Bulletin of Earthquake Engineering

Cluster Analysis for Informing Vulnerability Assessment of Masonry Churches to Natural Hazards

--Manuscript Draft--

Manuscript Number:	BEEE-D-24-00726R2
Full Title:	Cluster Analysis for Informing Vulnerability Assessment of Masonry Churches to Natural Hazards
Article Type:	Original Article
Keywords:	Churches archetypes; Cluster analysis; Structural assessment; Seismic vulnerability; Landslides
Corresponding Author:	Federica Del Carlo University of Pisa School of Engineering: Universita degli studi di Pisa Scuola di Ingegneria Pisa, ITALY
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	University of Pisa School of Engineering: Universita degli studi di Pisa Scuola di Ingegneria
Corresponding Author's Secondary Institution:	
First Author:	Federica Del Carlo
First Author Secondary Information:	
Order of Authors:	Federica Del Carlo
	Silvia Caprili
	Tiago Miguel Ferreira
	Pere Roca
	Marco Uzielli
Order of Authors Secondary Information:	
Funding Information:	
Abstract:	According to a census by the Catholic Church, Italy's territory hosts more than sixty thousand buildings of worship. Most of these buildings were built between the first and the nineteenth century A.D., with a load-bearing masonry structure that proved to be particularly prone to damage due to natural hazards. This investigation explores the use of clustering algorithms to identify and cluster typologies of buildings and archetypes. The aim is to define statistical models for the geometric and mechanical properties, to delineate a set of reference structures representative of the whole building stock, and finally select 'indicator attributes' that can be used in developing seismic and landslide vulnerability indicators. The proposed methodology is applied to a specific portfolio of seventy-one churches in the north-western area of the Tuscany region (Italy). The main geometric and mechanical features of the churches included in the portfolio are gathered using a new simplified Rapid Visual Survey form. A procedure is then proposed to define representative archetypes using three well-known clustering algorithms (K-Means, Gaussian Mixture Models, and Kernel-density). When analysed together, the identified archetypes can portray the variability of the geometric and mechanical properties in the selected portfolio, constituting a basis for developing new vulnerability models.
Suggested Reviewers:	Javier Ortega Heras javier.ortega@csic.es
	Rafael Ramirez Eudave rafael.ramirez@wsl.ch

Response to Reviewers:	A file has been attached with the comment rebuttal.

Cluster Analysis for Informing Vulnerability Assessment of Masonry Churches to Natural Hazards

Federica Del Carlo[1,3][0009-0005-6912-1884]*, **Silvia Caprili**[1][0000-0001-6650-4867], **Tiago Miguel Ferreira**[2][0000-0001-6454-7927], **Pere Roca**[3][0000-0001-5400-5817], and **Marco Uzielli**[4][0000-0001-7755-6144]

[1]*Department of Civil and Industrial Engineering, University of Pisa, Largo L. Lazzarino 1, 56122 Pisa, Italy*

* e-mail: federica.delcarlo@phd.unipi.it

[2] *University of the West of England (UWE Bristol). Frenchay Campus, Coldharbour Lane, Bristol BS16 1QY, United Kingdom*

[3]*Technical University of Catalonia. Jordi Girona 1-3, 08034, Barcelona, Spain*

[4]*University of Florence. Via di Santa Marta 3, 50139, Florence, Italy*

Abstract

According to a census by the Catholic Church, Italy's territory hosts more than sixty thousand buildings of worship. Most of these buildings were built between the first and the nineteenth century A.D., with a load-bearing masonry structure that proved to be particularly prone to damage due to natural hazards. This investigation explores the use of clustering algorithms to identify and cluster typologies of buildings and archetypes. The aim is to define statistical models for the geometric and mechanical properties, to delineate a set of reference structures representative of the whole building stock, and finally select 'indicator attributes' that can be used in developing seismic and landslide vulnerability indicators. The proposed methodology is applied to a specific portfolio of seventy-one churches in the north-western area of the Tuscany region (Italy). The main geometric and mechanical features of the churches included in the portfolio are gathered using a new simplified Rapid Visual Survey form. A procedure is then proposed to define representative archetypes using three well-known clustering algorithms (K-Means, Gaussian Mixture Models, and Kernel-density). When analysed together, the identified archetypes can portray the variability of the geometric and mechanical properties in the selected portfolio, constituting a basis for developing new vulnerability models.

Keywords: Churches archetypes, Cluster analysis, Structural assessment, Seismic vulnerability, Landslides.

1. Introduction and research aim

Religious buildings hold great cultural and historical relevance in the complex Italian architectural scenario. The Catholic Church alone owns around sixty thousand buildings of worship classified as heritage buildings throughout the Italian territory (BeWeB). Their geometrical and construction peculiarities (such as their long spans and heights), together with the fact that they were typically designed to gravity loads only, make these buildings prone to suffer extensive damages when subjected to exceptional events, as natural hazards. In the past, great attention has been given at investigating the vulnerability of these structures to natural hazards, as seismic events (Orduña et al. 2008; Cundari et al. 2017; Penna et al. 2019; Uva et al. 2019; Canuti et al. 2021) or soil displacements (Ferrero et al. 2021; Bertolin and Sesana 2023).

Although the state of research regarding the assessment of vulnerability of religious buildings is currently at different maturity stages depending on the hazard considered, two types of approaches to the problem can be identified in the literature. In the first approach, based on a case-by-case evaluation (De Matteis and Mazzolani 2010; Fortunato et al. 2017; Grazzini et al. 2019), detailed models and numerical simulations are used to assess the vulnerability of specific case studies. In these cases, the structural response of the buildings is usually evaluated through linear and nonlinear analyses on three-dimensional models (Betti and Vignoli 2008; Clementi et al. 2018; Noel et al. 2019; Formisano et al. 2022) reproducing, with sufficient reliability, their shape, geometry, material properties and boundary conditions. This approach is reliable and efficient when an in-depth evaluation of the current structural performance of the building is required and very detailed and reliable knowledge of materials and geometry are available. However, it is computationally intensive and, therefore, not viable to use to assess a large number of buildings. A second possible approach, particularly convenient when dealing with a large number of buildings, consists of identifying a suitable number of archetypes or building typologies representative of the portfolio of buildings under consideration (Palazzi et al. 2019; Sferrazza Papa et al. 2019; Sorrentino et al. 2019). Based on the thorough analysis of one or various representative typologies, it is then possible to carry out a relatively swift but meaningful and reliable evaluation of a large number of buildings of the same typology (i.e., with similar construction and structural characteristics). At a large-territorial scale, this approach allows the comparison of the expected vulnerability of a large set of buildings, thereby providing stakeholders with a tool to inform retrofit or maintenance measures (Marotta et al. 2018; Fuentes et al. 2021;

Pirchio et al. 2021).

Grouping buildings into typologies is undoubtedly advantageous when a large-scale overview of the vulnerability of the heritage buildings is required, but it involves opting not to delve into the detailed knowledge of specific technical features. The decision on which information to prioritise and collect and which to omit when cataloguing heritage for efficiency's sake comes with various challenges when dealing with ordinary buildings but involves an even higher level of complexity when heritage buildings are considered. Different vulnerability derivation taxonomies have been used in the past to assess the seismic vulnerability of ordinary and heritage buildings. Among the most well-known ones, the European Macroseismic Scale (EMS-98) (Grünthal and Levret 1998), an evolution of the MSK-64 scale (Medvedev–Sponheuer–Karnik scale) (Medvedev et al. 1965), group buildings into vulnerability classes, which range from A to D, based on the building's materials and load-bearing system. The ATC-13 (McCormack and Rad 1997), FEMA 154 (Agency (US) 2017) and HAZUS (Schneider and Schauer 2006) methodologies provided a new concept of a building classification system based on structural type, height, and local design code and construction practices calibrated to the United States construction portfolio. This led to a broader number of classes encompassing different building features in a more accurate way. For the European territory, examples of similar classification systems include the RISK-EU project (Mouroux and Le Brun 2008) and the Syner-G taxonomy (Crowley et al. 2012). Attempts have also been made to codify the heritage buildings on a worldwide scale. The Earthquake Engineering Research Institute (EERI) and the International Association for Earthquake Engineering (IAEE) have taken a big step forward in understanding and summarising worldwide construction types in a unique building inventory in the World Housing Encyclopaedia (WHE) project (BRZEV et al. 2004). Even though existing taxonomies are predominantly seismic vulnerability assessment-oriented, considerable efforts have been made in the recent past to enable the large-scale vulnerability assessment of buildings to other hazards (Santos et al. 2013; Cantarino et al. 2014; Guillard-Gonçalves et al. 2016; Del Zoppo et al. 2022).

One of the most recognised and adopted seismic vulnerability-oriented taxonomies for historical-religious buildings is the methodology based on twenty-eight damage mechanisms proposed by (Lagomarsino and Podestà 2004; Lagomarsino 2006) and incorporated in the Italian Guidelines for the seismic risk assessment mitigation of cultural heritage (DCP 2018). Despite its validity, the proposed method is still relatively time-consuming when applied in a large-scale context. Further, it is limited to seismic vulnerability (Sisti et al. 2023). Attempts to modify the method, moving to simpler, faster and less expensive procedures, have been recently conducted: for example, simplified approaches to carry out first screenings of risk have been proposed by (Lopez et al. 2019; Betti et al. 2021). On the other hand, (Sevieri et al. 2020; Arrighi et al. 2022; D'Ayala et al.) have proposed multi-hazard risk assessment frameworks specifically for cultural heritage buildings. Addressing multiple hazards entails a range of additional challenges. Hazard characteristics differ in nature, return period, and intensity; as a result, the methods used to analyse and quantify them are also heterogeneous (Zschau 2017). Moreover, the impacts hazards have on the elements at risk depend on different series of attributes, i.e. building geometric characteristics and material properties, that are different from hazard to hazard, leading to differences between vulnerability analysis methods (Kappes et al. 2012a; Julià and Ferreira 2021). Multi-hazard approaches may refer to a situation where more than one hazard is considered, but no hazard interactions are undertaken or to scenarios in which hazards are explicitly correlated, introducing cascading effects. The latter case, termed "multilayer single-hazard" analysis (Zschau 2017), is the most common approach considered, especially when addressing buildings. For example, reference (Fleming et al. 2016) proposed a framework that allows the estimation of the total risk arising from multiple independent hazards affecting an area by the harmonised comparison of risk curves from multiple natural hazards applied to the city of Cologne (Germany). Romão et al. (Romão et al. 2016) developed a methodology to perform the qualitative risk assessment intended to be used as a screening procedure for the preliminary assessment and identification of built and immovable cultural heritage assets. The procedure is based on a set of structured assessment flowcharts that address the likelihood of the hazard, the vulnerability of the asset to the hazard, and the capacity to recover from the event. Kappes et al. (Kappes et al. 2012b) presented a GIS-based approach that allows for assessing hazard-specific physical vulnerability towards multiple hazards. The study addressed hazard interactions in terms of spatial and temporal coincidence, but it did not consider the possible influence of one hazard on the others.

The present study is part of a broader research whose objective is to develop a multilayer, single-hazard, spatial-oriented approach for the vulnerability analysis of existing religious buildings to seismic and landslide events. The overarching objective is to establish a quantitative risk analysis approach to be applied at a territorial scale, where risk will parameterise the consequence of the identified hazards (namely, seismic and landslide hazards) on a set of religious buildings as the product of the hazard (i.e., the probability of occurrence of these events), the vulnerability of the buildings (grouped by typology) and their exposure. Vulnerability is estimated by assessing, either quantitatively or qualitatively, which geometric or material attributes have a greater influence on the structural performance in the face of the identified risk agents. It is important to note that the aim is not to study the interaction between seismic and landslide hazards, but rather to develop an indicator that facilitates comparability between analysis procedures for different hazards. While the broader research context encompasses both seismic and landslide vulnerabilities, this specific study focuses on validating the use of clustering techniques to group historical masonry churches into archetypes based on their geometric and material characteristics. Historical religious buildings are unique in their cultural, architectural, and structural characteristics. However, for the purposes of large-scale screening and preliminary vulnerability analysis, it is practical to group them based on common characteristics. The resulting archetypes are indeed to serve as a foundation for future vulnerability assessments, enabling the analysis of a reduced number of representative buildings that encapsulate the variability of the original dataset.

To achieve this, statistical analysis was employed to identify recurrent church typologies, with clustering techniques applied to determine archetypes within each typology. In this study, "church typologies" are defined based on in-plane layouts (e.g., one-nave, three-nave). Within each typology, clustering techniques were used to further identify "church archetypes" - representative structures that reflect the variability in geometric and material characteristics. Three clustering algorithms (K-Means, Gaussian Mixture Models, and Kernel-density) were investigated for grouping the geometrical attributes and identifying the representative archetypes. By grouping buildings into representative archetypes, this study contributes to a systematic approach to vulnerability analysis, laying the groundwork for informed decision-making in risk mitigation strategies. It should be noted, however, that while this study identifies archetypes as representative subsets of historical churches, further analysis is required to correlate these archetypes with specific failure mechanisms and damage patterns.

The proposed procedure is applied, as a proof of concept, to a portfolio of seventy-one churches located in the Tuscany region in West-central Italy. Information concerning geometric and mechanical properties was collected by using a simplified Rapid Visual Survey (RVS). The form was designed to be able to expeditiously recover the relevant attributes through both in-situ and remote investigations, reducing efforts when the procedure is applied at a large-scale/territorial level. The flexibility of the RVS form and the clustering techniques ensures that the approach can be extended to similar studies in different contexts, allowing other researchers to apply and adapt the methodology to their own datasets.

2. Religious building stock over the selected area

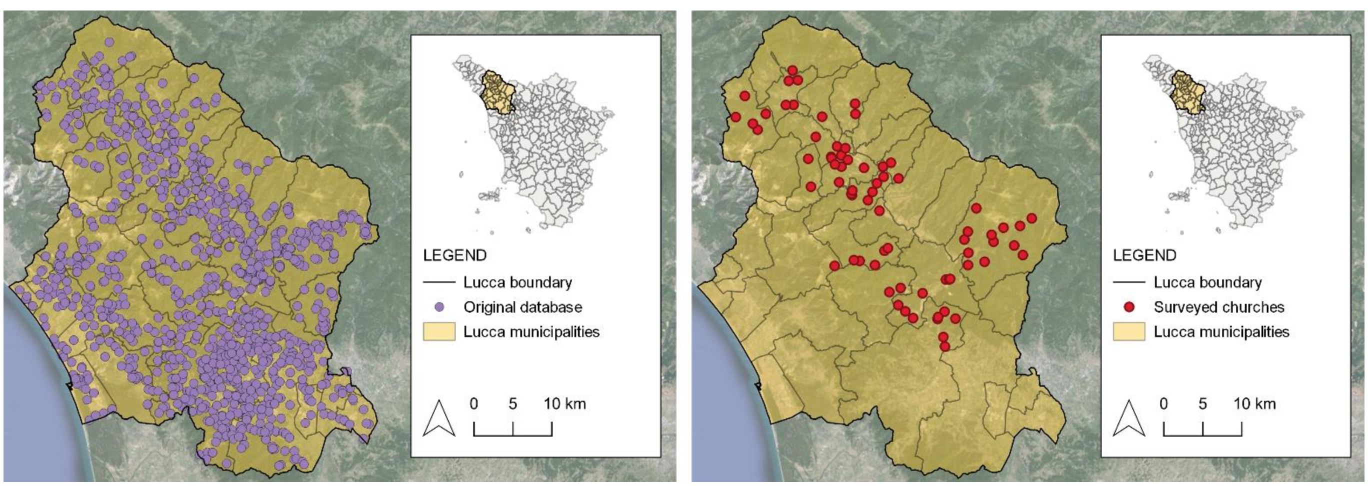
The Tuscany region covers a total area of around 23,000 $km^{2.}$ It is administratively divided into ten provinces, each of which is further subdivided into smaller municipalities (Firenze, Arezzo, Siena, Grosseto, Pisa, Livorno, Massa-Carrara, Pistoia, Prato and Lucca). The region is made up of various morphologies, but it is primarily distinguished by gently undulating hills and flat plains surrounded by mountainous ridges. Tuscany's lowlands are either coastal plains along the Tyrrhenian Sea or inner valleys like the Arno River. The area selected for the present research work is located in the Northern part of the region, more precisely in the Lucca district. It extends from the middle valley of the River Serchio to the Lunigiana-Garfagnana region, which includes a portion of the Northern Apennines. The area is hilly and mountainous and includes cities, industrial hubs connected by primary roads, and small settlements accessible by subsidiary roads. Earthquakes and slope instabilities are the two main hazards which have historically affected the area.

The several medium-to-high energy seismic events that occurred in the area throughout the last millennium (Camassi e Stucchi 1996; Gruppo di Lavoro CPTI 1999; Solarino, Ferretti, e Eva 2002) proved the complex distribution of the seismic activity in the studied area. The 1920 earthquake, which caused significant damage to the upper Serchio Valley, was the last well-documented, high-magnitude earthquake (M_w=6.5). With respect to the current Italian Technical Standards for Constructions (D.M.17.01.2018. 2018), the area falls into seismic zones 2 and 3, which are defined as areas with a seismic action value in terms of Peak Ground Acceleration (PGA) equal to 0.25g and 0.15g, respectively.

The area is also susceptible to slope instability due to the Serchio basin's geological, geomorphological and climatic characteristics (D'Amato Avanzi et al. 1993; Lo Presti et al. 2008). In addition to the National seismic zonation (OPCM 28 aprile 2006 n. 3519 2006) and slope instability inventories available at the regional level (Regione Toscana), the study involved the collection and analysis of existing seismic micro-zonation studies, which provide useful information about the patterns of ground motion amplification, liquefaction, surface fault ruptures and earthquake-induced slope instability. Reference (Karwacka et al. 2019) contributed a comprehensive database of all the religious buildings in the selected was available (Figure 1a).

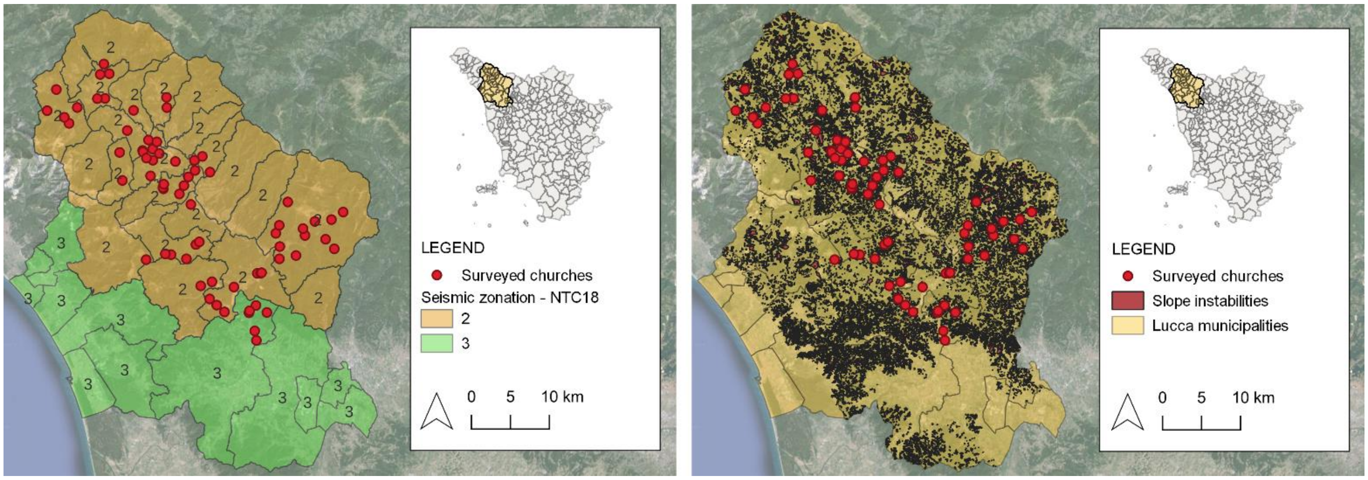
By means of the QGIS software, a sample of 71 churches (Figure 1b) was selected by mapping the National seismic hazard (Figure 2a) and the inventory of Tuscan landslide occurrences (Figure 2b) against the location of the listed historic religious buildings in the region. In doing so, it was possible to consider both the site characteristics and the location of the churches. The buildings in the dataset, which belong to the Romanesque architectural style, were constructed using unreinforced masonry (URM) techniques typical of their period and region. Reinforced concrete (RC) religious buildings located within the examined area were excluded from this study, as the focus is solely on URM religious structures. Nonetheless, the proposed methodology was designed and formulated to be applicable to RC buildings, provided a sufficiently large and representative database is available.

During the sample selection, preference was made for churches that are not plastered so that the wall texture was visible. Churches in very bad conservation states or abandoned were discarded due to the impossibility of accessing those structures safely. Churches that are part of monasteries or other religious complexes were also disregarded, as were the churches inserted within urban meshes, since their behaviour is strongly conditioned by their boundary conditions, an aspect that cannot be duly accounted for when using simplified assessment methodologies.



a) b)

Figure 1. Superimposition of (a) Identification of churches located in the selected area; and (b) Set of churches identified as samples.



a) b)

Figure 2. Superimposition of (a) Identified churches and Italian seismic zonation; and (b) identified churches and slope instability phenomena.

3. Characterisation of the churches' portfolio

The analysis of the churches' portfolio was conducted with the primary goal of defining the main church typologies and identifying the most suitable statistical models for characterizing their geometric and mechanical attributes. This procedure involved two key steps: *(i)* the in-situ inspection of the selected 71 religious buildings using an ad-hoc Rapid Visual Survey (RVS) form and *(ii)* the data gathering and subsequent analysis to identify the main typologies present in the building dataset.

3.1 Data collection

As previously mentioned, the in-situ surveys were performed adopting an ad-hoc RVS form (Del Carlo et al. 2023). The description of this form, together with information about assumptions and the criteria adopted to determine relevant attributes, are presented in detail by (Del Carlo et al. 2023). The selection of attributes for the RVS form was informed by a comprehensive review of existing literature and established vulnerability assessment methodologies. Particular attention was given to attributes commonly recognised as influential in determining the vulnerability of historical masonry buildings to seismic and landslide hazards. These attributes include factors such as wall thickness, plan and elevation geometry, and material properties, which are frequently cited in frameworks such as the EMS-98 (Grünthal and Levret 1998), the Italian Guidelines for Damage Assessment to Cultural Heritage (PCM-DPC Mibac 2006), and other multi-hazard assessment methodologies (Kappes et al. 2012b). Finally, the attributes were selected to be compatible with a fast screening of a considerable number of churches, which could be accomplished through on-site surveys or the collection of data and/or information through satellite imagery and census data. The building taxonomy, therefore, accounts for three main sections (i.e. general information, building layout, and façade layout) and nine sub-sections used to provide an overall description of the building. Specifically, the general data section includes details regarding the surveying process (date, location, accessibility, etc.), the material properties of the buildings, the type of horizontal load-bearing system, roof typology, and the relevant interventions that various components of the church may have undergone.

The building layout section of the RVS form encompasses information related to the building's layout (one nave, three naves, Latin-cross plan, etc.), as well as plan and elevation geometry and dimensions, and the presence of valuable decorative elements. The façade layout section gathers data on materials, façade type, quality and state of conservation, the presence or absence of a bell tower, construction materials, and other related features. Table 1 provides an overview of the main sections of the RVS form. While the form captures a broad range of attributes relevant to the vulnerability of the churches — encompassing geometric, material, and maintenance characteristics — in the subsequent statistical and cluster analysis, only a limited number of geometric attributes were considered. This decision was based on their direct measurability and statistical comparability across the dataset. Non-geometric attributes, such as wood quality and maintenance level, were collected but are reserved for potential future analyses. It is important to note that the identified attributes are not ranked in any particular order, meaning they are not ranked according to their relevance to the structural response of the building to seismic or landslide hazards.

Table 1. Church attributes selected as relevant for multi-hazard risk assessment.

Attribute	**Description**	**Hazard addressed**
General Data		
Survey data	Day, month, year	Hazard independent
Loci information	Province, municipality, sub-municipality (e.g., the Italian "Provincia", "Comune", and "Frazione")	Hazard independent
Geographic Information	Road type providing access to the building, position within the building context (e.g., isolated, attached to other buildings, at the corner of a block of buildings), GPS coordinates	Relevant to both seismic and landslide hazard
Vertical bearing structure	Vertical and lateral load resisting system, construction material, quality and maintenance level	Mainly relevant to seismic hazard
Roof structure	Roof shape, material, structural system, quality and maintenance level	Mainly relevant to seismic hazard
Interior ceiling	Ceiling shape, material and structure system, quality and maintenance level	Mainly relevant to seismic hazard

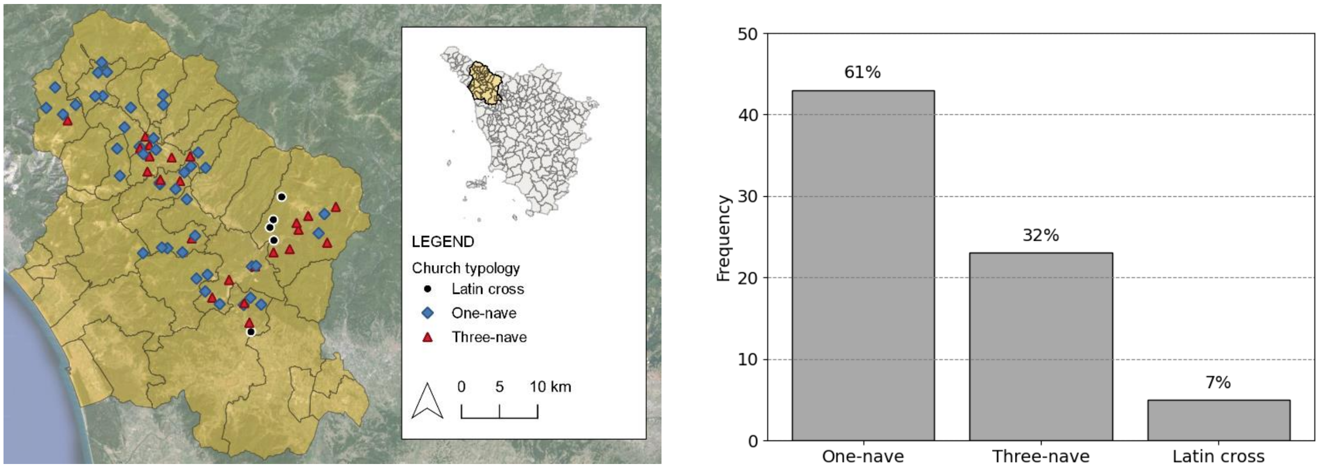
Intervention	Relevant interventions that various components of the church may have undergone (i.e., vaults reconstruction, re-covering, etc.)	Relevant to both seismic and landslide hazard
Building layout		
	Plan and elevation geometry and dimensions (i.e. one nave, three naves, Latin cross, etc.), height above the ground, presence of valuable decorative elements, orientation of the building	Relevant to both seismic and landslide hazard
Façade layout		
	Facade type, quality and maintenance level, presence or absence of a bell tower, construction material of the exterior walls	Mainly relevant to seismic hazard

3.2 Data analysis

3.2.1 Architectural typologies

The data collected using the RVS form, as described in this section, forms the basis for the subsequent clustering analysis. The primary objective of this analysis is to identify statistically representative archetypes that encapsulate the geometric variability within the surveyed typologies, thereby serving as a foundation for future vulnerability assessments.

As mentioned above, the dataset consists of 71 churches, which were grouped into three typologies based on their in-plane layouts: one-nave (61% of the sample), three-nave (32%), and Latin-cross (7%). For this study, only the one-nave and the three-nave typologies were subjected to clustering analysis to identify representative archetypes. The one-nave group included 43 churches, while the three-nave group consisted of 23 churches. The Latin-cross typology was excluded from further analysis due to its limited representation in the dataset, which precludes a statistically meaningful application of clustering techniques. However, this exclusion does not represent a limitation of the proposed methodology. With a larger dataset, the same approach could be extended to include Latin-cross churches and other building layouts. Figure 3a maps the distribution of the three layouts across the study areas, whereas Figure 3b includes a bar plot with the frequencies mentioned earlier.



a) b)

Figure 3. Church typologies distribution among the selected sample: (a) Map distribution; and (b) Histogram defined for the church typology distribution.

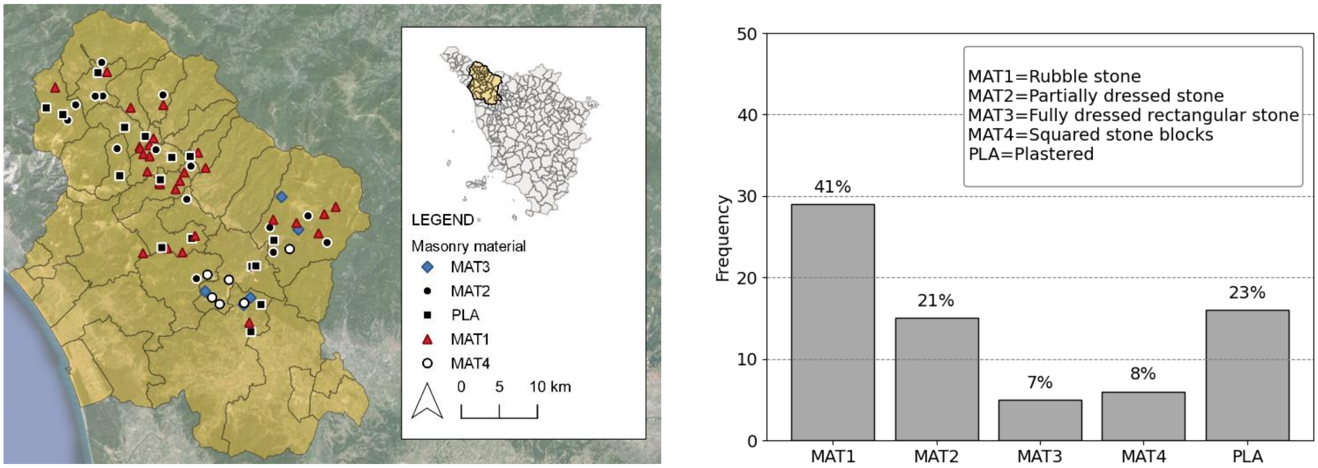
3.2.2 Material properties

The masonry material varies widely from church to church. However, in general, three recurring types were identified: rubble masonry, partially dressed stone masonry, and fully dressed stone masonry. There were a few instances where squared stone ashlars arranged in pseudo-isodomic square work were observed – – a masonry technique characterised by stones appearing to form regular courses but with slight variations in size or arrangement. The different types of masonry materials identified are summarised in Table 2. The connections between walls were typically made using stone ashlars, mostly at the external corners of the façade wall, to create alternate courses. In some cases, the ashlars were set to create continuous vertical joints, resulting in a weak connection with the side walls. It is noteworthy that rubble masonry emerged as the most frequently represented material in the entire church portfolio, accounting for 41% of the churches surveyed. Figure 4a maps the distribution of the masonry material across the study areas, whereas Figure 4b includes a bar plot with the masonry material frequencies (MAT1 = Rubble masonry, MAT2 = Partially dressed stone masonry, MAT3 = Fully dressed stone masonry, MAT4 = Squared stone masonry, PLA = Plastered). Variations in material texture were observed not only between different buildings but also frequently within the same building. For example, in the three-nave typology, the external upper portion of the central nave often exhibited a less regular texture and lower-quality material compared to the lower part of the building. One possible explanation for this distinction is that the external upper section of the central nave is not prominently visible from ground level, making aesthetic considerations less relevant. However, other factors, such as reconstructions, differing construction stages, or subsequent repairs, may also have contributed to these variations.

The mean wall thickness value across the entire church portfolio is approximately 71 cm. This value remains relatively consistent when considering churches of the same typology, measuring 70 cm for one-nave typologies and 71 cm for three-nave typologies. This minimal variation in mean wall thickness, regardless of the in-plane layout, indicates a consistent construction characteristic within the surveyed churches. Since complementary investigation tests such as endoscopy or Ground Penetration Radar (GPR) could not be performed, no information regarding the transversal section and stratigraphy of the walls (i.e., the presence of various connected layers) was possible to gather. There was one case, however, where a partially damaged wall made it possible to see that the wall was composed of three masonry leaves, with a total thickness of roughly 65 cm. A similar condition, i.e. a similar wall stratigraphy, can be expected in all the cases due to the thickness of the wall. While this assumption cannot be generalised to the entire sample without further investigation, it aligns with typical practices in historical masonry construction. According to (BeWeB), sandstone and limestone are the two materials most frequently identified in the outer layer of walls.

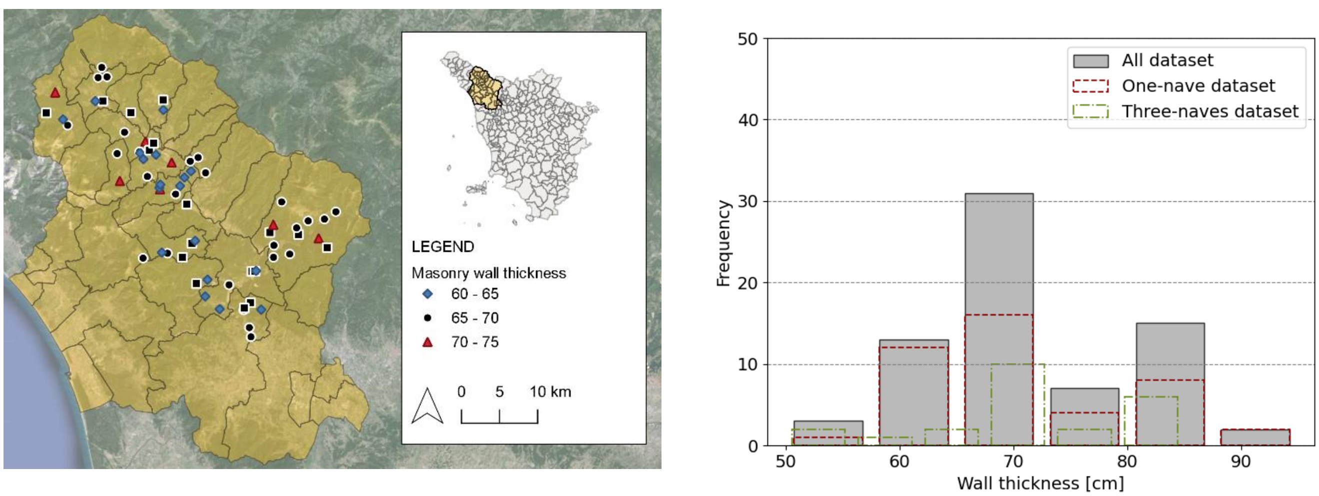
Table 2. Example of the four materials recognise d in the sample of churches examined.

Rubble masonry (MAT1)	Partially dressed stone masonry (MAT2)	Fully dressed stone masonry (MAT3)	Squared stone masonry (MAT4)
			



a) b)

Figure 4. Masonry material distribution among the selected sample: a) Territorial distribution; b) Histogram for the material distribution.



a) b)

Figure 5. Masonry thickness distribution among the selected sample: a) Territorial distribution; (b) Histogram for the wall thickness (tw).

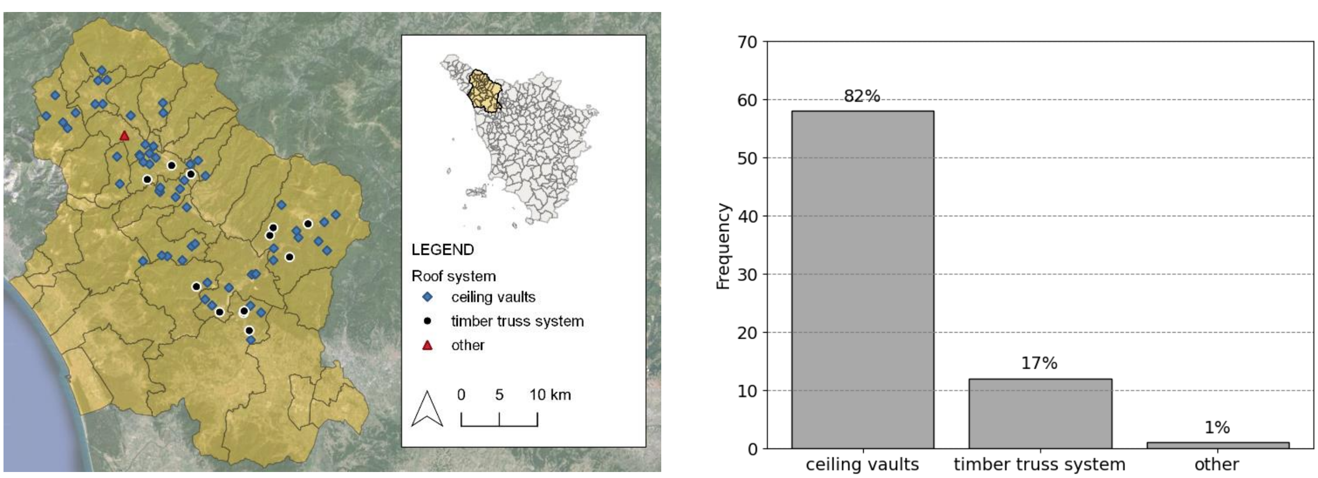
3.2.3 Vault and roofs morphologies

The presence of interior ceilings was widely observed in the surveyed religious buildings; vault interior ceilings were the most common (about 80%), supported by aligned longitudinal walls or circular and rectangular columns and with a square or rectangular architectural plan. Figure 6a maps the distribution of the roof type across the study areas, whereas Figure 6b includes a bar plot with the roof type frequencies.

Vaults' constitutive material was generally unknown. In three cases, however, a wide plaster detachment made it possible to see both the material and its texture. Brick *in folio* vaults were identified in two out of three cases, while a "canniccio" vault – a traditional Italian ceiling or vaulting technique involving woven reeds or canes covered with plaster – was found in the remaining case.

Concerning the shape, influencing the way the acting loads were redistributed on vertical elements, several typologies of vault shapes were found, namely barrel vaults (usually with transversal arches), groin vaults and sail vaults. Tie-rods were found in 75% of the churches surveyed, running in one (48%) or two directions (27%). Tie-rods placed along one of the two main plan orientations are positioned on the interior vault's ceiling springer. In the case of two directions, tie-rods were surveyed along the churches' longitudinal direction in addition to the ones described above.

The external roof structure was invariably found to be made of wooden beams. Where vaults were not detected (i.e. the 15% of the analysed churches), the roof proved to be constituted by a lightweight wooden structure supported by timber trusses, in the case of the one-nave type. A lightweight wooden cover supported by a timber truss system in the central nave and sloping wood beams in the lateral ones was found in the case of the three-nave type. No tie-rods were present when a timber truss system was found. The quality of the timber and its level of maintenance varied significantly across the surveyed buildings, ranging from very poor conditions – such as deteriorated or rotted wood – to new, well-maintained wooden trusses. The remaining 5% of the surveyed buildings featured flat, non-vaulted interior ceilings. Examples of roof structures and interior vault structures are provided in Figure 7.



a) b)

Figure 6. Roof type distribution among the selected sample: (a) territorial distribution; and (b) histogram for the roof type distribution.

a) b) c) d)

Figure 7. Examples of vault and roof structures. a) three-nave church, central nave vault ceiling; b) three-nave church, lateral nave vault ceiling; c) three-nave case, central nave timber trusses roof; and d) three-nave case, lateral nave sloping wood beams.

4. Statistical analysis of the main geometrical attributes

This section describes the statistical approach employed to analyse the geometric attributes of the surveyed churches, aiming to identify archetypes within the primary typologies. The proposed procedure is based on two hypotheses: (1) geometric attributes (e.g., wall thickness, span, height) exhibit clustering tendencies that can be utilised to group buildings into archetypes, and (2) clustering methods can effectively capture variability within the dataset while reducing its complexity, thereby enabling the analysis of a small, representative subset of structures. Figure 8 presents the framework flowchart, outlining the steps from the initial dataset to the identification of the final archetypes. There are several techniques to group data; these can usually be either supervised or unsupervised approaches. The main difference between the two lies in the fact that the former uses labelled data to help predict outcomes, while the latter does not (Sindhu Meena and Suriya 2020). In this study, clustering techniques were specifically selected for their ability to partition the dataset into homogeneous groups (*clusters*) based on similarity, without prior knowledge of the group labels (Jain et al. 1999; Xu and Tian 2015; Vanneschi and Silva 2023). This unsupervised approach is characterised by robustness in identifying underlying patterns in the data, making it ideal for exploring the inherent geometric in-plane and elevation configurations (i.e., archetypes) within the church typologies. General statistical correlations between different geometrical attributes characterising the in-plan and elevation layout of the one-nave and three-nave churches were preliminarily investigated through simple regression analysis. The relevant clusters inside the population with uncorrelated geometric attributes were then determined using three different well-known K-Means, Gaussian Mixture

Models, and Kernel-density clustering techniques.

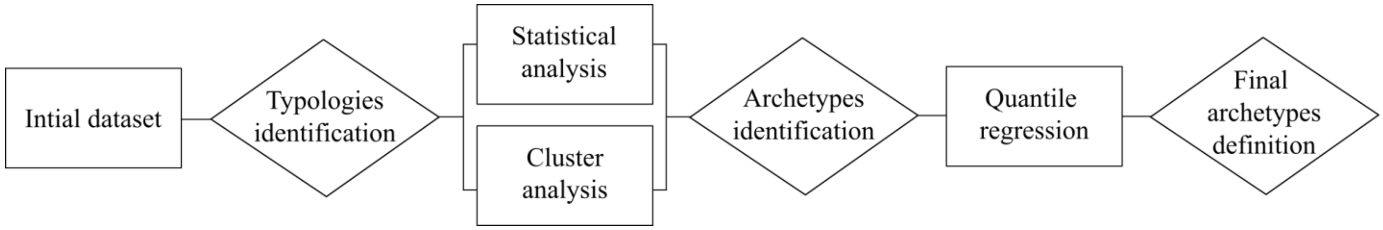


Figure 8 Workflow for archetypes definition.

4.1 Statistical dependency analysis

Histograms were used to analyse the frequency values for each geometric variable and define the lowest, maximum, and mean values of the key geometrical attributes. Additionally, the Coefficient of Variation (CoV), defined as the ratio of the standard deviation to the mean, expressed as a percentage, was calculated for each geometric variable, Table 3. This statistic provides an efficient measure of the relative variability of a dataset in relation to its mean. The high value of the Coefficient of Variation (CoV) obtained shows the high degree of variability in the dataset: this may suggest the fact that geometrical attribute datasets do not represent a single homogeneous population but contain distinct subgroups or clusters with different characteristics.

An investigation into the relationships among selected geometric attributes of one-nave and three-nave churches was conducted through Kendall's Tau (τ_{ken}) correlation coefficient (Temizhan et al. 2022). The coefficient is a nonparametric measure of association introduced by Kendall (Kendall 1938), and it could be used to evaluate the relationship between two ordinal variables. Kendall's Tau is arithmetically bound between −1 and +1, with -1 indicating a perfect negative relationship, 0 indicating no relationship, and 1 indicating a perfect positive relationship, Eq. 1.

$$\tau_{ken} = \frac{C - D}{\sqrt{(C + D + T) \cdot (C + D + U)}} \qquad \textit{Eq. 1}$$

where C is the number of concordant pairs, D is the number of discordant pairs, T is the number of ties only in x, and U is the number of ties only in y. The p-value was used to describe the statistical significance of the result, being able to quantify the comparison of the observed value of the statistic against the null distribution, i.e. the distribution of statistic values derived under the null hypothesis that the geometric attribute data are independent. In a two-sided test in which the statistic is positive, elements of the null distribution greater than the transformed statistic and elements of the null distribution less than the negative of the observed statistic are both considered "extreme values". If the p-value is less than or equal to 0.05, that is, if there is a low probability of sampling data from independent distributions that produce such an extreme statistic value, it is considered evidence that the geometric attributes are not independent.

Statistical independence or dependence was assessed among the main geometrical attributes characterising the in-plan and elevation layout of the churches, as the span (S), the length (L), the perimeter walls height (H_1), the overall building height at the main facade ridge (H_2) and masonry walls thickness (t_w) etc. Figure 9 shows the main geometrical feature considered, while Table 4 shows the investigation results. Regarding the one-nave typology, very low values of Kendall's Tau (τ_{ken}) correlation coefficient were observed when considering the relationship between the perimeter walls' height (H_1) and the other geometric measures. Conversely, the span (S), the length (L) and the perimeter wall thickness (t_w) could be considered related. The same trend is observable in the case of the three-nave typology, albeit in a less obvious manner. It should be noted that in the case of the three-nave church typology, the number of samples available is significantly lower than for the one-nave typology, as shown in Table 3. This reduced sample size introduces a limitation in the statistical analysis, making the result more sensitive to individual observations and more difficult to analyse their relationship. So, while trends are observable, proper statistical dependencies were not identified.

Where a dependency between the geometric variables - even if slight- could be identified, a comparative statistical analysis was performed to identify the most suitable regression model. Identifying the regression model that best represents the pairs of variables under investigation is of paramount importance in order to be able to identify a correlation domain and exclude those pairs of data that are outside certain lower and upper bounds of the domain, as will be explained later, §5.2. Linear, quadratic and exponential statistical models were examined. In the absence of a very large data set, the Leave-one-out cross-validation (LOOCV) method was chosen to understand which statistical model performed best on the given data. LOOCV estimates the error rate by splitting the set of observations into two parts (James et al. 2013). A single observation ($x_i - y_i$) is used to validate the data set while the remaining observations $\{(x_2, y_2), \ldots, (x_n, y_n)\}$ make up the training data set. The statistical learning method is fit on the $n - 1$ training observations, and a prediction $\dot{y}_1 = f(x_1)$ is made for the excluded observation, using its value x_1, where f is the statistical model under examination. Since (x_1, y_1) was not used in the fitting process, $SE_1 = (y_1 - \dot{y}_1)^2$ provides an approximately unbiased estimate for the error. The procedure is repeated n times, where n is the data number, obtaining SE_1, …, SEn. The LOOCV estimate for the data set is the average of these *n-squared* error estimates, Eq. 2:

$$CV = \frac{1}{n}\sum_{i=1}^{n} SEi \qquad \textit{Eq. 2}$$

The smaller the CV, the closer the predicted responses to the true values. Therefore, lower CV values indicate better model performance. Figure 10 shows two examples of the results obtained for the one-nave typology (a) and the three-nave typology (b). The noticeable dispersion in the correlation of geometric attributes shown in Figure 10 reflects the inherent variability within the dataset. While the correlation is weak, it remains statistically significant and provides a useful foundation for subsequent analyses. Specifically, clustering techniques, which form the core of this study, do not rely on strong correlations between variables. However, for the subsequent quantile regression, the observed linear relationship—despite its weakness—proves helpful in establishing statistically significant bounds and refining the archetypes. In general, an overview of all the statistical models shows that the linear model is the one that best captures the relationship between the geometric variables under examination.

Table 3. One-nave and Three-nave typologies: minimum, maximum, mean values and coefficient of variation (CoV).

	One-nave				Three-naves			
Plan	Data number	Mean	CoV	Range [m]	Data number	Mean	CoV	Range [m]
L	44	18 m	29%	[10,29]	22	20 m	27%	[11,29]
S	44	8 m	23%	[5,11]	22	14 m	18%	[10,19]
S_1	-	-	-	-	22	7 m	18%	[5,9]
S_2	-	-	-	-	22	4 m	24%	[2,5]
i	-	-	-	-	21	4 m	15%	[3,6]
t_w	44	70 cm	14%	[50,95]	22	71 cm	18%	[50,85]
Cross-	Data number	Mean	CoV	Range [m]	Data number	Mean	CoV	Range [m]
H_v	37	9 m	21%	[6,14]	-	-	-	-
$Hv1$	-	-	-	-	16	10 m	15%	[6,12]
$Hv2$	-	-	-	-	16	6 m	11%	[5,7]
H_1	31	8 m	22%	[6,13]	11	7 m	10%	[5,7]
H_2	33	9 m	21%	[7,15]	7	8 m	12%	[6,9]
H_3	-	-	-	-	10	10 m	14%	[8,12]
H_4	-	-	-	-	12	12 m	12%	[9,14]

Table 4. Kendall's tau and p-values for one-nave and three-nave typology geometrical attributes.

	One-nave			Three-naves		
Data pairs	Kendall's Tau (τ_{ken})	p-value	dependence	Kendall's Tau (τ_{ken})	p-value	dependence
L - S	0.33	0.0017	Yes	0.07	0.6718	No
L - t_w	0.33	0.0036	Yes	0.19	0.2387	No
S - t_w	0.41	0.0003	Yes	0.26	0.1116	No
L - H_1	0.07	0.5298	No	0.11	0.6367	No
S - H_1	0.10	0.3872	No	0.07	0.7528	No
H_1 - t_w	0.07	0.5801	No	0.10	0.6864	No

$H_1 - H_2$	0.83	3.61e-10	Yes	-	-	
$H_1 - H_4$	-	-		0.32	0.2037	No
$S_1 - S_2$	-	-		0.28	0.0743	No
$S - S_1$	-	-		0.59	0.0001	Yes
$S - S_2$	-	-		0.70	6.83 e-6	Yes

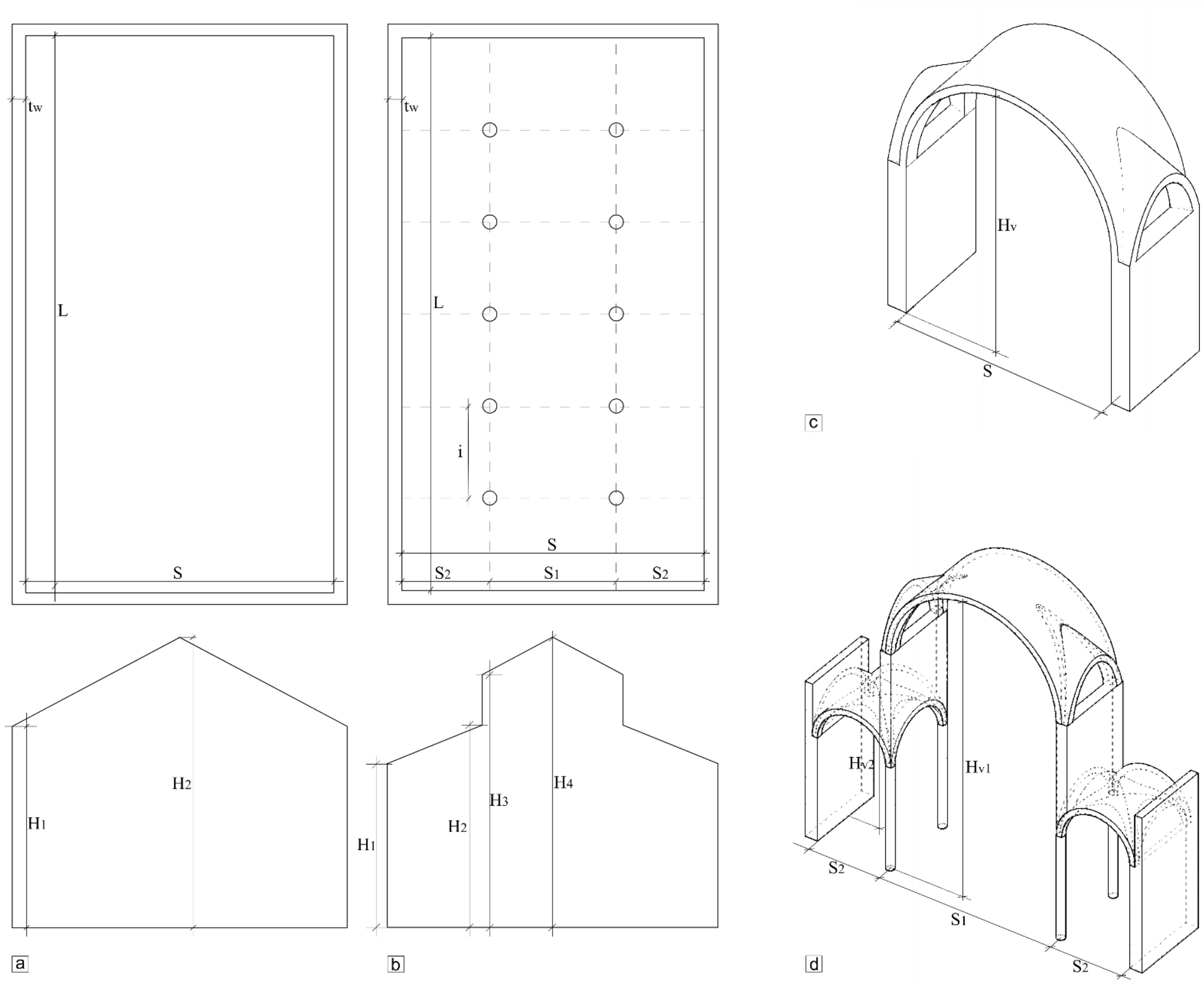
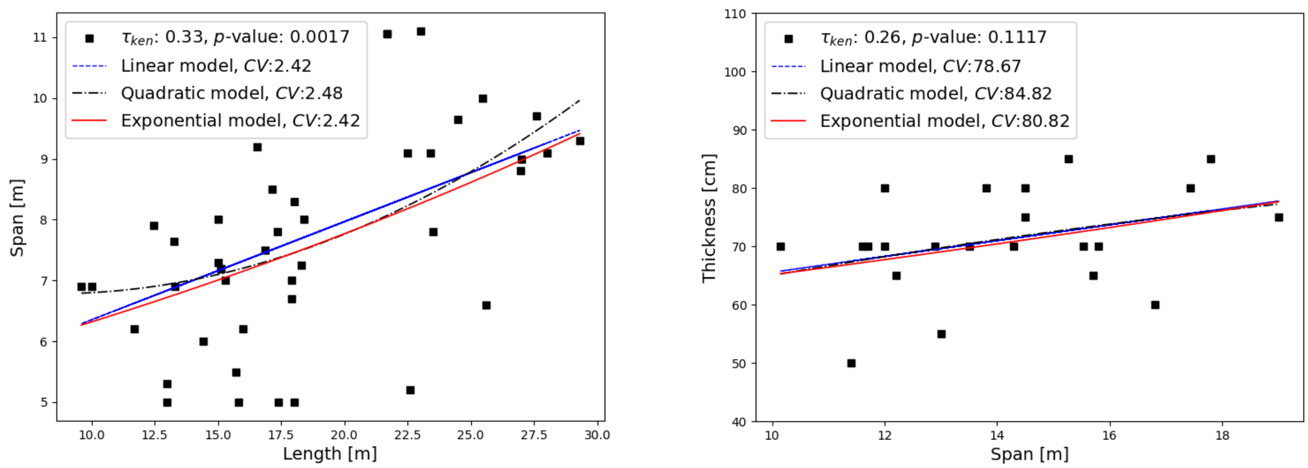


Figure 9. Geometrical attributes; a) one-nave type: plan and facade main geometrical attributes; b) three-nave type: plan and facade main geometrical attributes; c) one-nave type: cross-section main geometrical attributes; d) three-nave type: cross-section main geometrical attributes.



a) b)

Figure 10. Kendall's Tau correlation coefficient and statistical models for: (a) one-nave typology, Length and Span data set; and (b) three-naves typology, Span and Thickness data set.

5. Archetypes derivation

According to (Mata et al. 2014), archetypes (or archetypical buildings) are *'statistical composites of the features found within a category of buildings in the stock'*. This study evaluates the effectiveness of using an archetype aggregation approach to define representative buildings that typify the identified religious building typologies, with the objective of providing a foundation for assessing seismic and landslide hazards. Cluster analysis, or *clustering*, was conducted on churches classified under the one-nave and three-nave typologies, which represent the two most common in-plane layouts in the sample. The analysis aimed to determine whether the attributes could be grouped into relatively distinct clusters. The initial clustering analysis resulted in 36 archetypes for each group, reflecting the variability in the geometric and material attributes of the buildings. However, this initial clustering did not achieve a meaningful reduction in the number of representative buildings. To address this, quantile regression was applied to further refine the archetypes by eliminating statistically insignificant data pairs. This refinement reduced the number of archetypes to 12 for the one-nave group. However, no significant reduction was achieved for the three-nave group due to the limited dataset available. The two phases of this process are detailed in the following sections.

5.1 Cluster analysis

Three well-known clustering techniques were chosen and compared to assess the presence of sub-populations within the indicated attributes. The k-means analysis was initially performed. K-means is probably one of the best-known clustering algorithms and relies on the concept of *centroid* to assign data to a cluster (Sinaga and Yang 2020). The centroid of a cluster is a point whose coordinates in the feature space are the mean values of the respective features of all the observations that belong to the cluster. After placing a predefined number of random points in the feature space, representing the centroids of hypothetical clusters, the K-means algorithm randomly assigns each observation to the nearest centroid and then updates the centroid position based on these assignments until there are no more changes in the cluster (i.e. until the algorithm converged). The clustering solution adopted is the one that minimises the intra-cluster variance, i.e. the total sum of the squared error (SST) objective function. (i.e., the sum of squared Euclidean distances between the centroid and the elements of the cluster). Regarding the optimal number of clusters (K_i), the Elbow Method was adopted, consisting of calculating iterative, for $k=1$ to $k=n$, the Within-Cluster Sum of Squares (WCSS) value, where n is the total number of centroids. The optimal number of clusters corresponds to the k-value for which the WCSS starts decreasing in a linear fashion (i.e. the "elbow" in the graph).

The centroid positions obtained through the K-means method were then used as initial points to perform a Gaussian Mixture Model (GMM) algorithm. The K-means sub-population normality was assessed through the Shapiro-Wilk (SW-Statistic) test before performing GMM. The value of this statistic tends to be high (close to 1) for samples drawn from a normal distribution. GMM involves the mixture (i.e. the superposition) of multiple Gaussian PDFs, each one characterised by weights, mean, and variance (mean vector and covariance matrix for the multivariate case) (McLachlan et al. 2019). Attribute mixture components were obtained using the Expectation Maximization (EM) algorithm. The mean values of the multiple Gaussians were assumed to describe each cluster.

The results obtained through the first two methods were compared with the results obtained using the Gaussian Kernel Density Estimation (KDE) method. KDE is a probability density function estimation technique that generates a smooth empirical pdf based on the individual locations of the sample data (Węglarczyk 2018). Two concepts play a fundamental role in Kernel estimation: the kernel function and the coefficient of smoothness. The Gaussian symmetrical kernel function was used to estimate the probability distribution function *f(x)* of the series $\{x_1, x_2, ..., x_n\}$ of an independent and identically distributed sample of n observations taken from each geometric attribute population X. While the smoothness coefficient was carefully calibrated to obtain an amount of smoothing comparable with the number of clusters obtained in with the two previously presented methodologies. The maximum values corresponding to the peaks of the probability density functions are assumed for describing the sub-populations.

The results from these methods were compared with the mean, the values corresponding to plus and minus one standard deviation from the mean, and to plus and minus two standard deviations from the mean of the corresponding Gaussian probability density function. The goal was to understand how well the identified clusters correspond to the central tendency of the data, with the aim of defining archetypes that capture the full range of geometric variations within the sample.

The above cluster analysis methodologies were applied to the main geometrical attributes highlighted so far, leading to the results shown in Table 5. The mean, the values corresponding to plus and minus the standard deviation from the mean, and plus and minus 2·standard deviations from the mean of each geometric attribute Gaussian probability density function are shown in Table 6. Three or two main sub-populations were identified for each variable. Comparing Table 5 and Table 6, it is possible to note that the values obtained from the cluster analysis fall approximately within one standard deviation, indicating their proximity to the central tendency of the dataset. In view of identifying the geometric configurations representative of the two typologies, the values obtained were combined pairwise. Hence, for both typologies, considering the thickness of the walls (t_w), the length (L) and total span of the church (S) and the height of the perimeter walls (H_1) resulted in the identification of 36 archetypes. Aggregately, these archetypes represent all possible geometric variations within the analysed sample.

Table 5. Geometric attribute mean subpopulation values for the one and three-nave cases.

Attribute	**One-nave typology**				**Three-naves typology**			
	K-means	*SW-Statistic*	GMM	KDE	K-means	*SW-Statistic*	GMM	KDE
t_w [cm]	60.00	0.59	60.00	60.10	55.46	0.75	55.00	55.00
	71.33	0.79	70.00	70.20	69.97	0.64	70.00	70.00
	84.00	0.56	81.30	80.31	81.45	1.00	81.70	80.00
L [m]	15.12	0.93	15.20	16.03	15.92	0.94	15.03	15.71
	24.80	0.91	24.80	21.34	24.55	0.97	24.67	21.84
S [m]	5.44	0.80	5.13	5.26	11.79	0.92	11.74	11.84
	7.46	0.85	7.17	7.25	14.82	0.92	14.77	14.78
	9.76	0.94	9.38	8.98	18.03	0.89	18.01	18.45
H_1 [m]	7.48	0.95	7.52	7.65	5.62	0.95	5.62	5.49
	10.62	0.91	10.40	10.17	6.95	0.91	6.95	6.04

Table 6. Geometrical attributes normal distributions' key values for the one and three-naves cases.

	Attribute	**mean -2σ**	**mean -σ**	**mean**	**mean +σ**	**mean +2σ**
One-nave	t_w [cm]	50.38	60.31	70.30	80.20	90.10
	L [m]	7.92	13.10	18.30	23.50	28.60

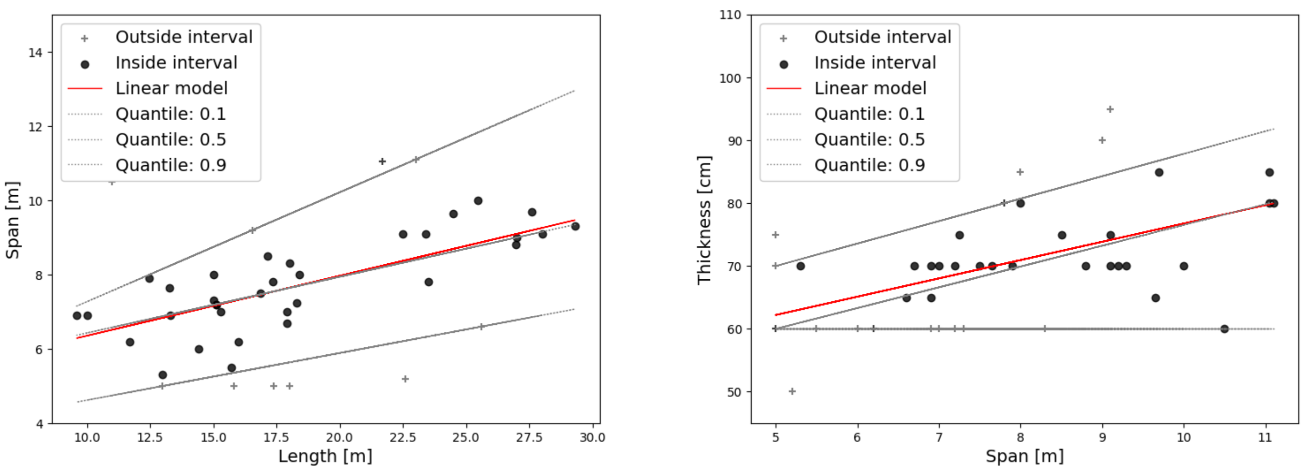
	S [m]	4.29	6.02	7.74	9.46	11.20
	H_1 [m]	5.00	6.74	8.48	10.20	12.00
Three-naves	t_w [cm]	53.83	62.49	71.14	79.79	88.44
	L [m]	9.54	14.67	19.79	24.92	30.04
	S [m]	9.42	11.91	14.40	16.90	19.39
	H_1 [m]	5.30	5.94	6.59	7.23	7.88

5.2 Quantile Regression

Although the initial clustering identified 36 archetypes for each typology, this number did not constitute a significant reduction from the original dataset of churches. To address this limitation, quantile regression was employed to filter out extreme or unrepresentative data pairs, focusing on the relationships among key geometrical attributes. Based on the result obtained from the clustering analysis, the mean values across the characteristic values of each geometric attribute (i.e. the centroid values in the case of the K-means analysis and the peak point values of the PDFs in the case of the GMM and KDE) were calculated. Considering three characteristic values for the thickness of the walls (t_w) and the total span of the church (S), and two for the length (L) and the height of the perimeter walls (H_1), thirty-six possible geometric combinations were obtained, as mentioned earlier. Together, these combinations represent the so-called archetypes and account for the inherent variability related to grouping buildings with different geometrical characteristics into the same building typology. To refine the results, data pairs which were deemed statistically "unrepresentative" were eliminated, thus decreasing the number of archetypes.

In detail, the relationship between the variables previously defined as dependent was studied with the aim of defining an upper and lower limit outside of which the geometric data pairs could be considered statistically unrepresentative. To define upper and lower limits, quantile regression was used to estimate a low and high quantile fixed at 10% and 90%, respectively. The Quantile Regression method is based on the minimisation of the mean absolute error (MAE) to estimate either the conditional median or other quantiles (Koenker and Hallock 2001; Yu et al. 2003). In contrast, the Linear Regression model used to describe the relationship among the dataset minimises the mean squared error (MSE) between the training and predicted targets to obtain the conditional mean. The difference in the object of estimation and the method to obtain it leads to the difference between the linear model and the quantile 0.5 visible in Figure 11. Choosing the 10% and 90% sample quantiles ensured considering data pairs that are representative of 80% of the surveyed one-nave churches population. Consequently, by excluding data pairs that fall outside of these two quantiles, it was possible to reduce the number of geometric combinations, i.e. the archetypes, from 36 to 12 combinations in the one-nave typology case. Table 7 includes a detailed summary of the geometric characteristics of the 12 archetypes. These configurations were selected to capture the variability within the dataset, ensuring that each archetype reflects key geometric distinctions relevant to structural performance. However, some archetypes remain unpopulated. This outcome is common in clustering techniques, as the statistical process may identify archetypes representing potential combinations of geometric attributes that do not correspond to any specific buildings in the dataset.

In the case of the three-nave church type, probably due to the limited amount of data available, it was not possible to identify a significant correlation between the geometric variables. Consequently, if, for each geometric variable, we consider excluding the values obtained from the cluster analysis that are above or below the mean plus or minus two standard deviations, respectively, it is not possible to further reduce the number of geometric combinations.



a) b)

Figure 11. Quantile regression for the one-nave typology: (a) length (L) and span (S); and (b) span (S) and thickness (t_w).

The identified archetypes capture the variations in geometric attributes that could influence structural behaviour under seismic loads. For instance, wall thickness and span directly affect the stiffness and stability of walls, while height influences the susceptibility to overturning. Although this study does not explicitly evaluate the relationship between archetypes and failure mechanisms, the archetypes represent a manageable and statistically representative subset of the original dataset. This approach facilitates large-scale analyses without necessitating building-by-building assessments.

Table 7 Geometric characteristics of one-nave archetypes.

Archetype ID	**t_w [cm]**	**L [m]**	**S [m]**	**H_1 [m]**	**N. of buildings assigned to each archetype**	**[%]**
1	60	15.45	7.30	7.55	11	25
2	70	15.45	7.30	7.55	10	23
3	70	23.65	7.30	7.55	0	0
4	80	23.65	7.30	7.55	0	0
5	70	23.65	9.40	7.55	3	7
6	80	23.65	9.40	7.55	3	7
7	60	15.45	7.30	10.40	2	5
8	70	15.45	7.30	10.40	6	14
9	70	23.65	7.30	10.40	1	2
10	80	23.65	7.30	10.40	0	0
11	70	23.65	9.40	10.40	4	9
12	80	23.65	9.40	10.40	4	9
				TOTAL	**44**	**100**

6. Conclusions

This study introduces a comprehensive methodology employing cluster techniques to analyse the geometric configuration of historical-religious masonry buildings, with a focus on a dataset of seventy-one churches in the Tuscany region (Italy). The methodology does not overlook the uniqueness of historical religious buildings; rather, it provides a practical solution for large-scale analysis by identifying recurring typological patterns and grouping buildings into archetypes. The resulting archetypes account for the intrinsic variability associated with grouping buildings with different geometrical characteristics into the same typology.

Through a two-phase approach, the study initially characterised the sample by identifying predominant typologies and key geometric attributes describing both the in-plane and elevation layout. Subsequently, cluster analysis helped to provide insights into representative geometric configurations for the two main typologies, resulting in several archetypes. Here, "Church archetypes” refer to combinations of geometric and non-geometric attributes that collectively capture the multiple geometrical characteristics and material mechanical properties inherent in a church typology.

To refine the results, upper and lower bounds of the correlation domain among geometric attributes were defined, excluding value pairs falling outside this domain. Quantile regression was then employed to establish low and high quantiles fixed at 10% and 90%, respectively. Geometric data pairs falling outside these quantiles were deemed statistically unrepresentative and subsequently eliminated. By applying clustering and refining the results using quantile regression, this study reduced the number of archetypes from 36 to 12 for the one-nave group, starting with a population of 43 churches. However, the limited data for three-nave churches impeded significant correlation identification, making it challenging to reduce geometric combinations for this group.

This study contributes to heritage risk mitigation by providing a systematic methodology to identify representative archetypes of historical churches. These archetypes serve as a practical framework for reducing the complexity of large-scale vulnerability assessments, enabling the prioritisation of conservation efforts. By identifying recurring geometric configurations and excluding statistically unrepresentative data, the study lays the groundwork for analysing the relationship between structural archetypes and failure mechanisms under seismic and landslide hazards. This approach supports the overarching goal of preserving cultural heritage by facilitating targeted risk mitigation strategies

In conclusion, this study validates the use of clustering techniques for grouping masonry churches into archetypes based on their geometric and material characteristics, provided a sufficiently large dataset is available. While this paper does not directly assess the structural implications of geometric variations, it lays a foundation for future research to enable vulnerability analyses at a large-scale level, where building-by-building investigation and analysis are impractical. Future research could encompass an enlarged set of case studies, which could constitute a more representative statistical sample. Additionally, investigating the correlation between the vulnerability of the archetypes identified against the main hazards they face, along with hazard intensity measures, could lead to the definition of specific fragility curves.

Acknowledgments

The authors would like to acknowledge the students Giacomo Bianchini and Stefano Maria Celardo for their contribution to the performance of on-site surveys, and to express their gratitude to the *Ufficio per l’Arte Sacra e i Beni Culturali Ecclesiastici dell’Arcidiocesi* di *Lucca* for providing full access to the religious buildings.

References

Agency (US) FEM (2017) Rapid visual screening of buildings for potential seismic hazards: A handbook. Government Printing Office

Arrighi C, Tanganelli M, Cristofaro MT, Cardinali V, Marra A, Castelli F, De Stefano M (2022) Multi-risk assessment in a historical city. Nat Hazards :1–32

Bertolin C, Sesana E (2023) Natural hazards affecting cultural heritage: assessment of flood and landslide risk for the 28 existing Norwegian stave churches. Int J Build Pathol Adapt

Betti M, Galano L, Lourenço PB (2021) Territorial seismic risk assessment of a sample of 13 masonry churches in Tuscany (Italy) through simplified indexes. Eng Struct 235:111479. https://doi.org/10.1016/j.engstruct.2020.111479

Betti M, Vignoli A (2008) Modelling and analysis of a Romanesque church under earthquake loading: Assessment of seismic resistance. Eng Struct 30(2):352–367

BeWeB Beni ecclesiastici in web. https://beweb.chiesacattolica.it/

BRZEV S, GREENE M, ARNOLD C, BLONDET M, CHERRY S, COMARTIN C, D'AYALA D, FARSI M, JAIN S, NAEIM10 F, others (2004) The web-based world housing encyclopedia: housing construction in high seismic risk areas of the world. In: 13th world conference on earthquake engineering, Paper

Cantarino I, Torrijo FJ, Palencia S, Gielen E (2014) Assessing residential building values in Spain for risk analyses – application to the landslide hazard in the Autonomous Community of Valencia. Nat Hazards Earth Syst Sci 14(11):3015–3030. https://doi.org/10.5194/nhess-14-3015-2014

Canuti C, Carbonari S, Dall'Asta A, Dezi L, Gara F, Leoni G, Morici M, Petrucci E, Prota A, Zona A (2021) Post-earthquake damage and vulnerability assessment of churches in the Marche Region struck by the 2016 Central Italy seismic sequence. Int J Archit Herit 15(7):1000–1021

Clementi F, Gazzani V, Poiani M, Antonio Mezzapelle P, Lenci S (2018) Seismic assessment of a monumental building through nonlinear analyses of a 3D solid model. J Earthq Eng 22(sup1):35–61

Crowley H, Özcebe S, Spence R, Foulser-Piggott R, Erdik M, Alten K (2012) Development of a European building inventory database. In: Proceedings of the 15th world conference on earthquake engineering. pp 24–28

Cundari GA, Milani G, Failla G (2017) Seismic vulnerability evaluation of historical masonry churches: Proposal for a general and comprehensive numerical approach to cross-check results. Eng Fail Anal 82:208–228

D'Amato Avanzi G, Puccinelli A, Trivellini M, others (1993) Slope stability maps in areas of particular seismic interest: a short report on the researches in Garfagnana and Lunigiana (Tuscany)

D'Ayala D, Galasso C, Putrino V, Garciano L, Oreta A, Yu K, Fanciullacci D, Barucco P, Fanciullacci V, Bronzino C MULTI–HAZARD RISK ASSESSMENT OF PRIORITY CULTURAL HERITAGE STRUCTURES IN THE PHILIPPINES1

DCP (2018) Italian Civil Protection Department. National Risk Assessment 2018. Overview of the potential major disasters in Italy. Updated

De Matteis G, Mazzolani FM (2010) The Fossanova Church: seismic vulnerability assessment by numeric and physical testing. Int J Archit Herit 4(3):222–245

Del Carlo F, Caprili S, Ferreira TM, Roca P, Uzielli M (2023) A Building classification system for the multi-hazard risk assessment of historical religious buildings. In: 9th ECCOMAS Thematic Conference on Computational Methods in Structural Dynamics and Earthquake Engineering, Paper

Del Zoppo M, Belliazzi S, Lignola GP, Di Ludovico M, Prota A (2022) Analytical structural vulnerability and classification of Italian residential building stock to tsunami hazard. Bull Geophys Oceanogr 63(4):555–574

D.M.17.01.2018. (2018) Aggiornamento delle norme tecniche per le costruzioni (NTC18), G.U. 20.02.2018 n. 42, Suppl. Ordinario n. 8.

Ferrero C, Cambiaggi L, Fenialdi A, Roca P, Vecchiattini R, Calderini C (2021) Slow-moving landslide damage assessment of historic masonry churches: some case-studies in Italy. In: SAHC 2021: 12th international conference on structural analysis of historical constructions. Artes Gráficas Torres SL Cornellà de Llobregat, pp 1256–1267

Fleming K, Parolai S, Garcia-Aristizabal A, Tyagunov S, Vorogushyn S, Kreibich H, Mahlke H (2016) Harmonizing and comparing single-type natural hazard risk estimations. Ann Geophys 59(2)

Formisano A, Vaiano G, Davino A, Citro S, D'Amato M (2022) Seismic vulnerability assessment of two territorial case studies of Italian ancient churches: Comparison between simplified and refined numerical models. Int J Mason Res Innov 7(1–2):172–216

Fortunato G, Funari MF, Lonetti P (2017) Survey and seismic vulnerability assessment of the Baptistery of San Giovanni in Tumba (Italy). J Cult Herit 26:64–78

Fuentes DD, Baquedano Julià PA, D'Amato M, Laterza M (2021) Preliminary Seismic Damage Assessment of Mexican Churches after September 2017 Earthquakes. Int J Archit Herit 15(4):505–525. https://doi.org/10.1080/15583058.2019.1628323

Grazzini A, Chiabrando F, Foti S, Sammartano G, Spanò A (2019) A multidisciplinary study on the seismic vulnerability of St. Agostino church in Amatrice following the 2016 seismic sequence. Int J Archit Herit

Grünthal G, Levret A (1998) European macroseismic scale 1998 (EMS-98) cahiers du centre Européen de géodynamique et de séismologie 15. Cent Eur Géodynamique Séismologie Luxemb

Guillard-Gonçalves C, Zêzere JL, Pereira S, Garcia RA (2016) Assessment of physical vulnerability of buildings and analysis of landslide risk at the municipal scale: application to the Loures municipality, Portugal. Nat Hazards Earth Syst Sci 16(2):311–331

Jain AK, Murty MN, Flynn PJ (1999) Data clustering: a review. ACM Comput Surv 31(3):264–323. https://doi.org/10.1145/331499.331504

James G, Witten D, Hastie T, Tibshirani R (2013) An introduction to statistical learning. Springer

Julià PB, Ferreira TM (2021) From single-to multi-hazard vulnerability and risk in Historic Urban Areas: a literature review. Nat Hazards 108(1):93–128

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

Kappes MS, Papathoma-Koehle M, Keiler M (2012b) Assessing physical vulnerability for multi-hazards using an indicator-based methodology. Appl Geogr 32(2):577–590

Karwacka E, Pozzobon E, Santini L, others (2019) Knowing, reviving and enhancing church buidings. Multi Criteria Decision Analysis applied to underused ecclesiastical properties. In:) Dio non abita più qui? Dismissione di luoghi di culto e gestione integrata dei beni culturali ecclesiastici/Doesn't God dwell here anymore? Decommissioning places of worship and integrated management of ecclesiastical cultural heritage. Artemide, pp 335–347

Kendall MG (1938) A new measure of rank correlation. Biometrika 30(1/2):81–93

Koenker R, Hallock KF (2001) Quantile Regression. J Econ Perspect 15(4):143–156. https://doi.org/10.1257/jep.15.4.143

Lagomarsino S (2006) On the vulnerability assessment of monumental buildings. Bull Earthq Eng 4:445–463

Lagomarsino S, Podestà S (2004) Seismic vulnerability of ancient churches: II. Statistical analysis of surveyed data and methods for risk analysis. Earthq Spectra 20(2):395–412

Lo Presti D, Fontana T, Marchetti D (2008) Slope Stability Analysis In Seismic Areas Of The Northern Apennines (Italy). In: AIP Conference Proceedings. American Institute of Physics, pp 525–534

Lopez S, D'Amato M, Ramos L, Laterza M, Lourenço PB (2019) Simplified formulations for estimating the main frequencies of ancient masonry churches. Front Built Environ 5:18

Marotta A, Sorrentino L, Liberatore D, Ingham JM (2018) Seismic Risk Assessment of New Zealand Unreinforced Masonry Churches using Statistical Procedures. Int J Archit Herit 12(3):448–464. https://doi.org/10.1080/15583058.2017.1323242

Mata É, Kalagasidis AS, Johnsson F (2014) Building-stock aggregation through archetype buildings: France, Germany, Spain and the UK. Build Environ 81:270–282

McCormack TC, Rad FN (1997) An earthquake loss estimation methodology for buildings based on ATC-13 and ATC-21. Earthq Spectra 13(4):605–621

McLachlan GJ, Lee SX, Rathnayake SI (2019) Finite Mixture Models. Annu Rev Stat Its Appl 6(1):355–378. https://doi.org/10.1146/annurev-statistics-031017-100325

Medvedev SV, Sponheuer W, Karnik V (1965) Seismic intensity scale version MSK 1964. United Nation Educ Sci Cult Organ Paris 7

Mouroux P, Le Brun B (2008) RISK-UE project: an advanced approach to earthquake risk scenarios with application to different european towns. In: Assessing and managing earthquake risk: geo-scientific and engineering knowledge for earthquake risk mitigation: developments, tools, techniques. Springer, pp 479–508

Noel MF, Moreira S, Briceño C, López-Hurtado E, Aguilar R (2019) Seismic assessment of the church of San Sebastian in Cusco, Peru by means of pushover nonlinear analysis. In: Structural Analysis of Historical Constructions: An Interdisciplinary Approach. Springer, pp 1462–1470

OPCM 28 aprile 2006 n. 3519 (2006) Ordinanza PCM 3519 del 28/04/2006. Criteri generali per l'individuazione delle zone sismiche e per la formazione e l'aggiornamento degli elenchi delle medesime zone, G.U. n.108 del 11/05/2006 (in Italian)

Orduña A, Preciado A, Galván F, Araiza JC (2008) Vulnerability assessment of churches at Colima by 3D limit analysis models

Palazzi NC, Rovero L, Tonietti U, de la Llera JC, Sandoval C (2019) Seismic vulnerability assessment of unreinforced masonry churches in central Chile. In: Structural Analysis of Historical Constructions: An Interdisciplinary Approach. Springer, pp 1172–1181

PCM-DPC Mibac (2006) Model A-DC Scheda per il rilievo del danno ai beni culturali—Chiese. PCM-DPC Mibac (Presidenza Consiglio dei Ministri—Dipartimento della Protezione Civile Ministero Beni e Attività Culturali)

Penna A, Calderini C, Sorrentino L, Carocci CF, Cescatti E, Sisti R, Borri A, Modena C, Prota A (2019) Damage to churches in the 2016 central Italy earthquakes. Bull Earthq Eng 17:5763–5790

Pirchio D, Walsh KQ, Kerr E, Giongo I, Giaretton M, Weldon BD, Ciocci L, Sorrentino L (2021) Seismic risk assessment and intervention prioritization for Italian medieval churches. J Build Eng 43:103061. https://doi.org/10.1016/j.jobe.2021.103061

Regione Toscana DB Geomorfologico. https://www502.regione.toscana.it/geoscopio/geomorfologia.html

Romão X, Paupério E, Pereira N (2016) A framework for the simplified risk analysis of cultural heritage assets. J Cult Herit 20:696–708

Rovida A, Locati M, Camassi R, Lolli B, Gasperini P, Antonucci A (2016) Catalogo parametrico dei terremoti italiani (cpti15). Napoli Ist Naz Geofis E Vulcanol Doi 10

Santos C, Ferreira TM, Vicente R, da Silva JRM (2013) Building typologies identification to support risk mitigation at the urban scale–Case study of the old city centre of Seixal, Portugal. J Cult Herit 14(6):449–463

Schneider PJ, Schauer BA (2006) HAZUS—Its development and its future. Nat Hazards Rev 7(2):40–44

Sevieri G, Galasso C, D’Ayala D, De Jesus R, Oreta A, Grio MEDA, Ibabao R (2020) A multi-hazard risk prioritisation framework for cultural heritage assets. Nat Hazards Earth Syst Sci 20(5):1391–1414

Sferrazza Papa G, Nollet M-J, Parisi MA, Youance S (2019) The seismic vulnerability assessment of Québec churches: considerations on territorial specificities. In: 12 th Canadian Conference on Earthquake Engineering. CAN, pp 1–8

Sinaga KP, Yang M-S (2020) Unsupervised K-means clustering algorithm. IEEE Access 8:80716–80727

Sindhu Meena K, Suriya S (2020) A survey on supervised and unsupervised learning techniques. In: Proceedings of international conference on artificial intelligence, smart grid and smart city applications: AISGSC 2019. Springer, pp 627–644

Sisti R, Cescatti E, Follador V, Da Porto F, Calderini C, Lagomarsino S, Morici M, Prota A (2023) Operational critical issues filling in the Italian form for the post-earthquake damage assessment of churches (A-DC 2006). Procedia Struct Integr 44:1848–1855

Sorrentino L, Marotta A, Liberatore D (2019) Multi-linear models for the rapid assessment of the vulnerability of unreinforced masonry churches involved in the 2016-2017 Central Italy seismic sequence. Multi-Linear Models Rapid Assess Vulnerability Unreinforced Mason Churches Involv 2016-2017 Cent Italy Seism Seq :84–92

Temizhan E, Mirtagioglu H, Mendes M (2022) Which correlation coefficient should be used for investigating relations between quantitative variables. Acad Sci Res J Eng Technol Sci 85:265–277

Uva G, Sangiorgio V, Ruggieri S, Fatiguso F (2019) Structural vulnerability assessment of masonry churches supported by user-reported data and modern Internet of Things (IoT). Measurement 131:183–192

Vanneschi L, Silva S (2023) Unsupervised Learning: Clustering Algorithms. In: Vanneschi L, Silva S (eds) Lectures on Intelligent Systems. Springer International Publishing, Cham, pp 289–331

Węglarczyk S (2018) Kernel density estimation and its application. ITM Web Conf 23:00037. https://doi.org/10.1051/itmconf/20182300037

Xu D, Tian Y (2015) A comprehensive survey of clustering algorithms. Ann Data Sci 2:165–193

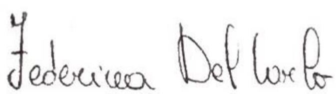
Yu K, Lu Z, Stander J (2003) Quantile regression: applications and current research areas. J R Stat Soc Ser Stat 52(3):331–350. https://doi.org/10.1111/1467-9884.00363

Zschau J (2017) 2.5 Where are we with multihazards, multirisks assessment capacities?

Statements and Declarations

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Federica Del Carlo



Silvia Caprili

Tiago Miguel Ferreira

Pere Roca

Marco Uzielli