ATLANTIS PRESS

Advances in Intelligent Systems Research, volume 148

International Conference on Information Technology and Management Engineering (ICITME 2018)

Mining Extreme Patterns of Seismic Signals in China Based on Heavy-Tailed Statistics

Peng Zhang*, Chunbo Fan, Yunxia Zhang and Yi Ding

National Disaster Reduction Center of China, No. 6, Guangbai East Road, Chaoyang District, Beijing, China

*Corresponding author

***Abstract*—Strong earthquakes are the most violent natural hazards, which have caused huge losses of lives and property to human beings in the last decade. Mining time patterns of seismic events, especially extreme ones, within a specific time window and region, is of great importance to policy making and disaster reduction. In this paper, we apply heavy-tailed statistics in extreme value theory to mining the temporal pattern of seismic events in China and build a computational model. We show that earthquakes of magnitude above Ms. 6.0 in China follow the generalized Pareto distribution. We use the proposed model to simulate the possible risk of extreme seismic magnitude in future time windows. Comparison with real-world records demonstrate the validity of our model.**

Keywords—pattern mining; extreme value theory; risk analysis; heavy-tailed statistics

I. INTRODUCTION

Extreme value theory, as a promising tool for analyzing the probability of extreme events, has been widely applied to various regions, including economics, life science, and disaster assessment. Natural disasters, especial extreme ones, are typical extreme events, which can be well modeled by extreme value theory. In recent years, disaster risk analysis based on this theory has drown extensive research interests, like risk assessment of earthquakes [1], loss assessment of floods [2], risk analysis of man-made architectures [3], to name just a few.

Critical earthquake events are of great threats to human beings, which have caused huge casualties and economic losses in the past decades. Mining the spatial and temporal pattern of seismic events, to predict the possibilities of their occurrence, is of great value to policy making in disaster emergency response and disaster risk reduction. This issue has been a hot research topic in the geoscience society, which is also a multi-discipline topic in pattern mining and signal processing.

Current advances on this topic focus on the prediction of possible extreme of seismic magnitude within a specific spatial and temporal window [4][5]. These methods are based on observations that strong earthquake events follow the independent Poisson distribution as a random process, and they use the generalized extreme distribution model to characterize events' statistical patterns. They successfully demonstrate that magnitude signals of strong earthquake events do follow the pattern of extreme statistics, however, recent research also shows that such methods are not so stable and statistically consistent in long-time temporal window [6][7].

In this paper, we apply heavy-tailed method in extreme value theory, which is a more robust and consistent approach in estimating extreme events, to model the magnitude signals of strong earthquake events above Ms. 6.0 in China since 1949. We test and validate the stability of magnitude series as a Poisson process and show that they are statistically consistent with a heavy-tailed distribution, that is, a generalized Pareto distribution. We present a straight-forward method to predict the possible extreme magnitude in a future time interval, and we build a computing framework for fast estimation and efficient application. Simulations on ground-truth data demonstrate the validity of our approach.

The rest parts of this paper are organized as follows. A detailed description of the proposed method is given in Section Ⅱ. Descriptions of data and numerical simulations are presented in Sections Ⅲ and Ⅳ, respectively. Some concluding remarks are given in Section Ⅴ.

II. HEAVY-TAILED MODEL OF SEISMIC SIGNALS

A. *Basci Assumptions*

Given a times series of seismic magnitudes of earthquake events, to apply heavy-tailed model to characterize their statistical patterns, two basic assumptions should be made.

Assumption 1. Earthquake events, whose magnitudes are above some given threshold, follow a stable Poisson process, that is, their occurrence has probability distribution function as

$$P(N(t+\tau)-N(t)=k)=\frac{e^{-\lambda t}(\lambda\tau)^k}{k!} \tag{1}$$

where parameter λ of event intensity is a constant or a linear function of t.

Assumption 2. The distribution of seismic magnitudes follows the generalized Pareto distribution (GPD) and has the following probability function

$$f(x|k,\sigma,\theta)=\frac{1}{\sigma}\left(1+k\frac{x-\theta}{\sigma}\right)^{-1-\frac{1}{k}} \tag{2}$$

B. Method Description

Based on the two assumptions in last subsection, the proposed method can be summarized as the following four steps.

Step 1. Setting a magnitude threshold. Heavy-tailed statistics are used to model extreme events; hence we need first determine a threshold to identify strong earthquake events. In this paper, we set Ms. 6.0 as a basic line and select events whose magnitude are no less than that.

Step 2. Testing the stability of events. Stability is one of the basic assumptions, which should be tested before fitting. We first fit the empirical cumulative distribution of λ and then test its consistency with the theoretical distribution $\lambda(t) = t$, using the K-S test. If the p-value is above 0.1, then the stability can be confirmed.

Step 3. Testing the heavy-tails. Heavy-tailed distribution is the other assumption that needs to be satisfied. Similar to Step 2, we first fit the empirical cumulative distribution of seismic magnitudes. Then we compute maximum likelihood estimates of parameters for the theoretical GPD. Next, empirical and theoretical distributions are tested using the K-S test. If the p-value is above 0.1, then heavy-tails can be confirmed.

Step 4. Computing extremes and probabilities. Given the length τ of a future time interval and confidence level q, the possible extreme magnitude can be computed using the following formula

$$Q_q(\tau) = \theta + \frac{\sigma}{k}\left(\left(log\frac{1}{q}\right)^{-k}(\lambda\tau)^k - 1\right) \tag{3}$$

III. Description of Data

We build up a data set of seismic magnitudes for strong earthquakes in China since 1949. Based on the catalog of earthquakes in China since 1949, which is released by the China Earthquake Administration, we sort out strong and destructive events with two thresholds, that are, magnitude no less than Ms. 6.0 and depth larger or equal to 60 km (i.e., shallow earthquakes). Furthermore, we use the sorting method in [9] to eliminate pre- and post- shocks and to reserve only the main shocks. After that, we obtain a set of 160 strong earthquake events. Some selected cases are shown in Table 1 and the scatter plot of the whole data set is shown in Figure 1.

TABLE I. SELECTED CASES OF EARTHQUAKES ABOVE MS. 6.0.

Date	Location	Magnitude	Depth
1966-03-22	Hebei	6.7	9
1971-09-14	Yunnan	6.2	33
1985-04-18	Yunnan	6.2	5
1992-07-30	Xizang	6.4	13
2002-06-29	Qinghai	6.1	22
2003-07-21	Yunnan	6.3	10
2008-05-12	Sichuan	8.0	14
2010-04-14	Qinghai	7.3	14
2013-04-20	Sichuan	7.0	17
2014-08-03	Yunnan	6.5	12

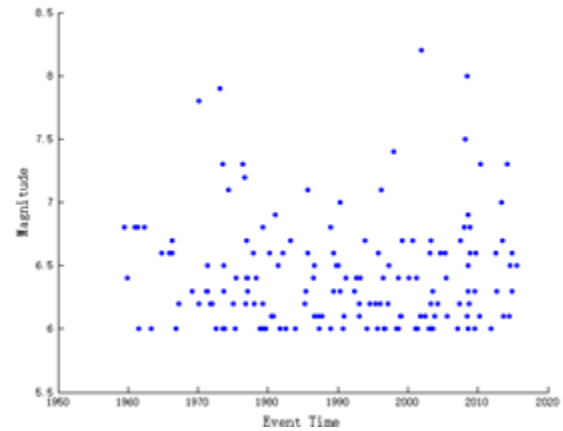

FIGURE I. SCATTER PLOT OF MAGNITUDE-VS-TIME DISTRIBUTION OF SHALLOW SHOCKS OVER MS. 6.0 IN CHINA SINCE 1949.

IV. Numercial Simulations

A. Estimation of Intensity and Test of Stability

The set of seismic magnitudes formulated in Section Ⅲ consists of a series of binary values (x_i, t_i), where x_i is the magnitude and t_i is the occurrence time. We apply the method proposed in Section Ⅱ-B to test the stability of earthquake events and estimate the intensity λ.

First, we normalize the occurrence time as

$$s_i = \frac{t_i - min_i t_i}{max_i t_i - min_i t_i} \tag{4}$$

and obtain the set of normalized time $\{s_i\}$.

Next, we fit the cumulative distribution function (CDF) of $\{s_i\}$ and use K-S test to validate the stability of event intensity λ, that is, we need to test whether the CDF of $\{s_i\}$ is consistent with the theoretical distribution $F(s) = s$. The test result shows a p-value of 0.3, which is larger than the confidence level 0.1 and confirms the stability of λ. The comparison of CDF and $F(s) = s$ is shown in Figure 2.

Finally, we estimate the intensity λ as

$$\hat{\lambda}_i = \frac{\text{Event\# up to } t_i}{t_i} \tag{6}$$

and obtain a series of intensities $\{\hat{\lambda}_i\}$, and we use $\hat{\lambda}_N$ as the estimate for the whole set. The plot of estimated intensity versus time is shown in Figure 3, where we can observe that intensity shows an obvious stable linear relationship to time.

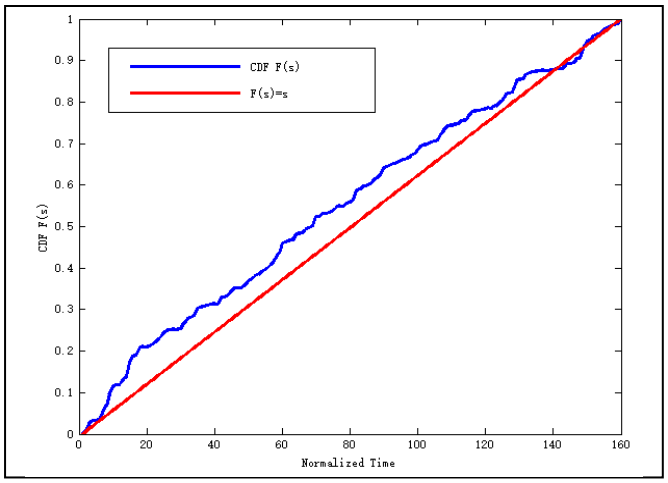

FIGURE II. FITTED EMPIRICAL DISTRIBUTION FUNCTION $F(S)$ AND THE STABLE DISTRIBUTION $F(S) = S$.

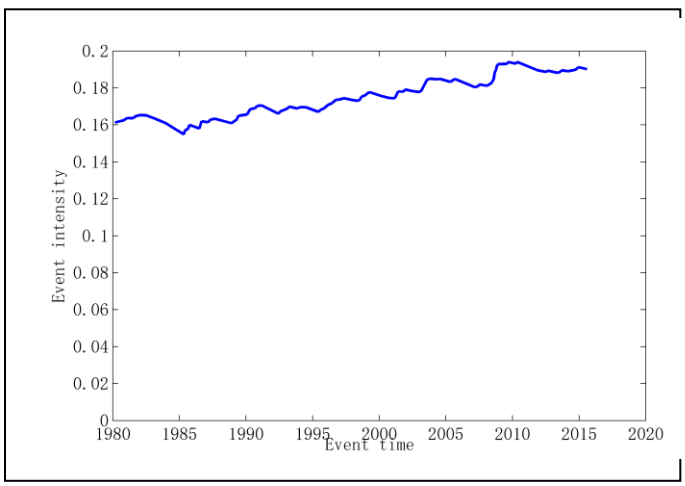

FIGURE III. PLOT OF ESTIMATED EVENT INTENSITY VERSUS TIME. SMOOTHED BY A 10-YEAR TIME WINDOW.

B. *Fitting the GPD and Estimation of Paramters*

We use the set (x_i, t_i) to fit the cumulative distribution function of events and the generalized Pareto distribution (GPD) function by using (2) and setting $\theta = 6.0$. These two fittings are demonstrated in Figure 4. We then use the K-S test to validate their consistency. The result shows a p-value of 0.97, which is significantly larger than the confidence level 0.1, showing that the magnitude data consistently follow the GPD with significant heavy-tail pattern. The parameters $\hat{k}$ and $\hat{\sigma}$ in GPD can also be obtained using maximum-likelihood during the fitting process. Estimated parameters are given in Table 2.

C. *Prediction for Future Strong Earthquakes*

We apply (3) to compute probabilities of extreme earthquake events in given time windows. All the estimated parameters and predefined values are given in Table 2 and simulation results are summarized in Table 3. Based on simulation results, we have the following observations.

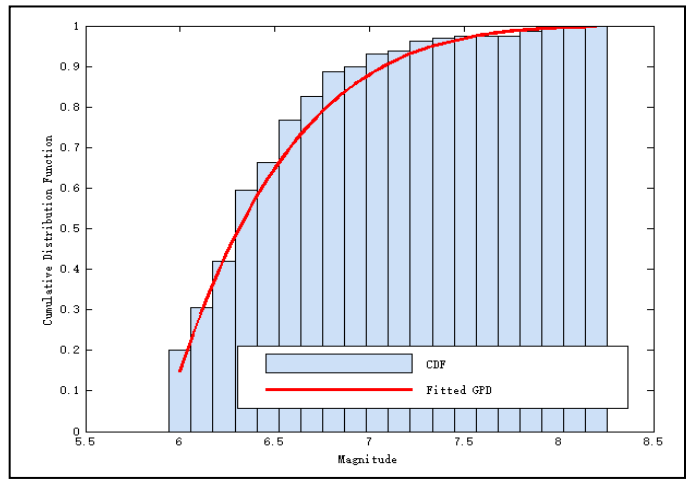

FIGURE IV. EMPIRICAL CUMULATIVE DISTRIBUTION FUNCTION (CDF) AND FITTED GENERALIZED PARETO DISTRIBUTION (GPD).

TABLE II. ESTIMATED PARAMETERS AND SETTINGS FOR COMPUTING MAXIMUM PROBABILITY OF CRITICAL EARTHQUAKES IN THE FUTURE.

Parameter	Value	Parameter	Value
$\hat{\lambda}$	2.8529	θ	6.0
$\hat{k}$	-0.1926	q	0.95
$\hat{\sigma}$	12	τ	1,2,3,4,5,10

TABLE III. ESTIMATED MAXIMUM MAGNITUDE OF CRITICAL EARTHQUAKES IN THE FUTURE

No.	Time Window	Maximum Magnitude	Confidence Level
1	1 Year	7.7	95%
2	2 Years	7.9	95%
3	3 Years	8.0	95%
4	4 Years	8.0	95%
5	5 Years	8.1	95%
6	10 Years	8.2	95%

- The statistical explanation for simulation results is that, the maximum magnitude of shallow earthquakes in China would not exceed the estimated value with confidence level 95%. Concretely, setting Jan. 1, 2016 as starting point, the maximum magnitude of shallow earthquake in China would not exceed 7.7 within 1 year, 7.9 within 2 years, 8.1 within 5 years, and 8.2 within 10 years, respectively.
- The highest magnitude of shallow earthquake in China is 6.7 within 1-year window since Jan. 1, 2016 (Ms. 6.7 quake in Xinjiang on 2016-11-25) and 7.0 within 2-year window (Ms. 7.0 quake in Sichuan on 2017-8-8).

The ground truth data are consistent with our predictions, which validate the effectiveness of the proposed model.

V. CONCLUSION

In this paper, we construct a set of strong earthquakes in China since 1949 and apply heavy-tailed statistical model to character the extreme patterns of this set. We analyze the stability of quake events and demonstrate that strong shallow earthquakes in China do follow the generalized Pareto distribution and show an obvious heavy-tail pattern. We further estimate the parameters of the GPD and predict possible extreme magnitude within given time windows in the future. The proposed method can provide valuable information for quake-proof designs and disaster reduction plans and can also be applied to statistical analysis of other extreme climate events, such as severe floods and typhoons/hurricanes.

ACKNOWLEDGMENT

This work was supported by the National Key R&D Program of China under Grant Nos. 2016YFC0803000 and 2016YFC0803005 and the National Natural Science Foundation (NNSF) of China under Grant No. 41201552.

REFERENCES

[1] X. Qian, F. Wang, G. Cao, and Q. Ren, “Application of the generalized extreme value distribution in seismic hazard analyasis,” Journal of Seismological Research, vol. 25, pp. 73-78, 2012.

[2] J. Liu, M. Wu, W. Jiang, and W. Zhan, “Flood disaster losses analysis based on the Poisson-lognormal compound extreme model,” Journal of Natural Disasters, vol. 19, pp. 61-66, 2010.

[3] L. Chen, Y. Wang, and J. Liu, “Application of extremal distribution to natural risk prediction of construction project,”, Journal of Natural Disasters, vol. 18, pp. 169-173, 2009.

[4] S. Chen, Y. Zhang, Y. You, and S. Huang, “An application and study of the extreme value theory in costal areas of Fujian earthquake prediction,” vol. 31, pp. 29-34, 2013.

[5] Y. Zhang, S. Chen, and K. Luo, “Extreme value theory application research about earthquake tendency prediction in Taiwan,”, South China Journal of Seismology, vol. 33, pp. 55-64, 2013.

[6] V.F. Pisarenko, M.V. Rodvin, “Statistical analysis of natural disasters and related losses,” SpringerBriefs in Earth Sciences, Springer, Dordrecht-Heidelberg-London-New York, 2014.

[7] V.F. Pisarenko, M.V. Rodvin, “Heavy-tailed distributions in disaster analysis,” Springer, Dordrecht-Heidelberg-London-New York, 2010.

[8] V.F. Pisarenko, M.V. Rodvin, “The new quantile approach: application to the seismic risk assessment,” In: Rascobic B, Mrdja S (eds) Natural disasters: prevention, risk factors and management. NOVA Publisher, New York, 141-174.

[9] L. Knopoff, Y. Kagan, R. Knopoff, “b-Values for foreshocks and aftershocks in real and simulated earthquake sequence,” Bulletin of the Seismological Society of America, vol. 72, pp. 1663–1675, 1982.