



Classifying buildings according to seismic vulnerability using Cluster-ANN techniques: application to the city of Murcia, Spain

J. Eduardo Meyers-Angulo¹ · Sandra Martínez-Cuevas¹ ·
Jorge M. Gaspar-Escribano¹

Received: 14 September 2022 / Accepted: 12 March 2023 / Published online: 3 April 2023
© The Author(s) 2023

Abstract

The seismic vulnerability of a city is a degree of its intrinsic susceptibility or predisposition to sustain damage or losses stemming from seismic events. In terms of physical vulnerability, one of the most important factors for assessing seismic risk, especially, for estimating losses, is the exposure of structures, particularly those structures intended for residential use. The present article outlines a methodology for classifying residential buildings based on the structural and non-structural components that ultimately determine the building typology and control the seismic performance. The proposed methodology is divided into three steps: first, spatial data are analysed using an official database that is supplemented by remote field work to verify, validate, and identify construction typologies and urban modifiers after incorporating the new observable data. During the second step, machine learning techniques based on Two-Step cluster analysis and neural networks are used to identify building typologies, using a multilayer perceptron to assess the representativeness of the building typologies identified. Finally, each building typology is defined, a vulnerability assessment is carried out, and vulnerability classes are ranked based on the macroseismic scale. The above-mentioned steps were applied to 7631 residential buildings in the city of Murcia, Spain. The methodology is scalable and may be automated, so it may be replicated in other urban areas with similar characteristics or adapted to different urban settings. This may help save time and reduce the cost of carrying out seismic risk studies, providing valuable information for both civil protection and regional and local governments.

Keywords Buildings classifications · Seismic vulnerability · Vulnerability index · Cluster analysis · Neural networks · Murcia Spain

✉ J. Eduardo Meyers-Angulo
eduardo.meyers.angulo@alumnos.upm.es

Sandra Martínez-Cuevas
sandra.mcuevas@upm.es

Jorge M. Gaspar-Escribano
jorge.gaspar@upm.es

¹ Universidad Politécnica de Madrid, Madrid, Spain

1 Introduction

The seismic vulnerability of a building or group of buildings reflects its intrinsic predisposition to sustaining damage in the event of a seismic motion and is directly correlated with its physical properties, structural design and soil-structure interaction (Barbat et al. 1998). In order to further a multidisciplinary and comprehensive approach to seismic risk assessment, we must first evaluate the potential physical damage that may result from the combination of hazard and physical vulnerability for exposed elements (Carreño 2006). Seismic risk assessment at the urban level must contemplate a large number of structures with their corresponding potential levels of damage and probabilities of sustaining said damage, while also bearing in mind that different buildings serve different purposes that are more or less relevant to city life (Basaglia et al. 2018). In this context, data quality and availability directly affect a realistic risk assessment, which serves as the basis for implementing informed disaster risk reduction actions, in line with the international agenda and the global targets set in the Sendai Framework, such as those that seek to make risk data and information more available to local communities and organizations (Torres et al. 2019).

The details of a building's construction make each building unique. In order to systematically characterize the different elements of the building, it is essential to record and update the data with the structural and non-structural attributes that can affect the seismic behaviour of the building. All these attributes are represented in an exposure model for vulnerability and seismic risk assessments, which can be adapted to a suitable taxonomy. This step poses the most significant challenges. For Dell'Acqua et al. (2013), the pitfalls of taking a detailed inventory of building stock include: (i) high investment in terms of time and money; (ii) access to data or public information that is incomplete or scattered across several entities; (iii) possible observation errors due to uncertainty; (iv) variable data formats or poor homogenization. Thus, the ongoing challenge of recording and enriching spatial data has led many researchers to propose methodologies based on primary data sources such as population and housing censuses, cadastres, and other official sources, combined with direct observations or computer-based surveys.

Examples of the above include the exposure model presented by Yepes-Estrada et al. (2017)—which, as part of the Global Earthquake Model's (GEM) South America Risk Assessment (SARA) Project, classifies typical building structures in seven South American countries based on population and housing census data—and the classification of structures in Central and South Asian countries presented by Lang et al. (2017), in which the authors managed to identify key building typologies based on data from rapid visual surveys. At the country level, the exposure model proposed by Santa María et al. (2017) provides a classification of residential building structures in Chile using census data as a primary source and identifying structures based on remote surveys. Another example of methodology, this time based on geomatics and statistical techniques, is the study by Torres et al. (2019) in which the researchers propose procedures for estimating seismic exposure and vulnerability by applying algorithms to remote-sensing data, thus providing accurate automation for these types of studies. Consequently, one current challenge is to adopt new automated methodologies for classifying building structures before assessing vulnerability.

Several global initiatives propose methods for assessing seismic vulnerability based on Building Typology Model (BTM) that correspond to predefined building classes with similar characteristics as far as their structural systems and seismic behaviour. Some of the most characteristic of these initiatives include: the taxonomy proposed by the Federal Emergency Management Agency (FEMA-154 1988, 2002), which contains 15 different

typological classes or BTMs based on building specifications typical in the United States; HAZUS (HAZUS-MH 2003), created by FEMA as a method for estimating seismic risk, with 15 BTMs subdivided according to the number of storeys; EMS-98 (Grünthal 1998), a European building classification based on empirical macroseismic intensity scales that identifies 15 BTMs based on the level of vulnerability in the event of a major earthquake; Risk-UE (Mouroux et al. 2004; Mouroux and Le Brun 2006), which integrates and homogenizes seismic risk projects in Europe, identifying 23 BTMs based on the most prevalent construction specifications; SYNER-G (2009), a comprehensive methodology for assessing the seismic vulnerability of buildings in Europe that contemplates the urban system and its interrelation with other systems and proposes new taxonomies for reinforced concrete and masonry; GEM Taxonomy; v.2.0 (Brzev et al. 2013), which comprises 13 attributes that describe the structural system, roofing, flooring, building envelope, and use, among other things. These attributes are divided into sub-levels corresponding to the specific characteristic of the main attribute, thus providing a more detailed physical description of the building and positioning the system as a global taxonomy. In this context, and in line with the Sendai Framework's goal of promoting availability of multi-hazard systems, recent taxonomies such as GED4ALL (Silva et al. 2018)—which presents a taxonomy based on GEM v2.0—have taken a multi-risk approach, clearly distinguishing between the concepts of exposure (common to all risks) and vulnerability (specific to each risk). Other use cases include European projects such as SERA (2017), urban exposure studies (Pittore et al. 2018), and studies based on modelling building typologies (Esteghamati et al. 2020).

Hence, the main objective of any building vulnerability rating is to identify those building typologies that might respond differently to seismic shaking and group them into classes to evaluate their response and the estimated extent of potential losses. Ongoing studies rate the variables that affect building vulnerability and formulate a statistical model using a discrimination index (Martínez-Cuevas et al. 2020) that makes it possible to identify habitable and non-habitable buildings. Recently published studies in the field of machine learning apply artificial neural networks (ANN) to the field of seismic engineering. Several authors (e.g., Stefanini et al. 2022; Ferreira et al. 2020) have used techniques based on artificial intelligence to evaluate seismic response and estimate damage. Others (Vazirizade et al. 2017; Tang et al. 2021) have applied machine learning to structural reliability assessments and rapid assessments of seismic risk and potential building loss.

This paper focuses on using data mining and machine learning techniques to obtain data and ultimately classify the predominant building construction patterns (including structural and non-structural aspects) in an urban area. The method was applied to the city of Murcia, which has one of the highest seismic risks of any city in Spain (IGN-UPM 2013; Gaspar-Escribano et al. 2015). The last significant earthquake in this region occurred in 2011 in the city of Lorca, 105 km from the city of Murcia and the two cities have similar urban settings. The literature on Lorca's post-earthquake vulnerability includes studies on seismic behaviour in masonry and reinforced concrete buildings (Basset-Salom and Guardiola-Víllora 2014; Gomez-Martinez et al. 2015; Ródenas et al. 2018) and analyses of building typologies and urban modifiers based on empirical data and the re-evaluation of macroseismic intensity estimates (Martínez-Cuevas and Gaspar-Escribano 2016; Martínez-Cuevas et al. 2017).

The proposed methodology uses multivariate statistical techniques and machine learning to classify building types according to seismic vulnerability. The study was carried out by performing data analysis at different resolutions, starting with primary data sources, and then verifying and collecting data via remote surveys using online map viewers, digital cartography analysis, and Geographical Information System (GIS) tools to obtain a

high-quality final database. Based on the data obtained, an initial building typology classification was carried out by applying a two-step cluster analysis and a multilayer perceptron ANN to the final evaluation of key building typologies in the study areas. Finally, seismic vulnerability was assessed and classified according to the EMS-98 macroseismic scale (Grünthal 1998).

2 General methodology

The present study proposes a methodology (Fig. 1) that is divided into three steps, with specific workflow stages in each phase.

The first step consisted on selecting a study area in the city of Murcia, compiling building data extracted from the Cadastre (<https://www.sedecatastro.gob.es/>) and setting up an initial database and geographic information system. This allowed us to preliminarily identify the different Construction Typologies (CT) and their variability within the selected study area. Preliminary data were checked for potential identification errors by means of a typology membership probability matrix for Murcia (RISMUR 2014), and in parallel, remote field work was carried out to identify urban modifiers for the buildings in the study area (Martinez-Cuevas et al. 2017). In this way, an initial database was configured and was thus enriched with the incorporation of new attributes (urban modifiers) coded according to the GEM taxonomy (<https://taxonomy.openquake.org/>). The initial CTs completed with the corresponding urban modifiers are called Building Typologies (BT) in this work.

The second step involved multivariate statistical techniques. A two-step cluster analysis was used to identify clusters, correlating previously identified construction typologies with the urban modifiers present in each building. For each CT, the resulting clusters, called Building Cluster Typologies (BCT) in this work, were internally similar but dissimilar compared to other clusters. The various BCT were evaluated using a multilayer perceptron neural network (ANN-MLP), which allowed us to assess the representativeness of the natural clusters obtained using the multivariate technique.

The third step consisted of conducting a vulnerability study of structural and non-structural components, developing a building taxonomy based on the BCT evaluated, calculating average vulnerability at the neighbourhood and census tract levels, and establishing vulnerability index ranges for each cluster analysed. The European Macroseismic Scale (EMS-98) was used to classify vulnerability classes, in addition the corresponding vulnerability curves are prepared for each BCT.

2.1 Study area and data

The study area comprises the urban area of the city of Murcia, whose local administration is divided into 28 neighbourhoods and 158 census blocks (CB). The former represents the largest scale at the urban level while the latter are lower-level territorial subdivisions that are useful for disseminating statistical data. The total area covered by the sum of all neighbourhoods analysed is approximately 11.93 km². The study area contains high-density residential buildings with unique urban characteristics and construction typologies attributable to the date of construction, urban planning, and the way construction techniques have evolved. The total building stock consists of 8698 buildings, of which 7631 are residential buildings. These buildings constitute the basis of this study, corresponding to 100% of the residential buildings within the city of Murcia. Figure 2 shows the study area and the

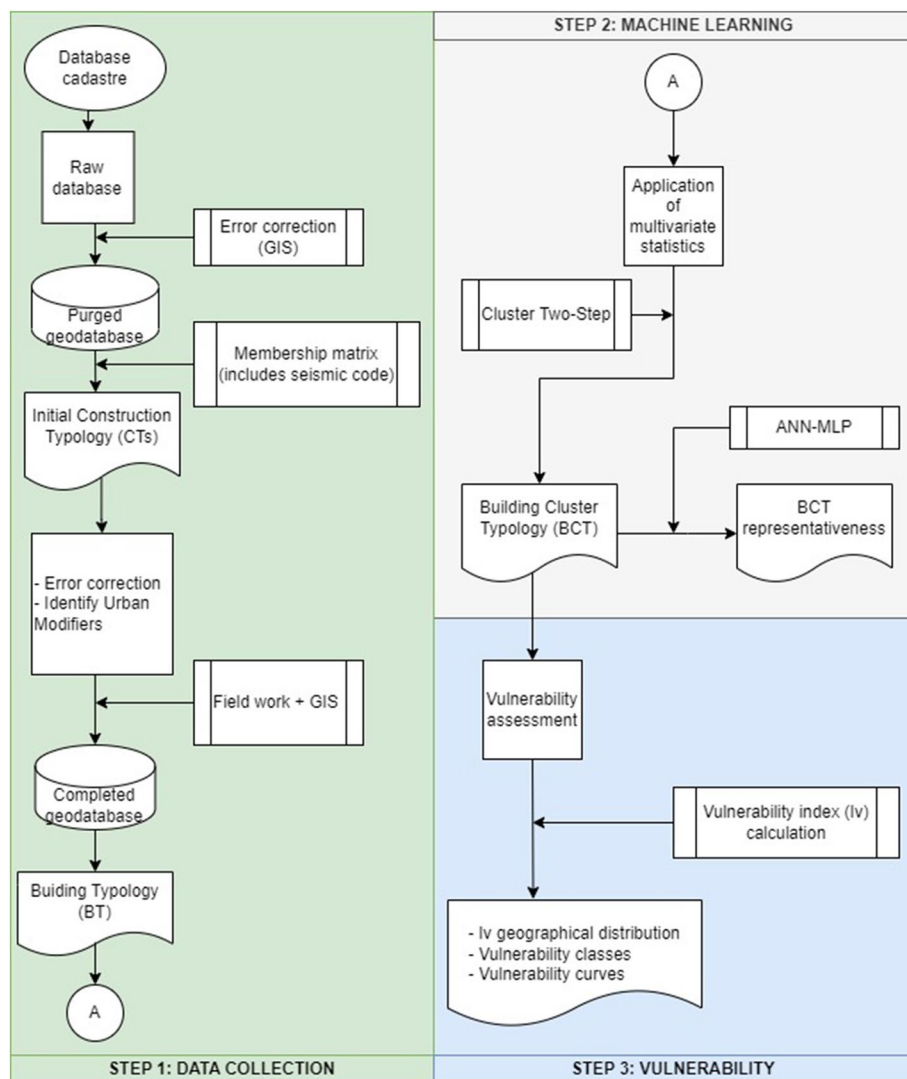


Fig. 1 Diagram showing the steps of the proposed methodology

correlation between neighbourhoods, census blocks, and the number of residential buildings analysed.

2.2 Database and geographical information

Initially, a building database and a GIS were set up using the primary data (shapefiles Building and BuildingPart) obtained from the Spanish Cadastre (<https://www.catastro.minhap.es/webinspire/index.html>) according to the INSPIRE Directive. The attributes extracted for each building included: identifier for each building and building parts, cadastral reference, geometries, year of construction, state of conservation, total number of

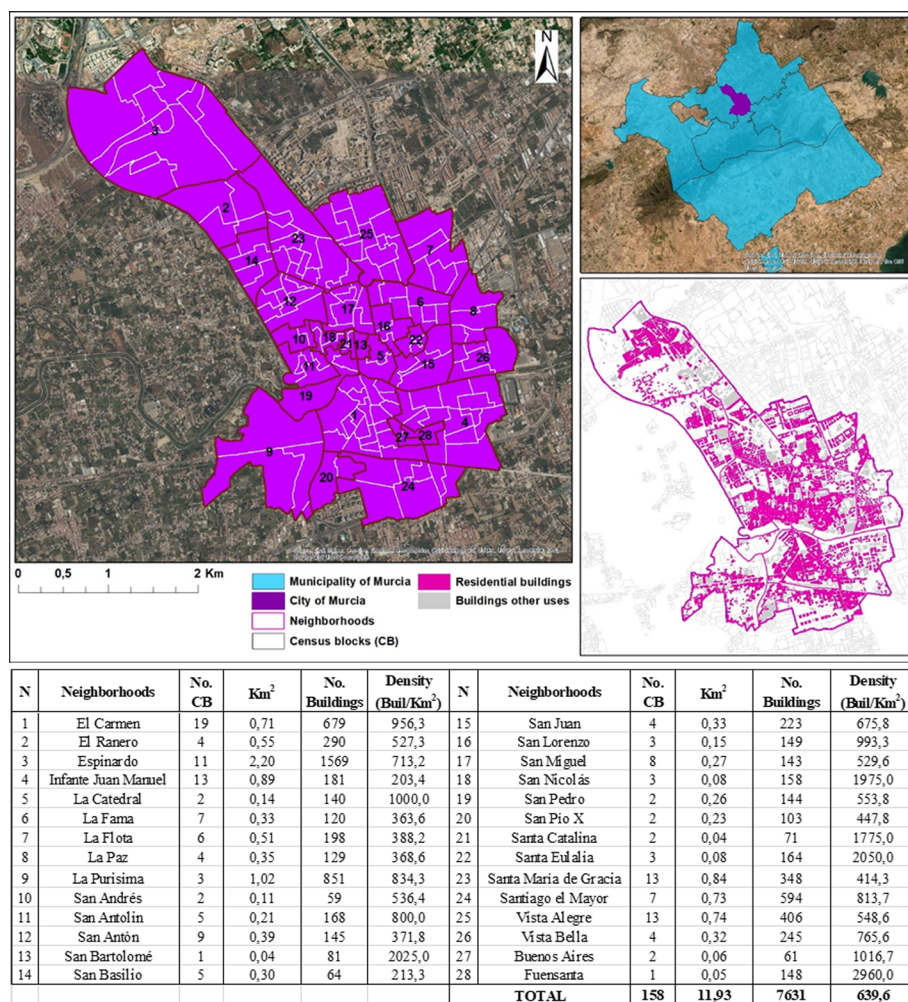


Fig. 2 Study area in the city of Murcia and a list of neighbourhoods, number of census blocks, number of buildings and densities

building, number of dwellings, number of storeys above and below ground level, reforms, areas, centroid coordinates, and precision. This database was depurated by specifying different IDs for buildings sharing the same cadastral reference, by correcting errors in contouring polygons and by removing nonessential data for the purpose of this work (retaining the number of storeys, year of construction, building use, number of dwellings, and renovation only). This process led to the configuration of the CT database, that was complemented with the attributes related to urban modifiers to obtain BT database.

The first step was to apply a building membership probability matrix (RISMUR 2014) that links the year of construction with the approximate validity periods of the corresponding earthquake-resistant building codes for the entire region of Murcia (Table 1). The initial matrix was randomly applied to the entire building database to obtain an initial identification of CTs. The CTs were classified as low-rise, medium-rise, and high-rise buildings

Table 1 Evolution of seismic codes applicable in Spain

Seismic regulations	BOE* publication	Earthquake-resistant design level	Enforcement period
No regulations	...	No seismic code	< 1964
MV-101/1962	February 9, 1963	Pre-code	1964–1969
PGS-1/1968	February 4, 1969	Pre-code	1970–1976
PDS-1/1974	November 21, 1974	Pre-code	1977–1996
NCSE-94	February 8, 1995	Medium-code	1997–2004
NCSE-02	October 11, 2002	Medium-code	> 2004

*Official Gazette of the Spanish State

according to the number of storeys. Depending on the year of construction and the applicable building code, concrete buildings were classified as pre-code and low-code (buildings with higher level of seismic design are not present in the study area).

It should be noted that, the NCSE-02 seismic-code regulation is an improved update of NCSE-94 based on probabilistic hazard studies. The NCSE-02 stands out for its focus on seismic safety and building lifespan, including detailed information on construction and conceptual design of projects. It also introduces the soil coefficient “S” factor, maintaining the same ground accelerations for Murcia (0.15 g). In our study, we have considered continuity in construction technology from the validity of NCSE-94, classifying buildings as CT RC3.1-low. During the years 1997–2004, the RC3.1-pre classification corresponds to 30%, by construction licenses under PDS-1 regulation. Table 2 offers a general description of each CT.

As the initial membership matrix was applied randomly to the baseline data, errors were observed in the initial identification of CTs, where an important piece of information is the number of storeys. For example, buildings identified as M1.1 and M3.1 were found to have more than seven storeys, which is more common in reinforced concrete structures. These errors were checked and rectified by means of remote inspection using the Street View tool in Google Maps. Based on the above, we hypothesized that some structures were near the lower and upper limits of the probability matrix’s date range, where uncertainty in identifying typology might be greater.

2.3 Remote field work

Using the data collection approach proposed by Martínez-Cuevas and Gaspar-Escribano (2016), remote fieldwork was carried out to check and correct preliminary CTs results. Google Maps was used to locate the buildings, and the Street View tool was used for visual inspection with a special emphasis on the dates of construction closest to the years in which seismic-resistant regulations changed. Approximately 60% (4607 out of 7693) of buildings were observed, with 64% (2952) of them incorrectly classified through the application of the initial matrix at random. Table 3 shows the matrix of CTs obtained after the remote fieldwork inspection. Figure 3 illustrates the number of buildings for each CT.

From the preliminary analysis and out of the total number of buildings, 62 buildings were listed as dilapidated. After checking the data through field work, the findings observed indicated that these buildings no longer existed, and the plots were vacant. On the other hand, 275 buildings had been completely renovated, leading us to re-assign

Table 2 CTs and height classification in the study area

	CT	Description	Height		
			Low-rise (L)	Medium-rise (M)	High-rise (H)
Masonry	M1.1	Unreinforced masonry-bearing rubble-stone walls with wooden slabs	1–2 storeys	3–5 storeys	> 5 storeys
	M3.1	Unreinforced masonry-bearing walls with wooden slabs			
	M3.4	Reinforced masonry with concrete slabs			
RC	RC3.1-pre	Regular infilled walls	1–3 storeys	4–7 storeys	> 7 storeys
	RC3.1-low	Regular infilled walls, low ductility class			

Table 3 Membership matrix of CT through field work

CTs	Period + regulation									
	<= 1900	1901–1920	1921–1940	1941–1950	1951–1963	MV-101 ^a 1964–1969	PGS-1 ^b 1970–1976	PDS-1 ^b 1977–1996	NCSE-94 ^c 1997–2004	NCSE-02 ^c > 2004
M1.1	0.55	0.42	0.22	0.14	0.02	0.00	0.00	0.00	0.00	0.00
M3.1	0.45	0.54	0.67	0.66	0.50	0.07	0.01	0.00	0.00	0.00
M3.4	0.00	0.04	0.12	0.20	0.36	0.48	0.36	0.12	0.00	0.00
RC3.1-pre	0.00	0.00	0.00	0.00	0.12	0.45	0.63	0.87	0.30	0.00
RC3.1-low	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.70	1.00

Seismic regulations applicable in Spain, depending on the construction period

^aPoor construction quality with no earthquake-resistant design

^bAcceptable construction quality with “Pre” level earthquake-resistant design

^cAcceptable construction quality with a low seismic-resistant design level

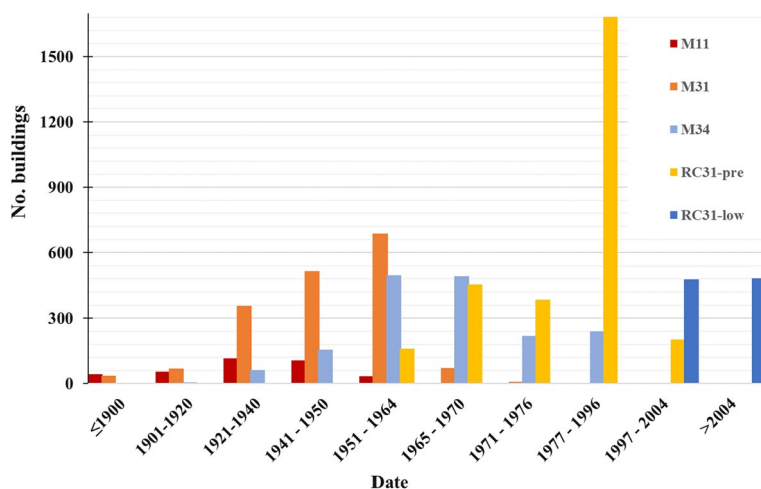


Fig. 3 CT results using remote field work

typologies based on the year of renovation, in accordance with current seismic-resistant regulations. Table 4 shows the amounts of buildings for each CT and general percentage within the study area. Figure 4 shows the percentage breakdown of the various CTs obtained for each neighbourhood.

3 Classification and characterization of urban modifiers

In the present study, urban modifiers reflect urban planning parameters that are applied to residential buildings and make buildings more or less vulnerable, depending on their location and other factors that may affect their behaviour in the event of a seismic event (Martinez-Cuevas et al. 2017). The GEM taxonomy Brzev et al. (2013) was used to code the various urban modifiers. Table 5 describes the derivation of each urban modifier based on different data sources, remote fieldwork, and GIS processes.

Below we describe each urban modifier and the procedure to obtain it:

Table 4 CTs frequencies and percentages in the study area

Construction typology	Buildings	%
M1.1	354	4.6
M3.1	1753	23.0
M3.4	1672	21.9
RC3.1-pre	2890	37.9
RC3.1-low	962	12.6
Total	7631	100

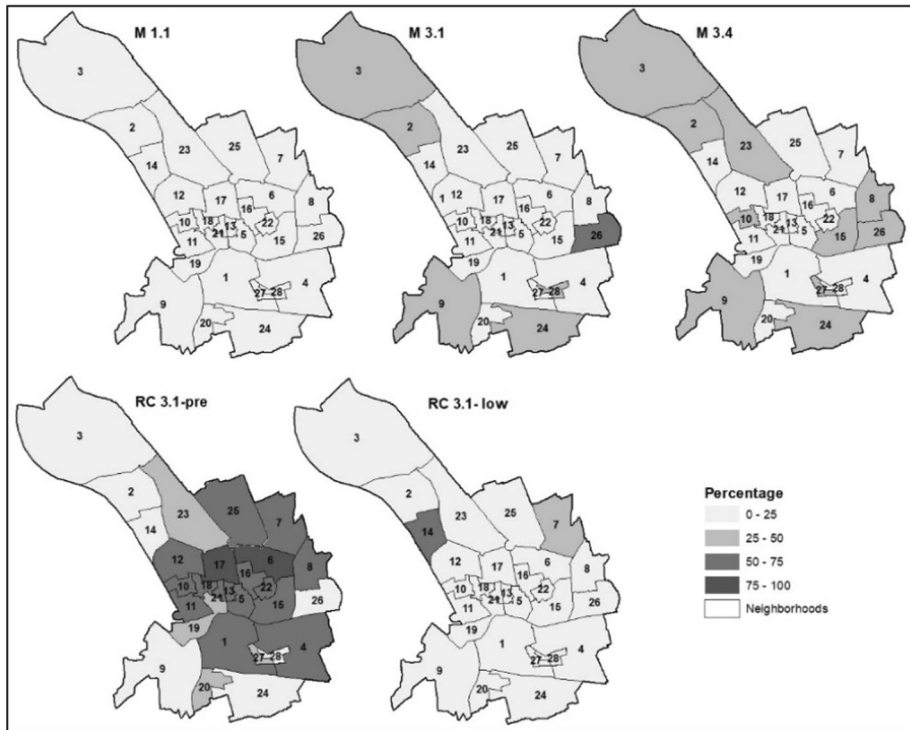


Fig. 4 Spatial distribution of CTs in the neighborhoods studied

3.1 Irregular floor plan (code PLF)

Floor plans with irregularities, mainly re-entrant corners, were evaluated based on cartographic data from the cadastre and using a GIS environment. Thus, a corner is considered re-entrant when the setback measures at least 15% of sections L and L' in the floor plan (Fig. 5). We also identified the various floor plan shapes for buildings (E, H, L, T, U) coded PLF (Brzev et al. 2013). Buildings with a setback $< 15\%$ are considered rectangular (R) or square (SQ), as the case may be.

3.2 Soft storey (code SOS)

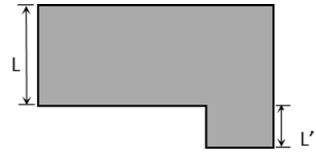
Structural layouts in which one storey of a building, usually the ground floor, is more ductile and/or weaker than the floors immediately above. Soft storeys are typically used as parking, retail, or office spaces (Brzev et al. 2013; Martinez-Cuevas et al. 2017). For the purposes of this study, mixed-use buildings with a second storey intended for non-residential use (as given by the cadastral database) are assumed to have a soft storey and are coded as SOS.

Table 5 Coding and identification of urban modifiers in the study area

Urban modifier	GEM code	Cadastral data ^a	Remotely viewed ^b	GIS process ^c	External data ^d
Regular floor plan	PLF	✓	✗		✗
Soft storey	SOS	✓	✗	✗	✗
Irregularities in the vertical structure	CHV	✓	✓	✓	✗
Short column	SHC	✓	✗	✓	✓
Residential building type	RES	✓	✗	✗	✗
Difference in height compared to adjacent Buildings	POP	✓	✓	✓	✗
Position of the building within the block	BP	✓	✓	✓	✗

^aRepresents the main database/databases and first option for the identification of urban modifiers^bRepresents a specific verification query^cRepresents a process after querying the main database

Fig. 5 Definition of an irregular re-entrant floor plan is where $L' > 0.15L$ or $L' \text{ and } L \geq 15\%$ (Charleson 2008)



3.3 Irregularities in the vertical structure (code CHV)

Variations in the vertical structure and/or the materials used in the vertical structures of the buildings were evaluated according to the approach of Lantada (2007), which compares the perfect and actual volumes of the building. These volumes were calculated using area data from the cadastral database with the support of GIS processes. The results of this process were validated with the dataset of remotely viewed buildings. In the remote visual inspection, we analysed buildings in which various parts of the exterior wall are displaced with respect to the limits of the vertical planes that enclose each building or its fachade (Martinez-Cuevas et al. 2017). The results of the visual inspection provided excellent reliability of the processes executed in GIS, which led us to assume as valid the results of these processes in non-inspected buildings. The code used to indicate variations in the vertical structure was CHV.

3.4 Short column (code SHC)

Irregularity related to variations in the distribution of columns within a building. This occurs when, at any given point, some columns are shorter or more fragile than others (Martinez-Cuevas et al. 2017). This concept also includes captive columns, a type of column whose free length is restricted by infill walls (Guevara and García 2001). Generally, a building has short columns if it has captive columns where non-structural walls reduce the height of the unsupported length of the column by at least half or the column aspect ratio to less than 2:1 (Brzev et al. 2013). This modifier was assigned upon analysing the site (see Sect. 6.1) for each reinforced concrete building that featured underground storeys (obtained from cadastral data). Based on the site analysis, buildings with underground storeys on slopes with gradients of $> 10\%$ were assigned code SHC to indicate short column irregularity.

3.5 Residential building type (code RES)

Residential buildings were divided into two main categories, according to cadastral data, based on the number of dwellings listed: single-family home or collective housing, using codes RES1 and RES2, respectively. It is worth mentioning that single-family residential subdivisions may include detached, semi-detached, attached, or closed-block buildings. Multifamily residential buildings, or collective housing, may include buildings separated by party walls or open-block buildings with façades facing the street.

3.6 Difference in height compared to adjacent buildings (code POP)

According to Brzev et al. (2013), when the space between buildings is less than 4% of the height of the shorter building, insufficient or non-existent spacing between adjacent buildings can potentially produce a pounding effect, leading to damage to both structures. For Lantada (2007), this irregularity is applied when there is a difference of at least 2 storeys between adjacent buildings. Based on this last criterion, the code POP was assigned to adjacent buildings with a difference of 2 or more storeys, according to a GIS inspection of the distribution of number of floors extracted from the cadastral database and validated for the buildings inspected in the field work.

3.7 Position of the building within the block (code BP)

Brzev et al. (2013), classify the position of a building within a block based on the adjoining or attached buildings on either side of the building to be classified. Accordingly, they establish four distinct categories: BPD buildings with no other adjoining or attached buildings, BP1 with adjoining or attached buildings on one side, BP2 with contiguous or attached buildings on two sides, and BP3 with adjoining or attached buildings on three sides. These can also be categorized as: detached buildings, header/terminal buildings at either end of a block, corner buildings, and intermediate buildings (Lantada 2007). Based on the above criteria, building location within the block was determined using GIS based on automatic neighbouring polygons techniques. The polygons for all buildings in the cadastre data set were analysed and, finally, a classification was assigned only to residential buildings.

Figure 6 illustrates the classification adopted. Sub-classifications were included to provide greater detail about the position of buildings within a block. This enabled us to adopt criteria, using the year of construction and the year in which renovations were completed as primary data. Our analysis resulted in the following classification: (i) semi-detached buildings with distinct cadastral reference numbers and identical structural typology that were built or renovated in the same year were assumed to have a continuous structure and classified as BPD; (ii) header/terminal buildings at either end of a row of attached buildings were classified as BP1; (iii) semi-detached buildings built or renovated in different years were assumed to have a discontinuous structure and classified as BP1_P, like the header/terminal buildings on either end of a row of attached buildings; (iv) attached buildings with adjacent units on two sides, leaving a corner of the building to be catalogued free, were classified as BP2_E corner buildings; (v) two adjacent buildings on opposite sides of the building to be catalogued were classified as BP2_I intermediate buildings; (vi) buildings with one side free were classified as BP3 intermediate buildings.



Fig. 6 Classification of buildings within blocks

3.8 Identifying urban modifiers in the study area

A descriptive statistical study was carried out to determine the frequency of each urban modifier in each CT. Figure 7 shows the relative presence of 1–2 storey masonry buildings and of 3–5 storey reinforced concrete buildings. Masonry CTs (M1.1, M3.1 and M3.4) tend to have highly irregular vertical structures (CHV) because they often have balconies and bay windows. They also tend to have regular floor plans. On the other hand, reinforced concrete buildings (RC31-pre and RC31-low) tend to have irregular floor plans and highly irregular vertical structures (CHV), and most have soft storeys (SOS). This makes it possible to infer certain Building Typologies (BT) from the combination of CT and urban modifiers. However, it can be useful to obtain more detailed information and establish a more direct relationship between CT and urban modifiers, as these can affect the seismic performance of buildings. In this study, we applied multivariate statistics and, more specifically, cluster analysis to identify subgroups, applying homogeneity criteria to help define representative BTs in the study area and, in turn, assess and compare their seismic vulnerability.

4 Identifying patterns based on a multivariate statistical study

A multivariate analysis was performed for each CT using the Two-Step Cluster analysis proposed by Chiu et al. (2001). IMB SPSS Statistics 26 software was used to perform the cluster analysis to identify similar and dissimilar multivariate patterns (urban modifiers) based on CTs.

The starting point was a set of five CTs for which we had access to information coded as a set of variables corresponding to the various urban modifiers identified previously. The first step was to select a limited number of variables, making sure not to duplicate the information contained in the data. The next step consisted of selecting the distance/dissimilarity measure, which must match the types of variables. In our case, the urban modifiers are qualitative variables, and the distance measures are for non-metric data. A grouping

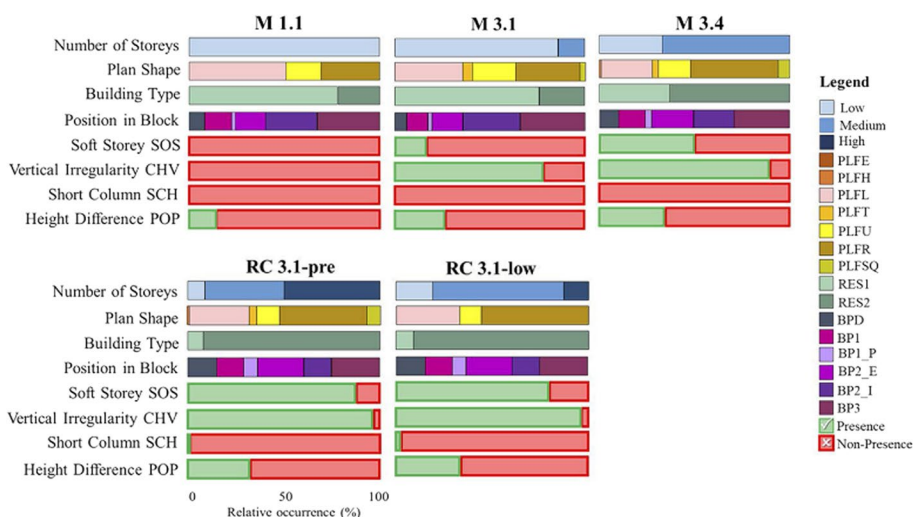


Fig. 7 Relative incidences of urban modifiers in each CT identified in the study area

criterion was applied to generate the various clusters with their corresponding predictor variables. This made it easier to form homogeneous groups. The last step consisted of evaluating the various clusters obtained.

This method was used because it can handle categorical and continuous variables, analyse large volumes of data using a cluster feature (CF) tree, and automatically deliver the natural and optimal number of clusters. The latter was carried out subject to the conditions of independent variables, with multinomial distribution for categorical variables and normal distribution for continuous variables. Nevertheless, internal empirical tests indicate that this multivariate analysis method is robust even when these conditions are not met, since the Two-Step cluster does not involve hypothesis testing or calculation of significance, leaving the decision as to whether or not the solution is satisfactory up to the researcher (Rubio-Hurtado and Baños 2017; Ballestar et al. 2018; Norusis 2011).

4.1 Types of variables

Based on the final data, the urban modifiers affecting each CTs were used as nominal qualitative variables, thus enabling us to obtain homogeneous clusters for each identified CT. A modifier was assigned a categorical variable when its presence was the only way a CT could feature this characteristic. If more than one modifier was present in each CT, it was coded as a dichotomous variable indicating the presence or absence of said characteristic (Table 6). Variables were incorporated in the analysis based on the principle of parsimony, thus limiting the selection of variables (Rubio-Hurtado, and Baños 2017). Consequently, E, H, L, T, and U floor plans were assumed to include re-entrant corners so that the presence of re-entrant corners in a floor plan did not have to be considered a variable.

4.2 Implementing two-step cluster analysis

The data was inputted in SPSS in a randomized manner, as recommended by Norusis (2008). The first stage of analysis consisted of sequence clustering to generate a CF tree structure of attributes using the notation described by Zhang et al. (1996), using the default CF-tree settings. This way, the variables (urban modifiers) for each CT was organized to form a branching tree structure, moving from the bottom, or root node, to the top level, or leaf node. The software process was sequential, counting the categories for each variable by scanning the records following the approach proposed by Theodoridis and Koutroumbas (2001). From the initial branching to the end of the process, the records were divided up to the leaf node by applying similarity/dissimilarity criteria to the data, thus generating a set of pre-clusters. The distance measure used corresponds to the maximum logarithmic likelihood using BIC clustering criterion (Schwarz 1978).

The number of clusters (Fig. 8) for which the BIC value drastically decreases, and the inflection point on the curve steadily declines from that point on is directly correlated with the formation of optimal clusters (Fraley and Raftery 1998; Norusis 2008). Given that our data set may include different points (bends in the curve) for each CT, thus altering the number and quality of the clusters, the inflection points defining each of the final clusters proved to be of good quality. After first verifying internal stability and specifying various fixed and maximum numbers of clusters, the various clusters were formed with the maximum number of clusters set to 10. Once the pre-clusters were formed, an exchange ratio was applied to the pre-cluster distance measure to determine the final clusters.

Table 6 List of variables included in the two-step cluster analysis

Variables	Description
Number of storeys	Categorical variable. Only possibility in range of storeys Range 1: 1–2 storeys Range 2: 3–5 storeys Range 3: 1–3 storeys Range 4: 4–7 storeys Range 5: > 7 storeys
Building shape in plan	Categorical variable. Only possibility of shape Shape 1: PLFE Shape 2: PLFH Shape 3: PLFL Shape 4: PLFT Shape 5: PLFU Shape 6: PLFR Shape 7: PLFSQ
Type of building	Categorical variable. Only possible type Type 1: RES1 Type 2: RES2
Position in block	Categorical variable. Only possible position Position 1: BPD Position 2: BP1 Position 3: BP1_P Position 4: BP2_E Position 5: BP2_I Position 6: BP3
Soft storey	Dichotomous variable. Possible presence of soft storey 1: Presence SOS 0: Non-presence SOS
Vertical irregularity	Dichotomous variable. Possible presence of vertical irregularity 1: presence CHV 0: non-presence CHV
Short column	Dichotomous variable. Possible presence of short column 1: Presence SHC 0: Non-presence SCH
Height difference	Dichotomous variable. Possible presence of height difference 1: Presence POP 0: Non-presence POP

Urban modifiers present in CTs and their correspondence with GEM taxonomy

The final setting for the clusters was assessed by means of silhouette validation (Rousseeuw 1987), which determines the cohesion of elements within a cluster and the distance between clusters. The silhouette coefficient can vary between -1 and 1 , where -1 is a poor model and 1 is a robust model. In general, a result > 0.5 indicates that the model is of good quality (Kaufman and Rousseeuw 1990). In our data, the silhouette

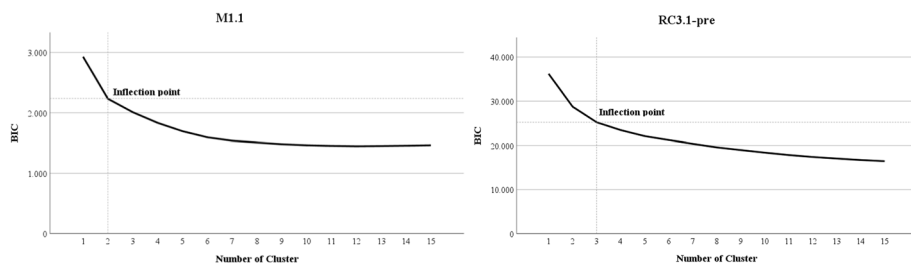


Fig. 8 Example of optimal number of clusters for M1.1 and RC3.1-pre based on the Schwarz Bayesian Criterion (BIC)

coefficient for the clusters formed was >0.5 (Fig. 9), indicating that the various modifiers for each typology generated optimal clusters.

4.2.1 Variables predictive of cluster formation

The variables introduced in this method were ranked in order of importance based on their specific contribution both within each cluster and between clusters. As may be observed in Fig. 10, building height was the key variable in most clusters, except in M1.1, for which height was not introduced as a variable. These buildings have a uniform height range of 1 to 2 storeys (see Fig. 8), for which the predictor variable is the residential building typology. This same correlation is also observed in RC3.1-low, for which the key predictor variable also corresponds to single-family or multi-family residential buildings. The other variables were less relevant to cluster formation, indicating that the existence or non-existence of certain modifiers is not representative when it comes to cluster formation. That is, the final weight of the variables is affected by the number of times that said value is present as building attributes (see Fig. 7). Nevertheless, it is useful to know the proportion in which modifiers are present in each CT so that we can subsequently assess the vulnerability associated with each cluster.

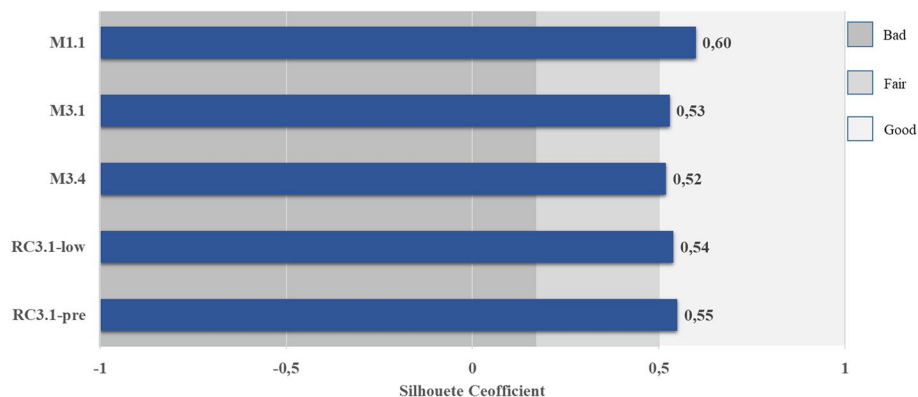


Fig. 9 Silhouette coefficients for cluster models applied to each CT

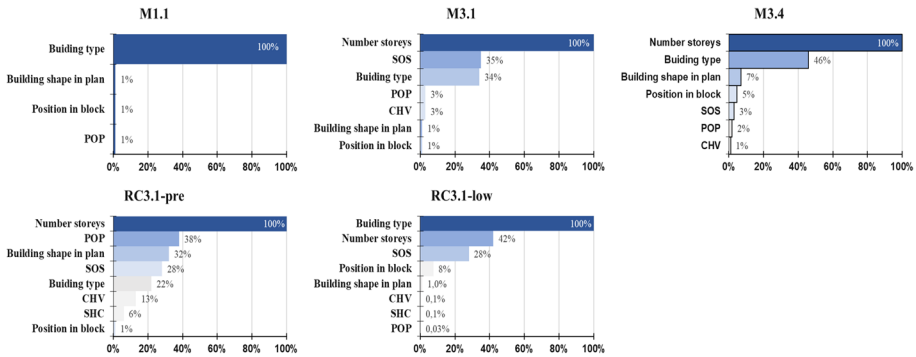


Fig. 10 Variables according to their degree of importance in identifying clusters

4.2.2 Buildings cluster typologies obtained

Each CT is associated, to a greater or lesser percentage, with urban modifiers that make up each BT. By applying the two-step method, it is possible to define subgroups of building typologies or Building Cluster Typologies (BCTs) for each BT. Figure 11 shows the final BCTs obtained and the participation percentages in each CT. The BTs were rearranged into homogeneous natural clusters, resulting in 11 BCTs.

Table 7 shows the description of frequencies and percentages for each urban modifier present in each BCT. To facilitate comprehension of each BCTs, the suffix “_C” followed by the cluster number was added to the type of construction. For example, M1.1

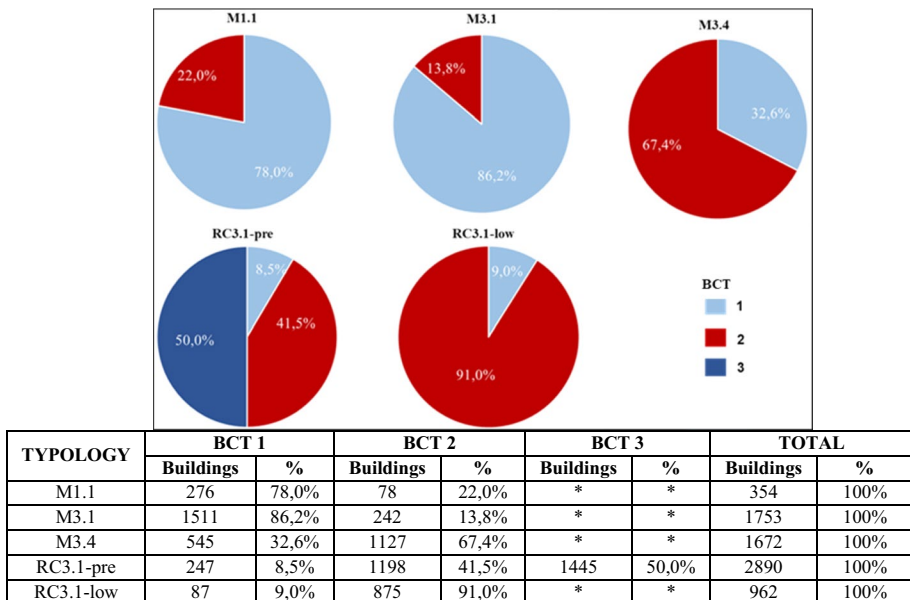


Fig. 11 Final BCTs obtained using the two-step cluster analysis

Table 7 Frequencies and percentages of urban modifiers in each BCT

Masonry urban modifier	M 1.1_C1		M 1.1_C2		M 3.1_C1		M 3.1_C2		M 3.4_C1		M 3.4_C2	
	Buildings	% Cluster	Buildings	% Cluster	Buildings	% Cluster	Buildings	% Cluster	Buildings	% Cluster	Buildings	% Cluster
1–2 storeys	276	100	78	100	1511	100	0	0	545	100	0	0
	*	*	*	*	0	0	242	100	0	0	1127	100
	*	*	*	*	*	*	*	*	0	0	4	0
	*	*	*	*	*	*	*	*	0	0	12	1
	145	53	35	45	550	36	78	32	182	33	270	24
	*	*	*	*	73	5	7	3	18	3	36	3
	53	19	13	17	366	24	38	16	123	23	169	15
	78	28	30	38	478	32	112	46	211	39	555	49
	*	*	*	*	44	3	7	3	11	2	81	7
	276	100	0	0	1280	85	47	19	437	80	181	16
	0	0	78	100	231	15	195	81	108	20	946	84
	22	8	6	8	85	6	14	6	68	12	119	11
	33	12	15	19	177	12	23	10	89	16	151	13
	6	2	2	3	37	2	3	1	13	2	36	3
	50	18	6	8	223	15	60	25	103	19	266	24
	75	27	22	28	476	32	53	22	133	24	214	19
	90	33	27	35	513	34	89	37	139	26	341	30
	*	*	*	*	121	8	154	64	85	16	757	67
	*	*	*	*	1149	76	223	92	464	85	1031	91
POP	42	15	9	12	353	23	102	42	122	22	447	40
Reinforced concrete urban modifier	RC 3.1-pre_C1		RC 3.1-pre_C2		RC 3.1-pre_C3		RC3.1-low_C1		RC3.1-low_C2			
	Buildings	% Cluster	Buildings	% Cluster	Buildings	% Cluster	Buildings	% Cluster	Buildings	% Cluster	Buildings	% Cluster
1–3 storeys	247	100	0	0	0	0	0	87	100	97	11	
4–7 storeys	0	0	1198	100	0	0	0	0	0	657	75	

Table 7 (continued)

Reinforced concrete urban modifier	RC 3.1-pre_C1		RC 3.1-pre_C2		RC 3.1-pre_C3		RC3.1-low_C1		RC3.1-low_C2	
	Buildings	% Cluster	Buildings	% Cluster	Buildings	% Cluster	Buildings	% Cluster	Buildings	% Cluster
> 7 storeys	0	0	0	0	1445	100	0	0	121	14
PLFE	0	0	0	0	1	0	*	*	*	*
PLFH	0	0	2	0	30	2	*	*	*	*
PLFL	84	34	380	32	435	30	36	41	281	32
PLFT	7	3	41	3	55	4	*	*	*	*
PLFU	24	10	141	12	168	12	1	1	96	11
PLFR	124	50	564	47	629	44	50	57	498	57
PLFSQ	8	3	70	6	127	9	*	*	*	*
RES1	174	70	51	4	0	0	87	100	0	0
RES2	73	30	1147	96	1445	100	0	0	875	100
BPD	46	19	112	9	292	20	12	14	121	14
BP1	28	11	185	15	188	13	10	11	156	18
BP1_P	21	9	71	6	120	8	6	7	65	7
BP2_E	38	15	323	27	335	23	16	18	215	25
BP2_J	71	29	183	15	164	11	23	26	130	15
BP3	43	17	324	27	346	24	20	23	188	21
SOS	60	24	1085	91	1377	95	10	11	750	86
CHV	218	88	1153	96	1396	97	84	97	836	96
SHC	1	0	10	1	4	0	1	1	16	2
POP	70	28	394	33	453	31	28	32	290	33

The asterisk (*) means that for the BCT, the urban modifier is not associated, that is, it does not have that urban modifier

was divided into two BCTs, which were named M1.1_C1 and M1.1_C2, representing two different BTs.

4.3 Using neural networks to assess BCTs

The results obtained using the Two-Step cluster method are natural outputs. In other words, the results were obtained using an unsupervised method and the various BCTs were defined based on similarities/dissimilarities among BTs. However, it is worth comparing these outputs using supervised learning techniques to evaluate and verify the representativeness of the BCTs obtained in the study area. This evaluation was carried out using a multilayer perceptron (MLP) artificial neural network (ANN) with IBM SPSS Statistics 26 software, as this ANN meets the characteristics of a universal approximator (Park and Sandberg 1991; Cybenko 1989) (Table 8).

In general, an ANN is a machine learning model that is based on the structure and function of the human brain (McCulloch and Pitts 1943). The ANN is composed of interconnected nodes, also known as artificial neurons, organised into layers (Rumelhart et al. 1986). These layers process and transform input data to produce output predictions, and the intermediate layers are known as hidden layers. The ANN is trained using a set of labelled examples, with the goal of reducing the difference between the predictions of the network and the actual values (LeCun et al. 2015; Hinton et al. 2012). In this context, the implementation of an ANN is essential in our methodology, as it allowed us to transfer the BCTs obtained through the Two-Step cluster method to the supervised outputs of the ANN-MLP. This allows to evaluate the importance and representativeness of each BCT in the study area, without discarding any, using the metrics presented in Table 9.

The ANN input predictor variables, that is to say the urban modifiers (Table 6) were the same variables used to define the various BCTs. The latter was the desired output set. A general ANN was applied, introducing all predictor variables as network input and the set of all BCTs as the output. The objective was to verify the behaviour of the final classification using an ANN and evaluate the degree of representativeness of each BCT. It is important to note that the input variables for each ANN were randomized, this is because, if the input variables are not randomly distributed, it can introduce biases into the model and affect its accuracy and ability to generalize (Goodfellow et al. 2016).

4.3.1 Neural network architecture

The ANN used has three layers: an input layer, a hidden layer, and an output layer. This architecture was sufficient to verify and assess the results obtained using the cluster method. Figure 12a and b illustrate the general forward propagation MLP model. The model consists of a hidden layer; the input layer (urban modifiers), formed by N neurons (sensors), and the output layer (BCT), which consists of M neurons with the previous classification.

The relationship between the urban modifier input set (R^N) and the BCT output set (R^M), is defined as:

$$(x_1, x_2, x_3, x_4, \dots, x_n) \in R^N \rightarrow (y_1, y_2, y_3, y_4, \dots, y_m) \in R^M$$

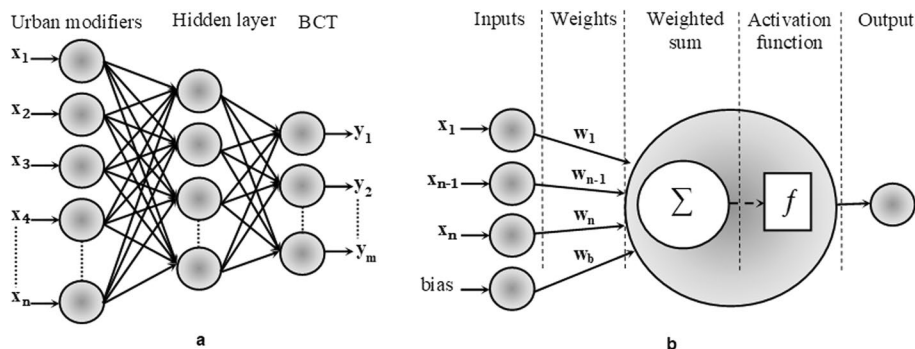
Table 8 Confusion matrix results after applying ANN-MLP

BCT	Sample	No. of cases	TP	TN	FP	FN
M1.1L_C1	Training	175	85	4244	109	90
M1.1L_C2		50	29	4465	13	21
M3.1L_C1		894	743	3264	370	151
M3.1M_C2		140	0	4385	3	140
M3.4L_C1		341	50	4126	61	291
M3.4M_C2		662	662	3724	142	0
RC3.1L-pre_C1		124	33	4380	24	91
RC3.1M-pre_C2		707	676	3441	380	31
RC3.1H-pre_C3		846	838	3629	53	8
RC3.1L-low_C1		52	40	4418	58	12
RC3.1L/M/H-low_C2		537	110	3942	49	427
Overall	59.30%	4528	3266	44,018	1262	1262
Accuracy	72.1%	Cross-entropy error		2750.03		
M1.1L_C1	Test	42	23	1491	39	19
M1.1L_C2		12	6	1552	8	6
M3.1L_C1		322	261	1140	110	61
M3.1M_C2		52	0	1517	3	52
M3.4L_C1		105	13	1446	21	92
M3.4M_C2		233	233	1286	53	0
RC3.1L-pre_C1		59	19	1493	20	40
RC3.1M-pre_C2		266	252	1178	128	14
RC3.1H-pre_C3		283	274	1268	21	9
RC3.1L-low_C1		21	13	1526	25	8
RC3.1L/M/H-low_C2		177	26	1371	24	151
Overall	20.60%	1572	1120	15,268	452	452
Accuracy	71.2%	Cross-entropy error		955.81		
M1.1L_C1	Holdout	59	25	1439	33	34
M1.1L_C2		16	6	1512	3	10
M3.1L_C1		295	246	1120	116	49
M3.1M_C2		50	0	1481	0	50
M3.4L_C1		99	10	1402	30	89
M3.4M_C2		232	232	1248	51	0
RC3.1L-pre_C1		64	22	1459	8	42
RC3.1M-pre_C2		225	218	1191	115	7
RC3.1-pre_C3		316	314	1192	23	2
RC3.1L-low_C1		14	11	1493	24	3
RC3.1L/M/H-low_C2		161	28	1354	16	133
Overall	20.10%	1531	1112	14,891	419	419
Accuracy	72.6%					

TP true positives; *TN* true negatives; *FP* false positives; *FN* false negatives

Table 9 F1-score values used to determine the degree of representativeness of a BCT

F1-score	Type	Description
0.00–0.20	Weak	BCT representativeness is very low or weak
0.21–0.60	Low	BCT with low representativeness
0.61–0.80	Medium	BCT with average representativeness
0.81–1.00	High	BCT with high representativeness

**Fig. 12** Neural network architecture **a** overall ANN-MLP, **b** detailed view of a neuron

The general supervised model (Fig. 12a) follows the function $f(\cdot) : \mathbb{R}^N \rightarrow \mathbb{R}^M$ and is applied to the entire set of building typologies, where the urban modifiers present define N vector spaces and M vector spaces are defined by the BCTs. The variables (urban modifiers) are inputted randomly, after testing stability. The sample set has a 3:1:1 ratio (approximately 60% used for training, 20% for testing, and 20% as a reserve sample), with on-line training. The optimization algorithm is correlated with the gradient slope, where the initial learning rate is 0.1, with Reduced rate of learning is 10 and a 0.9 impulse, the number of epochs was set automatically by SPSS.

Figure 12b shows the individual process for each neuron and indicates the linear weighted sum $\sum_{i=1}^n (x_i w_i + w_b)$, where w_i corresponds to synaptic weights and w_b to the bias weight. In general, the sum corresponds to the product of the matrix $XW + b$, where X represents the value of the input vectors, W the weight vectors and b the bias. The next step is to apply an activation function. In our study, the activation function of the hidden layer was the hyperbolic tangent (Eq. 1), whose output was bounded to the range $(-1, 1)$. In the output layer, activation was performed using the softmax function (Eq. 2), which calculates relative probabilities.

$$\tanh(\beta x) = \frac{e^{\beta x} - e^{-\beta x}}{e^{\beta x} + e^{-\beta x}} \quad (1)$$

$$\sigma(z)_j = \frac{e^{z_j}}{\sum_{k=1}^k e^{z_k}}; \quad \text{for } j = 1, \dots, k \quad (2)$$

In its general form, the MLP applied in our study, which contains a hidden layer and consists of L neurons, is defined as:

$$y_i = f_1 \left(\sum_{j=1}^L w_{ij} s_j \right) = f_1 \left(\sum_{j=1}^L w_{ij} f_2 \left(\sum_{r=1}^N t_{jr} x_r \right) \right) \quad (3)$$

where w_{ij} is the synaptic weight connecting the output neuron i with the j neuron of the hidden layer, the activation functions f_1 and f_2 represent the output and hidden layer units, and the synaptic weights t_{jr} connecting the hidden neuron j with the input neuron x_r . The backpropagation algorithm was used for the learning stages, enabling us to quantify the influence of each weight and bias in the ANN predictions (Rumelhart et al. 1986).

4.3.2 Results of BCT evaluation

Table 8 lists the results extracted from the confusion matrix after applying ANN-MLP and shows how the algorithm performed in supervised learning. The matrix represents the predictions in each BCT as a function of the actual number of cases compared to the number of cases predicted. After inputting all urban modifiers and assessing the ANN capability, fewer predictions were obtained for clusters with fewer cases, with overall accuracies of 72.1% for the training sample, 71.2% for the test sample, and 72.6% for the reserve sample, which generally indicates reliable performance. However, the following considerations should be considered when evaluating the various BCTs:

- In BCTs with high accuracy and low sensitivity or recall, this is indicative of a low rate of class detection that is nevertheless highly reliable when the class is actually detected.
- In BCTs with low accuracy and low sensitivity or recall, this is indicative of a high rate of class detection but includes cases in which the BTC are different.
- In BCTs with low accuracy and low sensitivity or recall, this is indicative of poor classification.
- In BCTs with high accuracy and high sensitivity or recall, this indicates that the model correctly recognizes the class.

In this study, we assessed the representativeness of each BCT in the study area using the F1-score metric, which indicates the harmonic mean between precision and sensitivity or recall and is highly reliable when there is an imbalance between these two metrics. The F1-score values used to evaluate each BCT ranged between 0 and 1, with values approaching 1 indicating a high degree of representativeness. The values used to determine the degree of representativeness of a BCT are listed in Table 7.

Table 8 shows the key metrics used to evaluate the ANN model and assess the representativeness of each BCT. Medium and high ratings correspond to the BCTs that performed best in the ANN-MLP.

5 Seismic vulnerability assessment

The present study adopted an approach based on the vulnerability index method (Benedetti and Petrini 1984; Angeletti et al. 1988; Milutinovic and Tredafiloski 2003), contemplating both building structure and non-structural components that can alter the seismic behaviour of a building.

The final plausible vulnerability index for each building (I_v) was calculated by adding the total scores for the urban modifiers (ΔM_c) and the variation in the basic vulnerability index (I_{vb}), as follows:

$$I_v = I_{vb-BTM} + \sum_{i=1}^n M_{ci} \quad (4)$$

where M_{ci} stands for the n urban modifiers present, which are added to the basic vulnerability index for the building I_{vb} that, in turn, belongs to a given *BTM*. Therefore, the final vulnerability index is a range of vulnerability values with a minimum value, maximum value, and buildings with vulnerability (I_{vx}) indexes, between the maximum and minimum possible values after adding any urban modifiers, that is to say:

$$I_v = [I_v^-, I_v^{++}] = \{I_{vx} \in I_v : I_v^- \leq I_{vx} \leq I_v^{++}\} \quad (5)$$

5.1 Seismic vulnerability estimate applied to BCTs (IvBCT)

The vulnerability index for each BCT was calculated using Eq. 6, as follows:

$$I_{vBCT} = I_{vCT} + \sum_{i=1}^n M_{ci} \quad (6)$$

where I_{vTC} is the basic vulnerability index for the CT.

5.1.1 Value of I_{vCT}

The CTs identified were classified according to height and assigned the vulnerability function and basic vulnerability index, the Building Typology Vulnerability Index (I_{vCT}), corresponding to the most probable vulnerability value proposed by Lagomarsino and Giovinazzi (2006). Table 9 shows the various CTs identified, and the basic vulnerability index used, given the corresponding construction technique and heights.

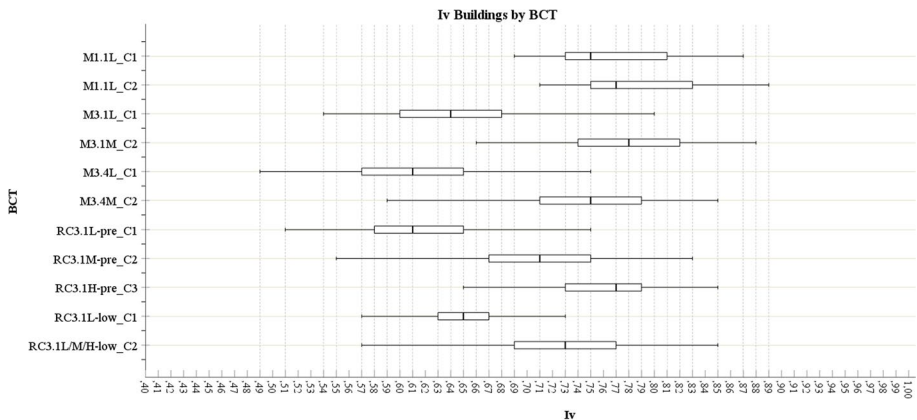
5.1.2 Modifiers by behaviour (M_c)

Building vulnerability can be influenced by the unique characteristics of each building's design and location. These characteristics, M_c , can either enhance or diminish the overall vulnerability of a building.

The M_c evaluated correspond to the urban modifiers observed. A score was assigned for the building type, soft storey, and short column, following the approach proposed by Milutinovic and Tredafiloski (2003). With regards to the presence of soft storeys, this M_c was

Table 10 Evaluation metrics for each BCT when applying ANN-MLP

BCT	Precision (%)	Recall (%)	Specificity (%)	F1-score	Type
M1.1L_C1	43	42	98	0.43	Low
M1.1L_C2	67	38	100	0.48	Low
M3.1L_C1	68	83	91	0.75	Medium
M3.1M_C2	0	0	100	0.00	Weak
M3.4L_C1	25	10	98	0.14	Weak
M3.4M_C2	82	100	96	0.90	High
RC3.1L-pre_C1	73	34	99	0.47	Low
RC3.1M-pre_C2	65	97	91	0.78	Medium
RC3.1H-pre_C3	93	99	98	0.96	High
RC3.1L-low_C1	31	79	98	0.00	Weak
RC3.1L/M/H-low_C2	64	17	99	0.27	Low

**Fig. 13** Boxplot with vulnerability spread calculated for each BCT

evaluated based on the difference between the total number of units in the building (Bu) and the number of residential units (Dw).

Evaluating short columns required implementing a GIS process to analyse soil morphology based on a digital terrain model (MTD25) obtained from Spain's National Centre for Geographic Information (CNIG) that enabled us to distinguish slopes with gradients greater or less than 10%. In evaluating reinforced concrete buildings, the combination of slopes with more than a 10% gradient and the presence of storeys below ground level indicates the probable presence of a short column.

The various classes of floor plan irregularity (E, H, L, T, U, R and SQ) were assigned a score based on evaluating the Ratio of Compactness/Circularity, RC (Udwin 1981 cited in Lantada et al. 2010). M_c values indicating vertical irregularity, differences in height, and position within the block were evaluated by assigning the corresponding score following the GIS procedures proposed by Lantada et al. (2010) Table 10 lists the scores for each M_c evaluated and taken from Martínez-Cuevas et al. (2017), Lantada et al. (2010) and Milutinovic and Trendafoloski (2003).

5.1.3 Calculating the I_{vBCT} value

The final I_{vCT} value for each building is the sum of the I_{vCT} and the M_c for each CT. Figure 13 shows the vulnerability distribution for each BCT as a range of maximum and minimum values. The distribution of I_{vBCT} values shows a concentration of buildings within the first and third quartile (Q1 and Q3), where 50% of buildings with these vulnerability ranges lie, corresponding to the probable I_v values. Twenty-five percent (25%) of the buildings are concentrated below Q1 and above Q3 with respectively less probable vulnerability values. The minimum and maximum vulnerability values (I_v^{++}) represent fewer probable values.

Table 11 shows the ranges described and the probable I_v value, which is the median value in Fig. 13.

Note that the I_v for reinforced concrete BCTs is very similar to that of some masonry buildings. This is because these buildings feature many urban modifiers, mainly soft storeys and vertical irregularities. For certain buildings with short columns, vulnerability approaches the extreme values (Tables 12 and 13).

5.2 BCT descriptions: building cluster typologies

Table 14 below describes each building typology (BCT) identified in this study and its degree of representativeness as indicated in Table 10.

6 Results and discussion

In this section, we will present the final results applied to the city of Murcia. Census blocks were used to visualize the results at a smaller scale, enhancing the final resolution and making them easier to interpret.

The methodological development of our study implied the correction of errors in the primary data source that were related to the digitization of polygons and updating of buildings. Through GIS processes and remote field work, errors were corrected,

Table 11 Basic vulnerability index of CTs

Construction typology	CT + heights	BTM	I_{vb}
M1.1	M1.1L	M1.w_L	0.77
M3.1	M3.1L	M5.w_L	0.62
	M3.1M	M5.w_M	0.70
M3.4	M3.4L	M6_L-PC	0.57
	M3.4M	M6_M-PC	0.65
RC3.1-pre	RC3.1L	RC3_L	0.57
	RC3.1M-pre	RC3_M	0.59
	RC3.1H-pre	RC3_H	0.63
RC3.1-low	RC3.1L-low	RC3-III_L	0.61
	RC3.1M-low	RC3-III_M	0.59
	RC3.1H-low	RC3-III_H	0.59

Table 12 Scores used to calculate vulnerability for each urban modifier

Urban modifiers	M_c	Description	Score
Building shape in plan	PLFE	$0.5 \leq RC < 0.7$	0.02
		$RC < 0.5$	0.04
	PLFH	$0.5 \leq RC < 0.7$	0.02
		$RC < 0.5$	0.04
	PLFL	$RC \geq 0.7$	0.00
		$0.5 \leq RC < 0.7$	0.02
	PLFT	$RC < 0.5$	0.04
		$RC \geq 0.7$	0.00
	PLFU	$0.5 \leq RC < 0.7$	0.02
		$RC < 0.5$	0.04
	PLFR	$RC \geq 0.7$	0.00
		$0.5 \leq RC < 0.7$	0.02
	PLFSQ	$RC < 0.5$	0.04
		$RC \geq 0.7$	0.00
Type of building	RES1	Single-family home	0.00
	RES2	Collective housing	0.02
Position in block	BDP	Detached buildings	0.00
	BP1	Header/terminal buildings	0.06
	BP1_P	Header/terminal semi-detached buildings	0.06
	BP2_E	Corner buildings	0.04
	BP2_I	Middle buildings between two buildings	– 0.04
	BP3	Middle buildings between three buildings	– 0.04
Soft storey	SOS	$Bu - Dw \neq 0$	0.06
Vertical irregularity	CHV	$\delta = 1$	0.00
		$0.5 \leq \delta < 1$	0.02
		$\delta < 0.5$	0.04
Short column	SCH	Storeys below ground + Slope*	0.06
Height difference	POP	Same height	0.00
		A lower building	0.02
		Building between two low buildings	0.04
		A taller building	– 0.02
		Building between two tall buildings	– 0.04

*SCH corresponds to the sum of the scores 0.04 and 0.02 for storeys below ground level and slopes with a gradient over 10% (Milutinovic and Tredafiloski 2003)

obtaining a refined database, which was enriched with data from urban modifiers, resulting in a final database of good quality. In this context, the proposed methodology must necessarily contemplate and correct common errors in the data sources used, since these errors will act as limitations in obtaining robust results. The final database can

Table 13 Vulnerability ranges for each cluster and vulnerability class

BCT	I_v^{--}	I_v^{-}	I_v	I_v^{+}	I_v^{++}
M1.1L_C1	0.69	0.73	0.75	0.81	0.87
M1.1L_C2	0.71	0.75	0.77	0.83	0.89
M3.1L_C1	0.54	0.60	0.64	0.68	0.80
M3.1M_C2	0.66	0.74	0.78	0.82	0.88
M3.4L_C1	0.49	0.57	0.61	0.65	0.75
M3.4M_C2	0.59	0.71	0.75	0.79	0.85
RC3.1L-pre_C1	0.51	0.58	0.61	0.65	0.75
RC3.1M-pre_C2	0.55	0.67	0.71	0.75	0.83
RC3.1H-pre_C3	0.65	0.73	0.77	0.79	0.85
RC3.1L-low_C1	0.57	0.63	0.65	0.67	0.73
RC3.1L/M/H-low_C2	0.57	0.69	0.73	0.77	0.85

be consulted in the *Catalogue of Residential Buildings and Classification According to Seismic Vulnerability in the City of Murcia, Spain* (Meyers-Angulo et al. 2023).

6.1 Distribution of variables and clusters in the city of Murcia

Applying the Two-Step clustering method to our final database generated clusters of buildings with similar characteristics for each CT.

In general, masonry buildings tend to be low or medium-rise, primarily single-family homes with structures up to two storeys high. The layout for most buildings features re-entrant corners, and there is a greater predominance of buildings attached on two or three sides, which explains the average level of the pounding effect. Note that medium-rise buildings have soft storeys and irregular vertical structures.

Most reinforced concrete buildings are medium or high-rise collective housing buildings and present a more varied range of structural designs. The shape of the floor plan and the position within the block show similar characteristics in all clusters with respect to the layouts. There is a high prevalence of soft storeys and irregular vertical structures as well as a high proportion of attached buildings, which directly affects the potential risk of generating the pounding effect. Table 15 shows the percentage distributions for each BCT in the study area.

Figure 14 illustrates the percentage distribution for each cluster relative to its prevalence in each census tract within the neighbourhoods in the study area. Note that the most varied range of BCTs is found closer to the city centre. In more peripheral areas, masonry structures and M3.1 and M3.4 construction types predominate and there is a lower proportion of modern, reinforced concrete buildings to be found. Nevertheless, this proportion is notably greater in certain census blocks, indicating new urban developments.

Table 14 Description of each BCT identified in the study area







BCT	Description
M1.1L_C1 	<p>Buildings with low representativeness: 1 to 2-storey stone or rubble masonry single-family residential buildings with wood girders and an L-shaped floor plan</p> <p>In addition, the floor plan for these buildings is predominantly rectangular and, to a lesser extent, U-shaped</p> <p>The position within the block is mainly within a row of attached buildings (BP1, BP2_E, BP2_I and BP3) of greater or equal height</p>
M1.1L_C2 	<p>Buildings with low representativeness: mainly 2-storey stone or rubble masonry multi-family residential buildings with wood girders and an L-shaped floor plan</p> <p>The floor plan for these buildings is predominantly rectangular and, to a lesser extent, U-shaped</p> <p>The position within the block is mainly within a row of attached buildings (BP1, BP2_I and BP3) of greater or equal height</p>
M3.1L_C1 	<p>Buildings with average representativeness: 1 to 2-storey unreinforced brick masonry wood-frame single-family residential buildings with irregular vertical structures (especially in 2-storey buildings) and L-shaped floor plans</p> <p>In addition, there is a predominance of collective housing with rectangular and, to a lesser extent, U-shaped floor plans</p> <p>The position within the block is mainly within a row of attached buildings (BP1, BP2_I and BP3) of greater or equal height</p>
M3.1M_C2 	<p>Buildings with very low or weak representativeness: 3 to 5-storey unreinforced brick masonry wood-frame multi-family residential buildings with rectangular floor plans, irregular vertical structures, and soft storeys</p> <p>In addition, there is a predominance of 3-storey single-family homes with L-shaped and, to a lesser degree, U-shaped buildings</p> <p>The position within the block is mainly within a row of attached buildings (BP1, BP2_I, BP2_E and BP3) of equal or lesser height</p>
M3.4L_C1 	<p>Buildings with very low or weak representativeness: 1 to 2-storey reinforced concrete brick masonry single-family residential buildings with rectangular floor plans and irregular vertical structures (especially in 2-storey buildings)</p> <p>In addition, there is a predominance of 2-storey collective housing with L-shaped and, to a lesser extent, U-shaped floor plans</p> <p>The position within the block is mainly within a row of attached buildings (BP1, BP2_I, BP2_E and BP3) of equal or greater height. There is a lower proportion of detached buildings (BPD)</p>
M3.4M_C2 	<p>Highly representative buildings: 3 to 5-storey reinforced concrete brick masonry multi-family residential buildings with rectangular floor plans, irregular vertical structures, and soft storeys</p> <p>In addition, there is a predominance of 3-storey single-family homes with L-shaped and, to a lesser degree, U-shaped buildings</p> <p>The position within the block is mainly within a row of attached buildings (BP1, BP2_I, BP2_E and BP3) of greater or lesser height. There is a lower proportion of detached buildings (BPD)</p>

Table 14 (continued)

BCT	Description
RC3.1L-pre_C1	<p>Buildings with low representativeness: 1 to 3-storey unreinforced brick masonry single-family residential buildings with regular concrete infilled porticos, rectangular floor plans, and irregular vertical structures (especially in buildings over 2-storeys high)</p> <p>In addition, there is a predominance of collective housing over 2-storeys high with L-shaped and, to a lesser extent, U-shaped floor plans</p> <p>The position within the block is mainly within a row of attached buildings (BP1, BP2_I, BP2_E and BP3) of greater or lesser height and with soft storeys. There is a lower proportion of detached buildings (BPD)</p>
RC3.1M-pre_C2	<p>Buildings with average representativeness: 4 to 7-storey unreinforced brick masonry multi-family residential buildings with regular concrete infilled porticos, rectangular floor plans, irregular vertical structures, and soft storeys</p> <p>In addition, there is a predominance of collective housing with L-shaped and, to a lesser extent, U-shaped floor plans</p> <p>The position within the block is mainly within a row of attached buildings (BP1, BP2_I, BP2_E and BP3) of equal or lesser height</p>
RC3.1H-pre_C3	<p>Highly representative buildings: more than 7-storeys high unreinforced brick masonry multi-family residential buildings with regular concrete infilled porticos, rectangular floor plans, irregular vertical structures, and soft storeys</p> <p>In addition, there is a predominance of collective housing with L-shaped and, to a lesser extent, U-shaped floor plans</p> <p>The position within the block is mainly within a row of attached buildings (BPD, BP1, BP2_I, BP2_E and BP3) of lesser height. There is a lower proportion of detached buildings (BPD)</p>
RC3.1L-low_C1	<p>Buildings with very low or weak representativeness: 1 to 3-storey unreinforced brick masonry single-family residential buildings with low seismic-resistant design, regular concrete infilled porticos, rectangular floor plans, and irregular vertical structures (especially in buildings with 2 or more storeys)</p> <p>In addition, there is a predominance of buildings with an L-shaped floor plan</p> <p>The position within the block is mainly within a row of attached buildings (BPD, BP1, BP2_I, BP2_E and BP3) of greater height. Lower proportion of detached buildings (BPD) and soft storeys among buildings over 2-storeys high</p>
RC3.1M/L/H-low_C2	<p>Buildings with low representativeness: 4 to 7-storey high unreinforced brick masonry multi-family residential buildings with low seismic-resistant design, regular concrete infilled porticos, rectangular floor plans, irregular vertical structures, and soft storeys</p> <p>In addition, there is a predominance of collective housing over 7-storeys high and, to a lesser extent, 2 to 3-storey collective housing. L-shaped and, to a lesser extent, U-shaped floor plans</p> <p>The position within the block is mainly within a row of attached buildings (BP1, BP2_I, BP2_E and BP3) of equal or lesser height. There is a lower proportion of detached buildings (BPD) and soft storeys</p>

Table 15 Percentage distribution of BCTs in study area

Construction typology	BCT	%
M1.1	M1.1L_C1	3.62
	M1.1L_C2	1.07
M3.1	M3.1L_C1	19.80
	M3.1M_C2	3.20
M3.4	M3.4L_C1	7.10
	M3.4M_C2	14.80
RC3.1-pre	RC3.1L-pre_C1	3.20
	RC3.1M-pre_C2	15.70
	RC3.1H-pre_C3	18.90
RC3.1-low	RC3.1L-low_C1	1.14
	RC3.1L/M/H-low_C2	11.47

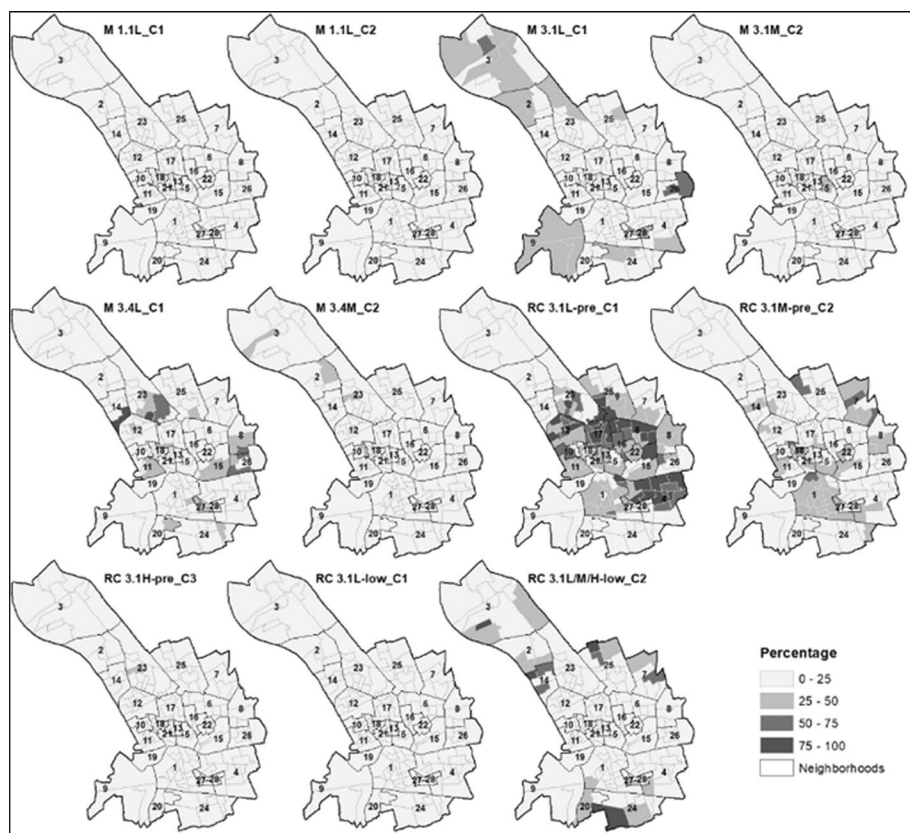
**Fig. 14** Percentage distribution of each BCT in the study area

Table 16 EMS-98 macroseismic scale reclassification applied to BCTs

Construction Typology	BCT	VULNERABILITY CLASS					
		A	B	C	D	E	F
MASONRY	M1.1	●.....				
		●.....				
	M3.1	●.....				
		●.....				
	M3.4	●.....				
		●.....				
		●.....				
		●.....				
REINFORCED CONCRETE	RC3.1-pre	●.....				
		●.....				
		●.....				
	RC3.1-low	●.....				
		●.....				

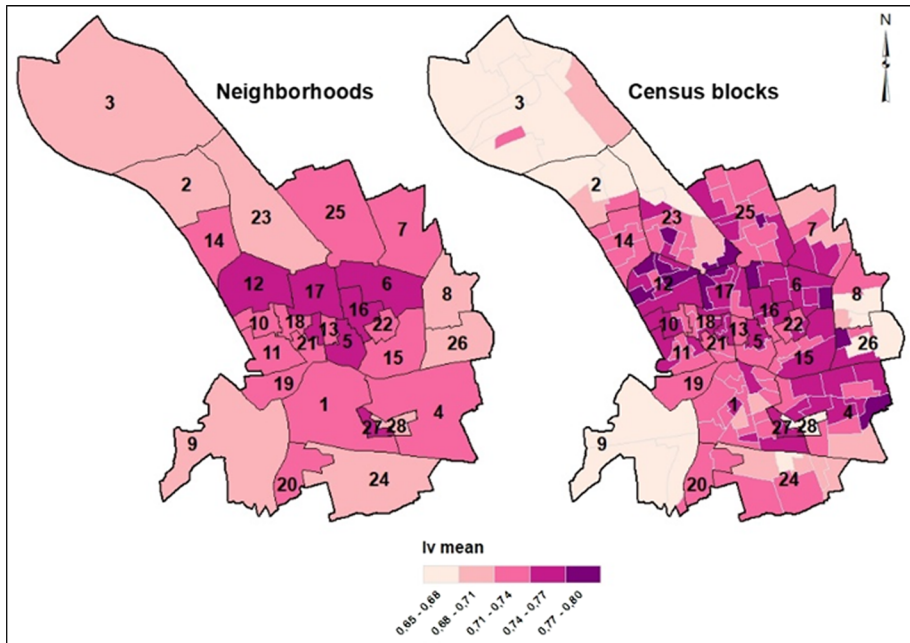


Fig. 15 Distribution of vulnerability in the study area by neighbourhoods and census blocks

6.2 Vulnerability distribution

Figure 15 shows total vulnerability by neighbourhoods and census blocks. The data are presented using the average vulnerability index for each neighbourhood and census tract to make it easier to understand the distributions. Note that the most vulnerable buildings are located closer to the centre of the study area, coinciding with the area with the greatest variety of construction types and corresponding BCTs. As the concentration of buildings is higher in this area, factors such as position within the block, height differences, and mixed uses (presence of soft storeys) have a greater weight in making these areas more vulnerable.

6.3 BCT vulnerability classes: macroseismic scale

The BCTs were adapted (see Fig. 13) to the EMS-98 scale (Grünthal 1998), which defines six vulnerability classes from A (most vulnerable) to F (least vulnerable). Table 16 shows the correlation between each the BCTs identified and the EMS-98 vulnerability classes. Note the low proportion of class A for masonry constructions. This is mainly due to the prevalence of older low-rise residential buildings and more modern medium-rise masonry buildings. The vulnerability class is higher for reinforced concrete constructions as the height and the number of applicable urban modifiers determine building typology, increasing seismic vulnerability.

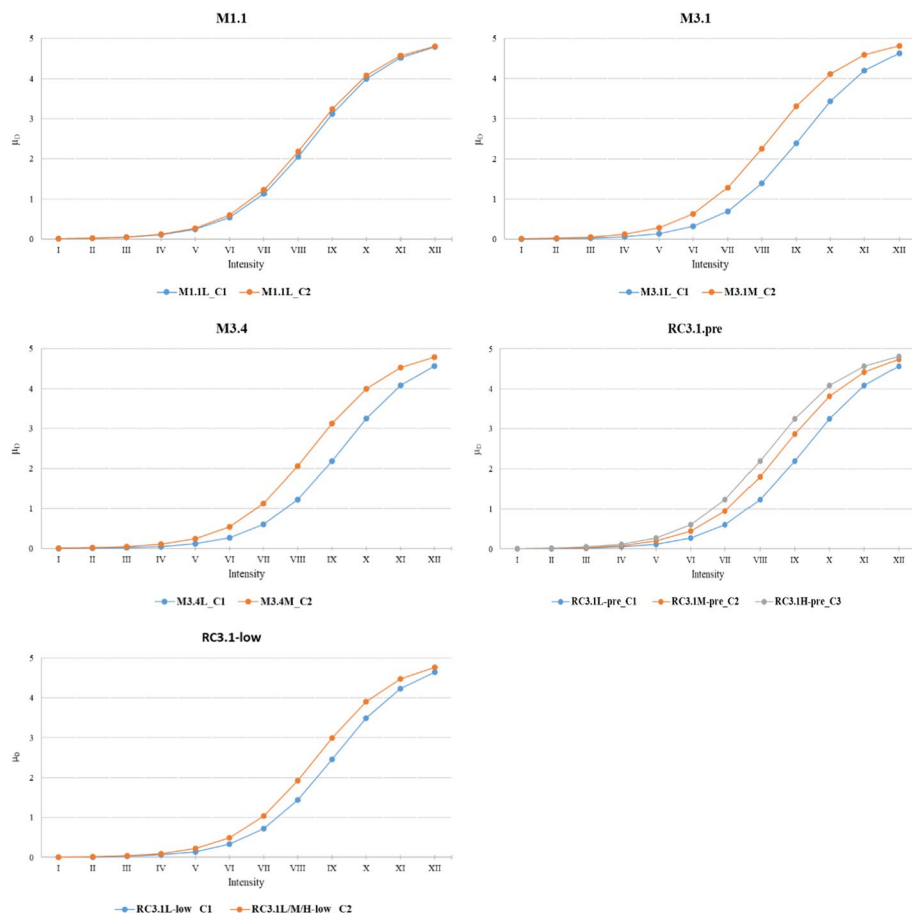


Fig. 16 Vulnerability curves for each BCT identified

6.4 Vulnerability curves applied to BCTs

In this study we have identified different BCTs, calculating the vulnerability for each one of them (Table 13) using the vulnerability index method, being able to evaluate the belonging of each BCT to a vulnerability class (Table 16). Figure 16 shows the vulnerability curves that relate, for each BCT, the mean damage (μD) with the macroseismic intensity and the I_v (Milutinovic and Trendafiloski 2003), defined in Table 13.

Previous research in the Region of Murcia, particularly in the city of Lorca, has assessed building vulnerability using empirical methodologies, identifying the urban modifiers that determine the irregular seismic response in buildings and evaluating their relationship with the damage caused after the 2011 Lorca earthquake (Martínez-Cuevas et al. 2017, 2020; Martínez-Cuevas and Gaspar-Escribano 2016). When comparing our findings in the city of Murcia with those from previous studies in Lorca, we observe similarities in relation to the identified CTs and BTs. Regarding the damage, buildings with greater irregularities (presence of urban modifiers) were found to be uninhabitable in both masonry and reinforced concrete CTs, according to data from the Lorca earthquake. This highlights the increased

vulnerability of buildings due to a greater presence of urban modifiers, leading to a higher likelihood of damage. In addition, this study examines additional typologies that are not present in those studies, which are related to the methodology used to identify the relationship between CTs and modifiers using machine learning techniques.

7 Conclusions

This study addresses the classification of construction types associated with the structural and non-structural components of buildings, following certain characteristic patterns. Our approach is based on the premise that each structure is more or less likely to be subject to unique urban modifiers that impact the seismic behaviour of a building. These patterns correspond to clusters that group the existing urban modifiers by construction type, a point of view that has seldom been studied and could help lay the foundation for automated classification and, subsequent, vulnerability assessments. The methodological approach is divided into three phases and includes specific procedures in each phase to obtain the final results, identifying building cluster typologies (BCT) and grouping the various building characteristics based on similarities and dissimilarities. Each phase and procedure are described in detail to lay a solid foundation that will enable rapid vulnerability assessments in the future, which would help shorten the duration of seismic risk studies and optimize the efficiency of response assessments in the aftermath of a seismic event.

The study was conducted based on a significant number of residential buildings in the metropolitan area of the city of Murcia, Spain. To enhance the resolution of the initial data, obtained from primary sources, the database was refined based on remote field work. The final data were coded following the Global Earthquake Model (GEM) taxonomy. A Two-Step cluster analysis was performed to find natural groupings within the data, and multilayer perceptron neural networks (ANN-MLP) were used to evaluate the degree of representativeness of the BCTs in the study area. Patterns were identified in five construction typologies: three types of masonry buildings and two reinforced concrete construction types. The relevance of the modifiers in each structure and their impact on the vulnerability distribution was analysed to obtain vulnerability index ranges for each BCT.

Based to the proposed methodology, we can assert that:

- Step 1 covers various procedures, from choosing the study area and obtaining baseline data to verifying and enriching the data through remote fieldwork and identifying urban modifiers. These procedures enabled us to obtain an expanded and more robust final database. All this was possible thanks to freely available online satellite and cartographic viewers that can be used to plan remote field work, establishing methodologies based on using initial data gathered from official primary sources and verifying and enriching cartographic data, thus reducing logistical and operational expenses. However, the time spent detecting possible errors in the original data source and incorporating new attributes and findings observed through remote visual inspection can make this a tedious and challenging task that requires training and a thorough familiarity with the various construction specifications, since adequate data collection will impact the robustness of the final database. Step 2 consisted of a machine learning study in which multivariate statistical techniques and ANN were used to define building typologies and assess their degree of representativeness within the study area. Finally, in Step 3,

we defined each BCT, we elaborate the vulnerability curves each BCT and conducted a vulnerability classification based on the EMS-98 scale.

- The uncertainty in the data obtained through remote visual inspection arises from the limitations in satellite, cartographic, and photographic updates of the viewers available online. Another key factor was access to technical construction specifications. This was not contemplated our study, as it would have required an in-depth analysis of technical data sheet. The aim is to propose a methodology that is easy to execute, is combined with remote observation, and relies on the technical expertise of the people in charge of carrying it out.
- Applying the cluster method required several tests to stabilize and increase the silhouette coefficients that indicate whether the cluster groupings are robust. When the original data was entered without doing any field work, the resulting classifications were inadequate, with silhouette values under 0.5, indicating average or poor-quality clusters. Thanks to remote fieldwork, as the volume of revised and new data increased, the quality of the model improved. Eventually, we obtained cluster models with silhouette values greater than the minimum optimal value, indicating that data robustness had been achieved.
- In our study, 60% of the total residential buildings in the cadastral database were inspected through remote field work. This relatively high percentage is due to the highly heterogeneous building stock present in Murcia, composed by buildings that cover a wide period (more than 300 yrs.), with different construction technologies and materials and with several changes in earthquake-resistant regulations. In addition, the relatively complete cadastral database, the low incidence of informal construction and the availability of Google street images helps carrying out virtual field campaigns.
- The application of the proposed methodology in other geographical areas will be contingent on the quality and reliability of the available data, including the availability of accurate cadastral data, as well as information gathered in housing censuses. Furthermore, access to high-quality street images from Google will also play a crucial role. It is important to consider potential errors or inaccuracies in primary data sources when adapting or applying the methodological procedures, especially in developing countries where additional challenges may arise due to the possibility of errors in cadastral data and a higher rate of self-construction, lack of information in google maps or other online viewers. Should remote fieldwork prove infeasible, it may be necessary to conduct in-situ fieldwork, which could entail higher costs in terms of time and logistics.
- Data behaviour and prediction using a general ANN-MLP applied to the entire set of modifiers and the eleven BCTs obtained, yielded error rates of 27.9% for the training sample, 28.8% for the test sample, and 27.4% for the reserve sample. These values indicate that the ANN may have failed to correctly identify overlaps among certain modifiers. It is worth noting that the highest error rates occurred in those clusters with the least number of cases. Based on the above, and since the ANNs are fed historical data during the learning process, it follows that the greater the number of cases for these particular clusters, the higher the prediction accuracy rates will be.
- While it is true that the ANN-MLP achieved high accuracy (72.6% for the reserve sample), due to imbalances between the accuracy and sensitivity metrics, this should not be taken to indicate that the overall ANN model is performing adequately for each CT. Given that each BCT identified is in fact a building typology present in the study area, the F1-score metric was used to find a harmonic mean between precision and sensitivity, and this enabled us to determine the degree of representativeness (Table 7@@) for each BCT within the study area.

- The vulnerability analysis was used to determine the relevance of the various modifiers in increasing the vulnerability indexes for each BCT identified, and these coincided with the expected ranges according to Risk-UE. EMS-98 macroseismic scale concepts were applied to obtain vulnerability classifications (Table 14) after factoring in the impact of the various urban modifiers present.

Based on the results, we obtained the BCT distributions for the urban area analysed at the census tract level. Compared to the BCTs, the vulnerability distributions were found to have higher resolution both in terms of individual building construction specifications and in terms of vulnerability distribution within the urban area of Murcia. Based on our study, obtaining up-to-date building typologies can help assess exposure based on different building characteristics grouped into specific types, and make it easier to perform future vulnerability assessments using predefined vulnerability index ranges and macroseismic class classifications (Table 11, Table 16 and Fig. 16).

In general, the democratization of cartographic and spatial data is a constant challenge that should require periodic updates based on methodologies adapted to the dynamics of urban development. In fact, the international agenda under the Sendai Framework, the New Urban Agenda, and the Sustainable Development Goals promotes periodic updates, data accessibility, data collection, and new technical studies. In our field and according to seismic risk studies, the foregoing leads to increased understanding of risk and enhanced the resilience of urban societies. In this context, the present study is in line with the international agenda and promotes updating data, in addition to presenting and proposing a methodology for classifying and assessing urban physical vulnerability at the local level.

The combination of procedures and techniques employed supports the use of data mining and machine learning techniques in the field of seismic engineering. The methodology can be replicated, in whole or in part, in other urban areas with similar characteristics or adapted to specific urban settings. In addition, seismic vulnerability was mapped to identify the most vulnerable areas with the aim of planning new areas for relocating people who might be left homeless as a result of a seismic event. This information can be very valuable for civil defence because it makes it possible to draw up contingency plans quantifying the number of people living in affected buildings and foreseeing their subsequent relocation. It can also be useful for local and regional governments (the City Council and Autonomous Community) to help them ensure that General Urban Development Plans include public spaces large enough to house all the people who might be left without housing.

As a final remark, it may be useful to assess structural resistance using analytical methods based on mechanical structural response models applied to each BCT identified. Furthermore, in future approaches, it would be interesting to consider social factors and urban fabric as they relate to building typologies. Specifically, the results for the proposed cluster model and vulnerability parameters may be applicable to other cities with similar characteristics. Similarly, in cities with different urban attributes, the methodology provides a basis for extrapolating methodological procedures. Of course, the theoretical basis of the present study can lead to computational developments that may expedite future seismic risk assessments.

Acknowledgements The authors thank the support and contribution of PhD Maribel Jiménez Martínez, PhD M. Carmen Morillo, and PhD Yolanda Torres Fernández. Additionally, we would like to extend our sincere appreciation to the anonymous reviewers for their valuable feedback and input in enhancing the quality of this article. This work is part of the activities of the TERRA Research Group (Geomatics, Natural Hazards and Risks) of Universidad Politécnica de Madrid.

Funding Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature. The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Declarations

Conflict of interest The authors declare that does not exist an interest conflict.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Angeletti P, Bellina A, Guagenti E, Moretti A, Petrini V (1988) Comparison between vulnerability assessment and damage index, some results. In: Proceedings of the 9th world conference on earthquake engineering. Tokyo-Kyoto, Japan, pp 181–186.
- Ballestar MT, Grau-Carles P, Sainz J (2018) Customer segmentation in e-commerce: applications to the cashback business model. *J Bus Res* 88:407–414. <https://doi.org/10.1016/j.jbusres.2017.11.047>
- Barbat AH, Lantada N, Pujades L, Carreño L, Cardona OD (1998) Evaluación del riesgo sísmico de Barcelona. SÍSMICA 2004—6° Congreso Nacional de Sismología e Ingeniería Sísmica
- Basaglia A, Aprile A, Spacone E, Pilla F (2018) Performance-based seismic risk assessment of urban systems. *Int J Archit Herit* 12(7–8):1131–1149. <https://doi.org/10.1080/15583058.2018.1503371>
- Basset-Salom L, Guardiola-Villora A (2014) Seismic performance of masonry residential buildings in Lorca's city centre, after the 11th May 2011 earthquake. *Bull Earthq Eng* 12:2027–2048. <https://doi.org/10.1007/s10518-013-9559-8>
- Benedetti D, Petrini V (1984) Sulla vulnerabilità sismica di edifici in muratura: proposte di un método di valutazione. *L'ind Delle Costru* 49:66–78
- Brzev S, Scawthorn C, Charleson AW, Allen L, Greene M, Jaiswal K, Silva V (2013) GEM building taxonomy version 2.0, GEM Technical Report 2013-02 V1.0.0. GEM Foundation, Pavia, Italy. DOI: <https://doi.org/10.13117/GEM.EXP-MOD.TR2013.02>.
- Carreño ML (2006) Técnicas innovadoras para la evaluación del riesgo sísmico y su gestión en centros urbanos: acciones ex ante y ex post. Tesis doctoral. Universidad Politécnica de Cataluña, Barcelona
- Charleson A (2008) Seismic design for architects. Architectural Press, UK, p 281
- Chiu T, Fang D, Chen J, Wang Y, Jeris C (2001) A robust and scalable clustering algorithm for mixed type attributes in large database environment. In Proceedings of the seventh ACM SIGKDD international conference on knowledge discovery and data mining (pp 263–268).
- Cybenko G (1989) Approximation by superpositions of a sigmoidal function. *Math Control Signal Syst* 2:303–314. <https://doi.org/10.1007/BF02551274>
- Dell'Acqua F, Gamba P, Jaiswal K (2013) Spatial aspects of building and population exposure data and their implications for global earthquake exposure modeling. *Nat Hazards* 68:1291–1309. <https://doi.org/10.1007/s11069-012-0241-2>
- Esteghamati MZ, Lee J, Musetich M, Flint MM (2020) INSSEPT: an open-source relational database of seismic performance estimation to aid with early design of buildings. *Earthq Spectra* 36(4):2177–2197. <https://doi.org/10.1177/8755293020919857>
- FEMA-154 (1988) Rapid visual screening of buildings for potential seismic hazards: a handbook, FEMA 154, Second Edition, C. Rojahn, Principal Investigator, C. Scawthorn, co-Principal Investigator, prepared by the Applied Technology Council, Redwood City, CA, (ATC-21), for the Federal Emergency Management Agency, Washington, D.C., USA

- Ferreira TM, Estêvão J, Maio R, Vicente R (2020) The use of artificial neural networks to estimate seismic damage and derive vulnerability functions for traditional masonry. *Front Struct Civil Eng* 14:609–622. <https://doi.org/10.1007/s11709-020-0623-6>
- Fraley C, Raftery AE (1998) How many clusters? Which clustering method? answers via model-based cluster analysis. *Comput J* 41(8):578–588. <https://doi.org/10.1093/comjnl/41.8.578>
- Gaspar-Escribano JM, Rivas-Medina A, Parra H, Cabañas L, Benito B, Barajas SR, Solares JM (2015) Uncertainty assessment for the seismic hazard map of Spain. *Eng Geol* 199:62–73. <https://doi.org/10.1016/j.enggeo.2015.10.001>
- Gómez-Martínez F, Pérez-García A, De Luca F, Verderame GM (2015) Comportamiento de los edificios de HA con tabiquería durante el sismo de Lorca de 2011: aplicación del método FAST. *Info Constr* 67(537):e065. <https://doi.org/10.3989/ic.12.110>
- Goodfellow I, Bengio Y, Courville A (2016) *Deep learning*. MIT Press, Cambridge
- Grünthal G (1998) European macroseismic Scale 1998. European Seismological Commission (ESC)
- Guevara LT, García LE (2001) La columna corta o la columna cautiva. *Revista Tecnología y Construcción*, N° 17-1, 2001, enero-abril 2001. Instituto de desarrollo experimental de la construcción, Universidad de Venezuela e Instituto de Investigaciones de la Facultad de Arquitectura de la Universidad de Zulia, Caracas
- Hazus MH (2003) Multi-hazard loss estimation methodology. Technical manual. Federal Emergency Management Agency (FEMA), Washington DC, p 712
- Hinton G, Salakhutdinov R (2012) Reducing the dimensionality of data with neural networks. *Science* 313(5786):504–507
- IGN-UPM (2013) Actualización de Mapas de Peligrosidad Sísmica de España 2012. Editorial Centro Nacional de Información Geográfica, Madrid, p 267
- Kaufman L, Rousseeuw P (1990) Finding groups in data: an introduction to cluster analysis. John Wiley, New York. <https://doi.org/10.2307/2532178>
- Lagomarsino S, Giovinazzi S (2006) Macroseismic and mechanical models for the vulnerability and damage assessment of current buildings. *Bull Earthq Eng* 4:415–443. <https://doi.org/10.1007/s10518-006-9024-z>
- Lang DH, Kumar A, Sulaymanov S, Meslem A (2017) Building typology classification and earthquake vulnerability scale of Central and South Asian building stock. *J Build Eng* 15:261–277. <https://doi.org/10.1016/j.jobe.2017.11.022>
- Lantada N, Irizarry J, Barbat AH et al (2010) Seismic hazard and risk scenarios for Barcelona, Spain, using the risk-UE vulnerability index method. *Bull Earthq Eng* 8:201–229. <https://doi.org/10.1007/s10518-009-9148-z>
- Lantada N (2007) Evaluación del riesgo sísmico mediante métodos avanzados y técnicas GIS. Aplicación a la ciudad de Barcelona. Tesis Doctoral. Dpto. Ingeniería del Terreno, Cartográfica y Geofísica. Universidad Politécnica de Cataluña, Barcelona, p 325.
- LeCun Y, Bengio Y, Hinton G (2015) Deep learning. *Nature* 521(7553):436–444
- Martínez-Cuevas S, Gaspar-Escribano JM (2016) Reassessment of intensity estimates from vulnerability and damage distributions: the 2011 Lorca earthquake. *Bull Earthq Eng* 14:2679–2703. <https://doi.org/10.1007/s10518-016-9913-8>
- Martínez-Cuevas S, Benito MB, Cervera J, Morillo MC, Luna M (2017) Urban modifiers of seismic vulnerability aimed at Urban Zoning regulations. *Bull Earthq Eng* 15:4719–4750. <https://doi.org/10.1007/s10518-017-0162-2>
- Martínez-Cuevas S, Morillo Balsera MC, Benito B, Torres Y, Gaspar-Escribano J, Staller A, García-Aranda C (2020) Assessing building habitability after an earthquake using building typology and damage grade. *Appl Lorca Spain J Earthq Eng* 26(7):3417–3439. <https://doi.org/10.1080/13632469.2020.1802370>
- McCulloch WS, Pitts W (1943) a logical calculus of the ideas immanent in nervous activity. *Bull Math Biophys* 5:115–133
- Meyers-Angulo E, Martínez-Cuevas S, Gaspar-Escribano JM (2023) Catalogue of residential buildings and classification according to seismic vulnerability in the city of Murcia (Spain). <https://doi.org/10.21950/HLF2US>
- Milutinovic ZV, Trendafiloski GS (2003) RISK-UE, An advanced approach to earthquake risk scenarios with applications to different European towns. Report to WP4: vulnerability of current buildings, p 109
- Mouroux P, Bertrand E, Bour M, Le Brun B, Depinois S, Masure P (2004) The European RISK-UE project: an advanced approach to earthquake risk scenarios. In: *Proceedings of the 13th World Conference on Earthquake Engineering*.

- Mouroux P, Le Brun B (2006) Presentation of RISK-UE project. *Bull Earthq Eng* 4(4):323–339. <https://doi.org/10.1007/s10518-006-9020-3>
- Norusis MJ (2011) IBM SPSS statistics 19 guide to data analysis. Prentice Hall, Upper Saddle River
- Norusis M (2008) Cluster analysis, Chapter 16, pp: 361–391. SPSS 17.0 statistical procedures companion, Marija Norusis
- Oterino BB, Medina AR, Escalante MP, Quirós LE, Barajas S, Escribano JG, Rey RH (2014) Servicio de actualización del análisis de riesgo sísmico en la Región de Murcia (RISMUR)
- Park, J., & Sandberg, I. W. (1991). Universal approximation using radial-basis-function networks. *Neural Computation*, 3(2), 246–257. <https://doi.org/10.1162/neco.1991.3.2.246>
- Pittore M, Haas M, Megalooikonomou KG (2018) Risk-oriented, bottom-up modeling of building portfolios with faceted taxonomies. *Front Built Environ* 4:41. <https://doi.org/10.3389/fbuil.2018.00041>
- Ródenas JL, García-Ayllón S, Tomás A (2018) Estimation of the buildings seismic vulnerability: a methodological proposal for planning ante-earthquake scenarios in urban areas. *Appl Sci* 8(7):1208. <https://doi.org/10.3390/app8071208>
- Rousseeuw P (1987) Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *J Comput Appl Math* 20:53–65. [https://doi.org/10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7)
- Rubio-Hurtado MJ, Baños RV (2017) El análisis de conglomerados bietápico o en dos fases con SPSS. *REIRE Rev D'innov Rec Educ* 10(1):118–126
- Rumelhart DE, Hinton GE, Williams RJ (1986) Learning internal representations by error propagation. In: Rumelhart DE, McClelland JL (eds) *Parallel distributed processing*. MIT Press, Cambridge, pp 318–362
- Santa María H, Hube MA, Rivera F, Yepes-Estrada C, Valcárcel JA (2017) Development of national and local exposure models of residential structures in Chile. *Nat Hazards* 86:55–79. <https://doi.org/10.1007/s11069-016-2518-3>
- Schwarz G (1978) Estimating the dimension of a model. *Ann Statist* 6(2):461–464. <https://doi.org/10.1214/aos/1176344136>
- SERA (2017) Seismology and earthquake engineering research. Infraestructura Alliance for Europe. <http://www.sera-eu.org/en/home/>. Accessed 02 Feb 2022.
- Silva V, Henshaw P, Huyck C, O'Hara M (2018) D5-Final Report. GEM Technical Report 2018-05, GEM Foundation, Pavia, Italy.
- Stefanini L, Badini L, Mochi G, Predari G, Ferrante A (2022) Neural networks for the rapid seismic assessment of existing moment-frame RC buildings. *Int J Disaster Risk Reduct* 67:102677. <https://doi.org/10.1016/j.ijdrr.2021.102677>
- SYNER-G D1.1 (2009) SYNER-G work plan. SYNER-G project: systemic seismic vulnerability and risk analysis for buildings, lifeline networks and infrastructures safety gain. WP1-Project coordination and management. <http://www.vce.at/SYNER-G/files/dissemination/deliverables.html>. Accessed 15 Jan 2022.
- Tang Q, Dang J, Cui Y, Wang X, Jia J (2021) Machine learning-based fast seismic risk assessment of building structures. *J Earthq Eng*. <https://doi.org/10.1080/13632469.2021.1987354>
- Theodoridis S, Koutroumbas K (2001) Pattern recognition and neural networks. In: Paliouras G, Karkaletsis V, Spyropoulos CD (eds) *Machine learning and its applications*. ACAI 1999. Lecture Notes in Computer Science, vol 2049. Springer, Berlin, Heidelberg. https://doi.org/10.1007/3-540-44673-7_8.
- Torres Y, Arranz J, Gaspar-Escribano J, Haghi A, Martínez-Cuevas S, Benito B, Ojeda J (2019) Integration of LiDAR and multispectral images for rapid exposure and earthquake vulnerability estimation. Application in Lorca, Spain. *Int J Appl Earth Obs Geoinf* 81:161–175. <https://doi.org/10.1016/j.jag.2019.05.015>
- Udwin D (1981) *Introductory spatial analysis*. Ed. Methuen, London, p 212
- Vazirizade SM, Nozhati S, Mostafa AZ (2017) Seismic reliability assessment of structures using artificial neural network. *J Build Eng* 11:230–235. <https://doi.org/10.1016/j.jobbe.2017.04.001>
- Yepes-Estrada C, Silva V, Valcárcel J, Acevedo AB, Tarque N, Hube MA, Coronel G, Santa María H (2017) Modeling the residential building inventory in South America for seismic risk assessment. *Earthq Spectra* 33(1):299–322. <https://doi.org/10.1193/101915eqs155dp>
- Zhang T, Ramakrishnan R, Livny M (1996) BIRCH: an efficient data clustering method for very large. *ACM SIGMOD Rec* 25(2):103–114. <https://doi.org/10.1145/235968.233324>