ORIGINAL ARTICLE

An empirical seismic vulnerability model

A. Rosti[1] · M. Rota[2] · A. Penna[1,2]

Abstract

This paper proposes a novel seismic vulnerability model for the classification of the existing residential building stock. The vulnerability model rests on a data-driven approach, taking advantage of observed seismic damages detected on several Italian building typologies, struck by the 2009 L'Aquila earthquake. Unsupervised machine learning techniques are exploited for clustering empirical damage data and objectively identifying vulnerability classes of decreasing vulnerability. The cascading use of different strategies, involving clustering analysis and probability theory, results in a comprehensive vulnerability model, which allows for determining, into a probabilistic framework, the degree of belonging of a given building typology to multiple vulnerability classes. The adoption of the peak ground acceleration for characterising the ground shaking is a further advantage of this study, overcoming several limitations related to the use of macroseismic intensity.

Keywords Vulnerability classes · Clustering analysis · Macroseismic approaches · Fragility curves · Post-earthquake damage data · L'Aquila seismic event

1 Introduction

Depending on the nature of the available data, procedures for seismic vulnerability assessment can be classified into empirical, analytical, expertise-based and hybrid. Among expertise-based methods, macroseismic approaches (e.g. Lagomarsino and Giovinazzi 2006; Bernardini et al. 2008, 2011) allow to classify the vulnerability of the exposed buildings by referring to the six vulnerability classes of the EMS-98 (Grünthal et al. 1998) and considering the uncertainty in the attribution of a given building typology to vulnerability classes by means of the fuzzy set theory. The operational implementation of macroseismic approaches takes place via closed-form analytical relations, correlating seismic input and the expected damage, as a function of the assessed vulnerability (e.g. Lagomarsino et al. 2021). Seismic input is represented by macroseismic intensity, which is however a

✉ M. Rota
  maria.rota@eucentre.it

1 Department of Civil Engineering and Architecture, University of Pavia, Via Ferrata 3,
   27100 Pavia, Italy

2 Department of Industrial Products, European Centre for Training and Research in Earthquake
   Engineering, Via Ferrata 1, 27100 Pavia, Italy

descriptive parameter resulting from the observation of the effects of an earthquake on the surrounding environment. Consequently, macroseismic intensity is affected by the characteristics, and thus by the vulnerability, of the existing building stock (e.g. Tertulliani et al. 2011; Graziani et al. 2019; Rossi et al. 2019). The identification of building typologies and their association to vulnerability classes can be limited by the regional variability of the built environment, driven by locally available construction materials, construction period and field experience gained over the centuries from the observation of damages caused by past earthquakes (e.g. Masi et al. 2021; Tocchi et al. 2021).

In Italy, the AeDES post-earthquake survey form (Baggio et al. 2007) is currently used for damage and usability assessment of ordinary buildings. Besides damage and usability information, the AeDES survey form allows for detecting metrical and typological attributes of buildings surveyed in the aftermath of an earthquake. Therefore, it can be an effective tool for supporting the definition of typological classification systems (e.g. Rota et al. 2008; Del Gaudio et al. 2017; Rosti et al. 2018) and capturing regional distinctive features of the building stock, which may affect seismic vulnerability. In this context, the interview-based CARTIS form (Zuccaro et al. 2015) can be also exploited for supporting the definition of building portfolios and vulnerability models (e.g. Polese et al. 2019, 2020; Brando et al. 2021).

Despite their methodological consistency and mathematical "elegance", macroseismic methods are limited by the need of resorting to uncertain and approximate laws to correlate intensity values and peak ground motion parameters (e.g. Bernardini et al. 2011; Maio et al. 2015) and make these models easily usable for vulnerability and risk applications (e.g. da Porto et al. 2021; Dolce et al. 2021).

Based on the above considerations, this study proposes an innovative and comprehensive empirical vulnerability model for masonry and RC buildings representative of the Italian building stock. Given the notable amount of available damage data and the possibility of suitably accounting for the negative evidence of damage, the L'Aquila (2009) post-earthquake damage database (Dolce et al. 2019) is exploited. The proposed approach preserves the main conceptual framework at the basis of the macroseismic approach (Lagomarsino and Giovinazzi 2006), which allows for an exhaustive vulnerability classification of the building stock by referring to vulnerability classes and by considering the uncertain association of building typologies to vulnerability classes. Novelties of this work are the adoption of the peak ground acceleration as physical parameter for the characterisation of ground motion severity, which makes the application of the proposed vulnerability model handy from the engineering perspective, and the use of unsupervised machine learning techniques for removing the subjectivity in the definition of vulnerability classes, by clustering seismic damages observed on the Italian built environment. The use of the AeDES survey form for the L'Aquila (2009) post-earthquake building inspections also allows for an enhanced definition of building typologies representative of the Italian building stock.

This paper is structured as follows. The adopted post-earthquake damage dataset is first described (Sect. 2), together with the main assumptions and interpretations, regarding the characterisation of the ground motion severity experienced at each building location, the classification of the observed seismic damage and the adopted building taxonomy. Typological fragility and mean damage curves are derived, as a function of the peak ground acceleration, by means of a suitable statistical model and fitting procedure. Empirically-derived mean damage data are then partitioned into a predefined number of clusters based on a data-driven approach. In this context, a soft clustering technique is used, permitting each data point to belong to multiple clusters with different membership degree (Sect. 3). This strategy allows for an objective

identification of vulnerability classes of decreasing vulnerability, for which fragility functions are subsequently derived. A total of ten vulnerability classes (i.e. A1, B1, C1, C2, D1, D2, E1, E2, F1 and F2) is considered, six of which refer to masonry buildings (i.e. A1, B1, C1, D1, E1 and F1) and four to RC buildings (i.e. C2, D2, E2 and F2). The expedient of distinguishing masonry from RC vulnerability classes arises from the different distance among damage states emerged from typological fragility functions.

A probabilistic framework is then set up allowing for the attribution of a given building typology to multiple vulnerability classes, based on an ad-hoc strategy, involving the use of probability theory and using empirically-derived typological fragility functions as a target (Sect. 4). Similarly to the EMS-98, a vulnerability table is proposed to differentiate the seismic vulnerability of the exposed built environment based on selected building attributes. In this context, the synthetic binomial parameter, representing the weighted mean vulnerability class, is indicated for each considered building typology, allowing for easily categorising the seismic vulnerability of masonry and RC buildings depending on their typological features. The feasibility of the proposed vulnerability model is then demonstrated by a case study application with reference to two selected building typologies (Sect. 5).

Final remarks and conclusions of this work are discussed in Sect. 6. The results obtained in the different methodological phases of this study, including the parameters of the fragility functions derived for both building typologies and vulnerability classes, the degrees of belonging of building typologies to vulnerability classes and the proposed vulnerability table, can be used in varied applications in the field of seismic vulnerability and risk, provided the similarity of seismic hazard and built environment.

2 Post-earthquake damage database

This work takes advantage of a robust post-earthquake database, gathering damage data of residential buildings hit by the 2009 L'Aquila seismic event (Dolce et al. 2019). Although the availability of several post-earthquake damage databases, the L'Aquila database represents a good candidate for different aspects, among which the significant number of inspected buildings and the considerable number of completely-surveyed municipalities (Fig. 1), identified by a completeness ratio (i.e. number of surveyed residential buildings over the total number of buildings evaluated from national building census, ISTAT 2001) higher than 90% (Rosti et al. 2021a, b). The possibility of accounting for the negative evidence of damage in the municipalities less affected by the ground shaking is a further advantage related to the use of this post-earthquake damage database. In this context, non-surveyed residential buildings, sited in 176 non-surveyed and 49 partially-surveyed (completeness ratio < 10%) municipalities of the Abruzzi region (Fig. 1), are reasonably assumed to be undamaged and used to integrate the post-earthquake damage database (Rosti et al. 2021a, b). Following these operations, the considered post-earthquake dataset includes damage data of 37′406 residential (masonry: 28′713 and RC: 8′693) buildings. Non-inspected residential buildings, from non-surveyed and partially-surveyed municipalities and assumed undamaged, are 175′152 (masonry) and 22′376 (RC).
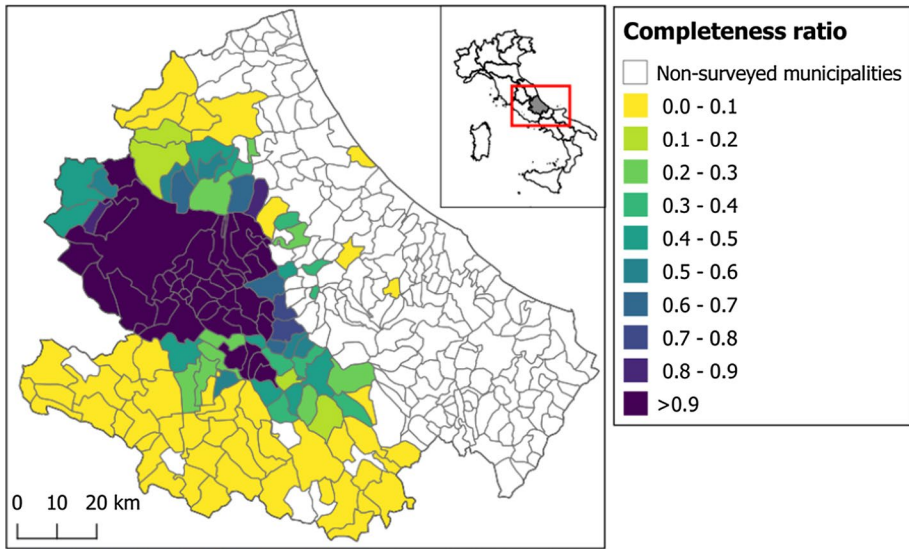
Fig. 1 Identification of the Abruzzi surveyed and non-surveyed municipalities. Colours denote the completeness ratio of each municipality

2.1 Ground motion characterisation

Consistently with the aim of this study, the ground motion severity experienced at each building location is quantified by PGA (e.g. Rosti et al. 2020a), extrapolated from updated INGV shakemap (Michelini et al. 2020). With respect to Michelini et al. (2008), the new shakemap configuration involves the use of recently developed ground motion models, selected based on a ranking procedure, an updated V_{s30} map for local site effects and the adoption of the newly developed USGS-ShakeMap version 4 (v.4) software (Worden et al. 2020). Also, the new shakemap defines isoseismic units of 0.02 g, allowing for a rather accurate seismic input characterisation of buildings subsequently involved in the fragility assessment (Fig. 2).

2.2 Classification of the observed seismic damage

Starting from the damage description provided by the post-earthquake survey form, grading the seismic damage observed on different building components based on both damage severity and extent, a unique global level of damage is assigned to each inspected building. Damage descriptions available from the post-earthquake survey form are mapped to the discrete damage levels of the EMS-98 using the damage conversion rules by Rota et al. (2008) and Del Gaudio et al. (2017) in case of structural and non-structural (i.e. masonry infills/partitions) damage, respectively. Global damage levels of masonry buildings are then given by the maximum level of damage detected on the vertical structure, intermediate diaphragms and roof (e.g. Rota et al. 2008; Rosti et al. 2018). In case of RC buildings, global damage levels are instead given by the
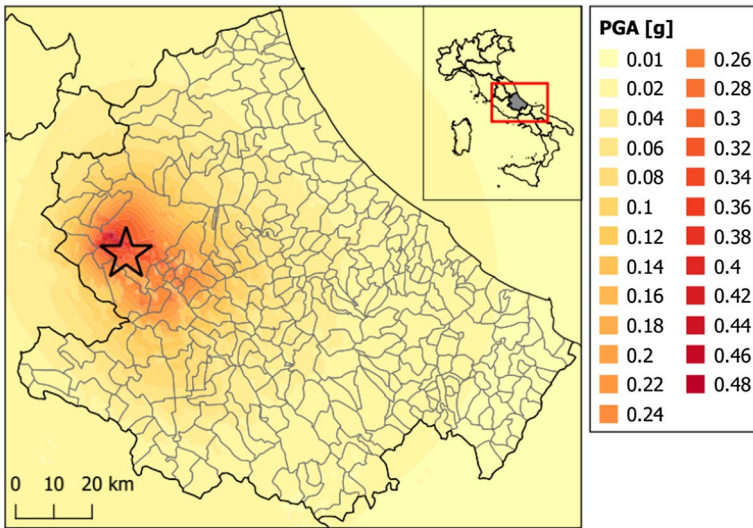
Fig. 2 INGV PGA shakemap of the 2009 L'Aquila seismic event (Michelini et al. 2020)

maximum level of damage observed on the vertical structure and masonry infills/partitions (e.g. Del Gaudio et al. 2017; Rosti et al. 2018).

Figure 3 shows the resulting damage distributions of residential masonry (**a**) and RC (**b**) buildings, with reference to the completely-surveyed municipalities.

2.3 Typological classification of the residential building stock

The typological classification of the residential building stock accounts for the main building attributes retrievable from the AeDES post-earthquake survey form and
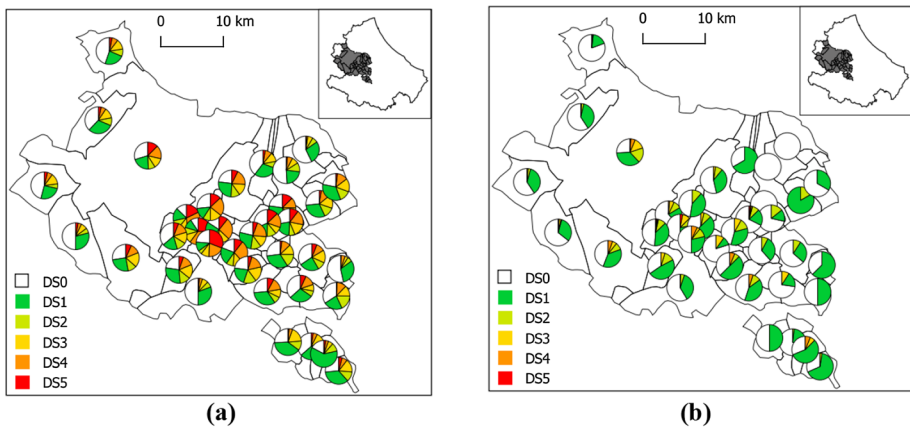

Fig. 3 Damage classification of the residential masonry (**a**) and RC (**b**) building stock for L'Aquila completely-surveyed municipalities

affecting the buildings' seismic behaviour. A first distinction is made based on the construction material (i.e. masonry/RC) and the number of storeys (i.e. 1, 2, 3 and ≥ 4 storeys for masonry buildings and 1, 2, 3, 4, ≥ 5 storeys for RC buildings). Masonry buildings are further classified based on the texture and quality of the masonry fabric (i.e. IRR: irregular layout or poor-quality; REG: regular layout and good-quality), in-plane flexibility of intermediate floor diaphragms (i.e. F: flexible; R: rigid) and presence (or lack) of connecting devices, such as tie-rods and/or tie-beams (i.e. CD: with connecting devices; NCD: without connecting devices), similarly to Rota et al. (2008).

Besides the building height, the typological classification of the RC building stock accounts for the level of seismic design (i.e. buildings seismically designed pre-1981 and post-1981, being 1981 a key date for the enforcement of relatively modern seismic design rules). Most of the municipalities in the L'Aquila region were classified as seismic prone in early twentieth century, hence RC buildings designed to gravity loads (and wind loads) only are basically missing in the dataset.

The considered building taxonomy leads to the identification of a total of 42 building typologies, 32 of which refer to masonry buildings and 10 to RC buildings. The higher level of detail of the adopted typological classification system is aimed at detecting possible differences or similarities in the observed seismic vulnerability of the existing building stock, given the presence (or absence) of specific building attributes or constructive details.

Masonry and RC buildings represent 77% and 23%, respectively, of the considered post-earthquake damage database. Irregular layout or poor-quality masonry constitutes about 68% of the considered masonry buildings, 32% of which are instead characterised by good-quality materials with regular layout. About 70% of the masonry buildings have flexible intermediate diaphragms, whereas 30% of the horizontal structures are rigid. Aseismic devices are present in 41% of the considered masonry buildings. Focusing on RC constructions, 36% and 64% are seismically designed pre- and post-1981, respectively.

Referring to the completely-surveyed municipalities, Fig. 4a shows the subdivision of the existing building stock based on the construction material. Figure 4 subdivides masonry (b) and RC buildings (c) based on the number of storeys. In Fig. 4d, e, f masonry buildings are classified based on the masonry type, in-plane stiffness of the intermediate diaphragms, presence (or lack) of aseismic devices (i.e. tie-rods/tie-beams).

The post-earthquake damage database is then enlarged by undamaged buildings from non-surveyed and partially-surveyed municipalities. As the building attributes considered by the national building census are construction material, construction age and number of storeys, the integration of undamaged RC buildings is straightforward. Conversely, mapping of undamaged masonry buildings to the predefined building typologies is carried out based on the typological composition of the masonry macro-categories, identified based on the age of construction (i.e. < 1919, 1919–45, 1946–61, 1962–71, 1972–81, 1982–91 and > 1991) and number of storeys (1, 2, 3, ≥ 4), as depicted in Fig. 5. Frequency values reported in Fig. 5, obtained by classifying residential masonry buildings of the considered post-earthquake damage dataset based on typological and census building attributes, both available from the AeDES survey form, are then applied to the undamaged building stock, for which only census-based information is available.

Table 1 reports the building typologies identified based on the adopted building taxonomy, together with the indication of the sample size, including both damaged and undamaged buildings.

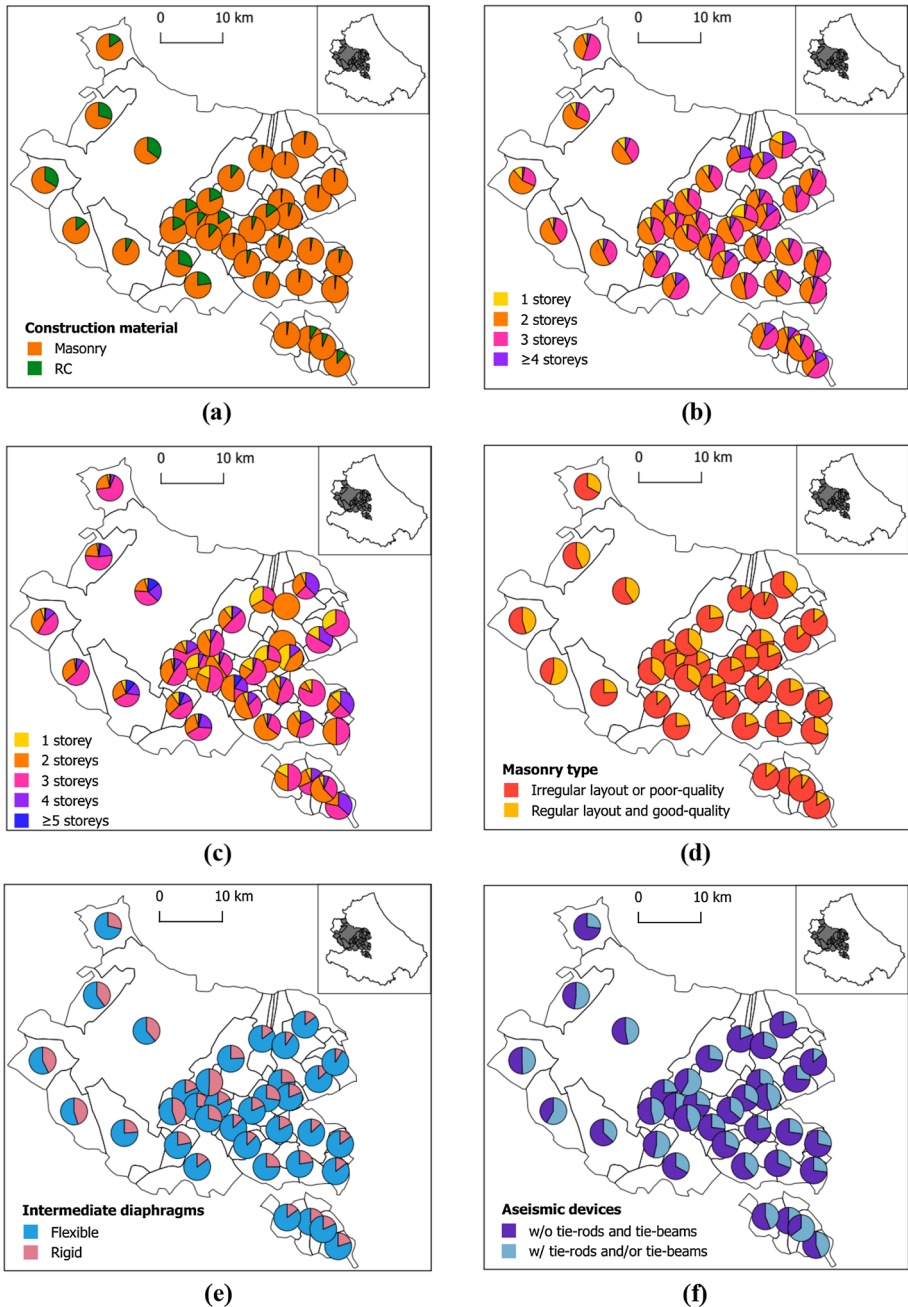**Fig. 4** Typological classification of the residential building stock for L'Aquila completely-surveyed municipalities. Construction material (**a**); subdivision of masonry (**b**) and RC (**c**) buildings based on the number of storeys; subdivision of masonry buildings based on the masonry type (**d**), in-plane stiffness of intermediate diaphragms (**e**), presence or absence of aseismic devices (**f**)
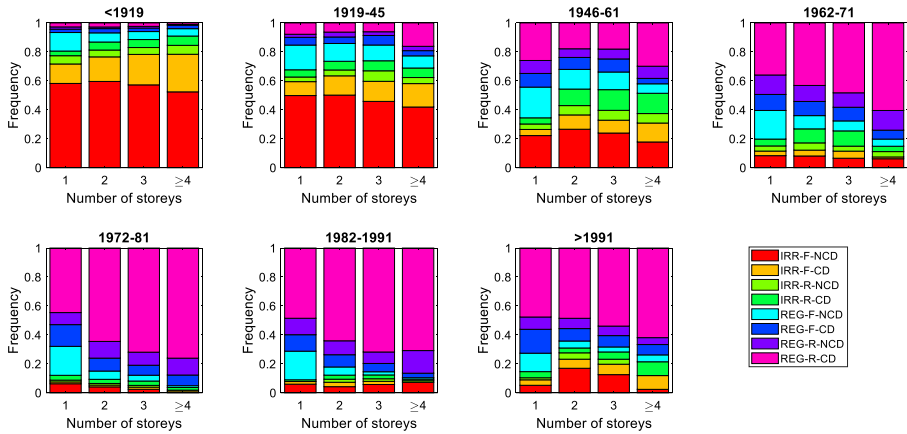
Fig. 5 Typological composition of the masonry macro-categories identified based on census building attributes

2.4 Seismic fragility assessment

Empirical fragility curves are derived for quantifying the seismic vulnerability of predefined building typologies, as a function of the ground motion severity. The cumulative lognormal distribution (e.g. Rossetto and Elnashai 2003; Rota et al. 2008; Del Gaudio et al. 2017; Ader et al. 2020; Rosti et al. 2020b) is employed for describing the probability of reaching or exceeding a given level of damage, P($ds \geq DS_i|PGA$), as a function of PGA:

$$P\left(ds \geq DS_i|PGA\right) = \Phi\left[\frac{\log\left(PGA/\theta_{DSi}\right)}{\beta}\right] \quad (1)$$

where θ_{DSi} is the median PGA value associated with damage level DS_i and β is the logarithmic standard deviation.

The buildings' subdivision in the different damage states, n_{ij}, given the jth PGA threshold, is approximated by the multinomial distribution (e.g. Agresti 2002; Charvet et al. 2014; Ioannou et al. 2021; Rosti et al. 2021a, b):

$$n_{ij} \sim \prod_{i=0}^{nDS} \frac{N_j!}{n_{ij}!} P\left(ds = DS_i|PGA_j\right)^{n_{ij}} \quad (2)$$

where N_j is the total number of buildings at the jth PGA threshold and P($ds = DS_i|PGA_j$) is the conditional probability of occurrence of damage level DS_i, defined as:

$$P\left(ds = DS_i|PGA_j\right) = \begin{cases} 1 - P\left(ds \geq DS_{i+1}|PGA_j\right) & i = 0 \\ P\left(ds \geq DS_i|PGA_j\right) - P\left(ds \geq DS_{i+1}|PGA_j\right) & 0 < i < nDS \\ P\left(ds \geq DS_i|PGA_j\right) & i = nDS \end{cases} \quad (3)$$

A common value of β is assumed to guarantee the ordinal nature of damage and preventing intersecting fragility functions (e.g. Lallemant et al. 2015; Ader et al. 2020; Rosti et al. 2021a, b). The same value of dispersion (β) is also enforced to all building typologies

(e.g. Coburn and Spence 2002; Karababa and Pomonis 2011). These conditions are met by simultaneously fitting the fragility curves on all damage levels and building typologies via the maximum likelihood estimate (MLE) approach:

$$(\boldsymbol{\theta}, \beta) = \arg\max[\log(L(\boldsymbol{\theta}, \beta)] = \arg\max\left[\log\left(\prod_{k=1}^{nTyp}\prod_{j=1}^{nPGA}\prod_{i=0}^{nDS}\frac{N_{jk}!}{n_{ijk}!}P(ds = DS_i|PGA_j, Typ_k)^{n_{ijk}}\right)\right]$$

(4)

where $nTyp$ is the number of building typologies, $nPGA$ is the number of PGA thresholds, nDS is the number of damage levels, N_{jk} is the total number of buildings of the kth building typology at the jth PGA threshold, n_{ijk} is the number of buildings of the kth building typology with damage level DS_i at the jth PGA threshold.

For each building typology, Table 2 summarises the parameters of the resulting lognormal fragility curves (i.e. θ_{DSi}, median PGA values and β, logarithmic standard deviation).

In line with existing studies (e.g. Braga et al. 1982; Dolce et al. 2003; Lagomarsino and Giovinazzi 2006), the mean level of damage, μ_D, attained at a given PGA threshold is defined as:

$$\mu_D(PGA_j) = \sum_{i=0}^{nDS} i \cdot P(ds = DS_i|PGA_j)$$

(5)

where nDS is the number of damage states and $P(ds = DS_i|PGA_j)$ is the probability of occurrence of the ith damage level, obtained from the previously determined fragility functions.

3 Identification of vulnerability classes and derivation of fragility curves via unsupervised machine learning techniques

The definition of vulnerability classes has been extensively addressed in the literature (e.g. Goretti and Di Pasquale 2004; Di Pasquale et al. 2005; Dolce et al. 2006; Dolce and Goretti 2015; Masi et al. 2021; Rosti et al. 2021a; Saretta et al. 2021). Most of those studies (e.g. Goretti and Di Pasquale 2004; Di Pasquale et al. 2005; Dolce et al. 2006; Dolce and Goretti

Table 1 Identification of building typologies based on the adopted building taxonomy. Numbers into brackets indicate the sample size

No of storeys	Masonry								RC	
	Irregular layout or poor-quality (IRR)				Regular layout and good-quality (REG)					
	Flexible diaphragms (F)		Rigid diaphragms (R)		Flexible diaphragms (F)		Rigid diaphragms (R)		Seismic design	
	NCD[a]	CD[b]	NCD	CD	NCD	CD	NCD	CD	Pre81	Post81
1	IRR-F-NCD-1 (11314)	IRR-F-CD-1 (2522)	IRR-R-NCD-1 (1344)	IRR-R-CD-1 (1537)	REG-F-NCD-1 (6909)	REG-F-CD-1 (3253)	REG-R-NCD-1 (2697)	REG-R-CD-1 (9101)	RC-Seismic-Pre81-1 (1426)	RC-Seismic-Post81-1 (1936)
2	IRR-F-NCD-2 (36704)	IRR-F-CD-2 (11243)	IRR-R-NCD-2 (5018)	IRR-R-CD-2 (7916)	REG-F-NCD-2 (10383)	REG-F-CD-2 (7486)	REG-R-NCD-2 (6886)	REG-R-CD-2 (28734)	RC-Seismic-Pre81-2 (5546)	RC-Seismic-Post81-2 (7691)
3	IRR-F-NCD-3 (15609)	IRR-F-CD-3 (5725)	IRR-R-NCD-3 (2062)	IRR-R-CD-3 (3124)	REG-F-NCD-3 (3099)	REG-F-CD-3 (2429)	REG-R-NCD-3 (1949)	REG-R-CD-3 (9595)	RC-Seismic-Pre81-3 (3612)	RC-Seismic-Post81-3 (5489)
4	IRR-F-NCD-4+ (2197)	IRR-F-CD-4+ (1091)	IRR-R-NCD-4+ (335)	IRR-R-CD-4+ (479)	REG-F-NCD-4+ (367)	REG-F-CD-4+ (282)	REG-R-NCD-4+ (429)	REG-R-CD-4+ (2056)	RC-Seismic-Pre81-4 (1635)	RC-Seismic-Post81-4 (1930)
5+									RC-Seismic-Pre81-5+ (1091)	RC-Seismic-Post81-5+ (713)

[a] NCD: without connecting devices (i.e. tie-rods and tie-beams)
[b] CD: with connecting devices (i.e. tie-rods and/or tie-beams)

Sample size	≤ 1000	> 1000 ≤ 2000	>2000 ≤4000	>4000 ≤8000	>8000 ≤16000	>16000

2015) resort to the association rule between structural typologies and vulnerability classes proposed by Braga et al. (1982), resulting from the best agreement between the Irpinia (1980) empirical damage data and the MSK scale. To remove possible subjectivity in the definition of vulnerability classes, Rosti et al. (2021a) employed a hierarchical agglomerative clustering algorithm for objectively merging predefined building typologies into three vulnerability classes of decreasing vulnerability. However, the methodological apparatus was not set into a probabilistic framework, as a univocal correspondence between structural typologies and vulnerability classes was proposed.

Aware of the advantage of removing subjectivity in the attribution of some choices and requiring a limited intervention of the expert in the definition of a suitable number of clusters in which partitioning the dataset, unsupervised machine learning algorithms have been used in disparate applications of earthquake engineering and seismology (e.g. Weatherhill and Burton 2009; Jayaram and Baker 2010; Rehman et al. 2014; Kotha et al. 2018; Mascandola et al. 2020; Xie et al. 2020). Clustering is one of the most common unsupervised machine learning algorithms, employed for drawing inferences from datasets of input data without labelled responses, based on some measures of similarity. In this context, K-means (Lloyd 1982) is an iterative algorithm, splitting data into a predefined number of mutually exclusive clusters. In other words, each observation is assigned to exactly one of the clusters, by minimising the distance between the data point and the centroid of the assigned cluster.

Fuzzy c-means (FCM) clustering (Bezdek 1981) is instead a soft version of K-means, where each data point has a fuzzy degree of belonging (u_{ij}) to each cluster. FCM is based on minimising the objective function, J_m:

$$J_m = \sum_{i=1}^{D} \sum_{j=1}^{N} u_{ij}^m \parallel x_i - c_j \parallel^2 \qquad (6)$$

where D is the total number of data points, N is the number of clusters, u_{ij} is the membership degree of x_i to the jth cluster, x_i is the ith data point, m controls the fuzzy overlapping among different clusters and c_j is the centroid of the jth cluster.

The algorithm proceeds as follows:

1. The u_{ij} values are randomly initialized
2. The centroids, c_j, of the clusters are computed as:

$$c_j = \frac{\sum_{i=1}^{D} u_{ij}^m x_i}{\sum_{i=1}^{D} u_{ij}^m} \qquad (7)$$

3. The u_{ij} values are updated:

$$u_{ij} = \frac{1}{\sum_{k=1}^{N} \left(\frac{\|x_i - c_j\|}{\|x_i - c_k\|} \right)^{\frac{2}{m-1}}} \qquad (8)$$

4. The objective function, J_m, is computed
5. Steps 2–4 are repeated until J_m improves by less than a specified minimum threshold or until after a specified maximum number of iterations.

Table 2 Parameters (i.e. median and logarithmic standard deviation) of the typological lognormal fragility curves

Building typology	θ_{DS1} (g)	θ_{DS2} (g)	θ_{DS3} (g)	θ_{DS4} (g)	θ_{DS5} (g)	β (-)
IRR-F-NCD-1	0.198	0.261	0.296	0.379	0.569	
IRR-F-NCD-2	0.134	0.190	0.228	0.304	0.490	
IRR-F-NCD-3	0.102	0.160	0.194	0.265	0.464	
IRR-F-NCD-4+	0.090	0.144	0.176	0.239	0.442	
IRR-F-CD-1	0.191	0.298	0.349	0.451	0.655	
IRR-F-CD-2	0.141	0.217	0.265	0.366	0.586	
IRR-F-CD-3	0.108	0.178	0.220	0.304	0.500	
IRR-F-CD-4+	0.081	0.140	0.176	0.247	0.463	
IRR-R-NCD-1	0.215	0.329	0.363	0.425	0.696	
IRR-R-NCD-2	0.160	0.240	0.287	0.371	0.590	
IRR-R-NCD-3	0.126	0.200	0.242	0.337	0.560	
IRR-R-NCD-4+	0.107	0.175	0.209	0.264	0.490	
IRR-R-CD-1	0.250	0.394	0.462	0.644	0.831	
IRR-R-CD-2	0.195	0.322	0.377	0.499	0.757	
IRR-R-CD-3	0.146	0.247	0.296	0.409	0.670	
IRR-R-CD-4+	0.111	0.168	0.211	0.314	0.675	
REG-F-NCD-1	0.311	0.457	0.546	0.725	0.943	
REG-F-NCD-2	0.185	0.305	0.361	0.494	0.800	
REG-F-NCD-3	0.146	0.266	0.324	0.427	0.678	
REG-F-NCD-4+	0.115	0.220	0.294	0.358	0.648	
REG-F-CD-1	0.405	0.711	0.903	1.301	1.857	0.705
REG-F-CD-2	0.239	0.469	0.538	0.704	1.154	
REG-F-CD-3	0.187	0.359	0.439	0.558	0.981	
REG-F-CD-4+	0.149	0.318	0.369	0.462	0.945	
REG-R-NCD-1	0.336	0.561	0.764	0.966	1.043	
REG-R-NCD-2	0.271	0.429	0.504	0.688	0.971	
REG-R-NCD-3	0.231	0.387	0.462	0.590	0.806	
REG-R-NCD-4+	0.189	0.352	0.409	0.582	0.748	
REG-R-CD-1	0.408	0.730	0.956	1.170	2.165	
REG-R-CD-2	0.366	0.661	0.822	0.945	1.248	
REG-R-CD-3	0.319	0.596	0.733	0.959	1.254	
REG-R-CD-4+	0.263	0.548	0.677	0.837	1.722	
RC-Seismic-Pre81-1	0.315	0.459	0.690	1.063	2.578	
RC-Seismic-Pre81-2	0.283	0.516	0.719	0.981	1.655	
RC-Seismic-Pre81-3	0.195	0.386	0.573	0.816	1.450	
RC-Seismic-Pre81-4	0.157	0.313	0.453	0.754	1.369	
RC-Seismic-Pre81-5+	0.128	0.226	0.337	0.743	1.545	
RC-Seismic-Post81-1	0.379	0.738	1.112	2.907	2.907	
RC-Seismic-Post81-2	0.303	0.681	0.958	1.390	2.230	
RC-Seismic-Post81-3	0.231	0.543	0.824	1.285	4.072	
RC-Seismic-Post81-4	0.187	0.400	0.663	1.098	2.290	
RC-Seismic-Post81-5+	0.142	0.270	0.502	1.128	2.354	

In this study, fuzzy c-means clustering is employed to split mean damage values, observed at preselected ground motion intensity levels, into a predefined number of clusters, representing the vulnerability classes. Differently from K-means clustering, the use of the FCM algorithm allows each observational mean damage data point to belong to multiple vulnerability classes with different membership degree.

According to the number of vulnerability classes considered by the EMS-98 (Grünthal et al. 1998), six vulnerability classes of decreasing vulnerability (from A to F) are considered, each split into two subgroups based on the construction material (i.e. masonry and RC). The reason of separating masonry from RC vulnerability classes derives from the different distance among damage levels observed in the fragility functions obtained for the two families of structural typologies (Table 2). Six vulnerability classes are considered for masonry (i.e. A1, B1, C1, D1, E1, F1) whereas four out of six vulnerability classes are introduced for RC buildings (i.e. C2, D2, E2, F2), for which higher vulnerable vulnerability classes (i.e. A2 and B2) lack. Vulnerability classes A2 and B2 could indeed refer to RC buildings without seismic design (i.e. RC buildings designed for gravity loads only), which are not available from the L'Aquila post-earthquake dataset, or to other more vulnerable RC building typologies not contemplated by the adopted building taxonomy.

Based on the above considerations, FCM clustering is separately applied to empirical mean damage values associated with masonry and RC building typologies. An overlapping coefficient, m, equal to 2 is considered, as higher values imply fuzzier boundaries among the different vulnerability classes, leading to less distinct fragility curves.

Considering the masonry dataset, the outermost vulnerability classes A1 and F1 are first defined, by respectively pooling together mean damage values of the two most vulnerable (i.e. IRR-F-NCD-4 + and IRR-F-CD-4+) and less vulnerable (i.e. REG-F-CD-1 and REG-R-CD-1) building typologies. The most and least vulnerable building typologies are identified by comparing observational typological mean damage values and then considering the two building typologies with higher and lower mean damage values, respectively. The remaining typological mean damage values are then split into four clusters (i.e. B1, C1, D1, E1) via FCM clustering. In this way, the presence of building typologies more and less vulnerable than the outermost vulnerability classes (i.e. A1 and F1) is avoided. A similar strategy is applied for the definition of the vulnerability classes for RC buildings. In this case, the number of clusters in which partitioning the RC mean damage dataset is set equal to two, allowing for the identification of the inner RC vulnerability classes (i.e. D2 and E2). The outermost vulnerability classes (i.e. C2 and F2) are instead obtained by considering empirical mean damage values of the most (i.e. RC-Seismic-Pre81-5 +) and least (i.e. RC-Seismic-Post81-1) vulnerable RC building typologies.

The implementation of FCM clustering provides, for a given PGA threshold, the membership degree of each mean damage value to different vulnerability classes. The highest membership degree denotes the most probable vulnerability class. As a result, empirical mean damage data points (Fig. 6a) are attributed to the most likely vulnerability class (i.e. the one with the higher membership degree) and to the other vulnerability classes, with different membership degree (Fig. 6b).

By interpreting the membership degree of a given mean damage value to the different vulnerability classes in terms of frequency, the number of buildings within a given building typology belonging to each vulnerability class is obtained, for each PGA threshold. Repetition of this procedure for all mean damage values and PGA thresholds leads to the definition of damage probability matrices of vulnerability classes, then approximated by fitting fragility functions (Fig. 7). Median PGA values associated with the different damage levels

are obtained via the maximum likelihood approach, by imposing the same dispersion value (β) resulting from the typological fragility curves (Sect. 2.4). Table 3 collects the parameters (i.e. θ_{DSi}, median PGA values and β, logarithmic standard deviation) of the fragility curves of the ten vulnerability classes. For each level of damage, resulting fragility curves are compared in Fig. 8, showing a clear hierarchy among the different vulnerability classes.

4 Proposed vulnerability model

Besides fragility functions, providing the expected damage distribution in the different levels as a function of the ground motion severity, a thorough vulnerability model should supply indications on the vulnerability classification of the exposed building stock. In this context, an ad-hoc strategy is built up to determine the degree of belonging of each building typology to multiple vulnerability classes. Fragility functions derived for vulnerability classes are linearly combined and optimal coefficients of the linear combination, representing the degrees of belonging of the selected building typology to vulnerability classes, are obtained by using typological fragility curves as a target.

Based on this procedure, the fragility curve associated with the ith damage level of the jth building typology can be approximated as:

$$\Phi\left[\frac{\log\left(PGA/\theta_{DSij}\right)}{\beta}\right] \approx \sum_{k=0}^{NClasses-1} w_{jk}\Phi\left[\frac{\log\left(PGA/\theta_{DSik}\right)}{\beta}\right] \tag{9}$$

where θ_{DSij} is the median PGA value of the fragility function of the ith damage level of the jth building typology, *NClasses* indicates the total number of the considered vulnerability classes, w_{jk} denotes the degree of belonging of the jth building typology to the kth vulnerability class, θ_{DSik} is the median PGA value of the fragility function of the ith damage level of the kth vulnerability class and β is the logarithmic standard deviation, which is constant among damage levels, building typologies and vulnerability classes.

The definition of the linear combination coefficients, w_{jk}, results from an optimisation problem, minimising the global deviation between the sets of approximating and empirically-derived typological fragility curves. As the optimisation problem aims at providing the coefficients of the linear combination better reproducing target fragility functions, optimal coefficients may refer to non-adjacent vulnerability classes. This issue is counteracted by introducing a probability distribution for describing the trend of the w_{jk} values. Considering the advantage of being fully described by a single parameter, the binomial model is selected. As two sets of different vulnerability classes are defined for masonry and RC typologies, two binomial distributions are introduced. One binomial distribution refers to masonry vulnerability classes and is fully described by the binomial parameter, y_{mas}, with k_{mas} ranging from 0 (F1) to 5 (A1), as indicated in Eq. (10). The other one refers to RC vulnerability classes and is fully determined by the binomial parameter, y_{RC}, with k_{RC} varying from 0 (F2) to 3 (C2), as per Eq. (11). The degree of belonging of the jth building typology to the kth vulnerability class, w_{jk}, is then expressed by jointly using the binomial distributions defined in Eqs. (10) and (11), suitably scaled by the factor c_{mas} (Eq. (12)). The scaling coefficient, c_{mas}, which is unknown together with the binomial parameters y_{mas} and y_{RC}, indicates the weight that the binomial distribution of "masonry" vulnerability classes takes in the global distribution of the w_{jk} coefficients. To ensure that the w_{jk} coefficients add up
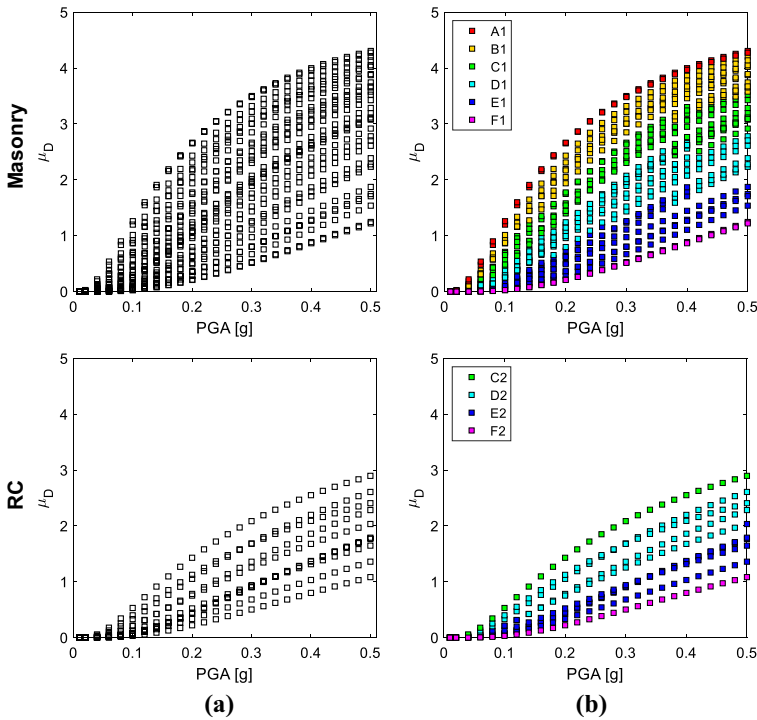
Fig. 6 Empirically-derived mean damage values of masonry and RC building typologies versus PGA (**a**) and identification of the most probable vulnerability class (i.e. higher membership degree) based on FCM clustering algorithm (**b**)

to 1, the "RC" binomial distribution is scaled by the complementary to 1 of the c_{mas} coefficient (Eq. (12)).

$$w_{jk,mas} = \frac{5!}{k_{mas}!\left(5 - k_{mas}\right)!}\left(\frac{y_{j,mas}}{5}\right)^{k_{mas}}\left(1 - \frac{y_{j,mas}}{5}\right)^{5-k_{mas}} \tag{10}$$

$$w_{jk,RC} = \frac{3!}{k_{RC}!\left(3 - k_{RC}\right)!}\left(\frac{y_{j,RC}}{3}\right)^{k_{RC}}\left(1 - \frac{y_{j,RC}}{3}\right)^{3-k_{RC}} \tag{11}$$

$$w_{jk} = c_{j,mas}w_{jk,mas} + \left(1 - c_{j,mas}\right)w_{jk,RC} \tag{12}$$

The joint use of two binomial models for describing the w_{jk} distribution allows for more flexibility. Besides the cases where masonry ($c_{mas}=1$) and RC ($c_{mas}=0$) vulnerability classes are used only, the combined use of masonry and RC vulnerability classes can be convenient for some building typologies, to suitably account for a different distance among the fragility curves of the different damage states.

By substituting Eqs. (10) and (11) into Eq. (12) and then replacing Eq. (12) into Eq. (9), the fragility curve of damage level DS_i of the jth building typology can be approximated as:
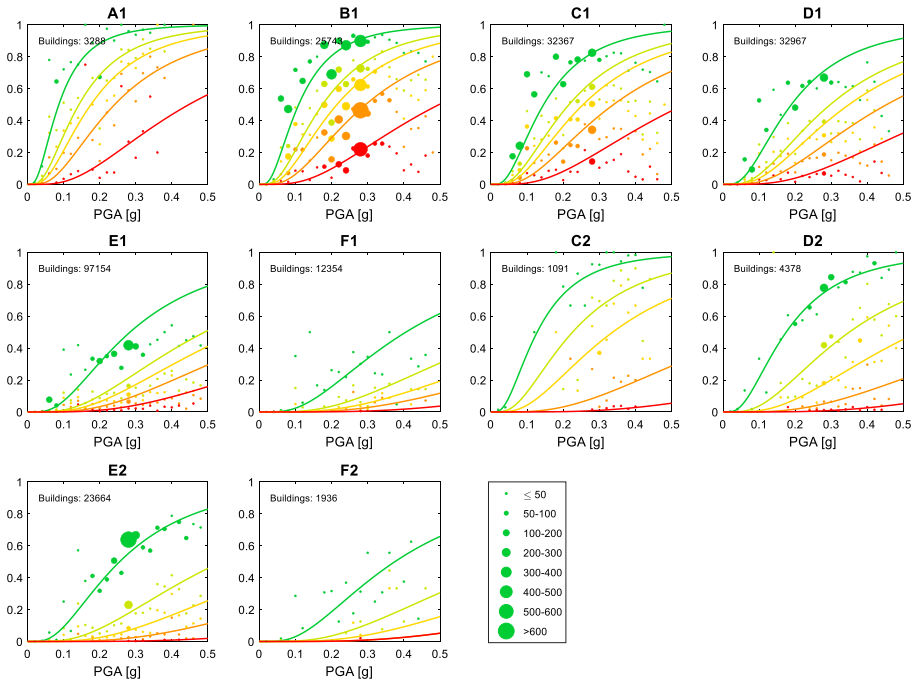
Fig. 7 Fragility curves of the ten vulnerability classes, resulting from FCM clustering of the observed mean damage values of Masonry (1) and RC (2) building typologies. Numbers in the legend indicate the sample size

Table 3 Parameters (i.e. median and logarithmic standard deviation) of the lognormal fragility curves of the ten vulnerability classes, resulting from FCM clustering of the observed mean damage values

Construction material	Vulnerability class	θ_{DS1} (g)	θ_{DS2} (g)	θ_{DS3} (g)	θ_{DS4} (g)	θ_{DS5} (g)	β (–)
Masonry	A1	0.087	0.143	0.176	0.241	0.447	
	B1	0.109	0.175	0.215	0.294	0.497	
	C1	0.146	0.216	0.255	0.339	0.534	
	D1	0.191	0.301	0.352	0.459	0.702	
	E1	0.286	0.499	0.600	0.751	1.034	
	F1	0.406	0.714	0.921	1.156	1.771	0.705
RC	C2	0.128	0.226	0.337	0.741	1.544	
	D2	0.175	0.350	0.542	0.886	1.579	
	E2	0.255	0.538	0.795	1.179	2.149	
	F2	0.375	0.716	1.021	1.576	1.576	

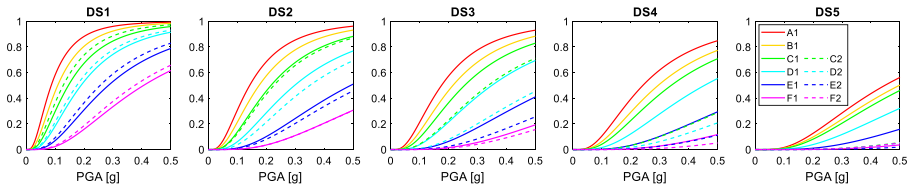**Fig. 8** Comparison of the fragility curves of the vulnerability classes, resulting from FCM clustering of the observed mean damage values

$$\Phi\left[\frac{\log\left(PGA/\theta_{DSij}\right)}{\beta}\right] \approx c_{j,mas}\sum_{k_{mas}=0}^{5}\left[\frac{5!}{k_{mas}!\left(5-k_{mas}\right)!}\left(\frac{y_{j,mas}}{5}\right)^{k_{mas}}\left(1-\frac{y_{j,mas}}{5}\right)^{5-k_{mas}}\right]$$

$$\Phi\left[\frac{\log\left(PGA/\theta_{DSik_{mas}}\right)}{\beta}\right] + \left(1-c_{j,mas}\right)\sum_{k_{RC}=0}^{3}\left[\frac{3!}{k_{RC}!\left(3-k_{RC}\right)!}\left(\frac{y_{j,RC}}{3}\right)^{k_{RC}}\left(1-\frac{y_{j,RC}}{3}\right)^{3-k_{RC}}\right]\Phi\left[\frac{\log\left(PGA/\theta_{DSik_{RC}}\right)}{\beta}\right]$$

$$(13)$$

For each building typology, the unknowns to be determined are three, i.e. the parameters of the binomial distribution defined for masonry, y_{mas}, and RC, y_{RC}, vulnerability classes and the fraction of the masonry binomial distribution, c_{mas}, to be considered within the global w_{jk} distribution. Optimal values of the parameters y_{mas}, y_{RC} and c_{mas} result from a constrained optimisation problem minimising the global deviation between the sets of approximating and target typological fragility functions. Parameters y_{mas}, y_{RC} and c_{mas} are constrained between 0 and 5, 0 and 3 and 0 and 1, respectively.

Table 4 collects the parameters c_{mas}, y_{mas}, y_{RC} which allow for defining the degrees of belonging of each building typology to vulnerability classes. Figures 9, 10 and 11 show the distribution of the degrees of belonging of building typologies to vulnerability classes. Figures also compare the approximating fragility curves (continuous lines) resulting from linearly combining the fragility functions of the different vulnerability classes, suitably accounting for their degree of belonging, with the empirically-derived typological fragility curves (dotted lines). Vertical dashed lines represent the weighted mean vulnerability class of the binomial distribution.

Fragility curves derived from the linear combination of the fragility functions of the vulnerability classes and accounting for the membership degree of the building typology to the vulnerability classes (continuous lines) generally well approximate the empirically-derived typological fragility curves (dotted lines), suggesting the suitability of the adopted strategy (Figs. 9, 10 and 11). In case of some masonry building typologies (i.e. IRR-R-CD-3/4+, REG-F-NCD-2/3/4+, REG-F-CD-2/3/4+, REG-R-NCD/CD-4+), the combined use of vulnerability classes defined for both masonry and RC buildings allows for better reproducing target typological fragility functions.

In line with the EMS-98, Fig. 12 provides the vulnerability classification of the existing building stock, resulting from the adopted procedure. The proposed vulnerability table supplies a synthetic vulnerability representation of each building typology, in terms of the weighted mean vulnerability class of the binomial distribution (squared markers), fully describing the degree of belonging of a given typology to multiple vulnerability classes. In the figure, black squared markers indicate the weighted mean "masonry" vulnerability class, whereas white squared markers denote the weighted mean "RC" vulnerability class. Bars, associated with each building typology, indicate the fraction of "masonry"/"RC" binomial distributions to be considered within the

global w_{jk} distribution. Grey solid hatch provides the value of the c_{mas} coefficient. In other words, fully grey bars correspond to the case of $c_{mas} = 1$ (i.e. only "masonry" vulnerability classes are considered), totally empty bars refer to the case of $c_{mas} = 0$ (i.e. only "RC" vulnerability classes are accounted for) whereas partially-filled bars correspond to the case of $0 < c_{mas} < 1$ (i.e. both "masonry" and "RC" vulnerability classes are considered). The proposed vulnerability table also permits to categorise and compare the seismic vulnerability of buildings based on the presence or absence of specific typological building attributes.

Results show that the layout and quality of the masonry fabric significantly affect the seismic vulnerability of masonry buildings, being irregular layout or poor-quality masonry building typologies more vulnerable than those made of good-quality materials with regular texture. Besides the characteristics of the masonry fabric, other building attributes, such as the in-plane stiffness of intermediate diaphragms and the presence of aseismic devices, impact the seismic vulnerability of masonry buildings. The presence of rigid diaphragms indeed improves the seismic behaviour of masonry constructions with respect to those with flexible horizontal structures. Also, the presence of connecting devices (i.e. tie-rods and/or tie-beams) has a beneficial effect on the seismic vulnerability of masonry buildings. Building code evolution improves the seismic response of RC buildings. Given the building height, RC buildings seismically-designed after 1981 are indeed less vulnerable than those seismically-designed according to obsolete (pre-1981) seismic prescriptions. The number of storeys strongly impacts the seismic vulnerability of both masonry and RC buildings,

Table 4 Parameters required for defining the degrees of belonging of each building typology to vulnerability classes

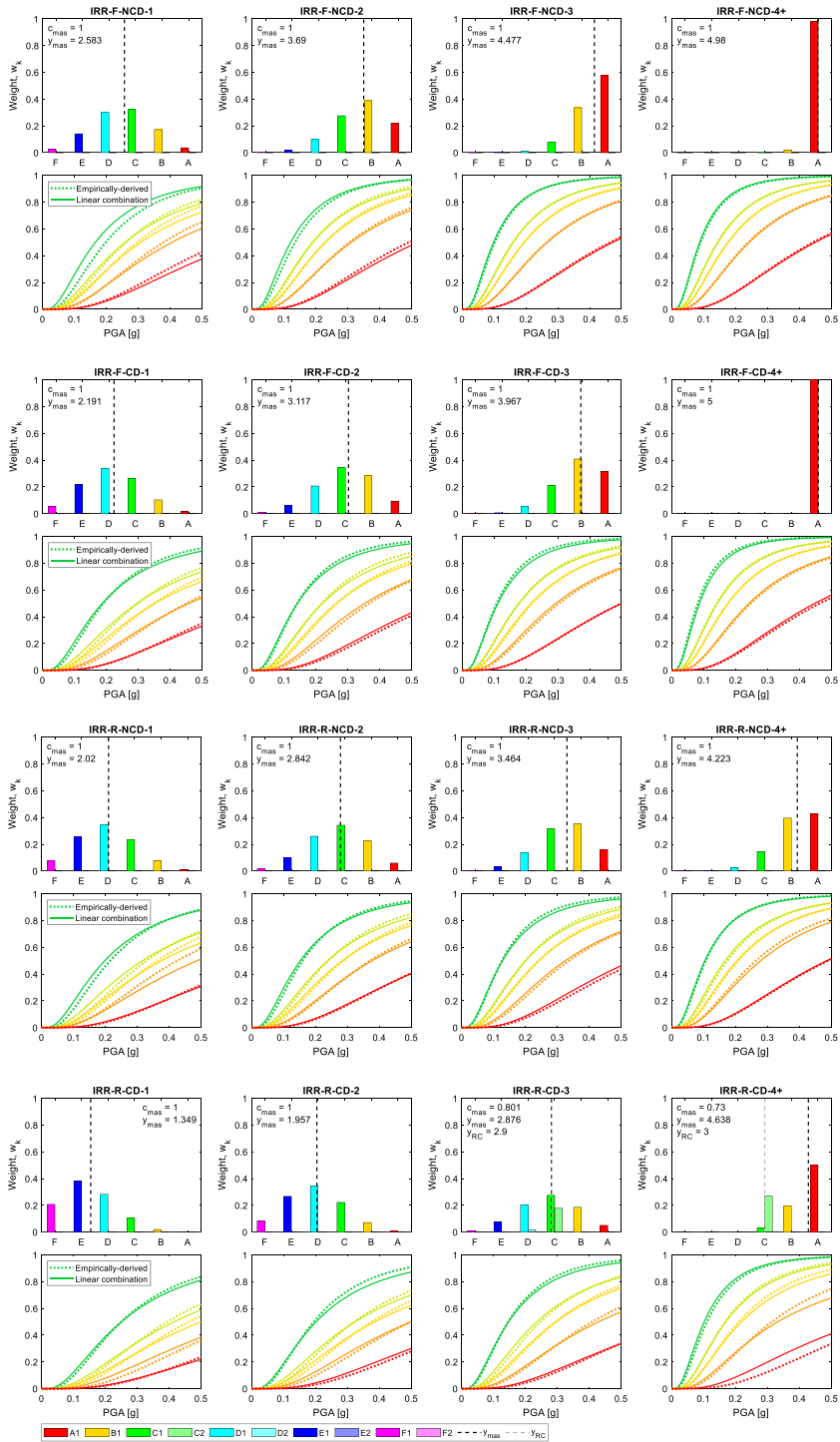| Building typology	c_{mas}	y_{mas}	y_{RC}	Building typology	c_{mas}	y_{mas}	y_{RC}
IRR-F-NCD-1	1	2.583	–	REG-F-CD-2	0.558	1.266	1.243
IRR-F-NCD-2	1	3.690	–	REG-F-CD-3	0.543	2.040	1.807
IRR-F-NCD-3	1	4.477	–	REG-F-CD-4+	0.546	2.781	2.061
IRR-F-NCD-4+	1	4.980	–	REG-R-NCD-1	1	0.454	–
IRR-F-CD-1	1	2.191	–	REG-R-NCD-2	1	1.125	–
IRR-F-CD-2	1	3.117	–	REG-R-NCD-3	1	1.459	–
IRR-F-CD-3	1	3.967	–	REG-R-NCD-4+	0.662	2.023	1.827
IRR-F-CD-4+	1	5.000	–	REG-R-CD-1	1	0.000	–
IRR-R-NCD-1	1	2.020	–	REG-R-CD-2	1	0.243	–
IRR-R-NCD-2	1	2.842	–	REG-R-CD-3	1	0.463	–
IRR-R-NCD-3	1	3.464	–	REG-R-CD-4+	0.296	0.956	0.897
IRR-R-NCD-4+	1	4.223	–	RC-Seismic-Pre81-1	0	–	0.856
IRR-R-CD-1	1	1.349	–	RC-Seismic-Pre81-2	0	–	0.873
IRR-R-CD-2	1	1.957	–	RC-Seismic-Pre81-3	0	–	1.739
IRR-R-CD-3	0.801	2.876	2.90	RC-Seismic-Pre81-4	0	–	2.307
IRR-R-CD-4+	0.730	4.638	3.00	RC-Seismic-Pre81-5+	0	–	2.999
REG-F-NCD-1	1	0.918	–	RC-Seismic-Post81-1	0	–	0.000
REG-F-NCD-2	0.796	2.079	2.684	RC-Seismic-Post81-2	0	–	0.411
REG-F-NCD-3	0.739	2.885	2.351	RC-Seismic-Post81-3	0	–	1.002
REG-F-NCD-4+	0.701	3.846	2.423	RC-Seismic-Post81-4	0	–	1.616
REG-F-CD-1	1	0.000	–	RC-Seismic-Post81-5+	0	–	2.391

Fig. 9 Degrees of belonging of irregular layout or poor-quality masonry building typologies to vulnerability classes
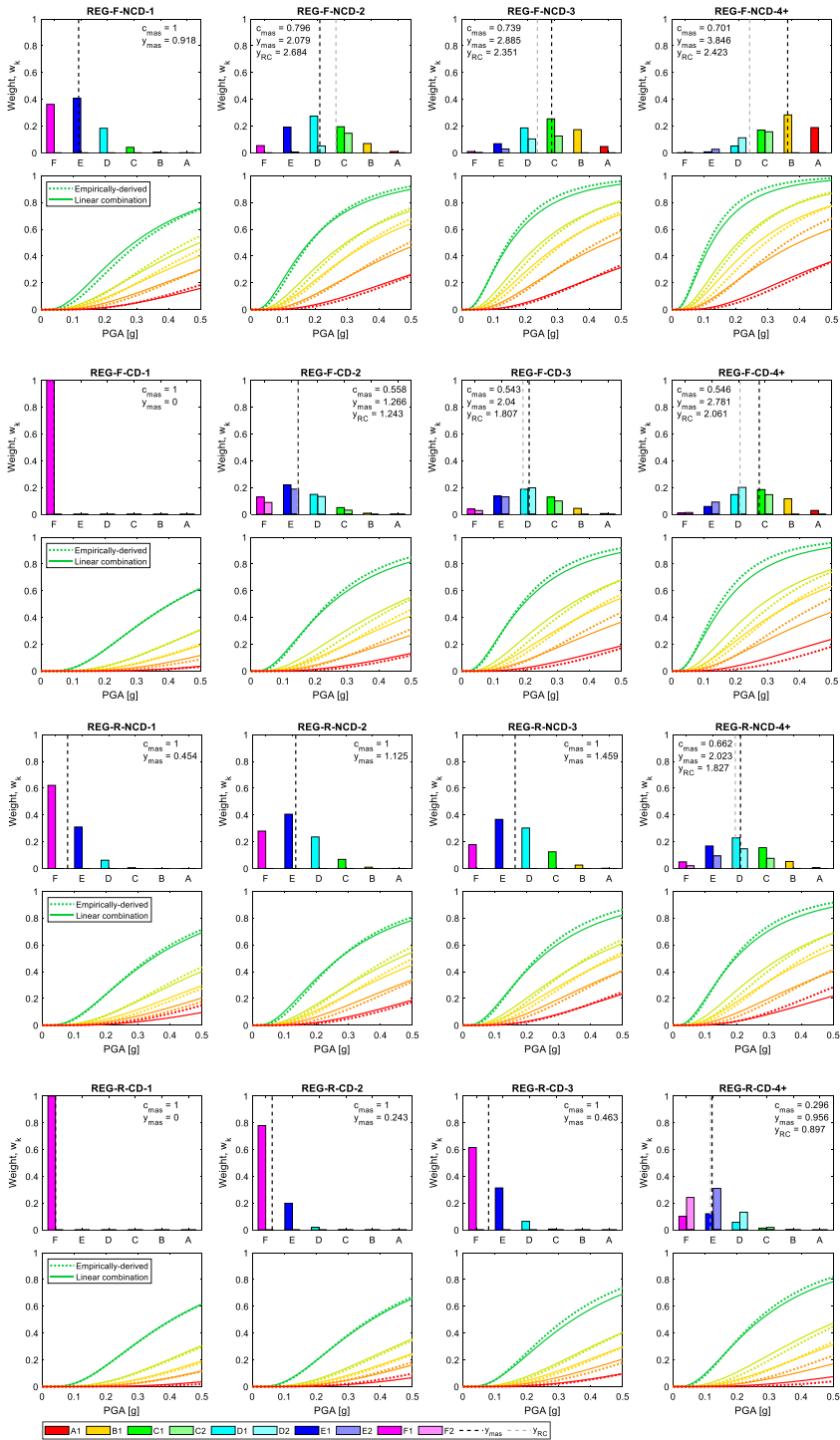
Fig. 10 Degrees of belonging of regular layout and good-quality masonry building typologies to vulnerability classes
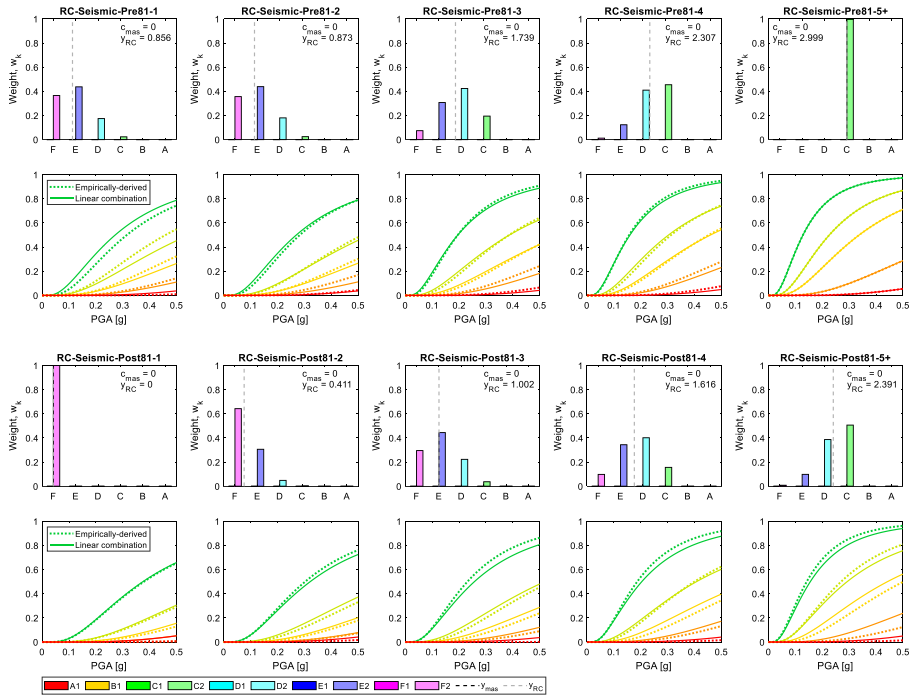
Fig. 11 Degrees of belonging of RC building typologies to vulnerability classes

suggesting the need and the importance of accounting for building height in the vulnerability classification of the existing building stock.

5 Example of application

To better clarify the steps involved in the application of the vulnerability model, let us consider two typologies of 3-storeys masonry buildings, both subjected to a PGA of 0.20 g:

- The first typology is made of undressed stone masonry, timber floor diaphragms, without suitable connections like steel ties or RC ring beams (i.e. belonging to the IRR-F-NCD-3 typology);
- The second one is made of clay brick masonry, timber floors and steel tie-rods (i.e. REG-F-CD-3 typology).

The probability of exceeding the five damage states for a PGA equal to 0.2 g for the different vulnerability classes is reported in Table 5. It is simply obtained using the lognormal model with parameters set in Table 3.

Table 4 provides for the first typology $c_{mas} = 1$ and $y_{mas} = 4.477$, whereas for the second one $c_{mas} = 0.543$, $y_{mas} = 2.040$ and $y_{RC} = 1.807$. The resulting weights resulting from the application of Eq. (12) are reported in Table 6.
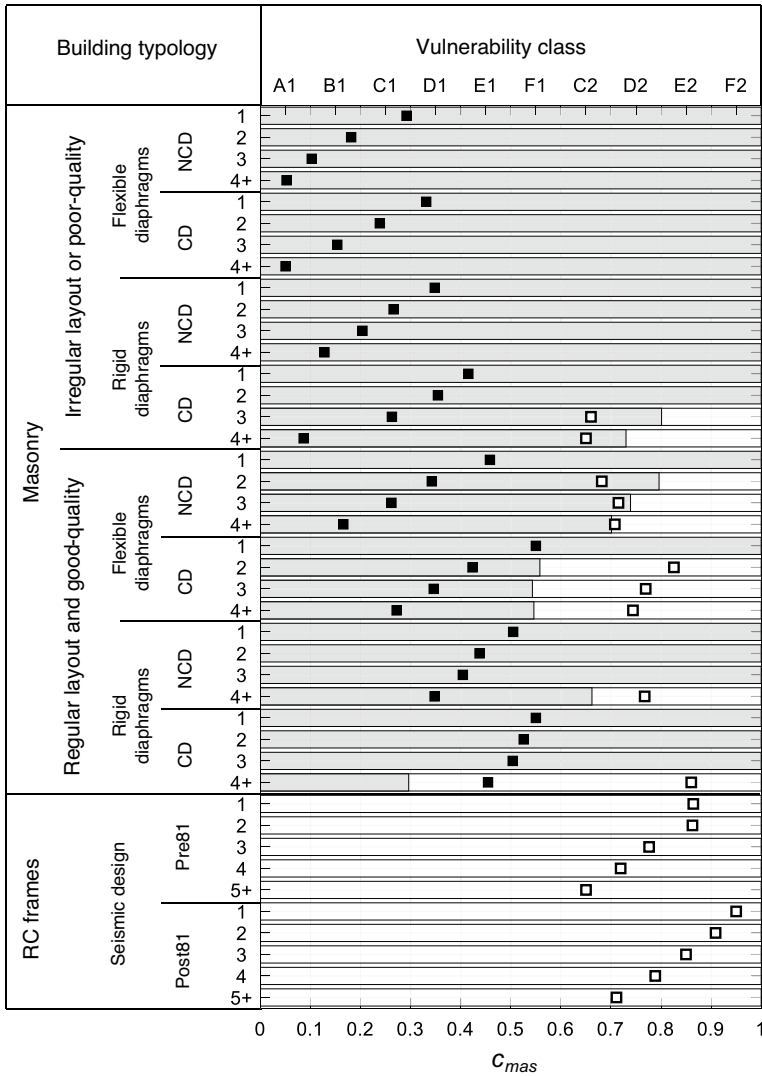
Fig. 12 Proposed vulnerability table for masonry and RC building typologies. Black and white squared markers indicate the weighted mean "masonry" and "RC" vulnerability classes, respectively, fully characterising the distribution of the degrees of belonging to vulnerability classes. Grey solid hatch provides the value of the c_{mas} coefficient

The weighted combination of the probabilities reported in Table 5 provides the following probabilities of exceeding the five damage states:

- $P_{DS1} = 0.83$, $P_{DS2} = 0.62$, $P_{DS3} = 0.51$, $P_{DS4} = 0.34$, and $P_{DS5} = 0.11$ for IRR-F-NCD-3;
- $P_{DS1} = 0.51$, $P_{DS2} = 0.25$, $P_{DS3} = 0.16$, $P_{DS4} = 0.08$, and $P_{DS5} = 0.03$ for REG-F-CD-3.

Table 5 Probability of exceeding the 5 damage states for the 10 vulnerability classes at PGA = 0.2 g

	A1	B1	C1	D1	E1	F1	C2	D2	E2	F2
$P_{DS1}(0.2g)$	0.881	0.805	0.672	0.526	0.306	0.158	0.737	0.575	0.365	0.186
$P_{DS2}(0.2g)$	0.683	0.575	0.457	0.281	0.097	0.036	0.431	0.214	0.080	0.035
$P_{DS3}(0.2g)$	0.572	0.459	0.365	0.211	0.060	0.015	0.230	0.079	0.025	0.010
$P_{DS4}(0.2g)$	0.396	0.292	0.227	0.119	0.030	0.006	0.032	0.017	0.006	0.002
$P_{DS5}(0.2g)$	0.127	0.098	0.082	0.037	0.010	0.001	0.002	0.002	0.000	0.002

Table 6 Weights for the combination of the fragility curves associated with the different vulnerability classes for the two building typologies considered in this sample application

Building Typology	w_{A1}	w_{B1}	w_{C1}	w_{D1}	w_{E1}	w_{F1}	w_{C2}	w_{D2}	w_{E2}	w_{F2}
IRR-F-NCD-3	0.576	0.336	0.079	0.009	0.000	0.000	–	–	–	–
REG-F-CD-3	0.006	0.045	0.129	0.188	0.136	0.039	0.100	0.198	0.130	0.029

The expected distribution of the buildings belonging to the two considered building types among the different damage states can be immediately deduced from the probabilities of exceedance. The results reported in Fig. 13 show the significant difference in vulnerability between these masonry building typologies subject to the same level of seismic shaking (PGA = 0.2 g).

6 Conclusions

This paper presents a novel and comprehensive vulnerability model for masonry and RC buildings, representative of the Italian built environment, relying on a data-driven approach. The proposed vulnerability model is based on the definition of classes characterised by fragility curves in PGA and on the relationship of belonging to the different vulnerability classes of structural types identified by essential attributes such as vertical and horizontal structure, structural details, level of seismic design and number of stories.

With respect to previous studies, peak ground acceleration, extrapolated from updated INGV shakemaps (Michelini et al. 2020), is employed for the seismic input definition, making the present model easily usable for vulnerability and risk applications. A robust post-earthquake damage database (Dolce et al. 2019), collected after the L'Aquila (2009) seismic event, is employed, allowing for suitable consideration of the completeness of the post-earthquake field surveys and for the negative evidence of damage in the municipalities less affected by the ground shaking. Attention is first devoted to data processing, involving seismic input characterisation, typological classification of the exposed building stock and definition of a damage metric based on the EMS-98 damage states. A suitable statistical model and fitting technique are then employed for deriving typological fragility curves and mean damage values, as a function of the selected intensity measure.

One of the original contributions of this work is the use of machine learning techniques for the objective identification of ten vulnerability classes, starting from seismic damages
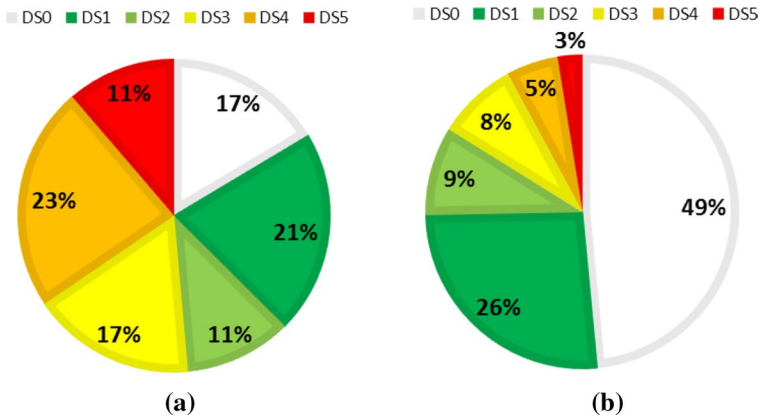
Fig. 13 Expected damage distribution for buildings belonging to IRR-F-NCD-3 (**a**) and REG-F-CD-3 (**b**) structural typologies subject to PGA = 0.2 g

observed on several Italian building typologies. Soft clustering is applied to empirically-derived mean damage values of the different building typologies, at preselected PGA thresholds. Damage distributions are hence obtained for each vulnerability class and then fitted by fragility functions. The adoption of a constant dispersion value for all damage levels and vulnerability classes ensures the hierarchy among the different damage levels, given the vulnerability class, and among the different vulnerability classes, which appear distinct and separated one from the other. In this study, vulnerability classes derived from masonry and RC building data are distinguished to account for the different distance among damage levels observed in the empirically-derived typological fragility functions.

In line with the conceptual framework of the macroseismic method (Lagomarsino and Giovinazzi 2006), accounting for the uncertainty in the attribution of building types to vulnerability classes, the degrees of belonging of each building typology to multiple vulnerability classes are determined. To this aim, a constrained optimisation problem, using empirically-derived typological fragility curves as a target, is set up. For a given building typology, the weighted mean vulnerability class is provided, allowing to fully characterise the distribution of the degrees of belonging to vulnerability classes, under the assumption of binomial distribution. Comparison of approximating and empirically-derived fragility functions of the considered building typologies shows the appropriateness of the adopted strategy, which results in a thorough vulnerability model consistently defined for both masonry and RC buildings.

Similarly to the EMS-98, a vulnerability table provides parameters to model the uncertain association of buildings belonging to typologies identified by selected structural features with the different seismic vulnerability classes. In this context, the availability of a robust post-earthquake database gathering both typological and damage information (Dolce et al. 2019) allows for improving the definition of structural types typical of the Italian building stock.

Comparison of the weighted mean vulnerability class of the different building typologies points out the higher vulnerability of irregular layout or poor-quality masonry buildings with respect to those with regular layout and good-quality materials, in line with observations from post-earthquake field surveys (e.g. Saatcioglu and Bruneau 1993; Penna et al. 2014; Sorrentino et al. 2019). The presence of rigid horizontal

structures as well as the presence of aseismic devices (e.g. appropriate wall-to-wall and wall-to-diaphragm connections) enhances the seismic response of masonry buildings. The level of seismic design has a clear beneficial effect on the seismic vulnerability of RC constructions, being RC buildings seismically-designed before 1981 more vulnerable that the corresponding ones complying with updated seismic design criteria. In line with other studies (e.g. Rota et al. 2011), the obtained results point out the role of the number of storeys on the empirical seismic vulnerability of both masonry and RC buildings, also suggesting consideration of this parameter as a possible future improvement of existing macroseismic scales.

Results provided in this paper (i.e. parameters of fragility curves derived for building typologies and for vulnerability classes, degrees of belonging of building typologies to vulnerability classes and the proposed vulnerability table) can be used for varied seismic vulnerability and risk applications, provided the similarity in the seismic hazard and exposed building stock of the area selected for application.

The availability of new damage data from post-earthquake observations will allow to further strengthen the vulnerability model and to verify or adapt its applicability to different territories and built environments. In particular, the integration of data on RC buildings without seismic design would allow completing the model towards the classes and types of greater vulnerability for which, in the current version, classes A2 and B2 have been envisaged.

Declarations

Conflict of interest All authors declare that they have no conflict of interest.

References

Ader T, Grant DN, Free M, Villani M, Lopez J, Spence R (2020) An unbiased estimation of empirical lognormal fragility functions with uncertainties on the ground motion intensity measure. J Earthq Eng 24(7):1115–1133

Agresti A (2002) Categorical data analysis. John Wiley & Sons, Inc., Hoboken, New Jersey, ISBN 0-471-36093-7

Baggio C, Bernardini A, Colozza R, Corazza L, Della Bella M, Di Pasquale G, Dolce M, Goretti A, Martinelli A, Orsini G, Papa F, Zuccaro G (2007) Field manual for post-earthquake damage and safety assessment and short-term countermeasures (AeDES). Translation from Italian: Rota M, Goretti A. JRC Scientific and Technical Reports, EUR 22868 EN-2007

Bernardini A, Salmaso L, Solari A (2008) Statistical evaluation of vulnerability and expected seismic damage of residential buildings in the Veneto-Friuli area (NE Italy). Boll Geofis Teor Appl 49(3–4):427–446

Bernardini A, Lagomarsino S, Mannella A, Martinelli A, Milano L, Parodi S (2011) Forecasting seismic damage scenarios of residential buildings from rough inventories: a case-study in the Abruzzo Region (Italy). Proc IMech E Part OJ Risk Reliab 224:279–296

Bezdek JC (1981) Pattern recognition with fuzzy objective function algorithms. Plenum Press, New York

Braga F, Dolce M, Liberatore D (1982) A statistical study on damaged buildings and an ensuing review of the M.S.K.-76 scale. Proceedings of the 7th European Conf on Earthq Eng, Athens, Greece

Brando G, Cianchino G, Rapone D, Spacone E, Biondi S (2021) A CARTIS-based method for the rapid seismic vulnerability assessment of minor Italian historical centres. Int J Disaster Risk Reduct 63:102478

Charvet I, Ioannou I, Rossetto T, Suppasri A, Imamura F (2014) Empirical fragility assessment of buildings affected by the 2011 Great East Japan tsunami using improved statistical models. Nat Hazards 73:951–973

Coburn A, Spence R (2002) Earthquake protection. Wiley, London, ISBN: 978-0-470-84923-1.

da Porto F, Donà M, Rosti A, Rota M, Lagomarsino S, Cattari S, Borzi B, Onida M, De Gregorio D, Perelli FL, Del Gaudio C, Ricci P, Speranza E (2021) Comparative analysis of the fragility curves for italian residential masonry and RC buildings. Bull Earthq Eng 19:3209–3252

Del Gaudio C, De Martino G, Di Ludovico M, Manfredi G, Prota A, Ricci P, Verderame GM (2017) Empirical fragility curves from damage data on RC buildings after the 2009 L'Aquila earthquake. Bull Earthq Eng 15(4):1425–1450

Di Pasquale G, Orsini G, Romeo RW (2005) New developments in seismic risk assessment in Italy. Bull Earthq Eng 3:101–128

Dolce M, Goretti A (2015) Building damage assessment after the 2009 Abruzzi earthquake. Bull Earthq Eng 13(8):2241–2264

Dolce M, Masi A, Marino M, Vona M (2003) Earthquake damage scenarios of the building stock of Potenza (Southern Italy) including site effects. Bull Earthq Eng 1:115–140

Dolce M, Kappos A, Masi A, Penelis G, Vona M (2006) Vulnerability assessment and earthquake damage scenarios of the building stock of Potenza (Southern Italy) using Italian and Greek methodologies. Eng Struct 28:357–371

Dolce M, Speranza E, Giordano F, Borzi B, Bocchi F, Conte C, Di Meo A, Faravelli M, Pascale V (2019) Observed damage database of past Italian earthquakes: the Da.D.O. WebGIS. Boll Geofis Teor Appl 60(2):141–164

Dolce M, Prota A, Borzi B, da Porto F, Lagomarsino S, Magenes G, Moroni C, Penna A, Polese M, Speranza E, Verderame GM, Zuccaro G (2021) Seismic risk assessment of residential buildings in Italy. Bull Earthq Eng 19:2999–3032

Goretti A, Di Pasquale G (2004) Building inspection and damage data for the 2002 Molise, Italy, earthquake. Earthq Spectra 20(S1):S167–S190

Graziani L, del Mese S, Tertulliani A, Arcoraci L, Maramai A, Rossi A (2019) Investigation on damage progression during the 2016–2017 seismic sequence in Central Italy using the European Macroseismic Scale (EMS-98). Bull Earthq Eng 17:5535–5558

Grünthal G (ed.), Musson RMW, Schwarz J, Stucchi M (1998) European Macroseismic Scale. Cahiers du Centre Européen de Géodynamique et de Séismologie, Vol. 15 - European Macroseismic Scale 1998. European Centre for Geodynamics and Seismology, Luxembourg

Ioannou I, Bertelli S, Verrucci E, Arcidiacono V, Rossetto T (2021) Empirical fragility assessment of residential buildings using data from the Emilia 2012 sequence of earthquakes. Bull Earthq Eng 19:1765–1795

Italian National Institute of Statistics, ISTAT (2001) http://dawinci.istat.it/jsp/MD/dawinciMD.jsp

Jayaram N, Baker JW (2010) Efficient sampling and data reduction techniques for probabilistic seismic lifeline risk assessment. Earthq Eng Struct Dyn 39:1109–1131

Karababa FS, Pomonis A (2011) Damage data analysis and vulnerability estimation following the August 14, 2003 Lefkada Island, Greece, Earthquake. Bull Earthq Eng 9:1015–1046

Kotha SR, Cotton F, Bindi D (2018) A new approach to site classification: Mixed-effects Ground Motion Prediction Equation with spectral clustering of site amplification functions. Soil Dyn Earthq Eng 110:318–329

Lagomarsino S, Giovinazzi S (2006) Macroseismic and mechanical models for the vulnerability and damage assessment of current buildings. Bull Earthq Eng 4:415–443

Lagomarsino S, Cattari S, Ottonelli D (2021) The heuristic vulnerability model: fragility curves for masonry buildings. Bull Earthq Eng 19:3129–3163

Lallemant D, Kiremidjian A, Burton H (2015) Statistical procedures for developing earthquake damage fragility curves. Earthq Eng Struct Dyn 44:1373–1389

Lloyd SP (1982) Least squares quantization in PCM. IEEE Trans Inf Theory 28:129–137

Maio R, Vicente R, Formisano A, Varum H (2015) Seismic vulnerability of building aggregates through hybrid and indirect assessment techniques. Bull Earthq Eng 13:2995–3014

Mascandola C, Barani S, Massa M, Paolucci E, Albarello D (2020) Clustering analysis of probabilistic seismic hazard for the selection of ground motion time histories in vast areas. Bull Earthq Eng 18:2985–3004

Masi A, Lagomarsino S, Dolce M, Manfredi V, Ottonelli D (2021) Towards the updated Italian seismic risk assessment: exposure and vulnerability modelling. Bull Earthq Eng 19:3253–3286

Michelini A, Faenza L, Lauciani V, Malagnini L (2008) ShakeMap implementation in Italy. Seismol Res Lett 79(5):688–697

Michelini A, Faenza L, Lanzano G, Lauciani V, Jozinovic D, Puglia R, Luzi L (2020) The New ShakeMap in Italy: progress and advances in the last 10 Yr. Seismol Res Lett 91(1):317–333

Penna A, Morandi P, Rota M, Manzini CF, da Porto F, Magenes G (2014) Performance of masonry buildings during the Emilia 2012 earthquake. Bull Earthq Eng 12:2255–2273

Polese M, d'Aragona MG, Prota A (2019) Simplified approach for building inventory and seismic damage assessment at the territorial scale: an application for a town in Southern Italy. Soil Dyn Earthq Eng 121:405–420

Polese M, Di Ludovico M, d'Aragona MG, Prota A, Manfredi G (2020) Regional vulnerability and risk assessment accounting for local building typologies. Int J Disaster Risk Reduct 43:101400

Rehman K, Burton PW, Weatherill GA (2014) K-means cluster analysis and seismicity partitioning for Pakistan. J Seismol 18:401–419

Rossetto T, Elnashai A (2003) Derivation of vulnerability functions for European-type RC structures based on observational data. Eng Struct 25(10):1241–1263

Rossi A, Tertulliani A, Azzaro R, Graziani L, Rovida A, Maramai A, Pessina V, Hailemikael S, Bufarini G, Bernardini F, Camassi R, Del Mese S, Ercolani E, Fodarella A, Locati M, Martini G, Paciello A, Paolini S, Arcoraci L, Castellano C, Verrubbi V, Stucchi M (2019) The 2016–2017 earthquake sequence in Central Italy: macroseismic survey and damage scenario through the EMS-98 intensity assessment. Bull Earthq Eng 17:2407–2431

Rosti A, Rota M, Penna A (2018) Damage classification and derivation of damage probability matrices from L'Aquila (2009) post-earthquake survey data. Bull Earthq Eng 16(9):3687–3720

Rosti A, Rota M, Penna A (2020a) Influence of seismic input characterisation on empirical damage probability matrices for the 2009 L'Aquila event. Soil Dyn Earthq Eng. https://doi.org/10.1016/j.soildyn.2019.105870

Rosti A, Del Gaudio C, Di Ludovico M, Magenes G, Penna A, Polese M, Prota A, Ricci P, Rota M, Verderame GM (2020b) Empirical vulnerability curves for Italian residential buildings. Boll Geofis Teor Appl 61(3):357–374

Rosti A, Rota M, Penna A (2021a) Empirical fragility curves for Italian URM buildings. Bull Earthq Eng 19:3057–3076

Rosti A, Del Gaudio C, Rota M, Ricci P, Di Ludovico M, Penna A, Verderame GM (2021b) Empirical fragility curves for Italian residential RC buildings. Bull Earthq Eng 19:3165–3183

Rota M, Penna A, Strobbia CL (2008) Processing Italian damage data to derive typological fragility curves. Soil Dyn Earthq Eng 28(10):933–947

Rota M, Penna A, Strobbia CL, Magenes G (2011) Typological seismic risk maps for Italy. Earthq Spectra 27(3):907–926

Saatcioglu M, Bruneau M (1993) Performance of structures during the 1992 Erzincan earthquake. Can J Civ Eng 20:305–325

Saretta Y, Sbrogiò L, Valluzzi MR (2021) Assigning the macroseismic vulnerability classes to strengthened ordinary masonry buildings: an update from extensive data of the 2016 Central Italy earthquake. Int J Disast Risk Re 62:102318

Sorrentino L, Cattari S, da Porto F, Magenes G, Penna A (2019) Seismic behaviour of ordinary masonry buildings during the 2016 Central Italy earthquakes. Bull Earthq Eng 17:5583–5607

Tertulliani A, Arcoraci L, Berardi M, Bernardini F, Camassi R, Castellano C, Del Mese S, Ercolani E, Graziani L, Leschiutta I, Rossi A, Vecchi M (2011) An application of EMS98 in a medium-sized city: the case of L'Aquila (Central Italy) after the April 6, 2009 Mw 6.3 earthquake. Bull Earthq Eng 9:67–80

Tocchi G, Polese M, Di Ludovico M, Prota A (2021) Regional based exposure models to account for local building typologies. Bull Earthq Eng. https://doi.org/10.1007/s10518-021-01242-6

Weatherill G, Burton PW (2009) Delineation of shallow seismic source zones using K-means cluster analysis, with application to the Aegean region. Geophys J Int 176:565–688

Worden CB, Thompson EM, Hearne M, Wald DJ (2020) ShakeMap Manual Online: technical manual, user's guide, and software guide, U. S. Geological Survey. http://usgs.github.io/shakemap/. https://doi.org/10.5066/F7D21VPQ

Xie Y, Sichani ME, Padgett JE, DesRoches R (2020) The promise of implementing machine learning in earthquake engineering: a state-of-the-art review. Earthq Spectra. https://doi.org/10.1177/8755293020919419

Zuccaro G, Dolce M, De Gregorio D, Speranza E, Moroni C (2015) La scheda CARTIS per la caratterizzazione tipologico-strutturale dei comparti urbani costituiti da edifici ordinari. Valutazione dell'esposizione in analisi di rischio sismico. Proceedings of GNGTS 2015, Trieste, Italy (in Italian)