Good practices in PSHA: declustering, b-value estimation, foreshocks and aftershocks inclusion; a case study in Italy

Matteo Taroni[1,*] and Aybige Akinci[1]

(1) Istituto Nazionale di Geofisica e Vulcanologia, INGV, Sezione Roma1, Via di Vigna Murata 605, 00143 Roma, Italy

*Corresponding author

Submitted to Geophysical Journal International, May 2020

Revised version

Abbreviated title: Good practices in PSHA

Contact details: Matteo Taroni; email: matteo.taroni@ingv.it

Summary

The classical procedure of the Probabilistic Seismic Hazard Analysis (PSHA) requires a Poissonian distribution of earthquakes. Seismic catalogs follow a Poisson distribution just after the application of a declustering algorithm that leaves only one earthquake for each seismic sequence (usually the stronger, i.e. the main shock). Removing earthquakes from the seismic

catalogs leads to underestimation of the annual rates of the events and consequently associate with low seismic hazard as indicated by several studies. In this study, we aim investigating the performance of two declustering methods on the Italian instrumental catalog and the impact of declustering on estimation of the *b*-value and on the seismic hazard analysis. To this end, first the spatial variation in the seismicity rate was estimated from the declustered catalogs using the adaptive smoothed seismicity approach, considering small earthquakes ($M_w \geq 3.0$). We then corrected the seismicity rates using new approach that allows for counting all events in the complete seismic catalog by simply changing the magnitude frequency distribution. The impact of declustering on seismic hazard analysis is illustrated using PSHA maps in terms of peak ground acceleration (PGA) and spectral acceleration (SA) in 2 s, with 10% and 2% probability of exceedance in 50 years, for Italy. We observed that the hazard calculated from the declustered catalogs was always lower than the hazard computed using the complete catalog. These results are in agreement with previous results obtained in different parts of the world.

Key words: Earthquake hazards, Probabilistic forecasting, Statistical seismology

Introduction

Probabilistic Seismic Hazard Analysis (PSHA) yields a common perspective estimating of the likelihood of earthquake shaking that may occur over a given time period considering both a model of earthquake occurrence and an estimate of the distribution of shaking that results from earthquakes. These results are often used for important actions and agreements such as those made by decision makers from governments and other industry end users, in relation to lives and

property. In standard PSHA, the occurrence of earthquakes is modelled using a Poisson process, where the occurrence of a future earthquake is independent of the occurrence of previous earthquakes from the same source (Cornell 1968). The Poisson assumption holds for declustered catalogs, where the declustering procedure keeps the largest events but removing those having a negligible contribution in terms of ground shaking. Moreover, complete (i.e. not-declustered) seismic catalogs can provide a biased view of the future spatial variability of seismicity rates, in particular for instrumental catalogs (Marzocchi and Taroni, 2014). Indeed, the instrumental catalogs, which contain the events of the last 30-40 years, show their greatest concentration in zones where seismic sequences have occurred (see Fig. 1 as example for the Italian catalog).

Declustering algorithms remove events from a sequence, typically using space-time windows that depends on the magnitude of the earthquake (the time window is proportional to the magnitude of the event) to identify aftershocks and foreshocks. One of the most widely used declustering algorithms for PSHA is the one developed by Gardner and Knopoff (1974) since it usually produces Poisson declustered catalogs. There exist many other declustering algorithms, for example the Reasenberg (1985) method, based on the space-time linked-windows, as defined in Omori's Law, or the stochastic declustering method of Zhuang et al. (2002) based on the epidemic type aftershock sequence (ETAS) model. In this work, we adopted Gardner and Knopoff and Reasenberg declustering methods, because they are the most commonly used approaches for seismic hazard studies. We prefer not to consider the stochastic declustering approach, since it provides a set of possible declustered catalogs, rather than a unique declustered one (Zhuang et al. 2002).

Although the standard seismicity models require a declustered earthquake catalog of independent events for calculation of time-independent earthquake rates, recent studies have shown that

declustering may lead to a significant underestimation of substantial hazard (Boyd 2012; Iervolino et al. 2014; Marzocchi and Taroni 2014; Teng and Baker 2019). These studies compared the classical declustered PSHA with a sort of "corrected" PSHA model that also include aftershocks. This correction was made with different techniques: by simulating aftershocks Boyd (2012), by integrating the aftershocks effect in the hazard formula (Iervolino et al. 2014), or by using theoretical considerations (Marzocchi and Taroni 2014). Therefore, in the present study, we incorporated foreshocks and aftershocks into the time-independent seismic hazard analysis, considering that the main shock, and each of its associated fore/after-shocks, have an opportunity to exceed a given amount of ground motion at a site.

We propose a straightforward approach that does not perform a series of simulations to correct for the removed events or missing rate due to declustering as given by Bold (2012). Similarly, it does not require modification of the PSHA equations as given by Iervolino et al. (2014), because we use the same theoretical considerations of Marzocchi and Taroni (2014).

We first declustered the instrumental catalog of Italian seismicity using two common approaches, Gardner and Knopoff (1974), hereafter GK74 and Reasenberg (1985), hereafter REAS85. After declustering the catalogs, we examined whether they conformed to a Poisson distribution using a Kolmogorov-Smirnov one sample test (Luen and Stark, 2012). Subsequently, the frequency distributions of magnitude were separately calculated from the two declustered catalogs and the complete one. The spatial distribution of the earthquake rates was then estimated using the adaptive smoothing method (Helmstetter et al. 2007), that relies on the seismic catalog and the allocation of seismic density over the study area. In fact, the adaptive smoothed seismicity uses individual smoothing distances for each earthquake, so that the smoothing distance varies with the distribution of the seismicity; the resulting smoothing

distances are relatively smaller in regions where seismic activity is higher compared with regions with sparse seismicity. Finally, we correct the seismicity rates for those removed events in the declustered catalogs through a simple magnitude-frequency relationship (that will be explained in detail in subsequent paragraphs) and investigate the impact of two declustering algorithms in the ground-motion seismic hazard in Italy. The results are presented with PSHA maps in terms of PGA and SA in 2 s, with 10% and 2% probability of exceedance in 50 years for Italy and as hazard curves for three different sites.

Instrumental Italian seismic catalog and declustering procedures

In this study, we used an Italian instrumental catalog that contains located earthquakes with a harmonized M_w magnitude over the past almost 37 years (1981-2017 April) (Gasperini et al. 2013). We considered only events above the magnitude of completeness suggested by the same author of the catalog, M_w 3.0, and with a depth ≤ 30 km; this catalog contains 7283 events. We then removed dependent events from the catalog, assuming that our forecasting results will have been affected at a given location by numerous foreshock/aftershock sequences associated with moderate and large main shocks, using the declustering algorithms proposed by GK74 and REAS85. Both for the GK74 and the REAS85 algorithms, we used the ZMAP tool (Wiemer, 2001); for the REAS85 algorithm, we adopted the default ZMAP parameters (τ_{min}=1, τ_{max}=10, x_k=0.5, x_{meff}=3.5, p_1=0.95, r_{fact}=10) without any adjustments (as in Luen and Stark 2012).

In Figure 1a-f, we present the catalog before and after declustering, to emphasize the results from different algorithms. The complete catalog shows four important seismic sequences in Italy; the Emilia 2012 sequence in northern Italy, and the Colfiorito 1997, L'Aquila 2009 and Amatrice-

Norcia 2016 sequences in central Italy (Fig. 1d). More than half of the earthquakes in the catalogue were identified as either aftershock or foreshock and the effect of the declustering increased with decreasing magnitude, as expected. GK74 removed more earthquakes compared to the REAS85 algorithm (Fig. 2). The number of events decreases to about 70% and 30% after the GK74 and the REAS85 declustering processes, respectively, using a cut-off magnitude of exceeding M_w 3.0. This was because the main shock windows are larger in space, but shorter in time, in the REAS85 algorithm compared to that of the GK74 algorithm. However, both declustering algorithms preserved a larger percentage of events of magnitudes between 4.5 and 5.5. The percentage of declustered event was approximately 45% and 30% for magnitudes 4.2-4.5 while it increased up to 55% and 40% for larger magnitudes, around 5.2-5.5 using the GK74 and REAS85 algorithms respectively.

We then tested the declustered catalogs to investigate whether they still exhibited a Poisson distribution. The Kolmogorov-Smirnov test was performed to check for a temporally homogeneous Poisson hypothesis (Luen and Stark 2012) of each declustered and complete catalog. In Figure 3, we present the number of earthquakes as a function of transformed time. This time is defined by the time of the event measured from the beginning of the catalog divided by the time length of the catalog, 36.33 years: it ranges between 0 and 1. The transformed time is calculated for the complete catalog and the two declustered catalogs from the GK74 and REAS85 algorithms, and the distribution is used to compute the *p*-values of the Kolmogorov-Smirnov tests for the temporally homogeneous Poisson hypothesis. The *p*-value associated with each test indicated that the GK74 declustered catalog, with a *p*-value of 0.11,was the only catalog with a Poisson distribution, while the REAS85 declustered catalog exhibited a very low *p*-value ($<10^{-10}$). The REAS85 failed to remove the aftershocks completely in the four major Italian

seismic sequences (Fig 1d, e). This confirmed that the Reasenberg declustering method does not generate a Poisson distribution for the Italian catalog, in agreement with the results obtained by Luen and Stark (2012), for the Californian catalog.

Calculating seismicity rates for the declustered and the complete catalogs

We calculated seismicity rates from the instrumental catalogs following the smoothed seismicity procedure, as detailed in the following section. Figure 4a shows the magnitude frequency distribution of the events, that is the cumulative annual rate of the events (in a $\log_{10}$ scale) *versus* the magnitude. As discussed in the previous paragraph, the annual rate of events for the GK74 declustered catalog was smaller than both the complete catalog and the REAS84 declustered catalog. The large number of removed events in the GK74 declustered catalog has a direct impact on the estimation of the *b*-value.

The *b*-values were calculated for each of the three catalogs as a uniform value over the study region, using the classical maximum-likelihood estimation method (Aki, 1965). Table 1 shows the estimated *b*-values for starting magnitude M_w 3.0, 4.0 and 4.5; interestingly, some *b*-values of declustered catalogs exceeded the ones of the complete catalog. These declustering algorithms remove 40-50% of events in the range M_w 5.0-5.5, leading to an increase of *b*-values for some magnitude thresholds (Fig. 2). This result concurs with the study of Christophersen et al. (2011) for New Zealand.

The most important result for the estimation *b*-value is shown in Figure 5. The estimated *b*-value strongly depends on the magnitude of completeness of the declustered catalog, especially if the catalog was declustered with the GK74 method. In general, if the magnitude of completeness is

small, the *b*-value estimation is biased and becomes smaller (see Appendix A for other details). As noted in Teng and Baker (2019) and Llenos and Micheal (2020), a lower *b*-value obtained from the declustered catalogs can lead to an unrealistic estimation of the number events of higher magnitudes, producing many more events than those of the complete catalog (Fig. 4b). However, the *b*-values estimated for the $M_w \geq 4.5$ events from all catalogs (the GK74, REAS84 declustered and the complete catalogs) yielded similar results (Table 1, Fig. 4c).

We note that the *b*-value estimated from the whole catalog may not be appropriate for some specific zones. For example, volcanic regions that infer generally different type of seismic behavior and result in higher *b*-values, as demonstrate by several studies for Italy and globally (Murru et al. 1999; Bridges and Gao 2006).

Constructing seismicity rate model from the declustered catalogs

To build an earthquake rate model for Italy, we used the adaptive smoothing seismicity approach described by Helmstetter et al. (2007), where the kernel bandwidth varies as a function of earthquake density. Since the adaptive smoothed model alters the smoothing distance with the local seismicity density, the resulting smoothing distances are smaller in regions of high seismicity than in regions with sparse seismicity (Moschetti 2015).

To construct the adaptive seismicity model, we first divided the whole observational region into a grid with spacing of 0.1° in latitude and longitude (about 10 km × 10 km grid size). We then counted the number of earthquakes with magnitudes exceeding the completeness magnitude (M_c) in each grid cell. Here, we used small earthquakes $M_w \geq 3.0$ from the instrumental seismicity to increase the spatial resolution of the forecast model and identify active fault structures that might

be locations of larger future events. Therefore, the seismicity rates were spatially smoothed using a two-dimensional Gaussian function. We adopted the nearest neighbor number equal to 2, chosen through maximum-likelihood optimization, as described in detail by Akinci et al. (2018) for Italy. Finally, the annual rates and *b*-values are estimated from $M_w \geq 4.5$ events; annual rates and *b*-values were 3.11 and 1.08 for GK74, 3.80 and 1.12 for RES85 (Table 1), respectively. We used the annual rates and *b*-values starting from M_w 4.5 since the hazard computation starts from this value (as explained further in the paper). As previously discussed, the *b*-value estimations from M_w 4.5 using the declustered catalogs, and in the GK74 method, were similar to that derived from the complete catalog.

Incorporating aftershocks and foreshocks into the seismicity rate model for PSHA purposes

To incorporate aftershocks and foreshocks into time-independent PSHA, we followed the idea proposed by Marzocchi and Taroni (2014), hereafter MT14, based on Le Cam's theorem (Le Cam, 1960) and its subsequent generalization (Serfling 1975); these theorems describe when the Poisson distribution can approximate a sum of random variables. In PSHA, these random variables are the ground-motion exceedances that can also exhibit a Poisson distribution, if generated by a seismic sequence that contains foreshock and aftershocks.

We can have a Poisson distribution of ground-motion exceedances from the complete catalog only if the conditions of "sufficiently high probability of taking zero value and sufficiently weak mutual dependence" are fulfilled. As shown by MT14, these conditions are met when we consider a probability of 10% or less of having an exceedance in 50 years; a common value used

for PSHA. To estimate the 10% (or less) probability of exceedance in 50 years, it is possible to leave foreshocks and aftershocks in the catalog, since the exceedance rate will remain as a Poisson distribution (Marzocchi and Taroni, 2014).

In more detail, the exceedance rate $\Lambda(IM > x)$ of an intensity measure IM is computed with the formula:

$$\Lambda(IM > x) = \sum_{i=1}^{N} \sum_{j=1}^{R} \sum_{k=1}^{M} \lambda_{ijk} P\left(IM > x | m_k, r_{ij}\right) \quad (1)$$

where N is the total number of spatial cells, R is the range of possible distances, M is the total number of magnitude bins m_k, λ_{ijk} is the earthquake rate corresponding to the *i-th* spatial cell, the *j-th* distance and *k-th* magnitude bin, and $P\left(IM > x | m_k, r_{ij}\right)$ is the probability that an event with magnitude m_k and distance r_{ij} can produce an $IM > x$ (this probability is given by the ground-motion prediction equations).

Equation (1) shows that the exceedance rate Λ of a certain measure of intensity is the sum of all the contributions λ_{ijk} from the N spatial cells and M magnitude bins. Cornell (1968) demonstrated that if all the λ_{ijk} come from independent Poisson distributions, their sum Λ will also follow a Poisson distribution. MT14 showed that the previous statement is a sufficient condition for the distribution of Λ to be Poisson, but is not a necessary condition. Using Le Cam's Theorem, they discovered that if we compute λ_{ijk} using a complete catalog, we can still have a Poisson distribution for their sum Λ, if we consider 10% or lower probability of exceedance. Since most of the PSHA maps globally deal with probability of exceedance of 10% or lower, in this work we follow the basic idea of MT14 and include all the events (foreshocks, mainshocks and aftershocks) into the computation of PSHA. Practically, and according to this

method, to compute a PSHA including all the seismic events, you merely have to substitute the declustered earthquake rate model with an earthquake rate model that considers the complete catalog.

In this study, we slightly modified the original MT14 approach. Instead of using their correction factor, we directly used the *b*-value and the annual rate of the complete catalog, according to the following rules:

1) Spatial distribution of earthquake rates: decluster the catalog (here we used both approaches from the GK74 and REAS85 algorithms) and obtain the spatial distribution of earthquakes over the study area. The spatial distribution of the smoothed seismicity needs to be calculated from the declustered catalog to remove the spatial bias of the complete catalog caused by the fact that the few clusters recorded in the instrumental catalogs are not representative of the whole distribution of all possible clusters (MT14).

2) Magnitude-frequency distribution: introduce the foreshock and the aftershock events into the earthquake rate distribution considering the *b*-value and annual rate estimated from the complete catalog.

Here, we used the estimation from the complete catalog considering magnitude $M_w \geq 4.5$, obtaining a *b*-value 1.05 and an annual rate of 5.59 (Table 1). In this way, we assumed that the entire study region has similar clustering properties; it is easy to change this assumption by adopting local estimation of the annual rate and *b*-value (still using the complete catalog) if some specific zones show a peculiar behavior (e.g. volcanic or geothermal zones).

With respect to the spatial distribution of events, usually the seismic catalogs used for smoothing are the instrumental catalogs, with a low completeness magnitude (e.g. M_w 3.0 or lower). These catalogs cover decades but rarely for more than one century, in this period of time, the seismic

sequences occur only in some zones and most of the events occur in these zones. Without declustering, the earthquake rate would be concentrated in the zones that have experienced seismic sequences, resulting in a biased view of the seismicity of the territory. Therefore, for the spatial distribution, we suggest using the declustered catalog, as in MT14. In contrast, for the magnitude-frequency distribution, one needs to use the complete catalog.

Seismic hazard calculations

The ground-shaking hazard is computed following the standard methodology (Cornell 1968), using the rate at each spatial grid node in conjunction with a Gutenberg-Richter magnitude frequency distribution, together with ground-motion prediction equations, to obtain the rate of exceedance at each ground-motion level.

We calculated the annual rate, $\Lambda(IM > x)$, of ground-motion IM exceeding x according to equation (1). For hazard calculations, we chose a low cutoff magnitude of M_{min} = 4.5, based on the observation that earthquakes with magnitude less than 4.5 are generally unlikely to cause significant damage in Italy; this value was also the value selected for the future Italian seismic hazard map (Akinci et al. 2018, Meletti et al. 2019). The upper limit, the maximum magnitude of the Truncated Gutenberg-Richter distribution was set to M 7.5, the same used by Akinci et al. (2018). This value can be inappropriate for the volcanic zones of Italy; a more detailed study of these regions is left for future work. The λ_{ijk} in each spatial bin (0.1° x 0.1°) and magnitude bin (0.1) were modeled using the smoothing distribution described above and according to the estimated annual rate and b-value. Finally, we calculated the 50 year probability of exceeding various ground-motion levels at a series of gridded sites following equation (1), using the λ_{ijk}

together with the selected ground-motion equations (it is possible to derive the $P(IM > x|m_k, r_{ij})$ in equation 1 directly from the GMPE). In the present study, we selected only one GMPE, derived from the Italian strong-motion data for the active shallow crustal region, given by Bindi et al. (2011), hereafter ITA10, to assess the ground-shaking hazard without exploring uncertainty resulting from the different choices of GMPEs.

PSHA results

We present the PSHA results for Italy first in a set of ground-motion intensity distribution maps from the declustered catalogs using GK74 and RES85 algorithms (Fig. 6a and 6b), and second, as hazard curves corresponding to three cities, Rome, Norcia and Tolmezzo (in Friuli region). Norcia and Tolmezzo were selected to reflect relatively high seismic density compared to with that of Rome, in the earthquake rate model developed in the present study. All results correspond to site class A, V_{s30}=800m/s, based on the shear-wave velocity intervals in the uppermost 30 m, V_{s30}, according to the EC8 (CEN, 2003), as defined in Bindi et al. (2011).

Figures 6a and 6b show the time-independent hazard maps of mean PGA values. These were calculated using the ITA10 GMPE, with an exceedance probability of 10% in 50 years, determined using the background smoothed seismicity from the instrumental seismic catalog declustered with the GK74 and RES85 algorithms, respectively. The maps for PGA show high accelerations in the Amatrice, Norcia and Irpinia areas located along the Apennine belt, around 0.40-0.50g from the REAS85 declustered catalog, while the ground motions are lower, around 0.30-0.35g, from the GK74 catalog because the latter approach removes many more events and result in a lower seismic hazard.

Impact of the foreshocks and aftershocks in PSHA

Figures 7a and b show the PGA map, constructed using the corrected seismic rates, which include foreshocks and aftershocks in the PSHA. The spatial distribution pattern of the ground motion, in terms of PGA, does not differ between the two maps determined from the declustered and the corrected models. However, the seismic intensity level increased significantly when the GK74 algorithm was applied, compared to when the REAS85 algorithm was applied. Typically, results depend on how the catalog has been declustered because each declustering technique, with its own parameters, eliminates different percentages of earthquakes. Figure 7a presents the corrected hazard map for the GK74 model; the maximum ground acceleration was approximately 0.45g, 0.1g higher than the corresponding PGA for the declustered hazard map (see Fig. 6a). This difference is lower for the declustered hazard map derived from the REAS85 model (Fig. 7b).

In Figures 6a, 6b, 7a and 7b, PSHA maps that were determined both from the declustered and corrected seismicity rates demostrated larger ground acceleration along the seismically active Apennine belt and as well as in volcanic regions, with similar PGA values. Naturally, in active tectonic regions, earthquakes occur with larger magnitudes and produce larger ground accelerations, than those observed in volcanic areas. We adopted a uniform distribution of *b*-value and maximum magnitude over the whole study region; however, many studies have demonstrated that *b*-value and maximum magnitude are often different in volcanic areas compared to typical tectonic earthquake sequences (e.g., Frohlich and Davis, 1993; Wyss et al. 1997; McNutt, 2005). Moreover, we used only one GMPE, ITA10, derived for active shallow crustal regions, for the whole territory, without considering a specific ground-motion relation for volcanic regions. Therefore, our PGA maps may not be appropriate for volcanic regions.

We also computed the differences in PGA value with an exceedance probability of 10% in 50 years, calculated with the corrected seismicity rates and the declustered seismicity rates, using the GK74 and REAS85 algorithms, ($PSHA_{COR}$-$PSHA_{DECL}$) as shown in Figure 8a and b. We observed significant effects on the seismic hazard caused from using different declustering algorithms. The GK74 algorithm removed many more events than the REAS85 algorithm, leading to a lower seismic hazard level, compared to that calculated from the catalog declustered by REAS85.

In Figure 9a and b, we present the percentage difference between the 10% in 50 years PGA calculated with the corrected seismicity rates for the declustered GK74 and the REAS85 catalogs, and the declustered seismicity, that is, (($PSHA_{COR}$-$PSHA_{DECL}$)/$PSHA_{DECL}$*100.). Our results demonstrated that the correction for foreshock and aftershock events introduced a significant contribution to the hazard maps computed with the declustered seismic catalogs. The hazard map represents up to 25-30% increase in PGA values given with an exceedance probability of 10% in 50 years. In the case of the REAS85 declustering approach, the relative difference due to the seismicity rate correction was 15% smaller, compared to using the GK74 procedure. The relative percentage variation was larger (up to 50%) at borderlines, while the absolute seismic hazard in terms of PGA was relatively small (0.01 or 0.02 g). This effect, which may arise from some approximations for small PGA values, will be investigated in future work.

In Figures 10a and b, we demonstrate the differences in PGA values with exceedance probability of 10% in 50 years between the corrected seismicity rates for the declustered GK74 and the REAS85 catalogs, that is, ($PSHA_{DEC\_REAS85}$ – $PSHA_{DEC\_GK74}$), and between the corrected catalogs, that is, ($PSHA_{COR\_REAS85}$ – $PSHA_{COR\_GK74}$). A large ground-motion difference in the PGA values of around 0.25g was observed in the Amatrice-Norcia area, where recent seismic

activity occurred (Fig. 10a and b). This difference was similar, around 0.30g, in the Etna volcanic region. The same order of differences in the ground acceleration value in these two zones may be due to problems of the model in volcanic zones, discussed earlier.

Hazard curves and percentage variability

Complete catalogs exhibited higher seismic hazard compared to the declustered catalogs both for a 475-year return period (10% probability of exceedance in 50 years), and a 2475-year return period (2% probability of exceedance in 50 years) (Fig. 11a, b, and c). This was observed in the PGA and SA (2 s), especially at Norcia and Tolmezzo (Fig. 11a and b). In the city of Norcia, seismic hazard was larger for the model derived from the REAS85 declustered catalog because the algorithm removes small number of events from the complete catalog, resulting in higher seismic activity compared to that obtained from the GK74 catalog (Fig. 11a). The PGA values were approximately 0.30g and 0.55g, inferred from the GK74 and REAS85 declustered models, and 0.45g and 0.60g from the corrected models, for the 475-year return periods, respectively. In the Tolmezzo city, where we also observed high seismic activity as in the Norcia area, seismic hazard curves provided similar PGA values calculated using the two declustered catalogs, GK74 and REAS85 (Fig. 11b). However, the seismic hazard, in terms of PGA, increased slightly for the corrected GK74 declustered catalog at the 475-year and 2475-year return periods. In contrast to Norcia, there was no high density of seismic activity caused by the large magnitude events and/or the active faults near the site in Rome (Fig.11c). The seismic hazard in Rome was mainly associated with two active seismogenic areas, the Alban Hills and the central Apennines regions, located about 20 km southeast and 80– 100 km east of central Rome, respectively (Akinci et al. 2009). Naturally, the effect of distant dominant sources becomes more significant than closer small sources for long period ground-motions. In fact, at this site the seismic hazard became

relevant also at longer spectral periods, SA (2 s), for the corrected models (Fig. 11c), being not very different from the seismic hazard estimated from the GK74 and REAS85 declustered catalogs.

Figures 12a, b and c present the relative difference in percentage between the complete and declustered seismic hazard curves, in terms of PGA and SA (2 s), as a function of return periods for Norcia, Tolmezzo, and Rome. The percentage variability was computed from differences between the PSHA exceeding the probability of 2% and 10% in 50 years, calculated with the corrected seismicity rates and with the declustered rates, normalized to the value of the declustered PSHA map.

The PGA and SA (2 s) ground motion corresponding to 475-year return period was ~22% and ~20% higher when derived from the complete catalog, compared to the catalog declustered by GK74, in Norcia. However, this percentage difference was about 17% both for PGA and SA (2 s) at a larger return period of 2475-year. The percentage difference was smaller in the case of REAS85.

At longer return periods (2475-year and larger), the contribution of foreshocks/aftershocks was similar for the two spectral ordinates, while at the shorter return periods (475-year), it was larger for PGA than SA (2 s) in Norcia for the GK74, where seismic hazard was more sensitive to higher frequencies than lower ones. This difference and the contribution of foreshocks/aftershocks were similar, both for the short-period and the long-period ground motion for REAS85 case. In contrast, the percentage variability of the PGA and SA (2 s) was larger, up to 27% and 21%, for the city of Rome, where seismic rate, as well as seismic hazard, was minor compared to those of the two other sites. Nevertheless, these two curves are well separated (Fig. 12c; dashed and continuous red curves). This reveals that different periods of seismic waves

contribute to hazard differently at each site, based on the frequency-magnitude distributions and distance from the relevant sources.

Conclusions

In this study, we incorporated foreshocks and aftershocks into time-independent seismic hazard analysis, by considering that not only the main shock, but also its associated shocks have a chance to exceed a given amount of ground motion at a site. We first built a seismic hazard model that produced probabilistic seismic hazard assessments for Italy, in terms of PGA and SA (2 s), using an adaptive spatially smoothed seismicity approach, along with the updated instrumental seismic catalog. To incorporate clustered events into time-independent PSHA, we followed the basic assumption and procedure described in Marzocchi and Taroni (2014). The method we proposed in this study is relatively simple and can be applied easily and globally. Our method modified and corrects the seismicity rates for the removed and missing events in the declustered catalogs through the magnitude-frequency relationship of the complete catalog. This method assumes that the rate of larger magnitudes that occurred during the instrumental catalog period is equal to the long-term rate; in Appendix B we show an example of the application of this method for an historical seismic catalog. A limitation of this method is that it assumes similar clustering properties for the entire territory: this limitation can be easily overcome by estimating the annual rate and the *b*-value in each specific sub-region (e.g. volcanic zones, geothermal fields or induced seismicity zones). Another limitation, that can't be overcome, is that this method works only for 10% or lower exceedance probability in 50 years (i.e. return period from 475-year) because it is based on the Le Cam theorem (Marzocchi and Taroni, 2014).Our first result was that the Italian catalog declustered by REAS85 did not exhibit a

Poisson distribution. Consequently, we recommend using the GK74 algorithm instead of the REAS85, to avoid a spatial distribution biased by events in the seismic sequences. Moreover, these results confirm the results obtained by Luen and Stark (2012) for the Californian catalog.

The second result concerned the estimation of the *b*-value for the purpose of PSHA. In this study, we demonstrated that estimation of *b*-value from the GK74 declustered catalog strongly depends on the completeness magnitude in the catalog. Generally, *b*-value smaller when completeness was of a lower magnitude (M_c=3.0, *b*-value=0.77), while it increased for larger magnitudes of completeness (M_c=4.5, *b*-value=1.08), and converges to that estimated for the complete catalog (*b*-value ~1.0). However, the *b*-value estimated form the catalog declustered by the REAS85 algorithm was similar to the *b*-value of the complete catalog, even for small magnitudes of completeness (M_c=3.0, *b*-value =1.01). For this reason, this may result in the REAS85 algorithm working better than the GK74 algorithm (Teng and Baker 2019). Therefore it becomes crucial to test the declustered catalogs to investigate whether they still exhibit Poisson distribution, although this test is not always applied in PSHA studies.

We demonstrated that for PSHA purpose, it is preferable to use a catalog with a lower completeness magnitude (in our case M_w 3.0) to estimate the spatial distribution of events, and to use a catalog with a higher completeness magnitude (in our case M_w 4.5) to estimate the magnitude-frequency distribution parameters (annual rate and *b*-value). In this context, the GK74 algorithm is a highly effective declustering algorithm.

The last result was that the seismic hazard, calculated using the complete catalogs, leads to increased rates of exceedance of ground motions in PGA around 0.1g at probability levels of 10% in 50 years. These findings are in accordance with the results obtained by Marzocchi and Taroni (2014). Using declustered catalogs leads to a general underestimation of the seismic

hazard, both for PGA and SA (2 s): the larger underestimations in seismic hazard correspond to areas where seismicity is denser. We observed that the impact of foreshocks and aftershocks considered through different declustering algorithms was high for the seismic hazard assessment; the ground motion at a given hazard level may increase by more than 25-30%, depending on the type of the declustering algorithm chosen.

For two selected sites, Norcia and Tolmezzo cities, we compare seismic hazard from complete and declustered catalogs. The increase, in terms of PGA and SA (2 s), was between approximately 24-25% and 16-18% for the two return periods (475-year and 2475-year), having a similar variation between the two spectral periods. This percentage change is much smaller in the declustered catalog calculated from the REAS85 algorithm, but as we highlighted previously, this algorithm is not suitable for the PSHA purposes, at least for the Italian catalogs. However, for the city of Rome where there is no high seismic density near site, the percentage variability of the PGA and the SA (2 s) values are well separated from each other, indicating that these two different spectral periods contribute to hazard differently at that site. Finally, we note that the *b*-value, maximum magnitude and GMPE adopted in this study may be inappropriate for volcanic regions, which needs further and detailed investigations.

Data and resources

The instrumental Italian seismicity catalog is described in Gasperini et al. (2013), and available upon request to the authors.

Acknowledgments

The authors are grateful for careful and useful reviews by Norman Abrahamson, Vincenzo Convertito and the Editor Margarita Segou. The authors thank to Giuseppe Falcone for helpful suggestions that improved our paper. This work was supported by the Centro per la Pericolosità Sismica (CPS) of INGV.

References

Aki, K., 1965. Maximum likelihood estimate of *b* in the formula log(*N*)=*a*-*bM* and its confidence limits, *Bull. Earthq. Res. Inst. Tokyo Univ.*, 43, 237-239.

Akinci, A., Galadini, F., Pantosti, D., Petersen, M., Malagnini, L. & Perkins, D., 2009. Effect of time dependence on probabilistic seismic-hazard maps and deaggregation for the central Apennines, Italy, *Bull. Seismol. Soc. Am*., 99(2A), 585–610.

Akinci, A., Moschetti, M. P. & Taroni, M., 2018. Ensemble Smoothed Seismicity Models for the New Italian Probabilistic Seismic Hazard Map, *Seismol. Res. Lett.,* 89 (4), 1277–1287.

Bindi, D., Pacor, F., Luzi, L., Puglia, R., Massa, M., Ameri, G. & Paolucci, R., 2011. Ground motion prediction equations derived from the Italian strong motion database, *Bull. Earthquake Eng.,* 9 (6), 1899–1920.

Boyd, O.S., 2012. Including foreshocks and aftershocks in time-independent probabilistic seismic hazard analyses, *Bull. Seismol. Soc. Am*., 102(3), 909–917.

Bridges, D.L., & Gao, S.S., 2006. Spatial variation of seismic b-values beneath Makushin Volcano, Unalaska Island, Alaska, *Earth Planet. Sci. Lett*., 245(1-2), 408-415.

CEN (Comité Européen de Normalisation), 2003. Eurocode 8: design of structures for earthquake resistance. Part 1: General rules, seismic actions and rules for buildings. Draft No 6, Doc CEN/TC250/SC8/N335

Christophersen, A., Gerstenberger, M. C., Rhoades, D. A., & Stirling, M. W., 2011. Quantifying the effect of declustering on probabilistic seismic hazard, *Proc. of the Ninth Pacific Conf. on Earthquake Engineering: Building an Earthquake-Resilient Society*.

Cornell, C.A., 1968. Engineering seismic risk analysis, *Bull. Seismol. Soc. Am.,* 58(5), 1583–1606.

Felzer, K.R., Becker, T.W., Abercrombie, R.E., Ekström, G. & Rice, J.R., 2002. Triggering of the 1999 Mw 7.1 Hector Mine earthquake by aftershocks of the 1992 Mw 7.3 Landers earthquake, *J. Geophys. Res. Solid Earth*, *107*(B9).

Frohlich, C., & Davis, S., 1993. Teleseismic b-values: or much ado about 1.0, *J. Geophys. Res.,* 98, 631–644.

Gardner, J.K., & Knopoff, L., 1974. Is the sequence of earthquakes in Southern California, with aftershocks removed, Poissonian? *Bull. Seismol. Soc. Am.,* 64(5), 1363–1367.

Gasperini, P., Lolli, B. & Vannucci, G., 2013. Empirical calibration of local magnitude data sets versus moment magnitude, *Bull. Seismol. Soc. Am.,* 103, 2227–2246.

Gruppo di Lavoro MPS, 2004. Redazione della mappa di pericolosità sismica prevista dallOrdinanza PCM 3274 del 20 marzo 2003, *Rapporto Conclusivo per il Dipartimento della Protezione Civile*, INGV, Milano-Roma, 5 pp.

Helmstetter, A., Kagan, Y.Y. & Jackson. D.D., 2007. High-resolution time-independent grid based forecast for M ≥ 5 earthquakes in California, *Seismol. Res. Lett.,*78, 78–86.

Iervolino, I., Chioccarelli, E. & Giorgio, M., 2018. Aftershocks' effect on structural design actions in Italy, *Bull. Seismol. Soc. Am.,* 104(2), 108(4), 2209–2220.

Iervolino, I., Giorgio, M. & Polidoro, B., 2014. Sequence-based probabilistic seismic hazard analysis, *Bull. Seismol. Soc. Am.,* 104(2), 1006–1012.

Kagan, Y. Y. (2002). Seismic moment distribution revisited: I. Statistical results. *Geophys. J. Int.*, *148*(3), 520-541.

Le Cam, L., 1960. An approximation theorem for the Poisson binomial distribution, *Pacific J. Math.* 10, 1181-1197.

Llenos, A.L., & Michael, A.J., 2020. Regionally Optimized Background Earthquake Rates from ETAS (ROBERE) for Probabilistic Seismic Hazard Assessment, *Bull. Seismol. Soc. Am*., doi: https://doi.org/10.1785/0120190279

Lombardi, A.M., 2003. The maximum likelihood estimator of b-value for mainshocks. *Bull. Seismol. Soc. Am*., *93*(5), 2082-2088.

Luen, B. & Stark, P.B., 2012. Poisson tests of declustered catalogues, *Geophys. J. Int.,* 189(1), 691–700.

Marzocchi, W., & Taroni, M., 2014. Some thoughts on declustering in probabilistic seismic-hazard analysis, *Bull. Seismol. Soc. Am.,* 104(4), 1838–1845.

McNutt, S.R, 2005. Volcanic seismology, *Annu. Rev Earth Planet Sci*. 32, 461–491.

Meletti, C., Marzocchi, W., D'amico, V., Luzi, L., Martinelli, F., Pace, B., ... & Visini, F., 2019. MPS19: the updated Italian Seismic Hazard model, *Geophys. Res. Abstr.* (Vol. 21).

Moschetti, M.P., 2015. A long-term earthquake rate model for the central and eastern United States from smoothed seismicity, *Bull. Seismol. Soc. Am.,* 105(6), 2928–2941.

Murru, M., Montuori, C., Wyss, M. & Privitera, E., 1999. The locations of magma chambers at Mt. Etna, Italy, mapped by b-values, *Geophys. Res. Lett*., 26(16), 2553-2556.

Ogata, Y., 1998. Space-time point-process models for earthquake occurrences, *Ann. I. Stat. Math.*, *50*(2), 379-402.

Petersen, M.D., Moschetti, M.P., Powers, P., Mueller, C.S., Haller, K.M., Frankel, A.D., Zeng, Y., Rezaeian, S., Harmsen, S.C., Boyd, O.S., et al. 2014. Documentation for the 2014 update of the National Seismic Hazard Maps, *U.S. Geol. Surv.*, Open-File Rept. 2014-1091.

Reasenberg, P., 1985. Second-order moment of central california seismicity, 1969–1982, *J. Geophys. Res. Solid Earth*, 90(B7), 5479–5495.

Rovida, A., Locati, M., Camassi, R., Lolli, B. & Gasperini, P., 2020. The Italian earthquake catalogue CPTI15, *B. Earthq. Eng*, 1-32.

Serfling, R.J., 1975. A general Poisson approximation theorem, *Ann. Probab*., 3, 726-731.

Teng, G. & Baker, J.W. 2019. Seismicity Declustering and Hazard Analysis of the Oklahoma–Kansas Region, *Bull. Seismol. Soc. Am.*, 109(6), 2356–2366.

Weichert, D.H., 1980. Estimation of the earthquake recurrence parameters for unequal observation periods for different magnitudes, *Bull. Seismol. Soc. Am.*, 70(4), 1337-1346.

Wiemer, S., 2001. A software package to analyze seismicity: ZMAP, *Seismol. Res. Lett.,*72, 373–382.

Wyss, M., Shimazaki, K. & Wiemer, S., 1997. Mapping active magma chambers by *b* values beneath the off-Ito volcano, Japan, *J. Geophys. Res.*, 102, 20413–20422.

Zhuang, J., Ogata, Y., & Vere-Jones, D., 2002. Stochastic declustering of space-time earthquake occurrences, *J. Am. Stat. Assoc.*, 97(458), 369-380.

Table 1 Catalog statistics: number of events, *b*-value, its uncertainty σ, computed through the Aki (1965) approach and annual rate.

Magnitude of Completeness	Declustering	Number of Events	*b*-value	σ (*b*-value)	Annual Rate
3.0	Complete	7283	1.00	0.01	200.45
	GK74	2012	0.77	0.02	55.38
	REAS85	5144	1.01	0.01	141.58
4.0	Complete	734	1.06	0.04	20.20
	GK74	369	1.00	0.05	10.16
	REAS85	521	1.11	0.05	14.34
4.5	Complete	203	1.05	0.07	5.59
	GK74	113	1.08	0.10	3.11
	REAS85	138	1.12	0.10	3.80

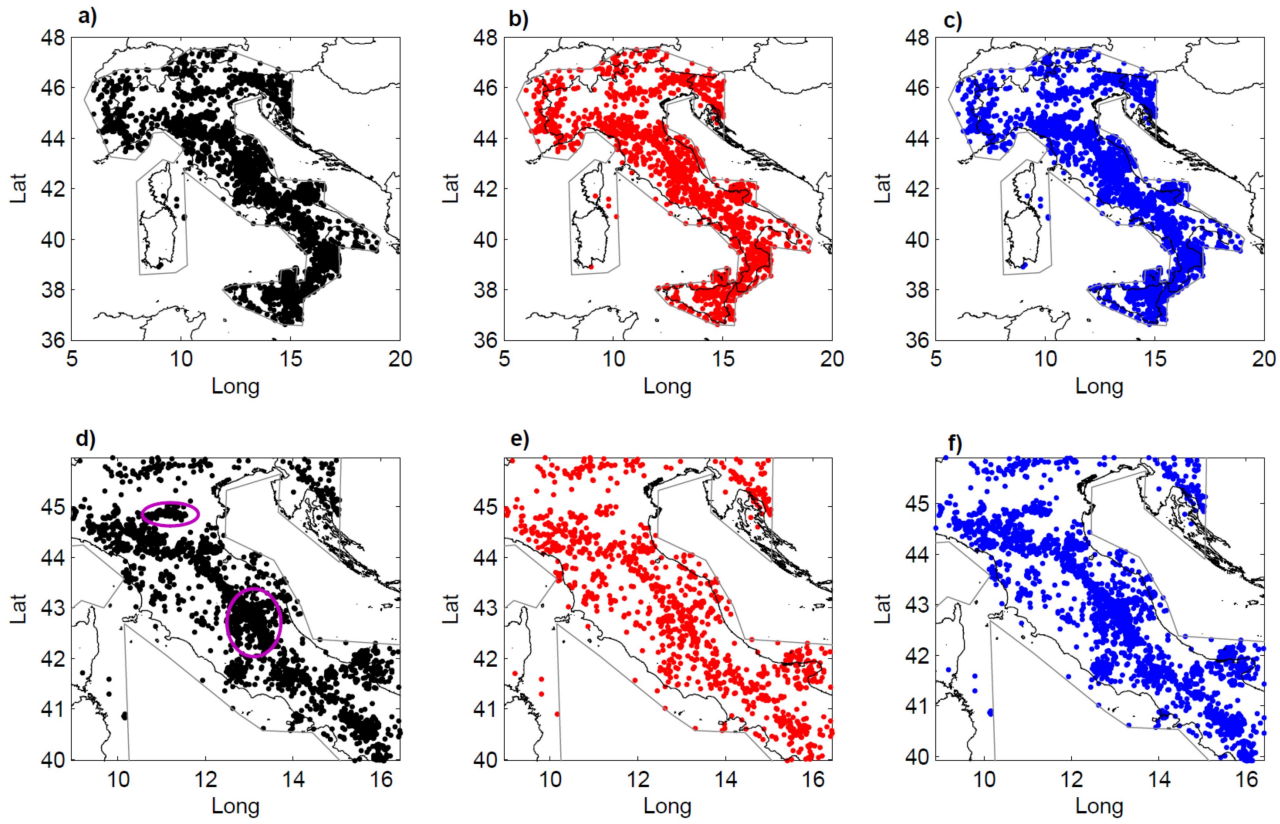

Figure 1 Location of events considered in the a) complete seismic catalog for the whole region of interest, b) declustred catalog using GK74, c) REAS85 declustering algorithm. Panel d), e) and f) are zoomed in the central part of Italy for a), b) and c) respectively. Purple ellipses in panel d) show the location of the major seismic sequences.

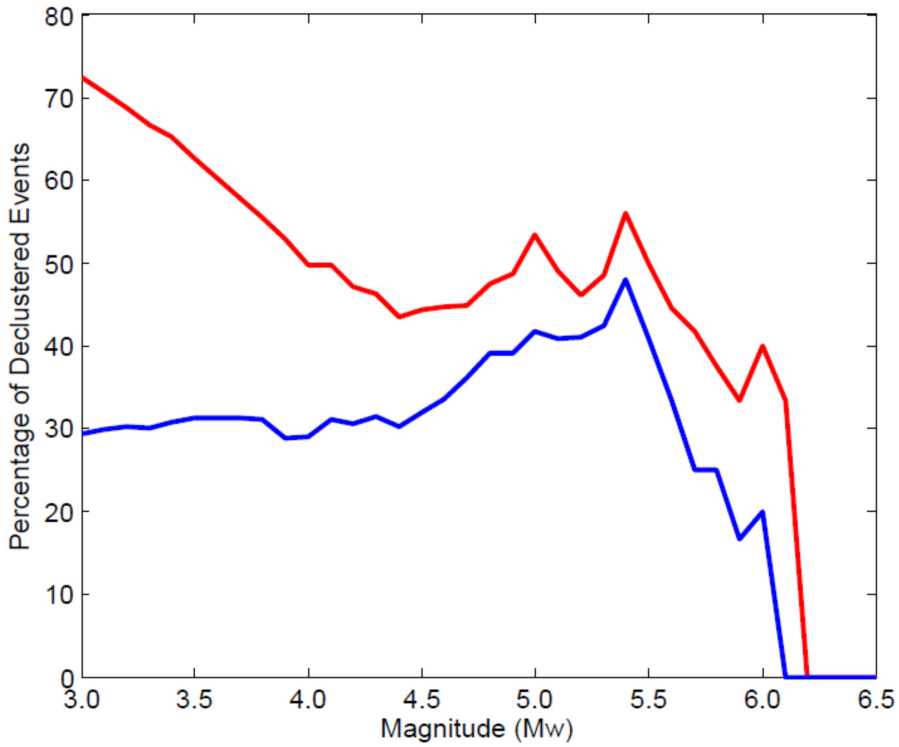

Figure 2 Percentage of declustered events as a function of magnitude estimated using GK74 (red curve) and REAS85 (blue curve) algorithm.

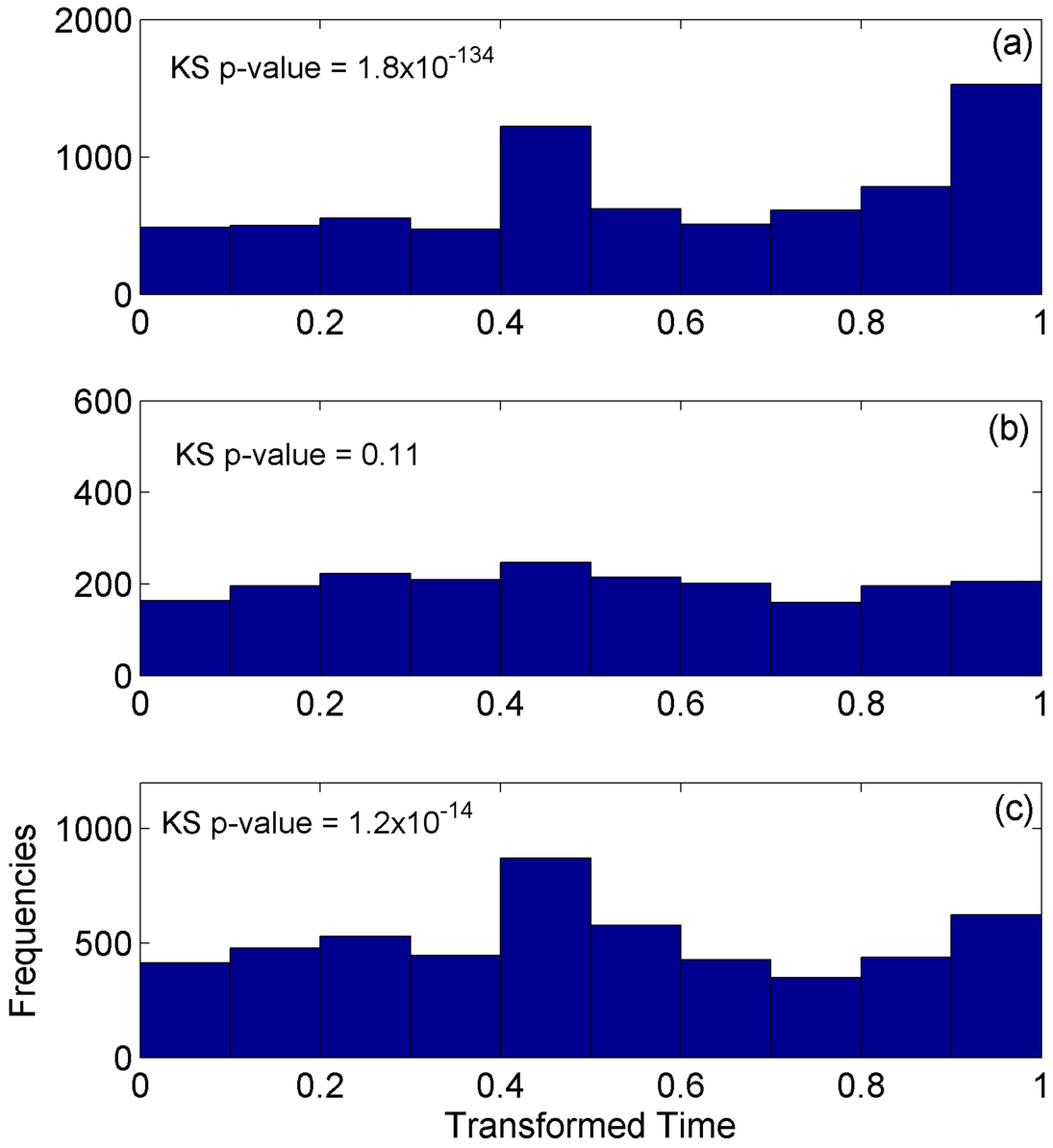

Figure 3 The number of earthquakes as function of transformed time (occurring time of the events divided by the time length of the catalog: this time span from 0 to 1) for a) the complete seismicity catalog; b) declustred catalog from the GK74 and c) declustred catalog from the REAS85 algorithm and the p-values of the Kolmogorov-Smirnov tests for the temporally homogeneous Poisson hypothesis of each catalog.

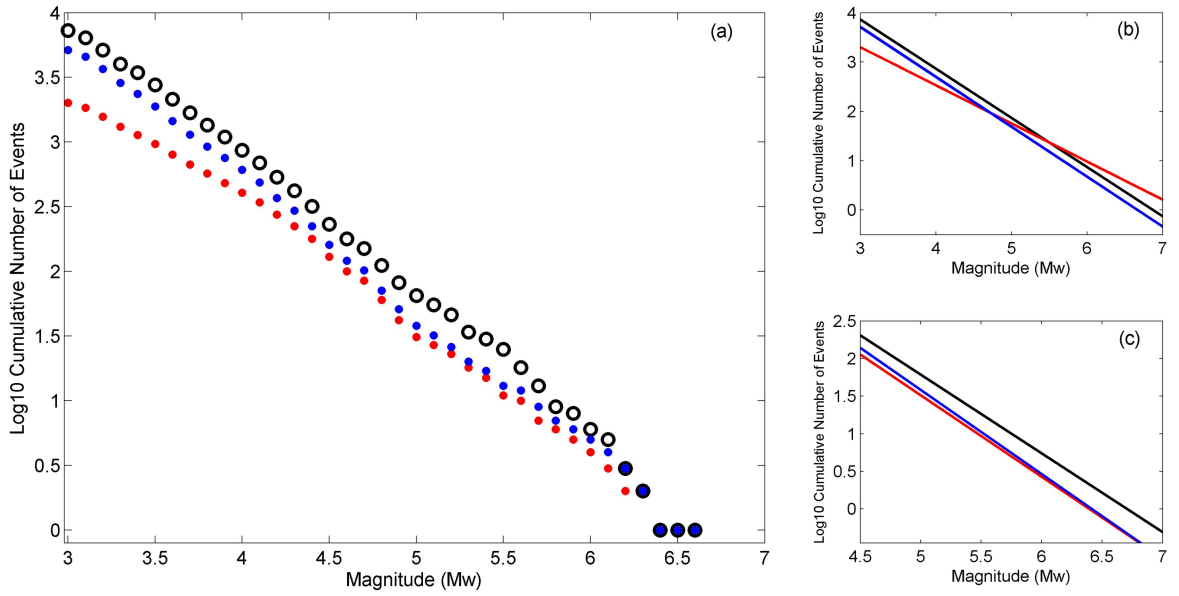

Figure 4 Magnitude frequency distribution of the events. (a) Cumulative number of events, in a Log10 scale, as a function of the magnitude. Black circles for the complete catalog, blue dots for the REAS85 declustered catalog and red dots for the GK74 declustered catalog. (b) Gutenberg-Richter estimated from M_w 3.0, black line for the complete catalog, blue line for REAS85 and red line for GK74. (c) As in (b), but from M_w 4.5.

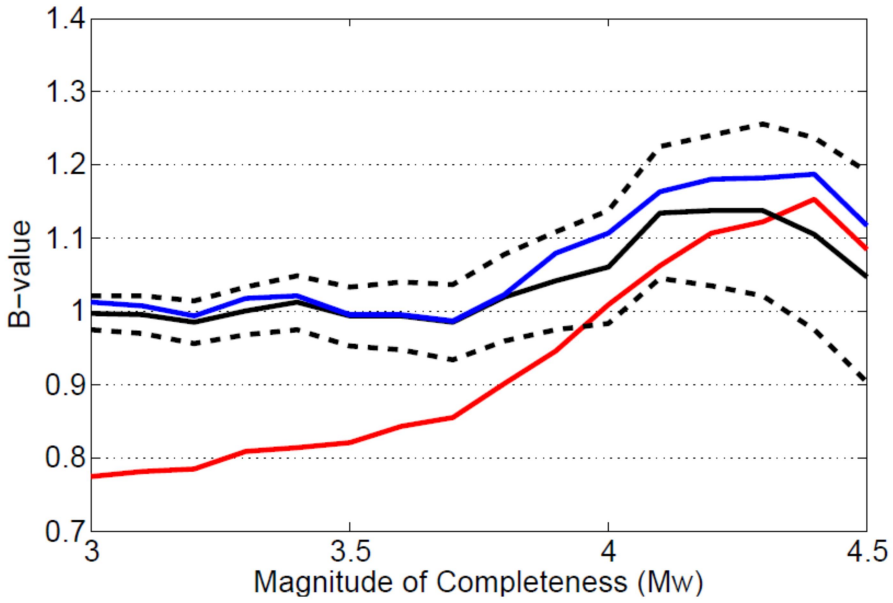

Figure 5 Estimated *b*-value of the Gutenberg-Richter law as a function of the magnitude of completeness of the catalog. Black solid line for the complete catalog (dashed black lines represent the 95% confidence interval of the estimation), blue line for the REAS85 declustered catalog and red line for the GK74 declustered catalog.

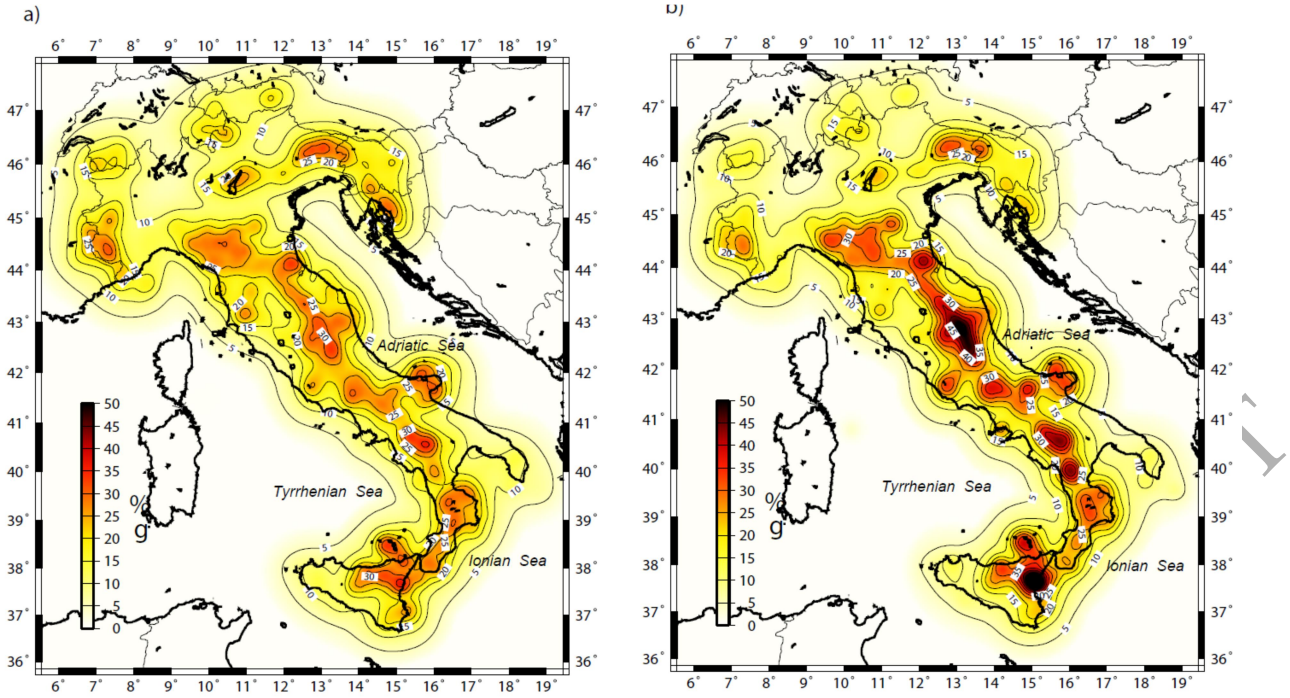

Figure 6 Expected peak ground acceleration (PGA) for 10% of probability of exceedance in 50 years using ITA10 GMPE and obtained from the instrumental catalogs declustered using a) the GK74 and (b) the REAS85 procedure.

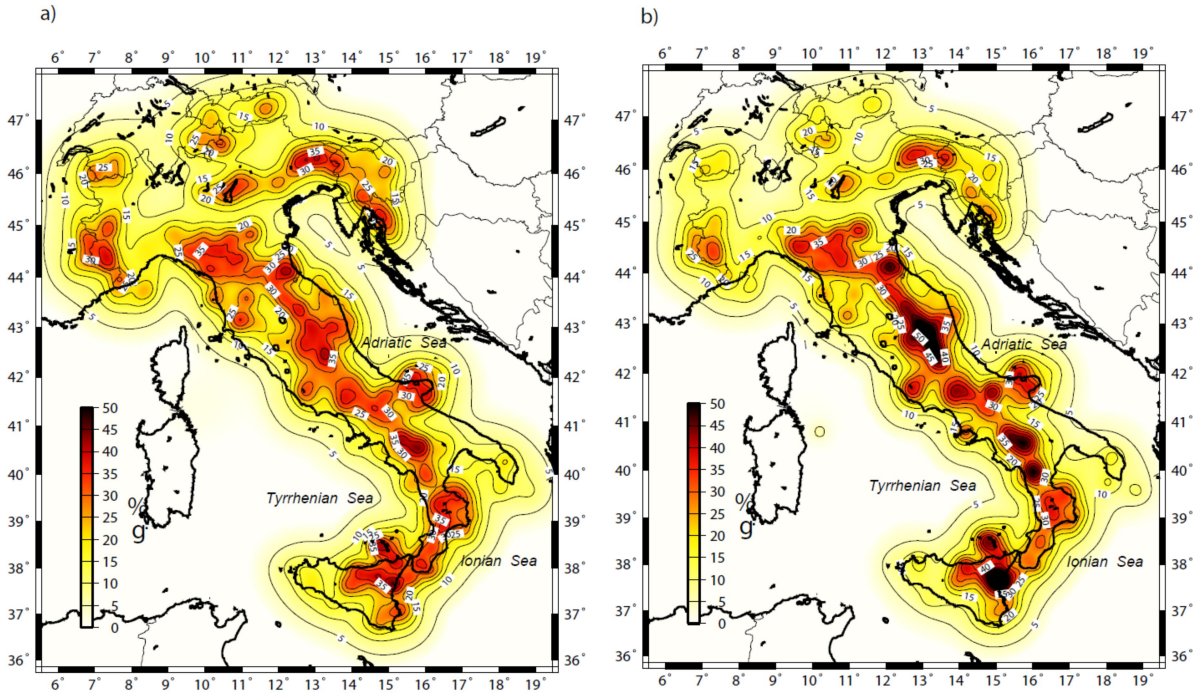

Figure 7 Expected peak ground acceleration (PGA) for 10% of probability of exceedance in 50 years using ITA10 GMPE and obtained from the corrected model a) GK74, b) REAS85.

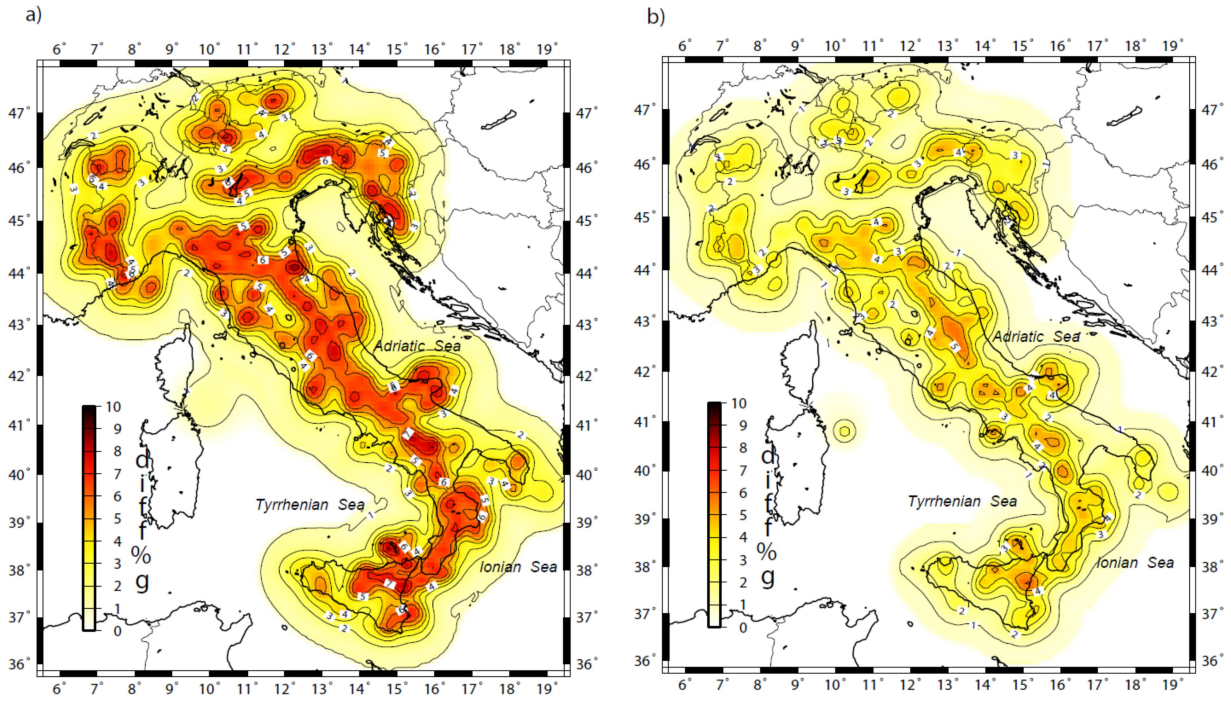

Figure 8 Differences in PGA value with exceeding probability of 10% in 50 years calculated with the corrected seismicity rates and the declustered seismicity rates using the a) GK74 and b) REAS85 algorithms, ($PSHA_{COR}$-$PSHA_{DECL}$).

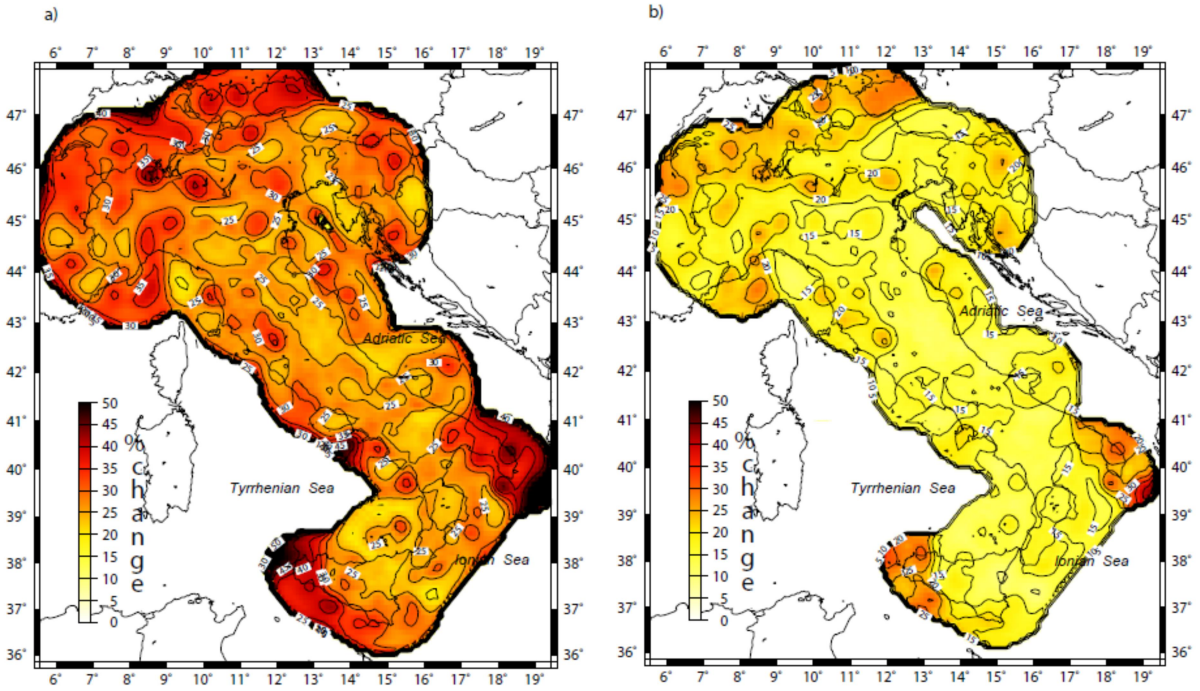

Figure 9 Relative percentage difference in PGA value with exceeding probability of 10% in 50 years calculated with the corrected seismicity rates and the declustered seismicity rates using the a) GK74 and b) REAS85 algorithms, $((PSHA_{COR}-PSHA_{DECL})/PSHA_{DECL}*100,)$.

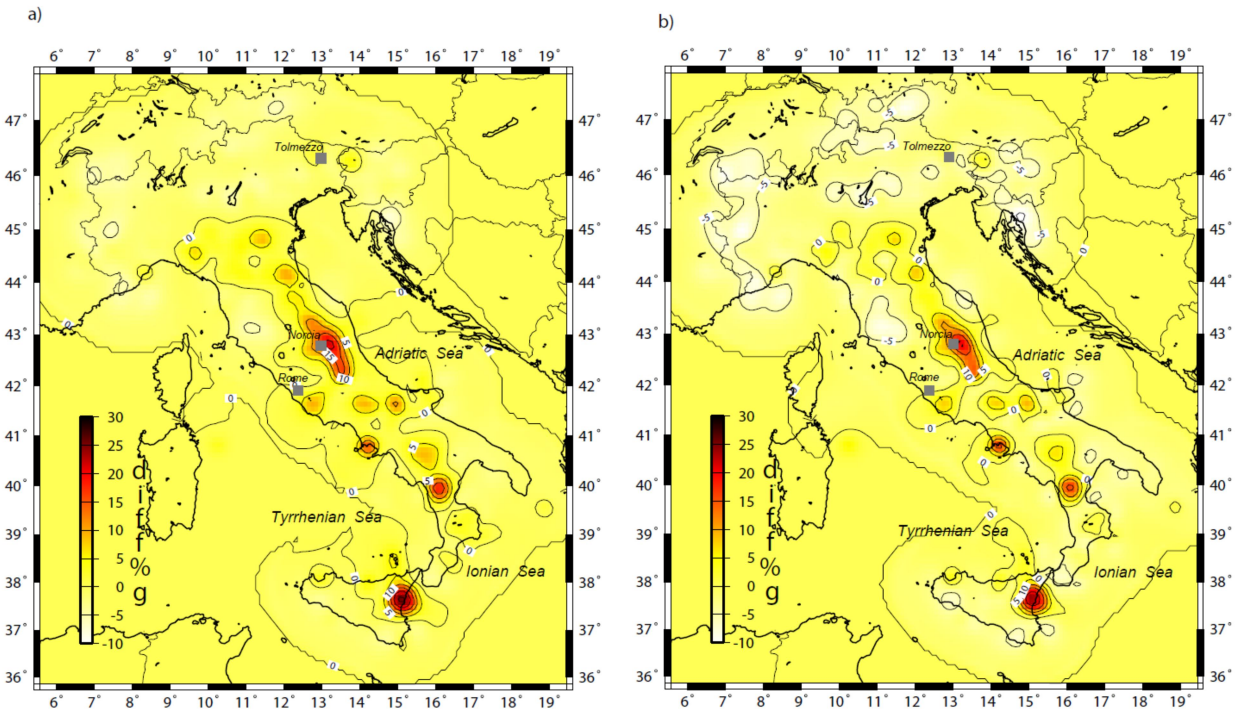

Figure 10 Differences in PGA value with exceeding probability of 10% in 50 years calculated with the declustered or the corrected seismicity rates for the declustered GK74 and the REAS85 catalogs; a) from declustered catalogs ($PSHA_{DECL\_REAS85} - PSHA_{DECL\_GK74}$)), b) from complete catalogs ($PSHA_{COR\_REAS85} - PSHA_{COR\_GK74}$)). Gray squares indicate sites where hazard curves are calculated.

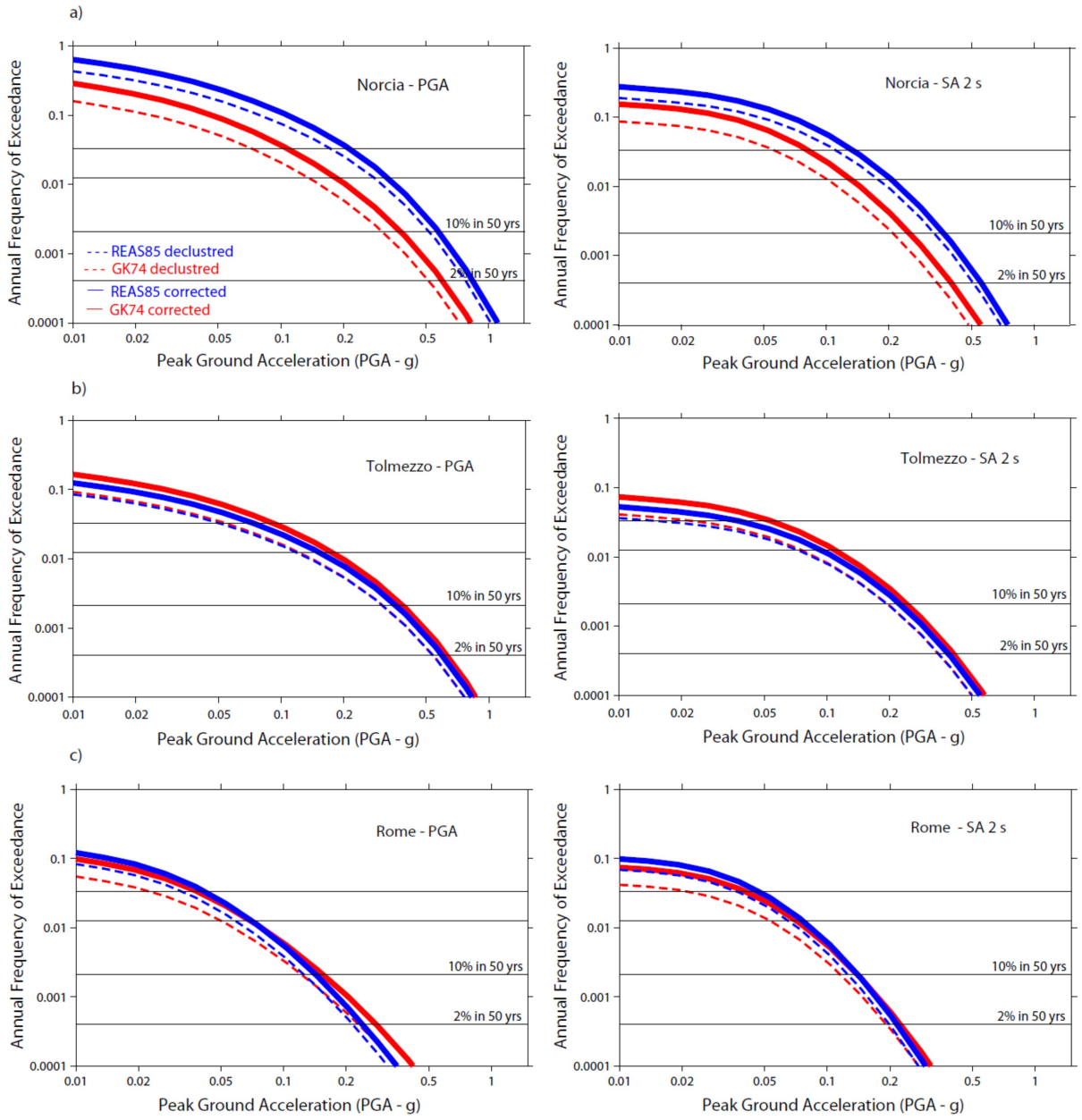

Figure 11 Hazard Curves in terms of PGA (left panels) and PSA (2 s, right panels) from the corrected (continuous lines) and declustered (dashed lines) GK74 (red) and REAS85 (blue) catalogs for three selected sites as shown in Figure 10 with grey squares a) Norcia town, in the central Apennines (42.80N, 13.10E); b) Tolmezzo town (in the Friuli region, north-east of Italy; 46.35N, 13.05E); and c) the city of Rome (41.95N, 12.35E).

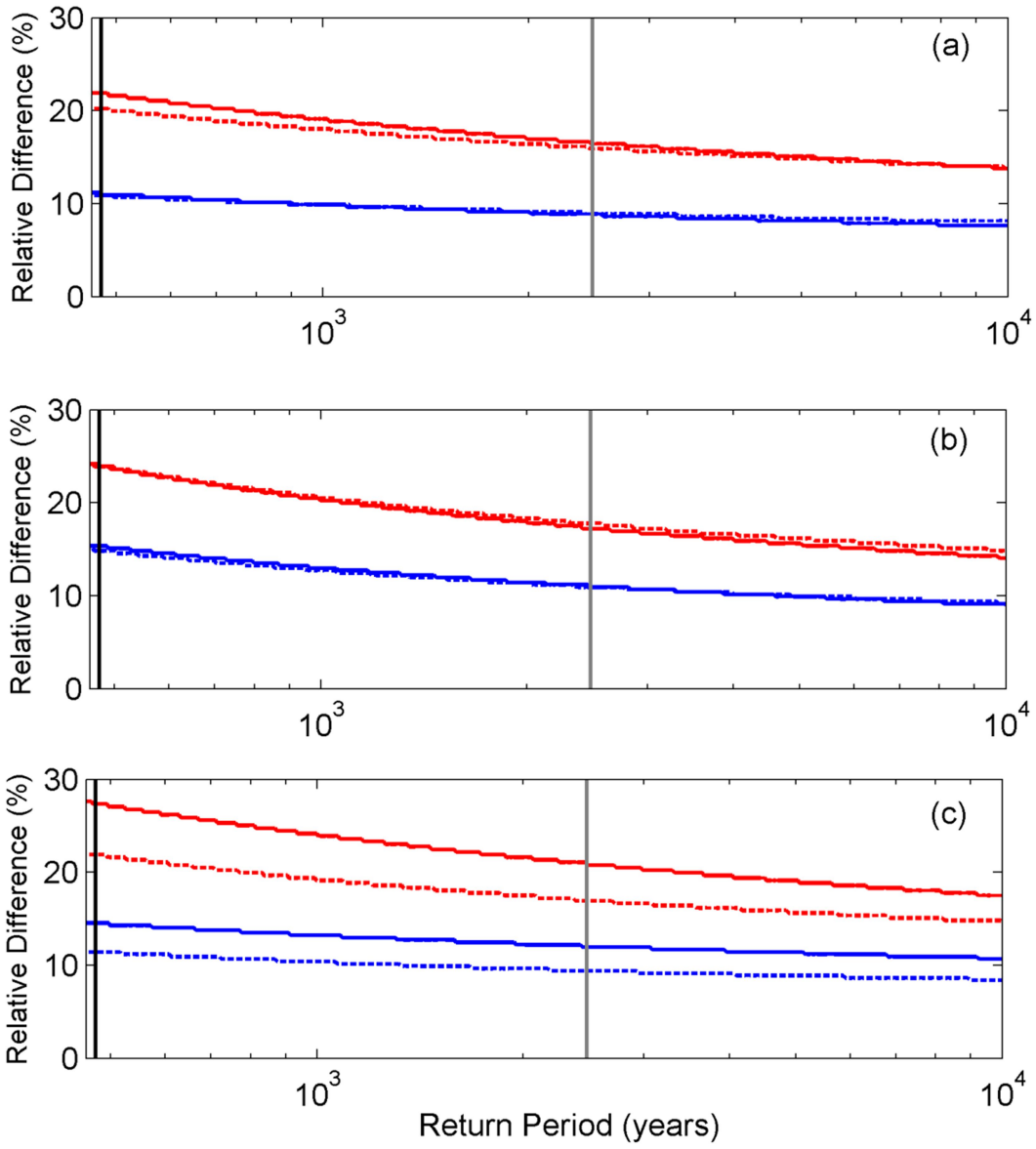

Figure 12 Relative difference (%) between hazard curves obtained from declustered GK74 and REAS85 (red and blue color, respectively) and complete catalogs for PGA (continuous lines) and SA (2 s, dashed lines) as a function of return periods, for a) Norcia (42.80N, 13.10E); b) Tolmezzo (in the Friuli region; 46.35N, 13.05E) and c) Rome (41.95N, 12.35E). Black vertical line represents the return period at 475-year (10% in 50 years), the grey vertical line presents the return period at 2475-year (2% in 50 years).

Appendix A

In this Appendix, we address the problem of variation in *b*-value for the declustered catalogs. Some studies assert that the *b*-value of the main shock is different from the *b*-value of the complete catalog for some physical reasons (e.g. the characteristic earthquake, Davison and Scholz 1985). In contrast, other studies demonstrated that the difference between the two *b*-values can be fully explained by statistical considerations (Lombardi, 2003), that is, the declustering algorithm artificially select the strongest events in each sequence.

We outline that our method is not based on the classical Gutenberg-Richter law: it is also possible to use different parametrizations of the magnitude-frequency distribution (e.g. Urban, 2016), or the tapered version of the Gutenberg-Richter law (Kagan, 2002), without breaking the assumptions of the method.

Here, we investigated in more detail, the effect of the declustering on the estimation of *b*-value. We simulated a synthetic seismic catalog with the ETAS model (Ogata, 1998) using the computer code of Felzer et al. (2002). We built 100 years long synthetic catalog using as magnitude-frequency distribution a Gutenber-Richter law, fixing the *b*-value equal to 1. To reproduce Fig. 5, we declustered the synthetic seismic catalog, both with the Gardner and Knopoff (1974) and the Reasenberg (1985) algorithm. We then estimated the *b*-value as function of the completeness magnitude, in order to compare the results of real and synthetic catalogs (the synthetic catalog is complete from M_w 3.0).

The pattern of the curves in Figure 1A is very similar to the curve of Figure 5: the *b*-value of the REAS85 declustered synthetic catalog is very similar to the *b*-value of the complete catalog, while the *b*-value of the GK74 declustered synthetic catalog is much smaller than the *b*-value of

the complete catalog. The *b*-value of the GK74 declustered synthetic catalog increased with increased magnitude of completeness, as occurred in the real case.

For the synthetic catalog, our results demonstrated that the GK74 declustering algorithm led to a biased estimation of the *b*-value, particularly for smaller magnitude of completeness.

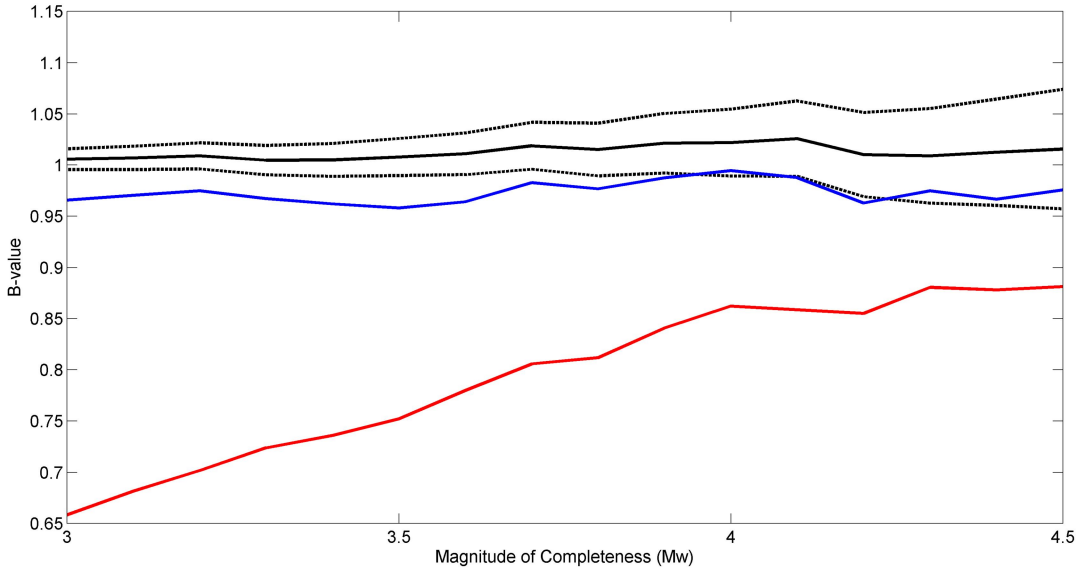

Figure 1A: estimated *b*-value of the Gutenberg-Richter law, as a function of the magnitude of completeness of the catalog. Black solid line for the complete synthetic catalog (dashed black lines represent the 95% confidence interval of the estimation), blue and red lines for the declustered synthetic catalogs from the REAS85 and the GK74 algorithms.

Appendix B

In this Appendix B, we showed an example of the application of our method to the historical events using the parametric catalog of Italian earthquakes (Catalogo Parametrico dei Terremoti Italiani, CPTI15, Rovida et al. 2020). The CPTI15 catalog lasts more than one thousand years (1000-2017), and contains 4720 events with magnitudes up to 7.3

We followed the same procedure as described in the main text; we first declustered the catalog using both the Gardner and Knopoff (1974) and the Reasenberg (1985) algorithms, and then selected only those events above the completeness period of time as given by Meletti et al. (2019). The cumulative annual rates are shown In Fig. 1B, while the estimation of the *b*-value and the annual rate using the Weichert (1980) approach are summarized in Table 1B. Similar to the instrumental catalog the annual rate was smaller for the declustered catalog, in particular for the GK74 declustered catalog. The *b*-value of the GK74 declustered catalog was smaller than the *b*-value of the complete catalog (0.98 and 1.05, respectively).

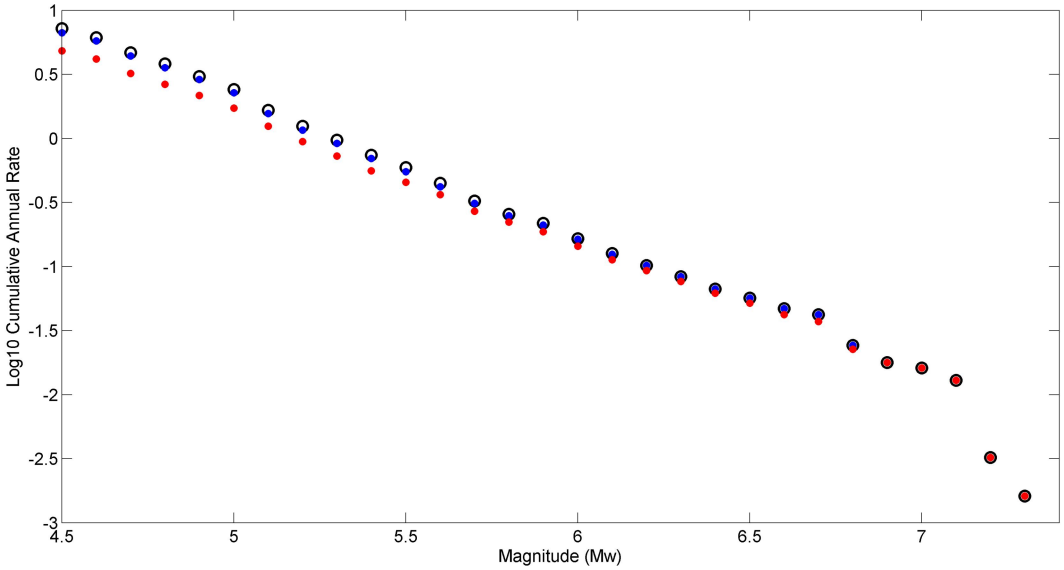

Figure 1B: Magnitude-frequency distribution of the events in the historical catalog, CPTI15; cumulative number of events (in a Log10 scale) as a function of the magnitude. Black, blue and red circles represent cumulative number of events (in a Log10 scale) as a function of the magnitude for the complete catalog, the REAS85 and the GK74 declustered catalogs, respectively.

Table 1B: estimated *b*-values and annual rates for the complete and declustered historical catalog.

Catalog	CPTI15 Complete	CPTI15 Declustered REAS85	CPTI15 Declustered GK74
***b*-value**	1.05	1.04	0.98
Annual Rate	7.15	6.63	4.78