Natural Hazards manuscript No.
(will be inserted by the editor)

Estimation of the upper bound of seismic hazard curve by using the generalised extreme value distribution

V. A. Pavlenko

The final publication is available at Springer 10.1007/s11069-017-2950-z

Received: 15 August 2016 / Accepted: 31 May 2017

Abstract The problem considered in this study is that of unrealistic ground motion estimates, which arise in the Cornell-McGuire method when the seismic hazard curve is calculated for extremely low annual probabilities of exceedance. This problem stems from using the normal distribution in the modelling of the variability of the logarithm of ground motion parameters. In this study, the database of the strong-motion seismograph networks of Japan was used to examine the distribution of the logarithm of peak ground acceleration (PGA). The normal distribution and the generalised extreme value distribution (GEVD) models were considered in the analysis, with the preferred model being selected based on statistical criteria. The results of the analysis demonstrated the superiority of the GEVD in the vast majority of considered examples. The estimates of the shape parameter of the GEVD were negative in every considered example, indicating the presence of a finite upper bound of PGA. Therefore, the GEVD provides a model that is more realistic for the scatter of the logarithm of PGA, and the application of this model leads to a bounded seismic hazard curve.

Keywords Probabilistic seismic hazard analysis · Ground motion prediction equation · Ground motion variability · Peak ground acceleration · Hazard curve

1 Introduction

Probabilistic seismic hazard analysis (PSHA) is an important field of modern seismology, related to the effects of strong earthquakes and their consequences. The main purpose of PSHA is to estimate the design ground motion that can be utilized in earthquake structural engineering to produce a structure that can withstand a certain level of shaking without severe damage, and thus reduce the negative effects of strong earthquakes, i.e. casualties and damage to infrastructure. Several PSHA methods exist (Cornell, 1968; Milne and Davenport, 1969; Molchan et al., 1970; Veneziano et al., 1984; Kijko and Graham, 1998; 1999; Ebel and Kafka, 1999; Shumilina et al., 2000); however, the most widely applied method is the Cornell-McGuire procedure (Cornell, 1968; 1971; McGuire, 1976; 1978). This method was formulated by C.A.Cornell and L.Esteva (Bommer and Abrahamson, 2006; McGuire, 2008) and was supplemented with computer programs developed by R.K.McGuire.

Although extensive further development has taken place since the Cornell-McGuire method was originally published, some controversial aspects remain in its mathematical apparatus. One of these is the

V. A. Pavlenko
University of Pretoria Natural Hazard Centre, Pretoria, South Africa
E-mail: pavlenko.vasily@gmail.com

limitlessness of the seismic hazard curves at very low annual probabilities of exceedance (APE). Pertinent examples of such a flaw in modern PSHA include seismic hazard studies that were performed for the Yucca Mountain nuclear waste repository in the USA (Stepp et al., 2001) and the PEGASOS project in Switzerland (Abrahamson et al., 2002). Both studies considered an extremely low APE (down to 10^{-8} and 10^{-7}, respectively) and provided unrealistically high values of ground motion parameters (Stepp et al., 2001; Corradini, 2003; Stamatakos, 2004; Klügel, 2005).

For very low APE, the hazard estimates are controlled by the tail of the distribution of the ground motion residuals (Anderson and Brune, 1999; Abrahamson, 2000; Wang, 2011). In current practice, the distribution of the residuals of the ground motion parameters is assumed to correspond to a log-normal distribution and, therefore, the distribution of logarithmic residuals is modelled by a normal (Gaussian) distribution. Since normal distribution is unlimited, the estimations of the ground motion parameters at very low APE are unlimited as well.

The determination of the possible upper bounds of seismic ground motions has been discussed in engineering seismology for a long time. The history of the development of this subject is discussed in Bommer et al. (2004) and Strasser and Bommer (2009). The values that were suggested as possible limits of ground motion parameters, such as peak ground acceleration (PGA) and peak ground velocity, have been gradually increasing because of the accumulation of strong motion records and the consequent increase of the observed maximum values of these parameters.

The Bayesian procedure for estimating the maximum value of a parameter of ground motion that would occur at a given site within the specified time interval was proposed by Pisarenko and Lyubushin (1997; 1999) and was applied by Lyubushin et al. (2002). This procedure utilizes the catalogue of seismic events and the ground motion prediction equation (GMPE); it is applicable for any ground motion parameter that may be estimated by using the GMPE (e.g., PGA, components of acceleration response spectra, seismic intensity).

Kijko and Graham (1999) have proposed two approaches for estimating the maximum value of PGA at a site of interest. The first approach suggests the straightforward estimation of PGA by using the appropriate GMPE, under assumption that the strongest possible earthquake (i.e., an earthquake with magnitude $\hat{M}_{\max}$, where $\hat{M}_{\max}$ is estimated by using one of the methods such as developed by Pisarenko et al., 1996; Kijko, 2004; Kijko and Singh, 2011) occurred very close to the site (e.g., at a distance of 10 km). The second approach utilises the distribution of logarithm of PGA at a site, derived by Kijko and Graham (1999). The derived distribution of logarithm of PGA at a site is of the same type as the distribution of earthquake magnitude, obtained under the assumption of applicability of the Gutenberg-Richter recurrence law (Gutenberg and Richter, 1944). The similarity of two distributions allows estimation of the maximum possible PGA at a site by modifying the techniques developed for assessing the upper limit of earthquake magnitude.

Some studies (e.g., Strasser et al., 2008) have sought a solution to the problem by truncation of the distribution of ground motion residuals. However, such a procedure has no clear physical meaning; moreover, to a certain extent it is arbitrary.

An example of such a study is that of Romeo and Prestininzi (2000), who proposed a truncation at two standard deviations above the median of the distribution. On the other hand, Strasser et al. (2004) indicated that the truncation should be performed at least at a level three standard deviations above the median. Conversely, McGuire (1976) suggested that the distribution of residuals should be truncated at a level six standard deviations above the median value, or truncation should be performed in such a way that the ground motion amplitude at a site could not be greater than the value at the epicenter (McGuire, 1977).

Bommer and Abrahamson (2006) have indicated that truncation at a level above three standard deviations has little effect on the hazard curves in the range of the return periods that are generally used in engineering design. If the seismic activity of the region is not very high, this effect would remain small, even for a return period of 10^4 years.

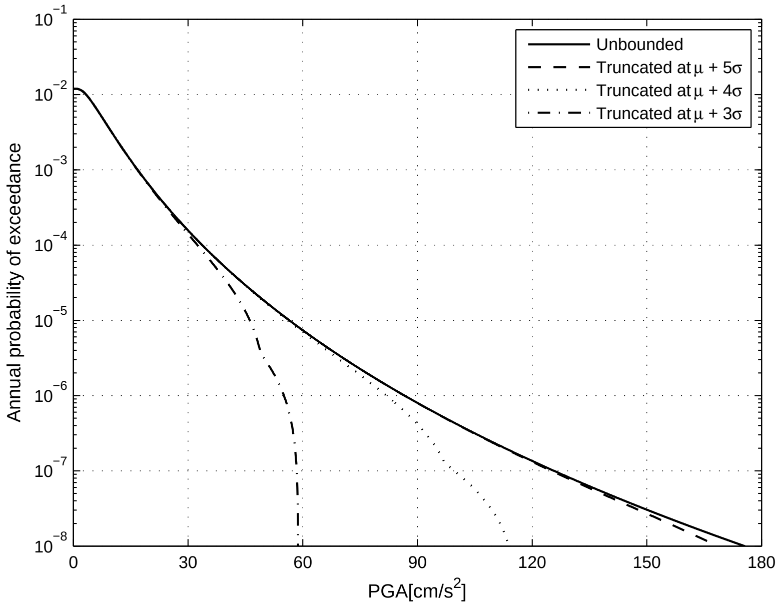

Fig. 1: Effect of truncation of the distribution of residuals on the hazard estimates

An interesting discussion on the influence of truncating the distribution of ground motion residuals on the results of PSHA can be found in Wu et al. (2011). These authors (Wu et al., 2011) concluded that the number of standard deviations depends on the range of APE under consideration and it should gradually increase with the decrease of the APE.

Hypothetical hazard curves, calculated by using the unbounded normal distribution, as well as the normal distribution, truncated at levels of three, four, and five standard deviations above the median value, are presented in Fig. 1

However, truncation of the distribution of the ground motion residuals potentially leads to an exclusion of certain strong motions that have low probability of occurrence from the scope of analysis. Moreover, it is exactly those particular strong motions that define the behavior of the seismic hazard curve at a long return period. Therefore, from this perspective, the truncation of ground motion variability seems doubtful.

On the other hand, some studies have focused on investigating the distribution of ground motion variability. Accordingly, Lavallée and Archuleta (2005) have investigated the distribution of the absolute values of PGA obtained from the strong motion records of the 1999 Chi-Chi earthquake and have suggested to model it by the Lévy distribution. Dupuis and Flemming (2006) have indicated that the distribution of residuals of PGA should theoretically correspond to the generalised extreme value distribution (GEVD). Dupuis and Flemming (2006) have performed regression analysis under assumption of the log-normal distribution of residuals, as well as under assumption of the residuals being distributed according to the GEVD. The results demonstrated that a superior fit to the data and, in turn, more accurate acceleration estimates are obtained under the latter assumption. Similar considerations were expressed by Raschke (2013), who criticized the log-normal assumption and noted that the GEVD is the natural distribution for residuals of maxima such as PGA.

Huyse et al. (2010) applied the peaks over threshold method to analyse both the raw PGA data and the logarithmic residuals of PGA, and concluded that the generalised Pareto distribution (GPD), with the negative shape parameter, provided a model that was more accurate for the tail fractions of both the studied datasets. Similar results were obtained by Pavlenko (2015), who found that the GEVD was a more appropriate model for logarithmic residuals of PGA.

In the present study, the analysis of the statistical properties of $\ln(\mathrm{PGA})$ is continued. The analysis is based on the data of the strong-motion seismograph networks of Japan (K-NET, KiK-net). Statistical criteria are used to compare the performance of the normal distribution and the GEVD models. The results indicate the superior performance of the GEVD in the vast majority of considered examples.

2 The Cornell-McGuire procedure

In PSHA, the seismic hazard is characterised by the probability $P(y \geq a_0, T) = P(a_0, T)$ that the ground motion parameter y will exceed the value a_0 at a given site at least once during a specified period of time T. It is usually assumed in PSHA studies that the sequence of major seismic events can be modelled by the Poisson process (e.g. Anderson and Brune, 1999). This assumption allows the calculation of $P(a_0, T)$:

$$P(a_0, T) = 1 - e^{-\lambda(a_0)T} \tag{1}$$

where $\lambda(a_0) = \lambda(y \geq a_0)$ is the mean annual rate of exceedance of ground motion level a_0 at the site.

For $T = 1$ year and for $(\lambda(a_0) \ll 1)$, Eq. (1) can be approximated as:

$$P(a_0, T = 1) = 1 - e^{-\lambda(a_0)} \cong \lambda(a_0) \tag{2}$$

Equation (2) is an approximation of the APE and $T = 1$ year is neglected on the right hand side of (2); therefore, both sides of this equation contain a dimensionless quantity (Wang, 2011). For a single seismic source, $\lambda(a_0)$ can be calculated as:

$$\lambda(a_0) = \nu P(y \geq a_0) \tag{3}$$

where ν is the annual rate of occurrence of earthquakes with magnitude greater than or equal to m_0, which is the lower threshold of magnitude of earthquakes capable of producing ground motions with $y \geq a_0$ at a site. The probability of exceedance $P(y \geq a_0)$, given the occurrence of an earthquake, can be calculated by using the total probability theorem:

$$P(y \geq a_0) = \iint P(y \geq a_0|m, r) f_M(m) f_R(r)\, dm\, dr \tag{4}$$

where $f_M(m)$ and $f_R(r)$ denote the probability density functions (PDFs) of magnitude and source to site distance, respectively, and $P(y \geq a_0|m, r)$ is the conditional probability that an earthquake of magnitude m would cause ground motion $y \geq a_0$ at distance r from the source.

The generalization of Eq. (3) for an instance of N seismic sources is straightforward, as the total annual rate of exceedance is the sum of the rates of individual seismic sources:

$$\lambda(a_0) = \sum_{i=1}^{N} \lambda_i(a_0) = \sum_{i=1}^{N} \nu_i P(y \geq a_0) \tag{5}$$

where subscript i indicates the i-th seismic source.

The substitution of Eq. (4) into Eq. (5) yields:

$$\lambda(a_0) = \sum_{i=1}^{N} \nu_i \iint P(y \geq a_0|m, r) f_{M_i}(m) f_{R_i}(r)\, dm\, dr \tag{6}$$

The variability of the ground motion was recognised as an important element of seismic hazard calculations (Bender, 1984), and integration over the distribution of possible values of y was included in the calculations (Cornell, 1971; McGuire, 1976).

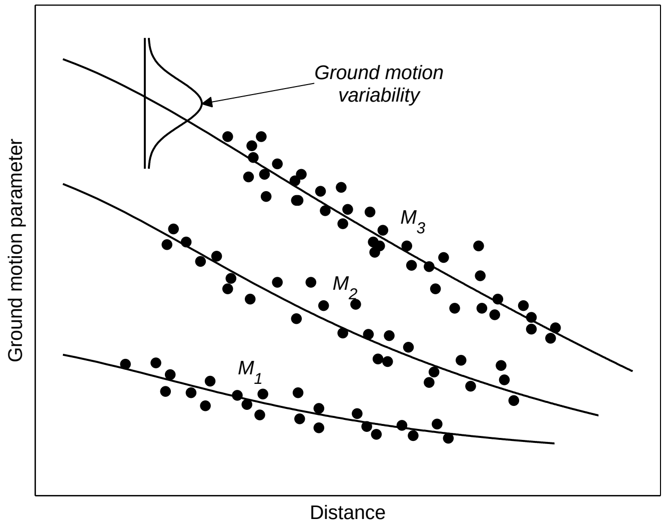

Fig. 2: Aleatory variability of ground motion

For a given combination of m and r, the GMPE allows estimation of the mean value of $\ln(y)$ and its standard deviation. The scatter of observed values of $\ln(y)$ around the mean is taken into account by considering the conditional probability distribution of $\ln(y)$ (Fig. 2). The standard assumption in current practice is that for the given m and r, y has a log-normal distribution (e.g. Abrahamson, 1988), or, equivalently, that $\ln(y)$ is normally distributed. Consequently, the conditional probability of exceedance $P(y \geq a_0|m,r)$ is calculated by using the normal distribution:

$$P(y \geq a_0|m,r) = \frac{1}{\sqrt{2\pi}\sigma} \int_{a_0}^{\infty} e^{-\frac{(\ln(u)-\mu)^2}{2\sigma^2}} du = 1 - \Phi(z) \tag{7}$$

where $z = [\ln(a_0) - \mu]/\sigma$ is the standardised normal random variable and $\Phi(z)$ is the standard normal cumulative distribution function (CDF).

As the normal distribution has unbounded support, any value of $\ln(y)$ receives a non-zero probability of being exceeded. Clearly, this is not appropriate, as the amount of energy released during an earthquake is finite, and, therefore, the ground motion should be bounded. As a result, truncating the distribution at some level above the median has become the standard practice. However, this practice has shortcomings, as was discussed above. A model that is more appropriate for the distribution of $\ln(y)$ would account for the finiteness of the ground motion induced by an earthquake of magnitude m at distance r.

3 Generalised extreme value distribution

The extreme value theory is devoted to the statistical analysis of rare events, and is used widely in various fields of knowledge, such as hydrology, meteorology, structural engineering, and earth sciences. This theory has a broad range of applications in the analyses of natural disasters (Pisarenko and Rodkin, 2010; 2014), and it is used in analysing the largest possible earthquakes (e.g. Epstein and Lomnitz, 1966; Kijko and Sellevoll, 1981).

The general result in the extreme value theory, the Fisher-Tippett-Gnedenko theorem (Fisher and Tippett, 1928; Gnedenko, 1943), postulates that a properly normalised maximum from a sample $\{X_n, n \geq 1\}$

of independent identically distributed random variables, with distribution function F, can only converge in distribution to one of the three possible limiting distributions. Specifically, assume a sequence of constants $a_n > 0$, and $b_n \in \mathbb{R}$ $(n \geq 1)$, such that a normalised sample maximum has a non-degenerate limiting distribution:

$$\lim_{n\to\infty} F^n(a_n x + b_n) = G(x) \tag{8}$$

Then, G must be one of the following three extreme value distributions:

$$\begin{array}{lll} \textit{Gumbel (type I):} & \Lambda(x) = \exp(-e^{-x}), & x \in \mathbb{R} \\ \textit{Fréchet (type II):} & \Phi_\alpha(x) = \exp(-x^{-\alpha}), & x > 0,\ \alpha > 0 \\ \textit{Weibull (type III):} & \Psi_\alpha(x) = \exp(-(-x)^{\alpha}), & x \leq 0,\ \alpha > 0 \end{array}$$

Three extreme value distributions can be combined into a single generalised form by introducing the shape parameter ξ so, that:

$$\xi = \begin{cases} 0 & \text{corresponds to } \Lambda(x) \\ \alpha^{-1} > 0 & \text{corresponds to } \Phi_\alpha(x) \\ -\alpha^{-1} < 0 & \text{corresponds to } \Psi_\alpha(x) \end{cases}$$

The following form is the GEVD, also called the Jenkinson-von Mises representation:

$$G_\xi(z) = \begin{cases} \exp\left(-(1+\xi z)^{-1/\xi}\right), & 1+\xi z > 0,\ \xi \neq 0 \\ \exp\left(-e^{-z}\right), & z \in \mathbb{R},\ \xi = 0 \end{cases} \tag{9}$$

where $z = (x-\mu)/\sigma$, μ and σ are the location and scale parameters, respectively.

The distribution function F belongs to the domain of attraction of G_ξ if (8) holds with $G = G_\xi$. In probability theory, heavy-tailed distributions are those probability distributions, whose tails tend to zero slower than an exponential function. Conversely, the distributions whose tails tend to zero faster than an exponential function are light-tailed distributions. The Gumbel domain of attraction $D(G_0)$ includes a large variety of distributions, of which the tails can differ significantly; ranging from moderately heavy, such as the log-normal distribution to very light, such as the normal distribution. The Fréchet domain of attraction $D(G_{\xi+})$ consists of the heavy-tailed distributions, of which the right tail behaves like a power law. Such distributions include the Pareto, Cauchy, Student's-t, and the Fréchet distributions. The Weibull domain of attraction $D(G_{\xi-})$ includes distributions with finite right endpoints (Fig. 3), for example, the uniform and the beta distributions.

Suppose a is the horizontal acceleration induced by an earthquake with magnitude m at distance r from the source. Let $a_{\max}$ denote the upper limit of a, then the possible values of $\ln(a)$ would be bounded by $(-\infty, \ln(a_{\max})]$, and the distribution of $\ln(a)$ would belong to the Weibull domain of attraction. As PGA is a maximum of a, it would be reasonable to expect that the distribution of $\ln(\mathrm{PGA})$ would converge to the Weibull extreme value distribution.

4 Applied procedure and data

The same procedure applied in Pavlenko (2015) was used in this study to examine the probability distribution of $\ln(\mathrm{PGA})$, with the normal distribution and the GEVD being considered as potential models in the current analysis. The data of the strong-motion seismograph networks of Japan (www.kyoshin.bosai.go.jp) were used in the study. The stations located on very dense soil (National Earthquake Hazards Reduction Program [NEHRP] class C) were selected, and the records of events with focal depths from 10 to 20 km and hypocentral distances from 46 to 54 km were used in the analysis. Some extension of the distance range is inevitable to gain enough data for statistical analysis. It is believed that the data obtained are

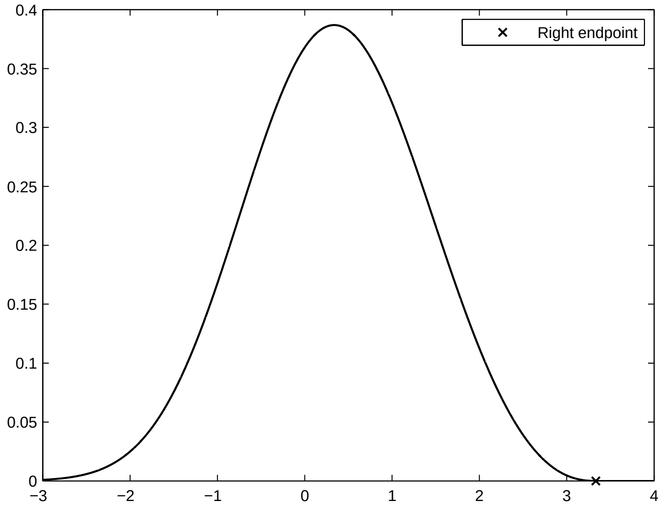

Fig. 3: PDF of a distribution with the support bounded on the right

representative of the random scatter in $\ln(\mathrm{PGA})$ for the given m and r. The data were grouped into bins according to magnitude, and the empirical distributions of $\ln(\mathrm{PGA})$ were modelled for the magnitude bins that contained the bulk of the data.

Various parameters have been used in ground motion studies to describe the level of horizontal acceleration, such as the square root of the sum of squares of two horizontal components (e.g. Kanno et al., 2006), and the geometric mean of the two components (e.g. Zhao et al., 2006). In this study, these two parameters are used for horizontal PGA, the first one is denoted PGA_{SR} (square root) and the second is denoted PGA_{GM} (geometric mean).

The PDFs of the normal distribution and the GEVD are given by:

$$\phi(z) = \frac{1}{\sqrt{2\pi\sigma}} \exp(-z^2/2),\ z \in \mathbb{R} \tag{10}$$

$$g(z) = \begin{cases} \exp\left(-(1+\xi z)^{-1/\xi}\right)(1+\xi z)^{-1/\xi-1}, & 1+\xi z > 0,\ \xi \neq 0 \\ \exp(-e^{-z}-z), & z \in \mathbb{R},\ \xi = 0 \end{cases} \tag{11}$$

The parameters of the normal distribution were estimated by the maximum likelihood method. The basic principle of this method is that the parameter values, θ, that maximise the likelihood of obtaining the given sample in a series of experiments should be taken as the most plausible estimates. In practice, this is achieved by maximisation of the likelihood function L, or its natural logarithm ℓ, called the log-likelihood function:

$$\ell = \ln(L) = \sum_{i=1}^{n} \ln[f(x_i|\theta)] \tag{12}$$

The details of the method for normal distribution are well known and therefore require no explanation. The estimators for parameters μ and σ are the sample mean and the sample standard deviation. The estimation of the parameters of the GEVD has its nuances; the support of G_ξ depends on the unknown values of the parameters μ, σ, and ξ. Furthermore, the applicability of the estimation methods depends on the value of ξ. Various methods exist for estimating ξ, which could be applied in different circumstances. For instance, the well-known Hill estimator (1975) is applicable only for positive values of ξ;

the maximum likelihood estimator is valid for $\xi > -0.5$; the probability-weighted moment estimator (Hosking et al., 1985) is valid for $\xi < 1$; and the Pickands estimator (1975) and the moment estimator proposed by Dekkers et al. (1989) can be applied in the general instance ($\xi \in \mathbb{R}$). Detailed reviews of these estimation techniques can be found in Embrechts et al. (1997), Beirlant et al. (2004), and de Haan and Ferreira (2006).

In the present study, the condition $\xi > -0.5$ was fulfilled for the analysed data and therefore the maximum likelihood method could be applied. The log-likelihood function of the sample $\{X_n, n \geq 1\}$ of GEVD random variables for the instance $\xi \neq 0$ is given by:

$$\ell = -n\ln(\sigma) - \left(\frac{1}{\xi}+1\right)\sum_{i=1}^{n}\ln(1+\xi z_i) - \sum_{i=1}^{n}(1+\xi z_i)^{-1/\xi} \tag{13}$$

where $z_i = (X_i - \mu)/\sigma$.

Differentiating (13) with respect to μ, σ, and ξ, yields the following likelihood equations:

$$\begin{cases} \sum_{i=1}^{n} a_i b_i = 0 \\ \sum_{i=1}^{n} z_i a_i b_i - n = 0 \\ \sum_{i=1}^{n} z_i a_i b_i + \frac{1}{\xi}\sum_{i=1}^{n}\ln(a_i)(b_i - \xi) = 0 \end{cases} \tag{14}$$

where $a_i = (1+\xi z_i)^{-1}$, $b_i = 1+\xi-(1+\xi z_i)^{-1/\xi}$.

These equations have no explicit solution and therefore should be solved numerically. Iterative numerical procedures were proposed for this purpose by, among others, Prescott and Walden (1980) and Hosking (1985). In the instance $\xi > -0.5$, the maximum likelihood method provides consistent, efficient, and asymptotically normal estimators. The covariance matrix of vector $\hat{\theta} = (\hat{\mu}, \hat{\sigma}, \hat{\xi})$ can be obtained from the inverse of the Fisher information matrix:

$$I_{ij} = -\left(\frac{\partial^2 \ell}{\partial\theta_i \partial\theta_j}\right)\Bigg|_{\theta=\hat{\theta}} \tag{15}$$

The Kolmogorov-Smirnov test (Kolmogorov, 1933) was used to test for significant deviations between the theoretical and the empirical distributions. The goodness of the fit of the models was compared by using the Akaike information criterion (AIC) (Akaike, 1974):

$$\mathrm{AIC} = -2\ln(L) + 2k \tag{16}$$

where L is the maximum value of the likelihood function, and k is a number of parameters of probability distribution. The AIC allows estimating the information loss as a result of using the particular model. When a set of candidate models is considered, with AIC_i denoting the AIC value of the i-th model, and $\mathrm{AIC}_{\mathrm{min}}$ denoting the minimum of those values, then, for the i-th model, the relative likelihood can be calculated as follows:

$$\tilde{L} = \exp[(\mathrm{AIC}_{\mathrm{min}} - \mathrm{AIC}_i)/2] \tag{17}$$

This quantity measures the relative probability of the i-th model to minimise the information loss.

5 Results and discussion

The histograms of $\ln(\mathrm{PGA}_{SR})$ and $\ln(\mathrm{PGA}_{GM})$, with fitted normal distribution, and the GEVD are shown in Figs. 4 and 5, and the relative likelihoods are listed in Table 1. The majority of histograms have a similar shape, with a slightly elongated right tail. There is good agreement between the results obtained

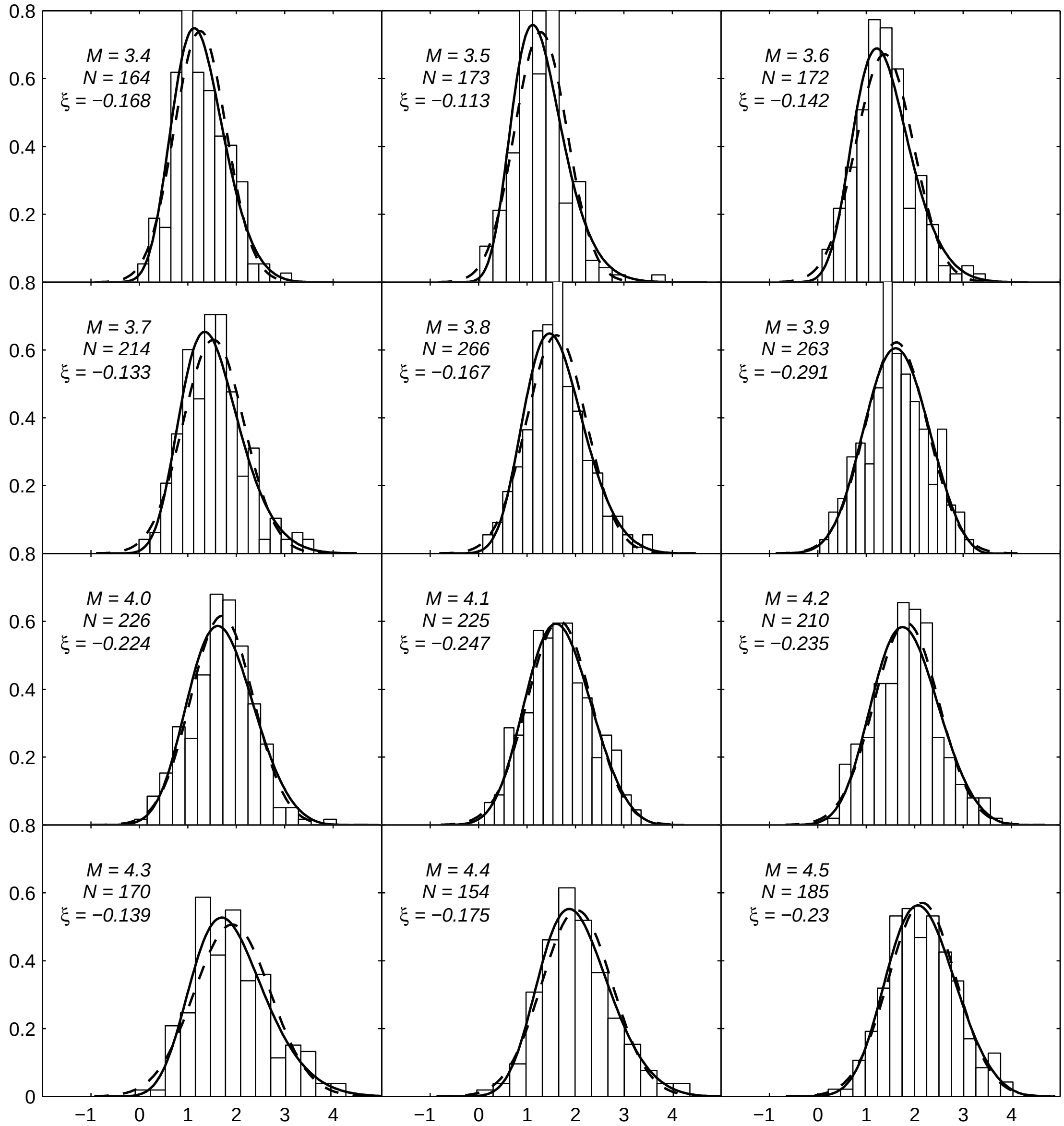

Fig. 4: Sample histograms of $\ln(\mathrm{PGA}_{SR})$. The PDFs of the normal distribution and the GEVD are shown by the broken and the solid lines, respectively. Magnitude (M), sample size (N) and estimated value of ξ are shown for each histogram

for $\ln(\mathrm{PGA}_{SR})$ and $\ln(\mathrm{PGA}_{GM})$, and, relevant to both parameters, the normal distribution performed better than the GEVD only in one instance out of twelve. Apparently, for this particular sample ($M = 4.0$) convergence to the GEVD was slower than for other samples. Although there is no obvious trend, all the estimates of ξ were negative, which confirms the convergence of data to the bounded Weibull extreme value distribution. Similar results were obtained by Huyse et al. (2010) and Pavlenko (2015).

The applicability of the GEVD for peak ground motion parameters is supported by the extreme value theory. The GEVD is a flexible distribution that can assume a variety of shapes and its shape parameter

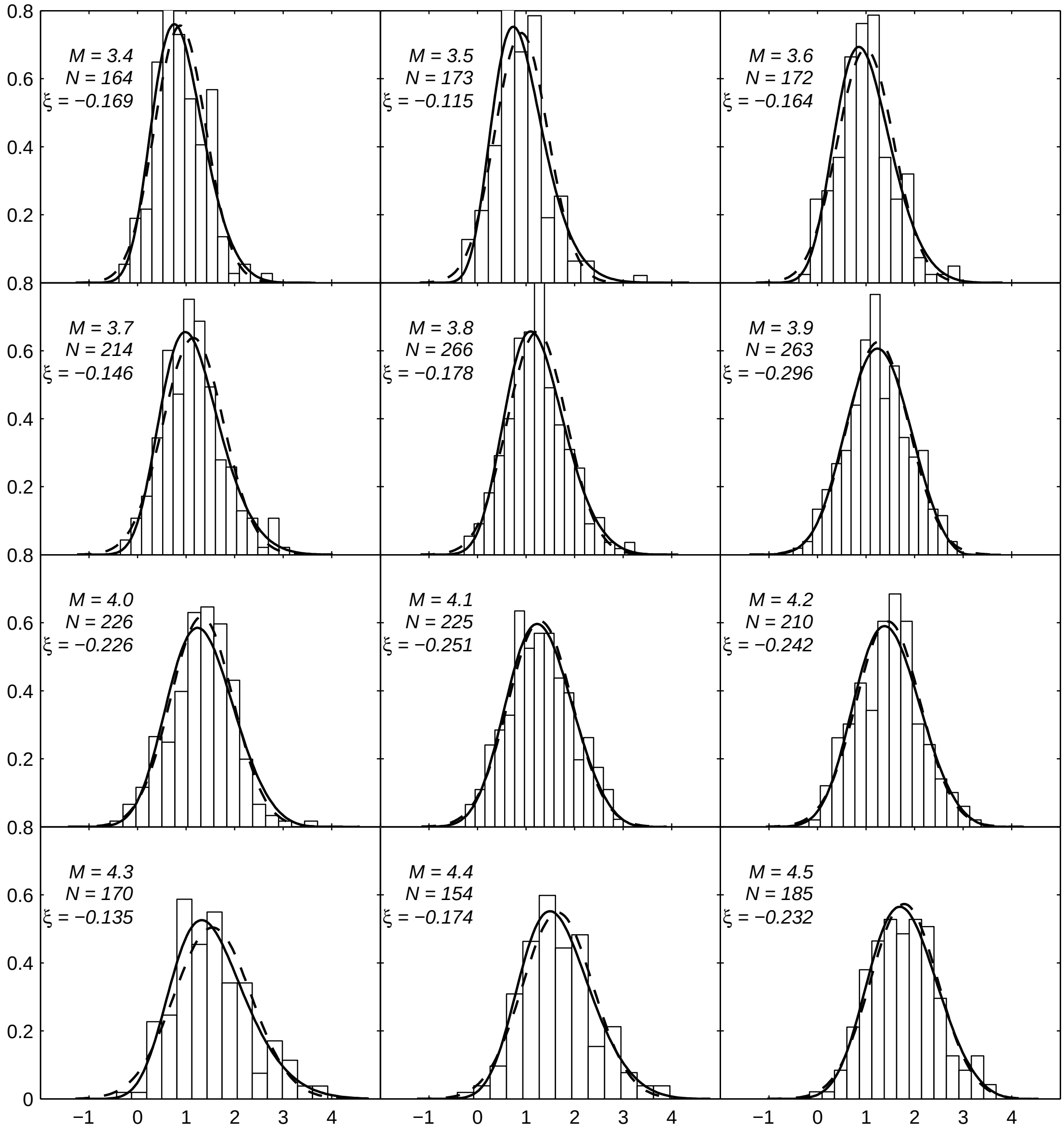

Fig. 5: The same as in Fig. 4 for $\ln(\mathrm{PGA}_{GM})$

ξ governs the decay of the tail of distribution. In instance $\xi < 0$, the support of the GEVD is bounded on the right. The inverse distribution function of the GEVD is defined for a probability $p \in [0, 1]$:

$$Q(p) = \begin{cases} \mu + \sigma\{[-\ln(p)]^{-\xi} - 1\}/\xi, & \xi \neq 0 \\ \mu - \sigma\{\ln[-\ln(p)]\}, & \xi = 0 \end{cases} \tag{18}$$

In instance $\xi < 0$, the right endpoint of the support is given by:

$$x^F = Q(1) = \mu - \frac{\sigma}{\xi} \tag{19}$$

	PGA$_{SR}$		PGA$_{GM}$	
M	$\tilde{L}(G_\xi)$	$\tilde{L}(\Phi)$	$\tilde{L}(G_\xi)$	$\tilde{L}(\Phi)$
3.4	1.0	0.497	1.0	0.313
3.5	1.0	0.133	1.0	0.092
3.6	1.0	0.344	1.0	0.113
3.7	1.0	0.032	1.0	0.013
3.8	1.0	0.173	1.0	0.084
3.9	1.0	0.479	1.0	0.394
4.0	0.016	1.0	0.028	1.0
4.1	1.0	0.312	1.0	0.321
4.2	1.0	0.847	1.0	0.872
4.3	1.0	0.029	1.0	0.034
4.4	1.0	0.402	1.0	0.414
4.5	1.0	0.758	1.0	0.699

Table 1: Relative likelihoods of the GEVD (G_ξ) and the normal distribution (Φ)

The estimate of x^F can be obtained by substituting the estimates of the parameters into Eq. (19). Thereby, the GEVD allows accounting for the finiteness of the seismic ground motion, and provides a rational way of estimating the maximum value of PGA for a specified earthquake scenario. This is a viable alternative to the common practice of using the truncated normal distribution (e.g. Strasser et al., 2008) for modelling the scatter of the logarithm of peak ground motion parameters.

6 Conclusion

In this study, the distribution of $\ln(\mathrm{PGA})$ was investigated by using the data of the strong-motion seismograph networks of Japan. The normal distribution and the GEVD were used for modelling the empirical distribution of $\ln(\mathrm{PGA})$. Two definitions of horizontal PGA were used, namely, the square root of the sum of squares of two horizontal components and the geometric mean of the two horizontal components. Similar results were obtained for both definitions, the GEVD provided a better fit than the normal distribution in eleven out of twelve instances. The estimated values of the shape parameter of the GEVD were negative in every instance, indicating that the support of the distribution is bounded on the right. Therefore, the GEVD provides a more realistic model for the scatter of the $\ln(\mathrm{PGA})$, which allows accounting for the finiteness of the ground motion induced by a specified earthquake scenario. The maximum value of PGA can be estimated directly from the parameters of the GEVD. Correct modelling of the ground motion parameters is important for realistic seismic hazard assessment and the studies on the statistical properties of these parameters should therefore be continued.

Acknowledgements I would like to thank Professor Andrzej Kijko for the review of the text, and the valuable suggestions that helped to improve its quality. I am grateful to two anonymous reviewers for their helpful comments and suggestions. I am grateful to the K-NET and KiK-net strong-motion seismograph networks of Japan for seismic ground motion records (www.kyoshin.bosai.go.jp).

References

Abrahamson NA (1988) Statistical properties of peak ground accelerations recorded by the SMART 1 array. Bull Seismol Soc Am 78(1):26–41

Abrahamson NA (2000) State of the Practice of Seismic Hazard Evaluation. In: Proceedings of GeoEng 2000, International Conference on Geotechnical & Geological Engineering, Melbourne, Australia, vol 1, pp 659–685

Abrahamson NA, Birkhauser P, Koller M, Mayer-Rosa D, Smit P, Sprecher C, Tinic S, Graf R (2002) PEGASOS - A Comprehensive Probabilistic Seismic Hazard Assessment for Nuclear Power Plants in Switzerland. In: Proceedings of the 12th European Conference on Earthquake Engineering, London, UK, paper No. 633

Akaike H (1974) A New Look at the Statistical Model Identification. IEEE Trans Autom Control 19(6):716–723, doi:10.1109/TAC.1974.1100705

Anderson JG, Brune JN (1999) Probabilistic Seismic Hazard Analysis without the Ergodic Assumption. Seismol Res Lett 70(1):19–28, doi:10.1785/gssrl.70.1.19

Beirlant J, Goedebeur Y, Teugels T, Segerus J, De Waal D, Ferro C (2004) Statistics of Extremes: Theory and Applications. Wiley, Chichester

Bender B (1984) Incorporating acceleration variability into seismic hazard analysis. Bull Seismol Soc Am 74(4):1451–1462

Bommer JJ, Abrahamson NA (2006) Why Do Modern Probabilistic Seismic-Hazard Analyses Often Lead to Increased Hazard Estimates? Bull Seismol Soc Am 96(6):1967–1977, doi:10.1785/0120060043

Bommer JJ, Abrahamson NA, Strasser FO, Pecker A, Bard PY, Bungum H, Cotton F, Fäh D, Sabetta F, Scherbaum F, Studer J (2004) The Challenge of Defining Upper Bounds on Earthquake Ground Motions. Seismol Res Lett 75(1):82–95, doi:10.1785/gssrl.75.1.82

Cornell CA (1968) Engineering seismic risk analysis. Bull Seismol Soc Am 58(5):1583–1606

Cornell CA (1971) Probabilistic analysis of damage to structures under seismic loads. In: Howells DA, Haigh IP, Taylor C (eds) Dynamic waves in civil engineering, Society for Earthquake and Civil Engineering, Wiley, New York, pp 473–488

Corradini ML (2003) Letter from chairman of the U.S. Nuclear Waste Technical Review Board to the Director of the Office of Civilian Radioactive Waste Management. http://www.nwtrb.gov/corr/mlc010.pdf

Dekkers ALM, Einmahl JHJ, de Haan L (1989) A moment estimator for the index of an extreme-value distribution. Ann Stat 17(4):1833–1855, https://projecteuclid.org/euclid.aos/1176347397

Dupuis DJ, Flemming JM (2006) Modelling peak accelerations from earthquakes. Earthq Eng Struct Dyn 35(8):969–987, doi:10.1002/eqe.565

Ebel JE, Kafka AL (1999) A Monte Carlo Approach to Seismic Hazard Analysis. Bull Seismol Soc Am 89(4):854–866

Embrechts P, Klüppelberg C, Mikosch T (1997) Modelling Extremal Events for Insurance and Finance. Springer-Verlag, Berlin

Epstein B, Lomnitz C (1966) A Model for Occurrence of Large Earthquakes. Nature 211:954–956, doi:10.1038/211954b0

Fisher RA, Tippett LHC (1928) Limiting forms of the frequency distribution of the largest and smallest member of a sample. Math Proc Camb Phil Soc 24(2):180–190, doi:10.1017/S0305004100015681

Gnedenko B (1943) Sur la distribution limite du terme maximum d'une série aléatoire (On the limiting distribution of the maximum term of random series). Ann Math 44(3):423–453, doi:10.2307/1968974. http://www.jstor.org/stable/1968974

Gutenberg B, Richter CF (1944) Frequency of earthquakes in California. Bull Seismol Soc Am 34(4):185–188

de Haan L, Ferreira A (2006) Extreme Value Theory: An Introduction. Springer Science and Business Media, LLC, New York

Hill BM (1975) A simple and general approach to inference about the tail of a distribution. Ann Stat 3(5):1163–1174, https://projecteuclid.org/euclid.aos/1176343247

Hosking JRM (1985) Algorithm AS 215: Maximum likelihood estimation of the parameters of the generalized extreme value distribution. Appl Stat 34(3):301–310, doi:10.2307/2347483. http://www.jstor.org/stable/2347483

Hosking JRM, Wallis JR, Wood E (1985) Estimation of the generalized extreme-value distribution by the method of probability-weighted moments. Technometrics 27(3):251–261, doi:10.2307/1269706. http://www.jstor.org/stable/1269706

Huyse L, Chen R, Stamatakos JA (2010) Application of Generalized Pareto Distribution to Constrain Uncertainty in Peak Ground Accelerations. Bull Seismol Soc Am 100(1):87–101, doi:10.1785/0120080265

Kanno T, Narita A, Morikawa N, Fujiwara H, Fukushima Y (2006) A New Attenuation Relation for Strong Ground Motion in Japan Based on Recorded Data. Bull Seismol Soc Am 96(3):879–897, doi:10.1785/0120050138

Kijko A (2004) Estimation of the Maximum Earthquake Magnitude, $m_{\max}$. Pure Appl Geophys 161(8):1655–1681, doi:10.1007/s00024-004-2531-4

Kijko A, Graham G (1998) Parametric-historic Procedure for Probabilistic Seismic Hazard Analysis. Part I: Estimation of Maximum Regional Magnitude $m_{\max}$. Pure Appl Geophys 152(3):413–442, doi:10.1007/s000240050161

Kijko A, Graham G (1999) "Parametric-historic" Procedure for Probabilistic Seismic Hazard Analysis. Part II: Assessment of Seismic Hazard at Specified Site. Pure Appl Geophys 154(1):1–22, doi:10.1007/s000240050218

Kijko A, Sellevoll MA (1981) Triple exponential distribution, a modified model for the occurrence of large earthquakes. Bull Seismol Soc Am 71(6):2097–2101

Kijko A, Singh M (2011) Statistical Tools for Maximum Possible Earthquake Magnitude Estimation. Acta Geophys Pol 59(4):674–700, doi:10.2478/s11600-011-0012-6

Klügel JU (2005) Problems in the application of the SSHAC probability method for assessing earthquake hazards at Swiss nuclear power plants. Eng Geol 78(3-4):285–307, doi:10.1016/j.enggeo.2005.01.007

Kolmogorov AN (1933) Sulla determinazione empirica di una legge di distribuzione (On the empirical determination of a distribution law). Giorn Ist Ital Attuar 4(1):83–91

Lavallée D, Archuleta RJ (2005) Coupling of the random properties of the source and the ground motion for the 1999 Chi Chi earthquake. Geophys Res Lett 32(L08311), doi:10.1029/2004GL022202

Lyubushin AA, Tsapanos TM, Pisarenko VF, Koravos GC (2002) Seismic Hazard for Selected Sites in Greece: A Bayesian Estimate of Seismic Peak Ground Acceleration. Nat Hazards 25(1):83–98, doi:10.1023/a:1013342918801

McGuire RK (1976) Fortran program for seismic risk analysis. U.S. Geological Survey, Open-File Report 76-67

McGuire RK (1977) Effects of uncertainty in seismicity on estimates of seismic hazard for the east coast of the United States. Bull Seismol Soc Am 67(3):827–848

McGuire RK (1978) FRISK: computer program for seismic risk analysis using faults as earthquake sources. U.S. Geological Survey, Open-File Report 78-1007
McGuire RK (2008) Probabilistic seismic hazard analysis: Early history. Earthq Eng Struct Dyn 37(3):329–338, doi:10.1002/eqe.765
Milne WG, Davenport AG (1969) Distribution of earthquake risk in Canada. Bull Seismol Soc Am 59(2):729–754
Molchan GM, Keilis-Borok VI, Vilkovich GV (1970) Seismicity and Principal Seismic Effects. Geophys J Roy Astron Soc 21(3):323–335, doi:10.1111/j.1365-246X.1970.tb01795.x
Pavlenko VA (2015) Effect of alternative distributions of ground motion variability on results of probabilistic seismic hazard analysis. Nat Hazards 78(3):1917–1930, doi:10.1007/s11069-015-1810-y
Pickands J (1975) Statistical inference using extreme order statistics. Ann Stat 3(1):119–131, http://www.jstor.org/stable/2958083
Pisarenko VF, Lyubushin AA (1997) Statistical estimation of maximum peak ground acceleration at a given point of a seismic region. J Seismol 1(4):395–405, doi:10.1023/a:1009795503733
Pisarenko VF, Lyubushin AA (1999) A Bayesian Approach to Seismic Hazard Estimation: Maximum Values of Magnitudes and Peak Ground Accelerations. Earthquake Res China (English Edition) 13(1):45–57, doi:10.1023/a:1013342918801
Pisarenko VF, Rodkin MV (2010) Heavy-Tailed Distributions in Disaster Analysis. Springer, Heidelberg
Pisarenko VF, Rodkin MV (2014) Statistical Analysis of Natural Disasters and Related Losses. Springer, Cham
Pisarenko VF, Lyubushin AA, Lysenko VB, Golubeva TV (1996) Statistical Estimation of Seismic Hazard Parameters: Maximum Possible Magnitude and Related Parameters. Bull Seismol Soc Am 86(3):691–700
Prescott P, Walden AT (1980) Maximum likelihood estimation of the parameters of the generalized extreme-value distribution. Biometrika 67(3):723–724, doi:10.1093/biomet/67.3.723
Raschke M (2013) Statistical modeling of ground motion relations for seismic hazard analysis. J Seismol 17(4):1157–1182, doi:10.1007/s10950-013-9386-z
Romeo R, Prestininzi A (2000) Probabilistic versus deterministic hazard analysis: An integrated approach for siting problems. Soil Dynam Earthquake Eng 20(1-4):75–84, doi:10.1016/S0267-7261(00)00039-7
Shumilina LS, Gusev AA, Pavlov VM (2000) An improved technique for determination of seismic hazard. J Earthq Predict Res 8:104–110
Stamatakos JA (2004) Review by the Office of Nuclear Material Safety and Safeguards of the U.S. Department of Energy's responses to key technical issue agreements SDS.2.01 and SDS.2.02 for a potential Geologic Repository at Yucca Mountain, Nevada, Project No: WM-011. Prepared for U.S. Nuclear Regulatory Commission Contract NRC-02-02-012. http://pbadupws.nrc.gov/docs/ml0428/ml042870242.pdf
Stepp JC, Wong I, Whitney J, Quittemeyer R, Abrahamson N, Toro G, Youngs R, Coppersmith K, Savy J, Sullivan T, Yucca Mountain PSHA Project Members (2001) Probabilistic Seismic Hazard Analyses for Ground Motions and Fault Displacements at Yucca Mountain, Nevada. Earthquake Spectra 17(1):113–151, doi:10.1193/1.1586169
Strasser FO, Bommer JJ (2009) Strong Ground Motions - Have We Seen the Worst? Bull Seismol Soc Am 99(5):2613–2637, doi:10.1785/0120080300
Strasser FO, Bommer JJ, Abrahamson NA (2004) The need for upper bounds on seismic ground motion. In: Proceedings of 13th World Conference on Earthquake Engineering, Vancouver, B.C., Canada, paper No. 3361
Strasser FO, Bommer JJ, Abrahamson NA (2008) Truncation of the distribution of ground-motion residuals. J Seismol 12(1):79–105, doi:10.1007/s10950-007-9073-z
Veneziano D, Cornell CA, O'hara T (1984) Historical Method for Seismic Hazard Analysis. Electric Power Research Institute, report, NP-3438, Palo Alto
Wang Z (2011) Seismic Hazard Assessment: Issues and Alternatives. Pure Appl Geophys 168(1-2):11–25, doi:10.1007/s00024-010-0148-3
Wu J, Mengtan G, Kun C, Bei H (2011) Discussion on the influence of truncation of ground motion residual distribution on probabilistic seismic hazard assessment. Earthquake Eng Eng Vib 10(3):379–392, doi:10.1007/s11803-011-0074-0
Zhao JX, Zhang J, Asano A, Ohno Y, Oouchi T, Takahashi T, Ogawa H, Irikura K, Thio HK, Somerville PG, Fukushima Y, Fukushima Y (2006) Attenuation Relations of Strong Ground Motion in Japan Using Site Classification Based on Predominant Period. Bull Seismol Soc Am 96(3):898–913, doi:10.1785/0120050122