



The Estimation of Probability of Extreme Events for Small Samples

V. F. PISARENKO¹ and M. V. RODKIN^{1,2}

Abstract—The most general approach to the study of rare extreme events is based on the extreme value theory. The fundamental General Extreme Value Distribution lies in the basis of this theory serving as the limit distribution for normalized maxima. It depends on three parameters. Usually the method of maximum likelihood (ML) is used for the estimation that possesses well-known optimal asymptotic properties. However, this method works efficiently only when sample size is large enough (~ 200 – 500), whereas in many applications the sample size does not exceed 50–100. For such sizes, the advantage of the ML method in efficiency is not guaranteed. We have found that for this situation the method of statistical moments (SM) works more efficiently over other methods. The details of the estimation for small samples are studied. The SM is applied to the study of extreme earthquakes in three large virtual seismic zones, representing the regime of seismicity in subduction zones, intracontinental regime of seismicity, and the regime in mid-ocean ridge zones. The 68%-confidence domains for pairs of parameter (ξ, σ) and (σ, μ) are derived.

1. Introduction

The statistical study of extreme events is necessary in many applications. Such extreme events are usually connected with the most powerful natural phenomena or with the most disastrous catastrophes. It seems that the most general and the most justified approach to statistical study of such events is based on the extreme value theory (EVT) and two of its main limit distributions: the Generalized Extreme Value Distribution (GEV) and the Generalized Pareto Distribution (GPD). Each of these distributions depends on three parameters.

The GEV parameters:

- the form parameter ξ , $-\infty < \xi < +\infty$;
- the location parameter μ , $-\infty < \mu < +\infty$;
- the scale parameter σ , $0 < \sigma < +\infty$.

The GPD parameters:

- the form parameter ξ , $-\infty < \xi < +\infty$;
- the lower threshold parameter h , $-\infty < h < +\infty$;
- the scale parameter s , $0 < s < +\infty$.

The GEV describes the limit distribution of normalized maxima of increasing samples, whereas the GPD appears as the limit distribution of scaled excesses over sufficiently high threshold h values. There is a close connection between these two distributions. A Poisson flow of events having a GPD distribution is distributed in accordance with GEV distribution. In this case, the form parameter ξ of both distributions is the same, but two other parameters differ. There are simple formulas connecting these parameters (see Embrechts et al. 1997; Pisarenko and Rodkin 2010). The practical use of the GPD and the GEV is aggravated often by small sample size of available data. The small size situation for the GPD was studied by Pisarenko et al. (2017) where the quantile method (QM) was found to be the most efficient. In this paper, we concentrate on the case of the GEV.

The GEV distribution function has a form:

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

If $\xi > 0$, then the function $1 - F(x)$ decreases as $x^{-1/\xi}$. If $\xi = 0$, then we have usual exponential distribution $F(x) = 1 - \exp \left\{ -\frac{x-\mu}{\sigma} \right\}$. It should be remarked that the normal distribution also leads to $\xi = 0$. If $\xi < 0$, then one has distribution restricted

¹ Institute of Earthquake Prediction Theory and Mathematical Geophysics, Russian Academy of Sciences, Profsoyuznaya 84/32, Moscow 117997, Russia. E-mail: rodkin@mitp.ru

² Institute of Marine Geology and Geophysics, FEB RAS, ul. Nauki, 1B, Yuzhno-Sakhalinsk 693022, Russia.

from above by $\mu - \sigma/\xi$. We denote this limit as $M_{\max} = \mu - \sigma/\xi$. It should be noted that the case of power-like decrease ($\xi > 0$) can be reduced to the case $\xi = 0$ by logarithmic transform. If a random variable X has power distribution (the Pareto distribution) $F(x) = 1 - (h/x)^b, x \geq h$, then $\log(X)$ has the exponential distribution. Thus, it is sufficient to consider only the case $\xi \leq 0$.

Some examples of GEV probability densities with non-positive ξ are shown in Fig. 1.

The most widespread method for parameter estimation is the standard maximum likelihood method. However, its application for the GEV is complicated by a few circumstances. First, the support of the underlying distribution function depends on the unknown parameters [see Eq. (1)]. Second, it was shown that the classical good properties of the maximum likelihood estimators (MLE) of the GEV parameters hold whenever $\xi > -0.5$ and are being lost for $\xi \leq -0.5$. Having this in mind we shall restrict our consideration by the former case. It should be noted that practically interesting cases usually meet this condition, and thus this restriction seems to be unessential. The third circumstance is connected with the sample size because the MLE

method is not guaranteed to be the most efficient for small samples. The MLE for the GEV were studied in (Smith and Naylor 1987; Buishand 1989; Smith 1990; Embrechts et al. 1997). The method of statistical moment estimation and the method of weighted moments were suggested by Hosking et al. (1985) and Christopheit (1994). However, the efficiency of these approaches for the case of small samples still remains uncertain, and this case is just very important for many practical situations where sample size consists only 30–100 [see (Pisarenko and Rodkin 2013)]. A very annoying property of small samples occurred because of a strong bias of the MLE and other methods. Thus, the new approaches for this situation are very crucial. In this paper we study in detail the case of small samples and suggest some new modifications of the statistical moment (SM) estimation procedure based on the bootstrap ideas. In the study by Pisarenko et al. (2017) it was found that the quantile method (QM) works more efficiently for small samples in the case of GPD. In this paper we study the problem of estimation of the GEV parameters for small and moderate (20–500) samples. We have found that for this situation the method of statistical moments (SM) works more efficiently than ML and QM estimates. We restrict ourselves below by the comparison of efficiency of SM and ML estimates.

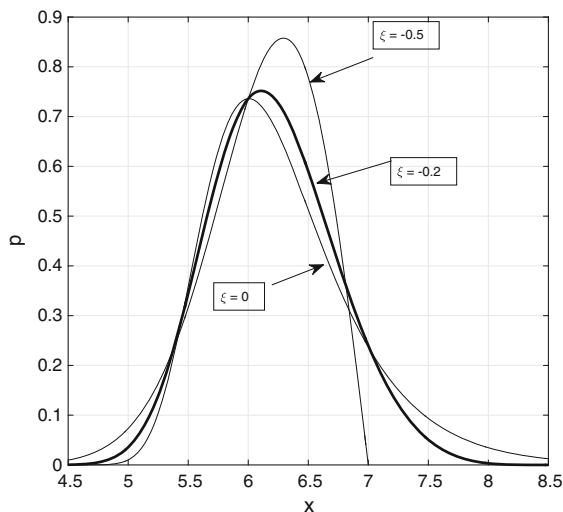


Figure 1

Examples of probability GEV densities $p(x)$ with a few characteristic non-positive ξ ; x values are chosen to be similar with m_w values; the most typical case $\xi = -0.2$ is given by a thick line; $\mu = 6; \sigma = 0.5$

2. The ML Estimates for the GEV

The GEV log-likelihood for a sample (x_1, \dots, x_n) has form

$$L(x|\xi, \sigma, \mu) = -n \cdot \log(\sigma) - \sum_{k=1}^n \left[1 + \xi \frac{x_k - \mu}{\sigma} \right]^{-1/\xi} - (1 + 1/\xi) \sum_{k=1}^n \log \left[1 + \xi \frac{x_k - \mu}{\sigma} \right]. \tag{2}$$

The ML estimators are found as values (ξ, σ, μ) maximizing (2) under conditions:

$$\sigma > 0, \quad 1 + \frac{\xi}{\sigma}(x_k - \mu) \geq 0. \tag{3}$$

3. The SM Estimates for the GEV

We denote the three first statistical moments by M_1, M_2, M_3 :

$$M_1 = \frac{1}{n} \sum_{k=1}^n x_k; \quad M_2 = \frac{1}{n} \sum_{k=1}^n x_k^2; \quad (4)$$

$$M_3 = \frac{1}{n} \sum_{k=1}^n x_k^3.$$

Their theoretic analogs are following:

$\Upsilon_1 = \mu - \sigma/\xi - \sigma \cdot \Gamma(-\xi)$; Γ is the gamma function;

$\Upsilon_2 = (\mu - \sigma/\xi)^2 - 2\sigma(\mu - \sigma/\xi) \cdot \Gamma(-\xi) - 2(\sigma^2/\xi) \Gamma(-2\xi)$;

$$\Upsilon_3 = (\mu - \sigma/\xi)^3 - 3\sigma(\mu - \sigma/\xi)^2 \cdot \Gamma(-\xi) - 6(\sigma^2/\xi)(\mu - \sigma/\xi) \cdot \Gamma(-2\xi) - 3(\sigma^3/\xi^2) \Gamma(-3\xi). \quad (5)$$

Equalizing (4) and (5) we get the system of three equations for three unknown parameters. This system is easily solved. In fact, the skew $E\left(\frac{X-EX}{\text{std}(X)}\right)^3$ depends only on ξ :

$$E\left(\frac{X-EX}{\text{std}(X)}\right)^3 = \frac{-\Gamma(1-3\xi) + 3\Gamma(1-\xi) - 2\Gamma^3(1-\xi)}{[\Gamma(1-2\xi) - \Gamma^2(1-\xi)^{3/2}]}, \quad (6)$$

where E is mathematical expectation symbol and std is a standard deviation. Equalizing (6) to the sample skew

$$\frac{M_3 - 3M_2M_1 + 2M_1^3}{(M_2 - M_1^2)^{3/2}}. \quad (7)$$

we get equation for ξ :

$$\frac{M_3 - 3M_2M_1 + 2M_1^3}{(M_2 - M_1^2)^{3/2}} = \frac{-\Gamma(1-3\xi) + 3\Gamma(1-\xi) - 2\Gamma^3(1-\xi)}{[\Gamma(1-2\xi) - \Gamma^2(1-\xi)^{3/2}]} \quad (8)$$

Equation (8) is solved numerically which provides the estimate $\hat{\xi}$ of parameter ξ .

Then we get estimators for two other parameters:

$$\hat{\sigma} = \left[\frac{M_2 - M_1^2}{-\Gamma^2(-\hat{\xi}) - \left(2/\hat{\xi}\right)\Gamma(1-2\hat{\xi})} \right]^{1/2}. \quad (9)$$

$$\hat{\mu} = M_1 \frac{\hat{\sigma}}{\hat{\xi}} + \hat{\sigma} \cdot \Gamma(-\hat{\xi}). \quad (10)$$

The third moment Υ_3 is finite for $\xi < 1/3$. Since we assumed $\xi \leq 0$, this condition is satisfied and we have included it as well into our numerical algorithm. Under this condition all three sample moments converge with probability one to corresponding theoretical moments as n tends to infinity. Thus, the SM estimates (8)–(10) converge as well to their true values with probability one.

We used the bootstrap variant with replacement (Efron 1979). In bootstrap sample (y_1, \dots, y_n) each y_k was randomly chosen from the initial sample (x_1, \dots, x_n) . We have taken usually 10,000 bootstrap samples and averaged obtained estimates. Bootstrap samples are not independent, but the mathematical expectation of a sum equals to sum of expectations for dependent terms too.

We have modified the SM by adding a very useful procedure. To estimate std we divided each bootstrap sample into two equal subsamples $(x_1^{(1)}, \dots, x_{n/2}^{(1)})$, $(x_1^{(2)}, \dots, x_{n/2}^{(2)})$. Then, we got estimate of variance:

$$\text{var} = \frac{1}{n} \sum_{k=1}^{n/2} (x_k^{(1)} - x_k^{(2)})^2, \quad (11)$$

taking into account that $E(x_k^{(1)} - x_k^{(2)})^2 = 2\text{var}$, so that var is an unbiased estimate of the true variance var. Then we averaged estimates (11) over all bootstrap ensembles. The similar method was used to estimate standard deviation of the quantiles $Q_q(\tau)$ (see below).

Similarly, to estimate the significance level of the Anderson–Darling distance (ADD) that was used in our procedure we put parameter estimates derived from the first subsample into the DF and used this DF as “theoretical DF” for the second subsample. Thus, each time theoretical DF and the used subsample were independent. Of course, there was some loss of efficiency (because the sample size is two times less), so that our procedure gives slightly lowered significance value.

4. Comparison of the MLE with the SM

We have carried out the comparison on artificial GEV samples generated with the help of computer random numbers.

The quality of statistical estimates is defined by three characteristics:

1. The mean-square error MSE

$$\text{MSE} = \left[E \left(\hat{\xi} - \xi \right)^2 \right]^{1/2}, \text{ where } E \text{ is mathematical expectation symbol.}$$

2. The bias

$$\text{BIAS} = E \left(\hat{\xi} \right) - \xi.$$

3. The standard deviation

$$\text{STD} = \left[E \left(\hat{\xi} - E(\hat{\xi}) \right)^2 \right]^{1/2}.$$

These statistical characteristics are related by equation:

$$\text{MSE}^2 = \text{BIAS}^2 + \text{STD}^2. \quad (12)$$

Thus, knowing two of them one can determine the third. The quality of an estimate is characterized fully by MSE, but sometimes two other statistical characteristics are of interest too. Analyzing estimates of the GEV parameters we concentrate mainly on the form parameter ξ since its estimation is most difficult. The estimation of parameter σ and especially parameter μ is much easier and certain. We carried out estimation of ξ within interval $-0.5 < \xi \leq 0$. Since the centered and normalized term $(x-\mu)/\sigma$ is entered in the GEV we can study estimates of parameter ξ for any fixed σ, μ . The statistical properties of ξ -estimate do not depend on them.

The results of the estimation are collected in Tables 1, 2, 3, 4, 5, and 6. The averaged over $\xi, -0.5 < \xi \leq 0$, MSE and absolute Bias for the Maximum Likelihood and the Moment Method are shown in Figs. 2 and 3. We see that ML has larger average MSE for $n \sim < 100$, and larger average bias over all considered range. These estimates for two other parameters σ, μ behave similarly. Thus, we can conclude that for small samples ($n \leq 100$) the SM estimates are more efficient than the ML estimates, at least within the ξ -range $(-0.5, 0]$ that is the most interesting for applications [see (Pisarenko and Rodkin 2010, 2013)].

5. The GEV Estimates for Different Seismic Zones

To illustrate the exposed SM-method we have formed three large samples of seismic magnitudes,

representing the seismic regime of three different seismic–tectonic situations: subduction zone, inter-continental seismicity and mid-ocean ridge zone. We have aimed both to illustrate the application of the SM-method and to check the conclusion (Kagan 1999) on the universality of the seismic moment–frequency relation in all seismic areas excepting the specific character of seismicity in the mid-ocean ridge zones.

We restricted the depth by 70 km. The moment magnitudes m_w were taken from the GCMT seismic moment catalog (<http://www.globalcmt.org/CMTsearch.html>); time interval 01.01.1976–12.01.2016 was examined. There were 23,379 events, $4.65 \leq m_w \leq 9.08$. We have cleaned the catalog from aftershocks and swarms with the help of the Kagan–Knopoff algorithm (Knopoff et al. 1982) which reduced the number of events down to 8954 (we call them main shocks). According to our method [see for details (Pisarenko and Rodkin 2010)] the catalog time interval is divided into n equal blocks, and maximum event m_k is taken in each block. Thus, we get a sample (m_1, \dots, m_n) . According to the theory of extreme events the distribution of m_k tends to the GEV as the length of block goes to infinity. Thus, we try to fit a GEV to this sample varying n and choosing block length that should be sufficiently large. We used both the ML and SM methods with the bootstrap modification described above. The number of bootstrap samples was 10,000. The goodness-of-fit was measured by the Anderson–Darling distance ADD between sample DF $F_n(x)$ and theoretical GEV $F(x)$:

$$\begin{aligned} \text{ADD} &= n \int_{-\infty}^{\infty} W(x) \cdot [F_n(x) - F(x)]^2 dF(x) \\ &= -n \\ &\quad - \frac{1}{n} \sum_{k=1}^n (2k-1) [\log(x_k) + \log(1-x_{n+1-k})], \end{aligned} \quad (13)$$

where $x < x_2 < \dots < x_n$ is ordered sample (m_1, \dots, m_n) ; the weighting function $W(x) = \frac{1}{F(x)[1-F(x)]}$ [see (Anderson and Darling 1954)]. We prefer to use just the ADD rather than more traditional measures of goodness-of-fit like, say, the Cramer-von Mises distance with $W(x) \equiv 1$ or the

Table 1
Mean-square error MSE values $MSE_{ML}(\hat{\xi})$ and $MSE_{SM}(\hat{\xi})$ as functions of the true values ξ and n

ξ	-0.499	-0.40	-0.30	-0.20	-0.10	0
$n = 20$	0.266	0.254	0.240	0.236	0.237	0.243
$MSE_{ML}(\hat{\xi})$						
$MSE_{SM}(\hat{\xi})$	<u>0.164</u>	<u>0.141</u>	<u>0.134</u>	<u>0.131</u>	<u>0.137</u>	<u>0.156</u>
$n = 50$	0.112	0.109	0.107	0.109	0.114	0.119
$MSE_{ML}(\hat{\xi})$						
$MSE_{SM}(\hat{\xi})$	<u>0.110</u>	<u>0.097</u>	<u>0.088</u>	<u>0.086</u>	<u>0.089</u>	<u>0.103</u>
$n = 100$	<u>0.069</u>	<u>0.067</u>	0.068	0.070	0.074	0.077
$MSE_{ML}(\hat{\xi})$						
$MSE_{SM}(\hat{\xi})$	0.081	0.071	<u>0.063</u>	<u>0.061</u>	<u>0.066</u>	<u>0.075</u>
$n = 200$	<u>0.044</u>	<u>0.043</u>	<u>0.045</u>	0.046	0.049	<u>0.052</u>
$MSE_{ML}(\hat{\xi})$						
$MSE_{SM}(\hat{\xi})$	0.058	0.051	<u>0.045</u>	<u>0.043</u>	<u>0.047</u>	0.056
$n = 500$	<u>0.026</u>	<u>0.026</u>	<u>0.026</u>	<u>0.027</u>	<u>0.030</u>	<u>0.032</u>
$MSE_{ML}(\hat{\xi})$						
$MSE_{SM}(\hat{\xi})$	0.037	0.033	0.029	<u>0.027</u>	<u>0.030</u>	0.038

Minimal values are underlined

Table 2
BIAS $_{ML}(\hat{\xi})$ and BIAS $_{SM}(\hat{\xi})$ as functions of the true values ξ and n

ξ	-0.499	-0.40	-0.30	-0.20	-0.10	0
$n = 20$	-0.099	-0.084	-0.063	-0.051	<u>-0.033</u>	<u>-0.023</u>
$BIAS_{ML}(\hat{\xi})$						
$BIAS_{SM}(\hat{\xi})$	<u>0.056</u>	<u>0.025</u>	<u>0.000</u>	<u>-0.030</u>	-0.058	-0.096
$n = 50$	-0.036	-0.030	-0.024	-0.020	<u>-0.012</u>	<u>-0.009</u>
$BIAS_{ML}(\hat{\xi})$						
$BIAS_{SM}(\hat{\xi})$	<u>0.026</u>	<u>0.019</u>	<u>0.001</u>	<u>-0.013</u>	-0.026	-0.051
$n = 100$	-0.020	-0.018	-0.014	-0.011	<u>-0.009</u>	<u>-0.004</u>
$BIAS_{ML}(\hat{\xi})$						
$BIAS_{SM}(\hat{\xi})$	<u>0.018</u>	<u>0.006</u>	<u>-0.001</u>	<u>-0.007</u>	-0.016	-0.029
$n = 200$	-0.012	-0.010	-0.008	-0.006	<u>-0.003</u>	<u>-0.002</u>
$BIAS_{ML}(\hat{\xi})$						
$BIAS_{SM}(\hat{\xi})$	<u>0.010</u>	<u>0.005</u>	<u>0.000</u>	<u>-0.003</u>	-0.007	-0.017
$n = 500$	<u>-0.006</u>	-0.005	-0.004	-0.003	<u>-0.002</u>	<u>-0.001</u>
$BIAS_{ML}(\hat{\xi})$						
$BIAS_{SM}(\hat{\xi})$	<u>0.006</u>	<u>0.002</u>	<u>0.000</u>	<u>-0.001</u>	-0.003	-0.007

Minimal absolute values are underlined

Kolmogorov–Smirnov distance $n^{1/2} \cdot \max |F_n(x) - F(x)|$ just because its weighting function stresses the fitting at the ends of distribution in study, which is important for our problem. We put both ML and SM sample estimates of GEV parameters in (1) and used this function as the theoretical $F(x)$. Finally, we used ADD for each bootstrap sample under given n and averaged it over the bootstrap ensemble. Varying

n we chose value n_0 giving minimum averaged ADD. To estimate significance level of the minimum ADD one can not use standard statistical ADD-tables since they are applicable only for exactly known theoretical DF with no estimated parameters. We used a simulation procedure to avoid this obstacle, repeating our method of ADD estimation 10,000 times. Then we estimated the p value equal to probability of

Table 3
 $MSE_{ML}(\hat{\sigma})$ and $MSE_{SM}(\hat{\sigma})$ as functions of the true values ξ and n

ξ	-0.499	-0.40	-0.30	-0.20	-0.10	0
$n = 20$	0.219	0.206	0.196	0.197	0.197	0.206
$MSE_{ML}(\hat{\sigma})$						
$MSE_{SM}(\hat{\sigma})$	<u>0.189</u>	<u>0.176</u>	<u>0.169</u>	<u>0.109</u>	<u>0.173</u>	<u>0.196</u>
$n = 50$	0.118	0.114	0.113	0.113	0.116	<u>0.119</u>
$MSE_{ML}(\hat{\sigma})$						
$MSE_{SM}(\hat{\sigma})$	<u>0.117</u>	<u>0.112</u>	<u>0.109</u>	<u>0.108</u>	<u>0.112</u>	0.127
$n = 100$	<u>0.082</u>	<u>0.080</u>	0.078	0.078	<u>0.080</u>	<u>0.083</u>
$MSE_{ML}(\hat{\sigma})$						
$MSE_{SM}(\hat{\sigma})$	0.085	<u>0.080</u>	<u>0.077</u>	<u>0.077</u>	<u>0.080</u>	0.092
$n = 200$	<u>0.057</u>	<u>0.056</u>	<u>0.055</u>	0.055	<u>0.056</u>	<u>0.058</u>
$MSE_{ML}(\hat{\sigma})$						
$MSE_{SM}(\hat{\sigma})$	0.060	0.057	<u>0.055</u>	<u>0.054</u>	0.058	0.067
$n = 500$	<u>0.036</u>	<u>0.035</u>	<u>0.034</u>	<u>0.034</u>	<u>0.035</u>	<u>0.036</u>
$MSE_{ML}(\hat{\sigma})$						
$MSE_{SM}(\hat{\sigma})$	0.038	0.036	0.035	<u>0.034</u>	0.037	0.045

Minimal values are underlined

Table 4
 $BIAS_{ML}(\hat{\sigma})$ and $BIAS_{SM}(\hat{\sigma})$ as functions of the true values ξ and n

ξ	-0.499	-0.40	-0.30	-0.20	-0.10	0
$n = 20$	<u>-0.015</u>	<u>-0.021</u>	<u>-0.034</u>	-0.038	-0.046	-0.053
$BIAS_{ML}(\hat{\sigma})$						
$BIAS_{SM}(\hat{\sigma})$	-0.064	-0.056	-0.050	<u>-0.032</u>	<u>-0.010</u>	<u>0.028</u>
$n = 50$	-0.036	-0.030	-0.024	-0.020	<u>-0.012</u>	<u>-0.009</u>
$BIAS_{ML}(\hat{\sigma})$						
$BIAS_{SM}(\hat{\sigma})$	<u>0.026</u>	<u>0.019</u>	<u>0.001</u>	<u>-0.013</u>	-0.026	-0.051
$n = 100$	<u>-0.011</u>	-0.043	-0.078	-0.018	-0.019	<u>-0.021</u>
$BIAS_{ML}(\hat{\sigma})$						
$BIAS_{SM}(\hat{\sigma})$	-0.025	<u>-0.023</u>	<u>-0.022</u>	<u>-0.014</u>	<u>0.010</u>	0.026
$n = 200$	-0.012	-0.010	-0.008	-0.006	<u>-0.003</u>	<u>-0.002</u>
$BIAS_{ML}(\hat{\sigma})$						
$BIAS_{SM}(\hat{\sigma})$	<u>0.010</u>	<u>0.005</u>	<u>0.000</u>	<u>-0.003</u>	-0.007	-0.017
$n = 500$	<u>0.000</u>	<u>-0.002</u>	<u>-0.003</u>	-0.005	-0.006	-0.007
$BIAS_{ML}(\hat{\sigma})$						
$BIAS_{SM}(\hat{\sigma})$	-0.006	-0.005	-0.004	<u>-0.003</u>	<u>0.001</u>	<u>0.005</u>

Minimal absolute values are underlined

exceeding observed ADD under null hypothesis that GEV is valid. Usually a hypothesis is rejected if p value is less than 0.1. In our examples, p values were less than the table values but still more than 0.1 (see Table 7) which admissibly justifies using the GEV.

We have applied this procedure to the seismic moment GCMT catalog. We have studied the

seismicity in three virtual zones representing the intracontinental seismicity, subduction zones seismicity, and the mid-ocean ridge zones seismicity. The earthquakes that occurred in these three virtual zones are presented in Fig. 4. We have chosen for examination the areas of high seismicity level corresponding to these three tectonic situations. The virtual subduction zone includes 18,299 events, the

Table 5
MSE_{ML}($\hat{\mu}$) and MSE_{SM}($\hat{\mu}$) as functions of the true values ξ and n

ξ	-0.499	-0.40	-0.30	-0.20	-0.10	0
$n = 20$	0.259	0.263	0.267	0.265	0.267	0.268
MSE _{ML} ($\hat{\mu}$)						
MSE _{SM} ($\hat{\mu}$)	<u>0.240</u>	<u>0.242</u>	<u>0.246</u>	<u>0.246</u>	<u>0.250</u>	<u>0.257</u>
$n = 50$	0.156	0.158	0.158	0.160	0.162	0.164
MSE _{ML} ($\hat{\mu}$)						
MSE _{SM} ($\hat{\mu}$)	<u>0.155</u>	<u>0.155</u>	<u>0.154</u>	<u>0.155</u>	<u>0.159</u>	<u>0.163</u>
$n = 100$	<u>0.109</u>	<u>0.108</u>	0.111	0.112	0.114	0.114
MSE _{ML} ($\hat{\mu}$)						
MSE _{SM} ($\hat{\mu}$)	0.110	<u>0.108</u>	<u>0.109</u>	<u>0.110</u>	<u>0.113</u>	<u>0.113</u>
$n = 200$	<u>0.076</u>	<u>0.077</u>	<u>0.078</u>	0.079	0.079	<u>0.079</u>
MSE _{ML} ($\hat{\mu}$)						
MSE _{SM} ($\hat{\mu}$)	0.077	0.078	<u>0.078</u>	<u>0.078</u>	<u>0.078</u>	0.080
$n = 500$	<u>0.048</u>	<u>0.050</u>	<u>0.049</u>	<u>0.050</u>	<u>0.049</u>	<u>0.051</u>
MSE _{ML} ($\hat{\mu}$)						
MSE _{SM} ($\hat{\mu}$)	0.049	<u>0.050</u>	<u>0.049</u>	<u>0.050</u>	<u>0.049</u>	0.052

Minimal values are underlined

Table 6
BIAS_{ML}($\hat{\mu}$) and BIAS_{SM}($\hat{\mu}$) as functions of the true values ξ and n

ξ	-0.499	-0.40	-0.30	-0.20	-0.10	0
$n = 20$	0.063	0.056	0.049	0.048	<u>0.039</u>	<u>0.037</u>
BIAS _{ML} ($\hat{\mu}$)						
BIAS _{SM} ($\hat{\mu}$)	<u>-0.015</u>	<u>0.004</u>	<u>0.018</u>	<u>0.034</u>	0.046	0.065
$n = 50$	0.026	0.021	0.019	0.019	<u>0.014</u>	<u>0.012</u>
BIAS _{ML} ($\hat{\mu}$)						
BIAS _{SM} ($\hat{\mu}$)	<u>-0.005</u>	<u>-0.001</u>	<u>0.006</u>	<u>0.014</u>	0.020	0.029
$n = 100$	0.013	0.014	0.012	0.010	<u>0.010</u>	<u>0.006</u>
BIAS _{ML} ($\hat{\mu}$)						
BIAS _{SM} ($\hat{\mu}$)	<u>-0.007</u>	<u>0.001</u>	<u>0.004</u>	<u>0.008</u>	0.013	0.016
$n = 200$	0.008	0.007	0.004	0.006	<u>0.004</u>	<u>0.004</u>
BIAS _{ML} ($\hat{\mu}$)						
BIAS _{SM} ($\hat{\mu}$)	<u>0.004</u>	<u>-0.001</u>	<u>0.000</u>	<u>0.004</u>	0.006	0.010
$n = 500$	0.004	0.004	0.003	0.002	<u>0.002</u>	<u>0.001</u>
BIAS _{ML} ($\hat{\mu}$)						
BIAS _{SM} ($\hat{\mu}$)	<u>-0.003</u>	<u>-0.001</u>	<u>0.000</u>	<u>0.001</u>	0.003	0.004

Minimal values are underlined

virtual intracontinental zone includes 2116 earthquakes, and the virtual mid-ocean ridge zone includes 4168 events.

The estimates of GEV parameters for three virtual zones are collected in Table 7. Corresponding distribution tails $1 - F(x)$ are shown in Figs. 5, 6, and 7. The optimal sample sizes occurred as follows: $n = 66$ (subduction), $n = 60$ (continents), $n = 60$ (ridges).

We have used the SM-estimation. Standard deviations were estimated by the method of two subsamples, described above. By the way, these std estimates were close to estimates obtained from the bootstrap ensembles. p values were estimated by the cross-validation of two subsamples too. It should be noted that p values calculated directly from the bootstrap ensembles (which is not legitimate because

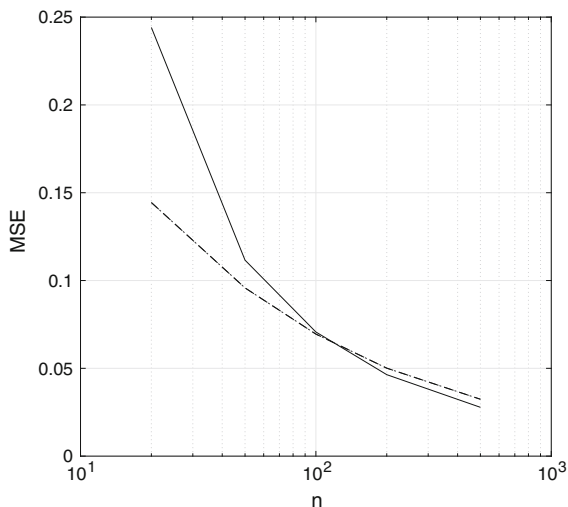


Figure 2

The averaged mean-square error MSE values for different sample sizes n ; the maximum likelihood (solid line) and the moment method (dashed line)

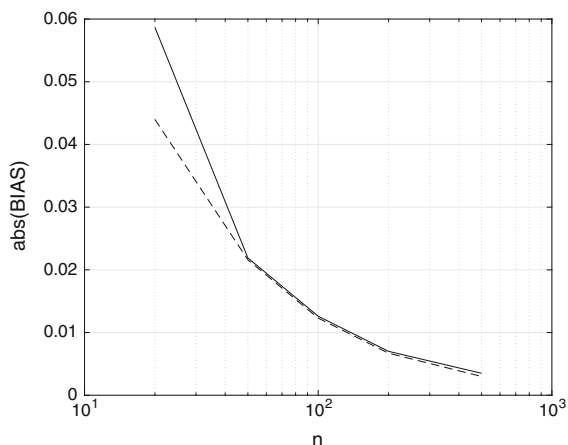


Figure 3

Averaged abs (BIAS) values for different sample sizes n ; the maximum likelihood (solid line) and the moment method (dashed line)

of inserted parameter estimates) in this case occurred much larger than the cross-validated p values. Thus, the cross-validation modification turned out to be quite valuable.

As it is seen from Table 7 the form parameter ξ is estimated with a large uncertainty in all three zones. The coefficient of variation C_v (ratio of sample std to absolute sample mean) varies from 0.29 up to 0.78 which testifies to a large uncertainty. The scale parameter σ is estimated more confidently. Its C_v ranges from 0.088 up

to 0.102. The location parameter μ is estimated the most certainly: $0.007 \leq C_v \leq 0.011$. The quantiles $Q_q(\tau)$ of maximum event in future τ years (see for details below) possess sufficient stability also. The sample quantile $Q_{0.5}(10)$ (median of maximum magnitude in future 10 years) has variation interval $0.014 \leq C_v \leq 0.017$ and $Q_{0.90}(10)$ has correspondingly $0.021 \leq C_v \leq 0.028$. For comparison, sample coefficient for statistic corresponding to maximum possible event, namely

$$\hat{M}_{\max} = \hat{\mu} - \frac{\hat{\sigma}}{c_{95}} \tag{14}$$

is extremely uncertain. Its C_v takes values much more than one. Thus, its use in practical purposes is highly questionable.

The statistical character of the SM estimates can be illustrated by 2D 68%-confidence domains for pairs of parameters (ξ, σ) and (σ, μ) . The confidence level 68% corresponds to \pm std intervals for Gaussian distribution. In our case, the quantile distribution is asymmetric, so that the 68% intervals are more adequate. These domains were derived as follows. First, we have found the non-parametric estimate of 2D-density of pair of estimates in question in form of

$$f(x, y) = \frac{1}{n} \sum_{k=1}^n K_{\sigma} \left(\frac{x - x_k}{s_x}, \frac{y - y_k}{s_y} \right), \tag{15}$$

where n is the size of artificial sample, generated by SM estimates of parameters in question (we took $n = 10,000$); K_{σ} is 2D symmetric kernel depending on a scale parameter σ ; $(x_k, y_k), k = 1, \dots, n$ is the sample; s_x, s_y are sample standard deviations of X, Y . We have used the Gaussian kernel

$$K_{\sigma}(x, y) = \frac{1}{2\pi\sigma^2} \exp \left(-\frac{x^2}{2\sigma^2} - \frac{y^2}{2\sigma^2} \right). \tag{16}$$

The scale parameter σ should be chosen as a compromise between too smoothed and too “sharp” estimates of the density. This choice depends on n and on structure of 2D sample points. We have taken $\sigma = 0.5$ which seems to meet both conditions. Now, let $D(h)$ denote domain on (x, y) -plane, where $f(x, y) > h$. We select h so that probability $P\{D\}$ of domain $D(h)$ would be equal to q , where q ($0 < q < 1$) is a given number:

Table 7
The GEV parameter estimates for three zones

	ξ	σ	μ	$Q_{0.50}(10)$	$Q_{0.90}(10)$	$p-v(ADD)$
Subduction $\Delta T = 221$ days $n_0 = 66$	-0.146 ± 0.068	0.509 ± 0.049	7.30 ± 0.08	8.58 ± 0.17	9.12 ± 0.30	0.34
Continents $\Delta T = 243$ days $n_0 = 60$	-0.212 ± 0.062	0.476 ± 0.048	6.49 ± 0.07	7.57 ± 0.13	7.96 ± 0.21	0.37
Ridges $\Delta T = 243$ days $n_0 = 60$	-0.102 ± 0.096	0.310 ± 0.037	6.41 ± 0.05	7.22 ± 0.12	7.60 ± 0.24	0.32

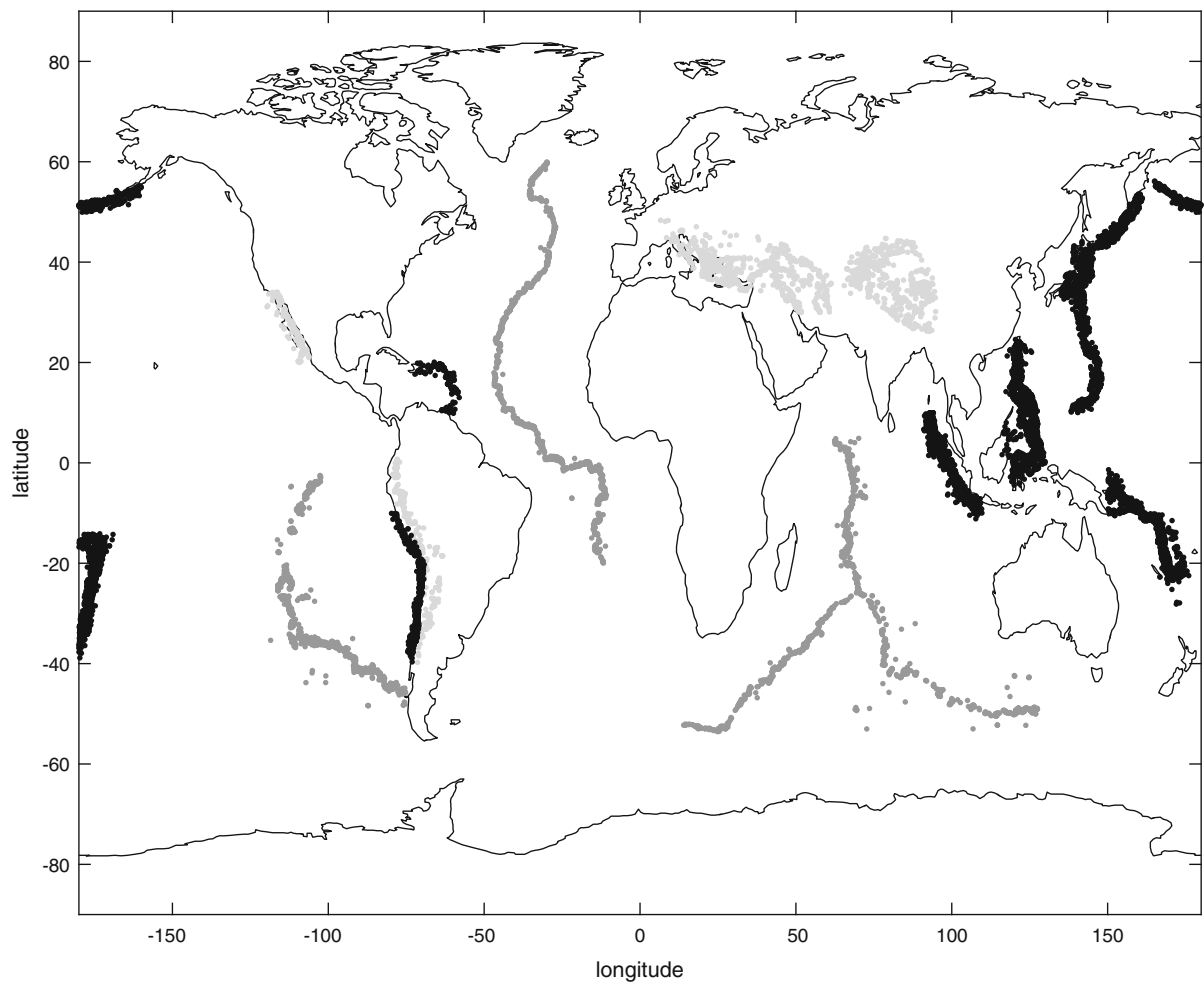


Figure 4

Earthquake epicenters of the virtual subduction zone (*black dots*), virtual intracontinental zone of seismicity (*grey dots*), and virtual mid-ocean ridge zone (*dark grey dots*)

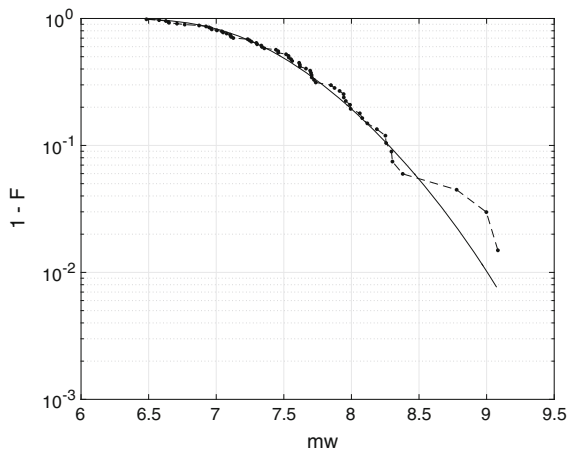


Figure 5

The complementary sample distribution function $1-F(x)$ vs. seismic moment m_w magnitudes for the virtual subduction zone

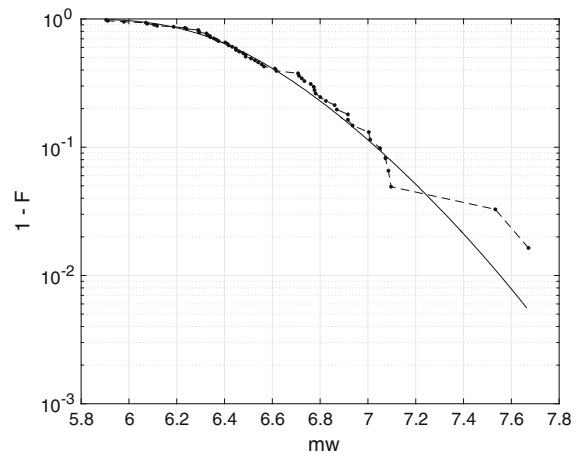


Figure 7

Distribution tail for the virtual mid-ocean ridge zone, the axis as in Fig. 5

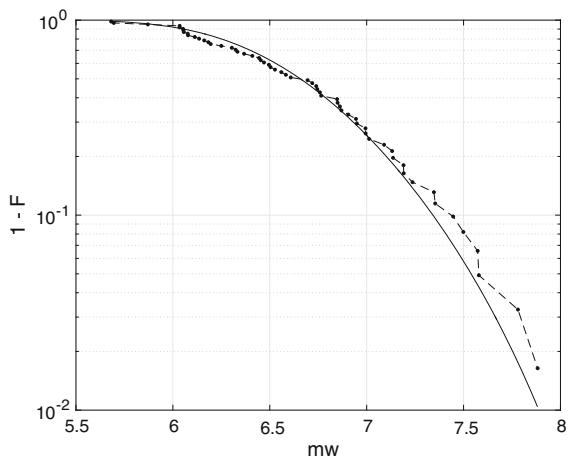


Figure 6

Distribution tail for the virtual intracontinental seismicity, the axis as in Fig. 5

$$P\{D\} = \iint_{D(h)} f(x, y) dx dy = q. \quad (17)$$

The 2D domain $D(h_q)$ is the q %-confidence domain for pair (X, Y) that is optimal (it consists of points of higher probability than other q %-domains). We have chosen $q = 68\%$ since in 1D case this confidence level corresponds to \pm std domain. The 68%-confidence domains for pairs (ζ, σ) and (σ, μ) are shown in Figs. 8 and 9.

We see the pair (ζ, σ) distinctly separates the Ridge Zone from two others, whereas the pair (σ, μ) separates definitely all three types of seismic zones.

One can say that parameters σ, μ have a resolving power for zone separation whereas parameter ζ has not.

The obtained estimates can be commented as follows. We cannot differentiate the cases of subduction zone seismicity, intracontinental seismicity, and mid-ocean ridge zone seismicity through the ζ parameter values, but these zones can be well differentiated by μ and σ parameter values (Fig. 9). It means that the magnitude distribution function in these zones is definitely different. Thus, basing on the available data (40 years catalog) we are not able to decide whether the ζ -parameter is universal or not,

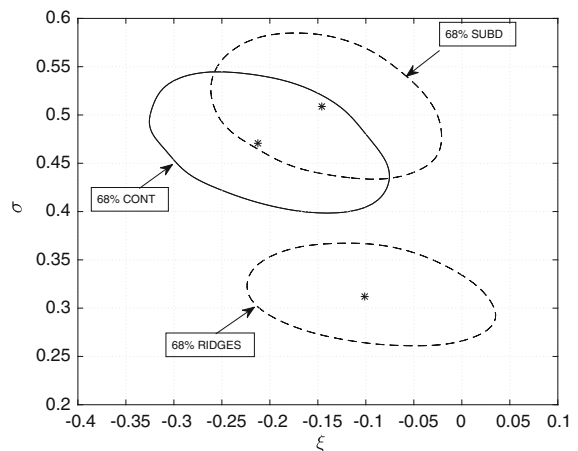


Figure 8

68%-confidence domains for pairs of parameters (ζ, σ) ; confidence domains are given by solid and dashed lines

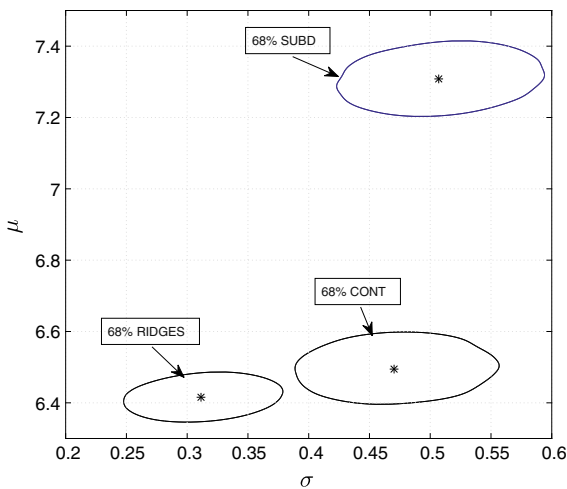


Figure 9
68%-confidence domains for pairs of parameters (σ, μ)

but longer catalogs perhaps will permit to decide this question.

For the full characterization of the seismic regime we have to add to the magnitude distribution discussed above one more parameter—the intensity of earthquake flow exceeding some threshold magnitude. This parameter appears to be rather robust. We have cleaned the catalog from aftershocks and swarms to make earthquake flow close to the Poisson point process. If the seismic region catalog covers time span T , then the intensity (the mean number of main events per time unit) can be estimated as $\hat{\lambda} = n/T$, where n is number of main events over some selected threshold h . Let us denote the true intensity as λ_0 . We have $E(\hat{\lambda}) = \lambda_0$. The variance of a Poisson random value n equals to its mean value:

$$\begin{aligned} \text{Var}(\hat{\lambda}) &= \text{Var}(n)/T^2 = E(n)/T^2 = \lambda_0 T/T^2 = \lambda_0/T; \\ \text{std}(\hat{\lambda}) &= \sqrt{\lambda_0}/T^{1/2}. \end{aligned} \tag{18}$$

For practical use for large n we can replace λ_0 for n/T :

$$E(\hat{\lambda}) \approx n/T; \quad \text{std}(\hat{\lambda}) \approx \sqrt{n}/T. \tag{19}$$

The ratio $\text{std}(\hat{\lambda})/E(\hat{\lambda}) \approx 1/n^{1/2}$ characterizes the variability of a random value. In our estimation of the intensity this ratio was usually less than 0.05–0.1 which can be interpreted as a rather high level of stability and robustness of estimate of intensity. In

our bootstrap experiments we have put sample size to be a Poissonian random number with parameter n (real sample size) and thus took into account the intensity variation.

6. The Quantiles $Q_q(\tau)$ as a Robust Estimate of the Seismic Hazard

The exposed above method can be used in the seismic hazard assessment. Usually for this aim, the maximum regional magnitude m_{\max} [see, e.g., (Kijko 2012)] is used. It is well known, however, that m_{\max} estimate is very unstable [see (Pisarenko and Rodkin 2010)]. In the GEV distribution, the maximum possible magnitude is given by expression $M_{\max} = \mu - \sigma/\xi$ (14) which explains the instability by small (or even very small) denominator ξ in most cases connected with seismic hazard assessment. In practical examples, one can obtain very improbable values M_{\max} like 20–100. Taking this fact into account we have recommended (Pisarenko and Rodkin 2007, 2010) to exploit instead of m_{\max} the quantiles $Q_q(\tau)$ level q of maximum event in future time interval τ . If one uses time blocks ΔT for measuring maxima then the theoretical $Q_q(\tau)$ has form:

$$Q_q(\tau) = \mu + \frac{\sigma}{\xi} \left[\left(\frac{\tau}{\Delta T \log(1/q)} \right)^\xi - 1 \right] \tag{20}$$

In our estimation we found an optimal number n_0 equal to $T/\Delta T$, where T was the catalog time span. Figures 10, 11 and 12 show the quantiles $Q_q(\tau)$ for our three zones with $q = 0.50$ (median) and $q = 0.90$ along with 68% confidence intervals. We see that for large future time intervals $\tau = 1000, 5000$ years the confidence intervals become very large. For the subduction zone they reach 1.2–1.5 magnitude unit (it should be noted also that the results for the ridge zone can be relatively less accurate because of lack of seismic stations near these zones). This fact testifies that it is hardly possible to speak definitely about maximum magnitudes occurring in such large future times. Quite satisfactory results can be obtained only for essentially less time intervals, e.g., if $\tau = 10$ years (short-term prediction) and $\tau = 50$ years (period of exploitation of usual

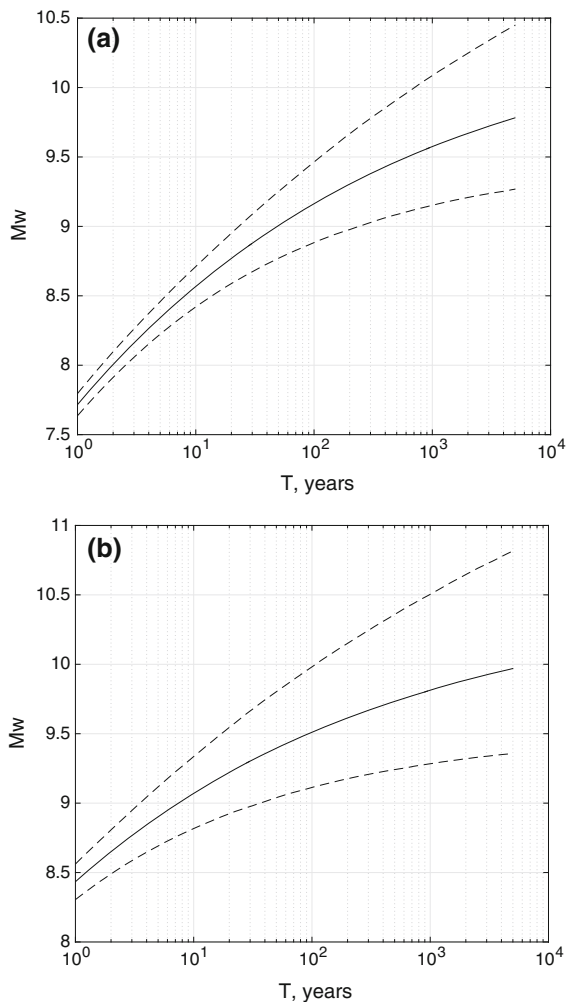


Figure 10

a Median and **b** 90%-quantiles of maximum m_w -value in future T years in the virtual subduction zone with 68% confidence levels

constructions). For $\tau = 10$, the 68% confidence intervals do not exceed 0.25 magnitude unit, and for case $\tau = 50$ years 68% confidence interval does not exceed 0.5 magnitude unit which is close to the standard error in the majority of hazard assessment schemes.

7. Summary

The most general theoretic approach to characterization of the tail distribution in range of rare extreme events is based on the Extreme Value theory

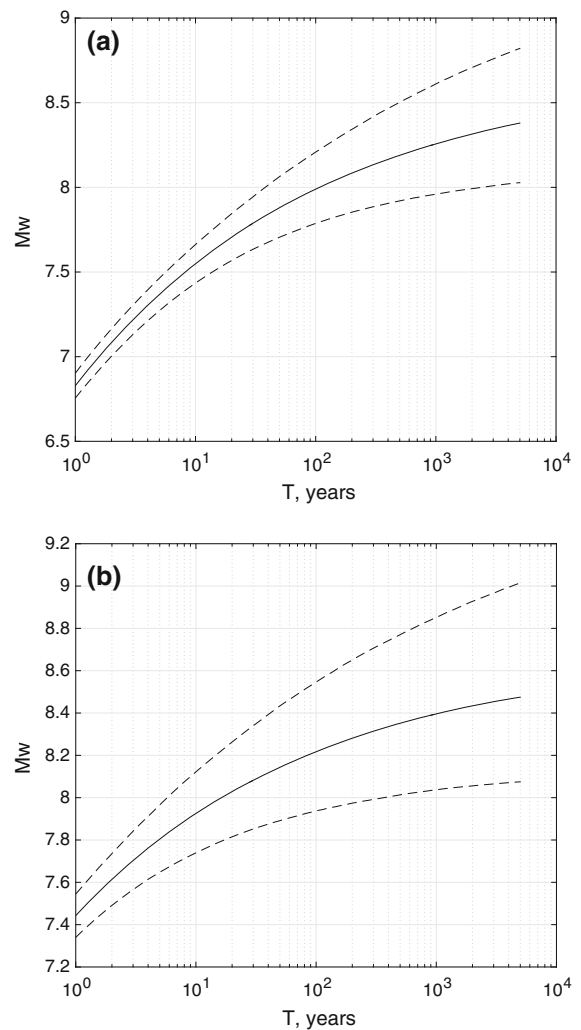


Figure 11

a Median and **b** 90%-quantiles of maximum m_w -value in future T years in the virtual intracontinental seismic zone with 68% confidence levels

(EVT). This theory has two main limit distributions: the GEV and the GPD. To use these distributions, one has to estimate statistically their parameters. But in many practical situations, one possesses limited experimental data for this aim, which makes the use of standard statistical tools like the method of maximum likelihood (ML) rather questionable and not reliable. We tried to study this rather typical situation of small sample size. The GPD was studied in our work (Pisarenko et al. 2017); the GEV has been considered in this paper. In the study by Pisarenko et al. (2017) we found for the GPD approach a good

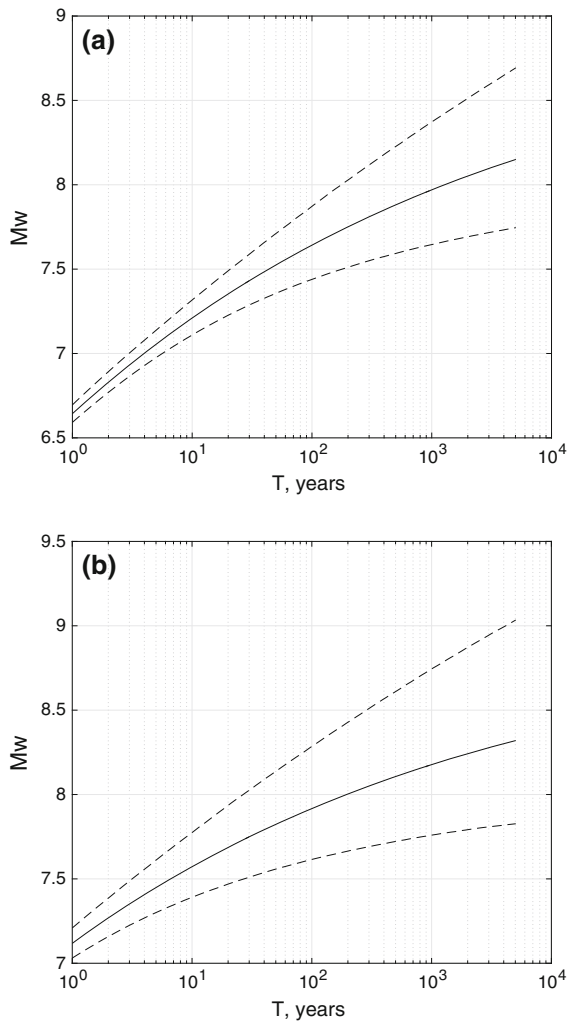


Figure 12

a Median and **b** 90%-quantiles of maximum m_w -value in future T years in the virtual MOR seismic zone with 68% confidence levels

competitor to the ML for small samples—the method of quantiles. Now we have established for the same situation that well-known method of statistical moments (SM) is preferable for the GEV. We tried the method of quantiles too, but for the GEV it turned out to be less efficient than the ML and SM methods. It was shown that for small samples ($n < 100$) the SM mostly overpasses other methods.

Using the discussed SM approach we have examined the Kagan (1999) conclusion on the universality of the seismic moment–frequency relation

and have found that we are able to discriminate three virtual seismic zones: the virtual subduction zone, the virtual intracontinental zone and virtual mid-oceanic ridge zone through μ and σ parameter values. It cannot be done, however, using the parameter ζ only. The change in μ and σ values of these zones correlates well with the change in earthquake magnitudes typical of such zones.

We have suggested the quantiles $Q_q(\tau)$ of maximum event in future time interval τ , as a stable statistical characteristic of the seismic regime and seismic danger. Its stability is explained by the fact that it is an integral characteristic of the distribution law in question as opposed to the unstable maximum possible magnitude M_{\max} .

Acknowledgements

The work was supported by the Russian Foundation for Basic Research, Grant Nos. 14-05-00866 and 14-05-00776.

REFERENCES

- Anderson, T. W., & Darling, D. A. (1954). A test of goodness-of-fit. *Journal of the American Statistical Association*, 49, 765–769.
- Buishand, T. (1989). Statistics of extremes in climatology. *Statistica Neerlandica*, 43, 1–30.
- Christopeit, N. (1994). Estimating parameters of an extreme value distribution by methods of moments. *Journal of Statistical Planning and Inference*, 41, 173–186.
- Efron, B. (1979). Bootstrap methods: another look at the Jackknife. *Annals Statistics*, 7(1), 1–26.
- Embrechts, P., Kluppelberg, C., & Mikosch, T. (1997). *Modelling extremal events* (p. 645). Berlin: Springer.
- Hosking, J. R., Wallis, J. R., & Wood, E. F. (1985). Estimation of the generalized extreme-value distribution by the method of probability-weighted moments. *Technometrics*, 27, 251–261.
- Kagan, Y. Y. (1999). Universality of the seismic moment–frequency relation. *Pure Applied Geophysics*, 155, 537–573.
- Kijko, A. (2012). On Bayesian procedure of maximum earthquake magnitude estimation. *Research in Geophysics*, 2(7), 46–54.
- Knopoff, L., Kagan, Y., & Knopoff, R. (1982). b-values for foreshocks and aftershocks in real and simulated earthquake sequences. *Bulletin of the Seismological Society of America*, 72, 1663–1675.
- Pisarenko, V., & Rodkin, M. (2007). Distributions with heavy tails: Application to the disaster analysis. In: *Computational Seismology*, iss.38 (p. 240). Moscow: GEOS. (in Russian).

- Pisarenko, V., & Rodkin, M. (2010). *Heavy-tailed distributions in disaster analysis. Advances in natural and technological hazards research, Volume 30*. Dordrecht: Springer.
- Pisarenko, V., & Rodkin, M. (2013). *Statistical analysis of natural disasters and related losses*. Dordrecht: Springer. (**Springer briefs in earth sciences**).
- Pisarenko, V., Rodkin, M., & Rukavishnikova T. (2017). The Estimation of probability of rare extreme events for the case of small samples, methods and examples of analysis of earthquake catalog. *Physics of the Earth*. (**in press**).
- Smith, R. (1990). Extreme value theory. In W. Ledermann (Ed.), *Handbook of Applicable Mathematics, Supplement* (pp. 437–472). Chichester: Wiley.
- Smith, R., & Naylor, J. (1987). A comparison of maximum likelihood and Bayesian estimators for the three-parameter Weibull distribution. *Applied Statistics*, 36, 358–369.

(Received December 25, 2016, revised February 5, 2017, accepted February 7, 2017, Published online February 24, 2017)