Effect of alternative distributions of ground motion variability on results of Probabilistic Seismic Hazard Analysis

V.A. Pavlenko

Abstract Probabilistic seismic hazard analysis (PSHA) is a regularly applied practice that precedes the construction of important engineering structures. The Cornell-McGuire procedure is the most frequently applied method of PSHA. This paper examines the fundamental assumption of the Cornell-McGuire procedure for PSHA, namely, the log-normal distribution of the residuals of the ground motion parameters. Although the assumption of log-normality is standard, it has not been rigorously tested. Moreover, the application of the unbounded log-normal distribution for the calculation of the hazard curves results in non-zero probabilities of the exceedance of physically unrealistic amplitudes of ground motion parameters. In this study, the distribution of the residuals of the logarithm of peak ground acceleration is investigated, using the database of the Strong-motion Seismograph Networks of Japan and the ground motion prediction equation of Zhao and co-authors. The distribution of residuals is modelled by a number of probability distributions, and the one parametric law that approximates the distribution most precisely is chosen by the statistical criteria. The results of the analysis show that the most accurate approximation is achieved with the generalized extreme value distribution for a central part of a distribution and the generalized Pareto distribution for its upper tail. The effect of replacing a log-normal distribution in the main equation of the Cornell-McGuire method is demonstrated by the calculation of hazard curves for a simple hypothetical example. These hazard curves differ significantly from one another, especially at low annual exceedance probabilities.

Keywords ground motion variability · probabilistic seismic hazard analysis · hazard curve · ground motion prediction equation · peak ground acceleration

1 Introduction

Probabilistic seismic hazard analysis (PSHA) is a complicated and crucial problem of modern seismology as it is related to the effects of strong earthquakes and their consequences for the inhabitants. PSHA is applied to estimate the possible amplitudes of destructive seismic ground motion and to provide the design loads for the construction of critical structures such as dams and power plants. The main goal of such analysis is to minimize the negative effect of future strong earthquakes. Although there are several methods of PSHA (Cornell 1968; Shumilina et al

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

2000; Kijko 2008), the most frequently used method is the Cornell-McGuire procedure (Cornell 1968). The theoretical foundations, formulated by C.A. Cornell and L.Esteva (McGuire 2008), were supplemented by the computer programs developed by R.K. McGuire (McGuire 1976, 1978), which led to a method of PSHA known as the Cornell-McGuire procedure.

Ground motion variability is an important component of this method (Bender 1984; Bommer and Abrahamson 2006). This component was introduced in the Cornell-McGuire procedure to account for the effect of the scatter of the amplitude of seismic ground motion at a site (Cornell 1971) and is included in the main equation of this procedure. The common assumption is that the ground motion variability can be modelled by a random variable with a log-normal distribution (Joyner and Boore 1981). This implies that the residuals of ground motion parameters are log-normally distributed about the predicted value or, equivalently, that the residuals of the logarithms of these parameters are normally distributed. However, this hypothesis has not been reliably tested. Moreover, the assumption of log-normally distributed residuals has become a standard, and as a result usually is not tested but is accepted as a given.

The evidence for a log-normal distribution was confirmed by the Kolmogorov-Smirnov (KS) test at the 90 per cent confidence limit (Campbell 1981). Nevertheless, although the hypothesis was not rejected by the KS test, it does not imply that the hypothesis is true. The KS test does perform well in a central part of a distribution, however, it is widely known that the test demonstrates poor sensitivity to deviations from the hypothesized distribution that occur in the tails.

The log-normal assumption is criticized in Raschke (2013), where author notes that the natural distribution for residuals of maxima, such as peak ground acceleration (PGA), is the generalized extreme value. The theory of extreme values is widely applied in the analysis of natural disasters in general and in the analysis of seismic hazard in particular. Pisarenko and Rodkin (2010) provides the results of the application of the extreme value theory for various aspects of the analysis of natural disasters.

In general, a PSHA is applied to estimate ground motions with an annual probability of exceedance down to 10^{-4}, a typical annual exceedance probability value designated for nuclear power plant design. However, in a PSHA performed for the Yucca Mountain nuclear waste repository, probabilistic hazard curves were extrapolated to an annual exceedance probability of 10^{-8}. The peak characteristics of ground motion corresponding to an annual exceedance probability of 10^{-7} were as high as 20 g for PGA and up to 1800 $\frac{cm}{s}$ for peak ground velocity (PGV) (Corradini 2003; Stamatakos 2004). This instance revealed a controversy in a fundamental assumption of the modern Cornell-McGuire method. As pointed out e.g. in Abrahamson (2000), at these low annual probabilities, the hazard estimates are controlled by the tail of the distribution of the ground motion residuals. Since log-normal distribution is unbounded, extrapolation of a hazard curve leads to the unlimited increase of the amplitudes of expected ground motions, with the decreasing of the annual probability of exceedance.

On the other hand, some recent studies of the results of the Global Seismic Hazard Assessment Program (GSHAP) revealed discrepancies between the observed seismicity and that predicted by the resulting maps of this program (e.g. Kossobokov and Nekrasova 2011, 2012; Wyss et al 2012). The authors of these studies concluded that the common methods of PSHA are inadequate and need to be revised and probably modified.

One probable source of the revealed inadequacy is the assumption of the log-normal distribution of residuals of the ground motion parameter (e.g. PGA). An upper tail of the distribution of the ground motion residuals controls the behaviour of hazard curves at long return periods. Therefore, an accurate modelling of this distribution, especially at an upper tail region, is a significant problem.

The current study is methodological in nature and its main purpose is to introduce a suitable method of studying the ground motion variability. In this study, an analysis of the distribution of the residuals of the logarithm of PGA is performed in order to select a parametric law that describes this distribution most accurately. Data obtained from the Japanese Strong-motion Seismograph Networks were used in the study. The Japanese database was chosen mainly because of a dense net of strong-motion stations that allows obtaining enough observations. A ground motion prediction equation (GMPE) of Zhao et al (2006) was used for the calculation of the forecast values of PGA. Statistical criteria show that the best approximation for the distribution of residuals of the logarithm of PGA is achieved with the generalized extreme value distribution (GEVD). The generalized Pareto distribution (GPD) is used to capture the behaviour of an upper tail more accurately.

The remainder of the paper is organized as follows. In Section 2, the methods used for data analysis are described. In Section 3, the main results of data analysis are presented and discussed. An implication for PSHA is demonstrated in Section 4. Concluding remarks are summarized in Section 5.

2 Methods

The method for studying the distribution of residuals is based on the sequential application of the KS test and the Akaike information criterion (AIC), and a quantile-quantile plot. The distribution of the residuals of the natural logarithm of PGA is modelled by a number of parametric distributions. The residual is defined as

$$\varepsilon = \ln(PGA_{observed}) - \ln(PGA_{predicted}) \tag{1}$$

where $PGA_{observed}$ is the observed value and $PGA_{predicted}$ is the value calculated by using an appropriate GMPE.

The typical GMPEs allow the calculation of median values of the ground motion parameters by using their dependence on the magnitude, source to site distance, local soil conditions at a site, source mechanism, and others. Such equations often have an empirical nature and are developed based on vast databases of observed values of ground motion parameters (Boore and Joyner 1982). The selection of the most appropriate GMPE is not a trivial task and some guidance and criteria for choosing the most appropriate GMPE for the application in a PSHA for a particular site can be found in Scherbaum et al (2009) and Arroyo et al (2014). A comprehensive list of GMPEs developed during the period 1964-2010 is presented in Douglas (2011).

In this study, data recorded by the Japanese Strong-motion Seismograph Networks were used. The GMPE of Zhao et al (2006) was used for the calculation of the forecast values of PGA. This GMPE was developed for the calculation of the ground motion parameters of subduction zone earthquakes, it allows calculating a geometrical mean of the horizontal components of PGA, or 5% damped acceleration response spectrum.

In this study, the logistic distribution, the Student's t-distribution and the GEVD were considered as alternatives to the normal distribution. Following a standard notation, where μ is a location parameter and σ is a scale parameter, PDFs of these distributions can be written as follows:

The PDF of Student's t-distribution is defined as

$$f_{\mu,\sigma,n}(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\sqrt{\pi n}\sigma}\left(1 + \frac{1}{n}\frac{(x-\mu)^2}{\sigma^2}\right)^{-\frac{n+1}{2}}, \quad x \in (-\infty; +\infty)$$

where Γ is an Euler's gamma function, n is a number of degrees of freedom.

The PDF of logistic distribution is defined as

$$f_{\mu,\sigma}(x) = \frac{\exp\left(-\frac{x-\mu}{\sigma}\right)}{\sigma\left(1+\exp\left(-\frac{x-\mu}{\sigma}\right)\right)^2}, \quad x \in (-\infty; +\infty)$$

The PDF of the GEVD is defined as

$$f_{\mu,\sigma,\xi} = \frac{1}{\sigma}\begin{cases} \left(1+\xi\frac{(x-\mu)}{\sigma}\right)^{-\frac{1}{\xi}-1}\exp\left\{-\left(1+\xi\frac{(x-\mu)}{\sigma}\right)^{-\frac{1}{\xi}}\right\}, & 1+\xi\frac{(x-\mu)}{\sigma} \geq 0, \xi \neq 0 \\ \exp\left\{-\exp\left(-\frac{x-\mu}{\sigma}\right)-\frac{x-\mu}{\sigma}\right\}, & x \in (-\infty; +\infty), \xi = 0 \end{cases}$$

where ξ is a shape parameter.

Statistical analysis was performed in the following order:

- Estimation of the distribution parameters by the maximum likelihood (ML) method.
- Testing the hypothesis that a sample belongs to the current distribution by the KS test at 0.05 significance level.
- Calculation of the AIC for hypotheses that were accepted by the KS test.

The application of the KS test (Massey 1951) for one sample allows rejecting the distributions that do not fit the empirical data. The test statistic of this test with the Bol'shev's amendment (Bol'shev and Smirnov 1965) is calculated by using the formula

$$S_k = \frac{6nD_n + 1}{6\sqrt{n}} \tag{2}$$

where $D_n = \max(D_n^+, D_n^-)$, $D_n^+ = \max\left\{\frac{i}{n} - F(x_i, \theta)\right\}$, $D_n^- = \max\left\{F(x_i, \theta) - \frac{i-1}{n}\right\}$; n is a sample size, $x_1, \ldots, x_n$ - elements of a sample, sorted in ascending order, $F(x, \theta)$ is a cumulative distribution function (CDF) of a parametric model that undergoes the test.

An attractive feature of this test is that the distribution of its test statistic itself does not depend on the underlying cumulative distribution function being tested. However, in composite hypotheses testing, when the parameters of the probability distribution are estimated on the analysed sample, the KS test loses this feature. In such instances the conditional distribution of a test statistic depends on a number of factors (such as form of $F(x, \theta)$, number of estimated parameters, method of parameter estimation etc). Lemeshko and Lemeshko (2009) presents the updated results (tables of percentage points and models of the distributions of statistics) for non-parametric goodness of fit tests in testing composite hypotheses in case of using ML estimations.

The KS test rejects hypotheses for which the maximum deviation of the theoretical CDF from the empirical CDF exceeds a critical value at a given significance level.

However, the KS test alone does not allow unambiguous conclusion about which parametric model approximates the empirical distribution most accurately. Such a conclusion can be made based on the calculation of the AIC (Akaike 1974) for hypotheses that were accepted by the KS test. The criterion is defined as

$$AIC = -2\ln(L) + 2k \tag{3}$$

where L is a maximised likelihood function, k is a number of parameters of the probability distribution model.

The parametric distribution for which the value of criterion is minimal is considered the best approximation among the considered alternatives for the empirical distribution.

The quantile-quantile plot allows comparing the quantiles of empirical and theoretical distributions. The conception of such a plot has emerged from the observation that for important classes of distributions, the quantiles are linearly related to the corresponding quantiles of a standard example from this class (Beirlant et al 2004). Linearity in a graph can be easily checked by the eye and can further be quantified by means of a correlation coefficient.

3 Results and discussion

The results of the statistical analysis show that the best approximation of the distribution of residuals is achieved with the GEVD. It is important to note that a similar conclusion was reached in Dupuis and Flemming (2006) from theoretical considerations. In Dupuis and Flemming (2006) the regression analysis was performed using both the GEVD and the normal distribution as a model for the distribution of residuals, it was demonstrated that a better fit to the data and in turn more accurate acceleration estimates are obtained with the use of the GEVD. A similar conclusion in regards of the distribution of the ground motion residuals was also reached in Raschke (2013).

Figure 1 demonstrates the histogram of residuals together with the fitted PDFs.
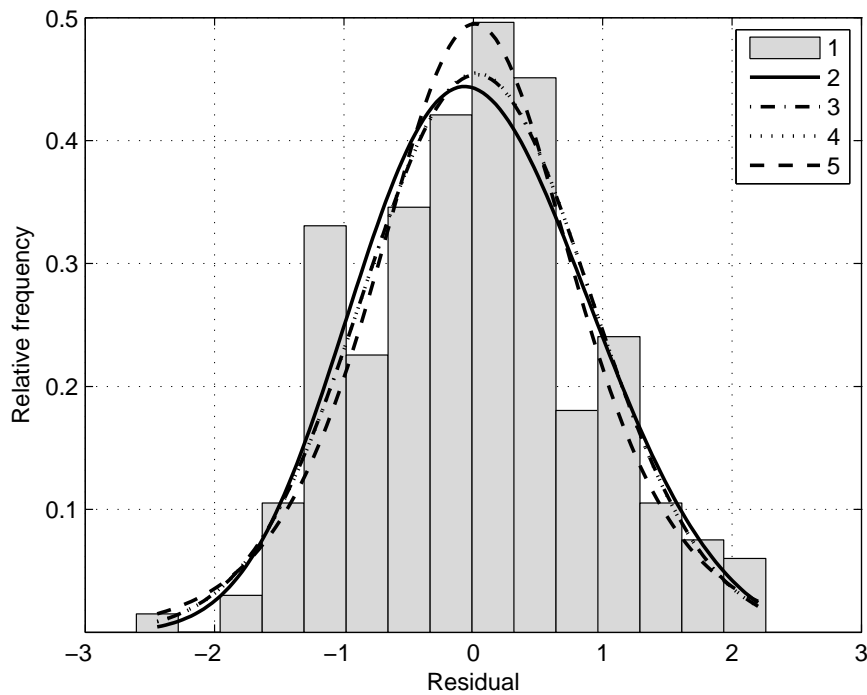

Fig. 1: Histogram of residuals $\left(\varepsilon = \ln(PGA_{observed}) - \ln(PGA_{predicted})\right)$ and the fitted PDFs. 1.Histogram 2.GEVD 3.Normal 4.Student's t 5.Logistic

Corresponding values of the AIC are presented in Table 1.

Model	AIC
GEVD	531.098
Normal	532.056
Student's t	534.054
Logistic	537.948

Table 1: Values of the AIC for considered distributions

The CDF of the GEVD is defined as

$$H_{\xi,\mu,\sigma}(x) = \begin{cases} \exp\left\{-\left(1+\xi\frac{(x-\mu)}{\sigma}\right)^{-1/\xi}\right\}, \xi \neq 0 \\[2ex] \exp\left\{-\exp\left(-\frac{x-\mu}{\sigma}\right)\right\}, \quad \xi = 0 \end{cases}$$

This is generalized form, also known as the Jenkinson-von Mises representation, which combines three types of extreme value distributions. When $\xi = 0$ it is equivalent to the Gumbel distribution (type I), when $\xi > 0$ it is equivalent to the Fréchet distribution (type II) and when $\xi < 0$ it is equivalent to the Weibull distribution (type III) (Embrechts et al 1997).

Figure 2 demonstrates the quantile-quantile plot for the quantiles of the GEVD. It can be seen from the plot, that an upper tail of the distribution of the residuals slightly deviates from the GEVD. Accurate modelling of that tail is an important problem because it defines the hazard curve at long return periods.
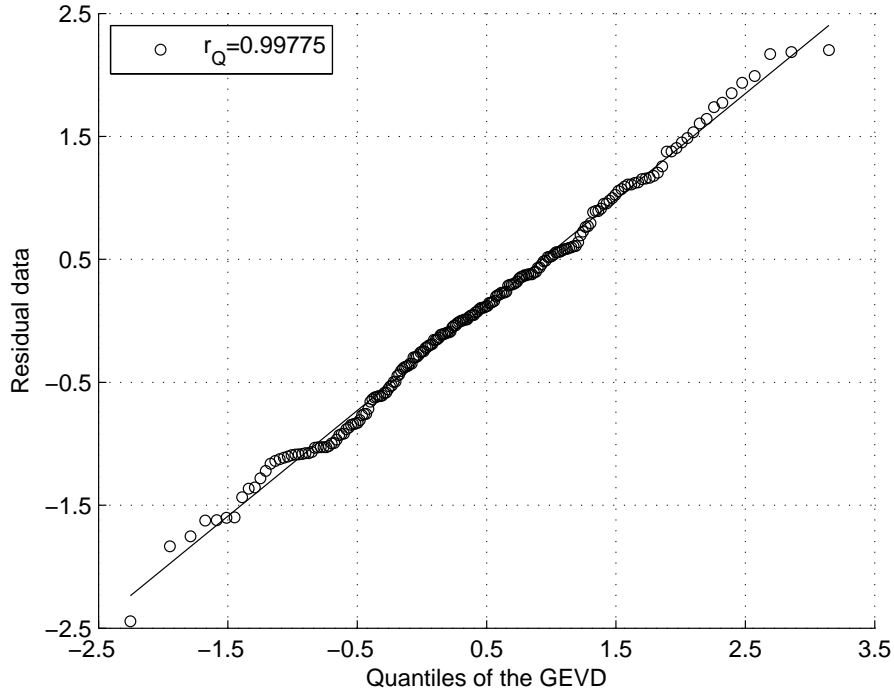

Fig. 2: Quantile-quantile plot of sample data vs quantiles of the GEVD

Therefore, the "peaks over threshold" method was applied to fit an upper tail of the distribution of residuals more precisely. This method is based on fitting the GPD to values that exceed a reasonably large threshold (Embrechts et al 1997). The GPD arises as a limiting distribution of the excesses for a sufficiently large threshold value and is often used for modelling the tails of empirical distributions. The CDF of the GPD is defined by the following function

$$
G_{\xi,\nu,\beta}(x) = \begin{cases} 1 - \left(1 + \xi\frac{(x-\nu)}{\beta}\right)^{-1/\xi}, \xi \neq 0 \\[2mm] 1 - e^{-\frac{x-\nu}{\beta}}, \qquad\qquad \xi = 0 \end{cases}
$$

Similar to the GEVD, the GPD is also characterised by three parameters, namely location ν, scale β and the shape parameter ξ. When the GPD is used as a model for a tail of some other distribution its parameter ν defines the threshold from which a tail region of that distribution begins. When $\xi = 0$ the GPD is equivalent to the exponential distribution, when $\xi > 0$ the GPD has a heavy tail, when $\xi < 0$ the GPD has a finite upper bound defined as $x_F = \nu - \frac{\beta}{\xi}$. Three possible types of tail of the PDF of the GPD, with different values of the shape parameter are shown in Figure 3.
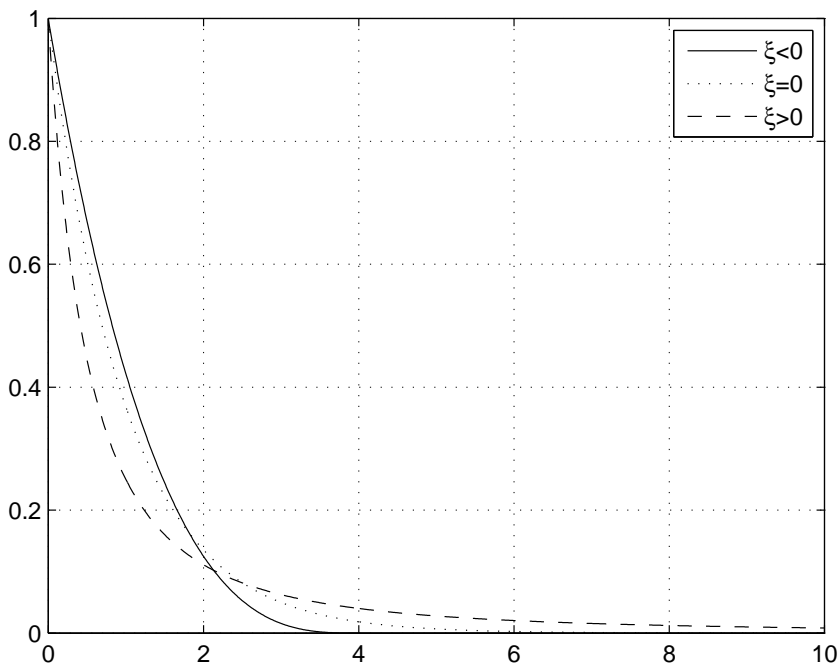

Fig. 3: Tails of PDF of the GPD with different values of parameter ξ

There are several methods for the estimation of the shape parameter. Well-known methods, such as the Hill estimator (Hill 1975) and the Pickands estimator (Pickands 1975) are both based on the asymptotic properties and require a significant number of observations. Applicability of

these methods to the real observations is doubtful (Pictet et al 1998). In this study, therefore, a shape parameter was estimated by using the ML method.

The robust estimation of the shape parameter $\hat{\xi}$ requires an optimal choice of the threshold value $\hat{\nu}$. If too high a value of $\hat{\nu}$ is chosen, too few exceedances and, consequently, high variance estimators will be the result. When $\hat{\nu}$ is too small, the estimators become biased. The procedure for the optimal determination of the threshold value is proposed in Embrechts et al (1997). This procedure utilises the linearity of the mean excess function for the GPD, which is defined as

$$e(\hat{\nu}) = E(X - \hat{\nu}|X > \hat{\nu}) = \frac{\hat{\beta} + \hat{\xi}\hat{\nu}}{1 - \hat{\xi}} \tag{4}$$

This procedure suggests selecting the threshold value $\hat{\nu}$ as a starting point of a linear segment of the mean excess graph. Such a graph for sample data is presented in Figure 4.
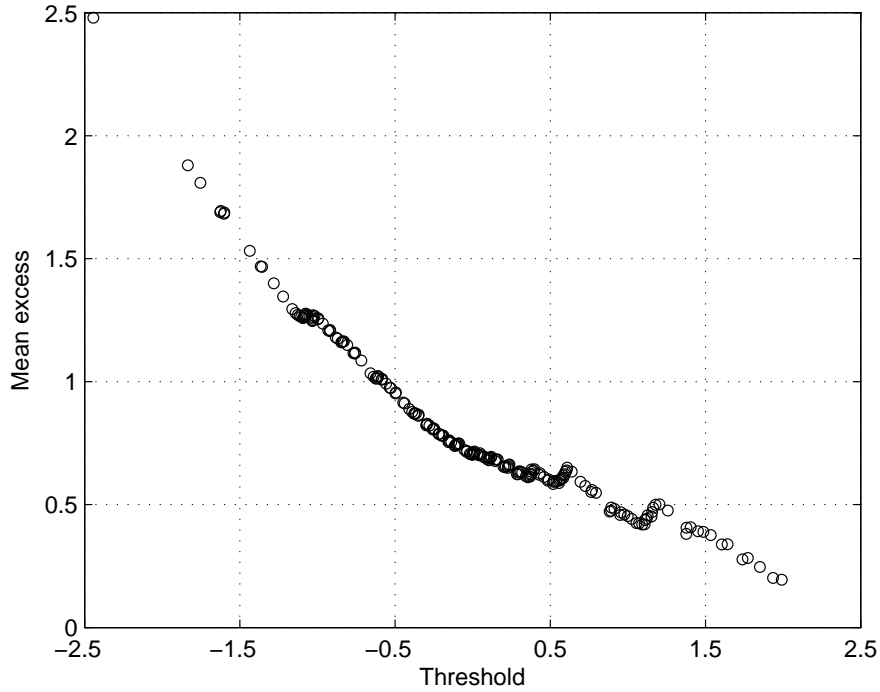

Fig. 4: Mean excess graph of sample data

By varying the threshold value and observing changes in the estimates of the rest of the parameters of the GPD, the optimal threshold value can be determined. Evidence of such a choice is the stabilization of the estimates of the scale and shape parameters. Once again, a quantile-quantile plot is used as a tool for comparing the data and the model. An estimator of the quantile of the GPD can be written as

$$\hat{x}_p = \hat{\nu} + \frac{\hat{\beta}}{\hat{\xi}}\left[\left(\frac{n}{N_{\hat{\nu}}}(1-p)\right)^{-\hat{\xi}} - 1\right] \tag{5}$$

where n and $N_{\hat{\nu}}$ are sample size and number of exceedances, respectively, and $\hat{\nu}$, $\hat{\beta}$ and $\hat{\xi}$ are estimates of the GPD parameters.

Quantile-quantile plots of a tail of residual data versus quantiles of the GPD, with the values of the shape parameter estimated by the ML method for the GEVD and the GPD are shown in Figure 5 and Figure 6. The estimates of a shape parameter differ for the instances where the GEVD is used as a model for a full range of residuals and where the GPD is used additionally to fit an upper tail more accurately. The estimates are $\hat{\xi} = -0.245$ for the first instance and $\hat{\xi} = -0.359$ for the second.

Therefore, the distribution of residuals is represented by a hybrid distribution model that consists of the GEVD in a central region and the GPD in a region of an upper tail.
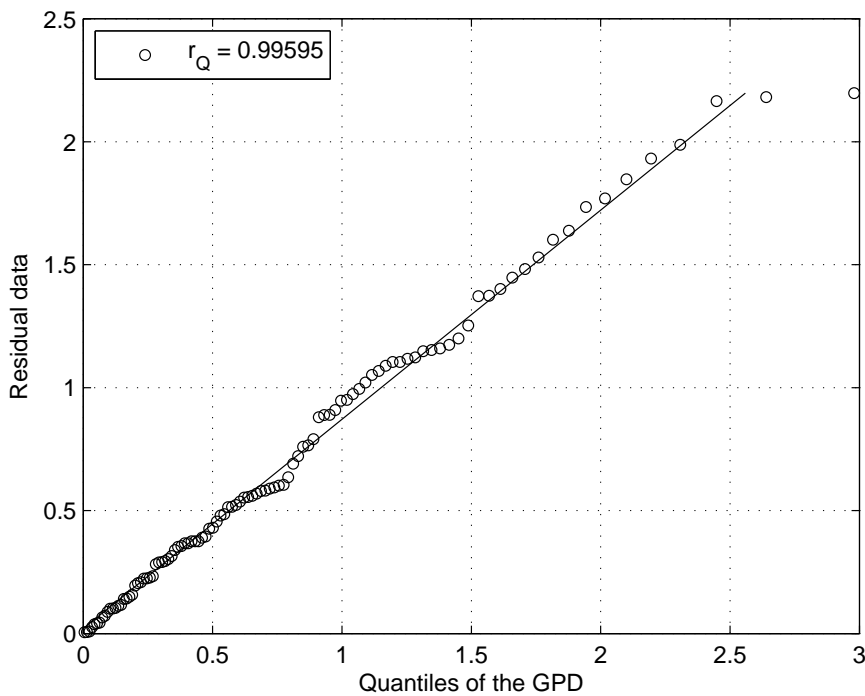

Fig. 5: Quantile-quantile plot of the tail fraction of residuals, ξ is estimated for the GEVD

A similar analysis was performed during this study by using the GMPEs of Atkinson and Boore (2003) and Kanno et al (2006) to check how generally applicable these results are. The results obtained with these GMPEs are very close to those presented in this study.

4 Implication for Probabilistic Seismic Hazard Analysis

For a demonstration of the effect of replacing a normal distribution, hazard curves were calculated in the following manner:
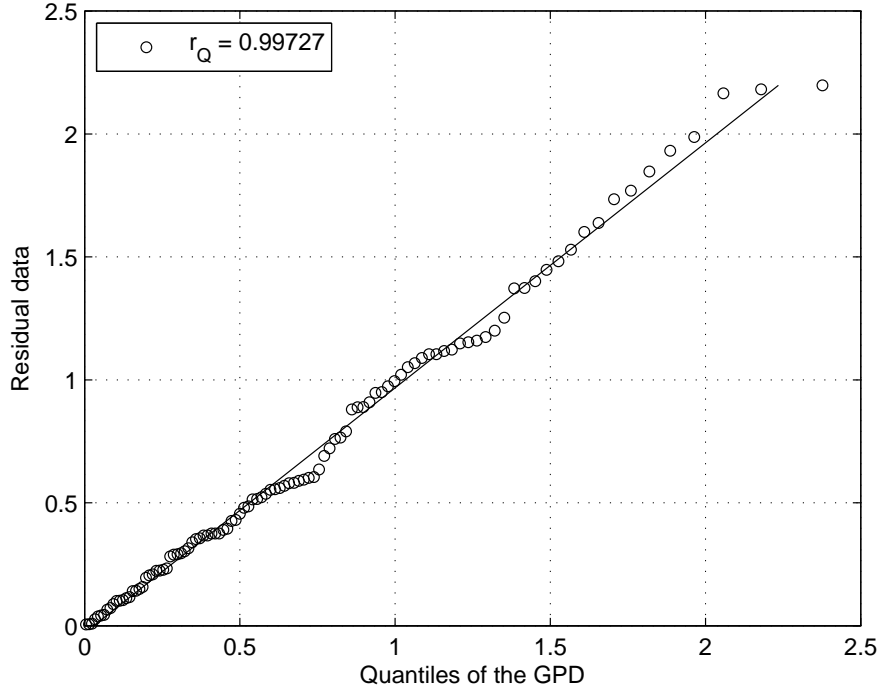
1. Using an unbounded normal distribution.

Fig. 6: Quantile-quantile plot of the tail fraction of residuals, ξ is estimated for the GPD

2. Using a normal distribution, truncated at a specified level of ground motion.
3. Using the GEVD.
4. Using a hybrid distribution model that consists of the GEVD for a central region and the GPD for an upper tail.

A brief review of the Cornell-McGuire probabilistic seismic hazard analysis procedure could be helpful for understanding the following material.

To begin with, recap of a PDF of log-normal distribution could be useful. If the distribution of a random variable is log-normal, its PDF has the following form

$$f_{\mu,\sigma}(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}}, \quad x > 0$$

The transformation $Y = \ln X$ leads to a normally distributed random variable with a location μ and a scale σ parameters. The values of these parameters are estimated by using an appropriate GMPE. Given an earthquake with magnitude m, the probability can be calculated that ground motion at distance r from the source will exceed a particular level a_0 by the following equation

$$P(y \geq \ln(a_0)|m,r) = \frac{1}{\sqrt{2\pi}\sigma} \int_{a_0}^{\infty} e^{-\frac{(y-\mu)^2}{2\sigma^2}} dy \tag{6}$$

This equation can be conveniently expressed in terms of the standard normal distribution

$$P(y \geq \ln(a_0)|m,r) = 1 - \Phi(z) \tag{7}$$

where $z = \frac{\ln(a_0) - \mu}{\sigma}$ is a standardised normal random variable and $\Phi(z)$ is the standard normal CDF.

Next, consider a site surrounded by N seismic sources. Each seismic source is characterized by magnitude M_i, distance to site R_i and annual activity rate ν_i. The parameters of future seismic events are yet unknown, therefore M_i and R_i are random variables with corresponding PDFs $f_{M_i}(m)$ and $f_{R_i}(r)$. The total annual rate of exceedance of a particular level of ground motion a_0 can be calculated as follows

$$\lambda(y \geq \ln(a_0)) = \sum_{i=1}^{N} \nu_i \iint P(y \geq \ln(a_0)|m,r) f_{M_i}(m) f_{R_i}(r) \, dr \, dm \tag{8}$$

From an assumption that the sequence of major seismic events can be modelled by the Poisson distribution it follows that the probability for a particular level of ground motion a_0 to be exceeded at least once during the time interval T can be calculated (Anderson and Brune 1999) as follows

$$P(y \geq \ln(a_0), T) = 1 - \exp(-\lambda(y \geq \ln(a_0)) \times T) \tag{9}$$

The equation 9 with $T = 1$ year defines the seismic hazard curve, the main result of the PSHA. For small values of the annual rate of exceedance ($\lambda(y \geq \ln(a_0)) \ll 1$), equation 9 can be approximated as

$$P(y \geq \ln(a_0), T = 1) = 1 - \exp(-\lambda(y \geq \ln(a_0))) \cong \lambda(y \geq \ln(a_0)) \tag{10}$$

As emphasized in Wang (2011), $T = 1$ year is neglected on the right side of 10, thus both sides of this equation contain a dimensionless quantity, i.e. the annual probability of exceedance.

As can be seen from equation 8, the ground motion variability is explicitly incorporated in the calculation of the seismic hazard. It is, namely used in a calculation of the conditional exceedance probability of a ground motion of a particular level a_0.

The normal distribution is unbounded, therefore the further a hazard curve extrapolated, the higher the level of ground motion is expected to be exceeded. The necessity of an upper bound of the ground motion, as well as the difficulties related to its determination are summarized in Bommer et al (2004). Strasser et al (2004) proposed the truncation of the distribution of residuals at a level of three standard deviations above the median as a measure to prevent the effect of unbounded normal distribution.

Given a normal distribution, truncated at a value a_T, the PDF has to be renormalized to satisfy the fundamental properties of PDF. Then, the probability that an earthquake with magnitude m will produce ground motion at distance r from the source that exceeds a particular level a_0 can be expressed as

$$P(y \geq \ln(a_0)|m,r) = \begin{cases} 1 - \frac{\Phi(z)}{\Phi(z_T)}, & y \leq a_T \\ 0, & y > a_T \end{cases} \tag{11}$$

where $z_T = \frac{\ln(a_T) - \mu}{\sigma}$.

After the replacement of a normal distribution by the GEVD, the same probability can be expressed as

$$P(y \geq \ln(a_0)|m,r) = 1 - H_\xi(z) \tag{12}$$

where H_ξ is a standardised CDF of the GEVD.

And after the replacement of a normal distribution by a hybrid model, this probability can be written as

$$P(y \geq \ln(a_0)|m, r) = \begin{cases} 1 - (1-p)\frac{H_\xi(z)}{H_\xi(z_\nu)}, \ y \leq \mu + \nu \\ p(1 - G_\xi(z)), \qquad y > \mu + \nu \end{cases} \tag{13}$$

where $z_\nu = \frac{\ln(a_\nu)-\mu}{\sigma}$, $a_\nu = \exp(\mu + \nu)$, $G_\xi(z)$ is a standardised CDF of the GPD, p is a fraction of the residual values that fall in a tail region.

For the purpose of demonstration, a simple hypothetical example was considered. This example is similar to an example used in Baker (2008) and assumes there are two seismic sources that may affect the site. Both sources are subduction slab sources. The first source is capable of producing an earthquake of magnitude $m_1 = 5.5$ every 100 years ($\nu_1 = 0.01$) and is located at a depth of $d_1 = 30 \ km$ and a distance of $r_1 = 140 \ km$ from the site. The second source is capable of producing an earthquake of magnitude $m_2 = 6.5$ every 500 years ($\nu_2 = 0.002$) and is located at a depth of $d_2 = 30 \ km$ and a distance of $r_2 = 200 \ km$ from the site. The soil conditions at a site are characterized as medium soil ($V_{S30} = 250 \ \frac{m}{s}$). For the given combinations of parameters, a GMPE of Zhao et al (2006) gives $\ln(PGA)$ values of $\mu_1 = 1.8404 \ \frac{cm}{s^2}$, $\mu_2 = 2.0233 \ \frac{cm}{s^2}$ and a standard deviation $\sigma = 0.6840$, which is a constant in this GMPE for seismic events generated by sources of identical type. With the defined earthquake scenarios, equation 8 simplifies to the following

$$\lambda(y \geq \ln(a_0)) = \nu_1 \times P(y \geq \ln(a_0)|m_1, r_1) + \nu_2 \times P(y \geq \ln(a_0)|m_2, r_2) \tag{14}$$

By repeating these calculations for a range of values of PGA, a total hazard curve can be constructed. Hazard curves calculated by using the above-mentioned distributions are represented in Figure 7.

As can be seen, the hazard curve calculated by using the GEVD displays the highest ground motion estimates, almost down to an annual exceedance probability of 10^{-6} where it crosses with the hazard curve calculated by using an unbounded normal distribution. The hazard curve calculated by using a hybrid distribution model is very close to the curve calculated by using a truncated normal distribution, down to an annual exceedance probability of 10^{-5}, after which it estimates higher ground motions and the difference gradually increases.

As can be seen from Figure 7, the hazard curves calculated by using the GEVD and a hybrid distribution model depend strongly on the shape parameter ξ. Although the method applied for statistical analysis in this study is satisfactory, the estimations of the shape parameter can only be called preliminary. These estimations were obtained based on the strong ground motion records from a particular region and, therefore, are valid only for this particular region, which, in this instance is Japan. Such analysis should be performed for a multiple number of datasets of the recordings of seismic ground motions that were induced by earthquakes of various types and magnitudes, and were recorded worldwide, for a possible generalization of these results.

5 Conclusion

In this study, the distribution of the residuals of $\ln(PGA)$ was modelled by a number of probability distribution laws, using the database of the Strong-motion Seismograph Networks of Japan and a GMPE of Zhao et al (2006). The results of the analysis indicate that the best approximation for the distribution of residuals was obtained with the GEVD. This result is consistent with the conclusions of Dupuis and Flemming (2006) and Raschke (2013). The "peaks
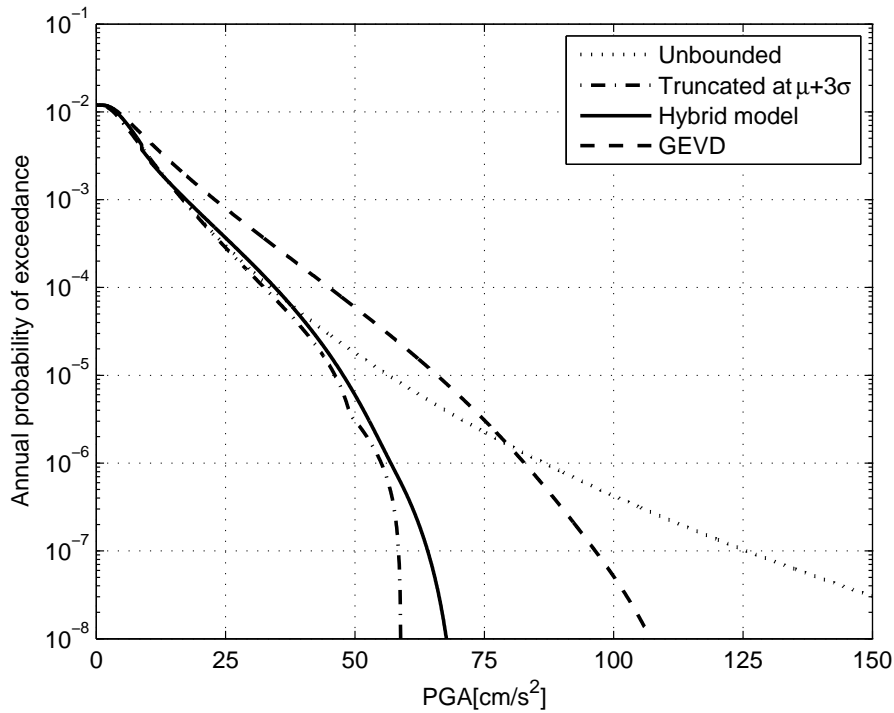
Fig. 7: Hazard curves calculated using different parametric distributions

over threshold" method was applied in an attempt to model an upper tail of the distribution of residuals more precisely. Thus, the resulting distribution of residuals is a hybrid model that consists of the GEVD in a central region and the GPD in a region of an upper tail. Similar analysis was performed during this study by using GMPEs of Atkinson and Boore (2003) and Kanno et al (2006), which demonstrated analogous regularities.

The estimations of the shape parameter of the GEVD and the GPD resulted in negative values, indicating that the distribution of residuals has a finite upper bound. Consequently, a maximum value of PGA can be associated with an earthquake scenario involved in the PSHA. This approach is preferred to the truncation procedures proposed in Strasser et al (2004), because a maximum value of PGA, unlike the truncation of a distribution, has a clear physical meaning.

Hazard curves were calculated for a simple hypothetical example to demonstrate the effect of the replacement of the normal distribution. Hazard curves were calculated by using the GEVD and a hybrid distribution model, which differ from each other and from the curves calculated by using the normal distribution. This difference is particularly evident at low annual exceedance probabilities.

References

Abrahamson N (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

Akaike H (1974) A new look at the statistical model identification. IEEE Transactions on Automatic Control 19(6):716–723

Anderson J, Brune J (1999) Probabilistic Seismic Hazard Analysis without the Ergodic Assumption. Seismological Research Letters 70(1):19–28

Arroyo D, Ordaz M, Rueda R (2014) On the selection of ground-motion prediction equations for probabilistic seismic-hazard analysis. Bulletin of the Seismological Society of America 104(4):1860–1875

Atkinson G, Boore D (2003) Empirical Ground-Motion Relations for Subduction-Zone Earthquakes and Their Application to Cascadia and Other Regions. Bulletin of the Seismological Society of America 93(4):1703–1729

Baker J (2008) An introduction to probabilistic seismic hazard analysis (PSHA) version 1.3. Available at http://web.stanford.edu/~bakerjw/Publications/Baker_(2008)_Intro_to_PSHA_v1_3.pdf

Beirlant J, Goedebeur Y, Teugels T, Segerus J, De Waal D, Ferro C (2004) Statistics of Extremes: Theory and Applications. John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex PO19 8SQ, England

Bender B (1984) Incorporating acceleration variability into seismic hazard analysis. Bulletin of the Seismological Society of America 74(4):1451–1462

Bol'shev L, Smirnov N (1965) Tables of Mathematical Statistics. Fizmatigiz, Moscow

Bommer J, Abrahamson N (2006) Why do modern probabilistic seismic-hazard analyses often lead to increased hazard estimates? Bulletin of the Seismological Society of America 96(6):1967–1977

Bommer J, Abrahamson N, Strasser FO, Pecker A, Bard P, Bungum H, Cotton F, Fäh D, Sabetta F, Scherbaum F, Studer J (2004) The challenge of defining upper bounds on earthquae ground motions. Seismological Research Letters 75(1):82–95

Boore D, Joyner W (1982) The empirical prediction of ground motion. Bulletin of the Seismological Society of America 72(6):S43–S60

Campbell K (1981) Near source attenuation of peak horizontal acceleration. Bulletin of the Seismological Society of America 71(6):2039–2070

Cornell C (1968) Engineering seismic risk analysis. Bulletin of the Seismological Society of America 58(5):1583–1606

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

Corradini M (2003) June 27 letter from Michael L. Corradini, Chairman, U.S. Nuclear Waste Technical Review Board, to Margaret S. Y. Chu, Director, Office of Civilian Radioactive Waste Management, on the Boards reactions to presentations at the February 2003 joint meeting of the Panel on the Natural System and the Panel on the Engineered System

Douglas J (2011) Ground-motion prediction equations 1964-2010. Final Rept. RP-59356-FR, Bureau de Recherches Gologiques et Minires (BRGM), Orlans, France, 444 pp. Available at http://peer.berkeley.edu/globalgmpe/wp-content/uploads/2011/03/douglas2011_RP-59356-FR.pdf

Dupuis D, Flemming J (2006) Modelling peak accelerations from earthquakes. Earthquake Engineering and Structural Dynamics 35:969–987

Embrechts P, Klüppelberg C, Mikosch T (1997) Modelling Extremal Events for Insurance and Finance. Springer-Verlag, Berlin, Heidelberg, New York, London, Paris, Tokyo, Hong Kong, Barcelona, Budapest

Hill B (1975) A simple and general approach to inference about the tail of a distribution. The Annals of Statistics 3(5):1163–1174

Joyner W, Boore D (1981) Peak horizontal acceleration and velocity from strong-motion records including records from the 1979 Imperial Valley, California, earthquake. Bulletin of the Seismological Society of America 71(6):2011–2038

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. Bulletin of the Seismological Society of America 96(3):879–897

Kijko A (2008) Data driven probabilistic seismic hazard assessment procedure for regions with uncertain seismogenic zones. NATO Science Series: IV: Earth and Environmental Sciences 81:237–251

Kossobokov V, Nekrasova A (2011) Global Seismic Hazard Assessment Program (GSHAP) Maps Are Misleading. Problems of Engineering Seismology 38(1):65–76, in Russian

Kossobokov V, Nekrasova A (2012) Global Seismic Hazard Assessment Program Maps are Erroneous. Seismic Instruments 48(2):162–170

Lemeshko B, Lemeshko S (2009) Distribution models for nonparametric tests for fit in verifying complicated hypotheses and maximum-likelihood estimators. Part 1. Measurement Techniques 52(6):555–565

Massey F (1951) The Kolmogorov-Smirnov test for goodness of fit. Journal of the American Statistical Association 46(253):68–78

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

McGuire R (1978) FRISK: computer program for seismic risk analysis using faults as earthquake sources. U.S. Geological Survey, open-File Report 78-1007

McGuire R (2008) Review. Probabilistic seismic hazard analysis: Early history. Earthquake Engineering and Structural Dynamics 37:329–338

Pickands J (1975) Statistical inference using extreme order statistics. The Annals of Statistics 3(1):119–131

Pictet O, Dacorogna M, Muller U (1998) Hill, bootstrap and jackknife estimators for heavy tails. In: Adler RJ, Feldman RE, Taqqu MS (eds) A Practical Guide to Heavy Tails: Statistical Techniques and Applications, Birkhäuser, Boston, pp 283–310

Pisarenko V, Rodkin M (2010) Heavy Tailed Distributions in Disaster Analysis. Springer, Dordrecht, Heidelberg, London, New York

Raschke M (2013) Statistical modeling of ground motion relations for seismic hazard analysis. J Seismol 17:1157–1182

Scherbaum F, Delavaud E, C R (2009) Model Selection in Seismic Hazard Analysis: An Information-Theoretic Perspective. Bulletin of the Seismological Society of America 99(6):3234–3247

Shumilina L, Gusev A, Pavlov V (2000) An improved technique for determination of seismic hazard. Journal of Earthquake Prediction Research 8:104–110

Stamatakos J (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. Center for Nuclear Waste Regulatory Analyses, San Antonio, Texas, prepared for U.S. Nuclear Regulatory Commission Contract NRC-02-02-012, available at http://pbadupws.nrc.gov/docs/ML0428/ML042870242.pdf

Strasser F, Bommer J, Abrahamson N (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

Wang Z (2011) Seismic Hazard Assessment: Issues and Alternatives. Pure and Applied Geophysics 168(1-2):11–25

Wyss M, Nekrasova A, Kossobokov V (2012) Errors in expected human losses due to incorrect seismic hazard estimates. Natural Hazards 62(3):927–935

Zhao J, Zhang J, Asano A, Ohno Y, Oouchi T, Takahashi T, Ogawa H, Irikura K, Thio H, Somerville P, Fukushima Y, Fukushima Y (2006) Attenuation Relations of Strong Ground Motion in Japan Using Site Classification Based on Predominant Period. Bulletin of the Seismological Society of America 96(3):898–913