

# Uncertainty analysis of strong earthquake hazard estimation based on the generalized Pareto distribution model: a case study of the northeastern Tibetan Plateau

Mengyi Ren

Institute of Geophysics, China Earthquake Administration, Beijing, China

Article history: received November 30, 2020; accepted July 12, 2021

## Abstract

A statistical method to analyze the uncertainties of strong earthquake hazard estimation is proposed from Generalized Pareto distribution (GPD) model using the northeastern Tibetan Plateau earthquake catalogue (1885–2017) data. For magnitude threshold of 5.5, the magnitude return levels in 20, 50, 100, 200, and 500 years are 7.19, 7.70, 7.99, 8.22, and 8.45, respectively. The corresponding 95% confidence intervals are [6.77, 7.60], [7.27, 8.12], [7.53, 8.44], [7.71, 8.72], and [7.84, 9.06], respectively. The upper magnitude limit obtained from this GPD model is 9.07 and its 80% confidence interval is [8.16, 9.98]. The sensitivity analysis by the Morris method indicates that the input magnitude threshold has a relatively large influence on the estimation results. Thus, threshold selection is important for the GPD model construction. The sensitivity characteristic ranking of input factors become increasingly stable with the increasing of return period, which implies that GPD model is more suitable for estimating strong earthquakes magnitude return levels and upper magnitude limit. The GPD modeling approach and qualitative uncertainty analysis methods for strong earthquake hazard estimations proposed in this paper can be applied to seismic hazard analysis elsewhere.

Keywords: Strong earthquake hazard estimation; Generalized Pareto Distribution; Uncertainty analysis; Northeastern Tibetan Plateau.

---

## 1. Introduction

Seismic activity models are often used to describe spatial, temporal and intensity distributions of earthquakes. This model forms the basis for seismic zoning and hazard analysis [Jiang and Dai, 1993]. The improvement of seismic activity models is thus an ongoing focus for seismologists. Generalized Pareto distribution (GPD) model from extreme value theory is developed in recent years, which uses historical large magnitude earthquake record. It assumes that magnitude distribution of large earthquakes conforms to the GPD and the earthquake occurrence

follows Poisson distribution over time. Pisarenko and Sornette [2003] used GPD model to analyze shallow earthquake distribution in 18 seismic regions from the Harvard earthquake catalogue. Qian et al. [2013] chosed GPD model to analyze the magnitude distribution of strong earthquakes in boundary zones of active blocks in China. Pisarenko et al. [2014] proposed a method to estimate the unknown parameters involved in the two limit theorems corresponding to the generalized extreme value distribution and GPD. They applied this method for the global Harvard and Fennoscandia catalogues. Tian et al. [2017] estimated the seismic hazard of potential earthquake tsunami source areas in the Manila subduction zone according to the GPD model. Ren [2018] used a GPD model to obtain the upper magnitude limit of earthquake in the Longmenshan region. Dutfoy [2019] combined the GPD and Poisson distribution to analyze the tail characteristics of the earthquake magnitude distribution in southern France and calculated the annual maximum earthquake magnitude there. GPD seismicity model can take full advantage of large earthquake data in a catalogue. There is no need to designate a lower earthquake magnitude limit a priori when processing an earthquake catalogue. It is suitable for areas with long historical records of big earthquakes.

However, it must be emphasized that the uncertainties of strong earthquake hazard estimation are inevitable when constructing a GPD model. Two input factors of the model should be considered: the earthquake catalog start time and magnitude threshold. The selection of these two model input factors is somewhat subjective. Dutfoy [2019] used a bootstrap statistical method to obtain the GPD model parameter distribution. He discussed a particular case where the probability density distribution of the GPD model shape parameter is bimodal, which may introduce uncertainties in the seismic hazard estimation. Nevertheless, the analysis of specific situation in which the seismicity model input parameters change independently is not enough but rather a systematic analysis of two situations where the input parameters change simultaneously and the input parameters interact with each other. Thus, the global sensitivity analysis method should be used when discussing the uncertainties. Morris method, one of the global sensitivity analysis method, as a process by which the uncertainty of strong earthquake hazard estimations can be analyzed. This method is more computationally efficient and it provides generally qualitative sensitivity measures, which is capable of ranking the input factors in order of importance.

In this paper, GPD is adopted to construct a strong earthquake activity model and strong earthquake hazard estimation (magnitude return level and upper magnitude limit and their corresponding confidence intervals) are obtained by using historical seismic data of large earthquakes from 1885 to 2017. The Morris method is used to qualitatively analyze the influence of input factor (earthquake catalog start time and magnitude threshold) uncertainties and the interaction of factors on the earthquake hazard estimation uncertainty. This approach is also applicable in seismic zoning and hazard analysis.

## 2. Method to model the GPD

Consider the model set  $X_1, X_2, \dots, X_n$  to be a sequence of independent random variables having a common distribution function. If there exist sequences of constants  $\{a_n > 0\}$  and  $\{b_n\}$  such that

$$P_r\{(M_n - b_n) / a_n \leq x\} \rightarrow G\{x\}, n \rightarrow \infty \quad (1)$$

where  $G$  is a nondegenerate distribution function, then  $G$  is called the extreme value distribution. Fisher and Tippett [1928] indicated the three major theorems of the extreme value distribution, pointing out that  $G$  must belong to one of the three extreme value distributions, namely, the Gumbel distribution, Fréchet distribution or Weibull distribution. The generalized extreme value theory unifies these three distribution types into a single family of models, having distribution functions of the form:

$$G(x) = \exp \left\{ - \left[ 1 + \xi \left( \frac{x-\mu}{\sigma} \right) \right]^{-1/\xi} \right\}, 1 + \xi(x-\mu) / \sigma > 0 \quad (2)$$

which is represented by three parameters, namely, the position parameter  $\mu$ , scale parameter  $\sigma$  and shape parameter  $\xi$  [Coles, 2001; Shi, 2006].

Then, for a large enough threshold  $u$ , the distribution function of  $(X - u)$ , conditional on  $X > u$ , is approximately obeys the GPD:

$$H(y) = 1 - \left(1 + \frac{\xi y}{\tilde{\sigma}}\right)^{-1/\xi} \quad (3)$$

defined on  $\{y: y > 0 \text{ and } (1 + \frac{\xi y}{\tilde{\sigma}}) > 0\}$ , where  $\tilde{\sigma} = \sigma + \xi(u - \mu)$ .

Two methods are available to select a reasonable threshold that provides a reasonable approximation for the model. One is an exploratory technique carried out prior to the model estimation, which means that the magnitude excess and magnitude threshold should satisfy a linear relationship in a plot of the magnitude mean excess distribution. The other method is a stability assessment of the parameter estimates based on the model fitting across a range of different thresholds. The shape parameter and modified scale parameter estimates should remain fairly constant with increasing threshold.

Suppose that a GPD with parameters  $\tilde{\sigma}$  and  $\xi$  is a suitable model for exceedances of a threshold  $u$  by a variable  $X$ . That is, for  $x > u$ ,

$$P_r\{X > x | X > u\} = \left[1 + \xi \left(\frac{x-u}{\tilde{\sigma}}\right)\right]^{-1/\xi} \quad (4)$$

It follows that,

$$P_r\{X > x\} = \zeta_u \left[1 + \xi \left(\frac{x-u}{\tilde{\sigma}}\right)\right]^{-1/\xi} \quad (5)$$

Where  $\zeta_u = P_r\{X > u\}$ . Hence, the level  $x_m$  that is exceeded on average once every  $m$  observations is the solution of

$$\zeta_u \left[1 + \xi \left(\frac{x_m-u}{\tilde{\sigma}}\right)\right]^{-1/\xi} = \frac{1}{m} \quad (6)$$

Rearranged,

$$x_m = u + \frac{\tilde{\sigma}}{\xi} [(m\zeta_u)^\xi - 1] \quad (7)$$

$x_m$  is the  $m$ -observation return level. The  $N$ -year return level is the level expected to be exceeded once every  $N$  years. If there are  $n_y$  observations per year, this corresponds to the  $m$ -observation return level, where  $m = N \times n_y$ . Hence, the  $N$ -year return level is defined by

$$x_N = u + \frac{\tilde{\sigma}}{\xi} [(Nn_y\zeta_u)^\xi - 1] \quad (8)$$

The parameters of the GPD can be estimated by maximum likelihood. Only negative values of the form parameter  $\xi$  are acceptable, which give the following estimation of upper limit value:

$$\hat{x} = u - \hat{\sigma} / \hat{\xi} \quad (9)$$

Standard errors or confidence intervals for  $\hat{x}$  can be derived by the delta method [Coles, 2001], that is, the approximate  $(1 - \alpha)$  confidence interval for the return level or upper limit would be:

$$\hat{x} \pm Z_{1-\frac{\alpha}{2}} \sqrt{\widehat{Var}(\hat{x})} \quad (10)$$

where  $\widehat{Var}(\hat{x}) \approx \nabla x^T V \nabla x$ ,  $V$  is the variance-covariance matrix of  $(\hat{\zeta}_u, \hat{\sigma}, \hat{\xi})$ ,  $\nabla x^T$  is the value of  $\left(\frac{\partial x}{\partial \zeta_u}, \frac{\partial x}{\partial \sigma}, \frac{\partial x}{\partial \xi}\right)$  evaluated at  $(\hat{\zeta}_u, \hat{\sigma}, \hat{\xi})$  and  $Z_{1-\frac{\alpha}{2}}$  is the  $1-\frac{\alpha}{2}$  quantile of the standard normal distribution.

### 3. Qualitative method of uncertainty analysis

The Morris method [Morris, 1991] is composed of individually randomized ‘one-factor-at-a-time’ experiments and evaluates the impact of changing one factor at a time. Each input factor may assume a discrete number of values, called levels, which are chosen within the factor variation range.

Assume that the model output factor is  $y = f(x_1, x_2, \dots, x_k)$  and  $x = (x_1, x_2, \dots, x_k)$  is an input factor with  $k$  components. Combining the probability distribution of each input factor, the Morris method maps the value range of each input factor into the interval  $[0, 1]$  and discretizes it into  $p$  levels. The Morris design begins by randomly selecting a ‘base’ value  $x^*$ . Each component  $x_i$  of  $x^*$  is sampled from  $\{0, 1/p-1, 2/p-1, \dots, 1\}$  where  $\Delta = 1/p-1$  is the predetermined amount of change. The first sampling point,  $x^{(1)}$ , is obtained by increasing one or more components of  $x^*$  by  $\Delta$ . The second sampling point is generated from  $x^*$  with the distinction that it differs from  $x^{(1)}$  in its  $i^{th}$  component, which has been either increased or decreased by  $\Delta$ . The design then produces a succession of  $(k+1)$  sampling points  $x^{(1)}, x^{(2)}, \dots, x^{(k+1)}$  with the key property that two consecutive points differ in only one component. The succession of sampling points defines what is called a trajectory in input space. Once a trajectory has been constructed and the model evaluated at its points, an elementary effect for each factor  $i$ ,  $i = 1, 2, \dots, k$ , can be computed. The elementary effect of the  $i^{th}$  input factor is defined as:

$$d_i(x) = \frac{[y(x_1, \dots, x_{i-1}, x_i + \Delta, x_{i+1}, \dots, x_k) - y(x)]}{\Delta} \quad (11)$$

The mean and standard deviation  $\sigma$  of the  $r$  elementary effects are defined as:

$$\mu = \sum_{i=1}^r d_i / r \quad (12)$$

$$\sigma = \sqrt{\sum_{i=1}^r (d_i - \mu)^2 / r} \quad (13)$$

Two sensitivity measures are proposed by Morris for each input factor: (1)  $\mu$ , which estimates the overall effect of the factor on the output and (2)  $\sigma$ , which estimates the ensemble of the second- and higher-order effects in which the factor is involved [Saltelli et al., 2004]. The absolute mean value of the elementary effect,  $\mu^*$ , is used instead of  $\mu$  because it is considered a more robust sensitivity metric, especially for the case of nonmonotonic functions [Campolongo et al., 2007]. The Euclidean distance between  $(\mu^*, \sigma)$  and the origin can also be used for parameter ranking [Pappas et al., 2013]. The key input factors of the GPD model can be determined based on a scatter plot of  $\mu^*$  and  $\sigma$  or their Euclidean distance for the output variables.

The method relies on the elementary effect that uses incremental ratios and is apparently a local measure. However, the final measures,  $\mu^*$  and  $\sigma$ , are calculated by averaging several elementary effects, which are obtained by the different trajectories. The first sampling point of the trajectory has randomness and the method can thus be regarded as global.

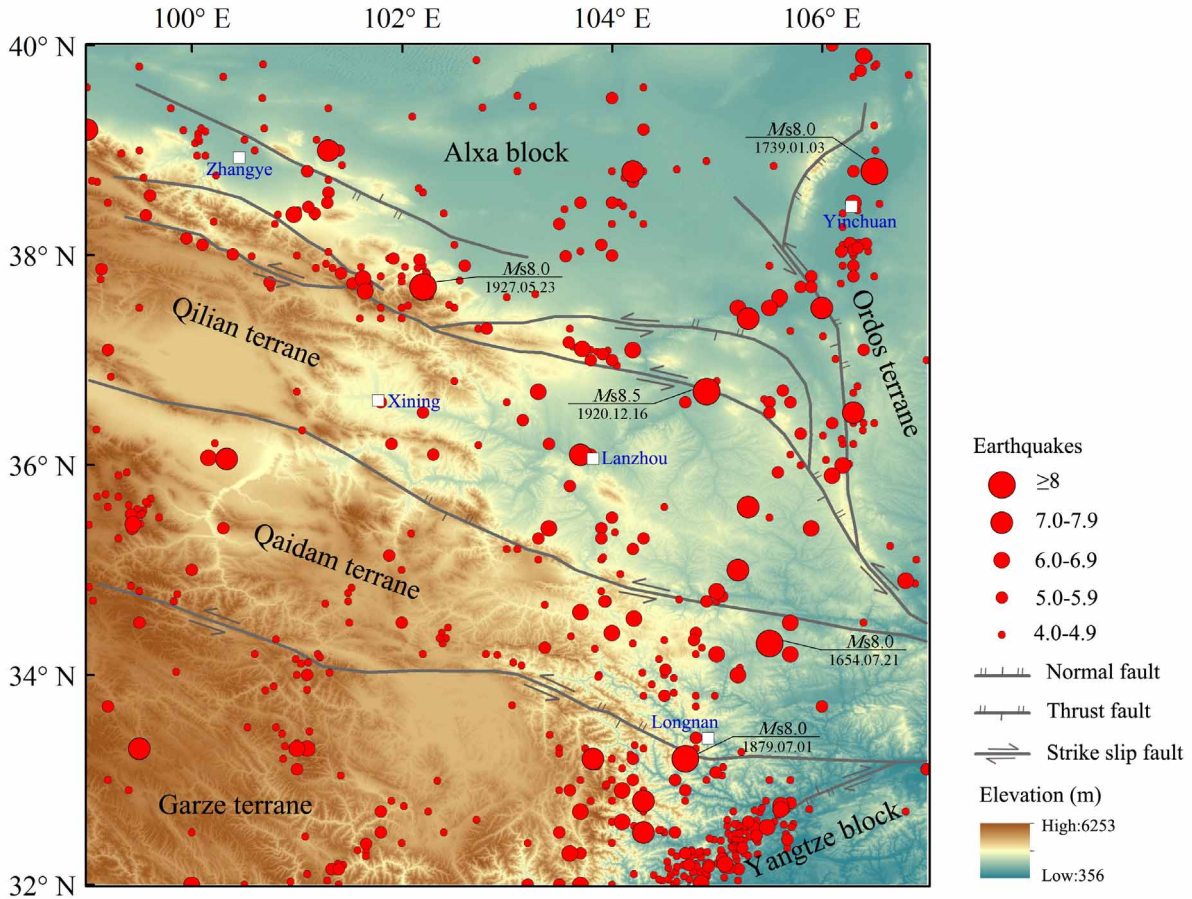
A large  $\mu^*$  value indicates that the factor has a strong effect on the output. A large  $\sigma$  value implies that either the factor is interacting with other factors or has nonlinear effects on the output. In contrast, a small  $\sigma$  value indicates very similar values of the elementary effects, implying that the effect of this factor is almost independent of the values obtained from the other factors. If  $\mu^*$  and  $\sigma$  are both large, the associated factor has a strong influence on the output involving nonlinearity or interactions; if the case is the opposite, the input factor has a negligible influence on the output [Ren et al., 2018].

## 4. Case study: Northeastern Tibetan Plateau

### 4.1 Seismic and geological background

The northeastern Tibetan Plateau is located at the intersection of the Qinghai-Tibetan Plateau block with the Ordos block and Alxa block. The present tectonics of northeastern Tibet resulted from SW–NE compression induced

by the collision between India and Eurasia, which initiated ~50 Ma [Taylor and Yinn, 2009; Yin and Harrison, 2000]. Recent global positioning system (GPS) measurements [Gan et al., 2007] show that approximately 10 mm/yr of NE–SW or NNE–SSW horizontal shortening is accommodated in northeastern Tibet. The intense tectonic activity and frequent seismic activities in this region have thus been a hotspot of geoscience research for many years [Zhao et al., 2004; Chen et al., 2005; Li et al., 2006; Chang et al., 2008; Tian and Zhang, 2013; Bao et al., 2013; Li et al., 2017; Zhu et al., 2018; Sun et al., 2019].



**Figure 1.** Map of main active faults and earthquake distribution ( $M_s \geq 4$ ) in northeastern Tibetan Plateau (according to Li et al. [2006] revised).

Five strong earthquakes with a magnitude greater than or equal to 8 are found in the historical record: Gulang ( $M_s$  8.0) in 1927; Haiyuan ( $M_s$  8.5) in 1920; Wudu ( $M_s$  8.0) in 1879; Pingluo ( $M_s$  8.0) in 1739; and Tianshui ( $M_s$  8.0) in 1654 (Figure 1). Huang et al. [1994] conducted a detailed analysis on the historical seismic data of most parts of mainland China and showed that the seismic data with  $M_s \geq 5$  from the northeastern Tibetan Plateau are essentially complete since 1885.

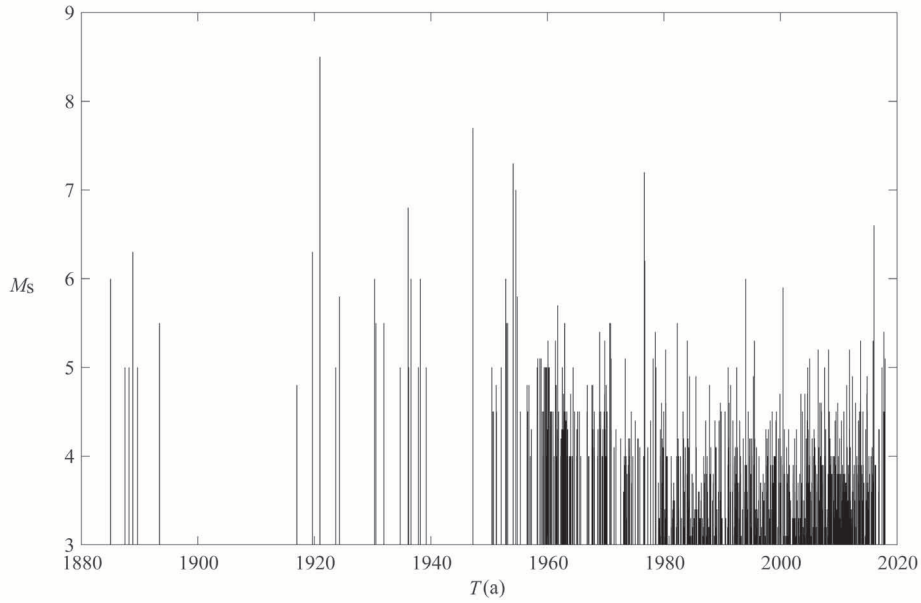
The data used in this study are mainly from the Chinese historical earthquake catalogue database and the National Seismic Network Earthquake Catalogue Database of the China Earthquake Data Center (<http://data.earthquake.cn>). The historical earthquake catalogue of the northeastern Tibetan Plateau region (99–107°E, 32–40°N) between 1100 and 2017 contains a total of 834 recorded earthquakes with a magnitude greater than 4. According to the collected seismic data, the northeastern Tibetan Plateau region is dominated by shallow earthquakes, including 5 earthquakes with  $M_s \geq 8$ , 15 with  $8 > M_s \geq 7$ , 50 with  $7 > M_s \geq 6$ , and 173 with  $6 > M_s \geq 5$  (Figure 1).

Aftershocks in the earthquake catalogue should be efficiently removed before using the model for seismic hazard estimation. Seismic hazard analysis from studies worldwide have widely adopted the space-time window method to remove aftershocks to account for the calculation simplicity and engineering operability.

In this method, the spatial window is determined according to:

$$\lg R = aM + b \tag{14}$$

where  $a$  and  $b$  are fixed parameter values, which for China are equal to 0.5 and  $-1.78$ , respectively [Liu et al., 1996].



**Figure 2.**  $M$ - $T$  map of the earthquakes from 1885 to 2017 in northeastern Tibetan Plateau ( $M_s \geq 4$ ).

The temporal window is shown in Table 1. After removing the aftershocks using the spatiotemporal window method, 560 earthquakes with a magnitude greater than 4 were collected from the northeastern Tibetan Plateau region.  $M$ - $T$  map of the earthquakes with  $M_s \geq 4$  from 1885 to 2017 is shown in Figure 2 according to the earthquake catalogue data after aftershocks deleting. Since 1960, with the update of seismic monitoring instruments, more and more small magnitude earthquakes have been fully detected. In this paper, earthquakes used in GPD model are greater than magnitude threshold, so the lack of small magnitude earthquakes in historical catalogue will not have influence on the seismic hazard estimation results.

Main shock/ $M_s$	Lasting days/d	Main shock/ $M_s$	Lasting days/d
4	42	6.5	790
4.5	83	7	915
5	155	7.5	960
5.5	290	8	985
6	510	8.5	985

**Table 1.** Aftershock time windows of different magnitudes [Console et al., 1979; Chen et al., 1998].

### 4.2 Strong earthquake hazard estimation

When constructing a GPD model of strong earthquakes, the magnitude threshold  $u$  is first selected according to the method introduced in section 2, and then the magnitude greater than the threshold  $u$  in the selected historical

seismic records in a certain period is taken as random variables. It is assumed that the magnitude satisfies the condition of independent and identically distribution and the magnitude excess conforms to the GPD. Assuming that there are  $n$  historical seismic record data in the selected  $t$  years, the  $n_y$  observation per year can be obtained by  $n/t$ , put it into Equation (8) to calculate the magnitude return level in the  $N$  years. The upper magnitude limit is calculated according to Equation (9). Using the delta method, the confidence intervals of the magnitude return levels and the upper magnitude limit can be calculated according to Equation (10).

This study analyzes the earthquake catalogue data between 1885 and 2017 to estimate the magnitude return level and upper magnitude limit of the northeastern Tibetan Plateau region using the GPD model. Figure 3 shows the mean excess distribution plot with an approximate 95% confidence interval for the earthquake data. The mean magnitude excess is approximately linearly distributed when  $u$  is in the range of (5.5, 6.0). To make the most reasonable selection of  $u$ , it is necessary to also consider the actual situation of the sample data. Too low of a threshold is likely to violate the asymptotic basis of the model, leading to bias; too high of a threshold will generate few excesses with which the model can be estimated, leading to high variance. In conjunction with Figure 4, the estimated values of the shape parameter and modified scale parameter in the range of (5.5, 6.0) remain relatively unchanged. The threshold  $u = 5.5$  is therefore selected.

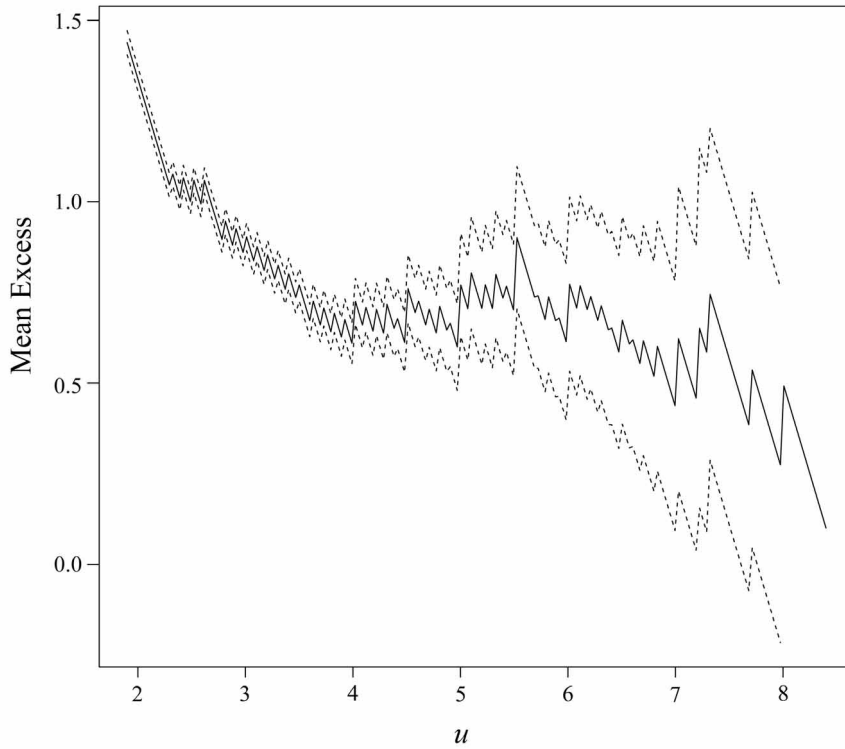


Figure 3. Plot of magnitude mean excess distribution.

The maximum likelihood estimation is used to estimate the GPD model parameters. The estimated values of the modified scale parameter and shape parameter are 1.23 and  $-0.35$ , respectively. Diagnostic plots for the fitted GPD model are shown in Figure 5. The probability and quantile plots indicate the fit between the model and data. The goodness-of-fit in the probability and quantile plots is highly convincing. The curve in the return level plot is convex and asymptotically approaches a finite value because the estimated value of the shape parameter is negative, which confirms that the upper magnitude limit of the northeastern Tibetan Plateau region can be obtained. The corresponding density estimate is generally consistent with the histogram of the data. All four diagnostic plots therefore support the fitted GPD model established in this study.

The estimated magnitude return levels for each return period and their corresponding confidence intervals with a 95% confidence coefficient can be obtained using the magnitude threshold  $u$  and estimated values of  $\tilde{\sigma}$  and  $\xi$  in Equation (8) (Table 2). Following Equation (9), the upper magnitude limit of the northeastern Tibetan Plateau region is 9.07.

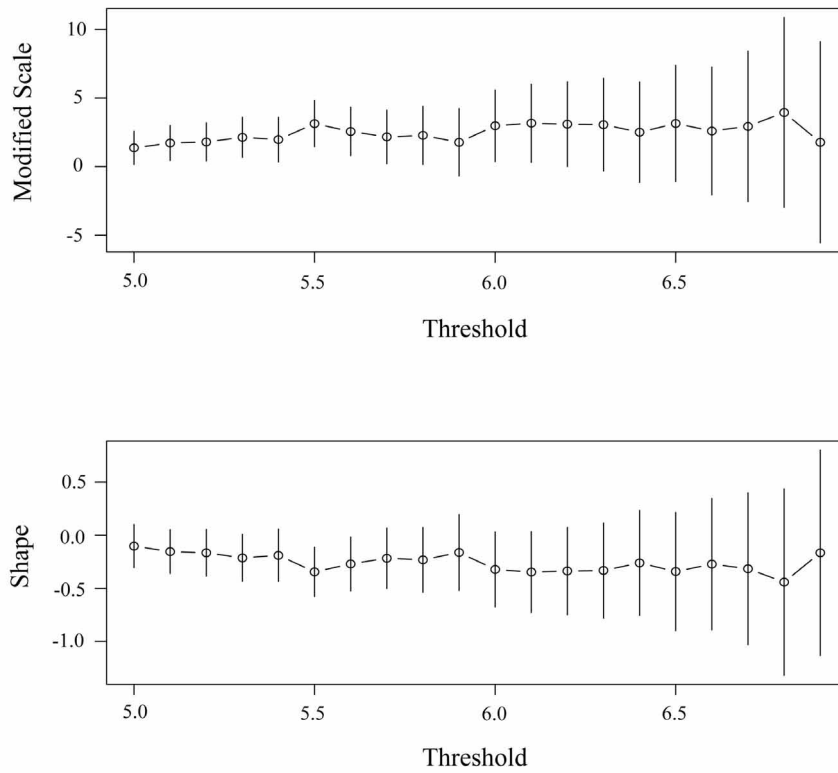


Figure 4. Parameter estimates against threshold for earthquake magnitude data.

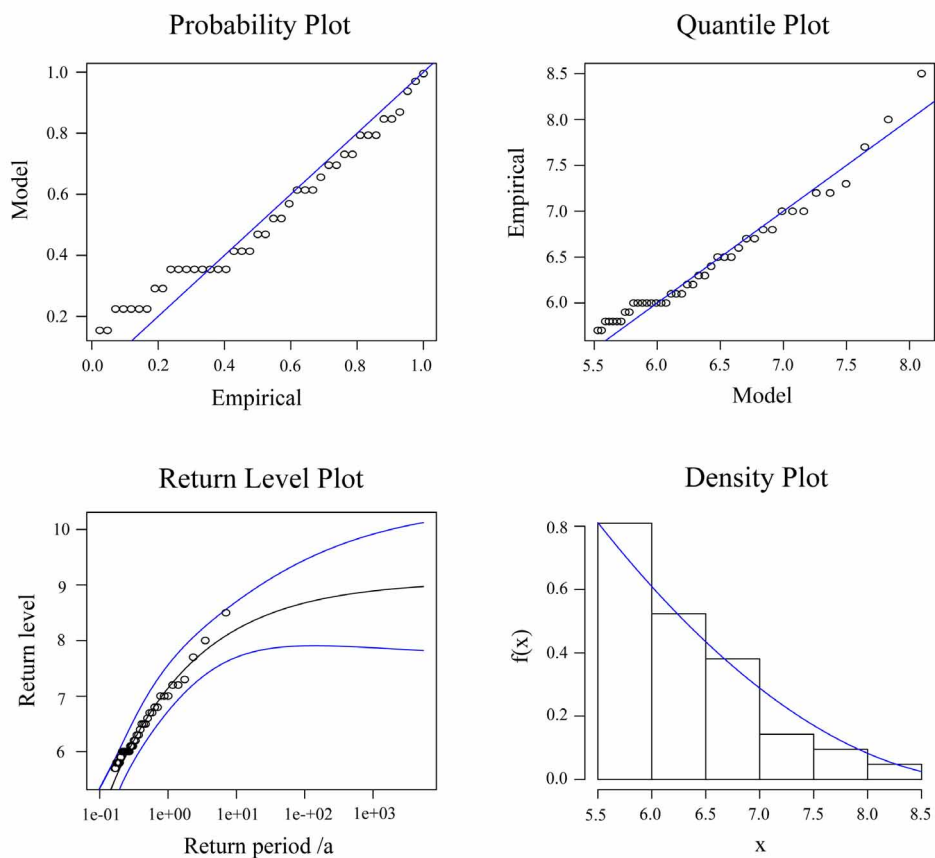


Figure 5. Diagnostic plots for threshold excess model fitted to magnitude data (a) Probability plot; (b) Quantile plot; (c) Return level plot; (d) Probability density plot.



The standard deviation of the upper magnitude limit is 0.71 and the confidence interval with an 80% confidence coefficient of the upper magnitude limit is [8.16, 9.98].

Return period /a	20	50	100	200	500
Magnitude	7.19	7.70	7.99	8.22	8.45
Confidence interval	[6.77, 7.60]	[7.27, 8.12]	[7.53, 8.44]	[7.71, 8.72]	[7.84, 9.06]

**Table 2.** Estimation of magnitude return level for northeastern Tibetan Plateau.

### 4.3 Uncertainty analysis of earthquake hazard estimation

The input factors of GPD model are earthquake catalog start time ( $st$ ) and magnitude threshold ( $u$ ). The outputs are magnitude return level from different periods, upper magnitude limit, and their corresponding confidence intervals. The range of start times for earthquake catalog in the northeastern Tibetan Plateau is based on a study of the seismic data integrity [Huang et al., 1994]. The range of magnitude threshold values is considered in the two methods discussed above to select the threshold (Section 2). Supposing that the range of  $st$  values is between 1880 and 1890 and the range of  $u$  values is between 5.5 and 6.0, the two input factors both follow a uniform distribution pattern, as shown in Table 3.

Input factors	Probability distribution
Earthquake catalog start time ( $st$ )	$st \sim U(1880, 1890)$
Magnitude threshold ( $u$ )	$u \sim U(5.5, 6.0)$

**Table 3.** Probability distribution of input factors.

The parameters of the Morris experiment are set to 6 levels and 10 trajectories. By using the Monte Carlo sampling method, 30 sets of GPD model parameters are obtained and the sampling results are listed in Table 4.

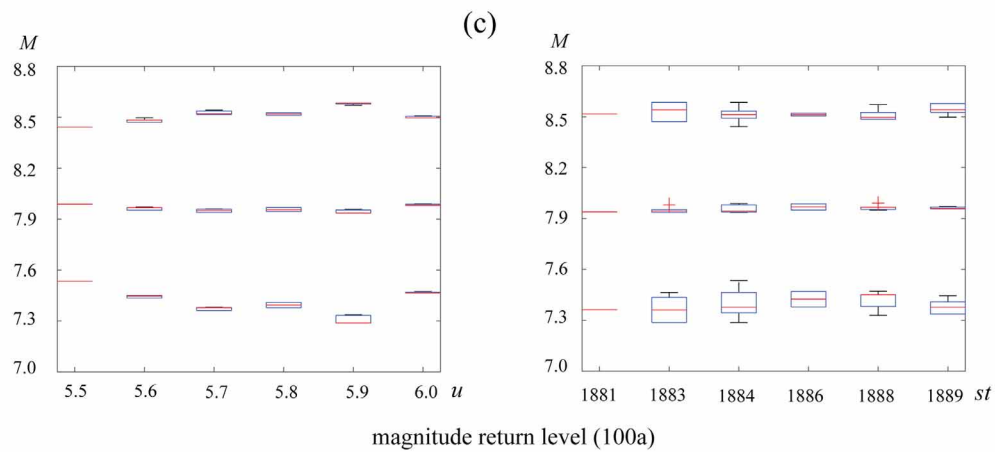
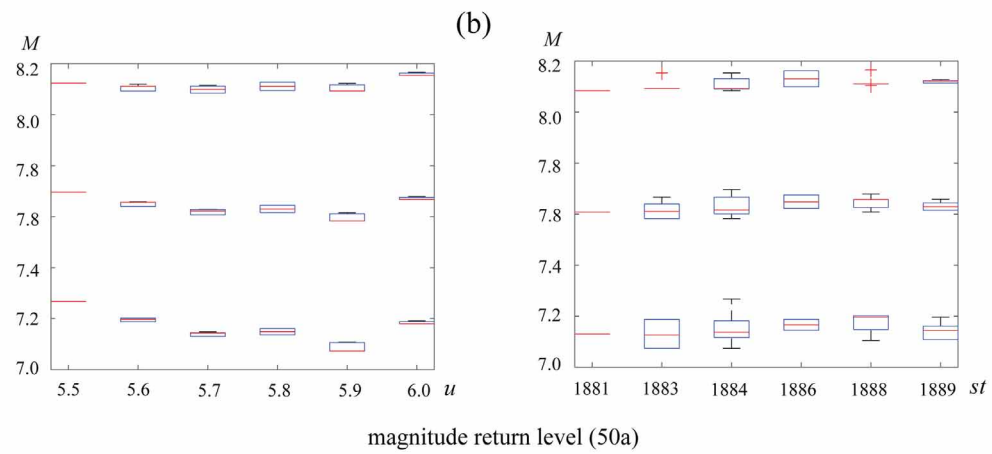
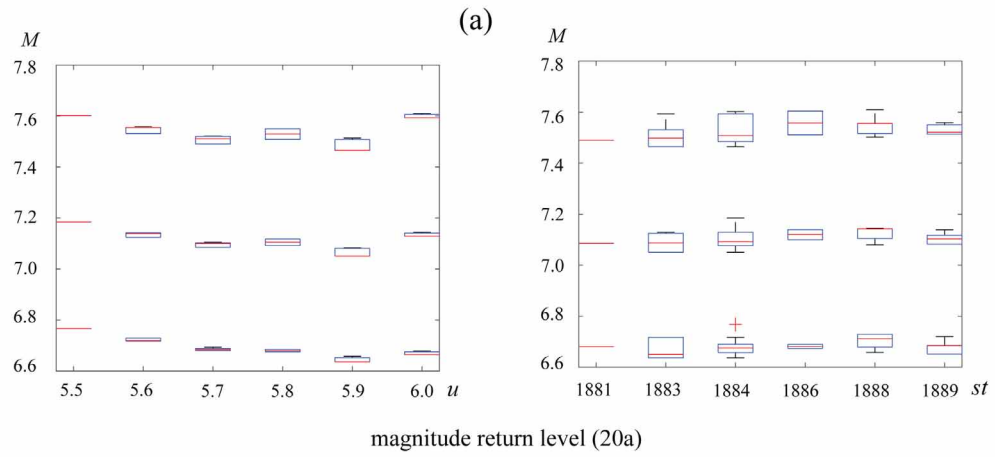
Trajectory	Samples		
1	(5.7,1888)	(6.0,1888)	(6.0,1883)
2	(5.7,1889)	(5.7,1884)	(6.0,1884)
3	(5.9,1883)	(5.9,1888)	(5.6,1888)
4	(5.6,1888)	(5.6,1883)	(5.9,1883)
5	(5.6,1888)	(5.6,1883)	(5.9,1883)
6	(5.9,1884)	(5.9,1889)	(5.6,1889)
7	(5.8,1889)	(5.8,1884)	(5.5,1884)
8	(5.7,1889)	(5.7,1884)	(6.0,1884)
9	(5.6,1884)	(5.9,1884)	(5.9,1889)
10	(6.0,1886)	(5.7,1886)	(5.7,1881)

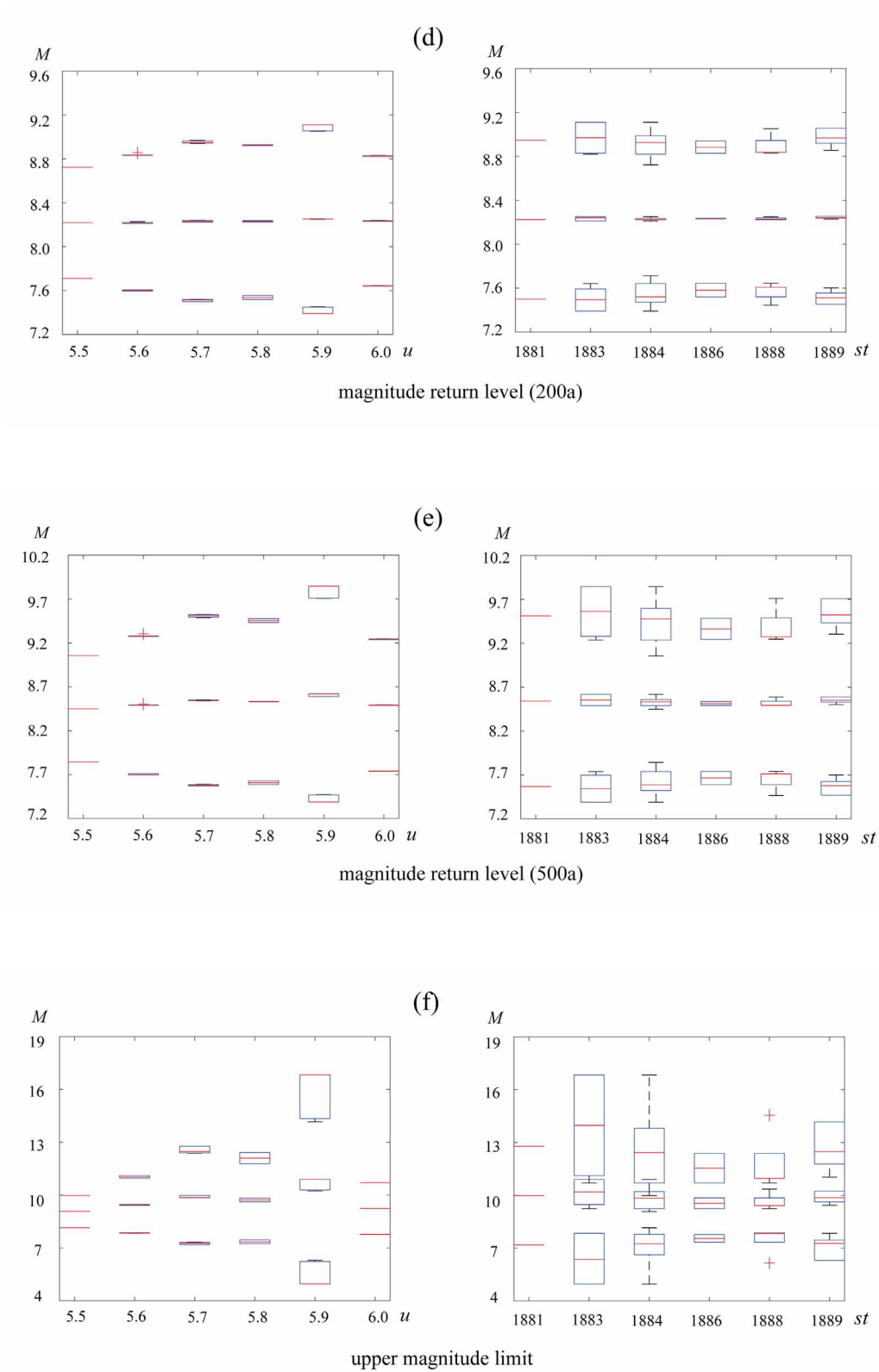
**Table 4.** GPD model parameters sampling combination.

Based on the sampling results used to construct the corresponding GPD models of the northeastern Tibetan Plateau, 30 sets of the earthquake hazard estimation results (Tables 5-10) are obtained. In order to show the distribution of the results more intuitively, the results are presented in the form of boxplots (Figure 6). The left subplots of Figure 6 show the distribution of the earthquake hazard estimation with different earthquake catalog start time  $st$  for fixed

## Mengyi Ren

magnitude threshold  $u$ , while the right subplots of Figure 6 show the distribution of the earthquake hazard estimation with different magnitude threshold  $u$  for fixed earthquake catalog start time  $st$ . The seismic hazard estimation in right subplots is more discrete than in left subplots. It shows that, compared to  $st$ , the different selection of  $u$  has greater influence on the dispersion degree of the estimation results.





**Figure 6.** Boxplots of the seismic hazard estimation based on the two input factors (Figures 6a-f are based on the results in the Table 5-10. X-axis in the left subplots show the magnitude threshold  $u$ , and X-axis in the right subplots show the earthquake catalog start time  $st$ . Each input factor is divided into 6 levels, so X-axis corresponds to 6 values. There are three groups of estimation values in each subplot, from small to large, corresponding to the low end of confidence interval, magnitude return level or upper magnitude limit, upper end of confidence interval).

Trajectory	Magnitude return level and its corresponding confidence interval					
1	7.10	[6.69,7.52]	7.14	[6.68,7.61]	7.13	[6.66,7.59]
2	7.10	[6.68,7.52]	7.09	[6.68,7.49]	7.13	[6.66,7.59]
3	7.05	[6.64,7.46]	7.08	[6.66,7.50]	7.14	[6.73,7.56]
4	7.14	[6.73,7.56]	7.12	[6.72,7.53]	7.05	[6.64,7.46]
5	7.14	[6.73,7.56]	7.12	[6.72,7.53]	7.05	[6.64,7.46]
6	7.05	[6.64,7.46]	7.08	[6.65,7.51]	7.14	[6.72,7.56]
7	7.12	[6.68,7.55]	7.09	[6.68,7.51]	7.18	[6.77,7.60]
8	7.10	[6.68,7.52]	7.09	[6.68,7.49]	7.13	[6.66,7.59]
9	7.12	[6.72,7.53]	7.05	[6.64,7.46]	7.08	[6.65,7.51]
10	7.14	[6.67,7.60]	7.10	[6.69,7.51]	7.09	[6.68,7.49]

**Table 5.** Magnitude return level in 20 years and its corresponding confidence interval with 95% confidence coefficient.

Trajectory	Magnitude return level and its corresponding confidence interval					
1	7.63	[7.15,8.10]	7.68	[7.19,8.17]	7.67	[7.18,8.15]
2	7.63	[7.14,8.11]	7.61	[7.13,8.08]	7.67	[7.18,8.15]
3	7.58	[7.07,8.09]	7.61	[7.10,8.11]	7.66	[7.20,8.11]
4	7.66	[7.20,8.11]	7.64	[7.19,8.09]	7.58	[7.07,8.09]
5	7.66	[7.20,8.11]	7.64	[7.19,8.09]	7.58	[7.07,8.09]
6	7.58	[7.07,8.09]	7.62	[7.11,8.12]	7.66	[7.20,8.12]
7	7.64	[7.16,8.13]	7.62	[7.14,8.09]	7.70	[7.27,8.12]
8	7.63	[7.14,8.11]	7.61	[7.13,8.08]	7.67	[7.18,8.15]
9	7.64	[7.19,8.09]	7.58	[7.07,8.09]	7.62	[7.11,8.12]
10	7.67	[7.19,8.16]	7.62	[7.14,8.10]	7.61	[7.13,8.08]

**Table 6.** Magnitude return level in 50 years and its corresponding confidence interval with 95% confidence coefficient.

Trajectory	Magnitude return level and its corresponding confidence interval					
1	7.95	[7.38,8.53]	7.99	[7.47,8.51]	7.98	[7.46,8.50]
2	7.96	[7.38,8.54]	7.94	[7.36,8.52]	7.98	[7.46,8.50]
3	7.94	[7.29,8.58]	7.95	[7.33,8.57]	7.97	[7.45,8.48]
4	7.97	[7.45,8.48]	7.95	[7.43,8.47]	7.94	[7.29,8.58]
5	7.97	[7.45,8.48]	7.95	[7.43,8.47]	7.94	[7.29,8.58]
6	7.94	[7.29,8.58]	7.96	[7.34,8.58]	7.97	[7.45,8.50]
7	7.97	[7.41,8.53]	7.94	[7.38,8.51]	7.99	[7.53,8.44]
8	7.96	[7.38,8.54]	7.94	[7.36,8.52]	7.98	[7.46,8.50]
9	7.95	[7.43,8.47]	7.94	[7.29,8.58]	7.96	[7.34,8.58]
10	7.99	[7.47,8.50]	7.95	[7.38,8.52]	7.94	[7.36,8.52]

**Table 7.** Magnitude return level in 100 years and its corresponding confidence interval with 95% confidence coefficient.

## Global sensitivity analysis based on GPD

Trajectory	Magnitude return level and its corresponding confidence interval					
1	8.23	[7.52,8.95]	8.24	[7.65,8.83]	8.23	[7.64,8.82]
2	8.24	[7.51,8.97]	8.22	[7.50,8.95]	8.23	[7.64,8.82]
3	8.25	[7.39,9.11]	8.25	[7.45,9.05]	8.22	[7.61,8.84]
4	8.22	[7.61,8.84]	8.21	[7.59,8.83]	8.25	[7.39,9.11]
5	8.22	[7.61,8.84]	8.21	[7.59,8.83]	8.25	[7.39,9.11]
6	8.25	[7.39,9.11]	8.25	[7.45,9.06]	8.23	[7.60,8.86]
7	8.24	[7.56,8.92]	8.22	[7.52,8.93]	8.22	[7.71,8.72]
8	8.24	[7.51,8.97]	8.22	[7.50,8.95]	8.23	[7.64,8.82]
9	8.21	[7.59,8.83]	8.25	[7.39,9.11]	8.25	[7.45,9.06]
10	8.24	[7.64,8.83]	8.23	[7.52,8.94]	8.22	[7.50,8.95]

**Table 8.** Magnitude return level in 200 years and its corresponding confidence interval with 95% confidence coefficient.

Trajectory	Magnitude return level and its corresponding confidence interval					
1	8.54	[7.59,9.49]	8.49	[7.74,9.25]	8.49	[7.74,9.24]
2	8.55	[7.58,9.52]	8.54	[7.57,9.51]	8.49	[7.74,9.24]
3	8.62	[7.39,9.85]	8.59	[7.47,9.71]	8.49	[7.71,9.27]
4	8.49	[7.71,9.27]	8.49	[7.70,9.28]	8.62	[7.39,9.85]
5	8.49	[7.71,9.27]	8.49	[7.70,9.28]	8.62	[7.39,9.85]
6	8.62	[7.39,9.85]	8.59	[7.47,9.71]	8.50	[7.70,9.30]
7	8.53	[7.63,9.43]	8.53	[7.59,9.48]	8.45	[7.74,9.24]
8	8.55	[7.58,9.52]	8.54	[7.57,9.51]	8.49	[7.74,9.24]
9	8.49	[7.70,9.28]	8.62	[7.39,9.85]	8.59	[7.47,9.71]
10	8.49	[7.74,9.24]	8.54	[7.59,9.48]	8.54	[7.57,9.51]

**Table 9.** Magnitude return level in 500 years and its corresponding confidence interval with 95% confidence coefficient.

Trajectory	Magnitude return level and its corresponding confidence interval					
1	9.85	[7.33,12.38]	9.23	[7.76,10.70]	9.23	[7.76,10.70]
2	9.87	[7.26,12.47]	9.98	[7.18,12.78]	9.23	[7.76,10.70]
3	10.89	[4.95,16.84]	10.35	[6.15,14.55]	9.42	[7.87,10.97]
4	9.42	[7.87,10.97]	9.47	[7.83,11.11]	10.89	[4.95,16.84]
5	9.42	[7.87,10.97]	9.47	[7.83,11.11]	10.89	[4.95,16.84]
6	10.89	[4.95,16.84]	10.24	[6.30,14.17]	9.43	[7.83,11.04]
7	9.62	[7.46,11.78]	9.83	[7.24,12.42]	9.07	[8.16,9.98]
8	9.87	[7.26,12.47]	9.98	[7.18,12.78]	9.23	[7.76,10.70]
9	9.47	[7.83,11.11]	10.89	[4.95,16.84]	10.24	[6.30,14.17]
10	9.23	[7.76,10.70]	9.85	[7.33,12.38]	9.98	[7.18,12.78]

**Table 10.** Upper magnitude limit and its corresponding confidence interval with 80% confidence coefficient.

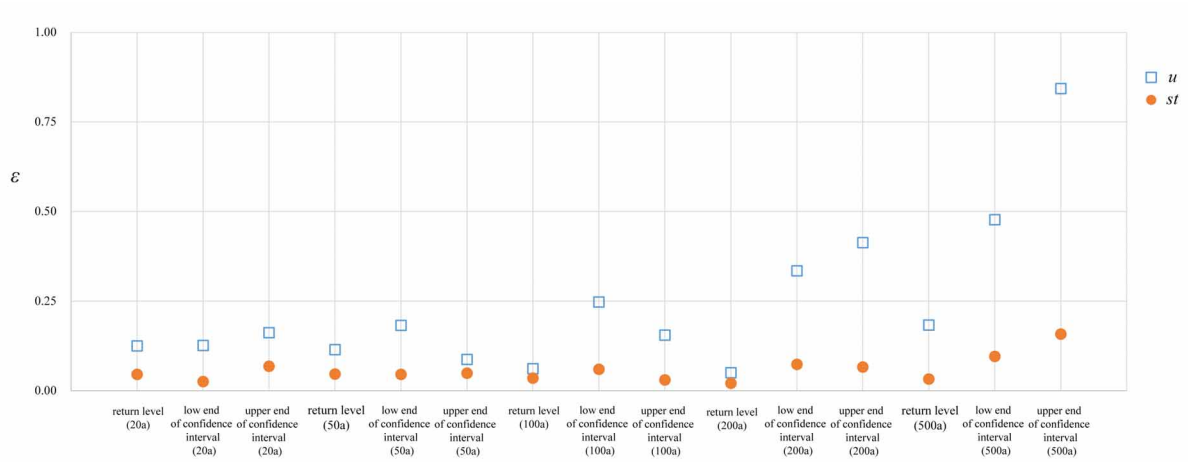


**Figure 7.** Absolute means ( $\mu^*$ ) and standard deviations ( $\sigma$ ) of the elementary effects of the two input factors for magnitude return levels, upper magnitude limit and their corresponding confidence intervals ( $u1$  and  $st1$  represent the input factors for magnitude return levels or upper magnitude limit.  $u2$  and  $st2$  represent the input factors for low end of confidence interval.  $u3$  and  $st3$  represent the input factors for upper end of confidence interval).

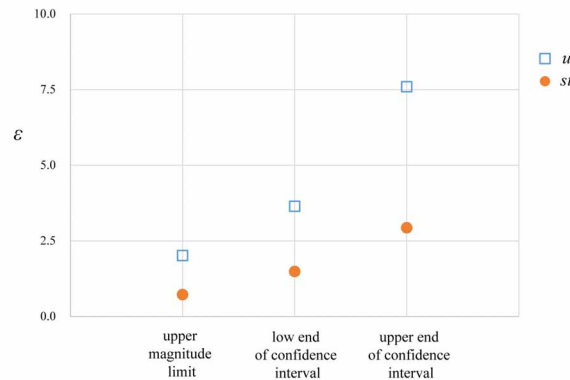
By using the Morris method, the sensitivity analysis results are shown in Figures 7-9, where the sensitivity measures ( $\mu^*$ ,  $\sigma$ ) of the two factors plotted for each output of interest. Recall that  $\mu^*$  reflects the overall influence of an input factor on the output, and  $\sigma$  indicates whether the factor has nonlinear effects or interacts with other factors. The Euclidean distance  $\varepsilon$  between ( $\mu^*$ ,  $\sigma$ ) and the origin is also used for parameter ranking.

A comparison of the two input factors (Figure 7) shows that  $u$  has a relatively high mean value  $\mu^*$ , which indicates a relatively strong influence on the estimation of the magnitude return levels and their corresponding confidence intervals. Moreover, with the increasing of return period,  $u$  has not only a high  $\mu^*$  value but also a relatively high  $\sigma$ , which indicates that the interaction with the other input factor also has a significant effect on the outputs.

Figure 8 shows that  $u$  has a more important impact than  $st$  on the uncertainty of the magnitude return level and its corresponding confidence intervals on the whole. The influence of  $st$  on the estimated results is relatively stable with increasing return period, whereas the influence of  $u$  on the estimated results significantly increases, especially at the both ends of the confidence intervals.



**Figure 8.** Euclidean distance  $\varepsilon$  of the two input factors for magnitude return levels and their corresponding confidence intervals (Blue box represents the input factor magnitude threshold  $u$ . Orange solid circle represents the input factor earthquake catalog start time  $st$ ).



**Figure 9.** Euclidean distance  $\varepsilon$  of the two input factors for upper magnitude limit and its corresponding confidence interval (Blue box represents the input factor magnitude threshold  $u$ . Orange solid circle represents the input factor earthquake catalog start time  $st$ ).

The effects of uncertainty of the two input factors on the uncertainty of the upper magnitude limit and its corresponding confidence interval are also shown in Figures 7 and 9, where  $u$  is well separated from the other value because of its high mean  $\mu^*$ . When considering  $\mu^*$  and  $\sigma$  together, one can conclude that  $u$  is not only important for the upper magnitude limit estimation and its corresponding confidence interval but also has a significant effect that involves either interaction or curvature. Hence, the rationality of the  $u$  selection should be carefully considered when estimating strong earthquake hazards using the GPD model. Moreover, the sensitivity characteristic ranking of the input factors become more stable with increasing return period, which also indicates that the GPD model is more suitable for estimating the strong earthquakes magnitude return level and upper magnitude limit using the tail feature of the magnitude sequence.

## 5. Discussion and Conclusions

In this paper, a qualitative uncertainty analysis process for strong earthquake hazard estimation is constructed. The selection of input factors for strong earthquake GPD model involves uncertainties, thus the seismic hazard estimation obtained by using this model is inevitably uncertain. Global sensitivity analysis method of strong earthquake hazard estimation based on GPD has not been previously reported. In this paper, the Morris method is introduced to handle the uncertainty of the GPD model. The influence of two input factors, earthquake catalog start time ( $st$ ) and magnitude threshold ( $u$ ), and their interaction on the hazard estimation uncertainty is analyzed.

For northeastern Tibetan Plateau seismic catalogue data, a GPD model is established to assess the magnitude return levels with different return periods and their corresponding confidence intervals. The model also predicts the upper magnitude limit and its corresponding confidence intervals for northeastern Tibetan Plateau. With magnitude threshold of 5.5, the magnitude excess distribution follows a GPD, and the modified scale and shape parameters are 1.23 and  $-0.35$ , respectively. The magnitude return levels in 20, 50, 100, 200, and 500 years are 7.19, 7.70, 7.99, 8.22, and 8.45, respectively. The corresponding 95% confidence intervals are [6.77, 7.60], [7.27, 8.12], [7.53, 8.44], [7.71, 8.72], and [7.84, 9.06], respectively. The upper magnitude limit from the GPD model is 9.07 and its 80% confidence interval is [8.16, 9.98]. The GPD modeling approach developed in this paper can make full use of historical earthquake catalogues and construct seismic activity models.

By using the Morris method, the influence of input factors and their interaction with the GPD model on the uncertainty of strong earthquake hazard estimations is qualitatively analyzed using the sensitivity index absolute mean ( $\mu^*$ ) and standard deviation ( $\sigma$ ). The analysis result indicates that the input magnitude threshold ( $u$ ) has a relatively large influence on the seismic hazard estimation. Thus, meaningful magnitude threshold selection is the most important step for the successful GPD modeling of seismic activities in a region. This method can be applied to the sensitivity analysis of other seismicity models. The results also can improve the understanding of seismicity models and increase the modeling efficiency.

**Acknowledgments.** We would like to thank Dr. Jiang Changsheng and Dr. Bian Yinju (Institute of Geophysics, China Earthquake Administration) for their assistance during this work, the reviewers for their helpful suggestions, and the editors for their hard work. This work has been supported by the National Key R&D Program of China (No. 2018YFC1503201-04).

## References

- Bao, X.W., X.D. Song, M.J. Xu, L.S. Wang, X.X. Sun, N. Mi, D.Y. Yu, H. Li (2013). Crust and upper mantle structure of the North China Craton and the NE Tibetan Plateau and its tectonic implications, *Earth Planet. Sci. Lett.*, 369-370, 129-137.
- Campolongo, F., J. Cariboni and A. Saltelli (2007). An effective screening design for sensitivity analysis of large models, *Environmental Modelling and Software*, 22, 10, 1509-1518.
- Chang, L.J., C.Y. Wang, Z.F. Ding, M.D. Zhou, J.S. Yang, Z.Q. Xu, X.D. Jiang, X.F. Zheng (2008). Seismic anisotropy of upper mantle in the northeastern margin of the Tibetan Plateau, *Chinese J. Geophys.*, 51, 2, 431-438.
- Chen, L., J. Liu, Y. Chen (1998). Aftershock deletion in seismicity analysis, *Chinese J. Geophys.*, 41, S1, 244-252.
- Chen, J.H., Q.Y. Liu, S.C. Li, B. Guo, Y.G. Lai (2005). Crust and upper mantle S wave velocity structure across Northeastern Tibetan plateau and Ordos block, *Chinese J. Geophys.*, 48, 2, 333-342.
- Console, R., C. Gasparini, B. De Simoni, L. Marcelli, M. Vecchi (1979). Preambolo al Catalogo Sismico Nazionale I criteri di informazione del CSN, *Annali Geofis.*, 32, 37-77.
- Coles, S. (2001). *An introduction to statistical modeling of extreme values*, Springer-Verlag.
- Dutfoy, A. (2019). Estimation of tail distribution of the annual maximum earthquake magnitude using extreme value theory, *Pure Appl. Geophys.*, 176, 2, 527-540.
- Fisher, R., Tippett, L.H. (1928). Limiting forms of the frequency distributions of the largest or smallest member of a sample, *Mathematical Proceedings of the Cambridge Philosophical Society*, 24, 180-190.
- Gan, W.J., P.Z. Zhang, Z.K. Shen, Z.J. Niu, M. Wang, Y.G. Wan, D.M. Zhou, J. Cheng (2007). Present-day crustal motion



- within the Tibetan Plateau inferred from GPS measurements, *J. Geophys. Res.: Solid Earth*, 112, B08416.
- Huang, W.Q., W.X. Li, X.F. Cao (1994). Research of seismic data integrity in mainland China II: the start year distribution image of basic integrity of the regional seismic data, *Acta Seismol. Sinica*, 16, 4, 423-432.
- Jiang, P., Dai, S.L. (1993). Introduction to engineering seismology, Beijing: Seismological Press.
- Li, Y.H., Q.J. Wu, Z.H. An, X.B. Tian, H.G. Li (2006). The Poisson ratio and crustal structure across the NE Tibetan Plateau determined from receiver functions, *Chinese J. Geophys.*, 49,5,1359-1368.
- Li, Y.H., Wang, X.C., R.Q. Zhang, Q.J. Wu, Z.F. Ding (2017). Crustal structure across the NE Tibetan Plateau and Ordos Block from the joint inversion of receiver functions and Rayleigh-wave dispersions, *Tectonophysics*, 705, 33-41.
- Liu, J., Q.F. Chen, Y. Chen (1996). Completeness analysis of the seismic catalog in north China region, *Earthquake*, 1, 59-67.
- Morris, M.D. (1991). Factorial sampling plans for preliminary computational experiments, *Technometrics*, 33, 2,161-174.
- Pappas, C., S. Fatichi, S. Leuzinger, A. Wolf, P. Burlando (2013). Sensitivity analysis of a process-based ecosystem model: Pinpointing parameterization and structural issues, *J. Geophys. Res.*, 118, 2, 505-528.
- Pisarenko, V.F., D. Sornette (2003). Characterization of the Frequency of Extreme Earthquake Events by the Generalized Pareto Distribution, *Pure Appl. Geophys.*, 160, 12, 2343-2364.
- Pisarenko, V.F., A. Sornette, D. Sornette, M.V. Rodkin (2014). Characterization of the Tail of the Distribution of Earthquake Magnitudes by Combining the GEV and GPD Descriptions of Extreme Value Theory, *Pure Appl. Geophys.*, 171, 8, 1599-1624.
- Qian, X.S., X.G. Cai, Q.Q. Ren (2013). Characteristics of the great earthquake magnitude distributions for active tectonic boundaries in Chinese mainland, *J. Earthquake Engin. Engin.Vibr.*, 33, 1, 212-220.
- Ren, M.Y. (2018). The establishment of Generalized Pareto Distribution model of seismicity in Longmenshan region, *J. Seismol. Res.*, 41, 2, 226-232.
- Ren, X.L., H.L. He, L. Zhang, F. Li, M. Liu, G.R. Yu, J.H. Zhang (2018). Modeling and uncertainty analysis of carbon and water fluxes in a broad-leaved Korean pine mixed forest based on model-data fusion, *Ecological Modelling*, 379, 39-53.
- Saltelli, A., S. Tarantola, F. Campolongo, M. Ratto (2004). Sensitivity Analysis in Practice: A Guide to Assessing Scientific Models, *J. Royal Stat. Soc.: Series A (Statistics in Society)*, Wiley, 232.
- Shi, D.Q. (2006). Practical extremum statistics methods, Tianjin: Science and Technology Press.
- Sun, Y.Q., G. Luo, L. Yin, Y.L. Shi (2019). Migration probability of big earthquakes and segmentation of slip rates on the fault system in northeastern Tibetan Plateau, *Chinese J. Geophys.*, 62, 5,1663-1679.
- Taylor, M., A. Yin (2009). Active structures of the Himalayan-Tibetan orogen and their relationships to earthquake distribution, contemporary strain field, and Cenozoic volcanism, *Geosphere*, 5, 3), 199-214.
- Tian, X.B., Z.J. Zhang (2013). Bulk crustal properties in NE Tibet and their implications for deformation model, *Gondwana Res.*, 24, 548-559.
- Tian, J.W., Z. Liu, L.C. Ren (2017). Seismic hazard estimation of the Manila trench subduction zone based on generalized Pareto distribution, *Earthquake*, 37, 1, 158-165.
- Yin, A., T.M. Harrison (2000). Geologic evolution of the Himalayan-Tibetan orogen, *Annual Rev. Earth Planet. Sci.*, 28, 211-280.
- Zhao, G.Z., J. Tang, Y. Zhan, X.B. Chen, X.J. Zhou, J.J. Wang, F. Xuan, Q.H. Deng, J.M. Zhao (2004). Relation between electricity structure of the crust and deformation of crustal blocks on the Northeastern margin of Qinghai-Tibet plateau, *Science in China (Series D)*, 34, 10, 908-918.
- Zhu, A.Y., D.N. Zhang, T. Zhu, Y.X. Guo (2018). Influence of mantle convection to the crustal movement pattern in the northeastern margin of the Tibetan Plateau based on numerical simulation, *Science China Earth Sci.*, 61, 1644-1658.

**\*CORRESPONDING AUTHOR: Mengyi REN,**

Institute of Geophysics, China Earthquake Administration,  
No.5 Minzudaxue Nanlu, Haidian District,  
Beijing 100081, China,  
e-mail: renmengyi@163.com

© 2021 the Author(s). All rights reserved.

Open Access. This article is licensed under a Creative Commons Attribution 3.0 International