WP 3: Representative models of Built Environment Typologies (BETs) prone to SUOD/SLOD. Case studies selection and data collection

T3.2 - Identification of BETs and their typical risks related to the selected SUOD/SLOD including typical users' exposure.

DELIVERABLE ID	D3.2.1
Deliverable Title	Basic BETs configuration and typical combinations
Date	15/10/2020
Last revision date	
Revision	
Main partner	RM
Additional partners	AN
Authors of the contribution	Currà Edoardo (UNIRM); D'Amico Alessandro (UNIRM); Russo Martina (UNIRM); Angelosanti Marco (UNIRM); Enrico Quagliarini (UNIPM); Gabriele Bernardini (UNIPM)
Deliverable type	report
Number of pages	

Abstract

According to the structures of the research project, …

Da scrivere

Keywords

SUOD, SLOD, basic BETs, statistical analysis, risk parameters.

Approvals

Role	Name	Partner	Date
Coordinator	Quagliarini Enrico	UNIPM	dd.mm.yyy
Task leader	Currà Edoardo	UNIRM	dd.mm.yyy

Revision versions

Revision	Date	Short summary of modifications	Name	Partner

Summary

1. Introduction

Disasters caused by natural hazards can trigger chains of multiple natural and man-made hazardous events over different spatial and temporal scales. Multi-hazard and multi-risk assessments make it possible to take into account interactions between different risks. Classes of interactions include triggered events, cascade effects, and the rapid increase of vulnerability during successive hazards (Scolobig et al. 2015). Recent research has greatly increased the risk assessment community's understanding of interactions between risks. As many regions of the world become subject to multiple hazards (Komendantova et al. 2014), also the number of people affected by them keep in raising. According to the report of the World Bank on the main hotspots of natural hazards (Dilley 2005), about 3.8 million km^2 and 790 million people in the world are relatively highly exposed to at least two hazards, while about 0.5 million km^2 and 105 million people to three or more hazards. Climate change is likely to further increase the exposure to multiple-risks affecting the magnitude, frequency and spatial distribution of hazardous and disastrous events (IPCC 2012). The overlap of different risks, is related also to the definition of Sudden-onset disasters (SUOD) (e.g. earthquakes, hurricanes, floods) and Slow-onset disaster (SLOD) (e.g. drought, pollution, heatwaves, epidemic disease) (PreventionWeb - UNDRR; WHO 2014; UNISDR 2015).

Existing risk assessment methods integrate large volumes of data and sophisticated analyses, as well as different approaches to risk quantification (Komendantova et al. 2014). As Kappes et al. stated in their review on multi-hazard risk we need to examine the frameworks employed in the field of risk management, as well as the interactions between science and practice in terms of knowledge transfer and the applicability of results (Kappes et al. 2012). The successful implementation of disaster risk reduction options and strategies demand not only comprehensive risk assessment schemes, but also an appropriate mechanism to communicate and transfer knowledge on risk and its underlying drivers to the various stakeholders involved in the decision-making process. The challenges for the development of multi-risk approaches are related not only to the applicability of results, but also to the link between risk assessment and decision making, the interactions between science and practice in terms of knowledge transfer, and more generally to the development of capacities at the local level (Gallina et al. 2016). Nevertheless, barriers to the application of multi-risk assessment remain.

As highilighted by Arosio, Martina, and Figueiredo, the approach to natural risk assessment needs to be holistic as "the whole is more than the sum of its parts" (Arosio et al. 2020). This calls for an expansion of the current quantitative risk assessment paradigm, and for measures to frame the study of interlinked disasters. The "multilayer single hazard approaches," in which the hazard potential or risk from one particular physical phenomenon is considered in isolation, pointed out that hazard and risk assessments often take (Gill and Malamud 2016). Indeed, several of those approaches combine distinct cloistered hazards through "standardization schemes" leading to the use of indices and semiquantitative approaches to address the issue of working with different reference units (Kappes et al. 2012). Another approach is to combine the hazards (with exposure and vulnerability) at the risk level using common metrics like monetary loss or probabilities. Monte Carlo simulation of dynamic exposure and vulnerability to represent the dynamic evolution of risk and hence introduce the potential for extreme loss events, define another significative approach to the issue (Mignan et al. 2014). In the end, the approach based on evacuation behaviours simulation, related to risk assessment and the definition of risk mitigation strategies in BEs (Bernardini et al. 2021).

A key question is related to the continue growing of losses from natural disasters if compared to the increasing scientific knowledge on multi-risk approaches (White et al. 2001; Komendantova et al. 2014). One reason is the increasing value of assets exposed to hazards. One means of planning for preparedness for multihazard disasters is by using scenarios that can test capabilities to mitigate or become resilient to disasters. The standard approach in this space is to use empirical models to inform deterministic disaster scenarios. Even though such disaster scenarios are often practical and useful, indeed a limited number of scenarios cannot test the full spectrum of capabilities at an appropriate impact level (Dunant et al. 2021). To summarize, the analysis of multi-risk is a complex problem with variety of challenges.

When facing multi-risk analysis one of the fundamental element to be considered is the risk-prone assets (Schmidt et al. 2011). And this represents the fundamental part to elaborate risk scenarios. Elements at risk or assets are spatio-temporal phenomena, valued by human society, and under threat to be damaged by hazards (i.e. buildings, streets, lifelines, people). They are specified by their spatial extent, location (necessary to determine exposure to hazards) and characterized by attributes describing their vulnerability relevant to the specific hazard. The vulnerability of an asset exposed to a particular hazard is dependent on the magnitude of the hazard exposure and the characteristics of the asset (Schmidt et al. 2011).

Among risk-prone assets, the open spaces have a significative role and the characterization of the built environment is therefore a decisive factor for risk assessment. It is directly connected with the exposure of the inhabitants, buildings vulnerability, and site hazard level (Russo et al. 2020). The resilience of a built environment (e.g., at a wider scale, of a city center) depends on both its fragility and the capacity of social agents to anticipate and to take action in order to adjust to changes and stresses, recognizing that their ability to act is constrained by access to resources and supporting systems. Built Environment (BE) that may be considered resilient exhibits a great number of characteristics, especially if considering wider scales like the urban ones, and also in respect of the critical risk conditions that they could face. Nowadays, the response of Built Environments (BE) to catastrophic events is closely related to the concepts of urban vulnerability and resilience. The classification of the built environment depends on both physical features of the BEs themselves and social aspects. In fact, while the flexibility, redundancy, and modularity of BE guarantee the organization of the strategic function, the preparedness of the social agents increases the ability to absorb shocks. Therefore, the effect of disasters is affected by users' presence combined with the features of the BE, especially while referring to BEs placed in urban areas, which are characterized by complexity issues on BEs correlation, overall layout, users' (i.e., population's) densities, and so on. In this context, existing BEs represent critical systems, because of these built-up areas shows conditions of inner vulnerability. Moreover, it can be evidenced how the characteristics of open spaces are common to national and European contexts, mainly to those related to BEs in urban and historical areas, which are worthy of investigation since they are affected by risk-increasing exposure and vulnerability conditions.

Several methods have been investigated in scientific literature to face the problem of multi-risk scenarios. Often, these researches based the analysis on same case studies, with their own peculiarities. Some of them, have tried to elaborate classification of Built Environments based on specific risk (Quagliarini et al. 2021) or on regional surveys of open spaces (Mandolesi and Ferrero 2001). In their research Mandolesi and Ferrero, have investigated all the open spaces in the Piceno Area of Marche Region of Italy, defining recurring typologies of this type of Built Environments. Even if the main objective of this research was not related to risk-assessment, it established a significative point to elaborate basic scenarios for further specific analysis.

The present work is part of an Italian PRIN (Projects of Relevant National Interest), entitled BE S^{2}ECURe "(make) Built Environment Safer in Slow and Emergency Conditions through behavioUral assessed/designed Resilient solutions" (Grant number: 2017LR75XK), supported by MIUR, the Italian Ministry of Education, University, and Research. According to the previous phases of BE S^{2}ECURE project, nine parameters emerged as significative to describe the BEs in physical terms and evaluable as relevant for the risk assessment. Those analysis were elaborated on the expert judgment and on a statistical analysis developed on a first sample of 133 square of the main Italian towns. Given the numbers of input variables, the final number of combinations was about 768 of BE possibilities. This large quantity would require a huge computational effort for detecting models in the urban environment, and hence, for evaluating the performance of the whole BE under disasters.

In this work, the aim is to elaborate ideal scenarios for thematic simulations related to safety and resilience of built environments in emergency conditions through behavioural approach. The characterization of these scenarios represents a significative passage from Built Environments (BEs) to Built Environments Typologies (BETs). Where BETs, could be defined as idealization of recurring characteristics of open spaces in BEs. In order to define the more suitable definition of BETs, the analysis of a great sample of open spaces, in this work intended as areal spaces (i.e. squares) needed to be investigated.

The actual digital instruments permitted this investigation through the use of the Geographic Information System (GIS). GIS technology is increasingly being used in spatial decision support systems (Tate et al. 2008). In the past few years, GIS emerged as a powerful risk assessment tool and is being used to assess risk from natural hazards. The information retrieved by querying the GIS database serves as inputs for the risk assessment models. GIS could be used also to acquire large amounts of data, in this case useful to define recurring characteristics of open spaces, and so is a tool that is well suited to this type of analysis.

To manage and investigate this large amount of data, cluster analysis appears to be a promising research tool. Cluster analysis is an unsupervised classification technique widely applied in different similar research fields, as the one investigated by Paliaga el al., where they used cluster analysis to classify catchments in terms of anthropogenic disturbance, and then trying to detect a possible link with floods frequency in a Mediterranean area (Paliaga et al. 2020).

In this report (Figure 1), we provide a characterization of the Italian BE open spaces, with the aims of: i) objectively define Italian recurring BE open spaces, defining the BETs, ii) provide a methodology to quantitatively assess of BETS in term of morphological and physical features related to risk iii) set ideal scenarios for further simulation based on behavioural approach.

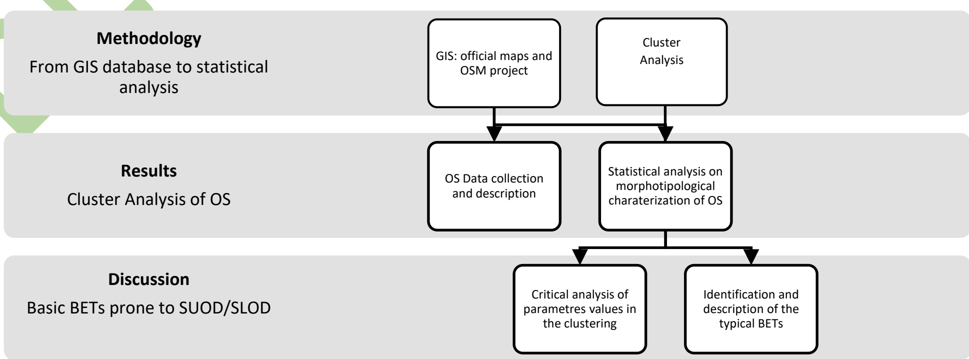


Figure 1: Synthesis of structure of the report and its contributions.

2. Methodology

The methodology is structured in 3 main sections following the scheme in

Figure 2.

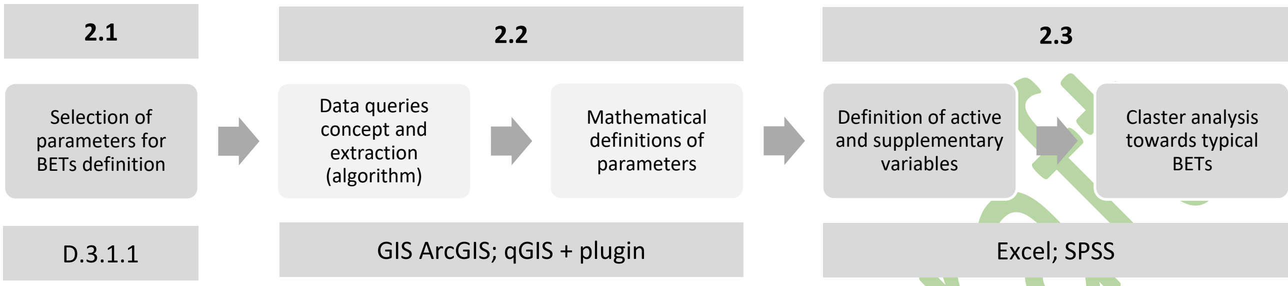


Figure 2: Representation of methodology workflow: identification of databases, import in GIS software, data queries concept, data extraction, statistical significant check and analysis of data, correlation of data towards basic BET definition.

2.1 Open Space parameters for BET description

Following what is set out in deliverable D.3.1.1., the analysis for the definition of BETs starts from the parameters illustrated in Table 1.

Table 1: The nine parameters to define BETs according to the D3.1.1. In light grey the first set of parameters, chose for the typical BETs definition; in yellow the second set of parameters, for the characterization of the case studies associated with typical BETs.

Parameters		Definition according to D3.1.1	Indicators
P1	**Morphology**	Prevalent shape of the OS, catalogued by typology	A. compact b. elongated c. very elongated
P2	**Height**	Comparison between maximum height (H_{max}) of the frontiers and OS medium width (W_{med});	$H_{max} > w_{med}$ $H_{max} \leq w_{med}$
P3	**Structural Type**	Related to the presence of Structural Aggregates along the frontiers of the OS	SA along all frontiers SA not along all frontiers
P4	**Accesses**	In terms of: numbers, width, position, and distance between them; indicators of permeability: α and λ	$\Sigma\ \alpha_i > 36° \setminus \lambda_{LS} > 0.06$ $\Sigma\ \alpha_i \leq 36° \setminus \lambda_{LS} \leq 0.06$
P5	**Special building**	Related to the presence of both building with a special function and building with peculiar structural quality	Yes No
P6	**Construction technique**	In terms of homogeneity of the construction technique, considering masonry as the prevalent type	Yes (all masonry) No
P7	**Porches**	Presence of porches along the frontiers of the OS (%)	Yes (>25%) No (<25%)
P8	**Slope**	Presence of sloped ground or different in the elevation (overhangs, cliffs, ramp/stairs)	Yes (slope >8%) No (slope <8%; flat ground)
P9	**Green**	Presence of green in terms of: trees, bushes/hedges, and grass.	Yes (in any) No

In order to achieve the objective of this report, the nine parameters will be divided into two groups: parameters with data already available on a wide scale (e.g. GIS data repository), suitable for accomplishing a fast statistical analysis, and parameters whose data need direct surveys to be collected, relevant of a detailed analysis. The parameters for fast analysis will be used to define the typical BETs (Table 1, light grey rows), while the parameters for detailed analysis will not be considered in this study (Table 1, yellow rows), but could be further included as additional attributes, in order to characterize deeply the case studies.

- Parameters for fast analysis: P1, P2, P4, P5, P8, P9 (related to morphological, geometric, functional characteristics);
- Parameters for detailed analysis: P3, P6, P7 (related to constructive characteristics).

Parameters have divided into "fast" and "detailed", according to the principle of quick collecting methods to assess BE scenarios. As highlighted by Quagliarini et al. easily available data sources are preferred to avoid costly in situ surveys, and permit to reproduce the workflow by non-expert technicians (Quagliarini et al. 2021). In this specific case, the "detailed" parameters, are not secondary for BE analysis, but need to be investigated trough in situ surveys, because no reliable database is available on these data, and therefore are not included in the present investigation.

This analysis is therefore based on the 5 fast parameters, described in Table 2. The definitions of the parameters has been updated according to the data extractable from GIS databases.

Table 2: The definitions of the fast parameters for the analysis, according to the data extractable from GIS databases .

First set of Parameters		Definition according to the GIS database
P1	**Morphology**	Prevalent shape of the OS, catalogued by typology, in terms of compactness and regularity of the shape
P2	**Height**	Comparison between maximum height (H_{max}) of the frontiers and OS minimum width (d_2);
P4	**Accesses**	In terms of number compared to the perimeter of the OS
P5	**Special building**	Related to the presence of building with a special function; in terms of numbers, according to four categories: places of worship, public buildings, education, cultural and tourism attractions.
P8	**Slope**	Presence of sloped ground, in terms of maximum difference in height
P9	**Green**	Presence of green in terms of percentage of green areas on the overall OS area

2.2 GIS: database and parameters extraction

Among available databases, the authors selected two specific databases containing information related to the description of the BETs, according to the parameters identified for the analysis (Table 3).

Table 3: The two databases selected and the information available in them.

Database	Type of database	Information
Open Street Maps (OSM)1	Open	Identification of Squares Function of building in OS frontier Identification of Streets

		Identification of Green spaces
Edificato dei capoluoghi di provincia - Ministero Ambiente (MinAmb)2	Official	Height of building in OS frontier Height above sea level

[1] www.openstreetmap.org;
[2] http://www.pcn.minambiente.it/mattm/scheda-metadati

One of the main characteristics of GIS software is the possibility to interrogate a source of data, in a spatial context, in order to get specific answers. These questions or "queries" are sets of commands and rules that are used to browse a database.
According to ArcGIS dictionary (ArcGIS 2021), a "query" is a request that examines feature or attributes in order to display those records that satisfy user-selected criteria. In GIS environment, queries are divided in attribute query and location query.
On the one hand, attribute queries ask for information from the tables associated with features or from standalone tables associated with the GIS. Attributes can be numeric values, text strings, Boolean values, or dates. On the other hand, the location or spatial queries are derived directly from the position of features on the map. In this way, inside a GIS environment, the features related to the records selected by the process are highlighted on the map as well as in the table of attributes.
All queries, both attribute and spatial ones, have three main parts: a source, a filter, and a relationship. The source can be a table or feature class. The filter can be an attribute value or a shape or feature. The relationship between the source and the filter is based on specific operators like the Comparison ones (=, <>, >, >=, <=), the logical ones (LIKE, AND, OR, and NOT) or spatial ones (Intersect, Are Within a Distance Of, Contain, Are Contained By…).
In the herein document, Spatial queries performed using Select by Location deal with vector data and use a shape as a filter and its relationship with features in the source layer. The choice of spatial operator (i.e. the relationship by the query) depends on the types of features that will be used for the source and filter.
After the query, it is worth introducing "spatial analysis" concept in GIS environment. Spatial analysis integrates spatial information to derive new and additional meaning from GIS data source. GIS Applications normally have spatial analysis tools for feature statistics or geoprocessing. The types of spatial analysis that can be used vary according to subject areas. In the following sections we will introduce the performed spatial analysis, carried out with vector data.
In order to achieve a proper GIS implementation of the parameters descripted above, suitable queries and processing tools are chosen inside QGIS software (Table 4).
In particular, the research focus on the query place=square (Figure 3) that is used to map a town or village square, usually defined as paved and open public space, generally of architectural significance, surrounded by buildings in a built-up area. According to its definition (Mooney and Minghini 2017), the majority of tag place=square is paved and suitable for open markets, concerts, political rallies, and other events that require a solid surface. They are also known as city, urban, public or market square and therefore they are usually named.
The following processing tools (algorithms) (Table 4) have been chose to extract the necessary information from GIS databases:

- the "minimum oriented bounding box", to calculate the rotated rectangle of minimum area that covers each OS element in an input GIS layer and extract its area (A_{MOBB});
- the "minimum circumscribed circumference", to calculate the minimum circumscribed circumference that covers each OS element and extract its diameter (d_1) and radius (r_1);

- the "inaccessibility pole" (Figure 4), to find the center of the maximum inscribed circumference for each OS and extract its diameter (d_2) and radius (r_2).

Table 4: Queries concept and structure used for the extraction of data.

Parameters		Source database	OSM Research query	Processing tools
P1	**Morphology**	OSM	place=square; admin_level=2; admin_level=4; admin_level=6; admin_level=8;	Minimum oriented Bounding Box;Minimum Circumscribed Circumference; Inaccessibility pole; Intersection; Fix geometry; Join attribute per position.
P2	**Height**	OSM + MinAmb	place=square; layer building (MinAmb).	Buffer; Intersection; Fix geometry; Join attribute per position.
P4	**Accesses**	OSM	place=square; Highway=pedestrian; Highway=residential; Highway=service; Highway=living_street.	Merge; Intersection; Fix geometry; Join attribute per position.
P5	**Special building**	OSM	place=square; TOURISM: tourism=attraction; tourism=museum. PUBLIC: amenity=townhall; amenity=police. RELIGION: amenity=place_of_worship; building=church; building=temple. INSTRUCTION: amenity=university; amenity=school; amenity=college.	Merge; Intersection; Fix geometry; Join attribute per position.
P8	**Slope**	OSM + MinAmb	place=square; layer building (MinAmb).	Buffer; Merge; Intersection; Fix geometry; Join attribute per position.
P9	**Green**	OSM	place=square; leisure=park; leisure=garden; landuse=forest.	Merge; Intersection; Fix geometry; Join attribute per position.

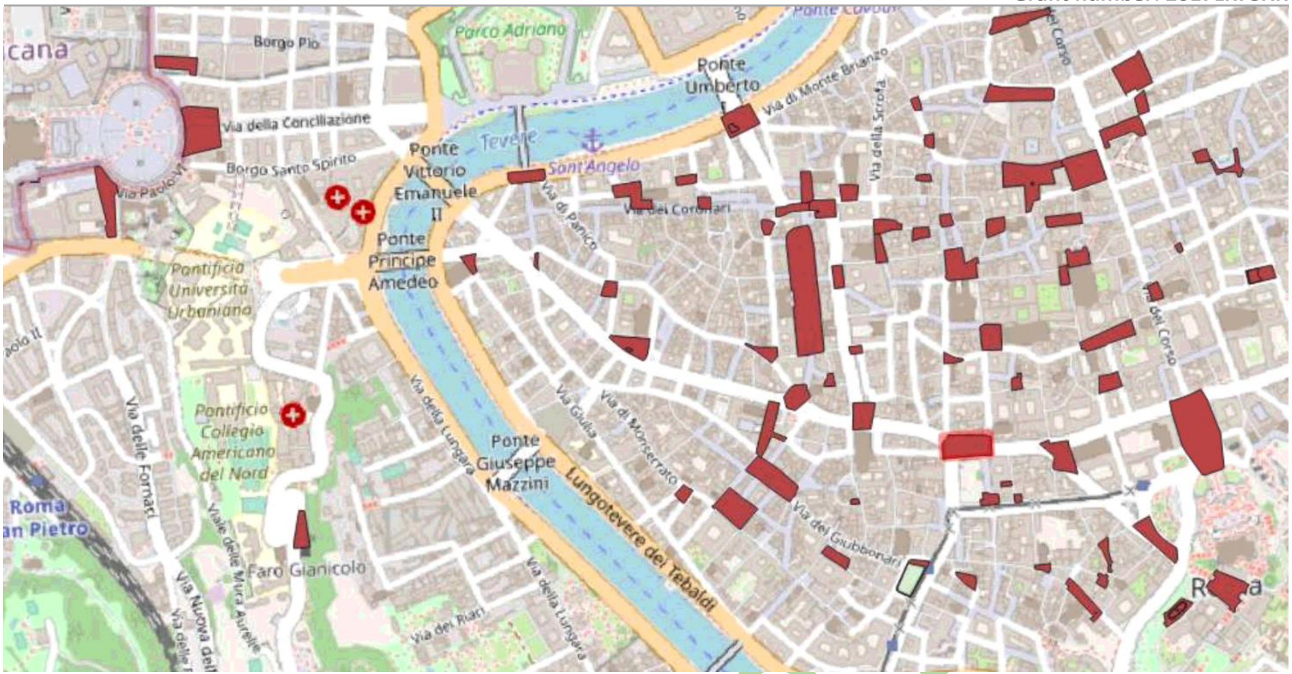


Figure 3: Italian OSM OS layer extraction from OSM performed by authors. In red query: place=square attribute.

P1

The parameter P1 was elaborated as product of the area regularity values and the radius ratio values (Equation 3).

$$P1 = P1a \; x \; P1b = \; (Area \; regularity \; x \; Radius \; ratio) \tag{1}$$

where "area regularity" and "radius ratio" are defined as follow.

$$P1a = Area \; regularity = \left(\frac{A_{AS} \; [m2]}{A_{MinimumOrientedBoundingBox} \; [m2]} \right) \tag{2}$$

Where:

- A_{AS} is the area [m^2] of the OS considered;
- A_{MOBB} is the area of the minimum oriented bounding box [m^2] of the OS considered;

The oriented bounding box shows indeed differences in approximation between more regular shapes and irregular or composite ones. In this way, the higher the "Area regularity" attribute, the more regular/quadrangular the OS shape.

The "radius ratio" was evaluated as a comparison between "minimum circumscribed circumference" and "Inaccessibility Pole" of the ASs.

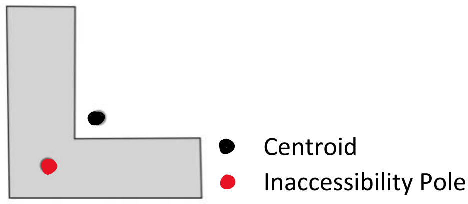


Figure 4: Difference among Centroid and Inaccessibility Pole.

On the one hand, “Minimum Circumscribed Circumference” processing tool algorithm calculates the minimum circumscribed circumference that covers each element in an input layer.On the other hand, “Inaccessibility Pole” processing tool (Garcia-Castellanos and Lombardo 2007; Agafonkin 2016) is herein defined as the center of the largest OS inscribed circle. It is the visual center, the point within an OS that is farthest from an edge. Contrary to centroid concept, if the shape is concave or has a hole, the inaccessibility pole will not fall outside of the shape. In the present deliverable, we use “Inaccessibility Pole” processing tool, which uses the poly-label algorithm (Agafonkin 2016). This tool is based on an iterative approach that guarantees to find the true inaccessibility pole coupled with a specified tolerance (in layer units). More precise tolerances require more iterations and will take longer to compute. The radius of the maximum inscribed circumference, calculated as distance from the pole to the edge of the polygon, will be stored as a new attribute in the output vector. In this sense, the concept of Inaccessibility Pole differs substantially from Centroid concept that creates a new vector of points, where the points represent the centroid of the geometries of an input vector.

$$P1b = Radius\ Ratio = \left(\frac{r2\ [m]}{r1\ [m]}\right) \tag{3}$$

Where:

- r_2 is the radius of the maximum inscribed circumference, calculated through the inaccessibility pole;
- r_1 is the radius of the minimum circumscribed circumference;

P2

P2 evaluates the heights of the buildings comparing to the width of the OS. To extract the buildings heights of the frontier, a specific WFS (Web Feature Service) layer “Edificato dei capoluoghi di provincia” has been imported in qGIS environment from “Geoportale Nazionale - Ministero dell'Ambiente e della Tutela del Territorio e del Mare”. With the aim of implementing in GIS the concept of BE frontier, the authors replace each OSM OS extracted geometry with 10 meters buffered one and perform a special intersection with the new building layer. Through the intersection among the OSM dataset with OS areas and the MinAmbiente dataset with heights of buildings (Figure 5), for each OS the authors considered H_{max} as the maximum height of buildings of the OS frontier:

$$P2 = \frac{Hmax}{d_2} \tag{4}$$

P2 can assume value lower than 1 (non-critical values) or greater than 1 (critical values)

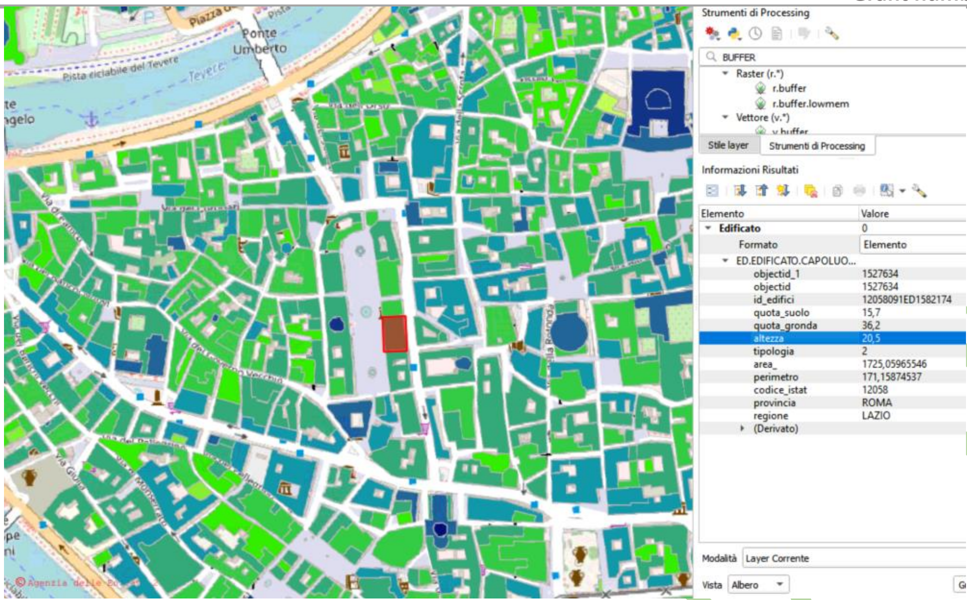


Figure 5: Italian principal city building height extraction from MINAMBIENTE.

P4

The parameter P4 evaluates the number of access point in a square, with respect to the perimeter of the reference OS. The authors defined P4 as number of accesses / perimeter of OS. For this reason, in order to extract the BE street network, the authors formulate specific query to OSM database (Table 4): Highway=pedestrian; Highway=residential; Highway=service; Highway=living_street. Indeed, in OSM the key highway=* identify linear geometry and is the main key used for any kind of road, street or path. The value of the key helps indicate the importance of the highway within the road network as a whole.
The previous step is a prerequisite for the spatial analysis in which the intersection between the OS and the linear elements of the street network identifies points in correspondence of the OS accesses (Figure 6).

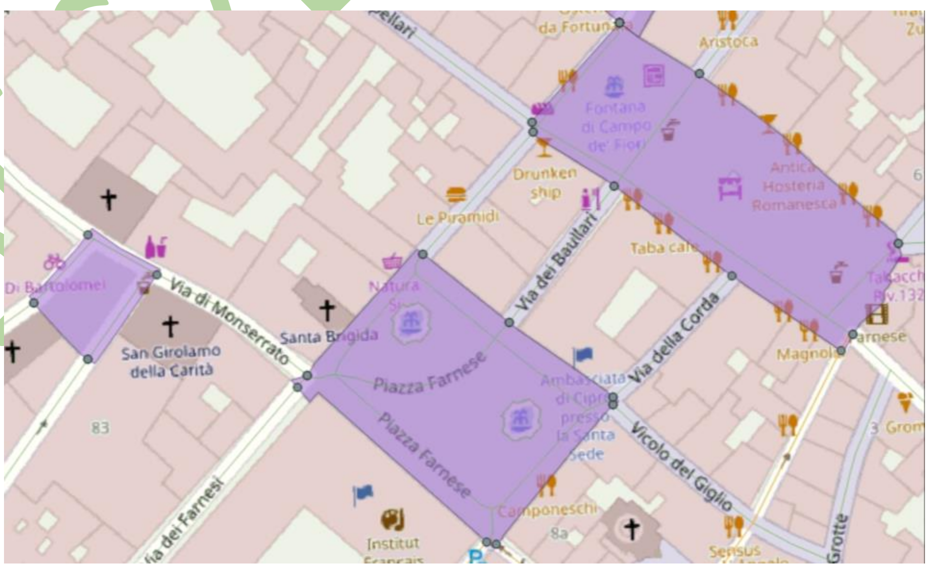


Figure 6: Example of access intersection between Italian OS and street network.

In order to avoid duplicates points overlapped, due to possible change of street network name from outside to inside the AS, a final merge of any point element has been processed. These points inherited the attributes from the AS and the street of intersection, allowing a rapid assessment of P4 by exporting the geometries attributes in .xls file.
P4 is defined as follow (equation 5):

$$P4 = \frac{\sum number\ of\ accesses}{AS\ perimeter} \quad (5)$$

P5

The parameter P5 indicates if any special buildings are present or not in the OS frontier. In order to evaluate this aspect, the authors chose suitable query (Table 4) to identify the presence of building dedicated to special function, according to specific classes:

- religious buildings (i.e. church or other place of worship; Figure 7);
- public buildings (i.e. town hall and police office);
- buildings for education (i.e. school, college and university);
- buildings with cultural or tourism importance (e.g. museum, palaces, castles).

Therefore, a cleaning of the data was processed to remove duplicates present in the OSM dataset, caused by redundant data implementation in the OSM database (Mooney and Minghini 2017). Using squares and buildings ID from the official MinAmbiente database, this process allowed removing any supposed duplicates.
P5 is defined as follow (equation 6):

$$P5 = \sum number\ of\ special\ building\ in\ the\ AS \quad (6)$$

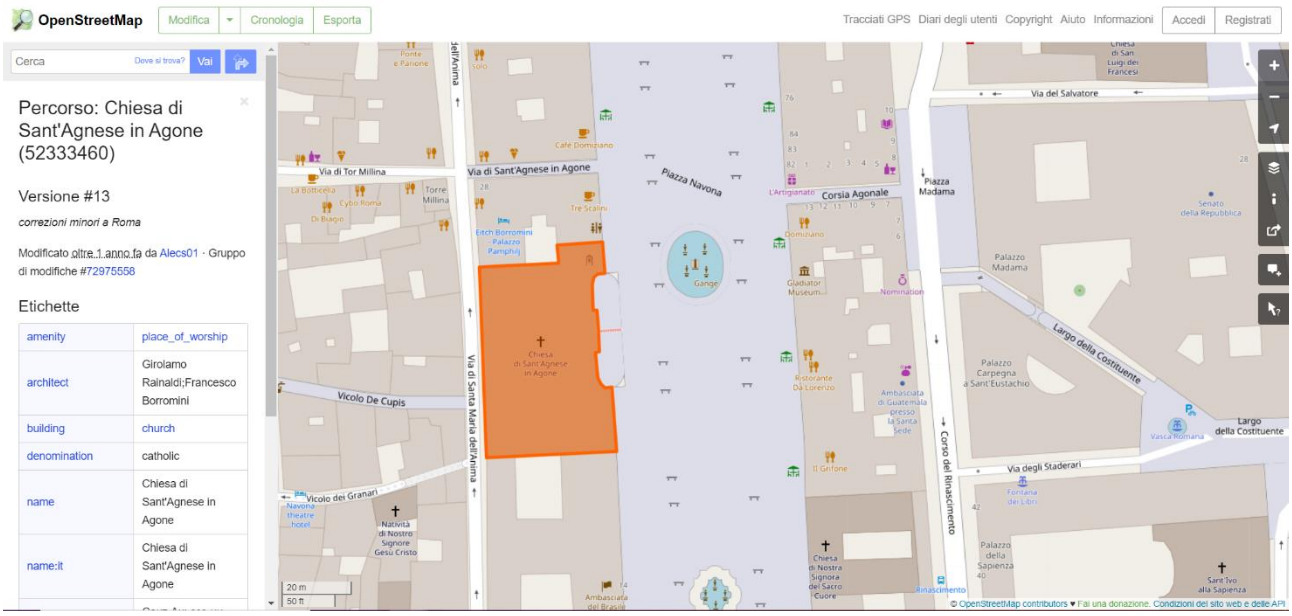


Figure 7: Example of the query amenity=place_of_worship to identify the religious building on the OS frontiers.

P8

According to D3.1.1, the parameter P8 indicates the presence of a sloped ground or different in the elevation (overhangs, cliffs, ramp/stairs). Through the information implemented in the chosen GIS dataset (Table 4), and in particular from “Edificato dei capoluoghi di provincia”, the authors could extract the value of minimum (HSL_{min}) and maximum (HSL_{max}) height above sea level for each OS. In this way it has been possible to evaluate the maximum diferrence of height above sea level for each OS of the OS (equation 6):

$$P8 = \Delta HSL = HSL_{max} - HSL_{min} \tag{7}$$

P9

The presence of green areas is indicated by the parameter P9. The authors interrogated the dataset to verify if any kind of green areas is present within the OS considered (Table 4). The result of this interrogation is a specification of the presence of green areas, defined by the use in the OPM database. These green areas could be of different kinds, including field surfaces, trees and brushes. It was possible to extract the overall number of OSs with green areas and the percentage of the green areas on the overall OS area (equation 7):

$$P9 = \%\text{Green area} = \frac{\sum green\ areas}{AS\ area} \tag{8}$$

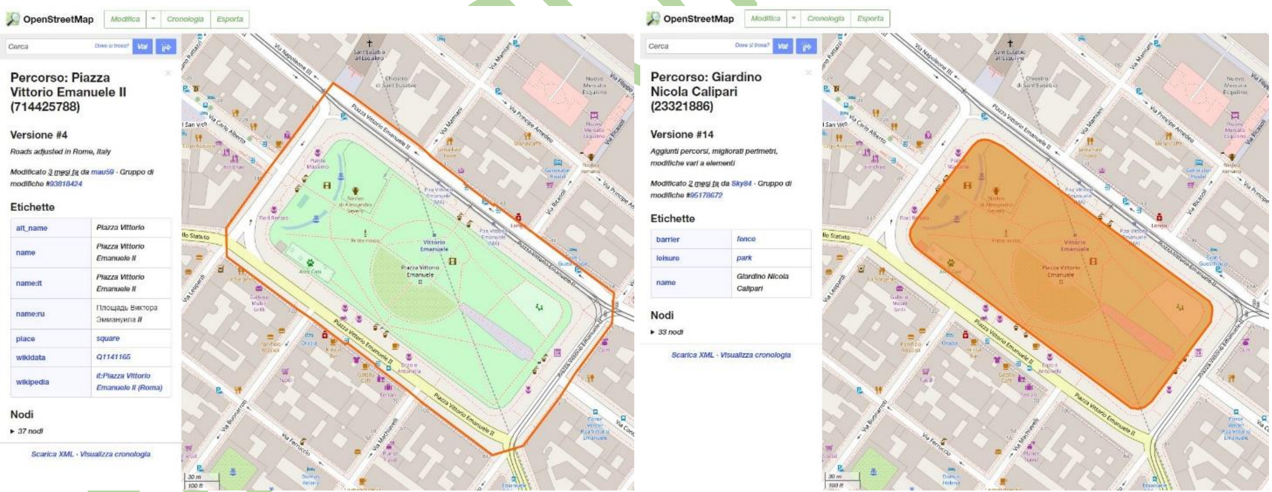


Figure 8: Example of green area (query leisure=park) within an OS; Piazza Vittorio, Roma.

2.3 Statistical analysis

Cluster analysis is a multivariate data mining technique (James and others 1967; Halkidi 2001) used in several research fields from medicine to sanitary engineering, psychology and economy. The focal concept is, given a large set of objects with data attribute associated, to identify groups where objects share a degree of similarity. The higher the similarity within a group and the differences between objects belonging to different groups, and the better is the degree of clustering. Cluster analysis belongs to the unsupervised classification techniques: no constrain or a priori condition is imposed, and the classification derives solely from the data (Halkidi 2001; Paliaga et al. 2020). Clustering processes may be hierarchical or partitional: in the first case clusters may have sub-clusters, while in the second one clustering is obtained dividing the set of elements in non-overlapping subsets (clusters). Considering the aim of the research, both hierarchical and partitional techniques have been investigated.

Firstly, a hierarchical process, based on single-linkage method, has been performed to individuate multivariate outliers in the database, and in the end a partitional technique has been chosen.

The agglomerative hierarchical clustering algorithm (single-linkage) produces a cluster tree that starts with every single data in a single cluster and it continues to merge with the closest pair of clusters according to some similarity criterion (Euclidian distance) until all the data are grouped in one cluster (Satari et al. 2017).

The study was intended to group catchments that share similarities in term of morphology of the Open Spaces (OS) (P1), the height of their fronts related to their width (P2), the number of access related to OS perimeter (P4), the presence of special buildings (P5), the difference in height of the OS (P8) and the incidence of green areas within the OS (P9).

Several algorithms have been developed to perform clustering. In the present research the authors tested different type of algorithms and finally choose the partitional k-means one (Halkidi 2001; Paliaga et al. 2020). Every object x_i is represented in the Euclidean n-dimension space defined by data attribute: the algorithm allows identifying a number of user defined *k* clusters represented by their centroid. Starting from *k* randomly chosen centroids every object is assigned to a cluster and then the centroid is updated; the process is iterated until converging, that is the centroids do not change anymore. Due to the nature of the n-dimension space, the distance between elements is defined in terms of Euclidean metric; formally the algorithm minimizes the sum of the squared error (SSE) that is used to evaluate the quality of clustering (equation 9):

$$SSE = \sum_{i=1}^{k} \sum_{x \in C_i} dist\ (c_i, x_i)^2 \qquad (9)$$

where *k* is the number of clusters, *x* is an object of the set, C_i is the i^{th} cluster, c_i is the centroid of cluster C_i and *dist* is the Euclidean distance.

The selected variables were standardized prior to performing cluster analysis, with Z-score method, and the number of clusters was determined on the basis of 2 criteria: the squared correlation ratio (R^2) and the pseudo F (pF) statistic. The “optimal” number of clusters corresponded to a consensus among these 2 criteria: pseudo F peaks combined with an inflection in the R^2 increase (Dujardin et al. 2004). F-values provide insight in the magnitude of univariate differences between clusters, while R^2 provides an explanation of separation between clusters.

The final goal of cluster analysis is to evaluate whether the cluster sizes and profiles are meaningful and interpretable, and clusters can be further characterized on variables which were eventually not included in the cluster analysis (Lee et al. 2005).

Statistical analysis was performed with SPSS Version 26.0 statistic software package (IBM Corp. 2019).

2.4 From cluster analysis to BET definitions

From the results of the cluster analysis, the authors defined the BETs through a critical analysis of the emerged clusters. An evaluation of the mean values of the parameters in the clusters was performed, and a division into classes of values characterizing the recurring BETs was elaborated from the critical comparison.

The parameters critical classes were set on the analysis of the sample of the individual clusters and the statistical values resulting from the cluster analysis.

Firstly, this analysis was based on the 5 active variables (P1, P2, P4, P8 and P9), and then the additional variable (P5) was also included. In addition, a more detailed analysis of the relationship between critical height and OS width allowed a further investigation of the results.

For the characterization and representation of different BETs, the mean values and the interquartile ranges of the parameters within each cluster were considered (Q1 = 25% and Q3 = 75%). According to these values, the classes of parameters value have been elaborated in two different type: tripartite on the IQ where significative differences arise, i.e. for P1, and dichotomic for other parameters based on the mean value.

3. Result and Discussion

Results include cluster analysis output and the following critical analysis conduct by the authors to define BETs. Section §**Errore. L'origine riferimento non è stata trovata.** describes the process to obtain the final dataset, starting from the preliminary extraction of the data from the selected databases. Section §**Errore. L'origine riferimento non è stata trovata.** illustrate the process of cluster analysis, including the elimination of the multivariate outlier (Section §**Errore. L'origine riferimento non è stata trovata.**), the clustering method comparison through the pF and R2 indicators result, and the output of the k-means method, chosen from the authors as the most suitable for the present research (Section §**Errore. L'origine riferimento non è stata trovata.**).
Subsequently, Section §**Errore. L'origine riferimento non è stata trovata.** shows the critical analysis made by the authors on the results of the k-means output, in order to assign specific classes to each parameter (Section §**Errore. L'origine riferimento non è stata trovata.**), describe the clusters through the active variables (Section §**Errore. L'origine riferimento non è stata trovata.**) and the supplementary variables (Section §**Errore. L'origine riferimento non è stata trovata.**), and define the final BETs in term of morphological, geometric and functional characteristics (Section §**Errore. L'origine riferimento non è stata trovata.**). Then significative examples of defined BETs are reported (Section §3.4). In the end, limitations of the approach are briefly discussed (Section §3.5).

3.1 Dataset description

The final sample is composed of 1.113 cases and includes OSs of Capitals of Provinces of the Italian Regions (Figure 9). Although the initial sample extracted from OSM through query place=square encompass 8.889 cases on the entire Italian territory, the author took into account only the cases included also in the MinAmbiente database, in order to ensure data coverage for all the BET parameters for fast analysis (P1, P2, P4, P5, P8, P9) (Table 5). Among these, some parameters dictated a further reduction of the database: it was possible to extract the information for parameters P2 and P8 for 1392 cases, the data relating to the calculation of P4 for a total of 1113 cases, while parameter P5 is correctly compiled in the OSM database for 476 cases only (Figure 9).

Table 5: Link between data source and BET parameters.

Parameters		Source database	Sample (num. AS included)	Geographic extension
P1	**Morphology**	OSM	8.889	Italian territory
P2	**Critical Height**	OSM + MinAmb	1.392	Capital of Province
P4	**Accesses**	OSM	1.113	Italian territory
P5	**Special building**	OSM	476	Italian territory
P8	**Slope**	OSM + MinAmb	1.392	Capital of Province
P9	**Green**	OSM	8.889	Italian territory

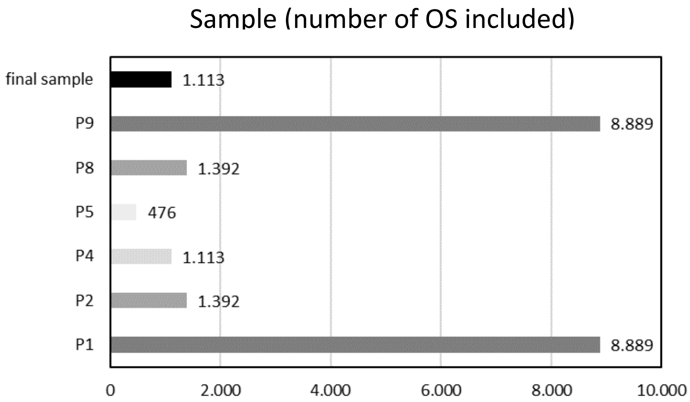
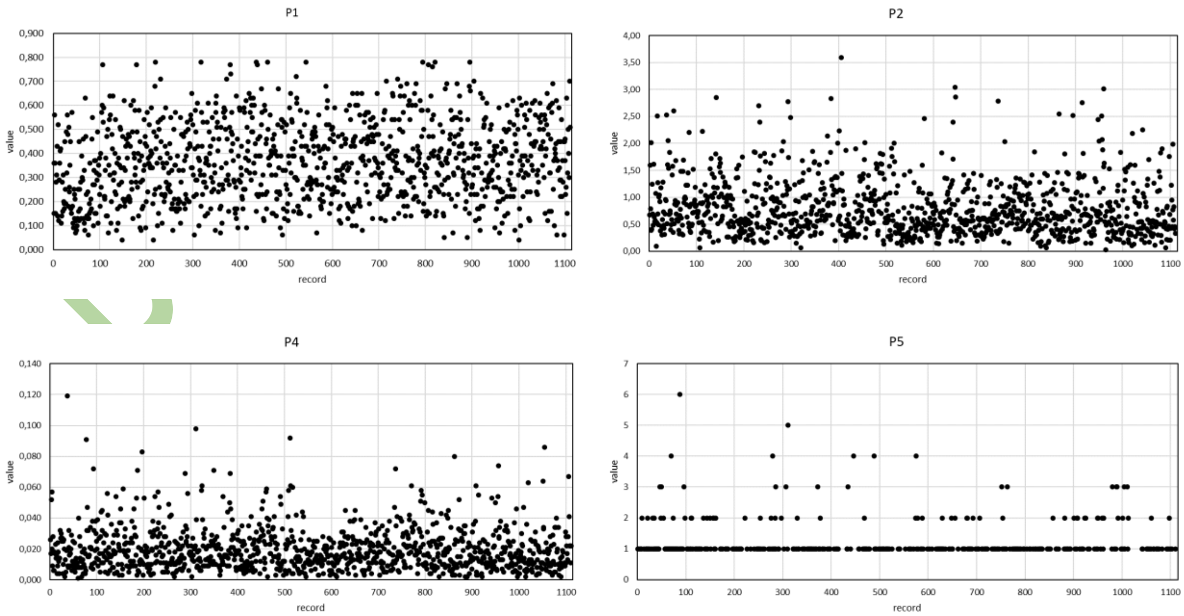


Figure 9: Numbers of cases included in both OMS and MinAmb databases for each parameters, compared to the final sample.

Given the nature of the samples being analyzed, the parameters are divided into *active* (P1, P2, P4, P8, P9) and *supplementary* (P5) for the cluster analysis. As can be seen in the graphs in Figure 10 and Figure 11, the values of the individual parameters are continuous and equally distributed in the total range that each of them can assume. There are no significant trends in the distribution of values at this stage.



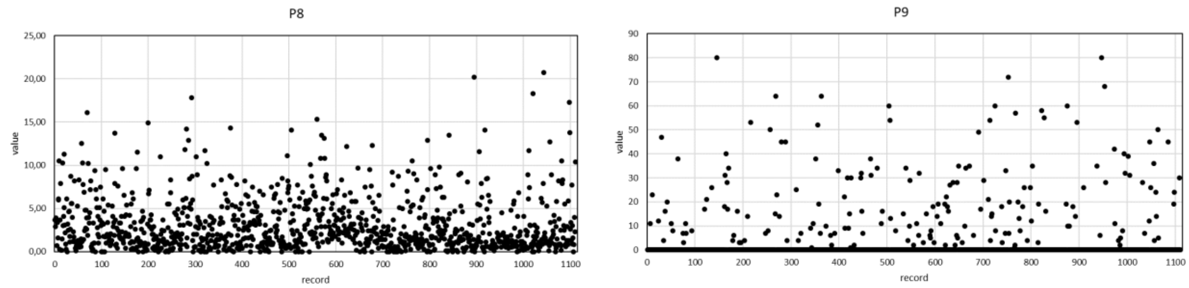


Figure 10: Scatterplots of single parameters, referred the final sample (1113 cases).

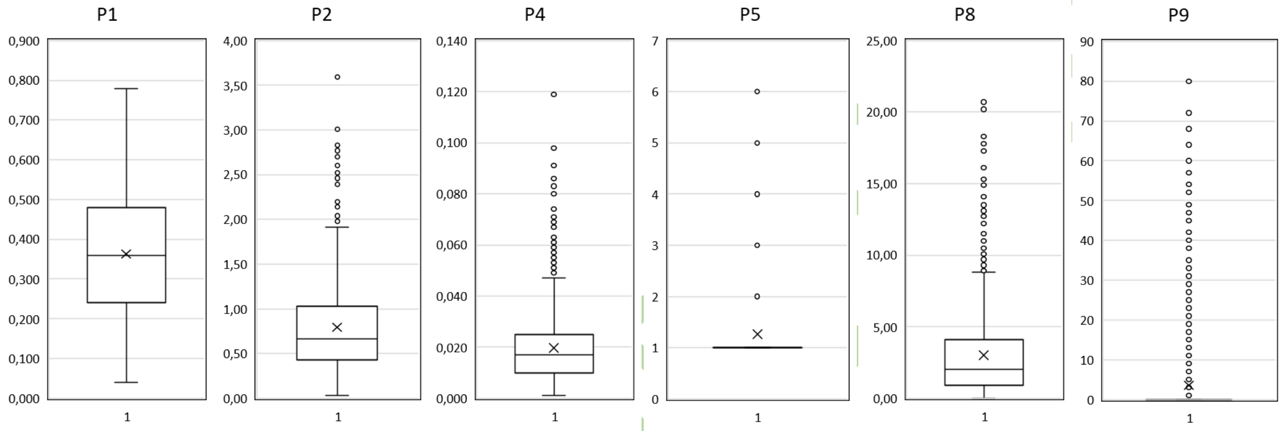


Figure 11: Boxplots of single parameters, referred the final sample (1113 cases).

3.2 Cluster Analysis

3.2.1 Single Linkage method and multivariate outliers

From the observation of the dendrogram (Figure 12) resulting from the analysis, a partition into 3 clusters allows the identification of the statistical units that can be considered multivariate outliers. The cluster division selected shows singular elements that merges to the vary last phases of the algorithm calculation.

In this way, 2 cases have been identified that appear to be exceptional urban spaces in terms of their geometric dimensions and character: one of these is a small widening space between dense fabrics on which very tall buildings overlook (Piazza dei Baroncelli in Florence); one is closer to being a tree-lined avenue than a square (Piazzale Giuseppe Mazzini in Padua) (Figure 13).

The two cases in question were eliminated from the sample, which thus reaches a final value of 1111 cases. The cluster analysis was performed on this final sample.

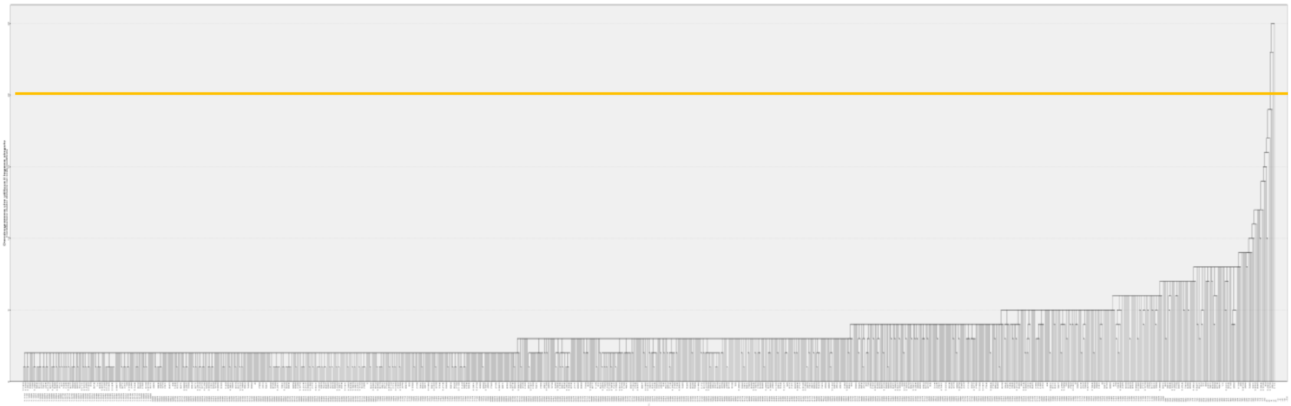


Figure 12. Dendogram produced with Single-linkage algorithms. Yellow line represent the selected cluster divisions.

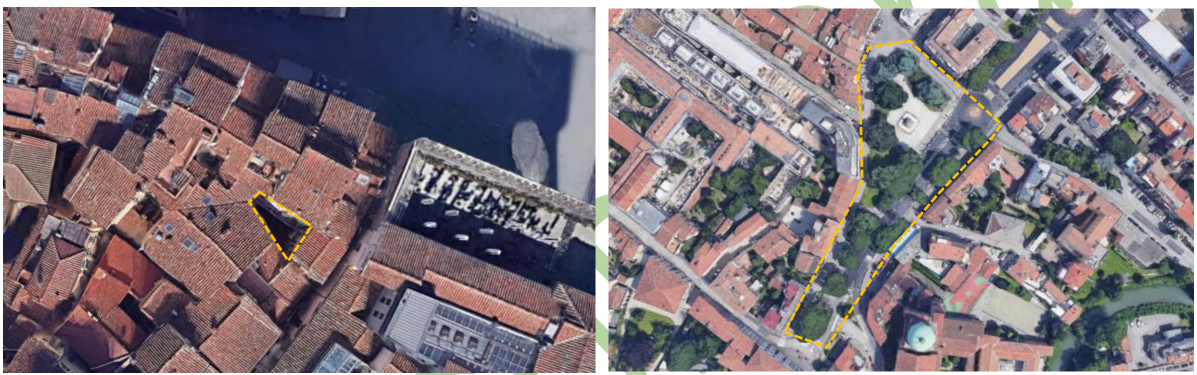


Figure 13: Multivariate outliers: Piazza dei Baroncelli in Florence (on the left) and Piazzale Giuseppe Mazzini in Padua (on the right).

3.2.2 Clustering methods comparison

Cluster analysis and relative pseudo F and R^2 calculation have been performed for the set of 5 parameters (P1, P2, P4, P8 and P9). Cluster analysis was conducted using four methods: k-means, Ward, average-linkage, and complete-linkage. The solution in the term of a number of clusters was searched from 3 to 12 groups for each method. The indicators pF and R^2 were used to compare the result and choose both the suitable number of groups and the methods for the clustering process. These indicators allow estimating the within and between variance among the result groups, pointing out the separation among the groups and the similarity of the elements within each group. The R^2 is a monotonic function and can assume values from 0 to 1; the higher is the values, the higher is the distance between the element in the groups (explained variance). The pF is not a monotonic function and has no predefined range; the highest value that the function assumes correspond to the number of clusters that present the highest similarity among the element within the groups.

Comparing the results, for each method the pF values indicate five clusters as the most optimize solution in terms of similarity of the elements. Moreover, the k-means method shows the maximum values of the R^2, suggesting this method is the most capable to characterize the peculiarities among the groups (Figure 14).

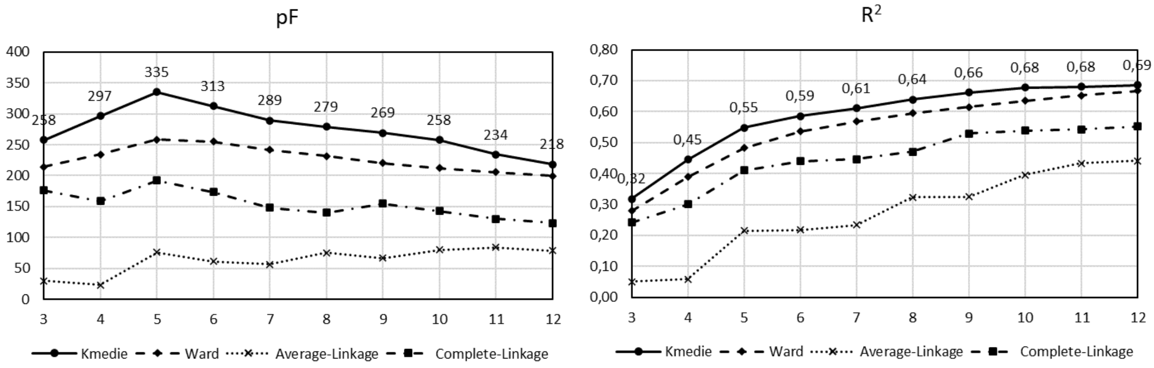


Figure 14. Values of pF and R2, calculated for different cluster divisions (from 3 to 12 groups) and compared between different cluster analysis methods.

In In blue, the values that deviate from the global means and represent the main characterization of each cluster.

Table 6 and Figure 15 the mean values of parameters for clusters individuated with K-means algorithm are reported. In blue, the values that deviate from the global means and represent the main characterization of each cluster.

	1	2	3	4	5	Total
P1	0,310	0,231	0,509	0,395	0,367	**0,362**
P2	0,882	0,830	0,538	1,308	0,692	**0,793**
P4	0,016	0,014	0,017	0,046	0,017	**0,020**
P8	9,037	1,976	2,038	2,257	2,930	**3,015**
P9	0,013	0,015	0,011	0,012	0,418	**0,036**
cases	146	391	369	141	64	1111

Table 6: Means of the variables for the 5 clusters, obtained with the k-means method; "total" indicates the values of the means of the entire dataset.

	1	2	3	4	5	Total
P1	0,310	0,231	0,509	0,395	0,367	**0,362**
P2	0,882	0,830	0,538	1,308	0,692	**0,793**
P4	0,016	0,014	0,017	0,046	0,017	**0,020**
P8	9,037	1,976	2,038	2,257	2,930	**3,015**
P9	0,013	0,015	0,011	0,012	0,418	**0,036**
cases	146	391	369	141	64	1111

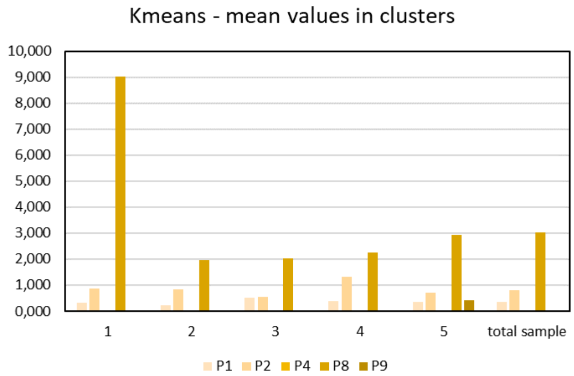


Figure 15. Mean values of parameters for clusters individuated with K-means algorithm.

3.3 BET definition

As a result of the cluster analysis and a sample check on the cases included in the final sample, it was possible to define classes for the parameters values estimation, in order to describe the typical OSs associated with each cluster.

3.3.1 Parameters values

From a critical analysis of the average values of the parameters divided into the clusters, the parameters classes for the analysis of open spaces were assigned (Table 7).

Table 7: Classes associated with the range of the values resulted from the k-mean clustering.

	Type	Class 1	Class 2	Class 3	Note
P1	Tripartite	X<0.25 low level of compactness and regularity	0.25<X<0.50 medium level of compactness and regularity	X>0.50 high level of compactness and regularity	Sample check of the data
P2	Dichotomous	X<1.00 without problems of overturning of the fronts	X≥1.00 with problems of overturning of the fronts	- -	1.00 is the limit value considering the nature of the values (see §**Errore. L'origine riferimento non è stata trovata.**)
P4	Dichotomous	X<0.02 critical ratio between number of accesses/perimeter	X≥0.02 no critical ratio between number of accesses/perimeter	- -	0.02 is the mean values
P8	Dichotomous	X<3.00 flat or slightly sloping ground	X≥3.00 sloping ground or changes in elevation	- -	3.00 is the mean values
P9	Dichotomous	X<0.30 No green areas	X≥0.30 Presence of green areas	- -	0.3 = 30% of OS surface with green area

P1 is process on both percentiles and a sample check of the data (Table 8); P2 is based on the critical value equal to 1.00, so two classes are defined for above e below this value; P4 and P8 are classified referred to mean value; P9 is defined by the threshold value of 30%, representing the surface with green area that distinguishes the clusters 1, 2, 3, and 4 (not characterized by the presence of green) from the cluster 5 (characterized by the presence of green).
In particular, for P1 a sample check process was performed to confirm the three classes highlight by the percentile. Figure 16, Figure 17, and Figure 18 show the correspondence of the definition given compared to the morphological classes.

Table 8: Sum of the percentiles used as reference for the clustering of P1.

		Percentile						
		5	10	25	50	75	90	95
Weighted arithmetic mean	**P1**	,1170	,1522	,2350	,3560	,4760	,5830	,6338
value used				0,25		0,50		

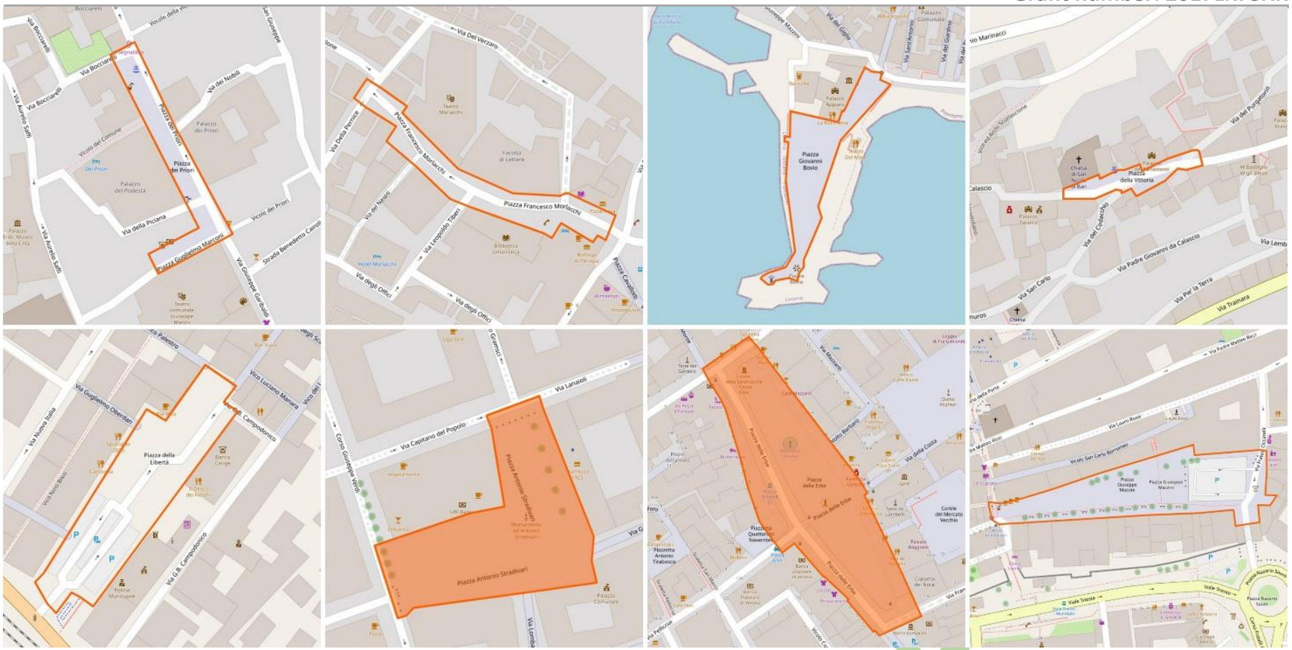


Figure 16: Example of the OS belonging to the class 1, not regular and compat (from top left to bottom right): Piazza dei priori, Narni, Terni (0,046); Piazza Francesco Morlacchi, Perugia (0,060); Piazza Giovanni Bovio, Livorno (0,079); Piazza della Vittoria, Calascio, L'Aquila (0,091); Piazza Antonio Stradivari, Cremona (0,215); Piazza delle Erbe, Verona (0,223); Piazza Giuseppe Mazzini, Macerata (0,242).

Figure 17: Example of the OS belonging to the class 2, medium regular and compact (from top left to bottom right): Piazza Bernardini, San Concordio, Lucca (0,289); Largo Giovanni Rodari, Trino, Vercelli (0,323); Piazza Giacomo Matteotti, Marsala, Trapani (0,342); Piazza Fausto Simonetti, Ascoli Piceno (0,354); Piazza del Duomo, Spoleto, Perugia (0,365); Piazza Santa Teresa, Cosenza (0,436); Piazza delle Scuderie, Alessandria (0,468); Piazza Dogana, Catanzaro Lido (0,487).

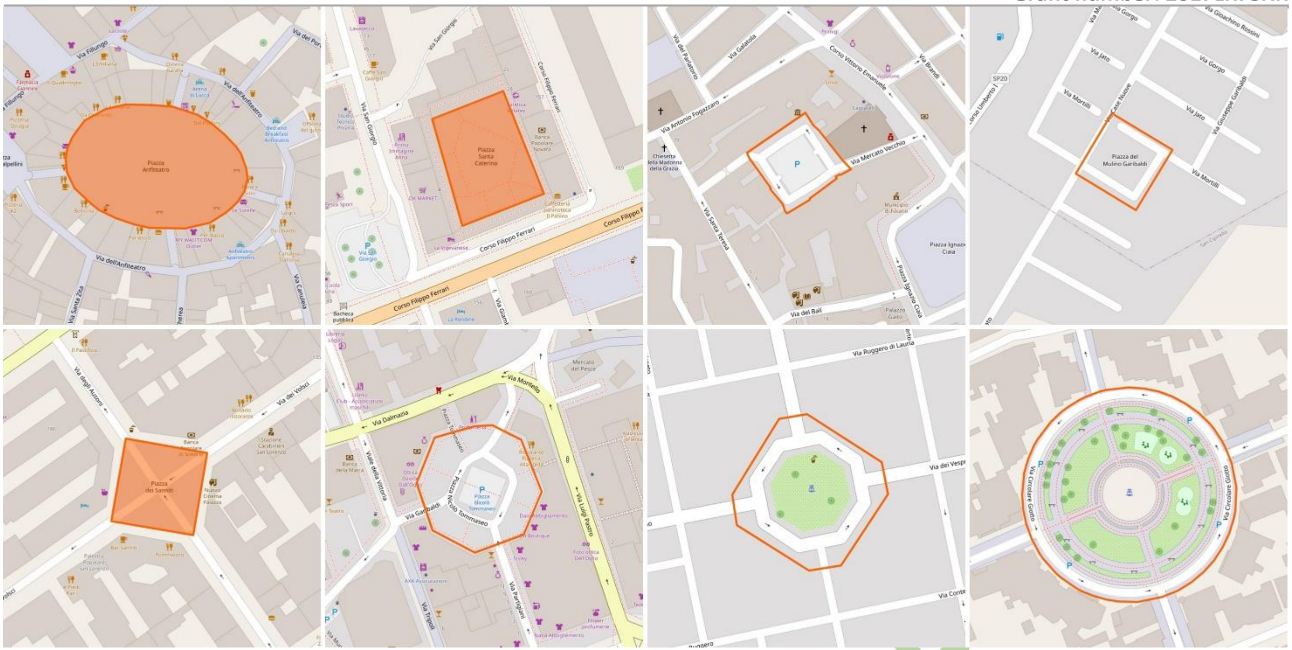


Figure 18: Example of the OS belonging to the class 3, very regular and compact (from top left to bottom right): Piazza dell'Anfiteatro, Lucca (0,522); Piazza Santa Caterina, Vigo, Savona (0,554); Piazza del Mercato Vecchio, Fasano, Brindisi (0,632); Piazza del Mulino Garibaldi, San Giuseppe Jato, Palermo (0,702); Piazza dei Sanniti, Roma (0,706); Piazza Nicolò Tommaseo, Montebelluna, Treviso (0,720); Piazza Sei Novembre, Carrabba, Catania (0,766); Piazza Circolare Giotto, Taurisano, Lecce (0,782).

Applying the exposed parameters thresholds (Table 7), the resulting cluster assume morpho-typological characterizations, as shown in Figure 19, which allow the identification of the BETs and their physical description exposed in the following sections.

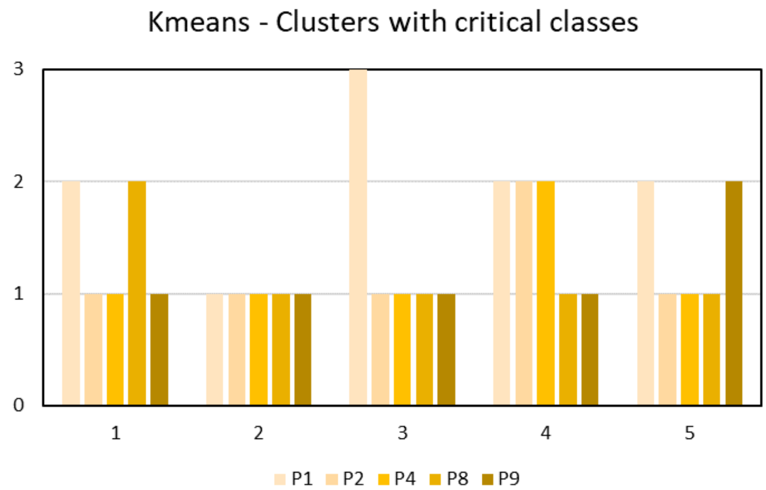


Figure 19. Cluster characterizations through the threshold values of the parameters.

3.3.2 BET description

Considering In blue, the values that deviate from the global means and represent the main characterization of each cluster.

Table 6, the values of the variables in each cluster are generally consistent with the overall means. Some specific variables diverge from the general means values, characterizing the clusters (see Figure 19 and Table 7). Boxplots for parameters values in each cluster are shown in Figure 20.

	1	2	3	4	5	Total
P1	0,310	0,231	0,509	0,395	0,367	**0,362**
P2	0,882	0,830	0,538	1,308	0,692	**0,793**
P4	0,016	0,014	0,017	0,046	0,017	**0,020**
P8	9,037	1,976	2,038	2,257	2,930	**3,015**
P9	0,013	0,015	0,011	0,012	0,418	**0,036**
cases	146	391	369	141	64	1111

Cluster 1: OS with a medium level of compactness and regularity of the morphology, without problems of overturning of the fronts, with a critical ratio between number of accesses and perimeter, on sloping ground or with changes in elevation, without green areas.
The values that diverge most from the means are those related to the morphology of the OS (P8; 9,037). This cluster is therefore mainly characterized by OS with sloping ground or with changes in elevation.

Cluster 2: OS with a low level of compactness and regularity of the morphology, without problems of overturning of the fronts, with a critical ratio between number of accesses and perimeter, on flat or slightly sloping ground, without green areas.
The values that diverge most from the means are those related to the morphology of the OS (P1; 0,231). This cluster is therefore mainly characterized by OS with low level of compactness and regularity of the shape.

Cluster 3: OS with a high level of compactness and regularity of the morphology, without problems of overturning of the fronts, with a critical ratio between number of accesses and perimeter, on flat or slightly sloping ground, without green areas.
The values that diverge most from the means are those related to the morphology of the OS (P1; 0,509). This cluster is therefore mainly characterized by OS with high level of compactness and regularity of the shape.

Cluster 4: OS with a medium level of compactness and regularity of the morphology, with problems of overturning of the fronts, without critical ratio between number of accesses and perimeter, on sloping ground or with changes in elevation, without green areas.
The values that diverge most from the means are those related to the numbers of access (P4; 1,308) and the ratio between height of the fronts and width of the OS (P2; 0,046). This cluster is therefore mainly characterized by OS with problems of overturning of the fronts, but with a suitable ratio between number of accesses and perimeter.

Cluster 5: OS with a medium level of compactness and regularity of the morphology, without problems of overturning of the fronts, with critical ratio between number of accesses and perimeter, on sloping ground or with changes in elevation, with green areas.
The value that diverges most from the means is P9 (0,418), indicating the presence of green areas. This cluster is therefore mainly characterized by OS with green areas.

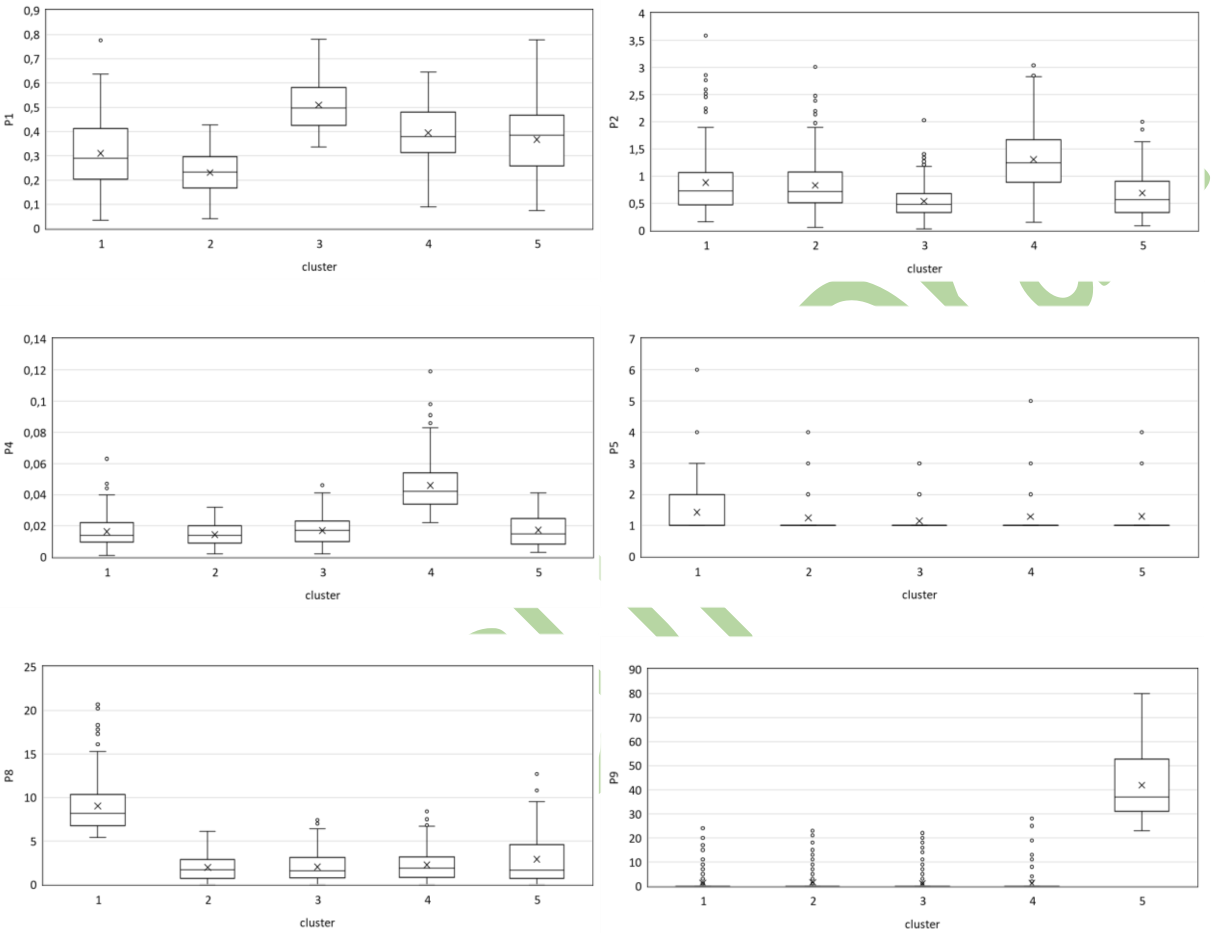


Figure 20. Boxplots of parameters values in single clusters (X mean value; ------ median value).

3.3.3 Supplementary variables

Once the claster analysis was completed using the active variables (P1, P2, P4, P8, P9), the defined clusters were explored and further characterized by analyzing the previously excluded supplementary variable (P5). A further analysis was also performed regarding the relationship between the height of the fronts and the width of the open space. With parameter P2 only the relationship between the maximum height of the built fronts and the width of the square was investigated, while here it was possible to deepen the investigation and compare the relationship between the median height of the fronts and the width of open space.

Parameter P5, which is related to the presence of special buildings in the analyzed open space, is significative for the definition of the BETs. By analyzing the parameter P5 in the individual clusters, it was possible to identify a further subdivision for some clusters. Cluster 1, 2 and 4 showed significant percentages of OS with the presence of special buildings (respectively 49% for cluster 1, 39% for cluster 2 and 47% for cluster 4) (Figure 21). Given that the absence of special buildings could be linked to a lack of data in the OSM database, or to an incorrect compilation of the same, rather than to an actual lack of special buildings, a splitting of the clusters was reasonably evaluated for the further definition of BETs.

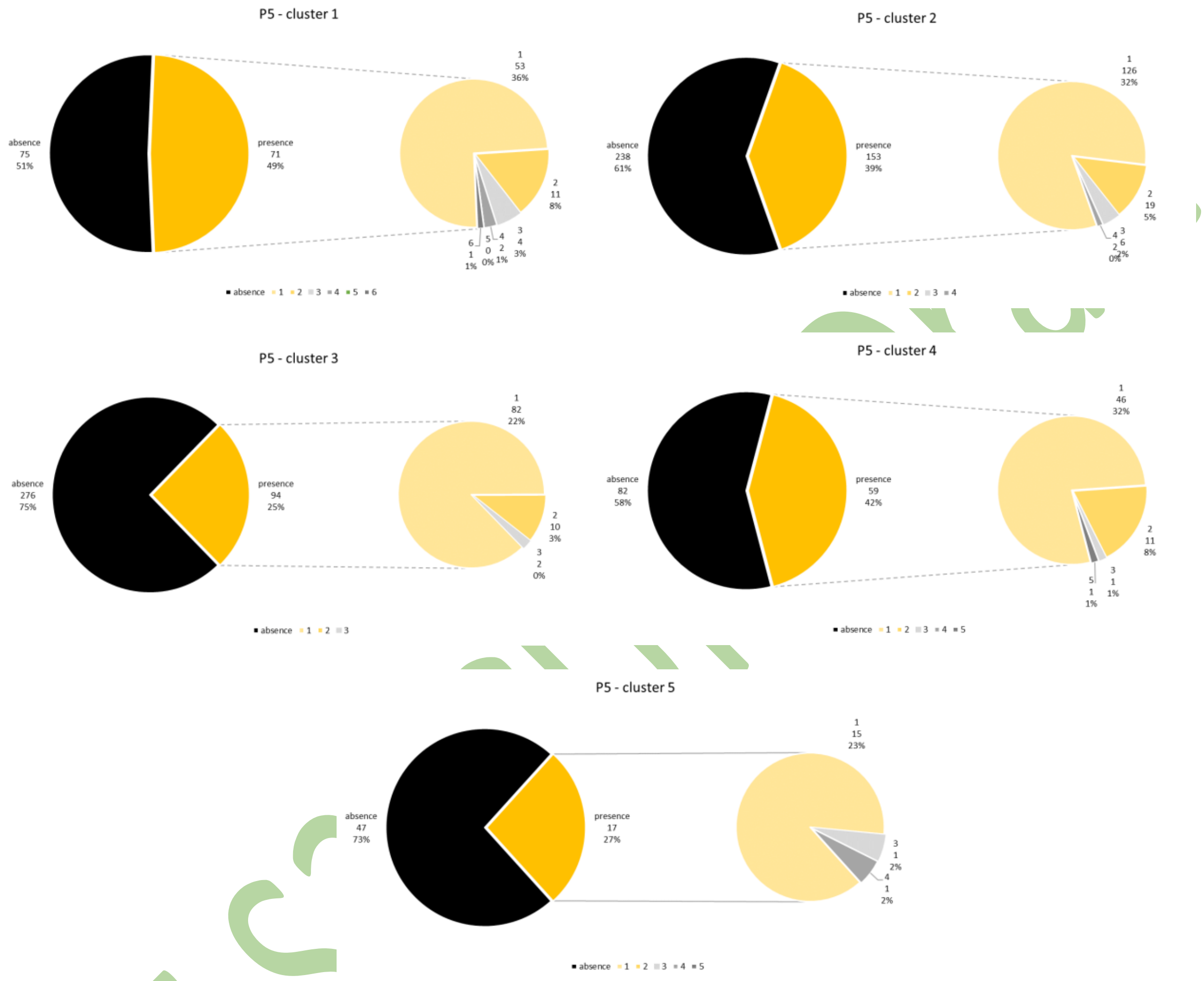


Figure 21. Analysis of supplementary variable (P5) within clusters from k-means algorithm.

The analysis of the relationship between the height of the built fronts and the width of the open space did not produce significant results except for cluster 4, already identified in the cluster analysis by a high value of the parameter P2. As a consequence of this result, the values of the cases with the presence of special buildings and the critical heights of the built fronts were interpolated. Figure 22 shows the descriptive graphs of this survey, where on the left are the critical height values (if related to the value of Hmax or Hmedian) and on the left the same analysis repeated on the subsample of cluster P4 which presents special buildings. The results thus obtained led us to the creation of 3 BETs deriving from cluster 4: 4a includes at least one special building with problems of overturning of the fronts caused by H max; 4b includes at least one special building with problems of overturning of the fronts caused by H median, 4c has not special buildings, but has problems of overturning of the fronts caused by H median.

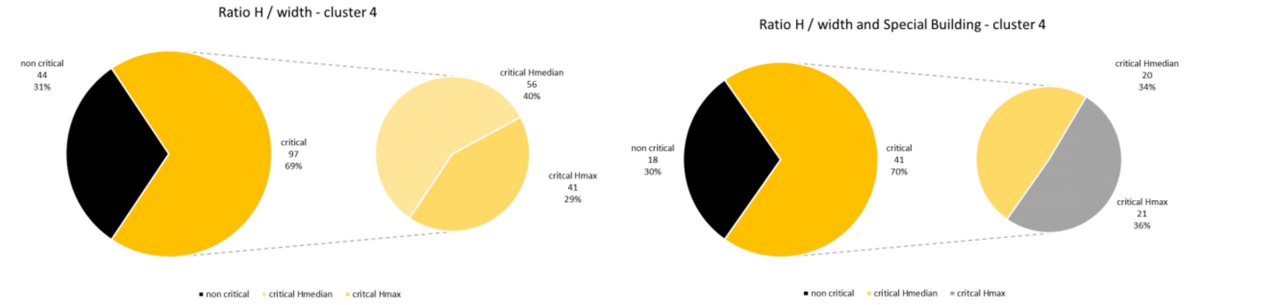


Figure 22. Analysis on cluster 4 regarding the critical Heigh related to the width of open spaces (on the left), and the same analysis for the subsample of cluster 4 with special buildings (P5).

3.3.4 BET representation

As a result of the critical analysis presented, the identified BETs are 9. The diagram in Figure 23 describes their definition starting from the identified clusters. Below are the representation and definition sheets of the individual BETs, accompanied by geometric descriptions.

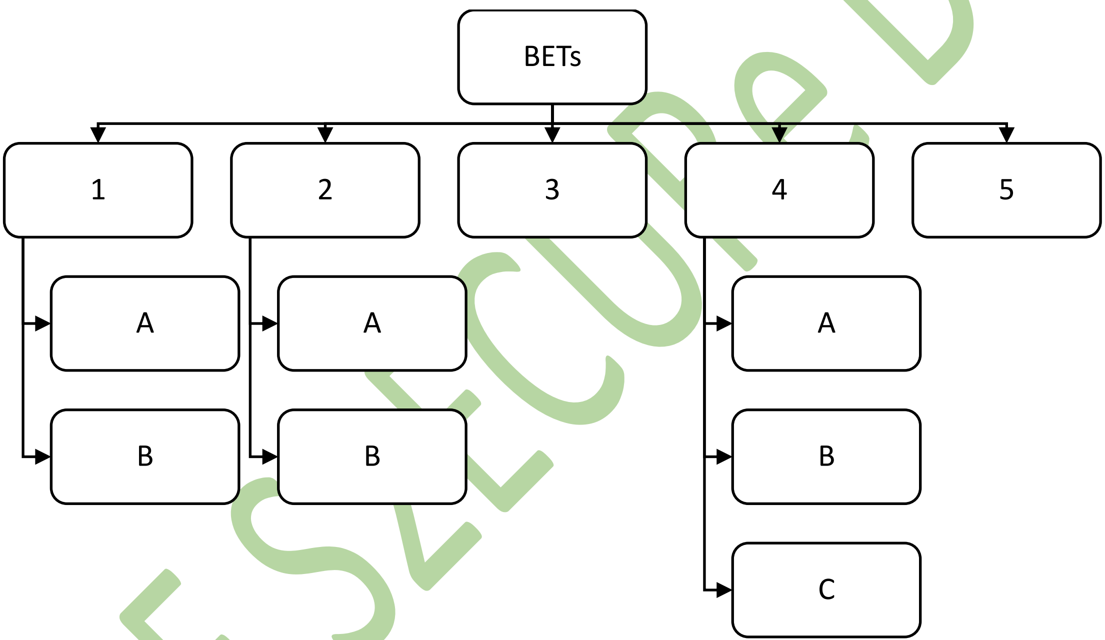
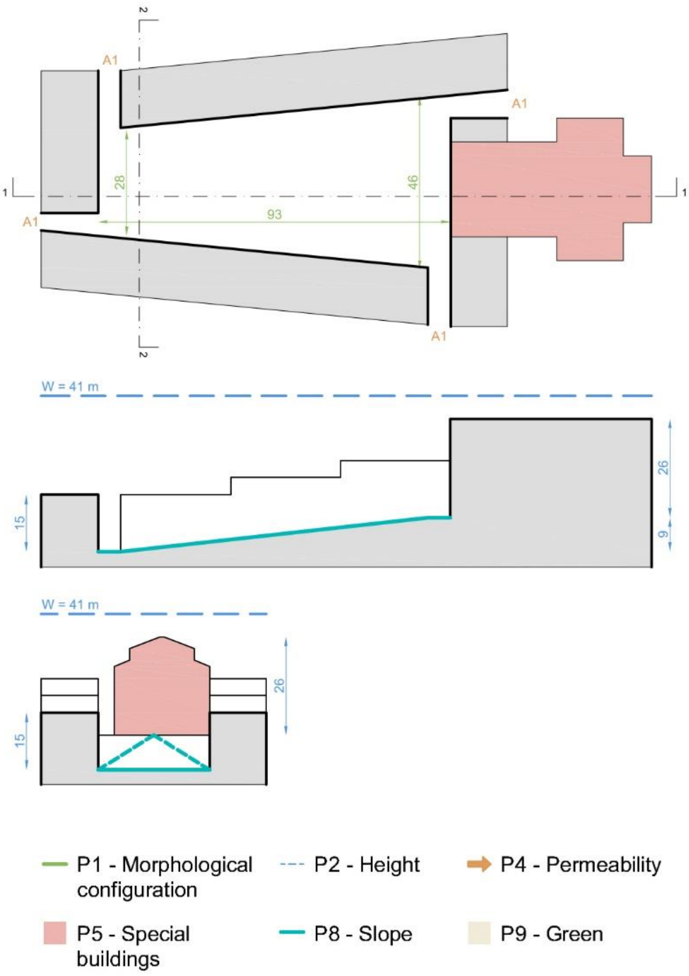
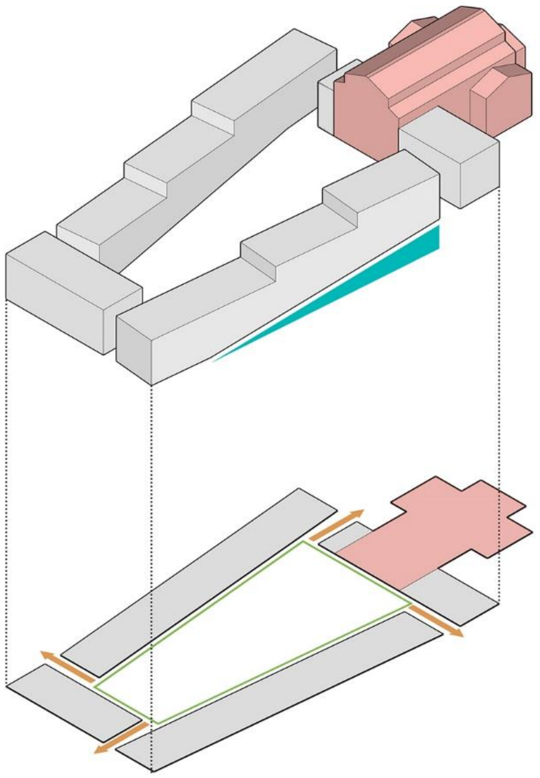


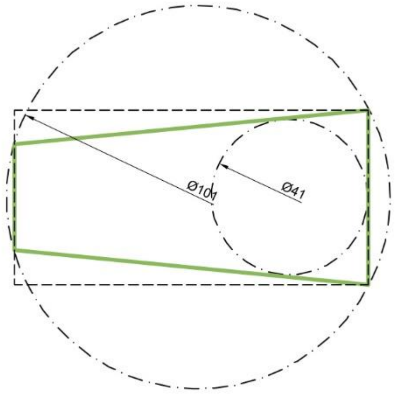
Figure 23. Diagram of BETs definition starting from the identified clusters.

BET 1A

AS with a medium level of compactness and regularity of the morphology, without problems of overturning of the fronts, with a critical ratio between number of accesses and perimeter, on sloping ground or with changes in elevation, without green areas. The AS includes at least one special building.





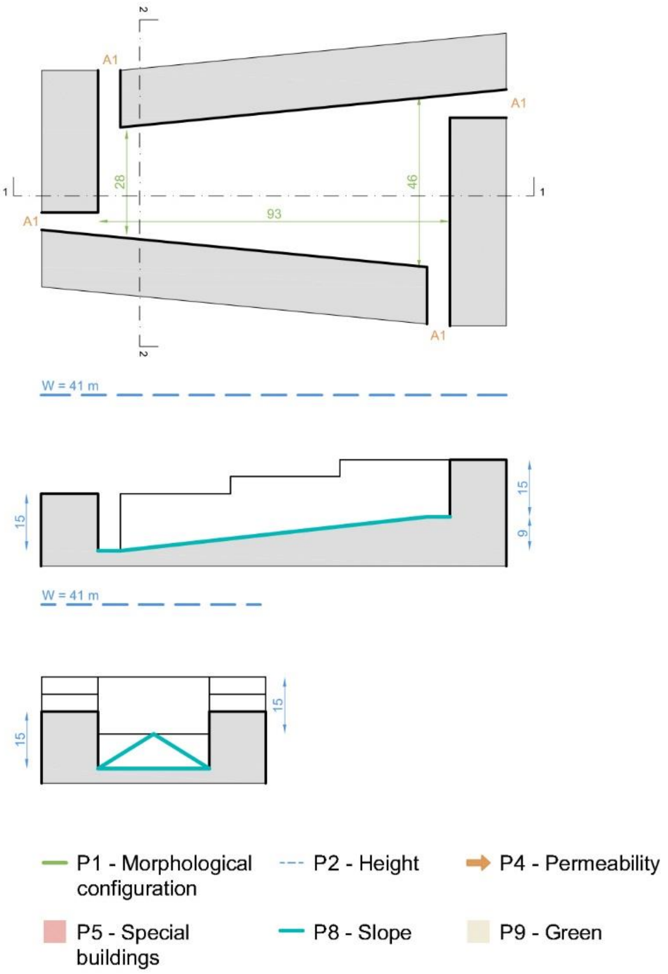


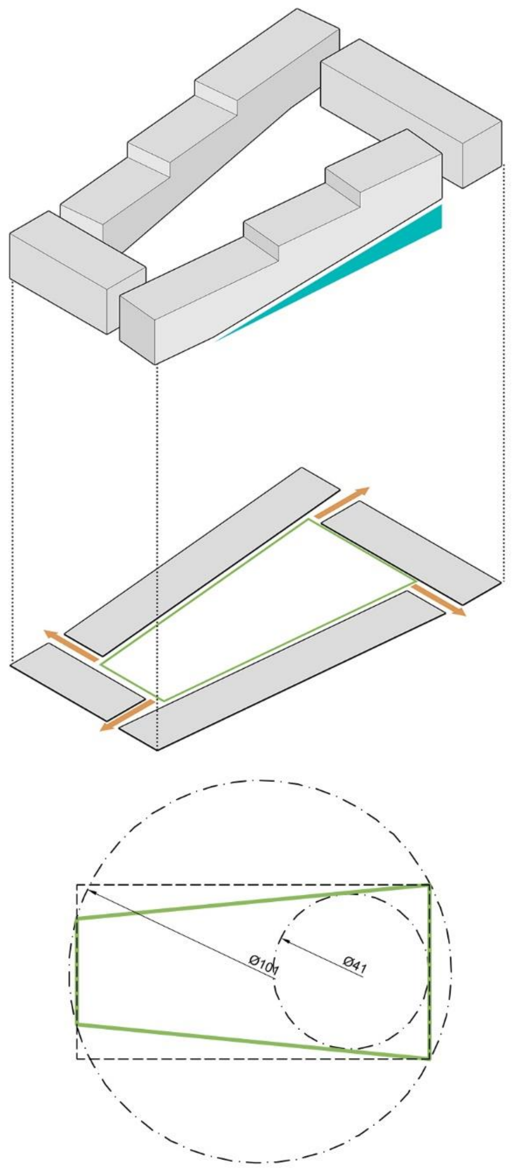
P1	Value	Range
D1 [m]	101	53 ÷ 127
D2 [m]	41	22 ÷ 56
D2/D1	0,41	0,41 ÷ 0,44
Area real [mq]	3.441	925 ÷ 4.940
Area BB [mq]	4.278	1.461 ÷ 7.299
Area real/Area BB	0,80	0,57 ÷ 0,81
P2		
H mean [m]	15	12 ÷ 19
H max (SB) [m]	-	-

P4	Value	Range
N. accesses [#]	3,8	2 ÷ 5
A1 n. [#] / dim. [m]	4 / 6	-
A2 n. [#] / dim. [m]	0 / 4	-
P5		
Special building [#]	0	1 ÷ 2
P8		
Δ HLS [m]	9	6,8 ÷ 10,3
P9		
Green area [%]	0	0

BET 1B

AS with a medium level of compactness and regularity of the morphology, without problems of overturning of the fronts, with a critical ratio between number of accesses and perimeter, on sloping ground or with changes in elevation, without green areas. The AS does not include any special building.



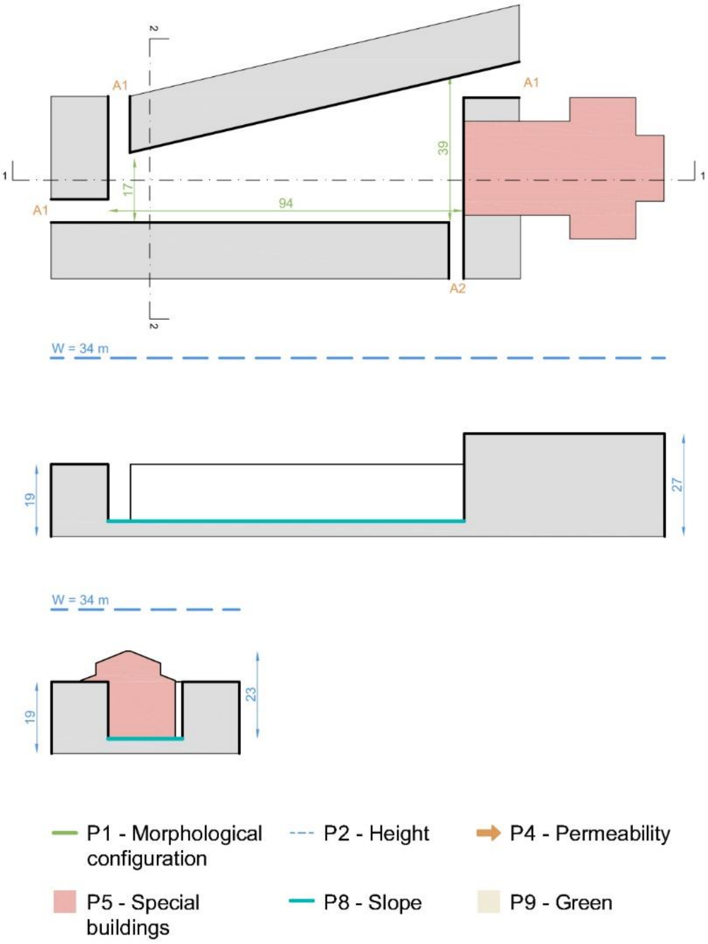


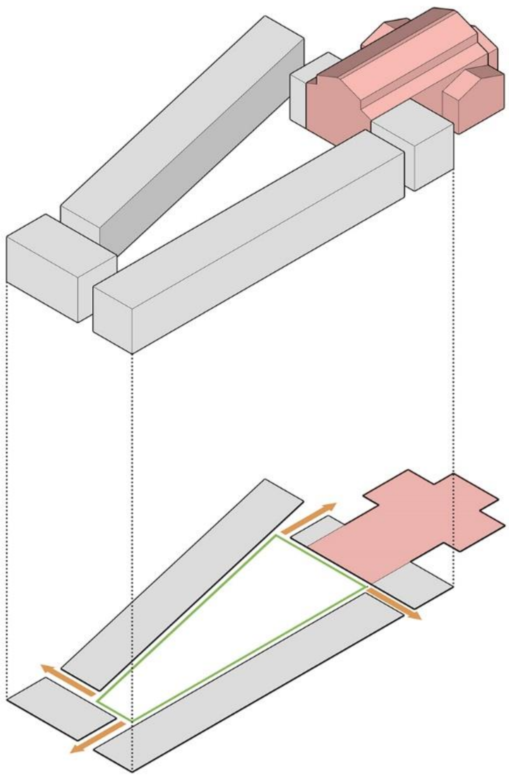
P1	Value	Range
D1 [m]	102	62 ÷ 127
D2 [m]	34	21 ÷ 42
D2/D1	0,33	0,33
Area real [mq]	2.632	1.022 ÷ 4.038
Area BB [mq]	3.666	1.534 ÷ 6.287
Area real/Area BB	0,72	0,55 ÷ 0,75
P2		
H mean [m]	15	10 ÷ 19
H max (SB) [m]	23	16 ÷ 28

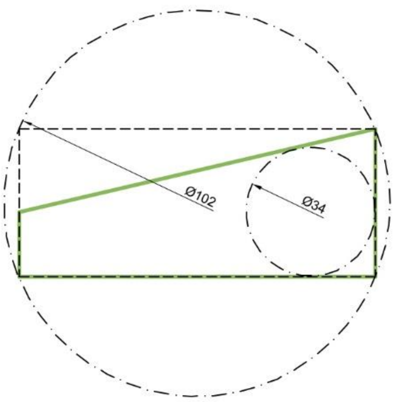
P4	Value	Range
N. accesses [#]	3,6	2 ÷ 5
A1 n. [#] / dim. [m]	3 / 6	-
A2 n. [#] / dim. [m]	1 / 4	-
P5		
Special building [#]	1	1
P8		
Δ HLS [m]	0	0,7 ÷ 2,9
P9		
Green area [%]	0	0

BET 2A

AS with a low level of compactness and regularity of the morphology, without problems of overturning of the fronts, with a critical ratio between number of accesses and perimeter, on flat or slightly sloping ground, without green areas. The AS includes at least one special building.





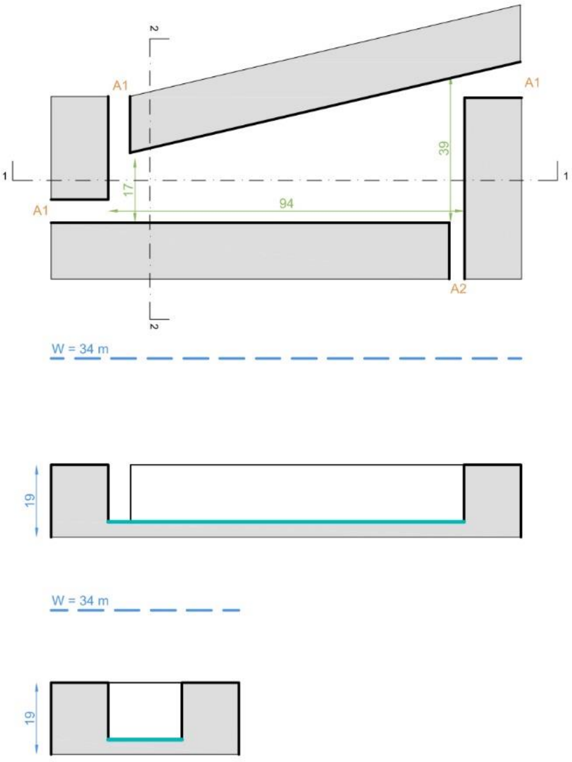


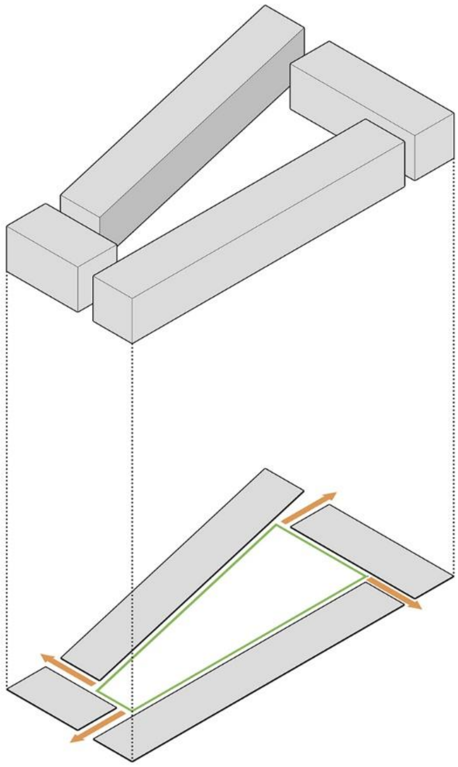
P1	Value	Range
D1 [m]	102	62 ÷ 127
D2 [m]	34	21 ÷ 42
D2/D1	0,33	0,33
Area real [mq]	2.632	1.022 ÷ 4.038
Area BB [mq]	3.666	1.534 ÷ 6.287
Area real/Area BB	0,72	0,55 ÷ 0,75
P2		
H mean [m]	15	10 ÷ 19
H max (SB) [m]	23	16 ÷ 28

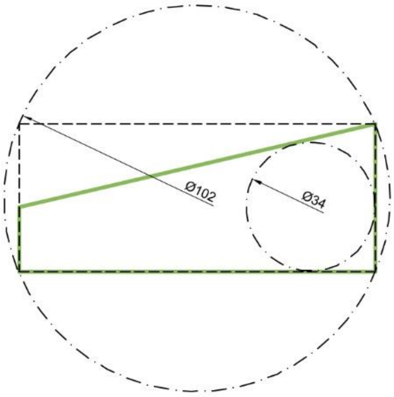
P4	Value	Range
N. accesses [#]	3,6	2 ÷ 5
A1 n. [#] / dim. [m]	3 / 6	-
A2 n. [#] / dim. [m]	1 / 4	-
P5		
Special building [#]	1	1
P8		
Δ HLS [m]	0	0,7 ÷ 2,9
P9		
Green area [%]	0	0

BET 2B

AS with a low level of compactness and regularity of the morphology, without problems of overturning of the fronts, with a critical ratio between number of accesses and perimeter, on flat or slightly sloping ground, without green areas. The AS does not include any special building.







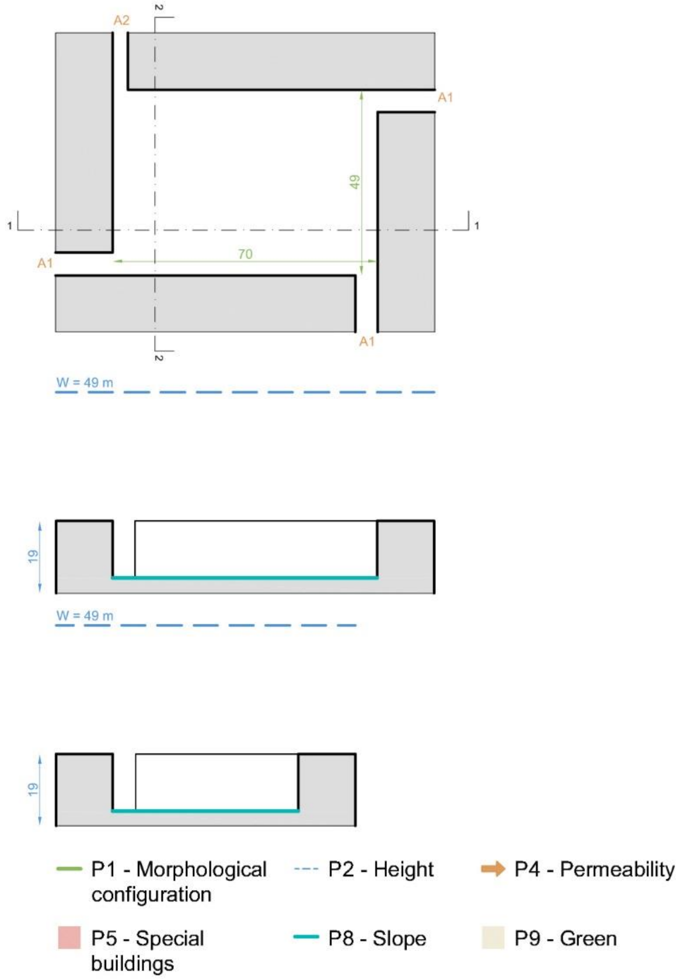
P1 - Morphological configuration — P2 - Height — P4 - Permeability
P5 - Special buildings — P8 - Slope — P9 - Green

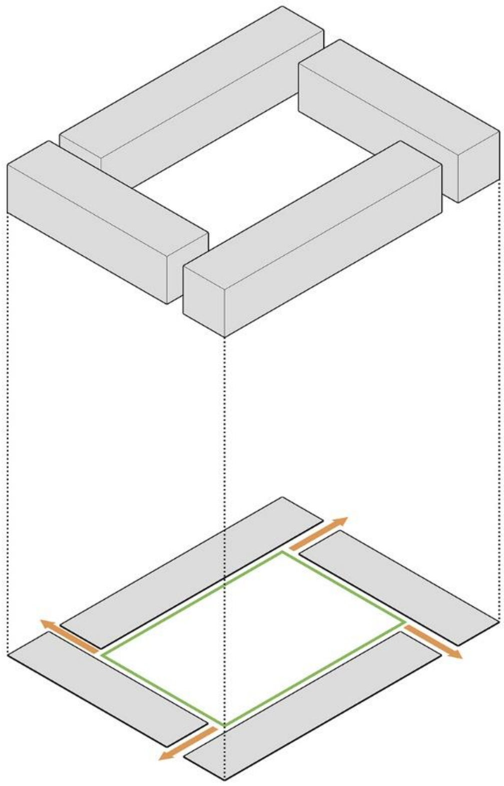
P1	Value	Range
D1 [m]	102	62 ÷ 127
D2 [m]	34	21 ÷ 42
D2/D1	0,33	0,33
Area real [mq]	2.632	1.022 ÷ 4.038
Area BB [mq]	3.666	1.534 ÷ 6.287
Area real/Area BB	0,72	0,55 ÷ 0,75
P2		
H mean [m]	15	10 ÷ 19
H max (SB) [m]	-	-

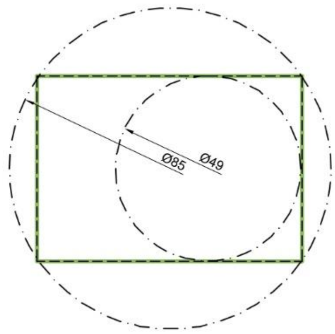
P4	Value	Range
N. accesses [#]	3,6	2 ÷ 5
A1 n. [#] / dim. [m]	3 / 6	-
A2 n. [#] / dim. [m]	1 / 4	-
P5		
Special building [#]	0	1
P8		
Δ HLS [m]	0	0,7 ÷ 2,9
P9		
Green area [%]	0	0

BET 3

AS with a high level of compactness and regularity of the morphology, without problems of overturning of the fronts, with a critical ratio between number of accesses and perimeter, on flat or slightly sloping ground, without green areas. The AS does not include any special building.





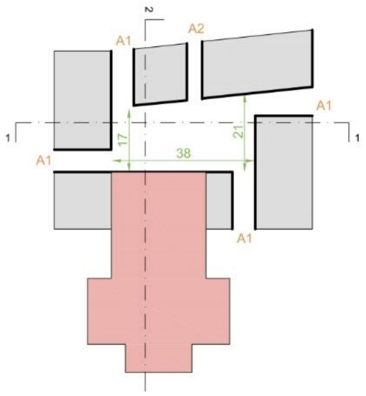


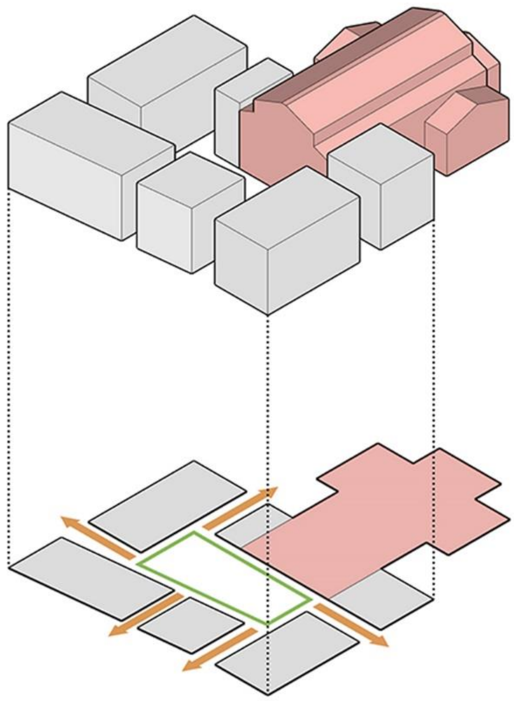
P1	Value	Range
D1 [m]	85	49 ÷ 105
D2 [m]	49	30 ÷ 60
D2/D1	0,56	0,52 ÷ 0,65
Area real [mq]	3.430	1.167 ÷ 4.738
Area BB [mq]	3.430	1.311 ÷ 5.441
Area real/Area BB	1	0,79 ÷ 0,95
P2		
H mean [m]	15	11 ÷ 19
H max (SB) [m]	-	-

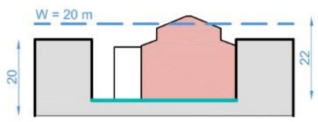
P4	Value	Range
N. accesses [#]	3,5	2 ÷ 4
A1 n. [#] / dim. [m]	3 / 6	-
A2 n. [#] / dim. [m]	1 / 4	-
P5		
Special building [#]	0	1
P8		
Δ HLS [m]	0	0,8 ÷ 3,2
P9		
Green area [%]	0	0

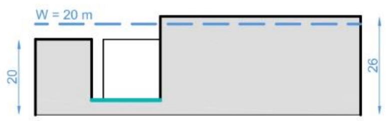
BET 4a

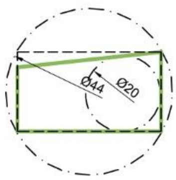
AS with a medium level of compactness and regularity of the morphology, with problems of overturning of the fronts caused by H max, without critical ratio between number of accesses and perimeter, on sloping ground or with changes in elevation, without green areas. The AS includes at least one special building.











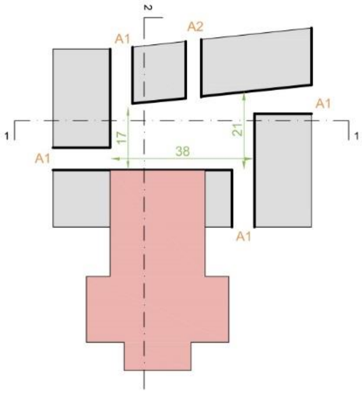
P1 - Morphological configuration
P2 - Height
P4 - Permeability
P5 - Special buildings
P8 - Slope
P9 - Green

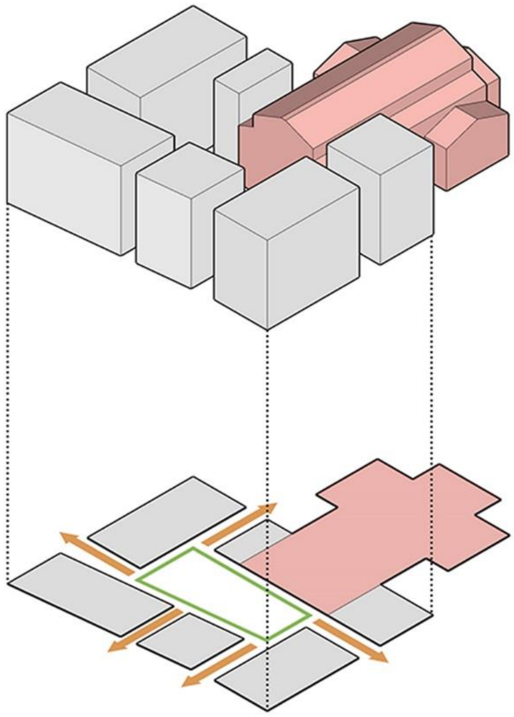
P1	Value	Range
D1 [m]	44	28 ÷ 50
D2 [m]	20	13 ÷ 24
D2/D1	0,46	0,42 ÷ 0,58
Area real [mq]	720	291 ÷ 775
Area BB [mq]	798	366 ÷ 1.127
Area real/Area BB	0,9	0,72 ÷ 0,90
P2		
H mean [m]	16	12 ÷ 19
H max (SB) [m]	22	18 ÷ 29

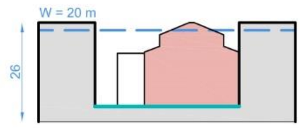
P4	Value	Range
N. accesses [#]	4,6	2 ÷ 4
A1 n. [#] / dim. [m]	4 / 6	-
A2 n. [#] / dim. [m]	1 / 4	-
P5		
Special building [#]	1	1
P8		
Δ HLS [m]	0	0,9 ÷ 3,2
P9		
Green area [%]	0	0

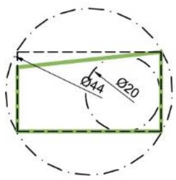
BET 4b

AS with a medium level of compactness and regularity of the morphology, with problems of overturning of the fronts caused by H median, without critical ratio between number of accesses and perimeter, on sloping ground or with changes in elevation, without green areas. The AS includes at least one special building.









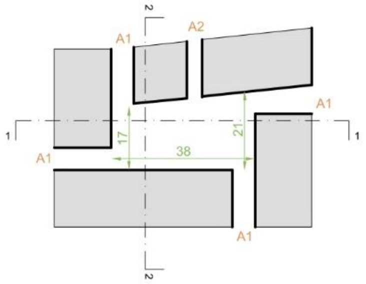
P1 - Morphological configuration
P2 - Height
P4 - Permeability
P5 - Special buildings
P8 - Slope
P9 - Green

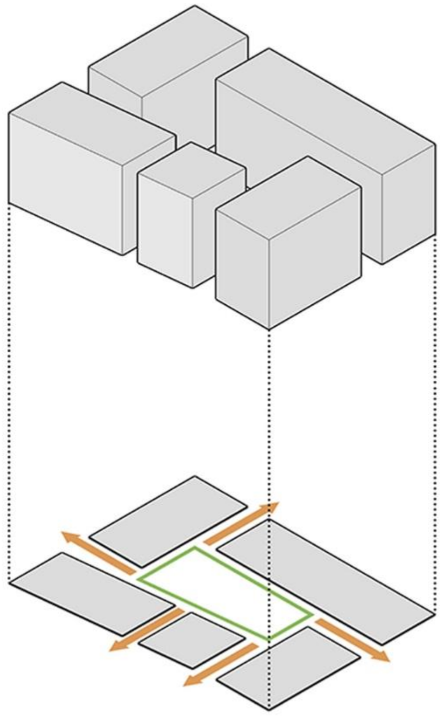
P1	Value	Range
D1 [m]	44	28 ÷ 50
D2 [m]	20	13 ÷ 24
D2/D1	0,46	0,42 ÷ 0,58
Area real [mq]	720	291 ÷ 775
Area BB [mq]	798	366 ÷ 1.127
Area real/Area BB	0,9	0,72 ÷ 0,90
P2		
H mean [m]	22	18 ÷ 29
H max (SB) [m]	22	18 ÷ 29

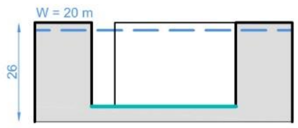
P4	Value	Range
N. accesses [#]	4,6	2 ÷ 4
A1 n. [#] / dim. [m]	4 / 6	-
A2 n. [#] / dim. [m]	1 / 4	-
P5		
Special building [#]	1	1
P8		
Δ HLS [m]	0	0,9 ÷ 3,2
P9		
Green area [%]	0	0

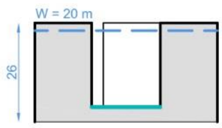
BET 4c

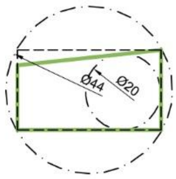
AS with a medium level of compactness and regularity of the morphology, with problems of overturning of the fronts caused by H median, without critical ratio between number of accesses and perimeter, on sloping ground or with changes in elevation, without green areas. The AS does not include any special building.











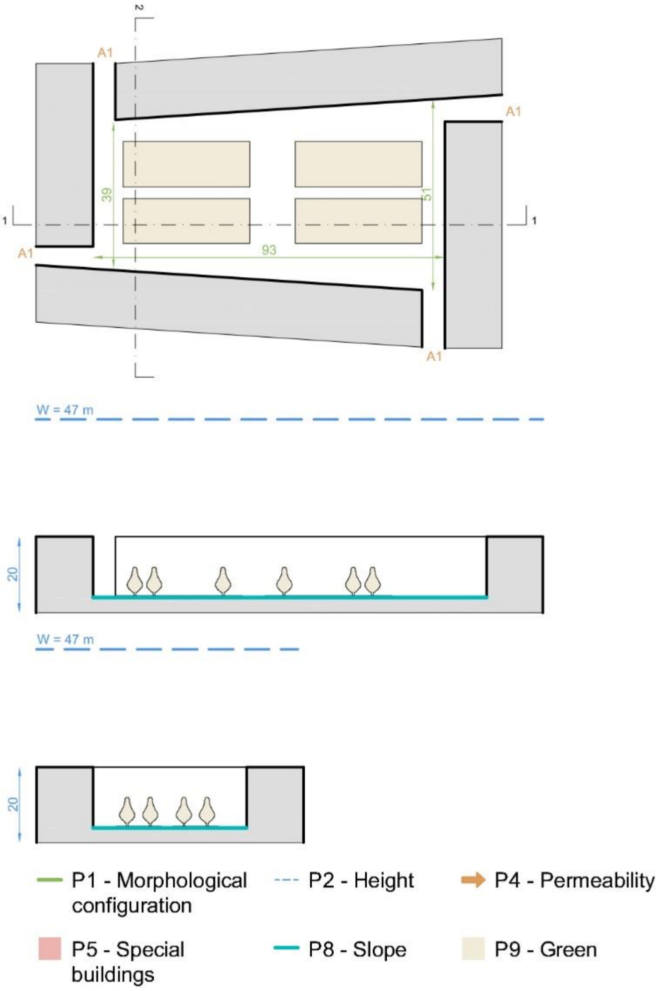
P1 - Morphological configuration — P2 - Height — P4 - Permeability
P5 - Special buildings — P8 - Slope — P9 - Green

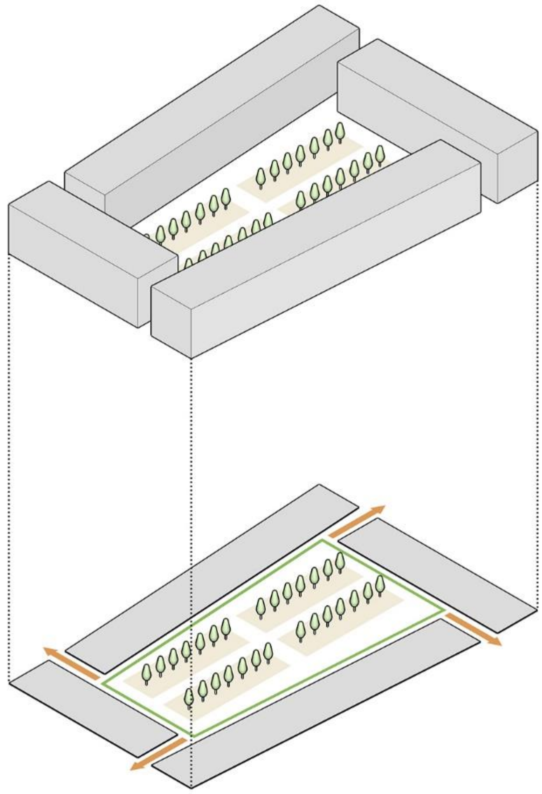
P1	Value	Range
D1 [m]	44	28 ÷ 50
D2 [m]	20	13 ÷ 24
D2/D1	0,46	0,42 ÷ 0,58
Area real [mq]	720	291 ÷ 775
Area BB [mq]	798	366 ÷ 1.127
Area real/Area BB	0,9	0,72 ÷ 0,90
P2		
H mean [m]	22	18 ÷ 29
H max (SB) [m]	-	-

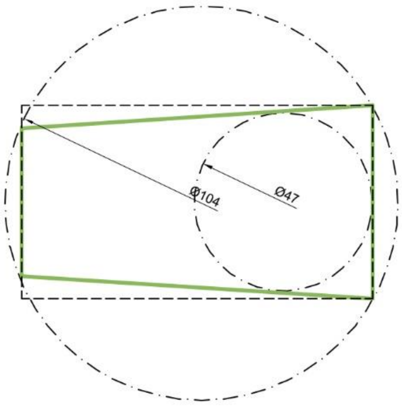
P4	Value	Range
N. accesses [#]	4,6	2 ÷ 4
A1 n. [#] / dim. [m]	4 / 6	-
A2 n. [#] / dim. [m]	1 / 4	-
P5		
Special building [#]	0	1
P8		
Δ HLS [m]	0	0,9 ÷ 3,2
P9		
Green area [%]	0	0

BET 5

AS with a medium level of compactness and regularity of the morphology, without problems of overturning of the fronts, with critical ratio between number of accesses and perimeter, on sloping ground or with changes in elevation, with green areas. The AS does not include any special building.





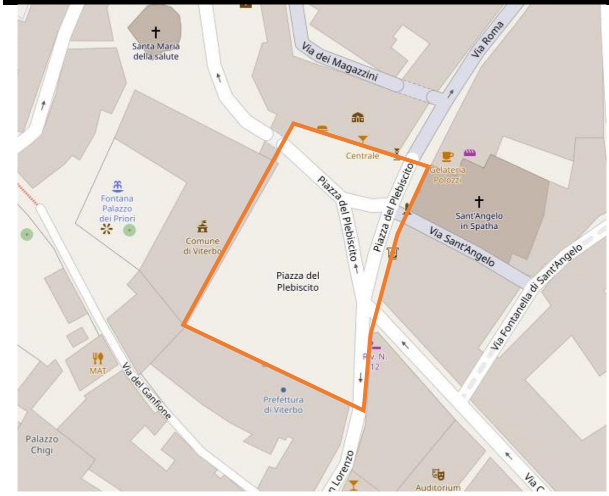


P1	Value	Range
D1 [m]	104	63 ÷ 134
D2 [m]	47	26 ÷ 57
D2/D1	0,45	0,38 ÷ 0,58
Area real [mq]	4.185	1.115 ÷ 5.210
Area BB [mq]	4.743	1.739 ÷ 7.453
Area real/Area BB	0,88	0,64 ÷ 0,90
P2		
H mean [m]	16	11 ÷ 21
H max (SB) [m]	-	-

P4	Value	Range
N. accesses [#]	4	2 ÷ 5
A1 n. [#] / dim. [m]	4 / 6	-
A2 n. [#] / dim. [m]	0 / 4	-
P5		
Special building [#]	0	1
P8		
Δ HLS [m]	0	0,7 ÷ 4,6
P9		
Green area [%]	40	31 ÷ 53

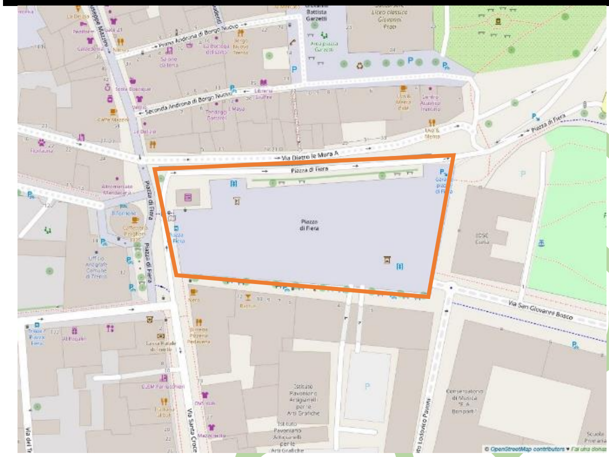
3.4 Example of BETs

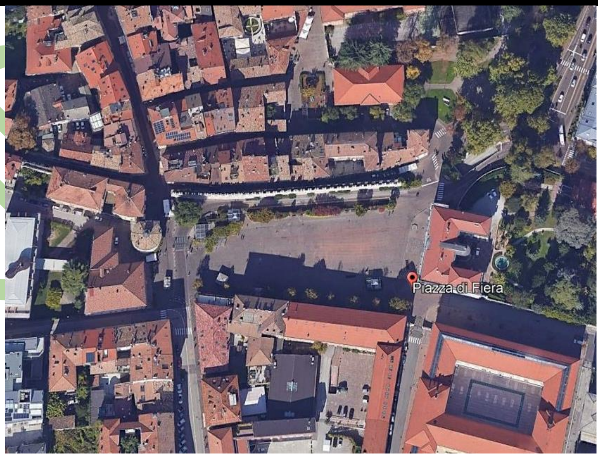
BET 1A



Piazza del Plebiscito, Viterbo (w33816241)

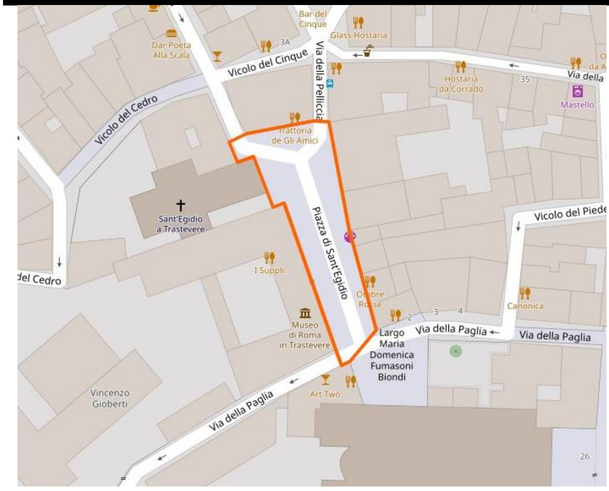
BET 1B





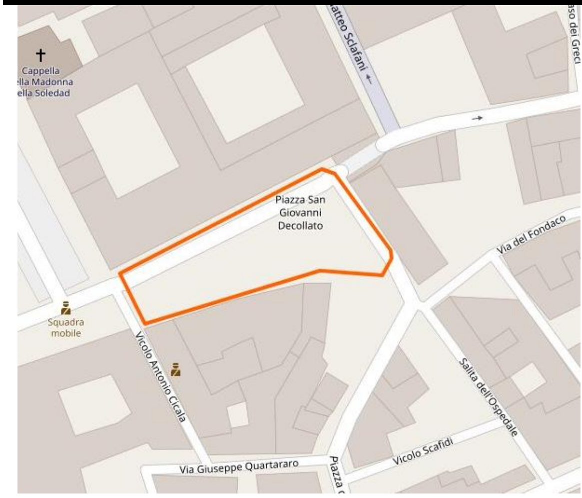
Piazza di Fiera, Trento (w26402532)

BET 2A



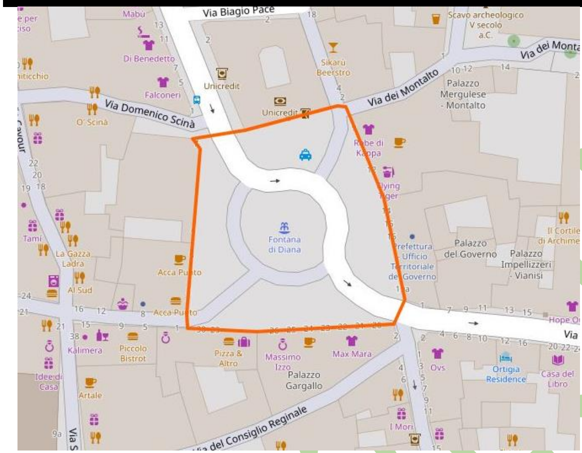
Piazza di Sant'Egidio, Roma (w203295672)

BET 2B



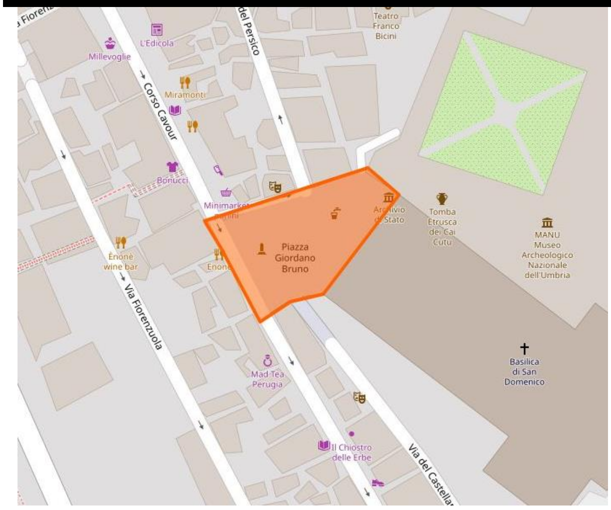
Piazza San Giovanni Decollato, Palermo (w86665670)

BET 3



Piazza Archimede, Siracusa (w173226095)

BET 4A

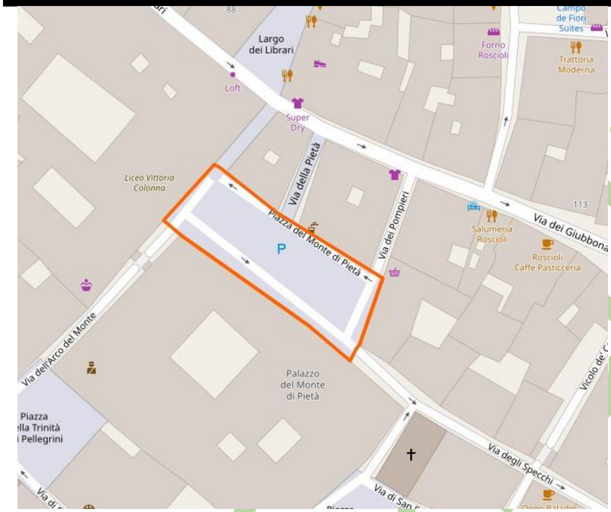


Piazza Giordano Bruno, Perugia (w456657168)

BET 4B

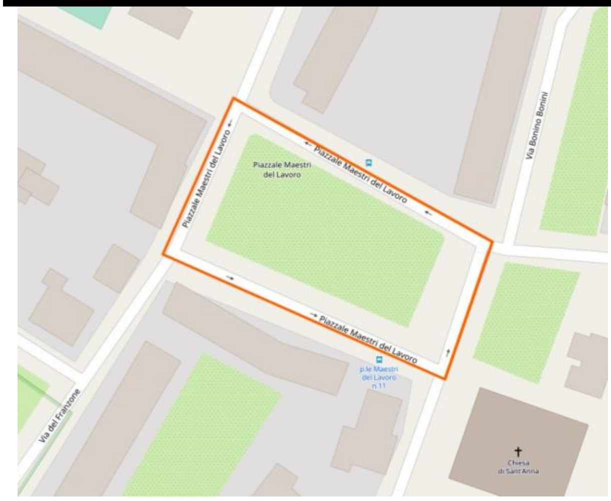
Campo Sant'Aponal, Venezia (w173196994)

BET 4C



Piazza del Monte di Pietà, Roma (w28867534)

BET 5



Piazzale Maestri del Lavoro, Brescia (w785362866)

3.5 Limitations

Although OpenStreetMap (OSM) is a valuable source of geographical data, some consideration about the limitation in term of data quality should be done. Since maps and data editors of OSM could have different level of experience and skills, OSM data could be vulnerable to errors and gaps. Quality issues of OSM database are well-known (Kaur et al. 2018), therefore using dataset extracted from OSM could be someway questionable. The main quality issues recognized include: geometric positional accuracy, completeness of the database, qualitative and quantitative information accuracy, topological consistency, and semantic accuracy.

In the context of this research some limitations and uncertainties could be point out, regarding the data availability and completeness, due to the "open" nature of the database, and the data accuracy, affected mostly by skills level of editors that improve the dataset, no top-down data quality assurance method, consistency of data implementation, and uncertainty about the data source. In particular, the authors highlight the incomplete data coverage regarded the P5 (special building) and P8 (green).

Some consideration should also be done about the data extraction procedures through queries and algorithms. Reducing queries is necessary to limit both the time require to elaborate the outcome and the not-suitable results, but it ended in a potential limitation of the findings (e.g. place=square). Although the algorithms performed allow an increase of the data extraction, they are subject to process miscalculations. In particular, in the present report the parameters P4 (access) was affected by the algorithms data extraction issues, due to the nature of the procedures used. In fact, in some cases the number of access associate with a street tangent to the square was not correctly count.

Cluster Analysis:

- number of cluster considered;

4. Conclusions

The identification of BE scenarios suitable for multi-risk assessment is a preliminary step towards the preparation of simulation models, with the aim to estimate the risk levels of OS in BE and the safety of its occupants. In this context, particular attention should be paid to the specific characteristics of Italian towns BE that affect SUOD (e.g., earthquake and terroristic attack) and SLOD (e.g., heatwave and air pollution) risks, which include morphological, geometrical, functional, and constructive aspects. Identifying potential scenarios characterized by recurring Italian BE features provides the opportunity to plan further significant analysis and simulations for risk levels description.

In this report, recurring typologies of BE (BET) describing multi-risk scenarios are identified through a cluster analysis based on data extracted from GIS databases. Six parameters (P1, morphology; P2, the height of the fronts; P4, number of accesses; P5, presence of special building; P8, the slope of the ground; and P9, presence of green area) characterizing the OS (e.g., squares) in BE of existing Italian towns have been selected to perform fast data extraction from GIS datasets, using proper queries and algorithms.

On the final database built by the authors, the analysis was carried out using five active variables (P1, P2, P4, P8, P9) and one supplementary variable (P5). The final dataset includes 1.113 case studies, then reduced to 1.111, after performing a hierarchical process based on the single-linkage method to find multivariate outliers.

Several algorithms to perform clustering were evaluated in this study. The k-means algorithm was chosen as the most appropriate for determining the solution. The results of the cluster analysis identified five groups of OSs, characterized by specific dimensional and functional characteristics. Combining the outcomes of the cluster analysis with an evaluation of the supplementary variable (P5) and a further insight on the height of the fronts (P2), nine final BETs were identified. To represent each BET, ranges of values were indicated for all parameters extracted from the dataset, selecting the interquartile ranges of the variables within each cluster considered (Q1 = 25% and Q3 = 75%).

Although the limitations in terms of extension and accuracy of open GIS data are recognized, the process proposed in this report enables a broad evaluation of the Italian BE characteristics, since the wide number of cases considered, reducing the time required to conventionally data collection. It is possible to replicate the process once the datasets considered have been increased, obtaining a deeper insight into the results. Moreover, limiting geographically the case studies, it would be possible to obtain the characterization of the BE of a specific Italian area. Finally, whether further datasets will be set describing additional aspects of the Italian BE (e.g., constructive aspects), the analysis could be repeated including more in-depth attributes of BE.

The multi-risk scenarios thus identified set the basis for future risk assessments of BE based on a statistical analysis of peculiar Italian town characteristics. The results can also provide a comparative assessment of the influence of individual features on the overall risk. Finally, such an approach could support the elaboration of specific actions for each case study, starting from the BET characteristics to which it belongs.

5. Abbreviations

AS - Areal Spaces

BE – Built Environment

BET – Built Environmental Typology

BIM – Building Information Model

GIS – Geographic Information System

LS - Linear Spaces

OS – Open Spaces

OSM – Open Street Maps

SLOD – Slow-onset disaster

SUOD - Sudden-onset disasters

6. References

Agafonkin V (2016) A new algorithm for finding a visual center of a polygon. In: Mapbox. https://blog.mapbox.com/a-new-algorithm-for-finding-a-visual-center-of-a-polygon-7c77e6492fbc. Accessed 4 Jan 2021

ArcGIS (2021) GIS Dictionary. https://support.esri.com/en/other-resources/gis-dictionary

Arosio M, Martina MLV, Figueiredo R (2020) The whole is greater than the sum of its parts: A holistic graph-based assessment approach for natural hazard risk of complex systems. Nat Hazards Earth Syst Sci 20:521–547. https://doi.org/10.5194/nhess-20-521-2020

Bernardini G, Romano G, Soldini L, Quagliarini E (2021) How urban layout and pedestrian evacuation behaviours can influence flood risk assessment in riverine historic built environments. Sustain Cities Soc 70:102876. https://doi.org/10.1016/j.scs.2021.102876

Dilley M (2005) Natural disaster hotspots a global risk analysis

Dujardin K, Defebvre L, Duhamel A, et al (2004) Cognitive and SPECT characteristics predict progression of Parkinson’s disease in newly diagnosed patients. J Neurol 251:1383–1392. https://doi.org/10.1007/s00415-004-0549-2

Dunant A, Bebbington M, Davies T, Horton P (2021) Multihazards Scenario Generator: A Network-Based Simulation of Natural Disasters. Risk Anal 0: https://doi.org/10.1111/risa.13723

Gallina V, Torresan S, Critto A, et al (2016) A review of multi-risk methodologies for natural hazards: Consequences and challenges for a climate change impact assessment. J Environ Manage 168:123–132. https://doi.org/10.1016/j.jenvman.2015.11.011

Garcia-Castellanos D, Lombardo U (2007) Poles of inaccessibility: A calculation algorithm for the remotest places on earth. In: Scottish Geographical Journal. Routledge , pp 227–233

Gill JC, Malamud BD (2016) Hazard interactions and interaction networks (cascades) within multi-hazard methodologies. Earth Syst Dyn 7:659–679. https://doi.org/10.5194/esd-7-659-2016

Halkidi M (2001) On Clustering Validation Techniques - Springer. 107–145

IBM Corp. (2019) IBM SPSS Statistics for Windows. Version 26

IPCC (2012) MANAGING THE RISKS OF EXTREME EVENTS AND DISASTERS TO ADVANCE CLIMATE CHANGE ADAPTATION. Cambridge University Press, Cambridge

James and others M (1967) Some methods for classification and analysis of multivariate observations. Proc fifth Berkeley Symp Math Stat Probab 1:281–297

Kappes MS, Keiler M, von Elverfeldt K, Glade T (2012) Challenges of analyzing multi-hazard risk: A review. Nat. Hazards 64:1925–1958

Kaur J, Singh J, Sehra SS, Rai HS (2018) Systematic literature review of data quality within openstreetmap. Proc - 2017 Int Conf Next Gener Comput Inf Syst ICNGCIS 2017 159–163. https://doi.org/10.1109/ICNGCIS.2017.35

Komendantova N, Mrzyglocki R, Mignan A, et al (2014) Multi-hazard and multi-risk decision-support tools as a part of participatory risk governance: Feedback from civil protection stakeholders. Int J Disaster

Risk Reduct 8:50–67. https://doi.org/10.1016/j.ijdrr.2013.12.006

Lee KM, Herrman TJ, Lingenfelser J, Jackson DS (2005) Classification and prediction of maize hardness-associated properties using multivariate statistical analyses. J Cereal Sci 41:85–93. https://doi.org/10.1016/j.jcs.2004.09.006

Mandolesi E, Ferrero A (2001) Piazze del Piceno. Gangemi

Mignan A, Wiemer S, Giardini D (2014) The quantification of low-probability-high-consequences events: Part I. A generic multi-risk approach. Nat Hazards 73:1999–2022. https://doi.org/10.1007/s11069-014-1178-4

Mooney P, Minghini M (2017) A Review of OpenStreetMap Data. In: Foody G, See L, Fritz S, et al. (eds) Mapping and the Citizen Sensor. Ubiquity Press, London, pp 37–59

Paliaga G, Faccini F, Luino F, et al (2020) A clustering classification of catchment anthropogenic modification and relationships with floods. Sci Total Environ 740:. https://doi.org/10.1016/j.scitotenv.2020.139915

PreventionWeb - UNDRR https://www.preventionweb.net/terminology#D

Quagliarini E, Lucesoli M, Bernardini G (2021) How to create seismic risk scenarios in historic built environment using rapid data collection and managing. J Cult Herit 48:93–105. https://doi.org/10.1016/j.culher.2020.12.007

Russo M, Angelosanti M, Bernardini G, et al (2020) Morphological systems of open spaces in built environment prone to Sudden-onset disasters. In: International Conference on Sustainability in Energy and Buildings SEB 2020. Split, Croatia, pp 1–10

Satari SZ, Muhammad Di NF, Zakaria R (2017) The multiple outliers detection using agglomerative hierarchical methods in circular regression model. J Phys Conf Ser 890:. https://doi.org/10.1088/1742-6596/890/1/012152

Schmidt J, Matcham I, Reese S, et al (2011) Quantitative multi-risk analysis for natural hazards: A framework for multi-risk modelling. Nat Hazards 58:1169–1192. https://doi.org/10.1007/s11069-011-9721-z

Scolobig A, Garcia-aristizabal A, Komendantova N, et al (2015) From Multi-Risk Assessment to Multi-Risk Governance: Recommendations for Future Directions. In: International Bank for reconstruction and Development (ed) UNDERSTANDING RISK, The Evolution of Disaster Risk Assessment. Washington DC, pp 163–167

Tate NJ, Fisher PF, Martin DJ (2008) Geographic Information Systems and Surfaces. Handb Geogr Inf Sci 239–258. https://doi.org/10.1002/9780470690819.ch13

UNISDR (2015) Sendai framework for disaster risk reduction 2015-2030

White GF, Kates RW, Burton I (2001) Knowing better and losing even more: The use of knowledge in hazards management. Environ Hazards 3:81–92. https://doi.org/10.1016/S1464-2867(01)00021-3

WHO (2014) Definitions: emergencies

Map data copyrighted OpenStreetMap contributors and available from https://www.openstreetmap.org