

Distribution of maximum earthquake magnitudes in future time intervals: application to the seismicity of Japan (1923–2007)

V. F. Pisarenko¹, D. Sornette^{2,3,4}, and M. V. Rodkin⁵

¹International Institute of Earthquake Prediction Theory and Mathematical Geophysics,
Russian Academy of Science, Profsoyuznaya 84/32, Moscow 117997, Russia

²D-MTEC, ETH Zurich, Kreuzplatz 5, 8032 Zurich, Switzerland

³Department of Earth Sciences, ETH Zurich, 8032 Zurich, Switzerland

⁴Institute of Geophysics and Planetary Physics, University of California, Los Angeles, CA 90095, USA

⁵Geophysical Center of Russian Academy of Sciences, Molodejnaya 3, Moscow, Russia

(Received August 16, 2009; Revised June 2, 2010; Accepted June 11, 2010; Online published August 31, 2010)

We have modified the new method for the statistical estimation of the tail distribution of earthquake seismic moments introduced by Pisarenko *et al.* (2009) and applied it to the earthquake catalog of Japan (1923–2007). The newly modified method is based on the two main limit theorems of the theory of extreme values and on the derived duality between the generalized Pareto distribution (GPD) and the generalized extreme value distribution (GEV). Using this method, we obtain the distribution of maximum earthquake magnitudes in future time intervals of arbitrary duration τ . This distribution can be characterized by its quantile $Q_q(\tau)$ at any desirable statistical level q . The quantile $Q_q(\tau)$ provides a much more stable and robust characteristic than the traditional absolute maximum magnitude M_{\max} (M_{\max} can be obtained as the limit of $Q_q(\tau)$ as $q \rightarrow 1$, $\tau \rightarrow \infty$). The best estimates of the parameters governing the distribution of $Q_q(\tau)$ for Japan (1923–2007) are the following: $\xi_{\text{GEV}} = -0.19 \pm 0.07$; $\mu_{\text{GEV}}(200) = 6.339 \pm 0.038$; $\sigma_{\text{GEV}}(200) = 0.600 \pm 0.022$; $Q_{0.90,\text{GEV}}(10) = 8.34 \pm 0.32$. We have also estimated $Q_q(\tau)$ for a set of q -values and future time periods in the range $1 \leq \tau \leq 50$ years from 2007 onwards. For comparison, the absolute maximum estimate $M_{\max,\text{GEV}} = 9.57 \pm 0.86$ has a scatter more than twice that of the 90% quantile $Q_{0.90,\text{GEV}}(10)$ of the maximum magnitude over the next 10 years beginning from 2007.

Key words: Extreme value theory, generalized extreme value distribution, generalized Pareto distribution, earthquake seismic moments, magnitude.

1. Introduction

The work presented in this article has two goals: (1) to adapt the method suggested by Pisarenko *et al.* (2009) for the statistical estimation of the tail of the distribution of earthquake magnitudes to catalogs in which earthquake magnitudes are reported in discrete values, and (2) to apply the newly developed method to the Japan Meteorological Agency (JMA) magnitude catalog of Japan (1923–2007) in order to estimate the maximum possible magnitude and other measures characterizing the tail of the distribution of magnitudes.

The method of Pisarenko *et al.* (2009) is a continuation and improvement of the technique suggested in Pisarenko *et al.* (2008). Both rely on the assumption that the distribution of earthquake magnitudes is limited to some maximum value M_{\max} , which is itself probably significantly less than the absolute limit imposed by the finiteness of the Earth. This maximum value M_{\max} may reflect the largest possible set of seismo-tectonic structures in a given tectonic region that can support an earthquake, combined with an extremal

occurrence of dynamical energy release per unit area. The simplest model embodying the idea of a maximum magnitude is the truncated Gutenberg-Richter (GR) magnitude distribution truncated at M_{\max} :

$$F(m) = C [1 - 10^{-b(m-m_0)}]; \quad m_0 \leq m \leq M_{\max}, \quad (1)$$

where $F(m)$ is the cumulative probability distribution of earthquake magnitudes, b is the slope parameter, m_0 is the lower known threshold above which magnitudes can be considered to be reliably recorded, M_{\max} is the maximum possible magnitude, and C is the normalizing constant (which depends on the unknown parameters b and M_{\max}) (Cosentino *et al.*, 1977; Kijko and Sellevol, 1989, 1992; Pisarenko *et al.*, 1996; Kijko, 2004). The parameter M_{\max} is a priori a very convenient tool for building engineers and the insurance business. However, multiple attempts to use M_{\max} have clearly shown that this parameter is unstable with respect to minor variations in the catalogs and, in particular, for use with incomplete regional catalogs, which are a rather common situation in seismology. Consequently, the parameter M_{\max} is an unreliable measure of the largest seismic risks. The truncated GR model can be contrasted with the various modifications of the GR law stretching to infinity. These modifications impose a finite-size constraint only on the statistical average of the energy released by

earthquakes (see, for example, Sornette *et al.*, 1996; Kagan, 1999; Kagan and Schoenberg, 2001), but they contradict the finiteness of seismogenic structures in the Earth and therefore have not been universally accepted.

The chief innovation, introduced first by Pisarenko *et al.* (2009) and extended here, is to combine the two main limit theorems of extreme value theory (EVT), which allows us to derive the distribution of T -maxima (maximum magnitude occurring in sequential time intervals of duration T) for arbitrary T . This distribution enables derivation of any desired statistical characteristic of the future T -maximum. The two limit theorems of EVT correspond to the generalized extreme value distribution (GEV) and to the generalized Pareto distribution (GPD), respectively. Pisarenko *et al.* (2009) established the direct relations between the parameters of these two distributions. The duality between the GEV and GPD provides a new approach to check the consistency of the estimation of the tail characteristics of the distribution of earthquake magnitudes for earthquakes occurring over arbitrary time intervals.

Instead of focusing on the unstable parameter M_{\max} , we suggest a new, stable, and convenient characteristic, $M_{\max}(\tau)$, defined as the maximum earthquake that can be recorded over a future time interval of duration τ . The random value $M_{\max}(\tau)$ can be described by its distribution function or by its quantiles $Q_q(\tau)$, which are, in contrast to M_{\max} , stable and robust characteristics. In addition, if $\tau \rightarrow \infty$, then $M_{\max}(\tau) \rightarrow M_{\max}$ with a probability of one. The methods for calculating $Q_q(\tau)$ are given in the following section. In particular, we can estimate $Q_q(\tau)$ for, say, $q = 10\%$, 5% , and 1% , as well as for the median ($q = 50\%$) for any desirable time interval τ . These methods are illustrated below on the magnitude catalog of the JMA, over the time period 1923–2007, for magnitudes $m \geq 4.1$.

We should stress that our method relies on the assumption that the distribution of earthquake magnitudes exhibits a regular limit behavior on its right (for large magnitudes)—even though there is no way to be absolutely certain that this is the case due to the limited data set for large and extreme earthquake sizes. Thus, in specific cases, seismologists are forced to accept the most appropriate assumption about the behavior of the magnitude distribution on its right end. The assumption used in our paper (which coincides with the assumption of the EVT: the existence of a non-trivial asymptotic distribution for centered and normalized maximum of sample) seems to be the least harmful and the most fruitful. It provides the three well-known types of possible limit distributions for the maximum (in our paper we use only one of these). Without such an assumption, it would scarcely be possible to obtain any useful result on the distribution of sample maxima.

2. The Method

The method developed here is based on the following assumptions:

- (1) the Poisson property of independence in time of the main shocks;
- (2) independence between the observed magnitudes M ;

- (3) regularity of the tail probability of the earthquake magnitudes M ;

We now present the elements that justify using these assumptions and then describe the specifics of the method.

2.1 Test of the Poisson hypothesis

Our analysis is performed for main shocks, following the application of a declustering method. We used the Kagan-Knopoff time-space window declustering method to remove the aftershocks. This method has a number of shortcomings, and other versions of aftershock cleansing are available, but these have no universally accepted advantages. There is a widespread opinion among seismologists that the overwhelming majority of main shocks can be considered to be independent random variables. This property is more evident when earthquake observations are considered on a global scale, but it is still a reasonable hypothesis for large seismic regions, such as Japan. The Japanese data that we use exhibit evident irregularities in the registration process, which are visible in Fig. 7. In particular, during the time interval 1945–1965, the lack of observations is clearly evident. Fortunately, this effect is not essential for the larger earthquakes, which are the focus of our work.

We note that the model of a Poisson flow of events corresponds to a renewal model with exponentially distributed intervals between successive events. Testing for the Poisson property is reduced to the study of the distribution of time intervals between successive main shocks. In our analysis, we are going to study this distribution for events in Japan with magnitudes larger than some chosen lower threshold. We will show that, at least for large earthquakes with $m \geq 7.0$, the exponential distribution cannot be rejected at a rather high statistical significance level. For earthquakes with $m \geq 6.0$, the exponential distribution can be accepted, at least since 1966. For earthquakes of smaller sizes, the deviations of the distribution of the time intervals from the exponential law becomes more pronounced; consequently, the renewal model with non-exponentially distributed time intervals is perhaps more appropriate. However, this is a rather irrelevant finding for our purpose of determining the distribution of maximum earthquake magnitudes, which is controlled mainly by the large earthquakes.

We analyze the empirical distributions of time intervals between successive events in sub-catalogs derived from the main catalog by selecting main shocks posterior to some time T_0 and with magnitudes larger than the lower threshold m_0 . In order to test the exponential hypothesis, we use the Kolmogorov distance KD-test modified by Stephens (1974) for the case where the unknown parameter is estimated from the same sample. We obtain the following results for different choices of m_0 and T_0 . The variable $\langle t_k - t_{k-1} \rangle$ is the mean inter-event waiting time.

$$m_0 = 6.0; T_0 = 1923.$$

$$\text{KD} = 1.27; n = 318; \langle t_k - t_{k-1} \rangle = 96.70 \text{ days};$$

$$p\text{-value} = 1.5\%;$$

$$m_0 = 6.0; T_0 = 1966.$$

$$\text{KD} = 0.91; n = 156; \langle t_k - t_{k-1} \rangle = 91.52 \text{ days};$$

$$p\text{-value} \cong 20\%;$$

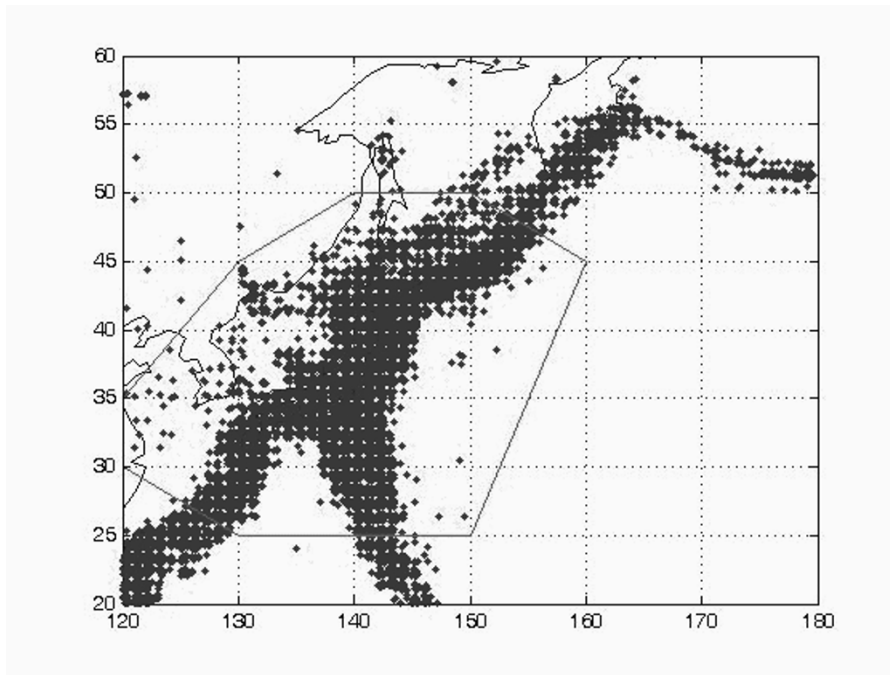


Fig. 1. Map of the region kept for our study; the coordinates of nodes of the polygon delimiting the area of study are [(160.00; 45.00); (150.00; 50.00); (140.00; 50.00); (130.00; 45.00); (120.00; 35.00) : (120.00; 30.00); (130.00; 25.00); (150.00; 25.00); (160.00; 45.00)].

$$m_0 = 7.0; T_0 = 1923.$$

$$\text{KD} = 0.74; n = 52; \langle t_k - t_{k-1} \rangle = 588 \text{ days};$$

$$p\text{-value} \cong 60\%.$$

$$m_0 = 7.5; T_0 = 1923.$$

$$\text{KD} = 0.53; n = 15; \langle t_k - t_{k-1} \rangle = 2,030 \text{ days};$$

$$p\text{-value} \cong 80\%.$$

The exponential Poisson hypothesis is thus acceptable (accepting, say, if $p\text{-value} > 0.1$) for $m \geq 6.0$ since 1966, and for $m \geq 7.0$ for the whole catalog starting from 1923.

2.2 Independence of the magnitudes

Figure 7 shows the magnitudes of the main shocks as a function of time from 1923 to 2007, inside the polygonal domain shown in Fig. 1, whose depths are < 70 km. Together with the test shown in the previous subsection, one can see that the magnitudes are approximately random above $m = 6$, which is the regime of interest for the application of the EVT. We also note that the GR distribution is rather well verified, as depicted in Figs. 2 and 4, confirming the standard one-point statistics of earthquakes.

2.3 Regularity of the tail probability of the earthquake magnitudes M

According to the EVT, the limit distribution of maxima can be obtained in two ways. The first, sometimes called the “peak over threshold” method, consists of increasing a threshold h above which observations are kept. The distribution of event sizes that exceed h tends—under an affine transformation—to the GPD as h tends to infinity. The GPD depends on two unknown parameters (ξ, s) and on the known threshold h (see, for example, Embrechts *et al.* (1997)). For the case of random values that are limited from

above, the GPD can be written as follows:

$$\text{GPD}(x|\xi, s, h) = 1 - [1 + (\xi/s) \cdot (x - h)]^{-1/\xi},$$

$$\xi < 0; s > 0; h \leq x \leq h - s/\xi. \quad (2)$$

Here, ξ is the form parameter, s is the scale parameter, and the combination $h - s/\xi$ represents the uppermost magnitude, that we shall denote M_{\max} :

$$M_{\max} = h - s/\xi, \quad \xi < 0. \quad (3)$$

We shall consider only this case of a finite M_{\max} , to capture the finiteness of seismo-tectonic structures in the Earth, as discussed in Introduction.

The second way to obtain the limit distribution of maxima consists of selecting directly the maxima occurring in sequences of n successive observations $M_n = \max(m_1, \dots, m_n)$ and in studying their distribution as n goes to infinity. In accordance with the main theorem of the EVT (see, for example, Embrechts *et al.*, 1997), this distribution, named the GEV, can be written (for the case of random values limited from above) in the form:

$$\text{GEV}(x|\zeta, \sigma, \mu) = \exp(-[1 + (\zeta/\sigma) \cdot (x - \mu)]^{-1/\zeta}),$$

$$\zeta < 0; \sigma > 0; x \leq \mu - \sigma/\zeta. \quad (4)$$

The conditions guaranteeing the validity of these two limit theorems include the regularity of the original distributions of magnitudes in their tail. These conditions ensure the existence of a *non-degenerate* limit distribution of M_n after a proper centering and normalization. Following the standard approach, we assume that the conditions for which a non-degenerate limit distribution of M_n exists are truly

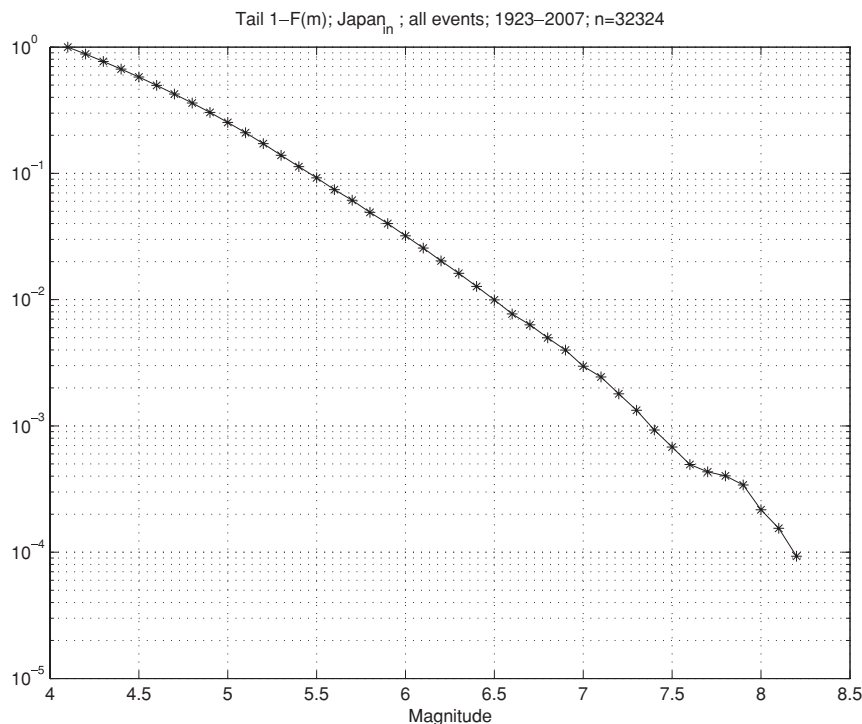


Fig. 2. Magnitude-frequency distribution of the 32,324 earthquakes that occurred in the region delimited by the polygon shown in Fig. 1 over the period 1923–2007.

valid. If this were not to be the case, we would not be able to perform any meaningful analysis. While this argument may appear circular, it is standard approach in statistics in general and in statistical seismology in particular. One can never really prove the validity of mathematical conditions solely from data. The model or theory can, however, be progressively validated by comparing its predictions with the results of precise tests (Sornette *et al.*, 2007, 2008). It is therefore the conclusions that we derive from our analysis that will support—or refute—the value of the analysis itself.

2.4 Formulation of the theory and procedure

In our analysis, we study the maximum magnitudes occurring in time interval $(0, T)$. We assume that the flow of main shocks is a Poissonian stationary process with some intensity λ . This property for main shocks was studied and confirmed in appendix A of Pisarenko *et al.* (2008) for the Harvard catalog of seismic moments over the time period 1 January 1977–20 December 2004. The term “main shock” refers here to the events that remain following the application of a suitable desclustering algorithm (see Pisarenko *et al.*, 2008, 2009, and below). In Subsection 2.1, we tested the Poisson hypothesis and confirmed that (1) for earthquakes with $m \geq 6.0$, the exponential distribution can be accepted—at least since 1966; (2) for large earthquakes with $m \geq 7.0$, the exponential distribution cannot be rejected with rather a high statistical significance level. We can then proceed with the description of the model.

Given the intensity λ and the duration T of the time window, the average number of observations (main shocks) within the interval $(0, T)$ is equal to $\langle n \rangle = \lambda T$. For $T \rightarrow \infty$, the number of observations in $(0, T)$ tends to infinity with a probability of one; we can therefore use Eq. (4) as

the limit distribution of the maximum magnitudes m_T of the main shocks occurring in time interval $(0, T)$ of growing sizes (Pisarenko *et al.*, 2008).

Pisarenko *et al.* (2009) showed that, for a Poissonian flow of main shocks, the two limit distributions, namely, the GPD given by relation (2) and the GEV given by relation (4), are related in a simple manner. Here, we briefly summarize the main points and refer the reader to Pisarenko *et al.* (2009) for details. If the random variable (rv) X has the GPD-distribution (relation (2)) and the maximum of a random sequence of observations X_k is taken:

$$M_T = \max(X_1, \dots, X_\nu), \quad (5)$$

where ν is a random Poissonian value with parameter λT , with $\lambda T \gg 1$, then M_T has the GEV-distribution (Eq. (4)) with the following parameters:

$$\zeta(T) = \xi; \quad (6)$$

$$\sigma(T) = s \cdot (\lambda T)^\xi; \quad (7)$$

$$\mu(T) = h - (s/\xi) \cdot [1 - (\lambda T)^\xi]. \quad (8)$$

These expressions are valid up to small terms of order $\exp(-\lambda T)$, which are neglected.

The inverse is true as well: if $M_T = \max(X_1, \dots, X_\nu)$ has the GEV distribution (Eq. (4)) with parameters ζ, σ, μ , then the original distribution of X_k has the GPD distribution (Eq. (2)) with parameters:

$$\xi = \zeta; \quad (9)$$

$$s = \sigma \cdot (\lambda T)^{-\xi}; \quad (10)$$

$$h = \mu + (\sigma/\xi) \cdot [(\lambda T)^{-\xi} - 1]. \quad (11)$$

The proof can be found in Pisarenko *et al.* (2009) where we see that the form parameter in the GPD and the GEV

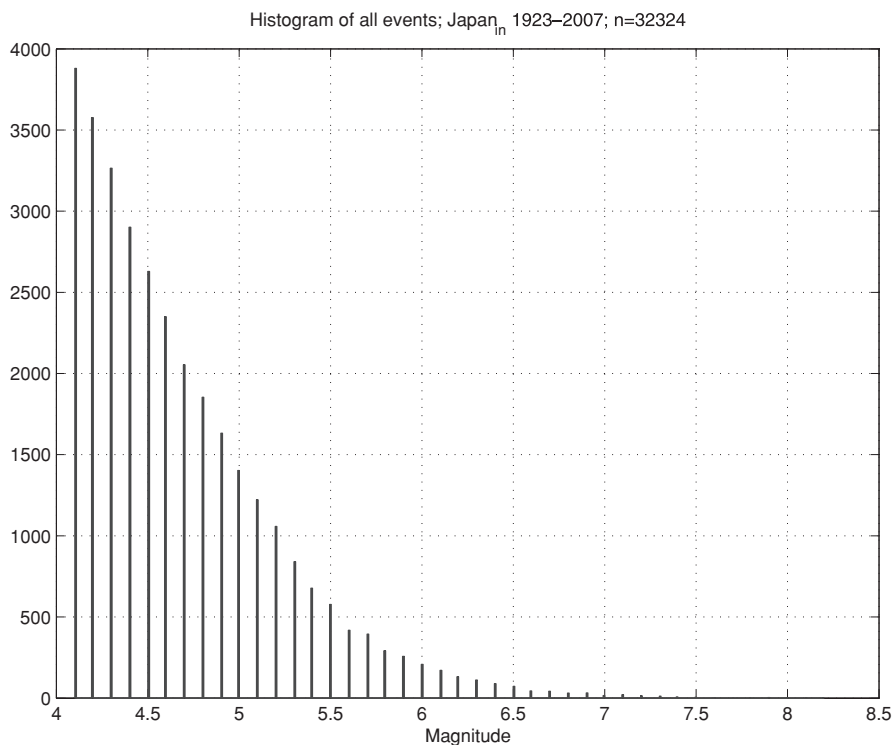


Fig. 3. Histogram of the magnitudes of earthquakes used in Fig. 2. The discrete 0.1-bins are clearly visible.

is always identical, whereas the centering and normalizing parameters differ.

Using Eqs. (6)–(11), one can recalculate the estimates $\zeta(T)$, $\sigma(T)$, $\mu(T)$ obtained for some T into corresponding estimates for another time interval of different duration τ :

$$\mu(\tau) = \mu(T) + (\sigma(T)/\xi) \cdot [(\tau/T)^\xi - 1]; \quad (12)$$

$$\sigma(\tau) = \sigma(T) \cdot (\tau/T)^\xi. \quad (13)$$

Equations (6)–(13) are very convenient, and we shall use them in our estimation procedures. In the following, we use the notation T to denote the duration of a window in the known catalog (or part of the catalog) used for the estimation of the parameters, whereas we use τ to refer to a future time interval (prediction).

From the GPD (relation (2)) or the GEV (Eq. (4)), we can obtain the quantiles $Q_q(\tau)$, which are proposed as stable robust characteristics of the tail distribution of magnitudes. These quantiles are the roots of equations:

$$\text{GPD}(x|\xi, s, h) = q; \quad (14)$$

$$\text{GEV}(x|\zeta, \sigma, \mu) = q. \quad (15)$$

Inverting Eqs. (14) and (15) for x as a function of q and using Eqs. (6)–(8), we obtain:

$$Q_q(\tau) = \mu(T) + (\sigma(T)/\xi) \cdot [a \cdot (\tau/T)^\xi - 1]; \quad (16)$$

from (14)

$$Q_q(\tau) = h + (s/\xi) \cdot [a \cdot (\lambda\tau)^\xi - 1], \quad (17)$$

from (15)

where $a = [\log(1/q)]^{-\xi}$.

3. Application of the GPD and GEV to the Estimation of T -maximum Magnitudes in Japan

3.1 Characteristics of the JMA data

The full JMA catalog covers the spatial domain delimited by $25.02 \leq \text{latitude} \leq 49.53^\circ$ and $121.01 \leq \text{longitude} \leq 156.36^\circ$ and by the temporal window 1 January 1923 to 30 April 2007. The depths of the earthquakes fall in the interval $0 \leq \text{depth} \leq 657$ km. The magnitudes are expressed in 0.1-bins and vary in the interval $4.1 \leq \text{magnitude} \leq 8.2$. There are 39,316 events in this space-time domain. The spatial domain covered by the JMA catalog covers the Kuril Islands and the east border of Asia.

Here, we focus our study on earthquakes occurring within the central Japanese islands. We thus restrict the territory of our study to earthquakes occurring within the polygon with coordinates [(160.00; 45.00); (150.00; 50.00); (140.00; 50.00); (130.00; 45.00); (120.00; 35.00); (120.00; 30.00); (130.00; 25.00); (150.00; 25.00); (160.00; 45.00)]. Figure 1 shows the map of the region delineated by the polygon within which we perform our study. There were 32,324 events within this area. The corresponding magnitude-frequency is shown in Fig. 2, and the histogram of magnitudes is shown in Fig. 3.

Next, we only keep “shallow” earthquakes whose depths are < 70 km and apply the declustering Knopoff-Kagan space-time window algorithm (Knopoff and Kagan, 1977). The remaining events constitute our “main shocks”, on which we are going to apply the GPD and the GEV methods described above. There are 6,497 main shocks in the polygon shown in Fig. 1 with depths < 70 km. The magnitude-frequency curve of these main shocks is shown in Fig. 4. It should be noted that the b -slope of the magnitude-frequency of main shocks is significantly smaller (by approx. 0.15)

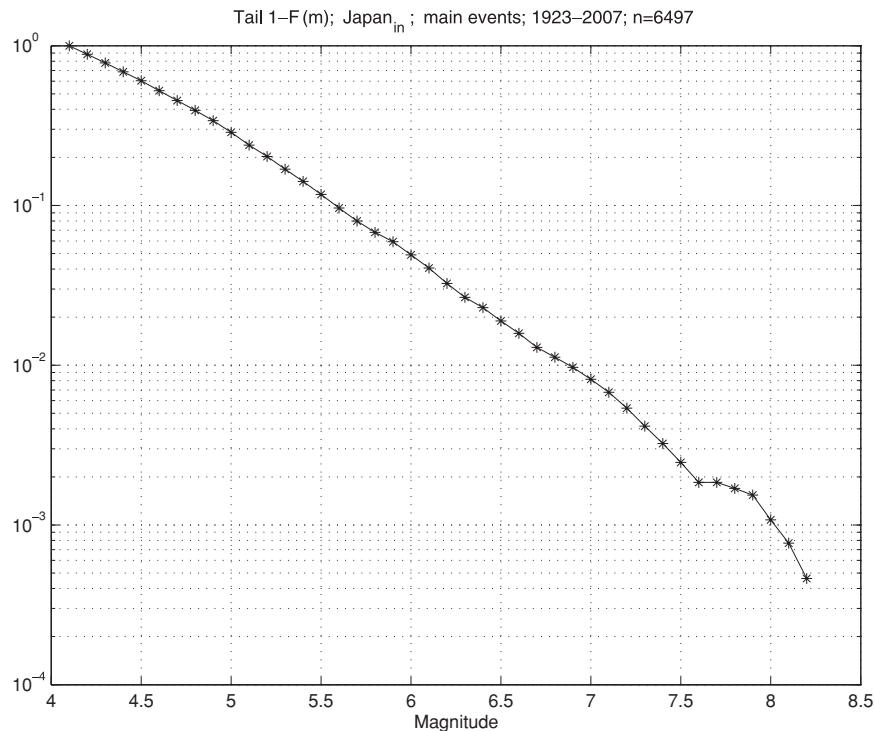


Fig. 4. Magnitude-frequency distribution of the 6,497 “main shocks” remaining in the domain delineated by the polygon shown in Fig. 1 over the period 1923–2007 that have depths of <70 km, after applying the Knopoff-Kagan declustering algorithm.

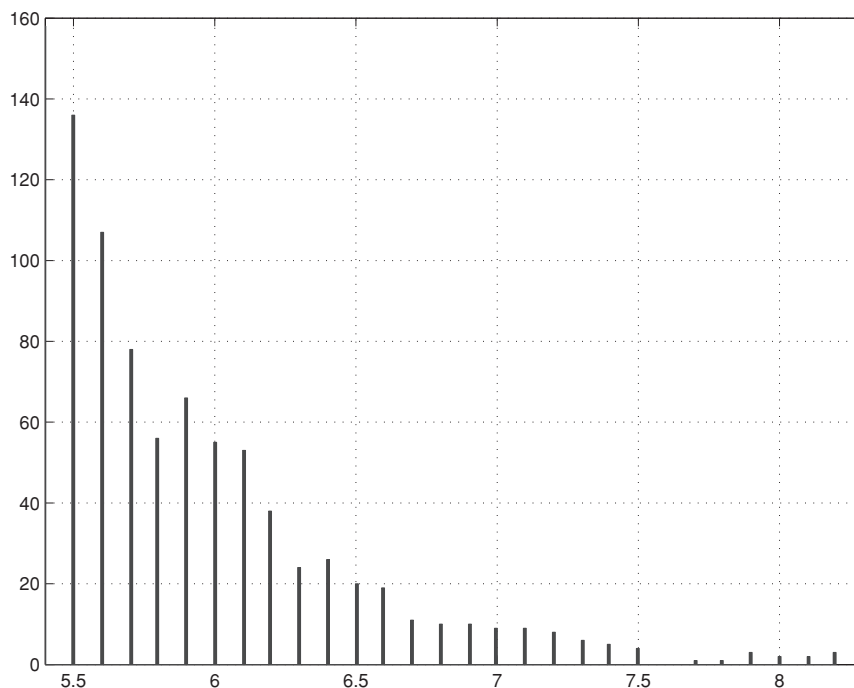


Fig. 5. Histogram of the magnitudes of the main shocks whose magnitude-frequency distribution is shown in Fig. 4. The discrete 0.1-bins are clearly visible. There are additional oscillations decorating the decay with magnitudes that require further coarse-graining, as explained in the text.

than the corresponding b -slope of the magnitude-frequency for all events. From the relatively small number of remaining main shocks, one concludes that the percentage of aftershocks in Japan is very high (about 80% according to the Knopoff-Kagan algorithm). The histogram of these main events with magnitudes $m \geq 5.5$ is shown in Fig. 5. This histogram of magnitudes is characterized by non-random

irregularities and a non-monotonic behavior. The irregularities force us to aggregate 0.1-bins into 0.2-bins, and the resulting discreteness in the magnitudes requires a special treatment (in particular, the use of the chi-square test), which is explained in the next subsection. On a positive note, no visible pattern associated with half-integer magnitude values can be detected. Thus, we accept that the use of

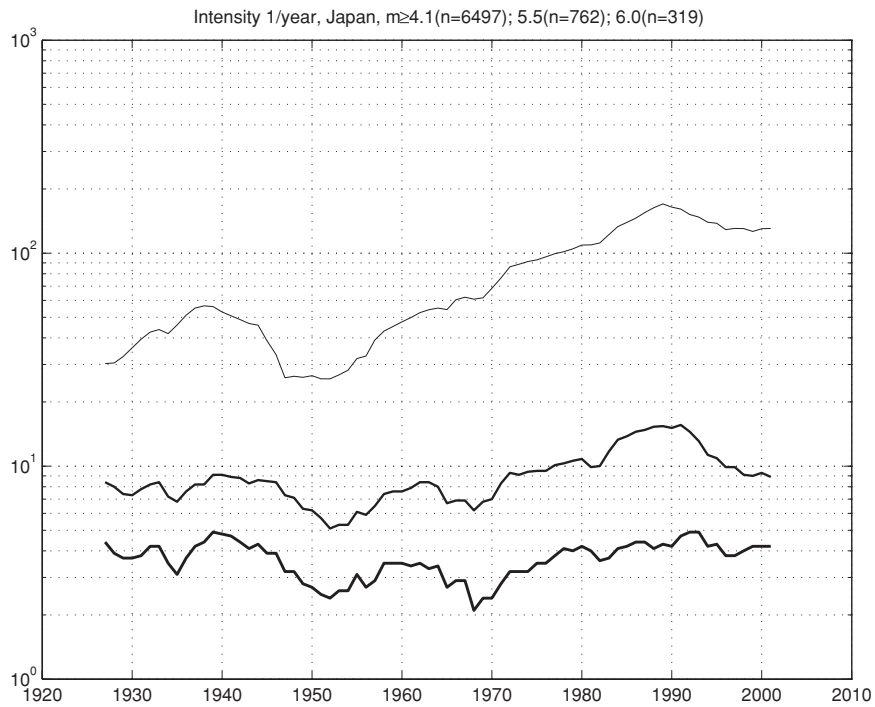


Fig. 6. Yearly number of earthquakes averaged over 10 years for three magnitude thresholds: $m \geq 4.1$ (all available events); $m \geq 5.5$; $m \geq 6.0$.

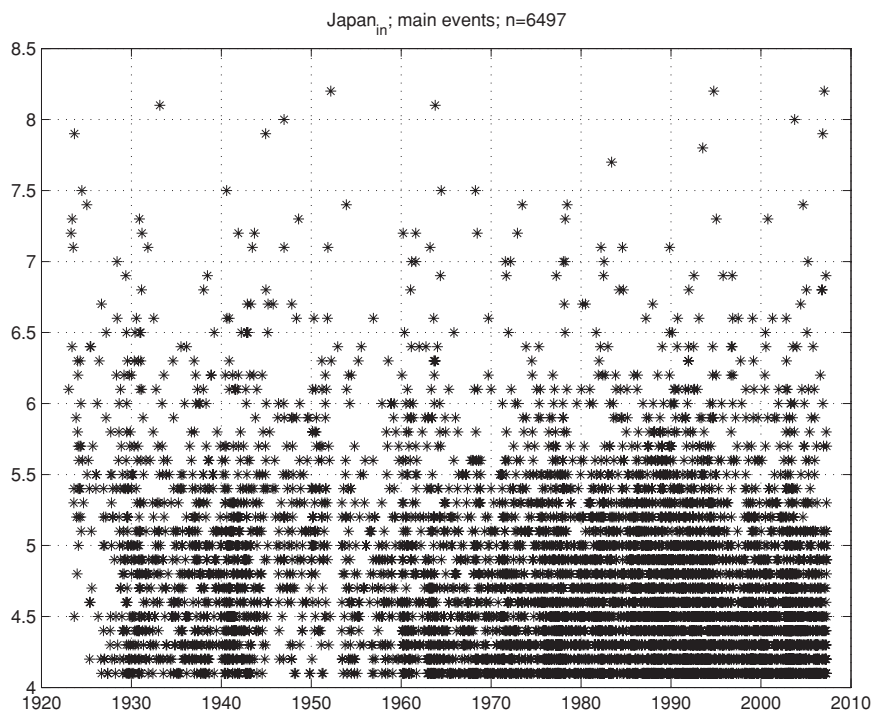


Fig. 7. Flow of main shocks from 1923 to 2007. Main shocks are defined as “shallow” earthquakes inside the polygonal domain shown in Fig. 1, whose depths are <70 km and which remain after applying the declustering Knopoff-Kagan space-time window algorithm (Knopoff and Kagan, 1977).

0.2-bins will be sufficient to remove the irregularities.

Figure 6 plots the yearly number of earthquakes averaged over 10 years for three magnitude thresholds: $m \geq 4.1$ (all available events); $m \geq 5.5$; $m \geq 6.0$. The latter time-series with $m \geq 6.0$ appears to be approximately stationary, with an intensity of about three to four events per year. Figure 7 shows the flow of main events (same variable as in Fig. 6 but for the main shocks obtained after applying the declustering

Knopoff-Kagan algorithm). For large events ($m \geq 6.0$), the flow is approximately stationary.

3.2 Adaptation for binned magnitudes

As shown in Figs. 3 and 5, the earthquake magnitudes of the JMA catalog are discrete. Moreover, the oscillations decorating the decay with magnitudes shown in Fig. 5 require further coarse-graining with bins of 0.2 units of magnitudes, as explained in the previous subsection. However,

all of the considerations identified in Section 2 refer to continuous random values, with continuous distribution functions. For discrete random variables, the EVT is not directly applicable. This contradiction is avoided as follows.

Consider a catalog in which the magnitudes are reported with a magnitude step Δm . In most existing catalogs, including that of Japan, in most cases $\Delta m = 0.1$. In some catalogs, two decimal digits are reported, but the last digit is fictitious unless the magnitudes are recalculated from seismic moments, themselves determined with several exact digits (such as for the m_W magnitude in the Harvard catalog). Here, we assume that the digitization is fulfilled exactly without random errors in intervals $((k-1) \cdot \Delta m; k \cdot \Delta m)$, where k is an integer. As a consequence, in the GPD approach, we should use only half-integer thresholds $h = (k-1/2) \cdot \Delta m$, which is not a serious restriction.

Furthermore, having a sample of observations exceeding some $h = (k-1/2) \cdot \Delta m$, and fitting the GPD to it, we need to test the goodness of fit of the GEV model to the empirical distribution. For continuous random variables, the Kolmogorov test or the Anderson-Darling test has been successfully used in earlier studies (Pisarenko *et al.*, 2008, 2009). For discrete variables, such statistical tools tailored for continuous random variables are incorrect. To demonstrate this, we calculated the Kolmogorov distances for $N = 1,000$ discrete artificial samples, each of them obeying the GEV. Our aim was to check the impact of discrete magnitudes on the Kolmogorov test. Specifically, we generate N times n synthetic random magnitudes $m_i, i = 1, \dots, n$, distributed according to the GEV distribution (relation (4)). Then, for each of the N set, we discretize the magnitudes by rounding off the random numbers with $\Delta m = 0.1$, thus mimicking the empirical data. For each of the N sets, we constructed the Kolmogorov statistic as follows. We estimated the empirical distribution function F_n for the n iid observations as $F_n(x) = (1/n) \sum_{i=1}^n I(m_i \leq x)$, where $I(m_i \leq x)$ is the indicator function, equal to 1 if $m_i \leq x$ and equal to 0 otherwise. The Kolmogorov statistics for the cumulative distribution function $\text{GEV}(x|\zeta, \sigma, \mu)$ is then given by

$$K_j = \text{Sup}_n |F_n(x) - \text{GEV}(x|\zeta, \sigma, \mu)|, \text{ where } j = 1, \dots, N$$

where Sup_n is the supremum of the set of distances. Having N realizations of K_j , we found that their distribution is very far from the true one (the Kolmogorov distances for discrete magnitudes are much larger than those for continuous random variables.).

This result shows that in our analysis we are forced to use statistical tools adapted to discrete random variables. We have chosen the standard Pearson chi-square (χ^2) method as it provides a way to both estimate unknown parameters and strictly evaluate the goodness of fit. The χ^2 -statistic is calculated by finding the difference between each observed and theoretical frequency for each possible magnitude bin, then squaring each difference, dividing it by the theoretical frequency, and taking the sum of the results. The χ^2 -statistics is then distributed according to the χ^2 -distribution with $n-3$ degrees of freedom (df) since we estimate three parameters in fitting the theoretical GEV distribution.

The chi-square test has two specific requirements:

1. In order to be able to apply the chi-square test, a sufficient number of observations is needed in each bin (we chose this minimum number as being equal to 8 (see discussion of this matter in Borovkov (1987)));
2. In order to compare two different fits (corresponding to two different vectors of parameters), it is highly desirable to have the same binning in both experiments in order to avoid large variations in the significance levels, which depend on the binning.

In general, the chi-square test is less sensitive and less efficient than the Kolmogorov test or the Anderson-Darling test due to the fact that the chi-square test coarsens data by placing data into discrete bins.

When using the GEV, the digitized GEV of the magnitude maxima in successive T -intervals is fitted using the χ^2 -method.

3.3 The GPD approach

Consider the discrete set of magnitudes registered with step Δm over threshold h ,

$$\begin{aligned} h + (k-1)\Delta m/2 \leq m < h + k\Delta m/2; \\ k = 1, \dots, r; \quad \Delta m = 0.1. \end{aligned} \quad (18)$$

The corresponding discrete probabilities read

$$\begin{aligned} p_k(\xi, s|h) &= P\{h + (k-1) \cdot 0.05 \leq m < h + k \cdot 0.05\} \\ &= \text{GPD}(h + k \cdot 0.05|\xi, s, h) \\ &\quad - \text{GPD}(h + (k-1) \cdot 0.05|\xi, s, h); \end{aligned} \quad (19)$$

$$p_{r+1}(\xi, s|h) = 1 - \text{GPD}(h + r \cdot 0.05|\xi, s, h). \quad (20)$$

The last $(r+1)$ -th bin covers the interval $(h + r \cdot 0.05; \infty)$. We use the following expression

$$\begin{aligned} \text{GPD}(x|\xi, s, h) &= 1 - [1 + (\xi/s)(x-h)]^{-1/\xi}, \\ h \leq x \leq h + s/\xi, \quad \xi < 0. \end{aligned} \quad (21)$$

Let us assume that the interval (Eq. (18)) contains n_k observations. Summing over the $r+1$ intervals, the total number of observations is $n = n_1 + n_2 + \dots + n_r + n_{r+1}$. The chi-square sum $S(\xi, s)$ is then written as:

$$S(\xi, s) = \sum_{k=1}^{r+1} [n_k - n \cdot p_k(\xi, s|h)]^2 / n \cdot p_k(\xi, s|h), \quad (22)$$

$S(\xi, s)$ should be minimized over the parameters (ξ, s) . This minimum value is distributed according to the χ^2 -distribution with $(r-2)$ df. The quality of the fit of the empirical distribution by expressions (19) and (20) is quantified by the probability $P_{\text{exc}} = P\{\chi^2(r-2) \geq \min(S)\}$, where $\chi^2(r-2)$ is the chi-square random value with $(r-2)$ df, i.e. P_{exc} is the probability of exceeding the minimum fitted chi-square sum. The larger the P_{exc} , the better the goodness of fit.

For magnitude thresholds $h \leq 5.95$ and $h \geq 6.65$, the chi-square sums $\min(S)$ happened to be very large, leading to very small P_{exc} values and indicating that such thresholds are not acceptable. For thresholds in the interval $(6.05 \leq h \leq 6.55)$, the results of the chi-square fitting procedure are shown in Table 1. In order to obtain confidence intervals, we also performed $N_b = 100$ bootstrapping procedures on

Table 1. Chi-square fitting procedure using the GPD approach.

h	6.05	6.15	6.25	6.35	6.45
r	7	7	6	6	6
degrees of freedom	5	5	4	4	4
ξ	-0.0468	-0.2052	-0.2137	-0.2264	-0.1616
s	0.5503	0.6420	0.6397	0.6264	0.6081
M_{\max}	17.87	9.43	9.31	9.11	10.20
$Q_{0.90}(10)$	8.73	8.32	8.29	8.24	8.52
P_{exc}	0.0753	0.2791	0.3447	0.3378	0.1747

The parameters are estimated by minimizing $S(\xi, s)$, defined by expression (21). M_{\max} is the rightmost point of the magnitude distribution given by expression (3). $Q_{0.90}(10)$ is the 90% quantile of the maximum magnitude distribution (T -maximum magnitude) in 10-year intervals.

our initial sample and averaged the results over the obtained estimates, as described in Pisarenko (2008, 2009).

As pointed out above, if the distribution of magnitudes over thresholds obeys the $GPD(x|\xi, s, h)$, then, for a Poissonian flow of events, the T -maxima have the GEV distribution:

$$GEV(x|\xi, \sigma_T, \mu_T) = \exp \left\{ -[1 + (\xi/\sigma_T)(x - \mu_T)]^{-1/\xi} \right\},$$

$$x \leq h - \sigma_T/\xi, \quad \xi < 0. \tag{23}$$

Thus, we can use an alternative approach, the GEV, to fit the sample of T -maxima derived from the same underlying catalog.

Having estimated the first triple (ξ, σ_T, μ_T) or the second triple (ξ, s, h) , we use these estimates to predict the quantile of τ -maxima for any arbitrary future time interval $(0, \tau)$, since these τ -maxima have the distribution $GEV(x|\xi, \sigma_\tau, \mu_\tau)$, as seen from Eqs. (6)–(13). Recall that, in Eqs. (6)–(13), λ denotes the intensity of the Poissonian flow of events whose magnitudes exceed the threshold h .

In Table 1, three thresholds $h = 6.15$; $h = 6.25$, and $h = 6.35$ give very close estimates. In contrast, the estimates obtained for the thresholds $h = 6.05$ and $h = 6.45$ have smaller goodness of fit (smaller P_{exc}), suggesting that the estimates corresponding to the highest goodness of fit ($h = 6.25$) should be accepted:

$$\xi_{\text{GPD}} = -0.2137; \quad s_{\text{GPD}} = 0.6397; \quad M_{\max, \text{GPD}} = 9.31;$$

$$Q_{0.90, \text{GPD}}(10) = 8.29. \tag{24}$$

These estimates are very close to their mean values obtained over the three thresholds $h = 6.15$; 6.25 ; 6.35 .

In order to estimate the statistical scatter of these estimates, we simulated our whole procedure of estimation $N_b = 100$ times on artificial GPD samples with known parameters. For a better stability, instead of sample standard deviations, we used the corresponding order statistics, namely, the difference of quantiles:

$$(Q_{0.84} - Q_{0.16})/2. \tag{25}$$

For Gaussian distributions, this quantity (Eq. (25)) coincides with its standard deviation (SD). For distributions

with heavy tails, the difference (Eq. (25)) is a more robust estimate of the scatter than the usual SD. Combining the scatter estimates (Eq. (25)) derived from simulations to the mean values (Eq. (24)), the final results of the GPD approach for the JMA catalog can be summarized by

$$\xi_{\text{GPD}} = -0.2137 \pm 0.1031; \quad s_{\text{GPD}} = 0.6397 \pm 0.0634;$$

$$M_{\max, \text{GPD}} = 9.31 \pm 1.14; \quad Q_{0.90, \text{GPD}}(10) = 8.29 \pm 0.49; \tag{26}$$

One can observe that the statistical scatter of M_{\max} exceeds the scatter of the quantile $Q_{0.90}(10)$ by a factor of more than two, thereby confirming once more our earlier conclusion on the instability of M_{\max} .

3.4 The GEV approach

In this approach, we divide the total time interval T_c from 1923 to 2007 covered by the catalog into a sequence of non-overlapping and touching intervals of length T . The maximum magnitude $M_{T,j}$ in each T -interval is identified. We have $k = [T_c/T]$ T -intervals, so the sample of our T -maxima has size k : $M_{T,1}, \dots, M_{T,k}$. We assume that T is large enough, so that each $M_{T,j}$ can be considered as being sampled from the GEV distribution $GEV(x|\xi, \sigma_T, \mu_T)$ with some unknown parameters (ξ, σ_T, μ_T) that should be estimated through the sample $M_{T,1}, \dots, M_{T,k}$.

The larger T is, the more accurate is the approximation for this observed sample, but one cannot choose too large a T because the sample size k of the set of T -maxima would be too small, resulting in an inefficient statistical estimation of the three unknown parameters (ξ, σ_T, μ_T) . Besides, we should keep in mind the restrictions mentioned above, imposed by the chi-square method, that the number of bins should be constant for all used T values and that the minimum number of observations per bin should not be < 8 . In order to satisfy these contradictory constraints, as a compromise, we had to restrict the T -values to be sampled in the rather small interval

$$200 \leq T \leq 300 \text{ days.} \tag{27}$$

It should be noted that, for all T -values > 50 days, the estimates of the parameters do not vary much and that only for $T \leq 40$ do the estimates change drastically. We have

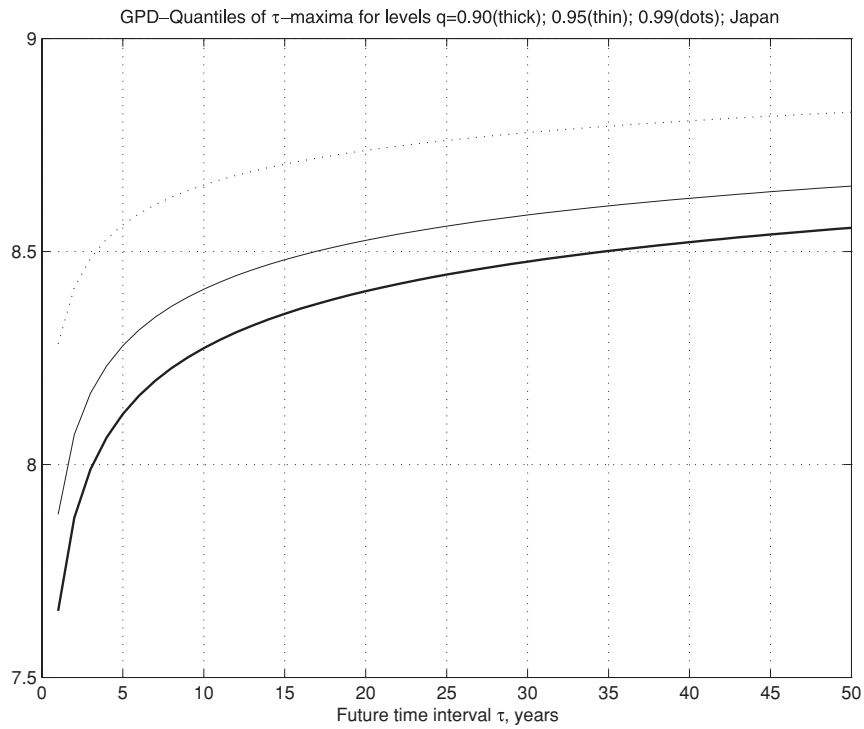


Fig. 8. Quantile $Q_q(\tau)$ of the distribution of maxima over a future time interval τ for three confidence levels q , defined by expression (16). The three curves use the parameters of the GPD estimated from the JMA catalog, as explained in the text.

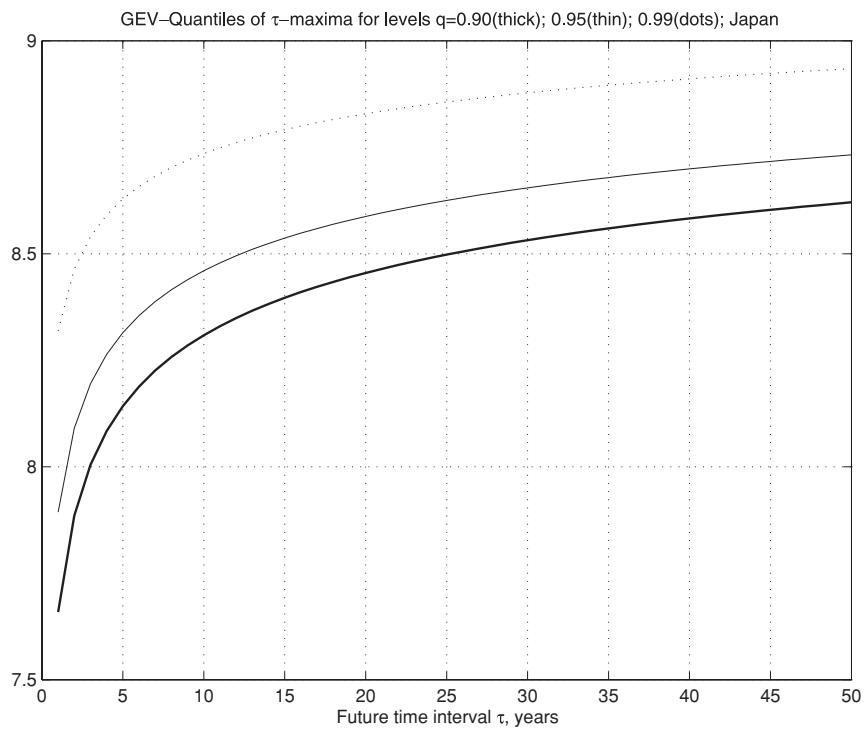


Fig. 9. Quantile $Q_q(\tau)$ of the distribution of maxima over a future time interval τ for three confidence levels q , defined by expression (17). The three curves use the parameters of the GEV estimated from the JMA catalog, as explained in the text.

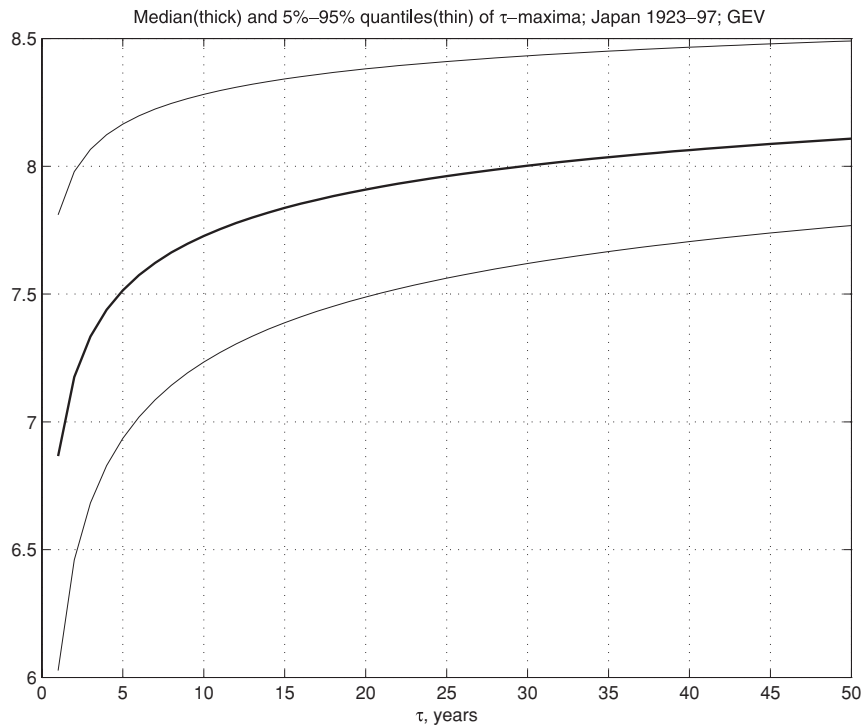


Fig. 10. Median (quantile $Q_q(\tau)$ for $q = 50\%$) of the distribution of the maximum magnitude over a future time interval τ , obtained by the GEV method, as a function of τ (years), together with the two accompanying quantiles 16% and 84% that correspond to the usual ± 1 SD.

chosen $T = 200$ and obtained the following estimates:

$$\begin{aligned} \xi_{\text{GEV}} &= -0.1901 \pm 0.0717; \\ \mu_{\text{GEV}}(200) &= 6.3387 \pm 0.0380; \\ \sigma_{\text{GEV}}(200) &= 0.5995 \pm 0.0223; \\ M_{\text{max,GEV}} &= 9.57 \pm 0.86; \\ Q_{0.90,\text{GEV}}(10) &= 8.34 \pm 0.32; \end{aligned} \quad (28)$$

The estimates of the scatter in Eq. (28) were obtained by the simulation method with 100 realizations, similar to the method used in the GPD approach. In estimating the parameters, we have used the *shuffling* procedure described in Pisarenko *et al.* (2009), which is similar to the bootstrap method, with $N_S = 100$ realizations. It should be noted that, in Eq. (28), the T -value for the parameters μ , σ is indicated in days ($T = 200$ days) whereas in the quantile Q , the τ -value is indicated in years ($\tau = 10$ years).

Comparing ξ , M_{max} and the Q -estimates obtained by the GPD and the GEV approaches, the GEV method is found to be somewhat more efficient (its scatter is smaller by a factor approximately equal to 0.7). This can be explained by the fact that the GEV approach uses the full catalog more intensively: all events with magnitude $m \geq 4.1$ participate (in principle) in the estimation, whereas the GPD approach throws out all events with $m < h$.

Finally, we show in Figs. 8 and 9 the dependence of the quantile $Q_q(\tau)$ as a function of τ , for $\tau = 1-50$ years, as estimated by our two approaches, respectively given by expressions (16) and (17). One can observe that the quantile $Q_q(\tau)$ obtained by the two methods are very close, which testifies to the stability of the estimations. Figure 10 plots the median (quantile $Q_q(\tau)$ for $q = 50\%$) of the distribution of the maximum magnitude as a function of the future

τ years, together with the two accompanying quantiles 16% and 84%, which correspond to the usual ± 1 SD. These quantiles $Q_q(\tau)$ can be very useful tools for pricing risks in the insurance business and for optimizing the allocation of resources and preparedness by state governments.

4. Discussion and Conclusions

We have adapted the new method of statistical estimation suggested by Pisarenko *et al.* (2009) to earthquake catalogs with discrete magnitudes. This method is based on the duality of the two main limit theorems of EVT. One theorem leads to the GPD (peak over threshold approach), and the other theorem leads to the GEV (T -maximum method). Both limit distributions must possess the same form parameter ξ . For the Japanese catalog of earthquake magnitudes over the period 1923–2007, both approaches provide almost the same statistical estimate for the form parameter, which is found to be negative; $\xi \cong -0.2$. A negative form parameter corresponds to a distribution of magnitudes that is bounded from above (by a parameter named M_{max}). This maximum magnitude corresponds to the finiteness of the geological structures supporting earthquakes. The density distribution extends to its final value M_{max} with a very small probability weight in its neighborhood, characterized by a tangency of a high degree (“duck beak” shape). In fact, the limit behavior of the density distribution of Japanese earthquake magnitudes is described by the function $(M_{\text{max}} - m)^{-1-1/\xi} \cong (M_{\text{max}} - m)^4$, i.e. by a polynomial of degree approximately equal to 4. This is the explanation of the unstable character of the statistical estimates of the parameter M_{max} : a small change in the catalog of earthquake magnitude can give rise to a significant

fluctuation in the resulting estimate of M_{\max} . In contrast, the estimation of the integral parameter $Q_q(\tau)$ is generally more stable and robust, as we demonstrate quantitatively for the Japanese catalog of earthquake magnitudes over the period 1923–2007.

The main problem in the statistical study of the tail of the distribution of earthquake magnitudes (as well as in distributions of other rarely observable extremes) is the estimation of quantiles that exceed the data range, i.e. quantiles of level $q > 1 - 1/n$, where n is the sample size. We would like to stress once more that the reliable estimation of quantiles of levels $q > 1 - 1/n$ can be made only with some additional assumptions on the behavior of the tail. Sometimes, such assumptions can be made on the basis of physical processes underlying the phenomena under study. For this purpose, we used general mathematical limit theorems, namely, the theorems of EVT. In our case, the assumptions for the validity of EVT amount to assuming a regular (power-like) behavior of the tail $1 - F(m)$ of the distribution of earthquake magnitudes in the vicinity of its rightmost point M_{\max} . Partial justification for such an assumption is the fact that, without it, there is no meaningful limit theorem in EVT. Of course, there is no a priori guarantee that these assumptions will hold in all real situations, and they should be discussed and possibly verified or supported by other means. In fact, because EVT suggests a statistical methodology for the extrapolation of quantiles beyond the data range, the question of whether such interpolation is justified or not in a given problem should be investigated carefully in each concrete situation.

Acknowledgments. This work was partially supported (V. F. Pisarenko, M. V. Rodkin) by the Russian Foundation for Basic research, grant 09-05-01039a, and by the Swiss ETH CCES project EXTREMES (DS).

References

- Borovkov, A. A., *Statistique Mathématique*, Moscow, Mir, 1987.
- Cosentino, P., V. Ficara, and D. Luzio, Truncated exponential frequency-magnitude relationship in the earthquake statistics, *Bull. Seismol. Soc. Am.*, **67**, 1615–1623, 1977.
- Embrechts, P., C. Kluppelberg, and T. Mikosch, *Modelling Extremal Events*, Springer, 1997.
- Epstein, B. C. and C. Lomnitz, A model for the occurrence of large earthquakes, *Nature*, **211**, 954–956, 1966.
- Kagan, Y. Y., Universality of the seismic moment-frequency relation, *Pure Appl. Geophys.*, **155**, 537–573, 1999.
- Kagan, Y. Y. and F. Schoenberg, Estimation of the upper cutoff parameter for the tapered distribution, *J. Appl. Probab.*, **38A**, 901–918, 2001.
- Kijko, A., Estimation of the maximum earthquake magnitude, *Mmax*, *Pure Appl. Geophys.*, **161**, 1–27, 2004.
- Kijko, A. and M. A. Sellevoll, Estimation of earthquake hazard parameters from incomplete data files. Part I, Utilization of extreme and complete catalogues with different threshold magnitudes, *Bull. Seismol. Soc. Am.*, **79**, 645–654, 1989.
- Kijko, A. and M. A. Sellevoll, Estimation of earthquake hazard parameters from incomplete data files. Part II, Incorporation of magnitude heterogeneity, *Bull. Seismol. Soc. Am.*, **82**, 120–134, 1992.
- Knopoff, L. and Y. Kagan, Analysis of the extremes as applied to earthquake problems, *J. Geophys. Res.*, **82**, 5647–5657, 1977.
- Pisarenko, V. F., A. A. Lyubushin, V. B. Lysenko, and T. V. Golubeva, Statistical estimation of seismic hazard parameters: maximum possible magnitude and related parameters, *Bull. Seismol. Soc. Am.*, **86**, 691–700, 1996.
- Pisarenko, V. F., A. Sornette, D. Sornette, and M. V. Rodkin, New approach to the characterization of M_{\max} and of the tail of the distribution of earthquake magnitudes, *Pure Appl. Geophys.*, **165**, 847–888, 2008.
- Pisarenko, V. F., A. Sornette, D. Sornette, and M. V. Rodkin, Characterization of the tail of the distribution of earthquake magnitudes by combining the GEV and GPD descriptions of extreme value theory, *Pure Appl. Geophys.*, (<http://arXiv.org/abs/0805.1635>), 2009.
- Sornette, D., L. Knopoff, Y. Y. Kagan, and C. Vanneste, Rank-ordering statistics of extreme events : application to the distribution of large earthquakes, *J. Geophys. Res.*, **101**, 13883–13893, 1996.
- Sornette, D., A. B. Davis, K. Ide, K. R. Vixie, V. Pisarenko, and J. R. Kamm, Algorithm for model validation: Theory and applications, *Proc. Natl. Acad. Sci. USA*, **104**(16), 6562–6567, 2007.
- Sornette, D., A. B. Davis, J. R. Kamm, and K. Ide, A general strategy for physics-based model validation illustrated with earthquake phenomenology, atmospheric radiative transfer, and computational fluid dynamics, in *Book series: Lecture Notes in Computational Science and Engineering, vol 62, Book Series: Computational Methods in Transport: Verification and Validation*, edited by F. Graziani and D. Swesty, pp. 19–73, Springer, New York (NY), (<http://arxiv.org/abs/0710.0317>), 2008.
- Stephens, M. A., EDF Statistics for Goodness of Fit and Some Comparisons, *J. Am. Statist. Soc.*, **69**(347), 730–737, 1974.

V. F. Pisarenko, D. Sornette (e-mail: dsornette@ethz.ch), and M. V. Rodkin