# Looking for undocumented earthquake effects: a probabilistic analysis of Italian macroseismic data

Andrea Antonucci[1], Andrea Rovida[1], Vera D'Amico[2] and Dario Albarello[3,4]

[1]Istituto Nazionale di Geofisica e Vulcanologia, Milano, 20133, Italy
[2]Istituto Nazionale di Geofisica e Vulcanologia, Pisa, 56125, Italy
[3]Department of Physics, Earth and Environmental Sciences, University of Siena, Siena, 53100, Italy
[4]Consiglio Nazionale delle Ricerche, Istituto di Geologia Ambientale e Geoingegneria, Roma, 00185, Italy

*Correspondence to*: Andrea Antonucci (andrea.antonucci@ingv.it)

**Abstract**. A methodology to detect local incompleteness of macroseismic intensity data at the local scale is presented. In particular, the probability that undocumented effects actually occurred at a site is determined by considering intensity prediction equations (in their probabilistic form) integrated by observations relative to known events documented at surrounding sites. The outcomes of this analysis can be used to investigate how representative and known are the seismic histories of localities (i.e., the list of documented effects through time). The proposed approach is applied to the Italian area. The analysis shows that, at most of the considered sites, effects of intensity ≥ 6 should most probably have occurred at least once, but they are not contained in the current version of the Italian macroseismic databases. In a few cases, instead, the lack of data may concern higher intensity levels (i.e., ≥ 8). The geographical distribution of potentially lost information reflects the heterogeneity of the seismic activity over the Italian territory.

## 1 Introduction

Extending the knowledge of the seismicity of an area as back as possible in time is essential in seismological research, including seismotectonic investigations and seismic hazard assessment. For these purposes, earthquake catalogues spanning hundreds of years represent an essential complement of instrumental data relative to the last decades. The compilation of these catalogues relies on the analysis of the effects documented on the human and natural environment during past earthquakes, interpreted and standardized in terms of macroseismic intensity scales (e.g., MCS – Sieberg, 1923; MMI – Wood and Neumann, 1931; MSK – Medvedev et al., 1964; EMS-98 - Grünthal, 1998). Intensity data available at different localities for a given earthquake (Intensity Data Points, hereafter IDPs) can be used to constrain respective epicentral location and magnitude with a variety of methodological approaches (e.g., in recent years Bakun and Wentworth, 1997; Gasperini et al., 1999; 2010; Provost and Scotti, 2020).

Compared to other regions of the world, the knowledge of European seismicity is particularly detailed and lengthy (Albini et al., 2013; Locati et al., 2014; Rovida et al; 2020a; 2022a), and Italy stands out from other European countries. The bulk of the current Italian Parametric Earthquake Catalogue – CPTI15 version 4.0 (Rovida et al., 2020b; 2022b) mostly derives from the harmonization and parametrization of intensity data contained in the Italian Macroseismic Database DBMI15 version 4.0 (Locati et. al., 2022). In fact, the majority of the earthquakes contained in CPTI15, which spans from 1000 to 2020 C.E., is supported by IDPs, in particular those in the pre-instrumental period, as a result of more than 45 years of research in the field of historical seismology, represented in the Italian Archive of Historical Earthquake Data ASMI (Rovida et al., 2017). Despite the increase in the number of macroseismic studies in time, the historical research remains anchored to the long tradition of national and regional seismological compilations, based mainly on local historiography, summarized in the pioneering work of Baratta (1901) and has later been influenced by projects commissioned for several scopes as, for example, the identification

of the sites for nuclear power plants in the 1980s (Stucchi, 1993; Camassi, 2004). As a consequence, several investigations often focused on specific, sometimes limited, geographical areas. This influences the content of the CPTI15 earthquake catalogue, the completeness of which needs to be analyzed from both the time and space points of view to evaluate how representative of the actual seismicity is (Albarello et al., 2001; Stucchi et al., 2004; Stucchi et al., 2011). In addition, earthquakes in a given completeness time span and area may show gaps in terms of documented effects at the sites. The assessment of earthquake parameters from intensity data is strictly connected to their reliability, number and spatial distribution. In Italy, as in the rest of Europe, there are many earthquakes attested by very few, or even single IDPs, which, of course, do not constrain the earthquake location and magnitude (Albini and Rovida, 2018; Albini, 2020; Rovida et al., 2020a). Moreover, the size of the earthquake and the number of IDPs are not related and many IDPs might support low magnitude earthquakes and vice-versa, with high magnitude earthquakes that might be represented by one or few IDPs that often correspond to the highest available intensities. This means that several places may not have documented the effects related to a given event, regardless of its size. Analyzing the undocumented earthquake effects and providing an estimate in terms of macroseismic intensity represent the basis for investigating the knowledge of the seismic history of a given site. Despite this, no such in-depth analysis is yet available at both European and regional scale.

The aim of this study is proposing a coherent probabilistic approach to detect sites where seismic effects of past earthquakes could be missing, for several reasons. Moreover, it also provides a deeper insight about the completeness level of data relative to historical earthquakes, and this may be useful to identify possible biases when incomplete macroseismic data are used for several seismological analyses. An application of this approach is here presented focusing on the Italian territory. In this area, a huge amount of macroseismic intensity data is available which have been extensively used for seismic hazard assessment and other seismological investigations (e.g., D'Amico and Albarello, 2008; Faenza and Michelini, 2010; Gomez Capera et al., 2020; Oliveti et al., 2022).

## 2 The Italian macroseismic data

The historical research in the field of macroseismology of the last decades led to a wealth of studies that present data on Italian earthquakes and surrounding areas in a variety of different formats. These studies are inventoried and gathered in the Italian Archive of Historical Earthquake Data – ASMI (https://emidius.mi.ingv.it/ASMI/index_en.htm, Rovida et al., 2017) which grants access to the information related to more than 6500 earthquakes occurred in the Italian area from 461 B.C. to 2020 C.E. The multiplicity of data contained in ASMI is used for the compilation of the Italian Macroseismic Database DBMI through an accurate selection of the dataset supporting each earthquake according to the content, update and thoroughness of the available studies.

The latest version of DBMI (DBMI15 v.4.0, Locati et. al., 2022) makes available 123981 IDPs, mostly expressed in the MCS macroseismic scale, related to 3229 earthquakes in the time-window 1000-2020 and referred to 15343 Italian localities. These data are the result of 191 different studies and most of IDPs (i.e., 60%) come from the recent (1980-2005) earthquakes provided by the ING and INGV Macroseismic Bulletin (e.g., Gasparini et al., 2003; Gasparini et al., 2011) and from the Catalogue of Strong Italian Earthquakes CFTI4Med (Guidoboni et al., 2007). The remaining part consists of intensity data from different studies, dealing with a great number of earthquakes (e.g., Molin et al., 2008; Camassi et al., 2011; Azzaro and Castelli, 2015), scientific papers on single earthquakes, areas or periods, and macroseismic field surveys of recent earthquakes (e.g., Tertulliani et al., 2012). The number of available data per earthquake and per locality is extremely variable. In particular, 5650 out of the 15343 sites contained in DBMI15 (36.8%) have only one intensity value, 2114 have two intensity data (13.8%) and 3207 have more than 10 intensity values (20.9%). Data contained in DBMI15 are used for compiling the seismic history of Italian

localities, that is the list of earthquake effects observed in a place through time, and for assessing macroseismic parameters (epicentral location and magnitude) of the events listed in the CPTI15 (https://emidius.mi.ingv.it/CPTI15-DBMI15/query_eq/, Rovida et al., 2020b; 2022b). This catalogue covers the entire Italian territory and neighboring areas and seas, and counts 4894 earthquakes in the period 1000-2020, with maximum intensity greater than or equal to 5 MCS and moment magnitude (Mw) greater than or equal to 4.

Despite of the enormous and unique amount of data, their diverse provenance, the possible gaps in the historical documentation and its non-systematic investigation affect their homogeneity in time and space. Assessing the completeness of intensity data requires a deep knowledge of the local history of the investigated locality, the preservation of the related documentation and its thorough analysis. In Italy, such complex and time-consuming investigations were performed for eighteen localities in the framework of dedicated projects (Albini et al., 2001; Albini and Stucchi, 2002; Albini et al. 2003). Stucchi et al. (2004) later extrapolated these results at the national scale for the assessment of the historical completeness of the Italian earthquake catalogue. Although the assessment of the historical completeness for all the localities in DBMI15 is hardly feasible, an overall picture is achievable through the identification and analysis of earthquake effects potentially occurred but not documented at a number of sites, as proposed in this work.

To this purpose, we applied the methodology developed by Antonucci et al. (2021) to a significant sample of Italian localities, selected as described in Section 4, and considered all the earthquakes in CPTI15 except volcanic events, because the attenuation of intensity in volcanic areas is different from that of the Italian territory (e.g., Carletti and Gasperini, 2003; Azzaro et al., 2006), and earthquakes with instrumental depth greater than 40 km, because they are generally slightly felt at the surface and thus are likely absent from the historical records.

## 3 Estimating intensity values for undocumented effects

Intensity data can be calculated where the effects of a given earthquake with known location and magnitude are missing. A common methodology relies on the use of Intensity Prediction Equations (IPEs) for computing an intensity value at a considered locality as a function of the source-to-site distance and the magnitude or epicentral intensity of an earthquake. The most recent IPE for the Italian area was published by Pasolini et al. (2008) and is based on a classical functional form, similar to that used for Ground Motion Prediction Equations (GMPEs), with the physical terms of geometric spreading and anelastic attenuation (Mak et al., 2015). This IPE, recently recalibrated by Lolli et al. (2019) using the data collected in the DBMI15 v.1.5 (Locati et al., 2016) and CPTI15 v.1.5 (Rovida et al., 2016; 2020b), was used in this study.

As extensively discussed in Albarello and D'Amico (2004) and Antonucci et al. (2021), a way to express the uncertainties related to each intensity estimation is using a probabilistic approach. In particular, the intensity value calculated at the site by the IPE is estimated through a normal probability distribution using the average $\mu$ and the standard deviation $\sigma$ determined by Pasolini et al. (2008) and Lolli et al. (2019) as a function of epicentral parameters contained in CPTI15.

Furthermore, when the intensity related to a given earthquake is not documented at the considered site but is available at close sites, the value estimated with the IPE can be constrained with such intensity values (Albarello et al., 2007) observed in at least one neighboring (within 20 km) locality. The distance of 20 km was selected through an analysis on more than 15000 Italian sites contained in DBMI15. We investigated the geographic distribution of these localities calculating the number of localities within a set of possible distance thresholds for every site (Antonucci et al., 2021). In particular, with a Bayesian approach, it is possible to estimate the discrete probability density distribution $pl(Is/Iv)$ at a given site, associating to each possible intensity degree $Is$ at the site $s$ a probability value conditioned by the occurrence of effects of intensity $Iv$ at nearby sites $v$:

$$p_l(I_s|I_v) = p_l(I_s) \frac{q(I_v|I_s)}{\sum_{I=1}^{12} p_l(I)q(I_v|I)} \qquad (1)$$

where $pl(Is)$ is the "prior" normal probability distribution estimated through the IPE and the conditional probability $q(Iv|Is)$ represents the correlation between intensity values at neighboring localities estimated empirically from a dataset of earthquakes and their observed IDPs. The latter probability is fixed and can be estimated from the relative frequencies of the differences between any pair of intensity values observed at the nearby sites as reported in DBMI15 (for details, see Antonucci et al., 2021). In function of the number of neighboring sites within 20 km of distance, Eq. (1) can be iteratively applied substituting the "prior" distribution $pl(Is)$ with the output of the preceding estimate. If the intensity documented at the close sites is uncertain (e.g., 6-7), an equal probability is assigned to each of the two contiguous degrees as explained in Antonucci et al. (2021) In other words, this approach (i) estimates an intensity value at the considered site from the epicentral location and magnitude of a given earthquake through an IPE expressed in a probabilistic form and (ii) uses the intensity values documented for the same event at close localities for constraining the value obtained through the IPE. Differently from the existing IPEs, this procedure is thought to better model the non-isotropic decay of intensity with distance, considering the values documented at nearby localities. In this way, the seismic history of any place can be integrated with an estimate of the number and severity of the earthquake effects that, although not documented, likely occurred at the site either because they are reported at nearby localities, or because earthquakes of given magnitudes took place within a certain distance from the place.

## 4 Selection of the sample sites

To analyze the number and the entity of undocumented macroseismic effects that might have occurred in Italy in the past, a dataset of sample localities was defined. The dataset was selected according to the geographical distribution of the localities and to the number of associated macroseismic observations in DBMI15, exclusively based on expert judgement without the use of automatic procedures. These sites had to present a homogeneous and dense distribution over the Italian territory also finding a good compromise between main cities and small villages. Moreover, the selection considered both the differences in the urbanization in Italy and the distance of 20 km among localities, that is adopted in the Bayesian procedure for estimating the intensities (Antonucci et al., 2021).

140

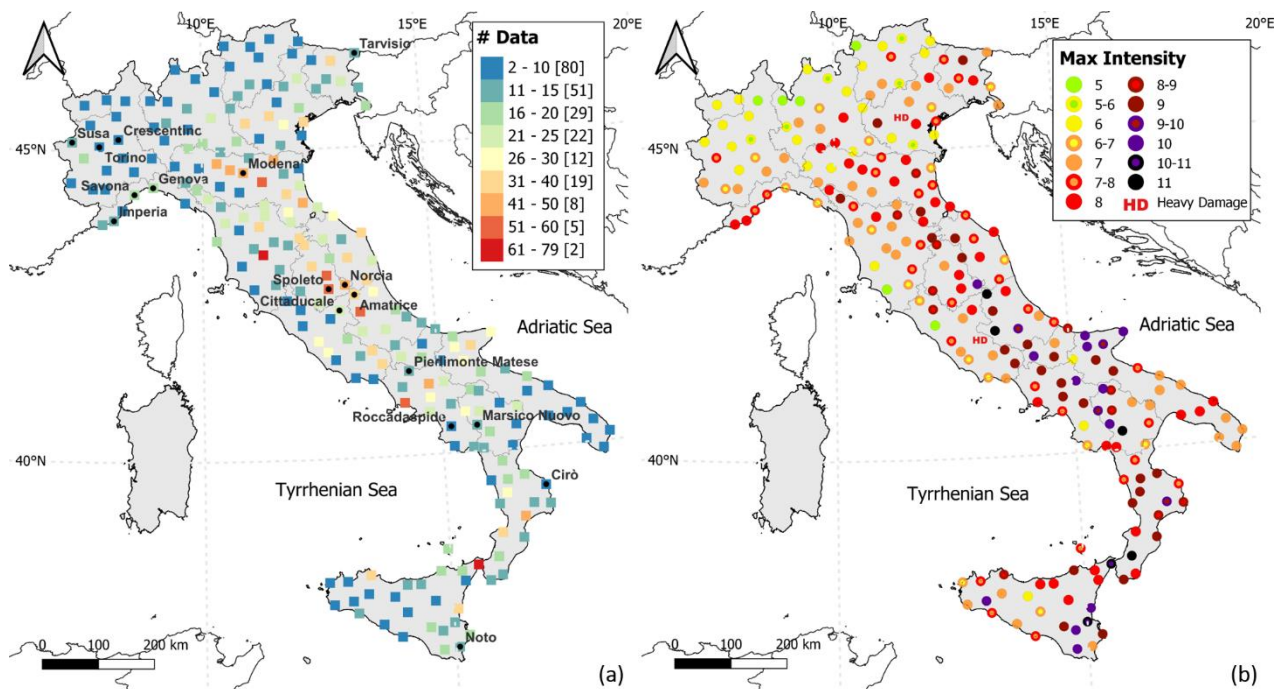



**Figure 1. Geographical distribution of the 228 selected sites. (a) Number of intensity data greater than or equal to 5 MCS per locality; numbers in squared brackets indicate the total of the localities in each class of data, (b) maximum intensity observed at each locality. The sites cited in the text and borders of administrative regional areas are reported in the maps.**

The selected dataset includes 228 sites (Table S1) distributed over the whole Italian territory, with the exception of localities

145    in the very low seismicity area of Sardinia island that present too few data, and represents a choice among the localities with: (i) the highest number of intensity data collected in DBMI15, and (ii) their geographical distribution also taking into account distances from one another. In particular, the seismic histories of the 228 localities have at least 2 intensity values greater than or equal to 5 MCS (Fig. 1) with a total of 10323 macroseismic data ranging from intensity 2 to 10-11 MCS. In addition, the 228 sites have 2201 data expressed with non-conventional descriptive codes (e.g., "HD" for heavy damage, see Locati et al.,

150    2022).

Focusing on the data with observed intensity ≥ 5 MCS (Fig. 1a), the number of macroseismic observations exceeds 50 at 7 localities only, and 80 sites have less than 10 intensity data, most of them located in Northern Italy. In addition, at almost the totality of selected sites (216 out of 228), effects of intensity ≥ 6 MCS have been documented. Figure 1b shows that some sites located in the areas with low seismicity (i.e., part of Northern Italy) have observed a maximum intensity equal to 5 and 6 MCS.

155    On the contrary, many localities placed in part of Central and Southern Italy, have suffered a maximum intensity greater than or equal to 10 MCS due to high seismicity areas.

## 5 Results

The entity of a given effect was computed at each selected site for each earthquake in CPTI15 when the respective IDP is lacking in DBMI15 (see Section 3). As a result, the intensities corresponding to undocumented effects are estimated in two

160    ways: (i) from earthquake parameters through the adopted IPE, i.e., effects not documented at the site nor in the sites nearby, but likely to have happened on the basis of the content of the earthquake catalogue; (ii) by integrating the above information with observations available at other localities within 20 km from the considered site (see Eq. 1). It is assumed that in the last case, probability that the considered intensity threshold has been actually reached is better constrained than in the former case. Intensity data inferred from the IPE either "corrected" with macroseismic observations available at nearby localities or not,

165 can be considered as "potentially lost" data because, although the locality likely experienced a given level of shaking in consequence of a known earthquake, this is not documented.

### 5.1 Site-by-site analysis

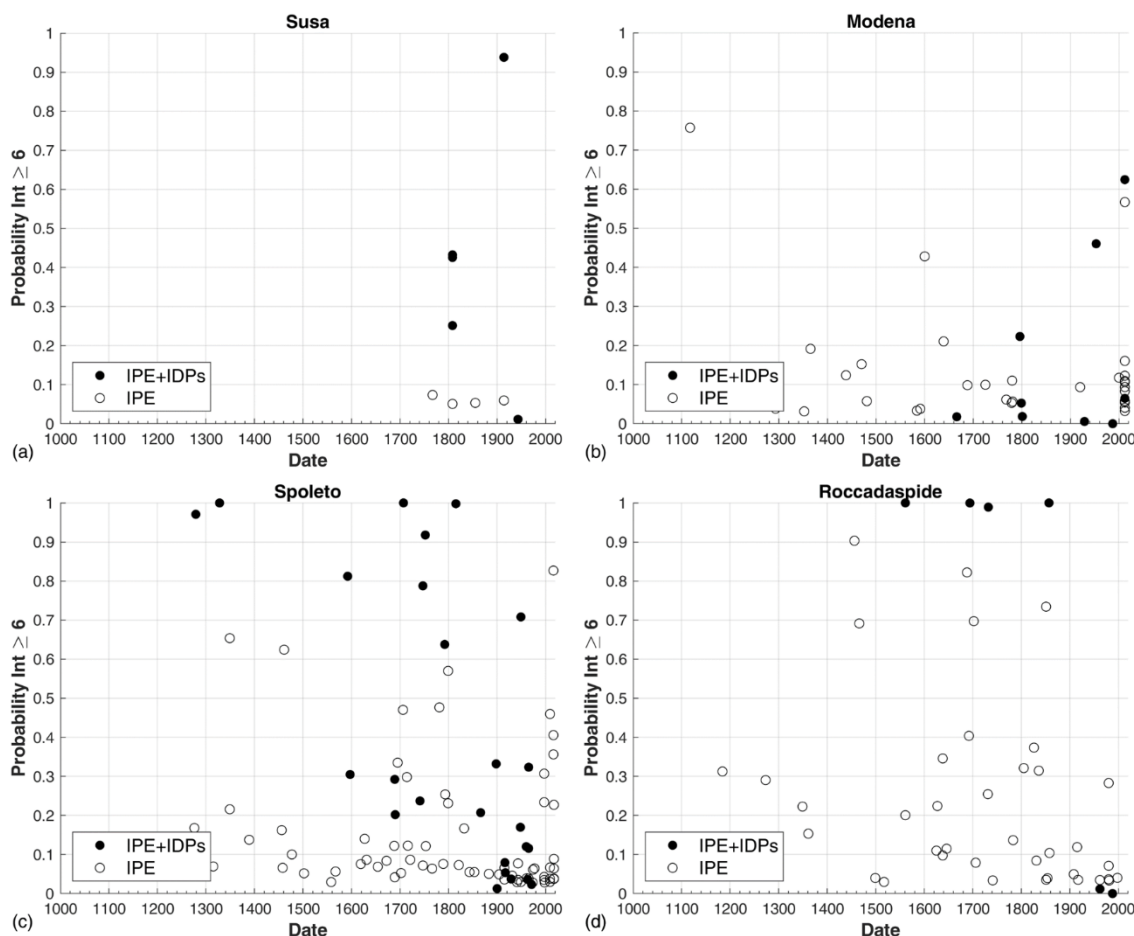
The earthquake effects at the selected sites can be analysed on a site-by-site basis in order to evaluate: (i) the number of undocumented effects at the considered sites, and (ii) the probability that each of these effects might have reached a given 170 intensity level. In other words, we estimated the probability of having an undocumented intensity value at each of the considered sites. Figure 2 shows, as an example, the results obtained at 4 sites in terms of probability of reaching or exceeding intensity 6 MCS, estimated through the Bayesian approach described above. These sites (see Fig. 1 for location) were selected to represent geographical areas characterized by different levels of seismicity: (i) Susa in the Western Alps; (ii) Modena in the Po Plain; (iii) Spoleto in the Central Apennines, and (iv) Roccadaspide in Southern Italy.

175

**Figure 2. Probability of undocumented effects with intensity ≥ 6 MCS estimated both with the only use of the IPE (white dots) and with the Bayesian approach considering also nearby IDPs (black dots) at: (a) Susa, (b) Modena, (c) Spoleto, and (d) Roccadaspide.**

As shown in Figure 2, the undocumented effects at the four sites are quite different in terms of both the total number and the exceedance probabilities. As regards the total number of effects, the highest numbers were estimated at Spoleto (Central 180 Apennines) and Roccadaspide (Southern Italy) with 93 and 45 effects with intensity ≥ 6 possibly lost, respectively. For the two localities in Northern Italy, the undocumented effects are 39 in Modena (Po Plain) and only 9 in Susa (Western Alps); for the latter, all the events occurred after 1760.

The number of estimated undocumented effects represents an overview of the analysis that does not consider the differences in terms of exceedance probability computed for each event. In fact, the number of effects estimated at each site changes considerably according to their probabilities. For example, Figure 2a shows that at Susa only one effect with intensity $\geq 6$ MCS can be considered potentially lost with a probability equal to 94%. This effect was estimated for the earthquake occurred on 26 October 1914 with Mw 5.2, located 26 km from the site. The value resulting from the IPE was constrained with 4 intensity data equal to 6 MCS and one equal to 6-7 MCS documented at close localities (https://emidius.mi.ingv.it/CPTI15-DBMI15/eq/19141026_0345_000, last access: 3 August 2022). Figure 2b shows that two events have higher probability to have produced effects greater than or equal to 6 at Modena. The first one has a probability of 76% and derives from one of the strongest earthquakes occurred in Northern Italy, i.e. the 3 January 1117 earthquake of Mw equal to 6.5 and located about 70 km from Modena (https://emidius.mi.ingv.it/CPTI15-DBMI15/eq/11170103_1515_000, last access: 3 August 2022). This probability was estimated through the only use of the IPE because the macroseismic intensity distribution in DBMI15 (from Guidoboni et al., 2007) shows very scattered data and none of these was documented at sites within 20 km from Modena. The second effect has a probability of 62% and is related to the main shock of the 2012 Emilia sequence (29 May 2012; Mw 5.9), which struck part of Northern regions (https://emidius.mi.ingv.it/CPTI15-DBMI15/eq/20120529_0700_000, last access: 3 August 2022). Several undocumented effects were estimated for the shocks of the 2012 sequence, because no intensities were assigned at Modena for any shock during the macroseismic survey (Tertulliani et al., 2012). Figure 2c shows that 8 effects with intensity $\geq 6$ MCS can be considered potentially lost at Spoleto with a probability greater than 75%, and are all estimated from earthquakes with epicenters located between 10 and 52 km. For four of these effects, the estimated probability is greater than 95%. All these probabilities are calculated constraining the value obtained by the IPE with intensity values documented at close localities, with the exception of the earthquake occurred on 26 October 2016 for which no IDPs are available within 20 km from Spoleto. As regards Roccadaspide (Fig. 2d), four effects have probabilities higher than 95% of reaching or exceeding intensity 6, all constrained with the Bayesian approach using the intensity data documented at close localities.

The exceedance probabilities for higher intensity levels, i.e. greater than or equal to 7, 8 and 9 MCS, were also calculated and analyzed. Figure 3 reports the results obtained for intensity greater than or equal to 8 MCS at the same four sites of Figure 2.

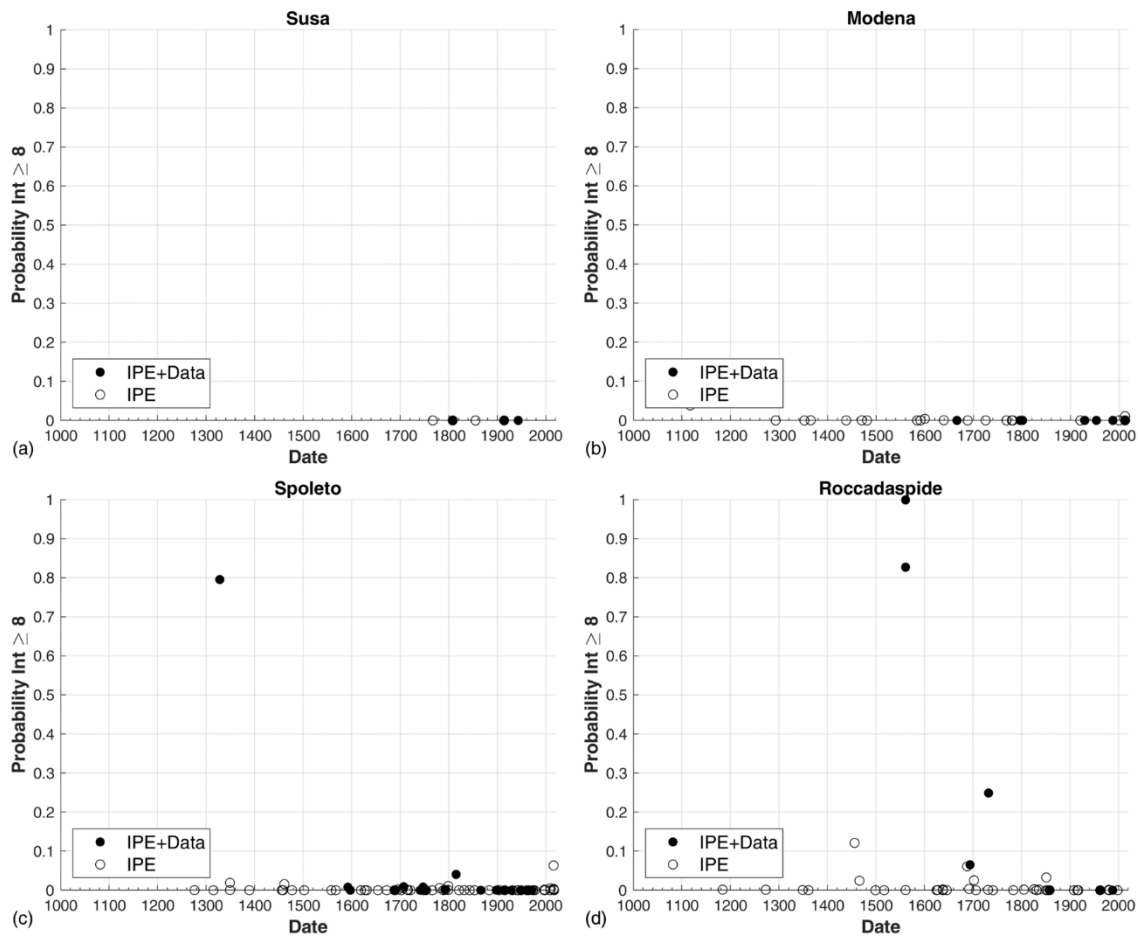



**Figure 3. Probability of undocumented effects with intensity ≥ 8 MCS estimated both with the only use of the IPE (white dots) and with the Bayesian approach considering also nearby IDPs (black dots) at: (a) Susa, (b) Modena, (c) Spoleto, and (d) Roccadaspide.**

In particular, Figure 3a and 3b show that the earthquakes that might have produced an intensity at least equal to 6 MCS (Fig. 2a and 2b) were not able to produce an intensity ≥ 8 MCS at Susa and Modena. On the other hand, as regards Spoleto (Fig. 3c), the earthquake occurred on 1 December 1328 with Mw 6.5, at about 27 km from the site, may have produced an intensity ≥ 8 MCS with 80% probability. In this case, the estimate provided by the IPE was constrained through the intensity 9 MCS documented (Monachesi, 1987) at a locality very close to Spoleto (https://emidius.mi.ingv.it/CPTI15-DBMI15/eq/13281201_0000_000, last access: 3 August 2022). Figure 3d shows that two effects with intensity ≥ 8 MCS can be considered potentially lost with high probability (> 80%) at Roccadaspide. These effects are related to two events occurred very close in time, on 31 July and 19 August 1561, that struck Southern Italy with Mw equal to 6.7 and 6.3, respectively. Both these events were located about 30 km from the considered site and macroseismic intensity data are available at nearby localities that allowed to constrain the intensity values obtained with the IPE.

**5.2 Geographical distribution of potentially lost effects**

As shown in the previous section, each undocumented effect estimated at a site has different exceedance probabilities for different intensity levels. In this respect, the number of effects potentially lost at the 228 selected sites can be quantified selecting a given probability threshold. For this purpose, at each site, we counted the number of undocumented effects with probabilities ≥ 75% (i.e., the third quartile of the entire probability distribution) of reaching or exceeding intensity levels 6, 7, 8 and 9 MCS (Fig. 4).

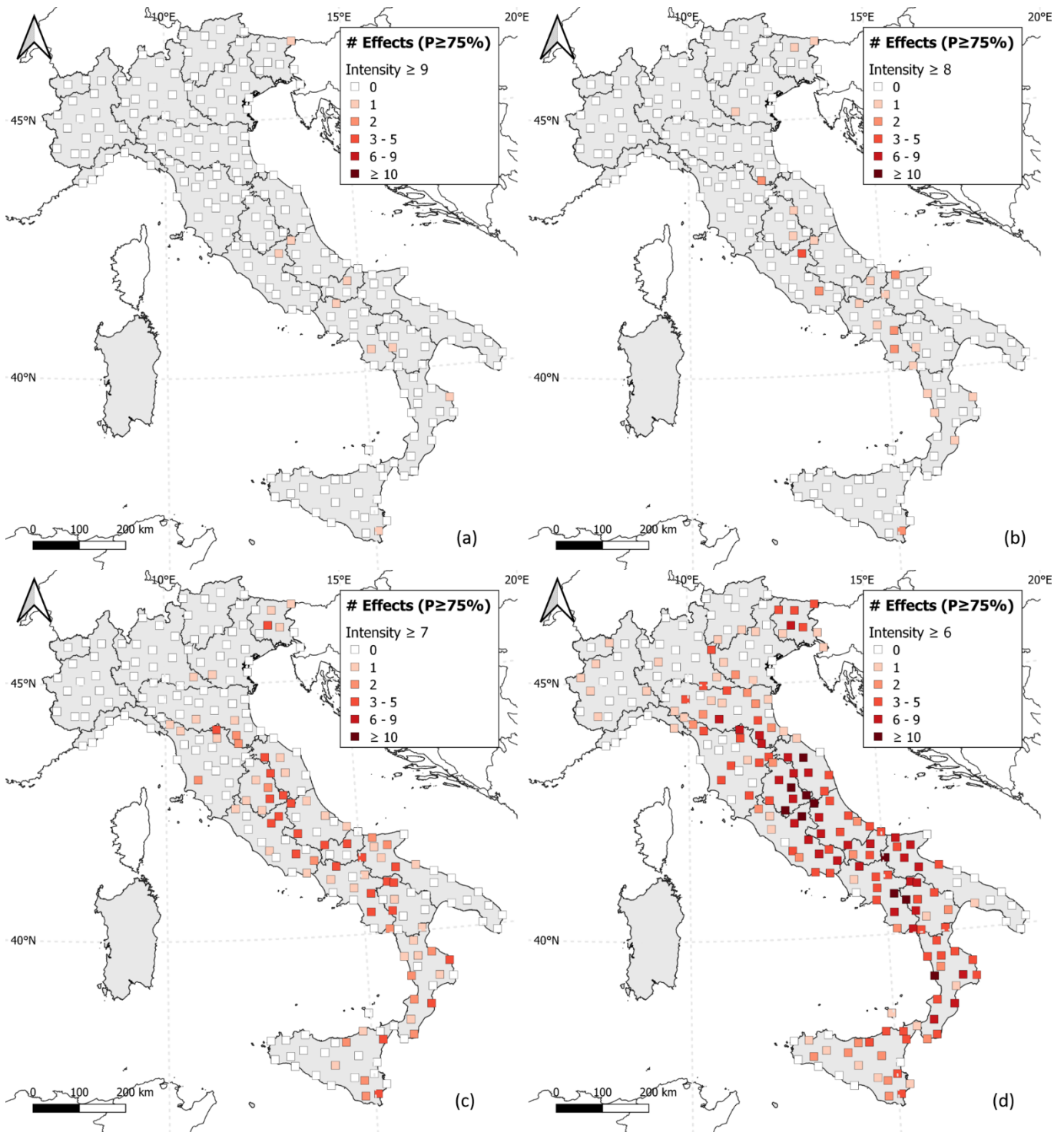



**Figure 4. Number of undocumented effects at each selected site with probabilities ≥ 75% of reaching or exceeding intensity (a) 9, (b) 8, (c) 7 and (d) 6 MCS.**

Figure 4a shows that one effect was potentially lost at nine localities assuming a probability threshold equal to 75% for intensity ≥ 9 MCS. For one of these sites, i.e., Tarvisio in Northeastern Italy (see Fig. 1), the estimated effect derived from the earthquake occurred on 25 January 1348 with Mw equal to 6.6, whose epicenter was very close to the site (less than 1 km, according to CPTI15). The undocumented effect estimated at Noto in Southeastern Sicily (see Fig. 1) is related to the Mw 7.3 earthquake occurred on 11 January 1693 (https://emidius.mi.ingv.it/CPTI15-DBMI15/eq/16930111_1330_000, last access: 3 August 2022). However, this effect should not be considered in this analysis because Noto was re-built and relocated after that event, which struck Sicilian island with total destruction of many sites included the site today known as Noto Antica located about 7

km from the present-day town. As regards intensities ≥ 8 MCS, Figure 4b shows that effects with exceedance probability ≥ 75% are calculated at 23 localities, for a total of 31 potentially lost effects. In detail, Cittaducale (see Fig. 1), in Central Italy, shows 3 potentially lost effects: one of these is related to the poorly constrained (17 IDPs; Guidoboni et al., 2007) event occurred on 9 September 1349 in the Central Apennines with Mw equal to 6.3 and the epicenter about 19 km from the site (https://emidius.mi.ingv.it/CPTI15-DBMI15/eq/13490909_0000_000, last access: 3 August 2022). Figure 4c shows that for the sites located in a large part of Northern Italy, in the Tyrrhenian regions and in the Southeastern region, no effects with intensity ≥7 MCS were computed; on the contrary, 150 potentially lost effects were estimated at 76 sites, principally located along the Central Apennines and in Southern Italy. The results obtained for intensity ≥ 6 MCS are completely different. In fact, as shown in Figure 4d, the number of potentially lost effects is 617 estimated at 153 out of 228 sites, located almost everywhere except for some areas in the Central Alps and in the Northwestern regions. The undocumented effects are more than ten for ten localities and, for two of these, i.e., Norcia and Amatrice (Central Italy; Fig. 1) are equal to 19 and 17, respectively. In general, the maps in Figure 4 show that for some localities placed in parts of Northwestern Italy and Central Alps no undocumented effects were estimated. In contrast, at least one effect that can be considered potentially lost for intensity ≥ 6 MCS was computed at most of 228 considered sites, except for 75 sites mostly located in the same low seismicity areas. This is consistent with the features of the seismicity of the Italian area, which shows low seismicity in the western and central Alps and strong and more frequent events in the central and southern Apennines and Sicily (see Rovida et al.; 2020b; 2022b).

Taking into account the different exceedance probabilities computed at each locality for $N$ undocumented earthquakes, it was possible to estimate the probability that at least one effect was not documented at the 228 selected sites for different intensity levels. This represents another way to analyze the outcomes obtained site-by-site (Sect. 5.1). Given the exceedance probabilities $P_l(I_s)$ relative to the intensity threshold $I_s$ for the $l$-$th$ of $N$ earthquakes, the probability $L(I_s)$ that at least one effect with intensity greater than or equal to $I_s$ has been lost, is given by:

$$L(Is) = 1 - \prod_{l=1}^{N}[1 - P_l(Is)] \tag{2}$$

The results of the analysis for different intensity levels $I_s$ are reported in Figure 5. In particular, the map shown in Figure 5a represents the probability that at least one effect of intensity ≥ 9 MCS was not documented at the selected sites: a probability greater than 5% was estimated at 41 sites, exceeding 95% only at five localities, principally located in Central and Southern Italy (Amatrice, Cirò, Marsico Nuovo, Piedimonte Matese and Noto; Fig. 1). Figure 5b shows that the probability of having at least one effect of intensity ≥ 8 MCS that can be considered potentially lost is greater than 95% at 16 sites in the Northeast, in Central and Southern Italy; on the contrary, low probabilities (< 5%) result at few localities placed in Southern and Central Italy and at most of the sites in the North. The results obtained for lower intensity levels (i.e., intensity 6 and 7) appear quite different. In fact, the map in Figure 5c shows that a probability greater than 50% of having at least one undocumented effect with intensity ≥ 7 MCS was estimated at 150 localities; out of these, 91 sites mostly located in Central and Southern Italy have probabilities greater than 95%; low probabilities (< 5%) result at 18 sites, principally in Northern Italy. As regards intensity threshold 6, Figure 5d shows that the probability of at least one potentially lost effect exceeds 50% at almost the totality of the considered sites (211 out of 228), whereas lower probabilities (< 25%) result at only 5 sites located in the Northwestern regions (Savona, Genova, Imperia, Crescentino, Torino; see Fig. 1 for their location).

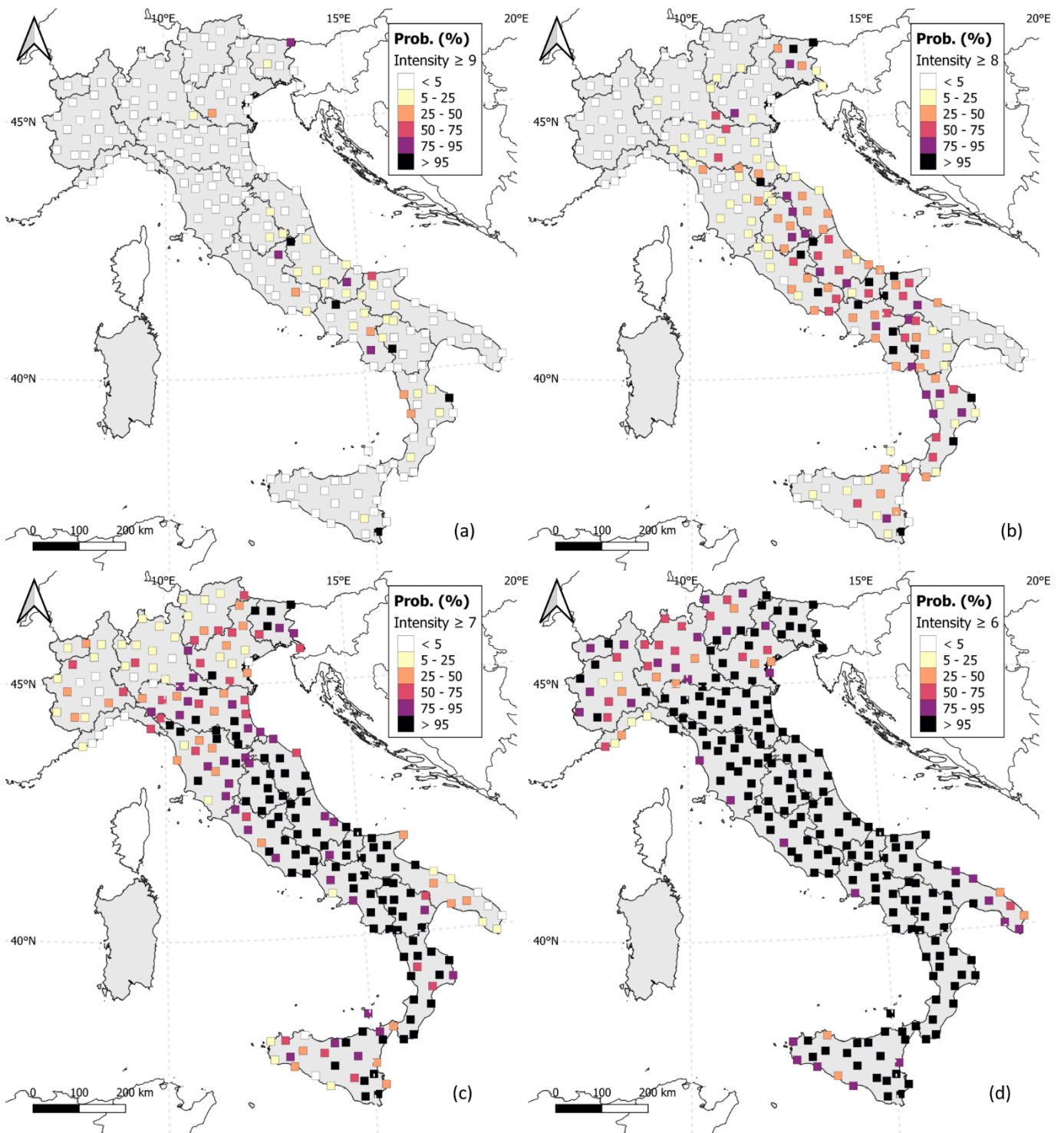



**Figure 5. Probability that at least one effect of intensity ≥ (a) 9, (b) 8, (c) 7, and (d) 6 was not documented at the 228 selected sites.**

## 6 Discussion and Conclusions

This work provides a probabilistic methodology devoted to quantitative estimate of the effects of past Italian earthquakes that can be considered potentially lost at a sample set of sites and analyzes the results both at the local (site-by-site) and national scale. The results show some gaps in the macroseismic data contained in DBMI15 despite their quality and quantity. Indeed, at least one damage effect with intensity ≥ 6 MCS could be potentially lost with a probability greater than 95% at many of the

selected sites (i.e., 173 out of 228). Considering the overall number of potentially lost effects (Fig. 4), they strongly decrease with increasing intensity, from 617 of intensity $\geq$ 6 MCS, to 31 of intensity $\geq$ 8 MCS, and just 9 of intensity $\geq$ 9 MCS. The reason is that heavy damage or destruction suffered at a locality, represented by intensities 8 and 9 MCS, are more likely to have been recorded by historical sources. On the contrary, slight damage, corresponding to intensity 6 MCS, may have left less significant traces in the historical record of a locality. Unreported macroseismic data of any intensity might be related to earthquakes of any size and period, including the most recent and strong ones. From a geographical point of view, few undocumented effects were computed at the sites located in a large part of Northwestern Italy, in the Central Alps and in the Southern Adriatic region and Sicily compared to those estimated in Central and Southern Italy (principally along the Apennines), independently of the considered intensity level. These discrepancies are likely representative of the differences in the seismicity of the Italian territory, with the number and strength of the earthquakes located in Central and Southern Italy clearly greater than in other areas. However, these results point out that the seismic history of one site might be different from the others, also within short distances. This probably depends on the relative importance that a locality had through time because the story of main towns is more documented with respect to minor ones.

The joint analysis of macroseismic intensity data observed through time in a place and those calculated through an IPE, constrained with data observed at nearby localities with the Bayesian procedure described in Antonucci et al. (2021), provides a general methodology to investigate the knowledge of seismic histories and to estimate the level of representativeness of each site in function of the seismicity of a given area. The results are given in a probabilistic form that allows considering both the uncertainties related to the assessment of intensity at a given site and the nature of macroseismic data (ordinal, discrete, and range limited). Such a procedure is repeatable and appliable to other regions and contexts. However, the outcomes are strictly dependent on the number of earthquakes and the reliability of the parameters contained in the input seismic catalogue and on the adopted IPE, with its specific functional form and parameters, and the associated uncertainties. In this regard, particular care should be paid to the interpretation of the results, considering that the content of a catalogue is progressively less representative of the actual seismicity going back in time, especially for small events. This implies that changing both the considered catalogue and IPE, the results in terms of calculated undocumented effects at the sites could considerably change.

Regardless of these limitations, the analyses show that the intensity data documented at a given site may not be representative of the actual shaking experienced through time even with enormous amounts of macroseismic data, such as in Italy. Consequently, the use of these data for several seismic analyses, such as intensity-based seismic hazard assessment at a local scale and testing of probabilistic seismic hazard models, should include a preliminary careful analysis of the representativeness and completeness of macroseismic data at the sites, regardless the study area considered. For this purpose, the main future goal will be checking the consistency of these results with those obtained through an in-depth historical investigation, which is the only way of providing robust quantitative estimates of the temporal completeness of the seismic history of a site for different intensity levels.

**Data availability**

DBMI15 is available at https://doi.org/10.13127/DBMI/DBMI15.4

CPTI15 is available at https://doi.org/10.13127/CPTI/CPTI15.4

## Author contribution

AA edited most parts of the paper and performed the statistical analyses. AR, VD and DA contributed to the manuscript and supervised the research.

## Competing interests

320 The authors declare that there is no conflict of interests.

## Acknowledgements

The authors wish to thank Carlos Hector Caracciolo, Josep Batlló, an anonymous reviewer, and the editor Animesh Gain for their comments and suggestions. The authors wish to warmly thank Mario Locati for his constant support, and Paola Albini for her comments and suggestions on this work and, more in general, for her precious and continuous teachings on historical
325 seismology.

## References

Albarello, D. and D'Amico, V.: Attenuation relationship of macroseismic intensity in Italy for probabilistic seismic hazard assessment, B. Geofis. Teor. Appl., 45, 271–284, 2004.

Albarello, D., Camassi, R. and Rebez, A.: Detection of space and time heterogeneity in the completeness of a seismic catalog
330 by a statistical approach: An Application to the Italian Area, B. Seismol. Soc. Am.; 91, 1694-1703, https://doi.org/10.1785/0120000058, 2001.

Albarello, D., D'Amico, V., Gasperini, P., Pettenati, F., Rotondi, R. and Zonno G.: Nuova formulazione delle procedure per la stima dell'intensità macrosismica da dati epicentrali o da risentimenti in zone vicine, Progetto DPC-INGV S1, http://esse1.mi.ingv.it/d10.html, 2007.

335 Albini, P.: Documenting Earthquakes in the United States of the Ionian Islands, 1815–1864, Seismol. Res. Lett., 91, 2554–2562, https://doi.org/10.1785/0220200056, 2020a.

Albini, P. and Rovida, A.: Earthquakes in southern Dalmatia and coastal Montenegro before the large 6 April 1667 event, J. Seismol., 22, 3, 721-754, https://doi.org/10.1007/s10950-018-9730-4, 2018.

Albini, P., Camassi, R., Castelli, V. and Stucchi, M. (ed.): Miglioramento della qualità delle informazioni macrosismiche per
340 un loro utilizzo nella valutazione della pericolosità sismica, Rapporto Tecnico INGV-MI per il Servizio Sismico Nazionale, Milano, 195 (Parte A, 43 pp.; Parte B, pp. 152.), 2001.

Albini, P., Stucchi, M. and Mirto, C.: Valutazione di completezza, su base storica, dei dati del catalogo sismico CPTI (1999) nell'area corrispondente alle ZS n. 5 e 6 in ZS4 (Belluno/Cansiglio e Asolo/Vittorio Veneto). Progetto GNDT «Scenari di danno in area veneto-friulana», Rapporto Tecnico INGV-MI, 25pp, 2003.

345 Albini, P., Locati, M., Rovida, A. and Stucchi, M.: European Archive of Historical EArthquake Data (AHEAD). Istituto Nazionale di Geofisica e Vulcanologia (INGV). https://doi.org/10.6092/ingv.it-ahead, 2013.

Antonucci, A., Rovida, A., D'Amico, V., and Albarello, D.: Integrating macroseismic intensity distributions with a probabilistic approach: an application in Italy, Nat. Hazards Earth. Syst. Sci., 21, 2299–2311, https://doi.org/10.5194/nhess-21-2299-2021, 2021.

350     Azzaro, R. and Castelli, V.: Materiali per un catalogo di terremoti etnei dal 1600 al 1831, Quaderni di Geofisica, 123, Istituto Nazionale di Geofisica e Vulcanologia (INGV), Roma, 284 pp, 2015.

Azzaro, R., Barbano, M.S., D'Amico, S. and Tuvè, T.: The attenuation of seismic intensity in the Etna region and comparison with other Italian volcanic districts, Ann. Geophys-Italy., 49, 4/5, https://doi.org/10.4401/ag-3113, 2006.

Bakun, W. H. and Wentworth, C. M.: Estimating earthquake location and magnitude from seismic intensity data, B. Seismol.
355     Soc. Am., 87, 1502–1521, 1997.

Baratta, M.: I terremoti d'Italia, Saggio di storia geografia e bibliografia sismica italiana, Torino, pp. 951, 1901.

Camassi, R.: Catalogues of historical earthquakes in Italy, Ann. Geophys-Italy., 47(2-3), https://doi.org/10.4401/ag-3329, 2004.

Camassi, R., Castelli, V., Molin, D., Bernardini, F., Caracciolo, C. H., Ercolani, E. and Postpischl, L.: Materiali per un catalogo
360     dei terremoti italiani: eventi sconosciuti, rivalutati o riscoperti. Quaderni di Geofisica, 96, Istituto Nazionale di Geofisica e Vulcanologia (INGV), Roma, 53 pp, 2011.

Carletti, F., and Gasperini, P.: Lateral variations of seismic intensity attenuation in Italy, Geophys. J. Int., 155, 839-856, https://doi.org/10.1111/j.1365-246X.2003.02073.x, 2003.

D'Amico, V. and Albarello, D.: SASHA: A Computer Program to Assess Seismic Hazard from Intensity Data, Seismol. Res.
365     Lett., 79 (5), 663–671, https://doi.org/10.1785/gssrl.79.5.663, 2008.

Faenza, L. and Michelini, A.: Regression analysis of MCS intensity and ground motion parameters in Italy and its application in ShakeMap, Geophys. J. Int., 180(3), 1138–1152. 10.1111/j.1365-246X.2009.04467.x, 2010.

Gasparini, C., Conte, S., Rocchetti, E., Saraceni, A. M. and Vannucci, C.: Bollettino macrosismico 1999, Istituto Nazionale di Geofisica (INGV), Roma, 99, 2003.

370     Gasparini, C., Conte, S. and Vannucci, C.: Bollettino macrosismico 2001-2005, Istituto Nazionale di Geofisica e Vulcanologia (INGV), Roma, CD-ROM, 2011.

Gasperini, P, Bernardini, F, Valensise, G., and Boschi, E.: Defining seismogenic sources from historical earthquake felt reports, B. Seismol. Soc. Am., 89, 94–110, 1999.

Gasperini, P., Vannucci, G., Tripone, D. and Boschi, E.: The Location and Sizing of Historical Earthquakes Using the
375     Attenuation of Macroseismic Intensity with Distance, B. Seismol. Soc. Am., 100, 2035-2066, http://dx.doi.org/10.1785/0120090330, 2010.

Gomez-Capera, A.A., D'Amico, M., Lanzano, G., Locati, M. and Santulin, M.: Relationships between ground motion parameters and macroseismic intensity for Italy, Bull. Earthq. Eng., 18, 5143–5164, 2020.

Grünthal, G.: European Macroseismic Scale 1998 (EMS-98), Cahiers du Centre Européen de Géodynamique et de
380     Séismologie, 13, 1–99, 1998.

Guidoboni, E., Ferrari, G., Mariotti, D., Comastri, A., Tarabusi, G. and Valensise, G.: CFTI4Med, Catalogue of Strong Earthquakes in Italy (461 B.C.-1997) and Mediterranean Area (760 B.C.-1500), INGV-SGA, http://storing.ingv.it/cfti4med/, 2007.

Locati, M., Rovida, A., Albini, P. and Stucchi, M: The AHEAD Portal: A Gateway to European Historical Earthquake Data. Seismol. Res. Lett., 85, 3, pp.727-734. https://doi.org/10.1785/0220130113, 2014.

Locati, M., Camassi, R., Rovida, A., Ercolani, E., Bernardini, F., Castelli, V., Caracciolo, C.H., Tertulliani, A., Rossi, A., Azzaro, R., D'Amico, S., Conte, S. and Rocchetti, E.: Database Macrosismico Italiano (DBMI15), Istituto Nazionale di Geofisica e Vulcanologia (INGV) [data set], https://doi.org/10.13127/DBMI/DBMI15.2, 2016.

Locati, M., Camassi, R., Rovida, A., Ercolani, E., Bernardini, F., Castelli, V., Caracciolo, C.H., Tertulliani, A., Rossi, A., Azzaro, R., D'Amico, S., Conte, S., Rocchetti, E. and Antonucci, A.: Database Macrosismico Italiano (DBMI15), versione 4.0, Istituto Nazionale di Geofisica e Vulcanologia (INGV) [data set], https://doi.org/10.13127/DBMI/DBMI15.2, 2022.

Lolli, B., Pasolini, C., Gasperini, P. and Vannucci, G.: Prodotto 4.8: Ricalibrazione dell'equazione di previsione di Pasolini et al. (2008), In: Meletti, C. e Marzocchi, W. (a cura di), Il modello di pericolosità sismica MPS19, rapporto finale, Centro Pericolosità Sismica, Istituto Nazionale di Geofisica e Vulcanologia, maggio 2019, Roma, 168, 2019.

Mak, S., Clements, R.A. and Schorlemmer, D: Validating Intensity Prediction Equations for Italy by Observations, B. Seismol. Soc. Am., 105, 6, 2942–2954, doi: https://doi.org/10.1785/0120150070, 2015.

Medvedev, S., Sponheuer, W., and Karník, V.: Neue seismische Skala Intensity scale of earthquakes, in: 7. Tagung der Europäischen Seismologischen Kommission vom 24.9. bis 30.9.1962, in Jena, Veröff. Institut für Bodendynamik und Erdbebenforschung in Jena, Deutsche Akademie der Wissenschaften zu Berlin, 77, 69–76, 1964.

Molin, D., Bernardini, F., Camassi, R., Caracciolo, C.H., Castelli, V., Ercolani, E. and Postpischl, L.: Materiali per un catalogo dei terremoti italiani: revisione della sismicità minore del territorio nazionale, Quaderni di Geofisica, 57, Istituto Nazionale di Geofisica e Vulcanologia (INGV), Roma, 75 pp, 2008.

Monachesi, G. (ed.): Revisione della sismicità di riferimento per i comuni di Cerreto d'Esi (AN), Esanatoglia (MC), Serra San Quirico (AN), Osservatorio Geofisico Sperimentale, Macerata, Internal report, 240 pp, 1987.

Oliveti, I., Faenza, L. and Michelini, A.: New reversible relationships between ground motion parameters and macroseismic intensity for Italy and their application in ShakeMap, Geophys. J. Int., 231, 2, 1117–1137, https://doi.org/10.1093/gji/ggac245, 2022.

Pasolini, C., Albarello, D., Gasperini, P., D'Amico, V. and Lolli, B.: The attenuation of seismic intensity in Italy, Part II: Modeling and validation, B. Seismol. Soc. Am., 98, 692-708, http://dx.doi.org/10.1785/0120070021, 2008.

Provost, L. and Scotti, O.: QUake-MD: Open-Source Code to Quantify Uncertainties in Magnitude–Depth Estimates of Earthquakes from Macroseismic Intensities, Seismol. Res. Lett., 91, 2520– 2530, https://doi.org/10.1785/0220200064, 2020.

Rovida, A., Locati, M., Camassi, R., Lolli, B. and Gasperini, P. (eds): Catalogo Parametrico dei Terremoti Italiani (CPTI15). Istituto Nazionale di Geofisica e Vulcanologia (INGV), [data set], https://doi.org/10.6092/INGV.IT-CPTI15, 2016.

Rovida, A., Locati, M., Antonucci, A. and Camassi, R.: Italian Archive of Historical Earthquake Data (ASMI), Istituto Nazionale di Geofisica e Vulcanologia (INGV), [data set], https://doi.org/10.13127/asmi, 2017.

Rovida, A., Albini, P., Locati, M. and Antonucci, A.: Insights into Preinstrumental Earthquake Data and Catalogs in Europe, Seismol. Res. Lett., 91(5), 2546–2553. https://doi.org/10.1785/0220200058, 2020a.

Rovida, A., Locati, M., Camassi, R., Lolli, B. and Gasperini, P.: The Italian earthquake catalogue CPTI15, B. Earthq. Eng, 18, 2953-2984, https://doi.org/10.1007/s10518-020-00818-y, 2020b.

420 Rovida, A., Antonucci, A., and Locati, M.: The European Preinstrumental Earthquake Catalogue EPICA, the 1000–1899 catalogue for the European Seismic Hazard Model 2020, Earth Syst. Sci. Data Discuss. [preprint], https://doi.org/10.5194/essd-2022-103, in review, 2022a.

Rovida, A., Locati, M., Camassi, R., Lolli, B., Gasperini, P. and Antonucci, A.: Catalogo Parametrico dei Terremoti Italiani (CPTI15), versione 4.0. Istituto Nazionale di Geofisica e Vulcanologia (INGV), [data set], 425 https://doi.org/10.13127/CPTI/CPTI15.4, 2022b.

Sieberg, A.: Geologische, physikalische und angewandte Erdbebenkunde. G. Fischer, Jena,, 1923.

Stucchi, M.: Historical seismology and seismic hazard, Ann. Geoph., 36(1), https://doi.org/10.4401/ag-4314, 1993.

Stucchi, M., Albini, P., Mirto, M., and Rebez, A.: Assessing the completeness of Italian historical earthquake data, Ann. Geophys.-Italy, 47, 2–3, https://doi.org/10.4401/ag-3330, 2004.

430 Stucchi, M., Meletti, C., Montaldo, V., Crowley, H., Calvi, G. M., and Boschi, E.: Seismic hazard assessment (2003–2009) for the Italian building code, B. Seismol. Soc. Am., 101, 1885–1911, https://doi.org/10.1785/0120100130, 2011.

Tertulliani, A., Arcoraci, L., Berardi, M., Bernardini, F., Brizuela, B., Castellano, C., Del Mese, S., Ercolani, E., Graziani, L., Maramai, A., Rossi, A., Sbarra, M. and Vecchi, M.: Emilia 2012 sequence: the macroseismic survey, in "The Emilia (northern Italy) seismic sequence of May-June, 2012: preliminary data and results", M. Anzidei, A. Maramai and P. Montone (Eds), 435 Ann. Geophys-Italy., 55, 4,  https://doi.org/10.4401/ag-6140, 2012.

Wood, H.O. and Neumann, F.: Modified Mercalli intensity scale of 1931. B. Seismol. Soc. Am., 21, 277–283, https://doi.org/10.1785/BSSA0210040277, 1931.