Structural Safety 99 (2022) 102264

Contents lists available at ScienceDirect

Structural Safety

journal homepage: www.elsevier.com/locate/strusafe

Asymptotic behavior of seismic hazard curves

Iunio Iervolino

Dipartimento di Strutture per l'Ingegneria e l'Architettura, Università degli Studi di Napoli Federico II, via Claudio 21, 80125 Naples, Italy

ARTICLE INFO

Keywords:
Extreme value distributions
Code calibration
Poisson process
Maximum seismic intensity
Seismic design
Probabilistic seismic hazard analysis

ABSTRACT

Hazard curves from probabilistic seismic hazard analysis (PSHA) are plots of the rate of earthquakes exceeding ground motion intensity values vs such threshold values, for a site of interest. In classical PSHA, these curves can be transformed to provide the probability of exceedance of ground motion intensity values in any time interval, utilizing the properties of the homogeneous Poisson process (HPP). In turn, these probability curves can be seen as the plot of the complementary cumulative distribution function of the maximum intensity observed, at the site, in the time interval of interest. One consequence of the HPP framework, within which PSHA is developed, is that, for large time intervals, it can be argued that these curves could asymptotically lead to a probabilistic model for extreme value (EV) random variables. This is discussed, with a simple engineering approach, in this short note, where it is found – via case studies – that exceedance hazard curves seem to converge towards an EV distribution (i.e., the EV type II or Fréchet), with a pace that is impacted by the discontinuity inherent to the curves. It is also seen that other common models, typically used to provide an analytical format to probabilistic curves, do not show the same level of convergence. Besides providing further insights on the results of PSHA, this study can possibly be useful for those cases where a closed-form equation for the hazard curve could be needed, such as reliability-based calibration of building codes, or seismic risk studies involving seismic hazard approximation/extrapolation.

1. Introduction

Probabilistic seismic hazard analysis (PSHA) [6,16] enables deriving the rate (typically annual) of earthquakes causing exceedance of a ground motion intensity measure threshold at a site of interest. Calculating these rates for a wide range of intensity thresholds leads to building a function whose plot is referred to as the (rate) *hazard curve* for the site. In classical PSHA, it can be assumed that the earthquakes causing exceedance at the site, of any intensity threshold, follow a homogeneous Poisson process (HPP). This enables one to simply compute the probability of exceedance of the threshold in any time interval. Mapping the probability of exceedance for to all the thresholds for which the hazard curve is developed, allows us to transform the latter into a curve representing the complementary cumulative distribution function (CCDF) of the maximum intensity observed in the considered interval. This is a distribution of interest to structural engineering, as it can be related to the structural seismic loads. However, because the hazard integral (to follow) is not typically amenable to a closed form expression, neither is the CCDF of the maximum intensity. Given the nature of the investigated phenomenon, and the inherent characteristics of PSHA, it can be argued whether, for large enough time intervals, so that the expected number of seismic events is large, the exceedance hazard curve should converge to a model for extreme-value (EV) random variables [12].

The motivation in investigating whether the distribution of the maximum ground motion intensity can be represented by some classical model is threefold: (i) to gather some further insights into the nature of PSHA results; (ii) because reliability-based calibration of building codes follows a semi-probabilistic approach, which benefits from stochastic modelling of uncertainty in the structural loads (e.g., [8]); (iii) in those cases of engineering relevance where approximation/extrapolation of hazard curves is employed.

To investigate if the EV distributions can be seen as the asymptote of the hazard curves is the scope of the discussion and the related investigations presented herein. Although the theory of EV has been already applied to seismic hazard assessment, it was more often used to calibrate some input model to PSHA (e.g., [20,24,25]), to propose an alternative approach to carry out probabilistic seismic hazard analysis (e.g., [23]), or to approximate its results (e.g., [11]), whereas the subject herein is specifically the relationship between the hazard curves computed according classical PSHA and the EV distributions.

It should be noted that Cornell [6] arrived at a distribution similar to

E-mail address: iunio.iervolino@unina.it.

https://doi.org/10.1016/j.strusafe.2022.102264
Received 4 May 2022; Received in revised form 24 July 2022; Accepted 4 August 2022
Available online 14 August 2022

an EV model for the hazard curve; however, this result was not general and not based on the extreme value theory, but rather because some of the assumptions taken in the applications of PSHA he developed. Reasoning about the EV theory applied to the earthquake amplitude can be found in another seminal paper from Milne and Davenport [18], dealing with the first seismic hazard assessment in Canada.

The present work considers two case study sites in Italy, featuring low and high seismic hazard, according to the country's standards. For each site the annual rate hazard curve is computed, and the probability density function (PDF) of the ground motion intensity given the occurrence of one *generic* earthquake (to follow) is derived, and its features are discussed. This is relevant, as this PDF represents the *parent* distribution for an EV model. Then the exceedance hazard curves for different time interval widths are fitted by type I and type II EV distributions. To evaluate the eventual convergence, some other common random variable (RV) models in civil engineering applications, such as the lognormal and the gamma ditributions, are also used as fitting candidates.

The remainder of this note is structured by first revisiting the basics of PSHA relevant for this study, that is, the hazard rate curve, the distribution of the ground motion intensity in one event, and the exceedance probability curves. Then, the PSHA for the two case study sites is presented. Subsequently, the interpretation of the exceedance hazard curves, and the shared features with the EV distributions, are discussed, along with other probabilistic models considered for the fitting. All the models are critically applied to the case study sites, discussing possible convergence with respect to the interval width. Final remarks close the paper.

2. Hazard rate curves, exceedance probability curves, and distribution of intensity in one event

2.1. Hazard rate curves

Given a ground motion intensity measure (IM) and a threshold (x), the expected number of earthquakes causing $IM > x$ at a site of interest in a unit time, that is, the exceedance rate ($\lambda_{IM>x}$), is computed in the classical approach to PSHA as:

$$\lambda_{IM>x} = \sum_{i=1}^{s} \nu_i \cdot \int_{m_{\min,i}}^{m_{\max,i}} \int_{r_{\min,i}}^{r_{\max,i}} P[IM > x|y,z]_i \cdot f_{M,R,i}(y,z) \cdot dy \cdot dz. \quad (1)$$

In the equation, commonly referred to as the *hazard integral*, which is written explicitly considering that multiple seismic sources, s, affect the site of interest, ν_i is the rate of earthquakes above a minimum magnitude (M) of interest, indicated as $m_{\min,i}$ for the i-th source. The $f_{M,R,i}$ term is the joint PDF of magnitude and source-to-site distance (R) RVs, for source i. Magnitude is usually bounded by the maximum magnitude considered possible for the source $(m_{\max,i})$, while the distance varies in the $(r_{\min,i}, r_{\max,i})$ range. Given the source, M and R are often considered stochastically independent, thus $f_{M,R,i}(y,z) = f_{M,i}(y) \cdot f_{R,i}(z)$, where $f_{M,i}(y)$, the magnitude distribution of the i-th source, is often a truncated exponential distribution following the Gutenberg-Richter model [13], while $f_{R,i}(z)$ is the distance distribution depending on the source geometry and site's location with respect to it. The $P[IM > x|y,z]_i$ term, provided by a ground motion prediction equation (GMPE), is the probability that $IM > x$, conditional on $M = y$ and $R = z$. GMPEs usually model the PDF of IM, conditional on $\{M = y, R = z\}$, $f_{IM|M=y,R=z}$, as a lognormal RV via the following equation:

$$\log(IM) = \mu_{\log(IM)}(y,z,\underline{\theta}) + \sigma_{\log(IM)} \cdot \varepsilon, \quad (2)$$

where $\mu_{\log(IM)}(y,z,\underline{\theta})$ is the mean of the logarithms of IM at the site for an earthquake with $M = y$ and $R = z$, $\underline{\theta}$ represents a vector of additional covariates, usually not treated as RVs (e.g., local soil site conditions), and $\sigma_{\log(IM)} \cdot \varepsilon$ is a zero mean and $\sigma^2_{\log(IM)}$ variance Gaussian RV. In fact, ε (*epsilon*) is also referred to as the *standardized residual*, as it measures the number of standard deviations that the logarithm of IM is away from its mean, conditional on $\{M = y, R = z\}$. It follows that the conditional distribution of IM, given magnitude and distance, is lognormal.

If Eq. (1) is (ideally) computed for all possible thresholds, then it is possible to map the exceedance rates against the corresponding thresholds, resulting in a function widely used in seismic risk assessment and the diagram of which is commonly referred to as the hazard curve. Unfortunately, because the hazard integral seldom yields closed form solutions, the hazard curves cannot generally be represented by simple equations (although, for some purposes, they can be locally approximated by simple relationships such as log-linear; e.g., [7]).

Finally, note that PSHA is often carried out via a so-called *logic tree* (e.g., [16]), where multiple alternative models are considered for each of the terms of the hazard integral. In this case, the discussion given in the following can at least be applied to each of the logic tree branches.

2.2. Distribution of the intensity in a generic event

Equation (1) can be always written in a more compact form:

$$\lambda_{IM>x} = \nu \cdot P[IM > x|E], \quad (3)$$

from which the hazard integral actually derives by an application of the total probability theorem. In the equation $P[IM > x|E]$ is the probability that $IM > x$ given that a generic earthquake (E) occurs on one of the seismic sources considered, that is, a seismic event of unspecified magnitude and location, and $\nu = \sum_{i=1}^{s} \nu_i$, is the total rate of earthquakes from the considered seismic sources. Equation (3) shows that the value to which the curve must approach, when the threshold goes to zero, is the total source rate:

$$\lim_{x \to 0} \lambda_{IM>x} = \nu. \quad (4)$$

It immediately follows from Eq. (3) that the hazard curve is closely related to the distribution of the ground motion intensity at the site, $f_{IM|E}$, given the occurrence of one generic earthquake. Such a PDF is just the derivative of the hazard curve divided by the rate of considered earthquakes:

$$\begin{cases} f_{IM|E}(x) = 0, & x \leqslant 0 \\ f_{IM|E}(x) = \dfrac{1}{\nu} \cdot \left| \dfrac{d(\lambda_{IM>x})}{dx} \right|, & x > 0 \end{cases}. \quad (5)$$

(The equation assumes that the ground motion intensity measure considered is non-negative, which is typical.) Note that this distribution does not generally yield a known probabilistic model for the same reasons the hazard integral doesn't. In fact, Eq. (1) shows that the distribution of the intensity in one event is a mixture of lognormal distributions with a truncated exponential in most cases (i.e., the magnitude distribution) and a distribution, that is, the distance distribution, whose shape cannot generally be anticipated, being dependent on the source geometry and the position of the site with respect to it. Nevertheless, it follows from the working hypotheses, according to which the hazard integral is developed, that the intensity RVs in different events can be considered independent and identically distributed (iid) according to Eq. (5).

2.3. Exceedance hazard curves

Although earthquakes generally occur in space time-clusters, the rates ν_i are usually obtained from a declustered seismic catalog (e.g., [9]), which only retains the mainshock (i.e., the largest magnitude event) among the earthquakes of each identified cluster (see also [14], for a discussion). Consequently, it can be assumed that the arrival of earthquakes considered in PSHA follows a HPP whose parameter is ν. Then, the rate $\lambda_{IM>x}$ can be used to obtain the probability that $IM > x$, in

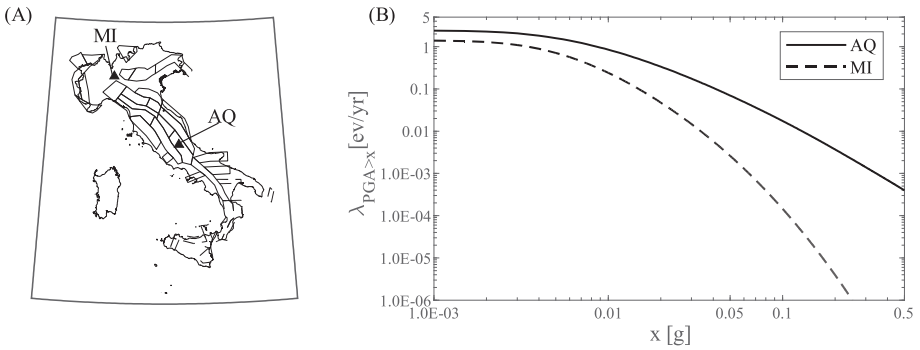

Fig. 1. Location of the considered sites and seismic source zones (A) and PGA (on rock) hazard curves (B).

any time interval $(t, t+\Delta t)$ of interest, via the following equation:

$$P[IM(t,t+\Delta t)>x]=P[IM(\Delta t)>x]=1-e^{-\lambda_{IM>x}\cdot\Delta t}. \tag{6}$$

This very well-known result can be derived from Eq. (3) considering that it is the probability to observe at least one event with $IM>x$ in the interval. Denoting as $N(t,t+\Delta t)$ and $N_{IM>x}(t,t+\Delta t)$ the RVs counting the earthquakes observed at the site (i.e., those mainshocks above the minimum magnitude) and those with $IM>x$, respectively, by virtue of the power series it results:

$$\begin{aligned}P[IM(t,t+\Delta t)>x]&=1-P[N_{IM>x}(t,t+\Delta t)=0]=1-\sum_{n=0}^{+\infty}P[IM\leqslant x|N=n]\cdot P[N(t,t+\Delta t)=n]=\\&=1-\sum_{n=0}^{+\infty}P[IM\leqslant x|E]^n\cdot\frac{(\nu\cdot\Delta t)^n}{n!}\cdot e^{-\nu\cdot\Delta t}=1-e^{-\nu\cdot\Delta t}\cdot e^{(1-P[IM>x|E])\cdot\nu\cdot\Delta t}=1-e^{-\lambda_{IM>x}\cdot\Delta t}.\end{aligned} \tag{7}$$

In fact, Eq. (3) shows that this result could be obtained more quickly by recognizing that the HPP counting the earthquakes causing exceedance is obtained from the HPP counting the mainshocks above the minimum magnitude, via an operation which is often referred to as *thinning* or *filtering* (e.g., [10]). (This also enables specifying that a generic earthquake is a seismic event of unspecified magnitude and location, yet among those considered by the source rates; i.e., above the minimum magnitude and being a mainshock.)

Mapping $P[IM(t,t+\Delta t)>x]$ to the corresponding x values leads to a transformation of the hazard curve which may be referred to as the *exceedance hazard* curve. It can be seen as the CCDF of the maximum intensity, $IM_{\max}$, observed in $(t,t+\Delta t)$ at the site:

$$P[IM(t,t+\Delta t)>x]=1-F_{IM_{\max}(t,t+\Delta t)}(x). \tag{8}$$

This is because observing at least one exceedance of x in $(t,t+\Delta t)$ coincides with the maximum intensity observed in the interval being larger than x (as also noted by [6]). Therefore, $F_{IM_{\max}(t,t+\Delta t)}$ is the cumulative density function of the $IM_{\max}(t,t+\Delta t)$ RV. In fact, recalling that probability distributions are always defined for $x\in(-\infty,+\infty)$, the CCDF in Eq. (8) should more precisely indicated as:

$$P[IM_{\max}(t,t+\Delta t)>x]=\begin{cases}1, & x<0\\ 1-e^{-\lambda_{IM>x}\cdot\Delta t}, & x\geqslant 0\end{cases} \tag{9}$$

It is also important to remark that $F_{IM_{\max}(t,t+\Delta t)}$ is a function of the time interval $(t,t+\Delta t)$ only via its width Δt, which is a consequence of the

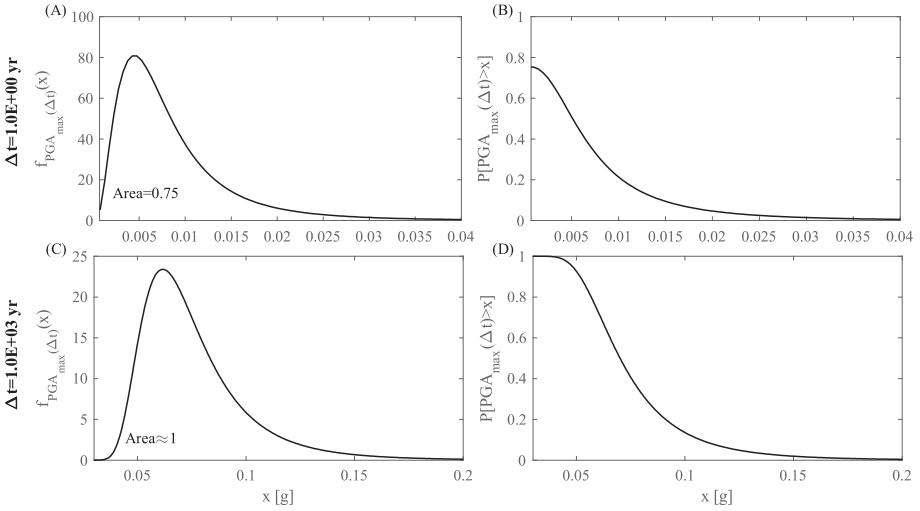

Fig. 2. PDF (left) and CCDF (right) of the maximum PGA (on rock) in Milan for time intervals with 10^0 year (top) and 10^3 years (bottom) widths.

fact that the HPP is *memoryless* and the maximum intensities in non-overlapping intervals of equal widths are iid RVs.

Equation (9) shows that the CCDF of the maximum intensity is discontinuous at $x = 0$, which follows from the fact the probability in Eq. (3), that is $P[IM > x|E]$, is the exceedance probability given the occurrence of one event, while there is a probability that the maximum intensity observed in the $(t, t + \Delta t)$ interval is zero because no earthquakes occur. In fact, the PDF of the maximum intensity is a *hybrid* distribution (e.g., [19]), leading to finite probability for $x = 0$, representing the probability that no earthquakes (mainshocks) occur in $(t, t + \Delta t)$, which is equal to $\exp\left(-\sum_{i=1}^{s} \nu_i \cdot \Delta t\right)$, and the derivative of Eq. (8) for $x > 0$. In other words, the PDF of $IM_{\max}(t, t + \Delta t)$ is:

$$\begin{cases} f_{IM_{\max}(t,t+\Delta t)}(x) = 0, & x < 0 \\ f_{IM_{\max}(t,t+\Delta t)}(x) = \delta(x) \cdot e^{-\sum_{i=1}^{s} \nu_i \cdot \Delta t}, & x = 0 \\ f_{IM_{\max}(t,t+\Delta t)}(x) = \dfrac{d\left(e^{-\lambda_{IM>x} \cdot \Delta t}\right)}{dx}, & x > 0 \end{cases}, \tag{10}$$

where $\delta(x)$ is Dirac's delta function centered at $x = 0$.

3. Hazard case studies

To explore the suitability of the EV distributions to describe the asymptotic behavior of hazard curves, two sites in Italy are considered: Milan (MI) and L'Aquila (AQ), which are characterized by comparatively low- and high-hazard in the country, respectively. More specifically, the peak ground acceleration (PGA) hazard curves, on rock soil conditions, are derived. PSHA is carried out via REASSESS [4], using the source model by Meletti et al. [17], which consists of thirty-six source zones, pictured in Fig. 1A together with the location of the sites (triangles). The sources are characterized in terms of annual rates of earthquakes referring to surface-wave magnitude bins with width equal to 0.3 magnitude units [5]. The considered GMPE is that by Ambraseys et al. [1], which is applicable to the 4.0–7.6 magnitude range, while the source-to-site distance, in terms of the *Joyner & Boore* metric [15] can be up to 200 km. For both sites considered, the minimum magnitude considered in PSHA is equal to 4.15, while the maximum is source-dependent. The dominant rupture mechanism indicated by Meletti et al. [17] for the source is also accounted for in the analyses, as suggested by Bommer et al. [3].

The resulting annual rate PGA hazard curves are given in Fig. 1B, where the difference in the hazard is apparent, with L'Aquila being systematically more seismically hazardous than Milan. The value to which the curves approach when x goes to zero, that is the total rate, is equal to 2.4 events/year in L'Aquila, meaning that more than two mainshock events above the minimum magnitude are expected per year, whereas it is 1.4 events/year in Milan.

Fig. 2 shows the PDF and the CCDF of the maximum PGA distribution for Milan when $\Delta t = 1$ *yr* and $\Delta t = 1{,}000$ *yr*. For simplicity, only the last terms of Eq. (9) and Eq. (10) are represented. The discontinuity is apparent in the value of the CCDF at zero and the area below the continuous part of the PDF, which is not equal to one for $\Delta t = 1$ *yr*. However, as it is also evident from the equations, the inherent discontinuity tends to disappear as $\Delta t \to \infty$, because the probability of not observing any events tends to zero, as the figure shows for $\Delta t = 1{,}000$ *yr*.

4. Extreme value distribution and other candidate fitting models

The EV distribution models are derived for the maximum among n independent and identically distributed RVs, when $n \to +\infty$ (e.g., [2]). These distributions represent the uncertainty in the individual observations, with the largest value among them being the maximum. The distribution, which the individual observations share, can be referred to as the parent distribution.

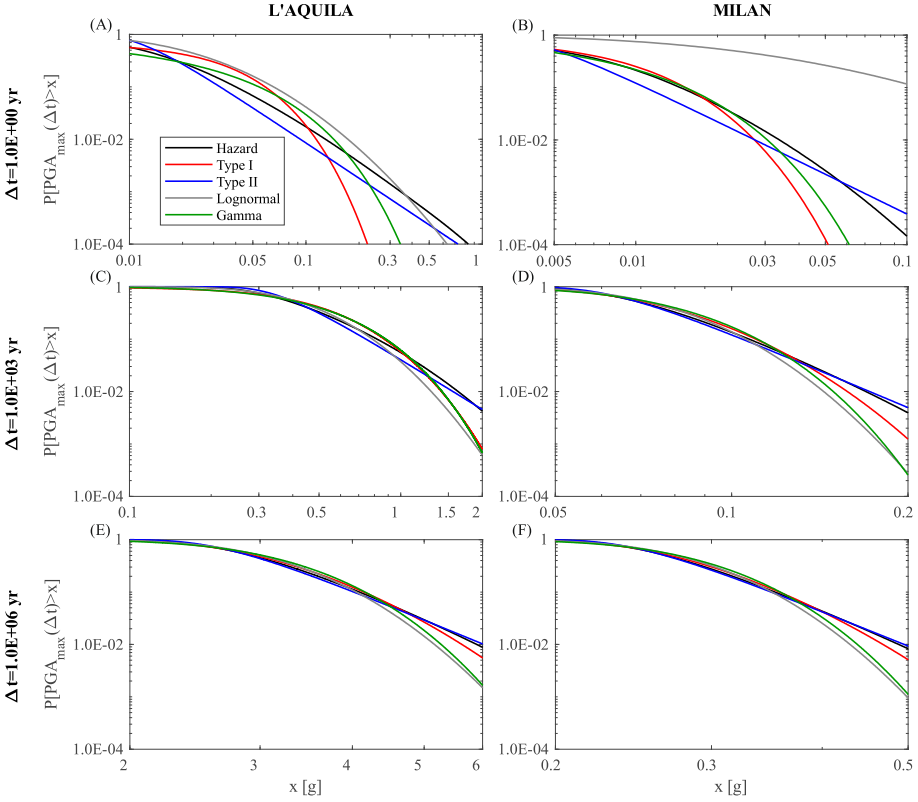

Fig. 3. Probability of exceedance curves for PGA on rock (i.e., maximum *IM* CCDFs) for L'Aquila (left) and Milan (right) referring to three interval widths: 10^0, 10^3, 10^6 years. The plot also shows some probabilistic models fitted to the curve.

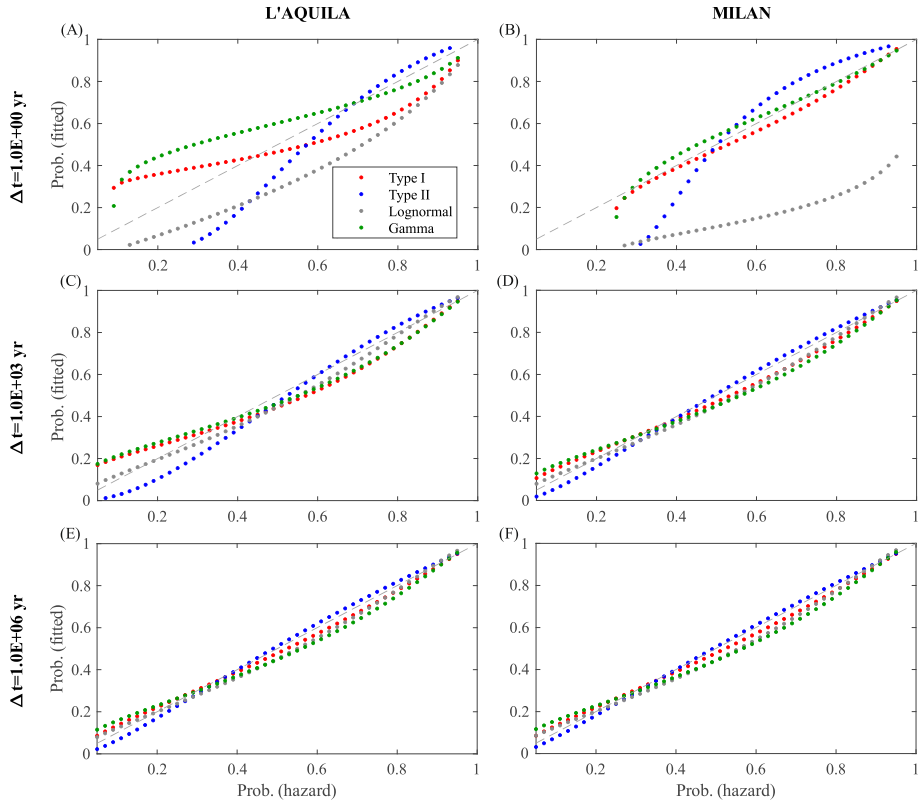

Fig. 4. P—P plots for the curves of Fig. 3 (fitted models vs hazard curves).

In the seismic hazard case, the maximum intensity in a time interval $(t, t+\Delta t)$ is the largest among those observed in the earthquakes occurring within the interval, if any. In mathematical terms, if these seismic events are n in number and IM_i, $i = 1, 2, \ldots n$, is each one's intensity, then:

$$IM_{\max}(t, t+\Delta t) = \max\{0, \max\{IM_1, IM_2, \ldots, IM_i, \ldots, IM_n\}\}, \tag{11}$$

where zero is the maximum intensity if no earthquakes occur in the interval. The parent distribution is the distribution of the intensity in one generic earthquake according to Eq. (3), which is derived from the hazard curve. It has been discussed already that in the hypothesis of classical PSHA, the intensities in different earthquakes, that is $IM_i\ \forall i$, are iid. Moreover, by virtue of the HPP, the expected value of the number of earthquakes observed in the time interval equals $\nu \cdot \Delta t$, thus it tends to infinity for $\Delta t \to +\infty$. Consequently, it could be investigated whether the distribution of the maximum in Eq. (5) asymptotically tends to an EV model.

When the parent distribution is unlimited on both the positive and negative axes, and the right tail has some specific features, then the asymptotic distribution, that is when $n \to +\infty$, of $IM_{\max}$ is:

$$F_{IM_{\max}(t,t+\Delta t)}(x) = e^{-e^{-\alpha \cdot (x-u)}},\ x \in (-\infty, +\infty), \tag{12}$$

where α and u are parameters related to the mean (μ) and the variance (σ^2) of $IM_{\max}(t, t+\Delta t)$ as $\mu \approx u + 0.577 \cdot \alpha^{-1}$ and $\sigma^2 \approx 1.645 \cdot \alpha^{-2}$. This distribution is called EV type I or Gumbel and, although being also defined for negative values of x, is often used to describe the maximum of non-negative variables.

A model for non-negative extreme values is the EV type II, or Fréchet, distribution, which is of the type:

$$F_{IM_{\max}(t,t+\Delta t)}(x) = e^{-\left(\frac{u}{x}\right)^k},\ x > 0, \tag{13}$$

applying to parent distributions with alternative shape features with respect to those for which the EV type I is derived. The mean and the variance are related to the parameters as $\mu = u \cdot \Gamma\left(1 - k^{-1}\right)$, with $k > 1$, and $\sigma^2 = u^2 \cdot \left[\Gamma\left(1 - 2 \cdot k^{-1}\right) - \Gamma^2\left(1 - k^{-1}\right)\right]$, with $k > 2$, and with $\Gamma(\cdot)$ being the *gamma function*.

These two distributions can be both considered candidates herein because the parent distribution has a site-specific shape, as the hazard integral in Eq. (1) shows. In fact, an EV type III distribution (related to the Weibull model) also exists, which is bounded in the right tail. Placing a cap on the maximum possible intensity is out of the scope of this study as this is a highly debated issue (e.g., [21,22]). Finally, it is worth recalling that all EV distributions can be seen as specific cases of the so-called generalized extreme value distribution [19].

For comparison, also the lognormal and gamma distributions, common in engineering or even to approximate hazard curves (e.g., [11]), are considered. The former is given in Eq. (14), while the latter in Eq. (15):

$$F_{IM_{\max}(t,t+\Delta t)}(x) = \Phi\left[\frac{\log(x) - \mu_{\log(IM_{\max})}}{\beta_{\log(IM_{\max})}}\right], x > 0, \tag{14}$$

$$F_{IM_{\max}(t,t+\Delta t)}(x) = \frac{1}{\Gamma(\alpha)} \cdot \int_0^{\gamma \cdot x} z^{\alpha-1} \cdot e^{-z} \cdot dz, x > 0. \tag{15}$$

In Eq. (14), the parameters $\mu_{\log(IM_{\max})}$ and $\beta_{\log(IM_{\max})}$ are the mean and the standard deviation of the logarithm of $IM_{\max}(t, t+\Delta t)$, while $\Phi(\cdot)$ is the standard (cumulative) Gaussian function. In Eq. (15) the parameters α and γ are related to the mean and variance of $IM_{\max}(t, t+\Delta t)$ as follows: $\mu = \alpha/\gamma$ and $\sigma^2 = \alpha/\gamma^2$.

4.1. Application to the case study sites

The hazard curves in Fig. 1B are transformed into probability curves

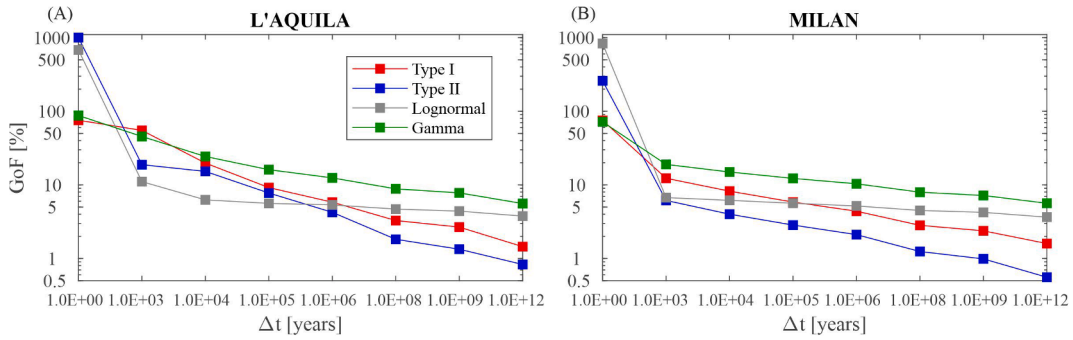

Fig. 5. Goodness of fit of some probabilistic models with respect to hazard curves for increasingly long intervals.

according to Eq. (6) considering three Δt values: 10^0, 10^3, and 10^6 years. (Although some may argue that the exceedance hazard curves for the longest of these intervals have no practical meaning, it must be recalled that, whatever the Δt, the information contained in $F_{IM_{\max}(t,t+\Delta t)}$ is always the same, so that changing the interval width can be interpreted simply as a scale change.).

Each of the four models from Eq. (12) to Eq. (15) was fitted to the exceedance curves; the fitting was obtained matching the mean and variance of the hazard distribution, properly accounting for the fact that the fitted distribution is a hybrid model, as Eq. (10) shows. The exceedance hazard curve and the fitted models are all displayed in Fig. 3 for the sites of L'Aquila and Milan, respectively.

Looking at L'Aquila, it can be qualitatively seen that the fit, when $\Delta t = 1\ yr$, is relatively poor for all the models, which is related to the fact that $e^{-\lambda_{IM>x}\cdot\Delta t}$, that is, the probability of not observing any event in the interval, is relatively large for small time intervals. Nevertheless, the model generally closer to the hazard curve seems to be the EV type II, closely followed by the EV type I. When $\Delta t = 1{,}000\ yr$, the fit seems to improve for all models, which is a consequence of the discontinuity jump of the distribution at zero becomes smaller, with the EV type II still apparently being the best performer. As the interval width becomes larger, the fitting of the EV models continues to improve with the same order of the two distributions, up to a point where the EV type II is almost completely overlapping the hazard curve. Conversely, the fitting of the other models does not seem to improve at the same pace, if at all. A similar situation is observed for the low-hazard site of Milan, indicating that – based on the considered cases – there is not a dependence of the asymptotic hazard behavior due to the hazard level of the considered site. This may be because, although the hazards at the two sites are different from the engineering point of view, the order of magnitude of the total rates is the same, which indicates the pace at which the number of observations tends to infinity as a function of the interval's width.

A more quantitative evaluation of the fit by the various considered models can be obtained via a probability vs probability (P—P) plots, which are provided in Fig. 4. The plots are obtained as follows: given an $IM_{\max}$ value in the hazard curve, which can be indicated as $x_{\text{hazard}}(p)$, the exceedance probability, say p', corresponding to such $IM_{\max}$ value in the fitted model is taken; then the p' is plotted versus p. Thus, a perfect fit would result in $p = p'$ for all the possible IM values.

Fig. 4 show, for both sites, a similar pace of possible convergence of the hazard curves towards the EV type I and type II model and that there is some convergence already for $\Delta t = 1{,}000\,yr$. It is also observed that the fit of the EV distribution always improves with increasing Δt, with the EV type II being the closest, while the same cannot be clearly claimed for the lognormal and gamma distributions.

The trend of the goodness of fit (GoF) can be quantified in several manners, with a possible (yet arbitrary and somewhat simplistic) metric being measuring the maximum relative difference between the quantiles of the hazard and fitted models:

$$\text{GoF} = \max_p \left\{ \left| \frac{x_{\text{model}}(p) - x_{\text{hazard}}(p)}{x_{\text{hazard}}(p)} \right| \right\} \cdot 100,\ p \in (0,1). \tag{16}$$

In the equation, $x_{\text{hazard}}(p)$ has been already defined, while $x_{\text{model}}(p)$ is the quantile of the fitted model corresponding to the same probability value as $x_{\text{hazard}}(p)$. The GoF practically measures the maximum relative distances of the PGAs corresponding to the same exceedance probability in the hazard curve and fitted distribution. Fig. 5 shows the GoF values (computed for the p values of the previous plots) for various Δt values, from 10^0 to 10^{12} years, including those three shown in the previous figures.

The trend of the hazard to be represented by EV distributions, in particular the EV type II, is quite evident, with maximum relative differences already around 10% or lower for relatively small Δt values. It is also confirmed that the two other models may provide some approximation, at least based on the GoF measure considered; however, they do not clearly emerge as asymptotes, at least in the cases made herein.

5. Final remarks

Engineering applications, such as reliability-based calibration of building codes or structural risk analyses, could benefit from stochastic modelling of loads acting on structures. One, often prominent, load for structural design derives from the seismic hazard of the construction site. However, probabilistic seismic hazard analysis does not typically yield a closed form for its results. On the other hand, hazard curves from classical PSHA, in their representation where the exceedance probability, in a given time interval, is a function of the ground motion intensity measure values, can be interpreted as the complementary cumulative distribution function of the maximum intensity in the interval. This is not a continuous distribution, because of the inherent features of the hazard curves from which it derives.

The intensities in various earthquakes among those considered by PSHA can be considered as independent and identically distributed random variables, and their common distribution can be directly derived by the hazard curves. Moreover, according to the stochastic Poisson process used to describe events' occurrence, the expected number of observations grows proportionally with the time interval's width. Consequently, because the extreme value distribution models are asymptotic distributions found under some similar hypotheses, it has been deemed interesting to investigate in this short note whether they could be a meaningful approximation of hazard curves. To this aim two sites in Italy, featuring different seismic hazard, have been taken as case studies. The PGA on rock hazard curves have been computed in terms of exceedance rate and of probability of exceedance in intervals with 10^0 to 10^{12} years. These curves have been fitted with four probabilistic models, EV type I, EV type II, lognormal, and gamma. The last two simply because they are common models for non-negative random variables.

The results quantitatively show that the hazard curves seem to tend to an EV model, which however needs time intervals longer than one year, a reference value in seismic hazard analysis, because for short time

interval the probability of not observing any event at all is relevant. More specifically, the best performing model seems to be the EV type II (Fréchet) somewhat closely followed by the EV type I. The other two common models considered, that is, the lognormal and the gamma, although still providing some approximation of the hazard curves, do not clearly emerge as asymptotes. These results hold independently of the seismic hazard level of the site, at least in the considered range of cases.

Declaration of Competing Interest

The author declares that he has no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

All data used in this study come from the listed references.

Acknowledgements

The study presented in this article was developed within the activities of the ReLUIS-DPC 2022–2023 research program, funded by the *Presidenza del Consiglio dei Ministri — Dipartimento della Protezione Civile* (DPC). (Note that the opinions and conclusions presented by the author do not necessarily reflect those of the funding entity.) The help from Pasquale Cito in developing the case studies herein shown, and the comments from Massimiliano Giorgio, Georgios Baltzopoulos (all at *Università degli Studi di Napoli Federico II*), Katsu Goda (*University of Western Ontario*), and an anonymous reviewer, are also gratefully acknowledged.

References

[1] Ambraseys NN, Simpson KA, Bommer JJ. Prediction of horizontal response spectra in Europe. Earthq Eng Struct Dyn 1996;25(4):371–400. https://doi.org/10.1002/(SICI)1096-9845(199604)25:4<371::AID-EQE550>3.0.CO;2-A.
[2] Benjamin J, Cornell CA. Probability, statistics, and decision for civil engineers. New York, NY: McGraw-Hill; 1970.
[3] Bommer JJ, Douglas J, Strasser FO. Style-of-faulting in ground-motion prediction equations. Bull Earthq Eng 2003;1:171–203. https://doi.org/10.1023/A:1026323123154.
[4] Chioccarelli E, Cito P, Iervolino I, Giorgio M. REASSESS V2.0: software for single- and multi-site probabilistic seismic hazard analysis. Bull Earthq Eng 2019;17(4):1769–93. https://doi.org/10.1007/s10518-018-00531-x.
[5] Cito P, Iervolino I. Rarity, proximity, and design actions: mapping strong earthquakes in Italy. Ann Geophys 2020;63(6):SE671. https://doi.org/10.4401/ag-8516.
[6] Cornell CA. Engineering seismic risk analysis. Bull Seismol Soc Am 1968;58(5):1583–606. https://doi.org/10.1785/BSSA0580051583.
[7] Cornell CA, Jalayer F, Hamburger RO, Foutch DA. Probabilistic basis for 2000 SAC federal emergency management agency steel moment frame guidelines. J Struct Eng 2002;128(4):526–33. https://doi.org/10.1061/(ASCE)0733-9445(2002)128:4(526).
[8] Ellingwood BR. Reliability-based condition assessment and LRFD for existing structures. Struct Saf 1996;18(2–3):67–80. https://doi.org/10.1016/0167-4730(96)00006-9.
[9] Gardner JK, Knopoff L. Is the sequence of earthquake in Southern California, with aftershocks removed, Poissonian? Bull Seismol Soc Am 1974. https://doi.org/10.1785/BSSA0640051363.
[10] Giorgio M, Iervolino I. On multisite probabilistic seismic Hazard analysis. Bull Seismol Soc Am 2016;106(3):1223–34. https://doi.org/10.1785/0120150369.
[11] Goda K. Nationwide Earthquake Risk Model for Wood-Frame Houses in Canada. In: *Front Built Environ*. 5. Frontiers Media S.A; 2019. p. 128. https://doi.org/10.3389/fbuil.2019.00128.
[12] Gumbel EJ. Statistics of extremes. New York: Columbia University Press; 1958.
[13] Gutenberg B, Richter CF. Frequency of earthquakes in California. Bull Seismol Soc Am 1944;34:185–8. https://doi.org/10.1785/BSSA0340040185.
[14] Iervolino I. Generalized earthquake counting processes for sequence-based Hazard. Bull Seismol Soc Am 2019;109(4):1435–50. https://doi.org/10.1785/0120180271.
[15] Joyner WB, Boore DM. Peak horizontal acceleration and velocity from strong-motion records including records from the 1979 Imperial Valley, California, earthquake. Bull Seismol Soc Am 1981;71(6):2011–38. https://doi.org/10.1785/BSSA0710062011.
[16] McGuire, R. K., 2004. *Seismic Hazard and Risk Analysis*. *Earthquake Engineering Research Institute, Oakland, CA* (MNO-10.). Oakland, CA: Earthquake Engineering Research Institute.
[17] Meletti C, Galadini F, Valensise G, Stucchi M, Basili R, Barba S, et al. A seismic source zone model for the seismic hazard assessment of the Italian territory. Tectonophysics 2008;450(1–4):85–108. https://doi.org/10.1016/j.tecto.2008.01.003.
[18] Milne WG, Davenport AG. Distribution of earthquake risk in Canada. Bull Seismol Soc Am 1969;59(2):729–54. https://doi.org/10.1785/BSSA0590020729.
[19] Mood AM, Graybill FA, Boes DC. *Introduction to the theory of statistics*. *McGraw-Hill* (3rd ed.). McGraw-Hill; 1974.
[20] Pisarenko VF, Sornette A, Sornette D, Rodkin MV. Characterization of the tail of the distribution of earthquake magnitudes by combining the GEV and GPD descriptions of extreme value theory. In: *Pure Appl Geophys*. 171. Birkhauser Verlag AG; 2014. p. 1599–624. https://doi.org/10.1007/S00024-014-0882-Z.
[21] Strasser FO, Bommer JJ, Abrahamson NA. Truncation of the distribution of ground-motion residuals. J Seismol 2007;12(1):79–105. https://doi.org/10.1007/s10950-007-9073-z.
[22] Suzuki A, Iervolino I. Italian vs worldwide history of largest PGA and PGV. Ann Geophys 2017;60(5):S0551. https://doi.org/10.4401/ag-7391.
[23] Weatherill G, Burton PW. An alternative approach to probabilistic seismic hazard analysis in the Aegean region using Monte Carlo simulation. In: *Tectonophysics*. 492. Elsevier; 2010. p. 253–78.
[24] Zentner I, Ameri G, Viallet E. Bayesian Estimation of the Maximum Magnitude mmax Based on the Extreme Value Distribution for Probabilistic Seismic Hazard Analyses. Pure Appl Geophys 2020;177(12):5643–60. https://doi.org/10.1007/S00024-020-02612-Y.
[25] Zoller G, Holschneider M, Hainzl S. The maximum earthquake magnitude in a time horizon: theory and case studies. Bull Seismol Soc Am 2013;103(2A):860–75. https://doi.org/10.1785/0120120013.