is partially confirmed by demonstrating that the identified archetypes correspond to real-world outcomes and causal mechanisms. This is supported by stakeholder engagement-based risk storylines, as outlined in Marciano et al. (2024). Marciano et al. (2024) present an exploratory case study using a participatory approach to develop multi-risk storylines, illustrating the cascading effects of a heatwave followed by intense rainfall in two Italian urban contexts: a peri-urban area and a metropolitan area. Findings reveal that peri-urban settlements face limited emergency resources and higher infrastructure failure risks, while metropolitan hubs have stronger emergency systems but face coordination challenges in managing large-scale events. The study highlights the varying levels of urban vulnerability across different archetypes. Further studies should explore the impacts of natural disasters on different archetypes, revealing key differences in vulnerability and response capabilities across the considered urban contexts.

External Validity assesses the generalizability of archetypes beyond the studied cases. While this study acknowledges the challenge of satisfying this dimension, we believe that further research is needed to validate the archetypes across different settings. The identification of archetypes in various Italian regions provides a preliminary basis for generalization, but additional studies are required to establish broader applicability, particularly in assessing the sensitivity of the proposed methodology with respect to the specific characteristics of authoritative data in comparable European countries.

Application Validity evaluates the practical usefulness of the archetypes. This study demonstrates the potential of urban archetypes in developing risk storylines, enhancing risk communication, and supporting stakeholder engagement. By categorizing urban areas into distinct archetypes, it becomes possible to assess how different hazard scenarios may unfold in each context, considering their specific vulnerabilities, exposure levels, and adaptive capacities. This structured approach enables policymakers to design tailored interventions and resilience strategies based on specific vulnerability profiles. However, to further strengthen urban resilience planning and develop targeted mitigation measures, it is crucial to consider not only exposure and vulnerability but also hazard data for each archetype—particularly the level of exposure of urban settlements to various natural hazards. Although in this study we did not yet integrate hazard information, there is a clear need for future research to incorporate this aspect and conduct GIS-based analyses for a more comprehensive assessment of urban risk (e.g., Tocchi et al., 2024).

7 Conclusion

This study presents urban settlements archetypes in Italy, based on geographic, demographic and socio-economic factors that cover different vulnerability dimensions. Cluster analysis revealed ten broad archetypes of urban settlements in Italy according to geographic and demographic urban features, further refined into 18 nested archetypes in regions with diverse socio-economic characteristics.

The proposed urban archetypes successfully address the six dimensions of validity outlined by Piemontese et al. (2022). Conceptual, construct, and internal validity are robustly established through scientifically sound research questions, careful attribute selection, and rigorous clustering methods. Empirical validity is supported by real-world outcomes and stakeholder engagement. While external and application validity require further investigation, this study lays a strong foundation for future research and practical applications in urban resilience planning.

Defining urban archetypes based on vulnerability factors may help identify areas with higher susceptibility to natural hazards and socio-economic challenges, supporting better resource allocation for disaster preparedness and response. Although this study does not incorporate hazard information, which requires further research and analysis, its findings offer a valuable tool for policymakers to design targeted interventions based on specific vulnerability profiles, ultimately supporting the development of urban resilience strategies tailored to different archetypes.

Data Availability

The complete dataset describing the urban archetypes associate to each municipality is available in:
Tocchi, G., Pittore, M., & Polese, M. (2025). Italian urban archetypes (Tocchi et al., 2025) (1.0) [Data set]. Zenodo. https://doi.org/10.5281/zenodo.14888733.

Acknowledgements

This study was carried out within the RETURN Extended Partnership and received funding from the European Union Next-GenerationEU (National Recovery and Resilience Plan – NRRP, Mission 4, Component 2, Investment 1.3 – D.D. 1243 2/8/2022, PE0000005)

Author contribution

All authors contributed to the conceptualization and design of the methodology. G. Tocchi carried out the formal analysis, led the research and investigation process, managed data visualization and presentation, and was responsible for writing the original draft along with subsequent editing. M. Pittore and M. Polese also contributed to the writing, review, and editing of the manuscript.

References

Alexander, D. (2012). Models of Social Vulnerability to Disasters*. *RCCS Annual Review*, *4*. https://doi.org/10.4000/rccsar.412

Alonso, W. (1964). *Location and Land Use*. Harvard University Press. https://doi.org/10.4159/harvard.9780674730854

Angeon, V., & Bates, S. (2014). Reviewing Composite Vulnerability and Resilience Indexes: A Sustainable Approach and Application. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2437980

ARDALAN, A., MAZAHERI, M., NAIENI, K. H., REZAIE, M., TEIMOORI, F., & POURMALEK, F. (2010). Older people's needs following major disasters: a qualitative study of Iranian elders' experiences of the Bam earthquake. *Ageing and Society*, *30*(1), 11–23. https://doi.org/10.1017/S0144686X09990122

Awah, L. S., Nyam, Y. S., Belle, J. A., & Orimoloye, I. R. (2024). A system archetype approach to identify behavioural patterns in flood risk management: Case study of Cameroon. *Environmental Development*, *51*, 101026. https://doi.org/10.1016/j.envdev.2024.101026

Balk, D., Leyk, S., Jones, B., Montgomery, M. R., & Clark, A. (2018). Understanding urbanization: A study of census and satellite-derived urban classes in the United States, 1990-2010. *PLOS ONE*, *13*(12), e0208487. https://doi.org/10.1371/journal.pone.0208487

Bertram, D., Chilla, T., & Hippe, S. (2023). Cross-border mobility: Rail or road? Space-time-lines as an evidence base for policy debates. *Journal of Borderlands Studies*, 1–18. https://doi.org/10.1080/08865655.2023.2249917

Bishara, A. J., & Hittner, J. B. (2017). Confidence intervals for correlations when data are not normal. *Behavior Research Methods*, *49*(1), 294–309. https://doi.org/10.3758/s13428-016-0702-8

Bruce, G. D. , & W. R. E. (1971). . Developing empirically derived city typologies: An application of cluster analysis. . *The Sociological Quarterly*, *12*, 238–246.

Brunkard, J., Namulanda, G., & Ratard, R. (2008). Hurricane Katrina Deaths, Louisiana, 2005. *Disaster Medicine and Public Health Preparedness*, *2*(4), 215–223. https://doi.org/10.1097/DMP.0b013e31818aaf55

Cardona, O.-D., van Aalst, M. K., Birkmann, J., Fordham, M., McGregor, G., Perez, R., Pulwarty, R. S., Schipper, E. L. F., Sinh, B. T., Décamps, H., Keim, M., Davis, I., Ebi, K. L., Lavell, A., Mechler, R., Murray, V., Pelling, M., Pohl, J., Smith, A.-O., & Thomalla, F. (2012). Determinants of Risk: Exposure and Vulnerability. In *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation* (pp. 65–108). Cambridge University Press. https://doi.org/10.1017/CBO9781139177245.005

Carnelli, F., & Frigerio, I. (2017). A socio-spatial vulnerability assessment for disaster management: insights from the 2012 emilia earthquake (italy). *SOCIOLOGIA URBANA E RURALE*, *111*, 22–44. https://doi.org/10.3280/SUR2016-111002

https://doi.org/10.5194/egusphere-2025-908
Preprint. Discussion started: 28 February 2025

Carroll, M., & Paveglio, T. (2016). Using community archetypes to better understand differential community adaptation to wildfire risk. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*(1696), 20150344. https://doi.org/10.1098/rstb.2015.0344

Centre for Research on the Epidemiology of Disasters (CRED). (2024). *EM-DAT -The international disaster database*. https://www.emdat.be/

Chen, W., Cutter, S. L., Emrich, C. T., & Shi, P. (2013). Measuring social vulnerability to natural hazards in the Yangtze River Delta region, China. *International Journal of Disaster Risk Science*, *4*(4), 169–181. https://doi.org/10.1007/s13753-013-0018-6

Cohen. (2013). *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*. Routledge. https://doi.org/10.4324/9780203774441

Cohen, A. J., Brauer, M., Burnett, R., Anderson, H. R., Frostad, J., Estep, K., Balakrishnan, K., Brunekreef, B., Dandona, L., Dandona, R., Feigin, V., Freedman, G., Hubbell, B., Jobling, A., Kan, H., Knibbs, L., Liu, Y., Martin, R., Morawska, L., … Forouzanfar, M. H. (2017). Estimates and 25-year trends of the global burden of disease attributable to ambient air pollution: an analysis of data from the Global Burden of Diseases Study 2015. *The Lancet*, *389*(10082), 1907–1918. https://doi.org/10.1016/S0140-6736(17)30505-6

Cutini, V. (2001). Centrality and Land Use: Three Case Studies on the Configurational Hypothesis. *Cybergeo*. https://doi.org/10.4000/cybergeo.3936

Cutler, A., & Breiman, L. (1994). Archetypal Analysis. *Technometrics*, *36*(4), 338. https://doi.org/10.2307/1269949

Cutter, S. L., Boruff, B. J., & Shirley, W. L. (2003). Social Vulnerability to Environmental Hazards *. *Social Science Quarterly*, *84*(2), 242–261. https://doi.org/10.1111/1540-6237.8402002

Cutter, S. L., & Finch, C. (2008). Temporal and spatial changes in social vulnerability to natural hazards. *Proceedings of the National Academy of Sciences*, *105*(7), 2301–2306. https://doi.org/10.1073/pnas.0710375105

Dalton, P. ; W. S. (2015). *Grouping Minnesota Cities: Using Cluster Analysis*.

Dickson, E., Baker, J. L., Hoornweg, D., & Asmita, T. (2012). *Urban Risk Assessments: An Approach for Understanding Disaster and Climate Risk in Cities* (The World Bank).

Dodman, D., Sverdlik, A., Agarwal, S., Kadungure, A., Kothiwal, K., Machemedze, R., & Verma, S. (2023). Climate change and informal workers: Towards an agenda for research and practice. *Urban Climate*, *48*, 101401. https://doi.org/10.1016/j.uclim.2022.101401

Douglas, J. (2007). Physical vulnerability modelling in natural hazard risk assessment. *Natural Hazards and Earth System Sciences*, *7*(2), 283–288. https://doi.org/10.5194/nhess-7-283-2007

DPS. (2013). *Le aree interne: di quali territori parliamo? Nota esplicativa sul metodo di classificazione delle aree (in Italian)*. https://www.agenziacoesione. gov.it/wp-content/uploads/2021/01/Nota_metodologica_Aree_interne-2-1.pdf.

Eriksen, S. H., & Kelly, P. M. (2007). Developing Credible Vulnerability Indicators for Climate Adaptation Policy Assessment. *Mitigation and Adaptation Strategies for Global Change*, *12*(4), 495–524. https://doi.org/10.1007/s11027-006-3460-6

European Environment Agency. (2024). *European Climate Risk Assessment* .

Eurostat. (2021). *Applying the degree of urbanisation : a methodological manual to define cities, towns and rural areas for international comparisons : 2021 edition*. Publications Office of the European Union.

Fan, C., Jiang, X., Lee, R., & Mostafavi, A. (2022). Equality of access and resilience in urban population-facility networks. *Npj Urban Sustainability*, *2*(1), 9. https://doi.org/10.1038/s42949-022-00051-3

FEMA. (2022). *Hazus 5.1, Hazus Flood Technical Manual*.

Flanagan, B. E., Gregory, E. W., Hallisey, E. J., Heitgerd, J. L., & Lewis, B. (2011). A Social Vulnerability Index for Disaster Management. *Journal of Homeland Security and Emergency Management*, *8*(1). https://doi.org/10.2202/1547-7355.1792

Fraiman, R., Justel, A., & Svarc, M. (2008). Selection of Variables for Cluster Analysis and Classification Rules. *Journal of the American Statistical Association*, *103*(483), 1294–1303. https://doi.org/10.1198/016214508000000544

Frigerio, I., Carnelli, F., Cabinio, M., & De Amicis, M. (2018). Spatiotemporal Pattern of Social Vulnerability in Italy. *International Journal of Disaster Risk Science*, *9*(2), 249–262. https://doi.org/10.1007/s13753-018-0168-7

Giuliano, G., & Narayan, D. (2003). Another Look at Travel Patterns and Urban Form: The US and Great Britain. *Urban Studies*, *40*(11), 2295–2312. https://doi.org/10.1080/0042098032000123303

Gordon, T. (1986). "Is the standard deviation tied to the mean?" *Teaching Statistics, 8*(2), 40–42.

Gower, J. C. (1971). A General Coefficient of Similarity and Some of Its Properties. *Biometrics*, *27*(4), 857. https://doi.org/10.2307/2528823

Guillard-Gonçalves, C., Cutter, S. L., Emrich, C. T., & Zêzere, J. L. (2015). Application of Social Vulnerability Index (SoVI) and delineation of natural risk zones in Greater Lisbon, Portugal. *Journal of Risk Research*, *18*(5), 651–674. https://doi.org/10.1080/13669877.2014.910689

Hamideh, S., Sen, P., & Fischer, E. (2022). Wildfire impacts on education and healthcare: Paradise, California, after the Camp Fire. *Natural Hazards*, *111*(1), 353–387. https://doi.org/10.1007/s11069-021-05057-1

Han, J., Kamber, M., & Pei, J. (2011). *Data Mining. Concepts and Techniques, 3rd Edition (The Morgan Kaufmann Series in Data Management Systems)*.

Harris, C. (1943). A Functional Classification of Cities in the United States. *Geographical Review*, *33*, 86–99.

Hoffman, R. N., Boukabara, S.-A., Kumar, V. K., Garrett, K., Casey, S. P. F., & Atlas, R. (2017). An Empirical Cumulative Density Function Approach to Defining Summary NWP Forecast Assessment Metrics. *Monthly Weather Review*, *145*(4), 1427–1435. https://doi.org/10.1175/MWR-D-16-0271.1

Huang, Z. (1998). Extensions to the k-Means Algorithm for Clustering Large Data Sets with Categorical Values. *Data Mining and Knowledge Discovery*, *2*(3), 283–304. https://doi.org/10.1023/A:1009769707641

Intergovernmental Panel on Climate Change (IPCC). (2022). *Climate Change 2022: Impacts, Adaptation and Vulnerability*.

Isard W. (1956). *Location and Space Economy*. MIT Press.

ISTAT. (2020). *Annuario statistico Italiano 2020* . https://www.istat.it/it/files//2020/12/C01.pdf

ISTAT, I. N. I. of S. (2011). *15° Censimento della popolazione.*

ISTAT, I. N. I. of Statistics. (2024). *Condizioni di vita e reddito delle famiglie - anno 2023 (in Italian)* . https://www.istat.it/wp-content/uploads/2024/05/REPORT-REDDITO-CONDIZIONI-DI-VITA_2023.pdf

James, Gareth., Witten, Daniela., Hastie, Trevor., & Tibshirani, Robert. (2017). *An introduction to statistical learning : with applications in R*. Springer : Springer Science+Business Media.

Jatain, A., Nagpal, A., & Gaur, D. (2013). Agglomerative Hierarchical Approach for Clustering Components of Similar Reusability. In *International Journal of Computer Applications* (Vol. 68, Issue 2).

Joshi, M. Y., Rodler, A., Musy, M., Guernouti, S., Cools, M., & Teller, J. (2022). Identifying urban morphological archetypes for microclimate studies using a clustering approach. *Building and Environment*, *224*, 109574. https://doi.org/10.1016/j.buildenv.2022.109574

Kappes, M. S., Papathoma-Köhle, M., & Keiler, M. (2012). Assessing physical vulnerability for multi-hazards using an indicator-based methodology. *Applied Geography*, *32*(2), 577–590. https://doi.org/10.1016/j.apgeog.2011.07.002

Kendra, J., Rozdilsky, J., & McEntire, D. A. (2008). Evacuating Large Urban Areas: Challenges for Emergency Management Policies and Concepts. *Journal of Homeland Security and Emergency Management*, *5*(1). https://doi.org/10.2202/1547-7355.1365

https://doi.org/10.5194/egusphere-2025-908
Preprint. Discussion started: 28 February 2025

Koren, D., Kilar, V., & Rus, K. (2017). Proposal for Holistic Assessment of Urban System Resilience to Natural Disasters. *IOP Conference Series: Materials Science and Engineering*, *245*, 062011. https://doi.org/10.1088/1757-899X/245/6/062011

La Sorte, F. A., Johnston, A., & Ault, T. R. (2021). Global trends in the frequency and duration of temperature extremes. *Climatic Change*, *166*(1–2), 1. https://doi.org/10.1007/s10584-021-03094-0

Lagomarsino, S., & Giovinazzi, S. (2006). Macroseismic and mechanical models for the vulnerability and damage assessment of current buildings. *Bulletin of Earthquake Engineering*, *4*(4), 415–443. https://doi.org/10.1007/s10518-006-9024-z

Lall, S. V., & Deichmann, U. (2012). Density and Disasters: Economics of Urban Hazard Risk. *The World Bank Research Observer*, *27*(1), 74–105. https://doi.org/10.1093/wbro/lkr006

Lavell, A., Rica, C., Oppenheimer, M., Diop, C., Moser, S., & Takeuchi, K. (2012). Climate Change: New Dimensions in Disaster Risk, Exposure, Vulnerability, and Resilience. In *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation*. A Special Report of Working Groups I and II of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, UK, and New York, NY, USA.

Loreti, S., Ser-Giacomi, E., Zischg, A., Keiler, M., & Barthelemy, M. (2022). Local impacts on road networks and access to critical locations during extreme floods. *Scientific Reports*, *12*(1), 1552. https://doi.org/10.1038/s41598-022-04927-3

MacQueen, J. B. (1967). *Some methods for classification and analysis of multivariate observations: Vol. Vol. 1*. Proc. 5-th Symp. Mathematical Statistics and Probability.

Marciano, C., Peresan, A., Pirni, A., Pittore, M., Tocchi, G., & Zaccaria, A. M. (2024). A participatory foresight approach in disaster risk management: The multi-risk storylines. *International Journal of Disaster Risk Reduction*, *114*, 104972. https://doi.org/10.1016/j.ijdrr.2024.104972

Marin Ferrer, M. , V. L. and P. K. (2017). *INFORM Index for Risk Management: Concept and Methodology, Version 2017. EUR 28655 EN. Luxembourg (Luxembourg). Publications Office of the European Union; 2017. JRC106949.*

Martin, J. A., Stiffler-Joachim, M. R., Wille, C. M., & Heiderscheit, B. C. (2022). A hierarchical clustering approach for examining potential risk factors for bone stress injury in runners. *Journal of Biomechanics*, *141*, 111136. https://doi.org/10.1016/j.jbiomech.2022.111136

Marzi, S., Mysiak, J., Essenfelder, A. H., Amadio, M., Giove, S., & Fekete, A. (2019). Constructing a comprehensive disaster resilience index: The case of Italy. *PLoS ONE*, *14*(9). https://doi.org/10.1371/journal.pone.0221585

Mesta, C., Cremen, G., & Galasso, C. (2022). Urban growth modelling and social vulnerability assessment for a hazardous Kathmandu Valley. *Scientific Reports*, *12*(1), 6152. https://doi.org/10.1038/s41598-022-09347-x

Miyazaki, T. (2022). Impact of Socioeconomic Status and Demographic Composition on Disaster Mortality: Community-Level Analysis for the 2011 Tohoku Tsunami. *International Journal of Disaster Risk Science*, *13*(6), 913–924. https://doi.org/10.1007/s13753-022-00454-x

Mulholland, E., & Feyen, L. (2021). Increased risk of extreme heat to European roads and railways with global warming. *Climate Risk Management*, *34*, 100365. https://doi.org/10.1016/j.crm.2021.100365

Nowak-Brzezińska, A., & Gaibei, I. (2022). How the Outliers Influence the Quality of Clustering? *Entropy*, *24*(7). https://doi.org/10.3390/e24070917

Oberlack, C., Pedde, S., Piemontese, L., Václavík, T., & Sietz, D. (2023). Archetypes in support of tailoring land-use policies. *Environmental Research Letters*, *18*(6), 060202. https://doi.org/10.1088/1748-9326/acd802

Oberlack, C., Sietz, D., Bürgi Bonanomi, E., de Bremond, A., Dell'Angelo, J., Eisenack, K., Ellis, E. C., Epstein, G., Giger, M., Heinimann, A., Kimmich, C., Kok, M. T., Manuel-Navarrete, D., Messerli, P., Meyfroidt, P., Václavík, T., & Villamayor-Tomas, S. (2019). Archetype analysis in sustainability research: meanings, motivations, and evidence-based policy making. *Ecology and Society*, *24*(2), art26. https://doi.org/10.5751/ES-10747-240226

O'Brien, K., Leichenko, R., Kelkar, U., Venema, H., Aandahl, G., Tompkins, H., Javed, A., Bhadwal, S., Barg, S., Nygaard, L., & West, J. (2004). Mapping vulnerability to multiple stressors: climate change and globalization in India. *Global Environmental Change*, *14*(4), 303–313. https://doi.org/10.1016/j.gloenvcha.2004.01.001

OECD, E. U. and E. C. J. R. C. (2008). *Handbook on Constructing Composite Indicators: Methodology and User Guide*. https://doi.org/10.1787/9789264043466-en.

Opach, T., Scherzer, S., Lujala, P., & Ketil Rød, J. (2020). Seeking commonalities of community resilience to natural hazards: A cluster analysis approach. *Norsk Geografisk Tidsskrift - Norwegian Journal of Geography*, *74*(3), 181–199. https://doi.org/10.1080/00291951.2020.1753236

Oppido, S., Ragozino, S., & Esposito De Vita, G. (2023). Peripheral, Marginal, or Non-Core Areas? Setting the Context to Deal with Territorial Inequalities through a Systematic Literature Review. *Sustainability*, *15*(13), 10401. https://doi.org/10.3390/su151310401

Papathoma-Köhle, M., Thaler, T., & Fuchs, S. (2021). An institutional approach to vulnerability: evidence from natural hazard management in Europe. *Environmental Research Letters*, *16*(4), 044056. https://doi.org/10.1088/1748-9326/abe88c

Perry, M., & Kader, G. (2005). Variation as Unalikeability. *Teaching Statistics*, *27*(2), 58–60. https://doi.org/10.1111/j.1467-9639.2005.00210.x

Piemontese, L., Neudert, R., Oberlack, C., Pedde, S., Roggero, M., Buchadas, A., Martin, D. A., Orozco, R., Pellowe, K., Segnon, A. C., Zarbá, L., & Sietz, D. (2022). Validity and validation in archetype analysis: practical assessment framework and guidelines. *Environmental Research Letters*, *17*(2), 025010. https://doi.org/10.1088/1748-9326/ac4f12

Preedy, V., & Watson, R. (2010). Aging Index. In *Handbook of Disease Burdens and Quality of Life Measures* (pp. 4140–4140). Springer New York. https://doi.org/10.1007/978-0-387-78665-0_5051

Riach, N., Glaser, R., Fila, D., Lorenz, S., & Fünfgeld, H. (2023). Climate risk archetypes. Identifying similarities and differences of municipal risks for the adaptation process based on municipalities in Baden-Wuerttemberg, Germany. *Climate Risk Management*, *41*, 100526. https://doi.org/10.1016/j.crm.2023.100526

Rocha, J., Malmborg, K., Gordon, L., Brauman, K., & DeClerck, F. (2020). Mapping social-ecological systems archetypes. *Environmental Research Letters*, *15*(3), 034017. https://doi.org/10.1088/1748-9326/ab666e

Rockstrom, J. (2013). *Balancing Water for Humans and Nature*. Routledge. https://doi.org/10.4324/9781849770521

Schwanen, T., Dieleman, F. M., & Dijst, M. (2004). The Impact of Metropolitan Structure on Commute Behavior in the Netherlands: A Multilevel Approach. *Growth and Change*, *35*(3), 304–333. https://doi.org/10.1111/j.1468-2257.2004.00251.x

Sibilia, A., Eklund, G., Marzi, S., Valli, I., Bountzouklis, C., Roeslin, S., Rodomonti, D., Salari, S., Antofie, T.-E., & Corbane, C. (2024). Developing a multi-level european-wide composite indicator to assess vulnerability dynamics across time and space. *International Journal of Disaster Risk Reduction*, *113*, 104885. https://doi.org/10.1016/j.ijdrr.2024.104885

Simon, C., Belyakov, A. O., & Feichtinger, G. (2012). Minimizing the dependency ratio in a population with below-replacement fertility through immigration. *Theoretical Population Biology*, *82*(3), 158–169. https://doi.org/10.1016/j.tpb.2012.06.009

Tariverdi, M., Nunez-del-Prado, M., Leonova, N., & Rentschler, J. (2023). Measuring accessibility to public services and infrastructure criticality for disasters risk management. *Scientific Reports*, *13*(1), 1569. https://doi.org/10.1038/s41598-023-28460-z

Thywissen, K. (2006). *Core terminology of disaster risk reduction: a comparative glossary. UNU Press, Tokyo*.

Tocchi, G., Polese, M., Del Gaudio, C., & Peresan, A. (2024). Multi-hazard exposure characterization of urban settlements: a clustering proposal using open source data. In *EGU General Assembly 2024, Vienna, Austria, 14–19 Apr 2024, EGU24-19858*. https://doi.org/10.5194/egusphere-egu24-19858

https://doi.org/10.5194/egusphere-2025-908
Preprint. Discussion started: 28 February 2025

Tocchi, G., Polese, M., Di Ludovico, M., & Prota, A. (2022). Regional based exposure models to account for local building typologies. *Bulletin of Earthquake Engineering*, *20*(1), 193–228. https://doi.org/10.1007/s10518-021-01242-6

UNDRR. (2023). *Terminology on Disaster Risk Reduction.* . Retrieved from, Https://Www.Undrr.Org/Drr-Glossary/Terminology.

United Nations. (2018). *World Urbanization Prospects: the 2018 Revision*.

Usman, D., & Stores, F. S. (2020). On Some Data Pre-processing Techniques for K-Means Clustering Algorithm. *Journal of Physics: Conference Series*, *1489*(1). https://doi.org/10.1088/1742-6596/1489/1/012029

Van Westen, C., & Woldai, T. (2012). The RiskCity training package on multi-hazard risk assessment. *International Journal of Applied Geospatial Research*, *3*(1), 41–52. https://doi.org/10.4018/jagr.2012010104

Vidal Merino, M., Sietz, D., Jost, F., & Berger, U. (2019). Archetypes of Climate Vulnerability: a Mixed-method Approach Applied in the Peruvian Andes. *Climate and Development*, *11*(5), 418–434. https://doi.org/10.1080/17565529.2018.1442804

Villagrán de León, J. C. (2006). *Vulnerability: A Conceptual and Methodological Review*.

Wachinger, G., Renn, O., Begg, C., & Kuhlicke, C. (2013). The Risk Perception Paradox—Implications for Governance and Communication of Natural Hazards. *Risk Analysis*, *33*(6), 1049–1065. https://doi.org/10.1111/j.1539-6924.2012.01942.x

Walter Gillis, P., Hugh, G., & Betty Hearn, M. (2012). *Hurricane Andrew* (W. Gillis Peacock, H. Gladwin, & B. Hearn Morrow, Eds.). Routledge. https://doi.org/10.4324/9780203351628

Wang, S., Zhang, M., Huang, X., Hu, T., Sun, Q. C., Corcoran, J., & Liu, Y. (2022). Urban–rural disparity of social vulnerability to natural hazards in Australia. *Scientific Reports*, *12*(1), 13665. https://doi.org/10.1038/s41598-022-17878-6

Wicki, S., Black, B., Kurmann, M., & Grêt-Regamey, A. (2024). Archetypes of social-ecological-technological systems for managing ecological infrastructure. *Environmental Research Letters*, *19*(1), 014038. https://doi.org/10.1088/1748-9326/ad1080

Wisner, B., Blaikie, P., Cannon, T., & Davis, I. (2004). *At risk: Natural hazards, people's vulnerability and disasters. 2nd Edition* (Routledge).

Zhao, X., Xu, W., Ma, Y., Qin, L., Zhang, J., & Wang, Y. (2017). Relationships Between Evacuation Population Size, Earthquake Emergency Shelter Capacity, and Evacuation Time. *International Journal of Disaster Risk Science*, *8*(4), 457–470. https://doi.org/10.1007/s13753-017-0157-2