16TH EUROPEAN CONFERENCE ON
EARTHQUAKE ENGINEERING THESSALONIKI
18 - 21 JUNE 2018

PROPOSED METHOD FOR SELECTING BLOCKS OF MAXIMA FOR PEAK GROUND ACCELERATION DATA BASED ON EXTREME VALUE THEORY

Shahin BORZOO[1], Morteza BASTAMI[2], Afshin FALLAH[3]

ABSTRACT

In modeling the risk of extremes, rare events and estimating the probability of an unusual event, extreme value theory, EVT, and generalized extreme value distribution are used. Generalized extreme value, GEV, distribution is three parametric continuous probability distributions based on EVT to combine three different of extremal distribution that is known as the Gumbel, Frechet and Weibull distributions for maxima or minima values. in this paper, GEV distribution is utilized for investigating the peak ground acceleration, PGA, that is with maximal nature. Also, a selection method for extreme values in fitting the GEV distribution of PGA data is presented and superiority of this distribution than other is shown. For evaluating generalized extreme value analysis two data selection method, are exist, block maxima and peak over the threshold. In this paper, block maxima method and how to choose the blocks in seismology is proposed. The common method for picking out the blocks in other events is yearly, but this period for PGA data is not logical. In the proposed method for selecting PGA blocks, two features of the equal size of blocks and balance between bias and variance are considered. After selecting the blocks, maximum values of each block are extracted and usual distributions are fitted to these values. The results show better fit for GEV distribution than commonly used distribution, such as log-normal for the PGA. In this investigation usual distribution for PGA data of Chi-Chi earthquakes are compared and GEV distribution has the highest maximum likelihood.

Keywords: Extreme value theory; block maxima; generalized extreme value distribution; peak ground acceleration

1. INTRODUCTION

The extreme values are data sets whose occurrence is rare in a statistical process. Although the occurrence of these phenomena is rare and the period of return is long, its occurrence in natural disasters such as earthquakes, floods, snow, and rain can have irreparable consequences. Therefore, the use of methods for predicting and preventing these events has been considered, frequently in various scientific fields. In recent years, using the extreme value theory has been increased in risk and hazard assessment. In the following sections, this statistical theory is described. In this theory, the extreme behavior is represented by three distributions of the extreme value, Gumbel, Frechet, and Weibull, which was presented by Fisher and Tippett (1928). Although, the first application of the extreme value distribution is given by Fuller (1914). Useful utilizations of the extreme value distribution is in meteorological information. The theory of extreme value depends on the statistical behavior of the extreme values in different processes. The theory of extreme value focuses on the statistical behavior of $Z_n = \max\{X_1,...,X_n\}$ in which $X_1,...,X_n$ is a sequence of random variables. In fact, $X_1,...,X_n$ is a block of data which can be related to rainfall, snow data, seismic data, financial parameters and etc.; and Z_n is the maximum measured value of a process caused by n observations

[1] phd candidate, IIEES, Tehran, Iran, s.borzoo@iiees.ac.ir
[2] Infrastructures Department, Associate professor, IIEES, Tehran, Iran, m.bastami@iiees.ac.ir
[3] Statistics Department, Associate professor, IKIU, Qazvin, Iran, afshin.ikiu@gmail.com

in a measured block. It can be shown that when the number of samples is large enough then the probability of not exceedance of Z_n than z , $P(Z_n \leq z)$, is presented by cumulative distribution function of the generalized extreme value as below:

$$G(z;\mu,\sigma,\xi)=\begin{cases}\exp\left\{-\left[1+\xi\left(\frac{z-\mu}{\sigma}\right)\right]^{-\frac{1}{\xi}}\right\} & 1+\xi\left(\frac{z-\mu}{\sigma}\right)>0 \\ 0 & 1+\xi\left(\frac{z-\mu}{\sigma}\right)\leq 0\end{cases} \quad (1)$$

The generalized extreme value distribution as it shown in Equation 1 has three parameters, μ as location, $\sigma > 0$ as scale, and ξ as shape parameter. The location and scale parameters are the center and extension of distribution respectively. The shape parameter, ξ is described for the tail of the distribution, which is based on its value, three types of extreme value distribution is formed as below:

If $\xi > 0$, then the tail is large and the domain of attraction is Frechet.

If $\xi = 0$, then the tail is short and the domain of attraction is Gumball.

If $\xi < 0$, then the distribution function is limited and the domain of attraction is Weibull.

The most important application of extreme value distribution is the calculation of the return period for the occurrence of random phenomena such as earthquakes, flood and etc. The return level is related to return period by

$$G(x_T)=1-\frac{1}{T}\Rightarrow x_T=\mu-\frac{\sigma}{\xi}\left[1-\left\{-\log\left(1-\frac{1}{T}\right)\right\}^{-\xi}\right] \quad (2)$$

where x_T and T are return level and return period, respectively.

2. APPLICATION OF EXTREME VALUE THEORY IN HAZARD ASSESSMENT

Extreme value theory in hazard assessment used in estimating the maximum magnitude of earthquakes by Nordquist (1945), calculation of the probability for earthquakes occurrence by Gumbel (1958), the occurrence of large earthquakes by Epstein and Lomnitz (1966) and earthquake risk by Lomnitz (1974). Almost in these studies, parameters of seismic zones are determined by extreme value theory. For example, the probability of earthquakes occurrence with magnitude smaller than M_i is as Gumbel cumulative distribution function:

$$P(M \leq M_i)=G(y)=\exp\{-\alpha\exp(-\beta\,\mathrm{y})\} \quad (3)$$

In Equation 3 α and β are related to a and b by equation 4 that are calculated by frequency-magnitude Gutenberg- Richter relationship as Figure1.

$$\alpha=\exp[a\ln 10] \quad , \quad \beta=\exp[b\ln 10] \quad (4)$$

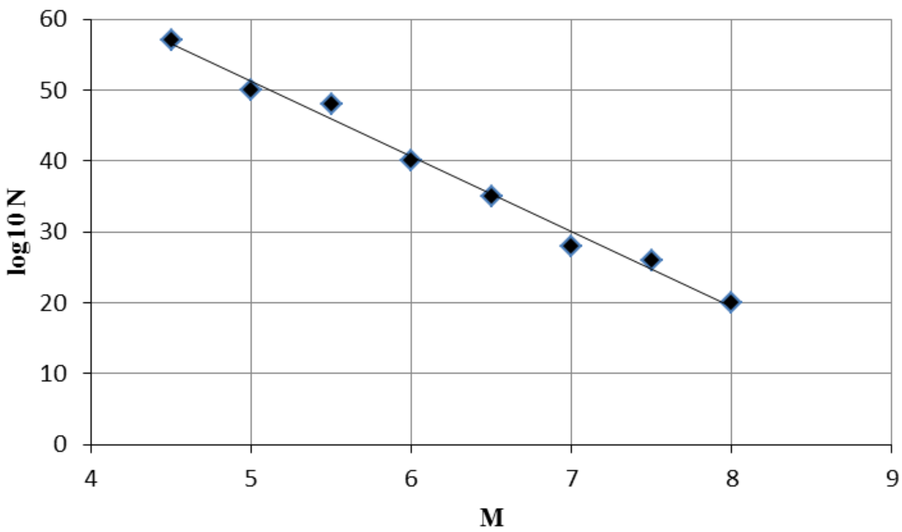

Figure 1. Calculation of a and b parameters by Gutenberg- Richter relationship

Also, in developing ground motion prediction equation, regression process has been done based on log- normality of ground motion. But since these values of ground motion are the maximum type, it is suggested to use extreme value distribution as alternative distribution (Dupuis and Flemming 2006, Raschke 2013, Pavlenko 2015). A major drawback of previous research is that these papers basic assumptions of extreme value theory for risk assessment. One of these assumptions is selecting blocks maxima, which is not presented by the algorithm in their seismic models. Below is a way of selecting block maxima to develop Ground Motion Prediction Equation (GMPE) by extreme value theory.

3. BLOCK MAXIMA SELECTION FOR GMPE

Attenuation models in risk assessment literature review presented for peak ground acceleration, peak ground velocity, and pseudo-spectral acceleration. PGA, which has the most application in seismic design, means the maximum acceleration value in a recorded accelerogram for each earthquake (Figure 2).

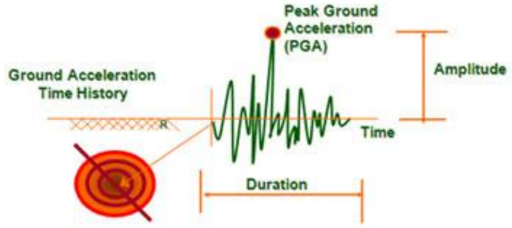

Figure 2. Peak ground acceleration in the accelerogram (tangri et al 2009)

In developing attenuation relationships for peak ground acceleration, PGA, various records are considered. The histogram of PGAs is skewed and for convenience calculations and statistical habits to use normal distribution, LOG conversion to use normal regression will be down. Since PGA is a maximal concept, extreme value theory usage for extreme values is suggested. But the main thing is choosing these maxima for regressions. The main problems in this direction are choosing blocks and block maxima. There are no specific parameters in the statistical literature for selecting blocks, but it is advised that the selected blocks have the same size.

The selection of the block size is also important, and in choosing it there should be a balance between the variance and the bias so that too small blocks lead to bias estimates of the parameters, and if the size of the blocks is large, it leads to Reducing the number of blocks and thus decrease the number of maxima, which results in the estimation of larger variance values.

As is seen in Figure 2, each earthquake and accelerogram has a duration of time depending on the magnitude of the earthquake. Therefore, selecting the peak ground acceleration values of each accelerogram, and placing it together to fit the extreme model, is not correct due to the time difference between each earthquake based on extreme value theory. It is therefore recommended intended PGAs will be chosen by accelerogram with almost equal duration. Therefore, the selection of PGAs should be based on accelerograms of a seismic event in which the duration of all accelerograms is almost same. In addition, the use of PGAs of a seismic event has another positive effect, and that is more accurate in modeling the prediction equation of ground motion. Accordance with results by Raschke (2013) use of event-specific attenuation relationships provides more precise prediction results for GMPE models.

4. DISTRIBUTION ASSESSMENT FOR PGA DATA OF CHI-CHI 1999 EARTHQUAKE

PGAs histogram and fitted distributions of Chi-Chi earthquakes are compared in Figure 3. As it seen in table1, the maximum likelihood function for GEV distribution is larger than other distributions. GEV distribution parameters for Chi- Chi PGAs data are as below:

$$\mu = 0.077 \qquad \sigma = 0.05 \qquad \xi = 0.48$$

Table 1.The log-likelihood function for different distributions

distribution	Log-likelihood
Lognormal	266.632
GEV	270.49
Gamma	242.944
Weibull	233.799
Inverse Gaussian	268.668
Exponential	224.974

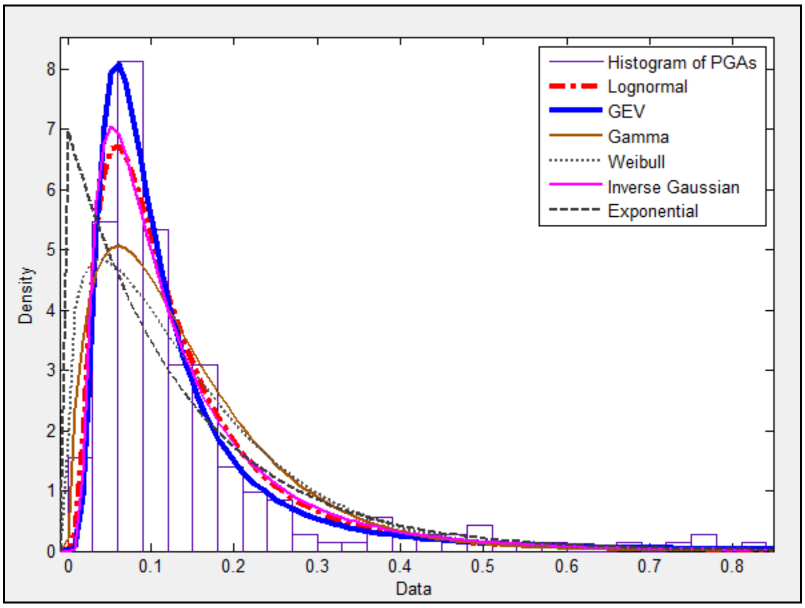

Figure 3. PGAs histogram and fitted distributions of Chi- Chi earthquake

Accordingly, the best distribution for the given data is GEV distribution. So must be better to use GEV distribution for PGA data of Chi- Chi earthquake.

This issue is done for twelve other earthquakes, and in other to compare GEV to lognormal distributions. Accordingly, it can be seen that in near-source earthquakes the GEV distribution have a better match with PGAs data, but for far source earthquakes, the lognormal distribution is better. It is related to the much more recorded amounts of PGAs, much more critical data in near source than far source. As is seen in Figure 4, in this case, the GEV distribution has shown the more appropriate level of hazard than lognormal distribution. In PGAs upper than 0.3g, survival function of GEV distribution is more than lognormal distribution. In fact GEV distribution shows more accurate of hazard levels in larger numbers of PGAs.

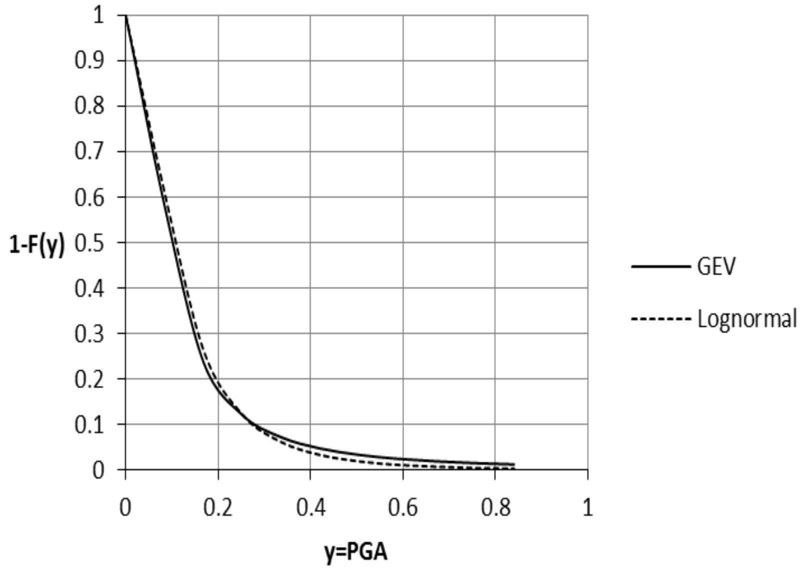

Figure 4.Survival plot for GEV and lognormal distributions of Chi-Chi PGAs.

Table 2. The value of the Log-likelihood function resulted from fitting GEV and lognormal distributions for different earthquakes

Type	**Earthquakes**	**Log Likelihood function**	
		GEV	**Lognormal**
Near source	Chi-Chi (1999)	270.49	266.632
	Iwate (2008)	113.79	112.595
	Northridge (1994)	99.4424	97.763
	10410337 (2009)	234.485	238.577
	Loma Prieta (1989)	56.1837	56.9091
	Whittier Narrows-01 (1987)	117.565	120.73
far source	Chuetsu-Oki (2007)	1701.38	1712.57
	Nigata (2004)	1449.42	1472.01
	Tottori (2000)	977.435	984.802
	El Mayor (2010)	850.696	861.236
	Iwate (2008)	619.564	626.416
	Chi-Chi (1999)	335.239	331.034

5. CONCLUSIONS

In this paper, suitable distribution for PGAs is introduced, based on extreme value theory. The block for maxima to select maxima values is selected as accelerograms of each earthquake, to have the almost same size of blocks. The results of twelve earthquakes maximum likelihood function indicate that the GEV distribution is more suitable for near-source data than for far source earthquakes. This issue, while confirming the validation of the block selection method, indicates that for near-source earthquake PGAs, the GEV distribution is better than lognormal. Since lognormal distribution is used in the preparation of all GMPEs, and GEV precision is higher than lognormal in fitting on PGA data; it can be proposed for developing GMPE of near-source earthquake PGAs, GEV regression is more suitable. But it should be noted that this distribution leads to developing a GMPE only for each earthquake, which is event-specific GMPE. This suggestion is being researched by the authors.

6. REFERENCES

Dupuis DJ, Flemming JM, (2006). Modeling peak acceleration from earthquakes, *Earthquake Engineering and Structural Mechanics*, 35: 969-987.

Epstein B, Lomnitz C, (1966). A model for the occurrence of large earthquakes, *Nature*, 211:954-956.

Fisher RA, Tippett LHC, (1928). Limiting forms of the frequency distributions of largest or smallest member of a sample. Proc. Cambridge Philos. Soc, 24: 180-190.

Gumbel EJ, (1958), Statistics of Extremes, Columbia University Press, New York.

Lomnitz C, (1974), Global Tectonics and Earthquake Risk, *Elsevier*.

Nordquist, JM, (1945). Theory of largest values applied to earthquake magnitudes, *Trans. Amer. Geophysics. Union*, 26: 29-31.

Pavlenko VA, (2015). Effect of alternative distributions of ground motion variability on results of Probabilistic Seismic Hazard Analysis, *Natural Hazards*, 78(3):1917–1930.

Raschke M, (2013). Statistical modeling of ground motion relations for seismic hazard analysis. *J Seismology*, 17(4):1157–1182.

Tangri R, Jena S, Roy S, (2009). Earthquake disaster management using GIS and probabilistic risk assessment. *Geospatial world*.