Alternative heavy tailed models in seismology

Miguel Felgueiras[1,2,3], **João Martins**[1,2], **Rui Santos**[1,2]

[1] School of Technology and Management, Polytechnic Institute of Leiria
[2] CEAUL Centre of Statistics and its Applications
[3] CARME, Polytechnic Institute of Leiria

E-mail: mfelg@ipleiria.pt

Abstract. Great earthquakes are commonly considered as the ones with moment magnitude (M_W) above or equal to 8.0. Since these earthquakes can destroy entire communities located near the epicentre, the search of physical laws that explain the energy released by them is an important issue. There is a connection between the radiated energy of an earthquake, its magnitude and its seismic moment (M_0). Thence, when fitting a heavy or an extremely heavy tailed distribution to a seismic moment dataset, we are in fact adjusting a mathematical model which explains the amount of energy released by these great seisms. Therefore, the main goal of this work is to study the more appropriated Pareto based models (the most used family in this field) when explaining the seismic moment of the great earthquakes. With this purpose in mind, we selected two different catalogs that accommodate recent events and are considered more accurate than other catalogs used in previous works. We conclude that the traditional Pareto distribution remains a good choice to deal with this kind of data, but Log-Pareto lead to higher p-values and Location-scale Pareto is better fitted to the biggest events.

1. Introduction

The great earthquakes are a rare phenomenon (only one occurrence in each five to ten years, in average) but can cause heavy human and material losses. As these seisms are still unpredictable, measuring their magnitude, which is directly connected to their released energy, is a relevant issue. Nowadays, the Moment Magnitude Scale, introduced by [11], is the most applied magnitude scale, although some variants may be considered. Previous to this measure, many others magnitude scales were adopted, such as the Local Magnitude Scale, the Surface Wave Magnitude Scale or the Body Wave Magnitude Scale [13]. These previously used scales were not appropriated to analyse the energy released by the great seisms, since they saturate above some threshold, and therefore seisms with magnitude above 8.8 were never recorded with those scales. This is usually denoted as the corner frequency problem. Thus, when studying the magnitude of great seisms and due to this lack of information, we must restrict our analysis to the XX and XXI centuries. This is the time period in which M_W information can be estimated with some accuracy [7]. Moreover, global data must be considered, since some regions have some specific characteristics, as the thickness of the seismogenic crust, which may imply non self-similarity between small and large events [14, 23] and a small maximum of possible energy released by the seisms in those regions. For example, for the Californian region, [1, 15] claim that a possible maximum lies somewhere between 8.1 and 8.3, and earthquakes with moment magnitude above 7.9 were never registered.

2. Connecting seismic moment, moment magnitude scale and radiated energy

Gutenberg and Richter [10] developed a relation between the radiated energy in ergs (E) and the surface-wave magnitude (M_S),

$$log_{10}E = 1.5M_S + 11.8, \tag{1}$$

and [22] connected the seismic moment with the radiated energy,

$$E = \frac{\Delta\sigma}{2\mu}M_0 \approx \frac{10^{-4}}{2}M_0 \tag{2}$$

where $\Delta\sigma$ is the average stress drop in the earthquake, μ is the rigidity of the elastic medium surrounding the fault and M_0 is the seismic moment in the dyne-centimetre scale. Later, [8] used the above relations, replacing (2) in (1) and M_S by M_W. This leads to

$$log_{10}M_0 = 1.5M_W + 16.1 \iff M_W = \frac{log_{10}M_0 - 16.1}{1.5}. \tag{3}$$

Thus, the previous authors created the M_W scale, which is related with M_0 and is based on a physical source model. When M_0 is measured in the Newton-meter (Nm) scale, $c = 16.1$ is replaced by $c = 9.1$ in the above equation, leading to

$$M_W = \frac{log_{10}M_0 - 9.1}{1.5}. \tag{4}$$

A more accurate M_0 calculation is based on the area of fault rupture, the average value of the final slip, and the rigidity modulus of the rocks and other material surrounding the fault (cf. [16]).

3. Pareto based distributions in the seismic moment fitting

Seismic moment can be converted into a power law [27] implying that the Pareto distribution is often chosen to model the energy released by earthquakes. For a state of the art under a geological perspective, and physical reasons to support this choice, see [1, 16, 20, 24] among others. However, there is usually some lack of fit in the higher magnitude earthquakes, which mainly arises from the non-similarity between great earthquakes and the others. In [28] the authors note that for some seismic regions the Pareto distribution underestimates the frequency of the very large earthquakes. Hence, in this work we are interested in analysing the energy released by the great earthquakes, commonly defined as the ones that usually result in total destruction, with M_W above 8.0 [18, 23].

3.1. The Pareto distribution

The Pareto distribution function with shape parameter $\alpha > 0$ is defined as

$$F_X(x) = 1 - x^{-\alpha}, \quad x \geq 1.$$

Since only earthquakes with some magnitude are detected (and even if detected, small earthquakes are irrelevant concerning their radiated energy), usually some truncation point t is considered [3, 12]. The truncated Pareto has distribution function given by

$$F_{X|X\geq t}(x) = 1 - \left(\frac{x}{t}\right)^{-\alpha}, \quad x \geq t \geq 1. \tag{5}$$

Pareto distribution is self-similar, that is,

$$\frac{X|X \geq t}{t} \sim Pareto(\alpha). \tag{6}$$

Self-similarity is very interesting in a mathematical point of view, but might be unrealistic when taking into account the specific characteristics of each region. Pareto is a scale free distribution, thus the scale parameter is the same whatever scale we look at in. For $t > 1$,

$$\overline{F}_X(tx) = t^{-\alpha}\overline{F}_X(x),$$

where $\overline{F}$ denotes the survival function. Previous works (such as [7, 21, 28]) suggested that in this context $\widehat{\alpha} \approx 0.67$ is an α estimate that is justified from a mathematical point of view and also from a scale convenience point of view, since with this alpha estimative $M_W \approx M_S$ for a moment magnitude $6.7 \leq M_W \leq 8.1$, where the major (but not the great) earthquakes occur.

3.2. The Log-Pareto distribution

Log-Pareto distribution (and also the next introduced distributions) is more heavy tailed than the Pareto distribution, and therefore might be more suitable for modelling the energy released by the great earthquakes, mainly outside the moment magnitude range $6.7 \leq M_W \leq 8.1$ where $M_W \approx M_S$ and $\widehat{\alpha} \approx 0.67$ when considering Pareto fitting. Preliminary work on this distribution properties and applications can be consulted in [16]. This distribution is merely a Pareto distribution transformation, because if $X \sim Pareto(\alpha)$ then $Y = e^X \sim Log - Pareto(\alpha)$. Thus, the α parameter has an interpretation similar to the one previously shown for the Pareto distribution, but now for the data logarithm. The Log-Pareto distribution function is

$$F_X(x) = 1 - (\ln x)^{-\alpha}, \quad x \geq e.$$

When a truncation point is set, the truncated Log-Pareto has distribution function

$$F_{X|X\geq t}(x) = 1 - \left(\frac{\ln x}{\ln t}\right)^{-\alpha}, \quad x \geq t \geq e, \tag{7}$$

and can also be transformed into a non-truncated Pareto,

$$\frac{\ln X | X \geq t}{\ln t} \sim Pareto(\alpha). \tag{8}$$

3.3. The extended slash family

The extended slash family is obtained considering the random variable

$$Y = \frac{X}{\Theta}$$

where $\Theta \sim Beta(\alpha, 1)$, $\alpha > 0$. The Y variable can also be seen as a Pareto scale mixture of the X variable [8], since $Y = \Theta^{-1}X$ where $\Theta^{-1} \sim Pareto(\alpha)$. With $X \sim Pareto(\alpha)$, [7] obtained a random variable whose distribution is designated as Extended Slash Pareto (ESP), and studied its properties and applications to the seismic moment fitting problem. The distribution function is

$$F_Y(x) = 1 - (1 + \alpha \ln x)\, x^{-\alpha}, \quad x \geq 1$$

and considering a truncation point t,

$$F_{Y|Y\geq t}(x) = 1 - \left(\frac{1 + \alpha \ln x}{1 + \alpha \ln t}\right)\left(\frac{x}{t}\right)^{-\alpha}, \quad x \geq t \geq 1. \tag{9}$$

Although it has a more complex distribution function expression, maximum likelihood estimation is straightforward for this distribution since an explicit estimator can be obtained.

Moreover, it is also relevant to note that the maximum likelihood estimator for the slash distribution family (this estimator is, in fact, a M-estimator for slash distribution family) has the highest breakdown point possible for regression equivariant estimators [17], thus leading to robust estimation. The α parameter has, under this distribution, a slightly different interpretation compared with the one previously presented for the Pareto distribution. However, and even considering this constraint, several authors already discussed the possibility of the Pareto relation might be too simple to adequately explain the seismic moment (see [27] for further references on this subject).

3.4. The Location-scale Pareto mixture

In [6] a Location-scale Pareto mixture (LSPM) was introduced. The LSPM was defined as $W = \mu + \sigma\Theta X$, where μ and σ are location and scale parameters, $\Theta \sim Pareto(\alpha)$ and $X \sim Pareto(1)$. Note that μ and σ define a truncation point, since the variable support is $S_W = [\mu + \sigma, \infty[$. This random variable has distribution function given by

$$F_W(x) = \frac{\alpha\left(\frac{x-\mu}{\sigma}\right)^{-2-\alpha}\left(x - \mu - \sigma\left(\frac{x-\mu}{\sigma}\right)^{\alpha}\right)}{\sigma(1-\alpha)}, \quad x \geq \mu + \sigma. \tag{10}$$

Maximum likelihood estimation is complex under this distribution, since it relies on iterative methods. However, in [6] the performance of the moment's method estimators was investigated using simulation, and the authors concluded that this method works reasonable well in most situations. Hence, the moment's method estimator seems to be an adequate estimation choice.

4. Catalogs and Data

We considered two different catalogs when adjusting the presented models. The first one is freely available in `http://earthquake.usgs.gov/earthquakes/search/`, and is offered by the United States Geological Survey (USGS catalog). The data is continuously updated and we collected information about all the events with $M_W \geq 7.95$ that occurred between 09-08-1901 (first detected event in the last century) and 26-05-2019 (last detected event until now) for a total of 95 seisms.

The second one can be requested by email in `http://www.isc.ac.uk/iscgem/request_catalogue.php` and is being developed by ISC-GEM Global Instrumental Earthquake Catalogue [4, 5] (ISC catalog). It contains data between 04-04-1904 and 31-12-2015 for a total of 105 seisms with $M_W \geq 7.95$. This is the last released version of the catalog (version 6.0 - released on 2019-03-07). Note that the version 1.05 used in [9] only had information about 86 seisms. Besides that, some of the data was updated, and therefore the data is now quiet different from the one analysed in [9].

Aside the number of events, the catalogs have other differences. In USGS most of the seisms magnitudes are registered with only one decimal place, implying that the data set have many repeated measures, while in ISC catalog the majority of the seisms magnitudes are indicated with two decimals places, leading to a data set with less repeated measures. Another relevant difference concerns the seisms magnitudes between catalogs. For instance, with $M_W \geq 9.0$ USGS catalog indicate five events with $M_W = (9; 9.1; 9.1; 9.2; 9.5)$ while ISC catalog only indicate four events with $M_W = (9.09; 9.3; 9.31; 9.6)$. In terms of the released energy, the increase in M_W from 9.5 to 9.6 is equivalent to a seism with $M_W = 9.24$. Therefore, and since both catalogs come from reputable sources, it is advisable to analyse them both and compare their results.

5. Fitting the Seismic moment of great earthquakes

For both catalogs we adjusted the Pareto based models with $M_W \geq 7.95$, $M_W \geq 8.05$ and $M_W \geq 8.15$. These truncation points, that roughly correspond to $M_0 \geq 10^{21.02}$, $M_0 \geq 10^{21.17}$

and $M_0 \geq 10^{21.32}$, were selected in order to work solely with great seisms, near the threshold where [23] estimates that distribution fitting still is an open issue, since the Pareto distribution family seems too light to fit the energy released by theses great seisms [7, 9, 26, 28].

The estimates for the parameters of the analysed models are presented in Table 1, together with the p-values of the Kolmogorov-Smirnoff (KS) and the Anderson-Darling (AD) goodness of fit tests. These tests were the selected ones since KS test is the most used goodness of fit test, and the AD test gives more weight to the distribution tails, preferable when the main interest is to fit extreme events (cf. [25]). Common model selection criterion, like the Akaike Information Criterion (AIC) or the Bayesian Information Criterion (BIC), cannot be used for the adjusted models since those models depend on different data set transformations, and therefore are incomparable with the usual information criteria [2]. Thus, parsimonious issues aren't numerically tackled in this work and therefore will depend on each scientist opinion.

Table 1. Summary results concerning the adjusted models.

	USGS Catalog			**ISC-GEM Catalog**		
M_w	7.95	8.05	8.15	7.95	8.05	8.15
M_0	$10^{21.025}$	$10^{21.175}$	$10^{21.325}$	$10^{21.025}$	$10^{21.175}$	$10^{21.325}$
Truncation M_0	$10^{21.02}$	$10^{21.17}$	$10^{21.32}$	$10^{21.02}$	$10^{21.17}$	$10^{21.32}$
n	95	70	45	105	74	48
Pareto α estimative	**0.9853**	**1.0333**	**0.9448**	**1.0302**	**1.0233**	**0.9403**
AD *p-value*	0.4463	0.3377	0.8562	0.7300	0.4248	0.9120
KS *p-value*	0.2156	0.1537	0.5826	0.6997	0.3623	0.8290
Log-Pareto α estimative	**48.6774**	**51.4309**	**47.3815**	**50.9288**	**50.966**	**47.193**
AD *p-value*	0.3934	0.3522	0.8761	0.7957	0.4344	0.9237
KS *p-value*	0.1568	0.1524	0.5876	0.8112	0.4756	0.9485
LSPM α estimative	**1.2169**	**1.3055**	**1.7113**	**1.2151**	**1.4983**	**1.2338**
AD *p-value*	0.2886	0.3145	0.8266	0.7665	0.3948	0.9210
KS *p-value*	0.1539	0.2207	0.4275	0.2840	0.4724	0.9450
ESP α estimative	**1.0051**	**1.0530**	**0.9643**	**1.0500**	**1.0431**	**0.9598**
AD *p-value*	0.4463	0.3377	0.8562	0.7300	0.4248	0.9120
KS *p-value*	0.2155	0.1537	0.5826	0.6997	0.3623	0.8290

The first striking conclusion is that none of the combinations between catalog, truncation point, model and goodness of fit test is rejected, with all $p - value \geq 0.15$. For this reason, all of the adjustments can be considered as valid. Also, for the same combination of truncation point, model and goodness of fit test, the results are always better when using the ISC catalog. This is probably due to the fact that in this catalog magnitudes are usually displayed with two decimal places, leading to smaller "jumps" in the distribution function that is adjusted to the data, and therefore to a better fit.

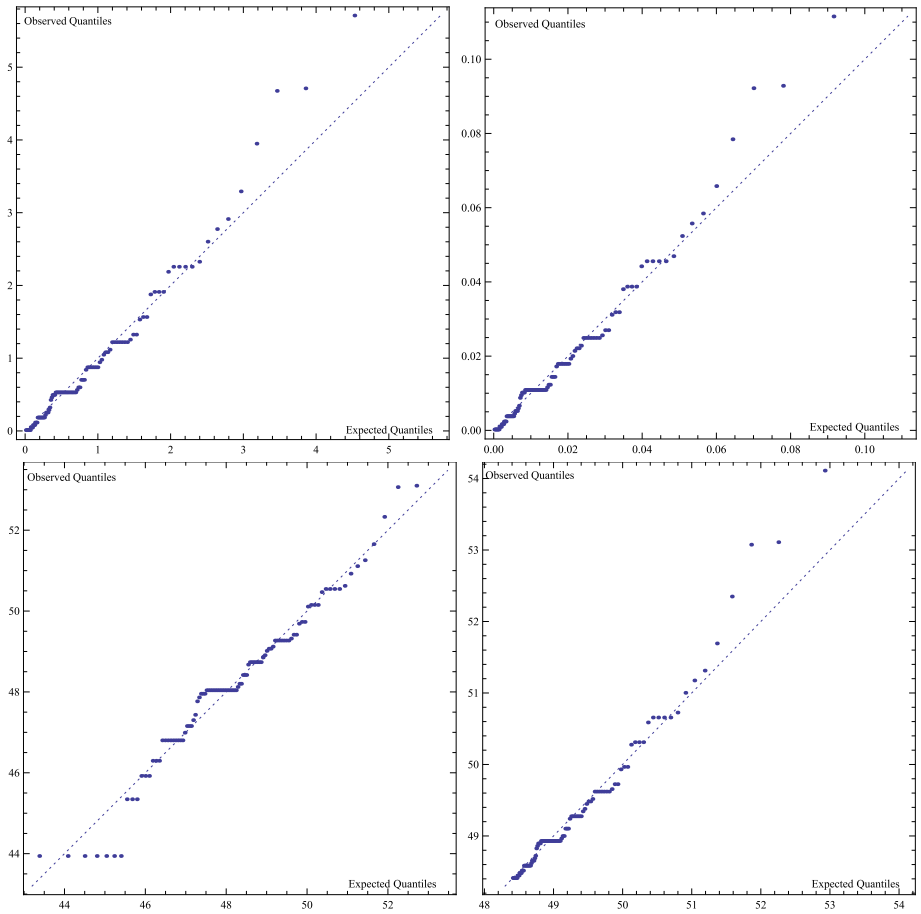
Another interesting conclusion is that α estimates are now substantially higher than the ones obtained in [7, 28] and that were recommended in works like [13, 21] due to important relations between magnitude scales. Aside from Log-Pareto distribution that does not have finite moments, all the other distributions have finite mean when $\alpha > 1$, which happens in the majority of the fitted models. Even considering, for instance, the 95% confidence interval for α in the Pareto fit for the ISC catalog with $M_W \geq 7.95$ (to work with a truncation point similar to the used in the above mentioned works), we obtain $\alpha \in [0.8088; 1.1868]$. Thus, the traditional assumption that $\alpha = 2/3$ in the Pareto fit is now rejected since $\alpha = 2/3$ is outside the confidence

interval. Moreover, in a first glance, higher values of α contradict the idea that the Pareto tails are to light for this kind of data.

Finally, Log-Pareto adjustments always led to slightly higher p-values (and therefore better fits) when using the ISC catalog, while when using the USGS catalog the situation varies according with the truncation point and the adjustment of fit test.

Another traditional way to assess the quality of the fit is to draw qq-plots. In the x-axis we represent the expected quantiles when using the introduced models, and in the y-axis the data quantiles. The points should scatter evenly in the neighbourhood of the line $y = x$, where theoretical quantiles are equal to observed quantiles. Points above the line show a quantile higher than the expected, while points below the line show the reverse situation. The expected quantiles were computed using equations (6), (8), (9) and (10), and the same transformations were applied to the data set. Since, according with the selected models, the graphs are all similar, we only present two sets of qq-plots. The first one is for the ISC catalog with $M_W \geq 7.95$, considering all the fitted models (Figure 1) and the second one for both catalogs with $M_W \geq 8.15$, considering Log-Pareto and LSPM models (Figure 2).

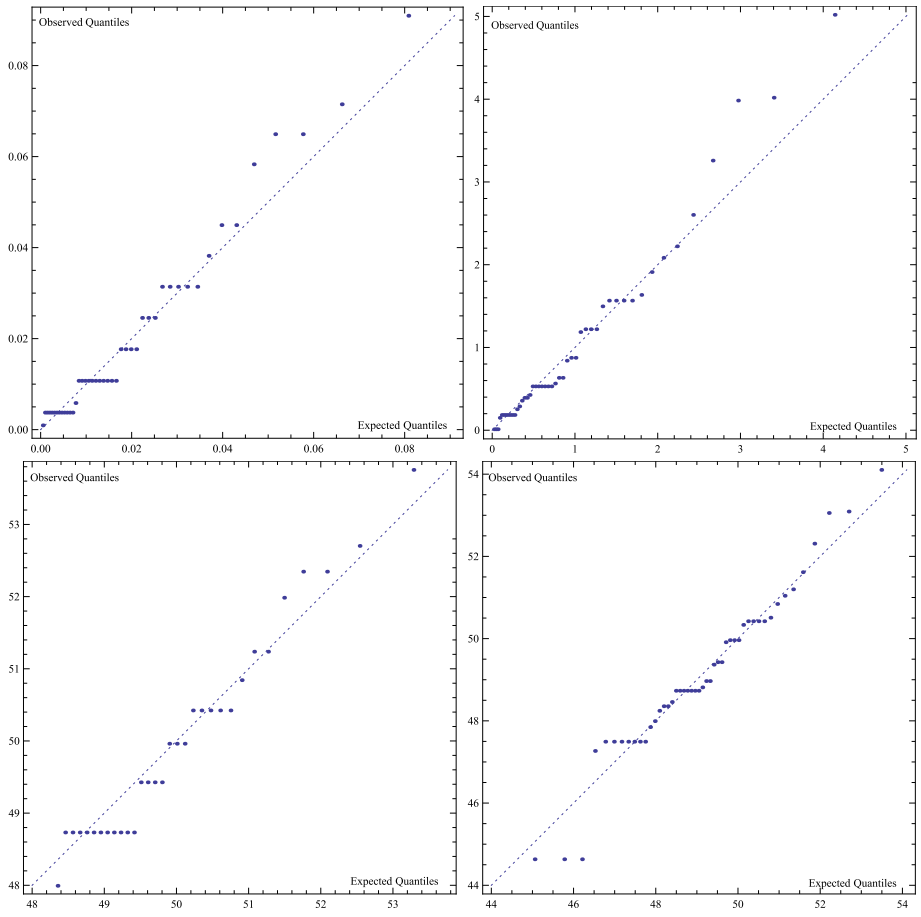
Figure 1. QQ-plots for the ISC catalog with $M_W \geq 7.95$ for the Pareto (top left), Log-Pareto (top right), LSPM (bottom left) and ESP (bottom right) models.



When analysing Figure 1, the adjustments seem good in a graphical evaluation since the points are randomly spread in the neighbourhood of the diagonal line for all the four graphs. However, and except for LSPM graph, there is a notorious lack of fit in the higher events since

all the eight greater seisms have observed quantiles above the expected ones. This was already observed in previous works like [7, 26, 28] and here it is more clear because the adjusted Pareto based models have a higher α parameter, implying lighter tales. The LSPM fit is the only one that, despite a worse fit in the lower events, captures better the energy released by the biggest seisms.

Figure 2. QQ-plots for the USGS catalog (left) and ISC catalog (right) with $M_W \geq 8.15$, using Log-Pareto (top) and LSPM (bottom) models.



Analysing Figure 2, it is now clear that the lower p-values obtained when using USGS catalog came from the excessive number of events with the same magnitude. The Log-Pareto fit seems better in the left graph, mainly because the biggest earthquake is recorded with $M_W = 9.5$ instead of $M_W = 9.6$ as in the right graph. Again, LSPM model reveals a more adequate fit for the biggest earthquakes independently from the used catalog.

6. Conclusions and future work

In this work several Pareto based models were applied in order to fit the energy released by the great earthquakes. This type of models is relevant since there are both physical and statistical reasons to support these models. Besides, the estimation of the radiated energy by great seisms remains an open issue, and updated data sets, as the analysed here, contribute to acquire a clearer picture of the mathematical shape of the phenomenon. Although the achieved results

were similar for the four considered models, Log-Pareto and LSPM models were the best ones. Log-Pareto because consistently led to higher p-values, and LSPM because it was better fitted to the energy released by the greatest earthquakes. As future work, different models for extreme events will be considered to deal only with the biggest seisms, together with a deep study about parsimonious models in the seismic moment fitting context.

Acknowledgments

Funded by FCT - Fundação para a Ciência e a Tecnologia through the project UIDB/00006/2020.

References

[1] Bird, P.; Kagan, Y.Y. (2004). Plate-Tectonic Analysis of Shallow Seismicity: Apparent Boundary Width, Beta, Corner Magnitude, Coupled Lithosphere Thickness, and Coupling in Seven Tectonic Settings. *Bull Seismol Soc Am*, **94 6**, 2380-2399.
[2] Burnham, K.; Anderson, D. (2002). *Model Selection and Multimodel Inference*. Springer.
[3] Clauset, A.; Shalizi, C. R.; Newman, M.E.J. (2009). Power-Law Distributions in Empirical Data. *Siam Rev*, **51 4**, 661-703.
[4] Di Giacomo, D., I. Bondr, D.A. Storchak, E.R. Engdahl, P. Bormann and J. Harris (2015). ISC-GEM: Global Instrumental Earthquake Catalogue (1900-2009): III. Re-computed MS and mb, proxy MW, final magnitude composition and completeness assessment. *Phys. Earth Planet. Int.*, **239**, 33-47
[5] Di Giacomo, D., E.R. Engdahl and D.A. Storchak (2018). The ISC-GEM Earthquake Catalogue (19042014): status after the Extension Project. *Earth Syst. Sci. Data*, **10**, 1877-1899.
[6] Felgueiras, M.; Santos, R. (2012). Exploring Pareto Scale Mixtures. *AIP Conf Proc*, **1479**, 1121-1124.
[7] Felgueiras, M. (2012). Explaining the seismic moment of large earthquakes by heavy and extremely heavy tailed models. *International Journal on Geomathematics*, **3 2**, 209-222.
[8] Felgueiras, M. (2013). Pareto Scale Mixtures. *Advances in Regression, Survival Analysis, Extreme Values, Markov Processes and Other Statistical Applications in Theoretical and Applied Statistics*, 281-288.
[9] Felgueiras, M.; Santos, R. (2015). Searching for the Corner Seimic Moment in Worldwide Data. *AIP Conf Proc*, **1702**, 030003.
[10] Gutenberg, B.; Richter, C.F. (1942). Earthquake magnitude, intensity, energy and acceleration. *Bull Seismol Soc Am*, **32**, 163-191.
[11] Hanks, T.C.; Kanamori, H. (1979). A Moment Magnitude Scale. *J Geophys Res*, **84**, 2348-2350.
[12] Kagan, Y.Y. (2005). Earthquake Slip Distribution: A Statistical Model. *J Geophys Res*, **110**, 1-15.
[13] Kanamori, H. (1983). Magnitude Scale and Quantication of Earthquakes. *Tectonophysics*, **93**, 185-199.
[14] Karakostas, V. (2009). Seismicity Patterns Before Strong Earthquakes in Greece. *Acta Geophys*, **57 2**, 367-386.
[15] Kijko, A. (2004). Estimation of the Maximum Earthquake Magnitude, mmax. *Pure Appl Geophys*, **161**, 1655-1681.
[16] Madariaga, R. (2009). Earthquake Scaling Laws. *Encyclopedia of Complexity and Systems*, 2581-2599, Springer.
[17] Mizera, I.; Mller, C.H. (1999). Breakdown points and variation exponents of Robust M-estimators in Linear Models. *Ann Stat*, **27 4**, 1164-1177.
[18] Monroe, J., Wicander R. (2014) *The Changing Earth: Exploring Geology and Evolution (7th Edition)*. Cengage Learning, USA.
[19] Neves, C.; Alves, I.F. (2008). Ratio of Maximum to the Sum for Testing Super Heavy Tails. *Advances in Mathematical and Statistical Modeling*, 181-194.
[20] Newman, M.E.J. (2005). Power laws, Pareto distributions and Zipfs law. *Contemp Phys*, **46 5**, 323-351.
[21] Okal, E.A.; Romanowicz, B.A. (1994). On the variation of b-values with earthquake size. *Phys Earth Planet In*, **87**, 55-76.
[22] Orowan, E. (1960). Mechanism of seismic faulting. *Geol Soc Am Bull*, **79**, 323-345.
[23] Pisarenko, V.F.; Sornette, D. (2004). Statistical detection and Characterization of a deviation from the Gutenberg-Richter distribution above Magnitude 8. *Pure Appl Geophys*, **161 4**, 839-864.
[24] Pisarenko, V.F.; Sornette, A.; Sornette, D.; Rodkin, M.V. (2014). Characterization of the tail of the distribution of earthquake magnitudes by combining the GEV and GPD descriptions of Extreme Value Theory. *Pure Appl Geophys*, **171 8**, 1599-1624.
[25] Razali, N.M.; Wah, Y.B. (2011). Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests. *Journal of Statistical Modeling and Analytics*, **2 1**, 21-33.

[26] Schmidt, M.; Lipson, H. (2009). Distilling Free-Form Natural Laws from Experimental Data. *Science*, **324**, 81-85.
[27] Utsu, T. (1999). Representation and Analysis of the Earthquake Size Distribution: A Historical Review and Some New Approaches. *Pure Appl Geophys*, **155**, 509-535.
[28] Zaliapin, I.V.; Kagan, Y.Y.; Schoenberg, F.P. (2005). Approximating the Distribution of Pareto Sums. *Pure Appl Geophys*, **162**, 1187-1228.