Pareto models for the energy released in earthquakes

MIGUEL FELGUEIRAS
ESTG and CARME
Polytechnic Institute of Leiria
CEAUL Lisbon
PORTUGAL

RUI SANTOS
ESTG
Polytechnic Institute of Leiria
CEAUL Lisbon
PORTUGAL

JOAO PAULO MARTINS
ESTG
Polytechnic Institute of Leiria
CEAUL Lisbon
PORTUGAL

Abstract: In this paper we explore Pareto based distributions to deal with the energy released by the major seisms. This is a relevant problem because great earthquakes can cause heavy losses, both human and material. The standard Pareto distribution, despite being usually well fitted to the data concerning the energy released by seisms reveals some lack of fit when dealing with the energy released by the great earthquakes. Besides the more traditional Pareto and Log-Pareto, we also consider the Extended Slash Pareto (ESP) and the Location-Scale Pareto Mixture (LSPM) distributions in this work. For the less studied ESP and LSPM distributions, we present the pa- rameters estimators and perform a simulation study in order to evaluate the estimators performance under different scenarios. Thenceforth, the four distributions are applied to two datasets (catalogs) containing information on the seisms magnitude, which has a direct connection to the energy released by the earthquakes (seismic moment). The used catalogs are considered as conveniently accurate and updated, and are being used in recent works. In conclusion, the Pareto distribution still is appropriate to fit this kind of data, but other distributions emerge as better models. The Log-Pareto distributions led to higher fitting p-values than the Pareto distribution, and LSPM also emerges as a strong competitor. LSPM is better fitted to the greatest observations and therefore givesamoreaccurateprevisionfortheenergyreleasedbythegreaterearthquakes.

Key–Words: Pareto based models, great earthquakes, seismic moment

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 [1], is the most frequently 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 [2]. These scales are not appropriate 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 [3]. 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 [4, 5] and a small maximum of possible energy released by the seisms in those regions. For example, for the Californian region, [6, 7] claim that a possible maximum lies somewhere between 8.1 and 8.3, and earthquakes with moment magnitude above 7.9 were never registered. Therefore, in this paper we analyse statistical models to fit the energy released by the great earthquakes, with large M_W magnitude. The models are restricted to the ones based in the Pareto distribution (see Section 3), because there are physical reasons to support the use of this distribution [8, 9]. The use of power law distributions (that is, Pareto based) is the rule when counting the number of events above some magnitude and the energy released by them, since these distributions usually lead to better fitting results [10]. Pareto distribution has been used in classic works in this subject [11, 12] but also in recent ones [8, 10].

2 Connecting seismic moment, moment magnitude scale and radiated energy

Gutenberg and Richter [11] developed a relation between the radiated energy in ergs (E) and the surface-wave magnitude (M_S),

$$log_{10}E = 1.5M_S + 11.8, \qquad (1)$$

and [13] connected the seismic moment with the radiated energy,

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

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, [1] used the above relations, replacing (2) in (1) and M_S by M_W. This leads to

$$log_{10}M_0 = 1.5M_W + 16.1 \qquad (3)$$

and

$$M_W = \frac{log_{10}M_0 - 16.1}{1.5}. \qquad (4)$$

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

$$M_W = \frac{log_{10}M_0 - 9.1}{1.5}. \qquad (5)$$

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. [14]).

3 Pareto based distributions in the seismic moment fitting

Seismic moment can be converted into a power law [12] 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 [6, 14, 15, 16] 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 [17] 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 near 8.0 or above [18, 5].

3.1 The Pareto distribution

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

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

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 [19, 20]. The truncated Pareto has distribution function given by

$$F_{X|X \geq t}(x) = 1 - \left(\frac{x}{t}\right)^{-\alpha}, \;\; x \geq t \geq 1. \qquad (6)$$

Pareto distribution is self-similar, that is,

$$\frac{X|X \geq t}{t} \sim \text{Pareto}(\alpha). \qquad (7)$$

Self-similarity is very interesting from 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 [3, 21, 17]) 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 α 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) has a heavier tail than the Pareto distribution. Therefore, it 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 $\widehat{\alpha} \approx 0.67$ in the Pareto fitting. Preliminary work on this distribution properties and applications is available in [22]. Log-Pareto is merely a Pareto distribution transformation, because if $X \sim \text{Pareto}(\alpha)$ then $Y = e^X \sim$

Log-Pareto (α). 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

$$F_{X|X \geq t}(x) = 1 - \left(\frac{\ln x}{\ln t} \right)^{-\alpha}, \quad x \geq t \geq e, \quad (8)$$

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

$$\frac{\ln X | X \geq t}{\ln t} \sim \text{Pareto}(\alpha). \quad (9)$$

3.3 The extended slash family

The extended slash family is obtained considering the random variable

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

where $\Theta \sim \text{Beta}(\alpha, 1)$, $\alpha > 0$. The Y variable can also be seen as a Pareto scale mixture of the X variable [23], since $Y = \Theta^{-1} X$ where $\Theta^{-1} \sim \text{Pareto}(\alpha)$. With $X \sim \text{Pareto}(\alpha)$, [3] 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. However, a detailed procedure leading to the maximum likelihood estimator, together with a simulation study, is now presented for the first time. The distribution function is

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

The ESP distribution can also be seen as a gamma distribution transformation, because if $X \sim \text{Gamma}(2, \alpha^{-1})$ then $e^X \sim \text{ESP}(\alpha)$. Setting a truncation point t,

$$F_{Y|Y \geq t}(x) = 1 - \left(\frac{1 + \alpha \ln x}{1 + \alpha \ln t} \right) \left(\frac{x}{t} \right)^{-\alpha}, \quad x \geq t \geq 1. \quad (10)$$

The α parameter has, under this distribution, a slightly different interpretation compared to the one previously presented for the Pareto distribution. However, and even considering this constraint, several authors have already discussed the possibility of the Pareto relation be too simple to adequately explain the seismic moment (see [12] for further references on this subject).

3.3.1 Estimation in the ESP

For the ESP and for the LSPM distributions (introduced in the next subsection) we also present the estimators for the α parameter, along with a simulation study to show how the process works. We decided to implement this procedure solely for the two above mentioned distributions because the estimation for the Pareto distribution is straightforward and is already implemented in most statistical software. For the Log-Pareto, it is only needed a simple transformation of the data, according to equation (9), and then α can be estimated.

Although ESP has a more complex distribution function expression, maximum likelihood estimation can be applied 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 [24], thus leading to robust estimation. The maximum likelihood estimator is obtained by solving

$$\frac{d}{d\alpha} \ln \left[\prod_{i=1}^{n} f_{Y|Y>t}(x_i) \right] = 0,$$

that is

$$\frac{d}{d\alpha} \ln \left[\prod_{i=1}^{n} \left(\frac{\alpha^2}{x_i} \frac{\ln x_i}{(\alpha \ln t + 1) \left(\frac{1}{t} x_i \right)^{\alpha}} \right) \right] = 0$$

which leads to

$$\frac{d}{d\alpha} \ln \left[\frac{\alpha^{2n} t^{\alpha n}}{(\alpha \ln t + 1)^n} \prod_{i=1}^{n} \left(\frac{\ln x_i}{x_i^{\alpha+1}} \right) \right] = 0.$$

Defining $\beta = \ln t$,

$$\frac{n\alpha^2\beta^2 + 2n\alpha\beta + 2n}{\beta\alpha^2 + \alpha} + \frac{d}{d\alpha} \sum_{i=1}^{n} \ln \left(\frac{\ln x_i}{x_i^{\alpha+1}} \right) = 0$$

and therefore

$$\frac{n\alpha^2\beta^2 + 2n\alpha\beta + 2n}{\beta\alpha^2 + \alpha} = \sum_{i=1}^{n} \ln x_i.$$

For $\gamma = \sum_{i=1}^{n} \ln x_i$ we finally obtain the estimator as

$$\widehat{\alpha} = \frac{-2n\beta + \gamma - \sqrt{-4(n\beta)^2 + 4n\beta\gamma + \gamma^2}}{2(n\beta^2 - \beta\gamma)}. \quad (11)$$

To evaluate the quality of the estimators, a simulation study was conducted. A total of $N = 10^6$ replicas were generated for each combination of magnitude M_W, sample dimension n and parameter α. The values of (M_W, n, α) presented in Table 1 only represent a part of the performed simulations. Since the results are all similar, we only present the situations where (M_W, n, α) are closer to the real ones (see Section 4). For all the above mentioned combinations, the observed $\widehat{\alpha}$ mean is presented together with the observed mean square error MSE, that is

$$\overline{\widehat{\alpha}} = \frac{\sum\limits_{i=1}^{N} \widehat{\alpha}_i}{N} \quad \text{and} \quad \text{MSE}\,(\widehat{\alpha}) = \frac{\sum\limits_{i=1}^{N} (\widehat{\alpha}_i - \widehat{\alpha})^2}{N}.$$

Table 1: Simulations results for the ESP model.

$\mathbf{M}_w$	$\mathbf{n = 25}$	$\boldsymbol{\alpha = 0.75}$	$\boldsymbol{\alpha = 1.00}$	$\boldsymbol{\alpha = 1.25}$
8.05	Mean	0.7705	1.0296	1.2887
	MSE	0.0161	0.0306	0.0502
8.15	Mean	0.7707	1.0295	1.2886
	MSE	0.0162	0.0306	0.0504
$\mathbf{M}_w$	$\mathbf{n = 50}$	$\boldsymbol{\alpha = 0.75}$	$\boldsymbol{\alpha = 1.00}$	$\boldsymbol{\alpha = 1.25}$
8.05	Mean	0.7602	1.0142	1.2685
	MSE	0.0074	0.0140	0.0230
8.15	Mean	0.7603	1.0141	1.2699
	MSE	0.0075	0.0140	0.0232
$\mathbf{M}_w$	$\mathbf{n = 100}$	$\boldsymbol{\alpha = 0.75}$	$\boldsymbol{\alpha = 1.00}$	$\boldsymbol{\alpha = 1.25}$
8.05	Mean	0.7546	1.0072	1.2595
	MSE	0.0036	0.0067	0.0110
8.15	Mean	0.7553	1.0071	1.2586
	MSE	0.0036	0.0067	0.0111

All the results show a suitable performance of the estimator, even when the sample is quite small ($n = 25$). There are no relevant differences in the results between $M_W = 8.05$ and $M_W = 8.15$. However, and as expected, the accuracy of the estimation increases with n. It is also important to note that the observed $\widehat{\alpha}$ mean is always greater than α. Thence, there is a small positive bias in the estimation process and therefore it is likely that the α estimates obtained in Section 5 are slightly above the real value of α. Despite that, and since all the obtained MSE values are small, the positive bias shall be almost irrelevant.

3.4 The Location-scale Pareto Mixture

In [25] a Location-scale Pareto mixture (LSPM) was introduced. The LSPM is now defined as

$$W = \mu + \sigma\Theta(X - 1), \tag{12}$$

where μ and σ are location and scale parameters, $\Theta \sim \text{Pareto}\,(\alpha)$ and $X \sim \text{Pareto}\,(1)$. Note that μ defines a truncation point, since the variable support is $S_W = [\mu, \infty[$. This happens because $\Theta(X - 1) > 0$, and this is the reason why we prefer to work with the random variable W defined by equation (12). The random variable $W^* = \mu + \sigma\Theta X$ would have $S_{W^*} = [\mu + \sigma, \infty[$ as support. Hence, when working with W, the μ value will be known as it represents the selected truncation point (M_W).

3.4.1 Estimation in LSPM

Maximum likelihood estimation is complex under this distribution, since it relies on iterative methods. An easy alternative is to use moments estimation method. Briefly, the moment's method consists in matching the populations moments (usually mean μ and variance σ^2 when dealing with two parameters) to the sample moments (usually mean $\overline{X}$ and variance S^2 when dealing with two parameters) and solve the equations in order to the unknown parameters (in this case α and σ) as they are functions of μ and σ.

In [25] the performance of the moment's method estimators was already investigated under simulation, and the authors concluded that this method works reasonably well in most cases. Other authors such as [26, 27] already point out that, in some situations, the moment's methods can be preferable. Hence, the moments estimation method seems to be an adequate choice. In our particular estimation problem, we are considering that μ is known, and therefore we only have to estimate two parameters (σ and α) instead of three as in [25].

To guarantee the existence of the two first moments, let us consider the random variable transformation

$$W' = \ln\,(W - \mu)\,. \tag{13}$$

The method of moments estimators are (cf. [25])

$$\widehat{\alpha} = \sqrt{\left(S'^2 - \frac{\pi^2}{3}\right)^{-1}} \quad \text{and} \quad \widehat{\sigma}' = \overline{W'} - \widehat{\alpha}^{-1}, \tag{14}$$

where S'^2 stands for the observed variance of W' and $\overline{W'}$ stands for the observed mean of W'.

To evaluate the quality of the estimators, a simulation study was conducted. A total of $N = 10^6$ replicas were generated for each combination of sample dimension n and parameters α and σ. Note that the $\mu = M_W$ parameter is irrelevant since it is removed before the estimation process begins. The values of (n, α, σ) presented in Table 2 only represent part of the performed simulations. Since the results are all

similar, we only present the situations where (α, σ) are closer to the actual values (see Section 4). For all the above mentioned combinations, the observed $\widehat{\alpha}$ and $\widehat{\sigma}$ means are presented together with the observed mean square error for both parameters.

Table 2: Simulations results for the LSPM model.

n = 25	(α, σ)	
	$(\mathbf{1.25}, \mathbf{1})$	$(\mathbf{1.25}, \mathbf{10})$
Mean	$(1.3559, 0.9335)$	$(1.3536, 9.3287)$
MSE	$(8.1329, 0.3480)$	$(10.6100, 34.913)$
	$(\mathbf{1.50}, \mathbf{1})$	$(\mathbf{1.50}, \mathbf{10})$
Mean	$(1.4341, 0.8492)$	$(1.4312, 8.4845)$
MSE	$(8.7415, 0.2959)$	$(6.6233, 29.6520)$
	$(\mathbf{1.75}, \mathbf{1})$	$(\mathbf{1.75}, \mathbf{10})$
Mean	$(1.4854, 0.7883)$	$(1.4819, 7.8795)$
MSE	$(8.9975, 0.2746)$	$(7.0397, 27.3810)$
n = 50	(α, σ)	
	$(\mathbf{1.25}, \mathbf{1})$	$(\mathbf{1.25}, \mathbf{10})$
Mean	$(1.4878, 0.9789)$	$(1.4743, 9.7871)$
MSE	$(10.2030, 0.2238)$	$(5.5486, 22.3940)$
	$(\mathbf{1.50}, \mathbf{1})$	$(\mathbf{1.50}, \mathbf{10})$
Mean	$(1.5983, 0.9018)$	$(1.5998, 9.2017)$
MSE	$(6.3259, 0.1900)$	$(5.9418, 19.0409)$
	$(\mathbf{1.75}, \mathbf{1})$	$(\mathbf{1.75}, \mathbf{10})$
Mean	$(1.6900, 0.8451)$	$(1.6879, 8.4438)$
MSE	$(8.9499, 0.1779)$	$(7.9860, 17.7540)$
n = 100	(α, σ)	
	$(\mathbf{1.25}, \mathbf{1})$	$(\mathbf{1.25}, \mathbf{10})$
Mean	$(1.5579, 1.0215)$	$(1.5547, 10.2232)$
MSE	$(7.0117, 0.1533)$	$(5.2399, 15.3830)$
	$(\mathbf{1.50}, \mathbf{1})$	$(\mathbf{1.50}, \mathbf{10})$
Mean	$(1.7534, 0.9577)$	$(1.7565, 9.5789)$
MSE	$(11.0434, 0.1288)$	$(8.8830, 12.9010)$
	$(\mathbf{1.75}, \mathbf{1})$	$(\mathbf{1.75}, \mathbf{10})$
Mean	$(1.9050, 0.9046)$	$(1.8843, 9.0372)$
MSE	$(60.9430, 0.1180)$	$(9.0497, 11.7940)$

The results presented in Table 2 are more complex to analyse. While the accuracy of the parameters estimates is reasonable, no evidence has been found to support, as expected, that the increase of the sample size leads to more accurate estimates, although the MSE almost always decreases when increasing n.

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 $M_W \geq 7.85$ 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 133 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 [28, 29] (ISC catalog). It contains data between 04-04-1904 and 31-12-2015 for a total of 139 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 [30] 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 [30].

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 has 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. This is also the reason that made us select earthquakes with $M_W \geq 7.85$ instead of $M_W \geq 7.9$. We assumed that under the USGS catalog many events with magnitude in $[7.85; 7.89]$ were registered with $M_W = 7.9$ and therefore the proposed truncated point does not contribute to an artificial difference between the catalogs.

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 both and compare their results.

5 Fitting the Seismic moment of great earthquakes

For both catalogs we adjusted the Pareto based models with $M_W \geq 7.85$, $M_W \geq 8.05$, $M_W \geq 8.15$ and $M_W \geq 8.35$, which correspond to $M_0 \geq 10^{20.875}$,

$M_0 \geq 10^{21.175}$, $M_0 \geq 10^{21.325}$ and $M_0 \geq 10^{21.625}$. These truncation points were selected in order to work solely with great seisms, near the threshold $M_W \in \,]7.8; 8.4[$ where [5] believes that distribution fitting remains an open issue, since the Pareto distribution family seems too lightly-tailed to fit the energy released by theses great seisms [3, 30, 31, 17].

The estimates $\widehat{\alpha}$ for the parameters of the analysed models are presented in Table 3 (USGS catalog) and Table 4 (ISC catalog), together with the p-values of the Kolmogorov-Smirnoff (KS) and the Anderson-Darling (AD) goodness of fit tests. These tests were selected 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. [32]). Common model selection criterion, such as the Akaike Information Criterion (AIC) or the Bayesian Information Criterion (BIC), cannot be used for the adjusted models since those models depend on different transformations of the data set, and therefore are incomparable with the usual information criteria [33]. Thus, parsimonious issues are not numerically tackled in this work and therefore will depend on each scientist opinion.

The first striking conclusion is that only one of the combinations between catalog, truncation point, model and goodness of fit test is rejected. This particular situation corresponds to the LSPM model with the USGS catalog, $M_W \geq 7.85$ and KS goodness of fit test. For all the other combinations we obtain p-value grater 0.05. For this reason, all the other 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, except for the AD test p-values when $M_W \geq 8.35$. 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 [3, 17] and that have been recommended in works like [2, 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 most of the fitted models. Even considering, for instance, the 95% confidence interval for α in the Pareto fit for the ISC catalog with $M_W \geq 8.05$ (to work

Table 3: Summary results concerning the adjusted models for the USGS catalog.

	USGS Catalog			
M_w	7.85	8.05	8.15	8.35
M_0	$10^{20.875}$	$10^{21.175}$	$10^{21.325}$	$10^{21.625}$
n	133	70	45	20
Pareto $\widehat{\alpha}$	**0.9874**	**1.0458**	**0.9552**	**0.7932**
AD *p-value*	0.2322	0.3402	0.8111	0.9662
KS *p-value*	0.0593	0.1032	0.4276	0.9650
Log-Pareto $\widehat{\alpha}$	**48.443**	**52.063**	**47.914**	**40.332**
AD *p-value*	0.2288	0.3590	0.8686	0.9662
KS *p-value*	0.0599	0.1043	0.5876	0.9650
LSPM $\widehat{\alpha}$	**1.1388**	**1.3926**	**2.0509**	**1.2900**
AD *p-value*	0.1319	0.3580	0.8220	0.9954
KS *p-value*	0.0419	0.1519	0.4291	0.9685
ESP $\widehat{\alpha}$	**1.0074**	**1.0655**	**0.9747**	**0.8124**
AD *p-value*	0.2588	0.3402	0.8111	0.9662
KS *p-value*	0.0830	0.1032	0.4276	0.9650

Table 4: Summary results concerning the adjusted models for the ISC catalog.

	ISC-GEM Catalog			
M_w	7.85	8.05	8.15	8.35
M_0	$10^{20.875}$	$10^{21.175}$	$10^{21.325}$	$10^{21.625}$
n	139	74	48	21
Pareto $\widehat{\alpha}$	**0.9842**	**1.0355**	**0.9506**	**0.7726**
AD *p-value*	0.9458	0.4479	0.9085	0.9189
KS *p-value*	0.7494	0.3611	0.8311	0.9813
Log-Pareto $\widehat{\alpha}$	**48.3184**	**51.5869**	**47.7211**	**39.3323**
AD *p-value*	0.9595	0.4213	0.9222	0.9189
KS *p-value*	0.7507	0.3611	0.9485	0.9813
LSPM $\widehat{\alpha}$	**1.9966**	**1.6561**	**1.2338**	**1.3375**
AD *p-value*	0.9123	0.4607	0.9210	0.9726
KS *p-value*	0.8434	0.4738	0.9450	0.9840
ESP $\widehat{\alpha}$	**1.0042**	**1.0553**	**0.9701**	**0.7917**
AD *p-value*	0.9463	0.4502	0.9085	0.9189
KS *p-value*	0.7494	0.3610	0.8311	0.9813

with a truncation point similar to the used in the above mentioned works), we obtain $\alpha \in [0.7786; 1.2298]$. 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, at a first glance, higher values of α contradict the idea that the Pareto's tails are too light for this kind of data.

Finally, LSPM adjustments led to slightly higher p-values (and therefore better fits) in eight of the mentioned combinations, Log-Pareto in six combinations and ESP in two combinations. It is important to note that the traditional Pareto model never presented the highest p-value and therefore it was never selected as the best fit.

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 (7), (9), (10) and (12), 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 $M_W \geq 8.05$, 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).

When analysing Figure 1, the adjustments seem good in a graphical evaluation, since the points are randomly scattered 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 such as [3, 31, 17] and here it is clearer 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, better captures the energy released by the biggest seisms.

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. Note that the used goodness of fit tests measure the distance between observed and expected quantiles. Therefore, and since the theoretical models are continuous, with increasing quantiles, many repeated measures with the same observed quantile lead to a gap between expected and observed quantiles. The Log-Pareto fit seems slightly better in the first graph, mainly because the biggest earthquake is recorded with $M_W = 9.5$ instead of $M_W = 9.6$ as in the second graph. Again, LSPM model reveals a more adequate fit for the biggest earthquakes independently from the used catalog (third and forth graphs).

6   Conclusion

We adjusted four Pareto models to the energy released by the major seisms. In addition, the estimation problem was considered when dealing with the not so common ESP and LSPM models. In what concerns to the estimation outcomes, our simulations showed better results for the ESP distribution. We highlight the relevance of Pareto models on the energy released by earthquakes, since there are both physical and statistical reasons to support these models. Moreover, the estimation of the radiated energy by great seisms re-

Figure 1: QQ-plots for the ISC catalog with $M_W \geq 8.05$ for the Pareto (first), Log-Pareto (second), LSPM (third) and ESP (forth) models.
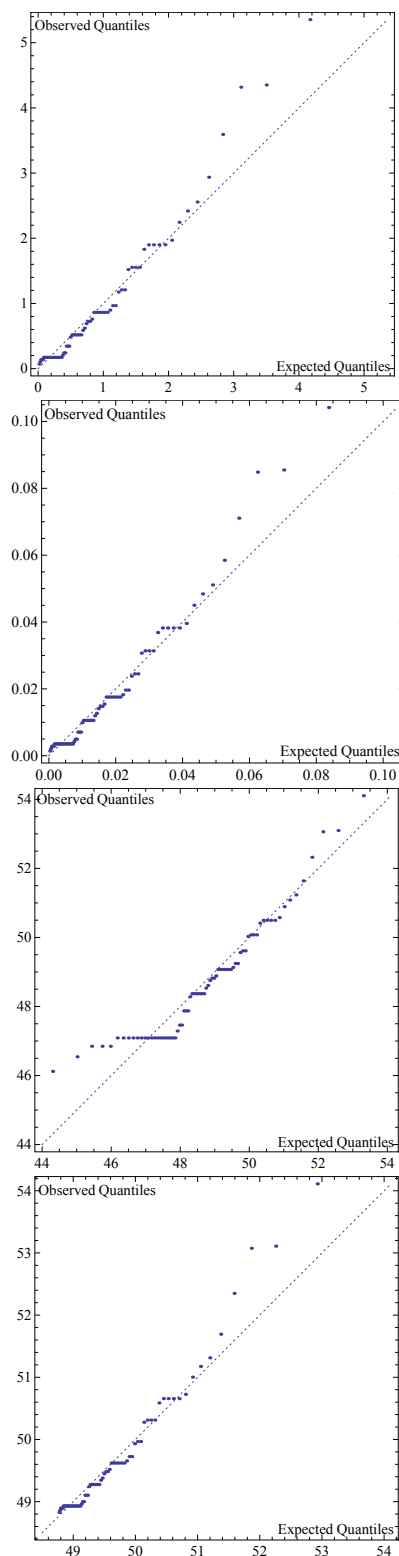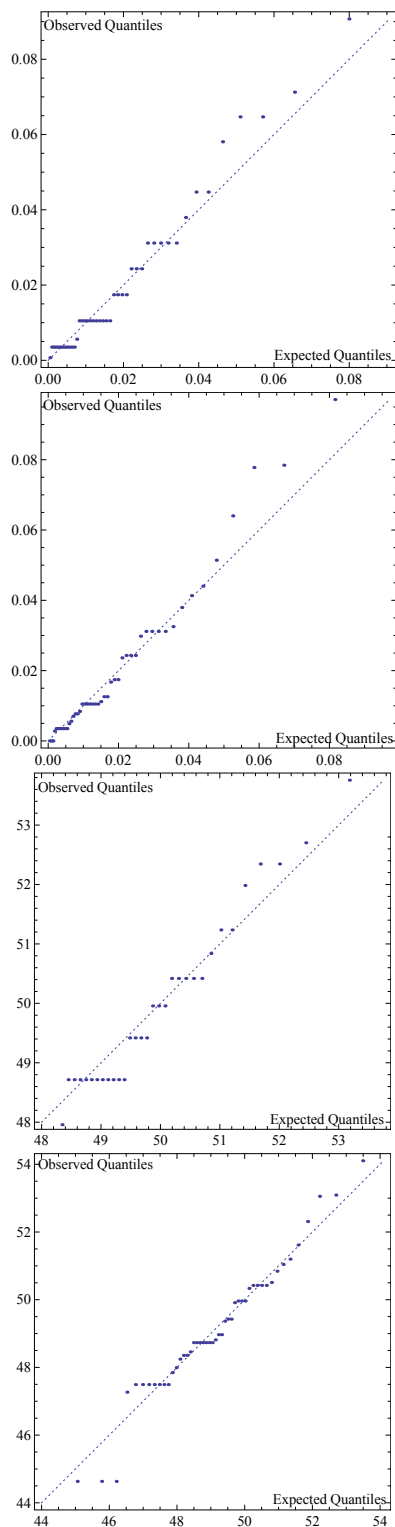
Figure 2: QQ-plots for the Log-Pareto model (top) and LSPM model (botom) with $M_W \geq 8.15$, using USGS catalog (first and third) and ISC catalog (second and forth).

mains an open issue. Therefore, updated data sets, as the analysed here, contribute to acquire a clearer picture of the mathematical shape of the phenomenon.

As expected, the achieved results were similar for all the considered models. This happens because the four models are based, as the literature recommends, in the Pareto distribution. Even though, Log-Pareto and LSPM models were the best ones. The Log-Pareto model is a simple transformation of the Pareto model and only has one parameter to be estimated, leading to a more robust model. Since this model had the highest p-values in six occasions, it is a candidate to be the best model. The LSPM model was better fitted to the energy released by the greatest earthquakes (see figures) and led to the best p-values in eight situations. The lack of fit for the larger events was one of the main reasons that led us to try to find different models, and in this particular LSPM model is clearly the best model. Even though, we need to estimate two parameters in the LSPM model. This is a drawback, because this model is more sensitive to sample changes since one extra parameter must be estimated.

As future work, different models for extreme events will be considered to deal only with the biggest seisms, together with a thorough study about parsimonious models in the seismic moment fitting context.

References:

[1] Hanks, T.C.; Kanamori, H. (1979). A Moment Magnitude Scale. *J Geophys Res*, **84**, 2348-2350.

[2] Kanamori, H. (1983). Magnitude Scale and Quantication of Earthquakes. *Tectonophysics*, **93**, 185-199.

[3] 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.

[4] Karakostas, V. (2009). Seismicity Patterns Before Strong Earthquakes in Greece. *Acta Geophys*, **57 2**, 367-386.

[5] 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.

[6] 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.

[7] Kijko, A. (2004). Estimation of the Maximum Earthquake Magnitude, mmax. *Pure Appl Geophys*, **161**, 1655-1681.

[8] Cheng, Q.; Sun, H. (2018). Variation of singularity of earthquake-size distribution with respect to tectonic regime. *Geoscience Frontiers*, **9**, 453-458.

[9] Gudmundsson, A. (2014). Elastic energy release in great earthquakes and eruptions. *Frontiers Earth Science*, **2**, 10.

[10] Gudmundsson, A.; Mohajeri, N. (2013). Relations between the scaling exponents, entropies, and energies of fracture networks. *Bulletin de la Societe Geologique de France*, **184**, 377-387.

[11] Gutenberg, B.; Richter, C.F. (1942). Earthquake magnitude, intensity, energy and acceleration. *Bull Seismol Soc Am*, **32**, 163-191.

[12] Utsu, T. (1999). Representation and Analysis of the Earthquake Size Distribution: A Historical Review and Some New Approaches. *Pure Appl Geophys*, **155**, 509-535.

[13] Orowan, E. (1960). Mechanism of seismic faulting. *Geol Soc Am Bull*, **79**, 323-345.

[14] Madariaga, R. (2009). Earthquake Scaling Laws. *Encyclopedia of Complexity and Systems*, 2581-2599, Springer.

[15] Newman, M.E.J. (2005). Power laws, Pareto distributions and Zipfs law. *Contemp Phys*, **46 5**, 323-351.

[16] 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.

[17] Zaliapin, I.V.; Kagan, Y.Y.; Schoenberg, F.P. (2005). Approximating the Distribution of Pareto Sums. *Pure Appl Geophys*, **162**, 1187-1228.

[18] Monroe, J., Wicander R. (2014) *The Changing Earth: Exploring Geology and Evolution (7th Edition)*. Cengage Learning, USA.

[19] Clauset, A.; Shalizi, C. R.; Newman, M.E.J. (2009). Power-Law Distributions in Empirical Data. *Siam Rev*, **51 4**, 661-703.

[20] Kagan, Y.Y. (2005). Earthquake Slip Distribution: A Statistical Model. *J Geophys Res*, **110**, 1-15.

[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] 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.

[23] 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.

[24] Mizera, I.; Mller, C.H. (1999). Breakdown points and variation exponents of Robust M-estimators in Linear Models. *Ann Stat*, **27 4**, 1164-1177.

[25] Felgueiras, M.; Santos, R. (2012). Exploring Pareto Scale Mixtures. *AIP Conf Proc*, **1479**, 1121-1124.

[26] Fielitz, B.; Myers, B.L. (1975). Estimation of parameters in the beta distribution. *Decision Sciences*, **6**, 1-13.

[27] Johnson, N.L.; Kotz, S.; Balakrishnan, N. (1994). *Continuous Univariate Distributions, Volume II*. Wiley.

[28] 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.

[29] 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.

[30] Felgueiras, M.; Santos, R. (2015). Searching for the Corner Seimic Moment in Worldwide Data. *AIP Conf Proc*, **1702**, 030003.

[31] Schmidt, M.; Lipson, H. (2009). Distilling Free-Form Natural Laws from Experimental Data. *Science*, **324**, 81-85.

[32] 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.

[33] Burnham, K.; Anderson, D. (2002). *Model Selection and Multimodel Inference*. Springer.