Fitting and Goodness-of-Fit Test of Non-Truncated and Truncated Power-Law Distributions

Anna DELUCA and Álvaro CORRAL

Centre de Recerca Matemàtica, Edifici C, Bellaterra, Barcelona, Spain

Departament de Matemàtiques, Universitat Autònoma de Barcelona, Cerdanyola, Spain
e-mails: ADeluca@crm.cat, ACorral@crm.cat

Abstract

Recently, Clauset, Shalizi, and Newman have proposed a systematic method to find over which range (if any) a certain distribution behaves as a power law. However, their method has been found to fail, in the sense that true (simulated) power-law tails are not recognized as such in some instances, and then the power-law hypothesis is rejected. Moreover, the method does not work well when extended to power-law distributions with an upper truncation. We explain in detail a similar but alternative procedure, valid for truncated as well as for non-truncated power-law distributions, based in maximum likelihood estimation, the Kolmogorov–Smirnov goodness-of-fit test, and Monte Carlo simulations. An overview of the main concepts as well as a recipe for their practical implementation is provided. The performance of our method is put to test on several empirical data which were previously analyzed with less systematic approaches. We find the functioning of the method very satisfactory.

Key words: power-law distribution estimation, goodness-of-fit tests, binning, seismic-moment distribution, waiting-time distribution, tropical-cyclone energy.

1. INTRODUCTION

Over the last decades, the importance of power-law distributions has continuously increased, not only in geoscience but also elsewhere (Johnson *et al.* 1994). These are probability distributions defined by a probability density (for a continuous variable x) or by a probability mass function (for a discrete variable x) given by

$$f(x) \propto \frac{1}{x^\alpha} \, , \tag{1}$$

for $x \geq a$ and $a > 0$, with a normalization factor (hidden in the proportionality symbol $\propto$) which depends on whether x is continuous or discrete. In any case, normalization implies $\alpha > 1$. Sometimes power-law distributions are also called Pareto distributions (Evans *et al.* 2000, Johnson *et al.* 1994) or Riemann zeta distributions in the discrete case (Johnson *et al.* 2005), although in other contexts the name Pareto is associated to a slightly different distribution (Johnson *et al.* 1994). So we stick to the clearer term of power-law distribution.

These have remarkable, non-usual statistical properties, as are scale invariance and divergence of moments. The first one means that power-law functions (defined between 0 and ∞) are invariant under (properly performed) linear rescaling of axes (both x and f) and therefore have no characteristic scale, and hence cannot be used to define a prototype of the observable represented by x (Christensen and Moloney 2005, Corral 2008, Newman 2005, Takayasu 1990). For example, no unit of distance can be defined from the gravitational field of a point mass (a power law), whereas a time unit can be defined for radioactive decay (an exponential function). However, as power-law distributions cannot be defined for all $x > 0$ but for $x \geq a > 0$, their scale invariance is not "complete" or strict.

A second uncommon property is the non-existence of finite moments; for instance, if $\alpha \leq 2$ not a single finite moment exists (no mean, no variance, *etc.*). This has important consequences, as the law of large numbers does not hold (Kolmogorov 1956, p. 65), *i.e.*, the mean of a sample does not converge to a finite value as the size of the sample increases; rather, the sample mean tends to infinite (Shiryaev 1996, p. 393). If $2 < \alpha \leq 3$ the mean exists and is finite, but higher moments are infinite, which means, for instance, that the central limit theorem, in its classic formulation, does not apply (the mean of a sample is not normally distributed and has infinite standard deviation) (Bouchaud and Georges 1990). Higher α's yield higher-order moments infinite, but then the situation is not so "critical". Newman reviews other peculiar properties of power-law distributions, such as the 80|20 rule (Newman 2005).

Although the normal (or Gaussian) distribution gives a non-zero probability that a human being is 10 m or 10 km tall, the definition of the probability density up to infinity is not questionable at all, and the same happens

with an exponential distribution and most "standard" distributions in probability theory. However, one already sees that the power-law distribution is problematic, in particular for $\alpha \leq 2$, as it predicts an infinite mean, and for $2 \leq \alpha < 3$, as the variability of the sample mean is infinite. Of course, there can be variables having an infinite mean (one can easily simulate in a computer processes in which the time between events has an infinite mean), but in other cases, for physical reasons, the mean should be finite. In such situations a simple generalization is the truncation of the tail (Aban *et al.* 2006, Burroughs and Tebbens 2001, Carrillo-Menéndez and Suárez 2012, Johnson *et al.* 1994), yielding the truncated power-law distribution, defined in the same way as before by $f(x) \propto 1/x^{\alpha}$ but with $a \leq x \leq b$, with b finite, and with normalizing factor depending now on a and b (in some cases it is possible to have $a = 0$, see next section). Obviously, the existence of a finite upper cutoff b automatically leads to well-behaved moments, if the statistics is enough to "see" the cutoff; on the other hand, a range of scale invariance can persist, if $b \gg a$. What one finds in some practical problems is that the statistics is not enough to decide which is the sample mean and one cannot easily conclude if a pure power law or a truncated power law is the right model for the data.

A well known example of (truncated or not) power-law distribution is the Gutenberg–Richter law for earthquake "size" (Kagan 2002, Kanamori and Brodsky 2004, Utsu 1999). If by size we understand radiated energy, the Gutenberg–Richter law implies that, in any seismically active region of the world, the sizes of earthquakes follow a power-law distribution, with an exponent $\alpha = 1 + 2B/3$ and B close to 1. In this case, scale invariance means that if one asks how big (in terms of radiated energy) earthquakes are in a certain region, such a simple question has no possible answer. The non-convergence of the mean energy can easily be checked from data: catastrophic events such as the Sumatra–Andaman mega-earthquake of 2004 contribute to the mean much more than the previous recorded history (Corral and Font-Clos 2013). Note that for the most common formulation of the Gutenberg–Richter law, in terms of the magnitude, earthquakes are not power-law distributed, but this is due to the fact that magnitude is an (increasing) exponential function of radiated energy, and therefore magnitude turns out to be exponentially distributed. In terms of magnitude, the statistical properties of earthquakes are trivial (well behaved mean, existence of a characteristic magnitude, *etc.*), but we insist that this is not the case in terms of radiated energy.

Malamud (2004) lists several other natural hazards following power-law distributions in some (physical) measure of size, such as rockfalls, landslides (Hergarten 2002), volcanic eruptions (Lahaie and Grasso 1998, McClelland *et al.* 1989), and forest fires (Malamud *et al.* 2005), and we can add rainfall (Peters *et al.* 2001, 2010), tropical cyclones (roughly speaking, hurricanes)

(Corral *et al.* 2010), auroras (Freeman and Watkins 2002), tsunamis (Burroughs and Tebbens 2005), *etc.* In some cases this broad range of responses is triggered simply by a small driving or perturbation (the slow motion of tectonic plates for earthquakes, the continuous pumping of solar radiation in hurricanes, *etc.*); then, this highly nonlinear relation between input and output can be labeled as crackling noise (Sethna *et al.* 2001). Notice that this does not apply for tsunamis, for instance, as they are not slowly driven (or at least not directly slowly driven).

Aschwanden (2013) reviews disparate astrophysical phenomena which are distributed according to power laws, some of them related to geoscience: sizes of asteroids, craters in the Moon, solar flares, and energy of cosmic rays. In the field of ecology and close areas, the applicability of power-law distributions has been overviewed by White *et al.* (2008), mentioning also island and lake sizes. Aban *et al.* (2006) provides bibliography for power-law and other heavy-tailed distributions in diverse disciplines, including hydrology, and Burroughs and Tebbens (2001) provide interesting geological examples.

A theoretical framework for power-law distributed sizes (and durations) of catastrophic phenomena not only in geoscience but also in condensed matter physics, astrophysics, biological evolution, neuroscience, and even the economy, is provided by the concept of self-organized criticality, and summarized by the sandpile paradigm (Bak 1996, Christensen and Moloney 2005, Jensen 1998, Pruessner 2012, Sornette 2004). However, although the ideas of self-organization and criticality are very reasonable in the context of most of the geosystems mentioned above (Corral 2010, Peters and Christensen 2006, Peters and Neelin 2006), one cannot rule out other mechanisms for the emergence of power-law distributions (Czechowski 2003, Dickman 2003, Mitzenmacher 2004, Newman 2005, Sornette 2004).

On the other hand, it is interesting to mention that, in addition to sizes and durations, power-law distributions have also been extensively reported in time between the occurrences of natural hazards (waiting times), as for instance in solar flares (Baiesi *et al.* 2006, Boffetta *et al.* 1999), earthquakes (Bak *et al.* 2002, Corral 2003, 2004b), or solar wind (Wanliss and Weygand 2007); in other cases the distributions contain a power-law part mixed with other factors (Corral 2004a, 2009b, Geist and Parsons 2008, Saichev and Sornette 2006). Nevertheless, the possible relation with critical phenomena is not direct (Corral 2005, Paczuski *et al.* 2005). The distance between events, or jumps, has received relatively less attention (Corral 2006, Davidsen and Paczuski 2005, Felzer and Brodsky 2006).

The importance of power-law distributions in geoscience is apparent; however, some of the evidence gathered in favor of this paradigm can be considered as "anecdotic" or tentative, as it is based on rather poor data analysis.

A common practice is to find some (naive or not) estimation of the probability density or mass function $f(x)$ and plot $\ln f(x)$ *versus* $\ln x$ and look for a linear dependence between both variables. Obviously, a power-law distribution should behave in that way, but the opposite is not true: an apparent straight line in a log-log plot of $f(x)$ should not be considered a guarantee of an underlying power-law distribution, or perhaps the exponent obtained from there is clearly biased (Bauke 2007, Clauset *et al.* 2009, Goldstein *et al.* 2004, White *et al.* 2008). But in order to discriminate between several competing theories or models, as well as in order to extrapolate the available statistics to the most extreme events, it is very important to properly fit power laws and to find the right power-law exponent (if any) (White *et al.* 2008).

The subject of this paper is a discussion on the most appropriate fitting, testing of the goodness-of-fit, and representation of power-law distributions, both non-truncated and truncated. A consistent and robust method will be checked on several examples in geoscience, including earthquakes, tropical cyclones, and forest fires. The procedure is in some points analogous to that of Clauset *et al.* (2009), although there are variations is some key steps, in order to correct several drawbacks of the original method (Corral *et al.* 2011, Peters *et al.* 2010). The most important difference is in the criterion to select the range over which the power law holds. As the case of most interest in geoscience is that of a continuous random variable, the more involving discrete case will be postponed to a separate publication (Corral *et al.* 2012).

2. POWER-LAW FITS AND GOODNESS-OF-FIT TESTS

2.1 Non-truncated and truncated power-law distributions

Let us consider a continuous power-law distribution, defined in the range $a \leq x \leq b$, where b can be finite or infinite and $a \geq 0$. The probability density of x is given by

$$f(x) = \frac{\alpha - 1}{a^{1-\alpha} - 1/b^{\alpha-1}} \left(\frac{1}{x}\right)^{\alpha}, \qquad (2)$$

the limit $b \to \infty$ with $\alpha > 1$ and $a > 0$ provides the non-truncated power-law distribution, also called here pure power law; otherwise, for finite b one has the truncated power law, for which no restriction exists on α if $a > 0$, but $\alpha < 1$ if $a = 0$ (which is sometimes referred to as the power-function distribution (Evans *et al.* 2000)); the case $\alpha = 1$ needs a separate treatment, with

$$f(x) = \frac{1}{x \ln(b/a)} \, . \qquad (3)$$

We will consider in principle that the distribution has a unique parameter, α, and that a and b are fixed and known values. Remember that, at point x, the probability density function of a random variable is defined as the probability per unit of the variable that the random variable lies in a infinitesimal interval around x, that is,

$$f(x) = \lim_{\Delta x \to 0} \frac{\text{Prob}[x \leq \text{random variable} < x + \Delta x]}{\Delta x} \, , \qquad (4)$$

and has to verify $f(x) \geq 0$ and $\int_{-\infty}^{\infty} f(x)dx = 1$; see for instance Ross (2002).

Equivalently, the distribution can be also characterized by its (complementary) cumulative distribution function,

$$S(x) = \text{Prob}[\text{random variable} \geq x] = \int_{x}^{\infty} f(x')dx'. \qquad (5)$$

For a truncated or non-truncated power law this leads to

$$S(x) = \frac{1/x^{\alpha-1} - 1/b^{\alpha-1}}{a^{1-\alpha} - 1/b^{\alpha-1}} \, , \qquad (6)$$

if $\alpha \neq 1$ and

$$S(x) = \frac{\ln(b/x)}{\ln(b/a)} \, , \qquad (7)$$

if $\alpha = 1$. Note that although $f(x)$ always has a power-law shape, $S(x)$ only has it in the non-truncated case ($b \to \infty$ and $\alpha > 1$); nevertheless, even not being a power law in the truncated case, the distribution is a power law, as it is $f(x)$ and not $S(x)$ which gives the name to the distribution.

2.2 Problematic fitting methods

Given a set of data, there are many methods to fit a probability distribution. Goldstein *et al.* (2004), Bauke (2007), White *et al.* (2008), and Clauset *et al.* (2009) check several methods based in the fitting of the estimated probability densities or cumulative distributions in the power-law case. As mentioned in the first section, $\ln f(x)$ is then a linear function of $\ln x$, both for non-truncated and truncated power laws. The same holds for $\ln S(x)$, but only in the non-truncated case. So, one can either estimate $f(x)$ from data, using some binning procedure, or estimate $S(x)$, for which no binning is necessary, and then fit a straight line by the least-squares method. As we find White *et al.*'s (2008) study the most complete, we summarize their results below, although those of the other authors are not very different.

For non-truncated power-law distributions, White *et al.* (2008) find that the results of the least-squares method using the cumulative distribution are reasonable, although the points in $S(x)$ are not independent and linear regression should yield problems in this case. We stress that this procedure only can work for non-truncated distributions (*i.e.*, with $b \to \infty$); truncated ones yield bad results (Burroughs and Tebbens 2001).

The least-squares method applied to the probability density $f(x)$ has several variations, depending on the way of estimating $f(x)$. Using linear binning one obtains a simple histogram, for which the fitting results are catastrophic (Pueyo and Jovani 2006, Bauke 2007, Goldstein *et al.* 2004, White *et al.* 2008). This is not unexpected, as linear binning of a heavy-tailed distribution can be considered as a very naive approach. If instead of linear binning one uses logarithmic binning the results improve (when done "correctly"), and are reasonable in some cases, but they still show some bias, high variance, and bin-size dependence. A fundamental point is to avoid having empty bins, as they are disregarded in logscale, introducing an important bias.

In summary, methods of estimation of probability-distribution parameters based on least-squares fitting can have many problems, and usually the results are biased. Moreover, these methods do not take into account that the quantity to be fitted is a probability distribution (*i.e.*, once the distributions are estimated, the method is the same as for any other kind of function). We are going to see that the method of maximum likelihood is precisely designed for dealing with probability distributions, presenting considerable advantages in front of the other methods just mentioned.

2.3 Maximum likelihood estimation

Let us denote a sample of the random variable x with N elements as x_1, x_2, ..., x_N, and let us consider a probability distribution $f(x)$ parameterized by α. The likelihood function $L(\alpha)$ is defined as the joint probability density (or the joint probability mass function if the variable were discrete) evaluated at x_1, x_2, ..., x_N in the case in which the variables were independent, *i.e.*,

$$L(\alpha) = \prod_{i=1}^{N} f(x_i) \,. \tag{8}$$

Note that the sample is considered fixed, and it is the parameter α what is allowed to vary. In practice it is more convenient to work with the log-likelihood, the natural logarithm of the likelihood (dividing by N also, in our definition),

$$\ell(\alpha) = \frac{1}{N} \ln L(\alpha) = \frac{1}{N} \sum_{i=1}^{N} \ln f(x_i) \,. \tag{9}$$

The maximum likelihood (ML) estimator of the parameter α based on the sample is just the value of α that yields the maximum of $\ell(\alpha)$ (which coincides with the maximum of $L(\alpha)$, obviously). For a given sample, we will denote the ML estimator as α_e (e is from empirical), but it is important to realize that the ML estimator is indeed a statistic (a quantity calculated from a random sample) and therefore can be considered as a random variable; in this case it is denoted as $\hat{\alpha}$. In a formula,

$$\alpha_e = \arg\max_{\forall \alpha} \ell(\alpha) , \tag{10}$$

where argmax refers to the argument of the function ℓ that makes it maximum.

For the truncated or the non-truncated continuous power-law distribution we have, substituting $f(x)$ from Eqs. (2)-(3) and introducing $r = a/b$, disregarding the case $a = 0$,

$$\ell(\alpha) = \ln \frac{\alpha - 1}{1 - r^{\alpha-1}} - \alpha \ln \frac{g}{a} - \ln a , \quad \text{if } \alpha \neq 1, \tag{11}$$

$$\ell(\alpha) = -\ln \ln \frac{1}{r} - \ln g , \quad \text{if } \alpha = 1 ; \tag{12}$$

where g is the geometric mean of the data, $\ln g = N^{-1} \sum_1^N \ln x_i$, and the last term in each expression is irrelevant for the maximization of $\ell(\alpha)$. The equation for $\alpha = 1$ is necessary in order to avoid overflows in the numerical implementation of Eq. (11). Remember that the distribution is only parameterized by α, whereas a and b (and r) are constant parameters; therefore, $\ell(\alpha)$ is not a function of a and b, but of α.

In order to find the maximum of $\ell(\alpha)$ one can derive with respect α and set the result equal to zero (Aban *et al.* 2006, Johnson *et al.* 1994),

$$\frac{d\ell(\alpha)}{d\alpha}\bigg|_{\alpha=\alpha_e} = \frac{1}{\alpha_e - 1} + \frac{r^{\alpha_e-1} \ln r}{1 - r^{\alpha_e-1}} - \ln \frac{g}{a} = 0 , \tag{13}$$

which constitutes the so-called likelihood equation for this problem. For a non-truncated distribution, $r = 0$, and it is clear that there is one and only one solution,

$$\alpha_e = 1 + \frac{1}{\ln(g/a)} , \tag{14}$$

which corresponds to a maximum, as

$$L(\alpha) = e^{N\ell(\alpha)} = \frac{1}{a^N}(\alpha - 1)^N e^{-N\alpha \ln(g/a)} , \tag{15}$$

has indeed a maximum (resembling a gamma probability density, see next subsection). Figure 1 illustrates the log-likelihood function and its derivative, for simulated power-law data.
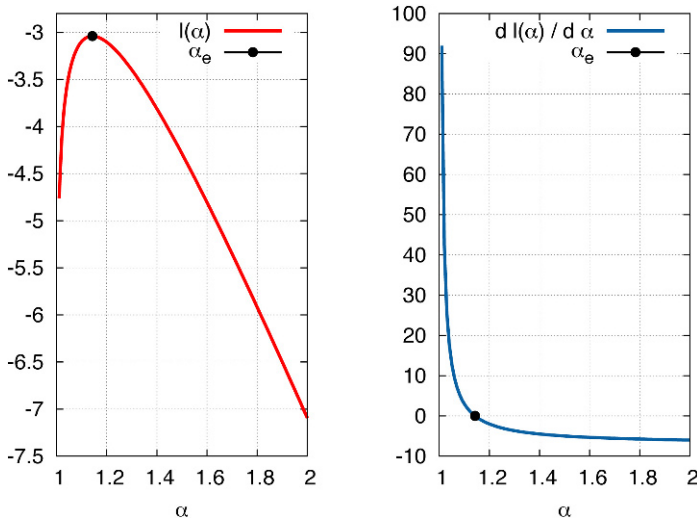
Fig. 1. Log-likelihood $\ell(\alpha)$ and its derivative, for simulated non-truncated power-law data with exponent $\alpha = 1.15$ and $a = 0.001$. The total number of data is $N_{\text{tot}} = 1000$. The resulting estimation yields $\alpha_e = 1.143$, which will lead to a confidence interval $\alpha \pm \sigma = 1.143 \pm 0.005$.

In the truncated case it is not obvious that there is a solution to the likelihood equation (Aban *et al.* 2006); however, one can take advantage of the fact that the power-law distribution, for fixed a and b, can be viewed as belonging to the regular exponential family, for which it is known that the maximum likelihood estimator exists and is unique; see Barndorff-Nielsen (1978, p. 151) or del Castillo (2013). Indeed, in the single-parameter case, the exponential family can be written in the form

$$f(x) = C^{-1}(\alpha)H(x)e^{\theta(\alpha)\cdot T(x)} , \tag{16}$$

where both $\theta(\alpha)$ and $T(x)$ can be vectors, the former containing the parameter α of the family. Then, for $\theta(\alpha) = -\alpha$, $T(x) = \ln x$, and $H(x) = 1$ we obtain the (truncated or not) power-law distribution, which therefore belongs to the regular exponential family, which guarantees the existence of a unique ML solution.

In order to find the ML estimator of the exponent in the truncated case, we proceed by maximizing directly the log-likelihood $\ell(\alpha)$ (rather than by solving the likelihood equation). The reason is a practical one, as our procedure is part of a more general method, valid for arbitrary distributions $f(x)$, for which the derivative of $\ell(\alpha)$ can be difficult to evaluate. We will use the downhill simplex method, through the routine "amoeba" of Press *et al.* (1992), although any other simpler maximization procedure should work, as the problem is one-dimensional, in this case. One needs to take care when the value of α gets very

close to one in the maximization algorithm, and then replace $\ell(\alpha)$ by its limit at $\alpha = 1$,

$$\ell(\alpha) \rightarrow_{\alpha \to 1} - \ln \ln \frac{1}{r} - \alpha \ln \frac{g}{a} - \ln a \,, \tag{17}$$

which is in agreement with the likelihood function for a (truncated) power-law distribution with $\alpha = 1$.

An important property of ML estimators, not present in other fitting methods, is their invariance under re-parameterization. If instead of working with parameter α we use $\nu = h(\alpha)$, then the ML estimator of ν is "in agreement" with that of α, *i.e.*, $\hat{\nu} = h(\hat{\alpha})$. Indeed,

$$\frac{d\ell}{d\nu} = \frac{d\ell}{d\alpha}\frac{d\alpha}{d\nu} \,, \tag{18}$$

so, the maximum of ℓ as a function of ν is attained at the point $h(\hat{\alpha})$, provided that the function h is "one-to-one". Note that the parameters could be multidimensional as well. Casella and Berger (2002) study this invariance with much more care.

In their comparative study, White *et al.* (2008) conclude that maximum likelihood estimation outperforms the other fitting methods, as always yields the lowest variance and bias of the estimator. This is not unexpected, as the ML estimator is, mathematically, the one with minimum variance among all asymptotically unbiased estimators. This property is called asymptotical efficiency (Bauke 2007, White *et al.* 2008).

2.4 Standard deviation of the ML estimator

The main result of this subsection is the value of the uncertainty σ of $\hat{\alpha}$, represented by the standard deviation of $\hat{\alpha}$ and given by

$$\sigma = \frac{1}{\sqrt{N}} \left[\frac{1}{(\alpha_e - 1)^2} - \frac{r^{\alpha_e - 1} \ln^2 r}{(1 - r^{\alpha_e - 1})^2} \right]^{-1/2} \,, \tag{19}$$

(Aban *et al.* 2006). This formula can be used directly, although σ can be computed as well from Monte Carlo simulations, as explained in another subsection. A third option is the use of the jackknife procedure, as done by Peters *et al.* (2010). The three methods lead to essentially the same results. The rest of this subsection is devoted to the particular derivation of σ for a non-truncated power-law distribution, and therefore can be skipped by readers interested mainly in the practical use of ML estimation.

For the calculation of $\hat{\alpha}$ (the ML estimator of α) one needs to realize that this is indeed a statistic (a quantity calculated from a random sample) and therefore it can be considered as a random variable. Note that α denotes the

true value of the parameter, which is unknown. It is more convenient to work with $\alpha - 1$ (the exponent of the cumulative distribution function); in the non-truncated case $(r = 0$ with $\alpha > 1)$ we can easily derive its distribution. First let us consider the geometric mean of the sample, g, rescaled by the minimum value a,

$$\ln \frac{g}{a} = \frac{1}{N} \sum_{i=1}^{N} \ln \frac{x_i}{a} . \tag{20}$$

As each x_i is power-law distributed (by hypothesis), a simple change of variables shows that $\ln(x_i/a)$ turns out to be exponentially distributed, with scale parameter $1/(\alpha - 1)$; then, the sum will be gamma distributed with the same scale parameter and with shape parameter given by N (this is the key property of the gamma distribution (Durrett 2010)). Therefore, $\ln(g/a)$ will follow the same gamma distribution but with scale parameter $N^{-1}(\alpha - 1)^{-1}$.

At this point it is useful to introduce the generalized gamma distribution (Evans *et al.* 2000, Johnson *et al.* 1994, Kalbfleisch and Prentice 2002), with density, for a random variable $y \geq 0$,

$$D(y) = \frac{|\delta|}{c\Gamma(\gamma/\delta)} \left(\frac{y}{c}\right)^{\gamma-1} e^{-(y/c)^{\delta}} , \tag{21}$$

where $c > 0$ is the scale parameter and γ and δ are the shape parameters, which have to verify $0 < \gamma/\delta < \infty$ (so, the only restriction is that they have the same sign, although the previous references only consider $\gamma > 0$ and $\delta > 0$); the case $\delta = 1$ yields the usual gamma distribution and $\delta = \gamma = 1$ is the exponential one. Again, changing variables one can show that the inverse $z = 1/y$ of a generalized gamma variable is also a generalized gamma variable, but with transformed parameters,

$$\gamma, \delta, c \rightarrow -\gamma, -\delta, \frac{1}{c} . \tag{22}$$

So, $\hat{\alpha} - 1 = z = 1/\ln(g/a)$ will have a generalized gamma distribution, with parameters $-N$, -1, and $N(\alpha - 1)$ (keeping the same order as above). Introducing the moments of this distribution (Evans *et al.* 2000),

$$\langle y^m \rangle = c^m \frac{\Gamma\left(\frac{\gamma+m}{\delta}\right)}{\Gamma(\gamma/\delta)} \tag{23}$$

(valid for $m > -\gamma$ if $\gamma > 0$ and for $m < |\gamma|$ if $\gamma < 0$, and $\langle y^m \rangle$ infinite otherwise), we obtain the expected value of $\hat{\alpha} - 1$,

$$\langle \hat{\alpha} - 1 \rangle = \frac{N(\alpha - 1)}{N - 1} . \tag{24}$$

Note that the ML estimator, $\hat{\alpha}$, is biased, as its expected value does not co-incide with the right value, α; however, asymptotically, the right value is re-covered. An unbiased estimator of α can be obtained for a small sample as $(1 - 1/N)\alpha_e + 1/N$, although this will not be of interest to us.

In the same way, the standard deviation of $\hat{\alpha} - 1$ (and of $\hat{\alpha}$) turns out to be

$$\sigma = \sqrt{\langle(\hat{\alpha} - 1)^2\rangle - \langle\hat{\alpha} - 1\rangle^2} = \frac{\alpha - 1}{(1 - 1/N)\sqrt{N - 2}}, \qquad (25)$$

which leads asymptotically to $(\alpha - 1)/\sqrt{N}$. In practice, we need to replace α by the estimated value α_e; then, this is nothing else than the limit $r = 0$ ($b \to \infty$) of the general formula stated above for σ (Aban *et al.* 2006). The fact that the standard deviation tends to zero asymptotically (together with the fact that the estimator is asymptotically unbiased) implies that any single estimation converges (in probability) to the true value, and therefore the estimator is said to be consistent.

2.5 Goodness-of-fit test

One can realize that the maximum likelihood method always yields a ML esti-mator for α, no matter which data one is using. In the case of power laws, as the data only enters in the likelihood function through its geometric mean, any sample with a given geometric mean yields the same value for the estimation, although the sample can come from a true power law or from any other dis-tribution. So, no quality of the fit is guaranteed and thus, maximum likelihood estimation should be rather called minimum unlikelihood estimation. For this reason a goodness-of-fit test is necessary (although recent works do not take into account this fact (Baró and Vives 2012, Kagan 2002, White *et al.* 2008)).

Following Goldstein *et al.* (2004) and Clauset *et al.* (2009) we use the Kolmogorov–Smirnov (KS) test (Chicheportiche and Bouchaud 2012, Press *et al.* 1992), based on the calculation of the KS statistic or KS distance d_e between the theoretical probability distribution, represented by $S(x)$, and the empirical one, $S_e(x)$. The latter, which is an unbiased estimator of the cumula-tive distribution (Chicheportiche and Bouchaud 2012), is given by the stepwise function

$$S_e(x) = n_e(x)/N, \qquad (26)$$

where $n_e(x)$ is the number of data in the sample taking a value of the variable larger than or equal to x. The KS statistic is just the maximum difference, in absolute value, between $S(x)$ and $n_e(x)/N$, that is,

$$d_e = \max_{a \leq x \leq b} |S(x) - S_e(x)| = \max_{a \leq x \leq b} \left| \frac{1}{1 - r^{\alpha_e - 1}} \left[\left(\frac{a}{x}\right)^{\alpha_e - 1} - r^{\alpha_e - 1} \right] - \frac{n_e(x)}{N} \right|, \quad (27)$$
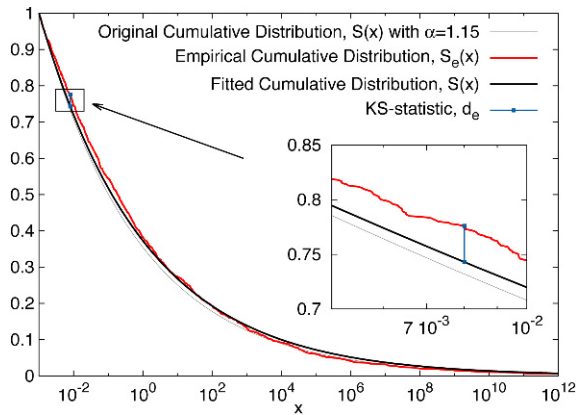
Fig. 2. Empirical (complementary) cumulative distribution for a simulated non-truncated power-law distribution with $\alpha = 1.15$, $a = 0.001$, and $N_{\text{tot}} = 1000$, together with its corresponding fit, which yields $\alpha_e = 1.143$. The maximum difference between both curves, $d_e = 0.033$, is marked as an illustration of the calculation of the KS statistic. The original theoretical distribution, unknown in practice, is also plotted. Colour version of this figure is available in electronic edition only.

where the bars denote absolute value. Note that the theoretical cumulative distribution $S(x)$ is parameterized by the value of α obtained from ML, α_e. In practice, the difference only needs to be evaluated around the points x_i of the sample (as the routine "ksone" of Press *et al.* (1992) does) and not for all x. A more strict mathematical definition uses the supremum instead of the maximum, but in practice the maximum works perfectly. We illustrate the procedure in Fig. 2, with a simulation of a non-truncated power law.

Intuitively, if d_e is large the fit is bad, whereas if d_e is small the fit can be considered as good. But the relative scale of d_e is provided by its own probability distribution, through the calculation of a p-value. Under the hypothesis that the data follow indeed the theoretical distribution, with the parameter α obtained from our previous estimation (this is the null hypothesis), the p-value provides the probability that the KS statistic takes a value larger than the one obtained empirically, *i.e.*,

$$p = \text{Prob}[\text{KS statistic for power-law data (with } \alpha_e) \text{ is } > d_e] ; \qquad (28)$$

then, bad fits will have rather small p-values.

It turns out that, in principle, the distribution of the KS statistic is known, at least asymptotically, independently of the underlying form of the distribution, so

$$p_Q = Q(d_e\sqrt{N} + 0.12d_e + 0.11d_e/\sqrt{N})$$

$$= 2\sum_{j=1}^{\infty}(-1)^{j-1}\exp[-2j^2(d_e\sqrt{N} + 0.12d_e + 0.11d_e/\sqrt{N})^2] , \qquad (29)$$

for which one can use the routine "probks" of Press *et al.* (1992) (but note their Eq. (14.3.9) is not right). Nevertheless, this formula will not be accurate in our case, and for this reason we use the symbol p_Q instead of p. The reason is that we are "optimizing" the value of α using the same sample to which we apply the KS test, which yields a bias in the test, *i.e.*, the formula would work for the true value of α, but not for one obtained by ML, which would yield in general a smaller KS statistic and too large p-values (because the fit for α_e is better than for the true value α) (Clauset *et al.* 2009, Goldstein *et al.* 2004). However, for this very same reason the formula can be useful to reject the goodness of a given fit, *i.e.*, if p_Q obtained in this way is already below 0.05, the true p will be even smaller and the fit is certainly bad. But the opposite is not true. In a formula

$$\text{if } p_Q < 0.05 \Rightarrow \text{ reject power law} , \tag{30}$$

otherwise, no decision can be taken yet. Of course, the significance level 0.05 is arbitrary and can be changed to another value, as usual in statistical tests. As a final comment, perhaps a more powerful test would be to use, instead of the KS statistic, the Kuiper's statistic (Press *et al.* 1992), which is a refinement of the former one. It is stated by Clauset *et al.* (2009) that both tests lead to very similar fits. In most cases, we have also found no significant differences between both tests.

2.6 The Clauset *et al.*'s recipe

Now we are in condition to explain the genuine Clauset *et al.*'s (2009) method. This is done in this subsection for completeness, and for the information of the reader, as we are not going to apply this method. The key to fitting a power law is neither the ML estimation of the exponent nor the goodness-of-fit test, but the selection of the interval $[a, b]$ over which the power law holds. Initially, we have taken a and b as fixed parameters, but in practice this is not the case, and one has to decide where the power law starts and where ends, independently of the total range of the data. In any case, N will be the number of data in the power-law range (and not the total number of data).

The recipe of Clauset *et al.* (2009) applies to non-truncated power-law distributions $(b \to \infty)$, and considers that a is a variable which needs to be fit from the sample (values of x below a are outside the power-law range). The recipe simply consists in the search of the value of a which yields a minimum of the KS statistic, using as a parameter of the theoretical distribution the one obtained by maximum likelihood, α_e, for the corresponding a (no calculation of a p-value is required for each fixed a). In other words,

$$a = \text{the one that yields minimum } d_e. \tag{31}$$

Next, a global p-value is computed by generating synthetic samples by a mixture of parametric bootstrap (similarly to what is explained in the next subsection) and non-parametric bootstrap. Then, the same procedure applied to the empirical data (minimization of the KS distance using ML for fitting) is applied to the syntetic samples in order to fit a and α.

These authors do not provide any explanation of why this should work, although one can argue that, if the data is indeed a power law with the desired exponent, the larger the number of data (the smaller the a-value), the smaller the value of d_e, as d_e goes as $1/\sqrt{N}$ (for large N, see previous subsection). On the other hand, if for a smaller a the data departs from the power law, this deviation should compensate and overcome the reduction in d_e due to the increase of N, yielding a larger d_e. But there is no reason to justify this overcoming.

Nevertheless, we will not use the Clauset *et al.*'s (2009) procedure for two other reasons. First, its extension to truncated power laws, although obvious, and justifiable with the same arguments, yields bad results, as the resulting values of the upper truncation cutoff, b, are highly unstable. Second, even for non-truncated distributions, it has been shown that the method fails to detect the existence of a power law for data simulated with a power-law tail (Corral *et al.* 2011): the method yields an a-value well below the true power-law region, and therefore the resulting p is too small for the power law to become acceptable. We will explain an alternative method that avoids these problems, but first let us come back to the case with a and b fixed.

2.7 Monte Carlo simulations

Remember that we are considering a power-law distribution, defined in $a \leq x \leq b$. We already have fit the distribution, by ML, and we are testing the goodness of the fit by means of the KS statistic. In order to obtain a reliable p-value for this test we will perform Monte Carlo simulations of the whole process. A synthetic sample power-law distributed and with N elements can be obtained in a straightforward way, from the inversion or transformation method (Devroye 1986, Press *et al.* 1992, Ross 2002)

$$x_i = \frac{a}{[1 - (1 - r^{\alpha_e - 1})u_i]^{1/(\alpha_e - 1)}}, \tag{32}$$

where u_i represents a uniform random number in $[0, 1)$. One can use any random number generator for it. Our results arise from "ran3" of Press *et al.* (1992).

2.8 Application of the complete procedure to many syntetic samples and calculation of p-value

The previous fitting and testing procedure is applied in exactly the same way to the synthetic sample, yielding a ML exponent α_s (where the subindex s stands from synthetic or simulated), and then a KS statistic d_s, computed as the difference between the theoretical cumulative distribution, with parameter α_s, and the simulated one, $n_s(x)/N$ (obtained from simulations with α_e, as described in the previous subsection), *i.e.*,

$$d_s = \max_{a \leq x \leq b} \left| \frac{1}{1 - r^{\alpha_s - 1}} \left[\left(\frac{a}{x}\right)^{\alpha_s - 1} - r^{\alpha_s - 1} \right] - \frac{n_s(x)}{N} \right|. \tag{33}$$

Both values of the exponent, α_e and α_s, should be close to each other, but they will not be necessarily the same. Note that we are not parameterizing $S(x)$ by the empirical value α_e, but with a new fitted value α_s. This is in order to avoid biases, as a parameterization with α_e would lead to worse fits (as the best one would be with α_s) and therefore to larger values of the resulting KS statistic and to artificially larger p-values. So, although the null hypothesis of the test is that the exponent of the power law is α_e, and synthetic samples are obtained with this value, no further knowledge of this value is used in the test. This is the procedure used by Clauset *et al.* (2009) and Malmgren *et al.* (2008), but it is not clear if it is the one of Goldstein *et al.* (2004).

In fact, one single synthetic sample is not enough to do a proper comparison with the empirical sample, and we repeat the simulation many times. The most important outcome is the set of values of the KS statistic, d_s, which allows to estimate its distribution. The p-value is simply calculated as

$$p = \frac{\text{number of simulations with } d_s \geq d_e}{N_s}, \tag{34}$$

where N_s is the number of simulations. Figure 3 shows an example of the distribution of the KS statistic for simulated data, which can be used as a table of critical values when the number of data and the exponent are the same as in the example (Goldstein *et al.* 2004).

The standard deviation of the p-value can be calculated just using that the number of simulations with $d_s \geq d_e$ is binomially distributed, with standard deviation $\sqrt{N_s p(1 - p)}$ and therefore the standard deviation of p is the same divided by N_s,

$$\sigma_p = \sqrt{\frac{p(1 - p)}{N_s}}. \tag{35}$$

In fact, the p-value in this formula should be the ideal one (the one of the whole population) but we need to replace it by the estimated value; further, when doing
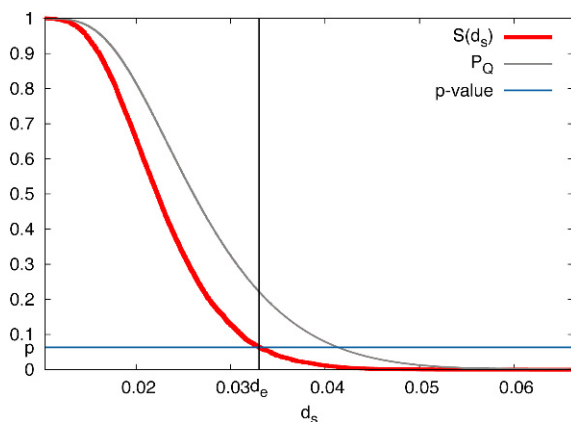
Fig. 3.   Cumulative (complementary) distribution of the Kolmogorov–Smirnov statistic for simulated non-truncated power-law distributions with $\alpha = \alpha_e = 1.143$, $a = 0.001$, and $N_{\text{tot}} = 1000$. The original "empirical" value $d_e = 0.033$ is also shown. The resulting p-value turns our to be $p = 0.060 \pm 0.008$. The "false" p-value, p_Q, arising from the KS formula, leads to higher values for the same d_e, in concrete, $p_Q = 0.22$. Colour version of this figure is available in electronic edition only.

estimation from data, N_s should be $N_s - 1$, but we have disregarded this bias correction. It will be also useful to consider the relative uncertainty of p, which is the same as the relative uncertainty of the number of simulations with $d_s \geq d_e$ (as both are proportional). Dividing the standard deviation of p by its mean (which is p), we obtain

$$CV_p = \sqrt{\frac{1-p}{pN_s}} \simeq \sqrt{\frac{1-p}{\text{number of simulations with } d_s \geq d_e}} \tag{36}$$

(we will recover this formula for the error of the estimation of the probability density).

In this way, small p-values are associated to large values of d_e, and therefore to bad fits. However, note that if we put the threshold of rejection in, let us say, $p \leq 0.05$, even true power-law distributed data, with exponent α_e, yield "bad fits" in one out of 20 samples (on average). So we are rejecting true power laws in 5% of the cases (type I error). On the other hand, lowering the threshold of rejection would reduce this problem, but would increase the probability of accepting false power laws (type II error). In this type of tests a compromise between both types of errors is always necessary, and depends on the relative costs of rejecting a true hypothesis or accepting a false one.

In addition, we can obtain from the Monte Carlo simulations the uncertainty of the ML estimator, just computing $\bar{\alpha}_s$, the average value of α_s, and from here its standard deviation,

$$\sigma = \sqrt{\overline{(\alpha_s - \bar{\alpha}_s)^2}}, \tag{37}$$

where the bars indicate average over the N_s Monte Carlo simulations. This procedure yields good agreement with the analytical formula of Aban *et al.* (2006), but can be much more useful in the discrete power-law case.

2.9 Alternative method to the one by Clauset *et al.*

At this point, for given values of the truncation points, a and b, we are able to obtain the corresponding ML estimation of the power-law exponent as well as the goodness of the fit, by means of the p-value. Now we face the same problem Clauset *et al.* (2009) tried to solve: how to select the fitting range? In our case, how to find not only the value of a but also of b? We adopt the simple method proposed by Peters *et al.* (2010): sweeping many different values of a and b we should find, if the null hypothesis is true (*i.e.*, if the sample is power-law distributed), that many sets of intervals yield acceptable fits (high enough p-values), so we need to find the "best" of such intervals. And which one is the best? For a non-truncated power law the answer is easy, we select the largest interval, *i.e.*, the one with the smaller a, provided that the p-value is above some fixed significance level p_c. All the other acceptable intervals will be inside this one.

But if the power law is truncated, the situation is not so clear, as there can be several non-overlapping intervals. In fact, many true truncated power laws can be contained in the data, at least there are well know examples of stochastic processes with double power-law distributions (Boguñá and Corral 1997, Corral 2003, 2009a, Klafter *et al.* 1996). At this point any selection can be reasonable, but if one insists in having an automatic, blind procedure, a possibility is to select either the interval which contains the larger number of data, N (Peters *et al.* 2010), or the one which has the larger log-range, b/a. For double power-law distributions, in which the exponent for small x is smaller than the one for large x, the former recipe has a tendency to select the first (small x) power-law regime, whereas the second procedure changes this tendency in some cases.

In summary, the final step of the method for truncated power-law distributions is contained in the formula

$$[a, b] = \text{the one that yields higher} \left\{ \begin{array}{c} N \\ \text{or} \\ b/a \end{array} \right\} \text{provided that } p > p_c , \qquad (38)$$

which contains in fact two procedures, one maximizing N and the other maximizing b/a. We will test both in this paper. For non-truncated power-law distributions the two procedures are equivalent.

One might be tempted to choose $p_c = 0.05$; however, it is safer to consider a larger value, as for instance $p_c = 0.20$. Note that the p-value we are using is
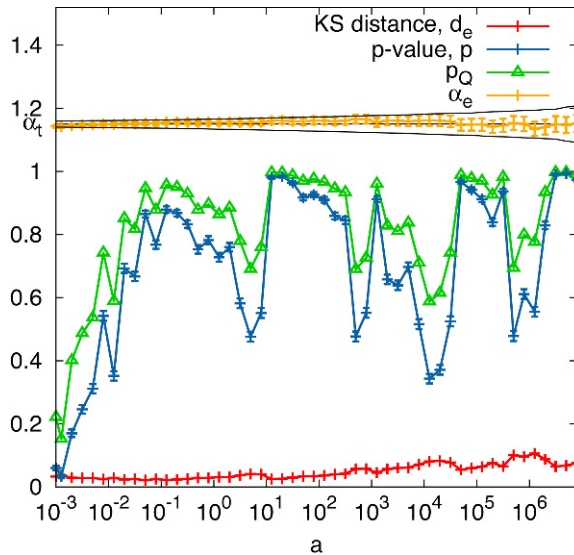
Fig. 4. Evolution as a function of a of the KS statistic, the false p-value p_Q, the true p-value (for fixed a), and the estimated exponent. The true exponent, here called α_t and equal to 1.15, is displayed as a thin black line, together with a 2σ interval. Colour version of this figure is available in electronic edition only.

the one for fixed a and b, and then the p-value of the whole procedure should be different, but at this point it is not necessary to obtain such a p-value, as we should have already come out with a reasonable fit. Figure 4 shows the results of the method for true power-law data.

2.10 Truncated or non-truncated power-law distribution?

For broadly distributed data, the simplest choice is to try to fit first a non-truncated power-law distribution. If an acceptable fit is found, it is expected that a truncated power law, with $b \geq x_{\max}$ (where $x_{\max}$ is the largest value of x) would yield also a good fit. In fact, if b is not considered as a fixed value but as a parameter to fit, its maximum likelihood estimator when the number of data is fixed, *i.e.*, when b is in the range $b \geq x_{\max}$, is $b_e = x_{\max}$. This is easy to see (Aban *et al.* 2006), just looking at the Eqs. (11) and (12) for $\ell(\alpha)$, which show that $\ell(\alpha)$ increases as b approaches $x_{\max}$. (In the same way, the ML estimator of a, for fixed number of data, would be $a_e = x_{\min}$, but we are not interested in such a case now.) On the other hand, it is reasonable that a truncated power law yields a better fit than a non-truncated one, as the former has two parameters and the latter only one (assuming that a is fixed, in any case).

In order to do a proper comparison, in such situations the so-called Akaike information criterion (AIC) can be used. This is defined simply as the difference

between twice the number of parameters and twice the maximum of the log-likelihood multiplied by the number of data, *i.e.*,

$$\text{AIC} = 2 \times (\text{number of parameters}) - 2N\ell(\alpha_e) \, . \tag{39}$$

In general, having more parameters leads to better fits, and to higher likelihood, so, the first term compensates this fact. Therefore, given two models, the one with smaller AIC is preferred. Note that, in order that the comparison based on the AIC makes sense, the fits that are compared have to be performed exactly over the same data. So, in our case this can only be done for non-truncated power laws and for truncated power laws with $b \geq x_{\max}$. Nevertheless, due to the limitations of this paper we have not performed the comparison.

3. ESTIMATION OF PROBABILITY DENSITIES AND CUMULATIVE DISTRIBUTION FUNCTIONS

The method of maximum likelihood does not rely on the estimation of the probability distributions, in contrast to other methods. Nevertheless, in order to present the results, it is useful to display some representation of the distribution, together with its fit. This procedure has no statistical value (it cannot provide a substitution of a goodness-of-fit test) but is very helpful as a visual guide, specially in order to detect bugs in the algorithms.

3.1 Estimation of the probability density

In the definition of the probability density,

$$f(x) = \lim_{\Delta x \to 0} \frac{\text{Prob}[x \leq \text{random variable} < x + \Delta x]}{\Delta x} \, , \tag{40}$$

a fundamental issue is that the width of the interval Δx has to tend to zero. In practice Δx cannot tend to zero (there would be no statistics in such case), and one has to take a non-zero value of the width. The most usual procedure is to draw a histogram using linear binning (bins of constant width); however, there is no reason why the width of the distribution should be fixed (some authors even take $\Delta x = 1$ as the only possible choice). In fact, Δx should be chosen in order to balance the necessity of having enough statistics (large Δx) with that of having a good sampling of the function (small Δx). For power-law distributions and other fat-tailed distributions, which take values across many different scales, the right choice depends of the scale of x. In these cases it is very convenient to use the so-called logarithmic binning (Hergarten 2002, Pruessner 2012). This uses bins that appear as constant in logarithmic scale, but that in fact grow exponentially (for which the method is sometimes called

exponential binning instead). Curiously, this useful method is not considered by classic texts on density estimation (Silverman 1986).

Let us consider the semi-open intervals $[a_0, b_0), [a_1, b_1), \ldots, [a_k, b_k), \ldots$, also called bins, with $a_{k+1} = b_k$ and $b_k = Ba_k$ (this constant B has nothing to do with the one in the Gutenberg–Richter law, Section 1). For instance, if $B = \sqrt[5]{10}$ this yields 5 intervals for each order of magnitude. Notice that the width of every bin grows linearly with a_k, but exponentially with k, as $b_k - a_k = (B - 1)a_k = a_0(B - 1)B^k$. The value of B should be chosen in order to avoid low populated bins; otherwise, a spurious exponent equal to one appears (Pruessner 2012).

We simply will count the number of occurrences of the variable in each bin. For each value of the random variable x_i, the corresponding bin is found as

$$k = \text{int}\left(\frac{\ln(x_i/a_0)}{\ln B}\right), \tag{41}$$

where the function int denotes the integer part of its argument. Of course, a_0 has to be smaller than any possible value of x. For a continuous variable the concrete value of a_0 should be irrelevant (if it is small enough), but in practice one has to avoid that the resulting values of a_k coincide with round values of the variable (Corral *et al.* 2011).

So, with this logarithmic binning, the probability density can be estimated (following its definition) as the relative frequency of occurrences in a given bin divided by its width, *i.e.*,

$$f_e(x_k^*) = \frac{\text{number of occurrences in bin } k}{(b_k - a_k) \times \text{number of occurrences}}, \tag{42}$$

where the estimation of the density is associated to a value of x represented by x_k^*. The most practical solution is to take it in the middle of the interval in logscale, so $x_k^* = \sqrt{a_k b_k}$. However, for sparse data covering many orders of magnitude it is necessary to be more careful. In fact, what we are looking for is the point x_k^* whose value of the density coincides with the probability of being between a_k and b_k divided by the width of the interval. This is the solution of

$$f(x_k^*) = \frac{1}{b_k - a_k}\int_{a_k}^{b_k} f(x)dx = \frac{S(a_k) - S(b_k)}{b_k - a_k}, \tag{43}$$

where f and S are the theoretical distributions. When the distribution and its parameters are known, the equation can be solved either analytically or numerically. It is easy to see that for a power-law distribution (truncated or not) the solution can be written

$$x_k^* = \sqrt{a_k b_k}\left[(\alpha - 1)\frac{B^{\alpha/2-1}(B - 1)}{B^{\alpha-1} - 1}\right]^{1/\alpha}, \tag{44}$$

where we have used that $B = b_k/a_k$ (if we were not using logarithmic binning we would have to write a bin-dependent B_k). Note that for constant (bin-independent) B, *i.e.*, for logarithmic binning, the solution is proportional but not equal to the geometric mean of the extremes of the bin. Nevertheless, the omission of the proportionality factor does not alter the power-law behavior, just shifts (in logarithmic scale) the curve. But for a different binning procedure this is no longer true. Moreover, for usual values of B the factor is very close to one (Hergarten 2002), although large values of B (Corral *et al.* 2011) yield noticeable deviations if the factor in brackets is not included; see also our treatment of the radionuclide half-lives in Section 3, with $B = 10$. Once the value of B is fixed (usually in this paper to $\sqrt[5]{10}$), in order to avoid empty bins we merge consecutive bins until the resulting merged bins are not empty. This leads to a change in the effective value of B for merged bins, but the method is still perfectly valid.

The uncertainty of $f_e(x)$ can be obtained from its standard deviation (the standard deviation of the estimation of the density, f_e, not of the original random variable x). Indeed, assuming independence in the sample (which is already implicit in order to apply maximum likelihood estimation), the number of occurrences of the variable in bin k is a binomial random variable (in the same way as for the p-value). As the number of occurrences is proportional to $f_e(x)$, the ratio between the standard deviation and the mean for the number of occurrences will be the same as for $f_e(x)$, which is,

$$\frac{\sigma_f(x)}{f_e(x)} = \sqrt{\frac{q}{\text{mean number of occurrences in } k}} \simeq \frac{1}{\sqrt{\text{occurrences in } k}} , \quad (45)$$

where we replace the mean number of occurrences in bin k (not available from a finite sample) by the actual value, and q, the probability that the occurrences are not in bin k, by one. This estimation of $\sigma_f(x)$ fails when the number of counts in the bin is too low, in particular if it is one.

One final consideration is that the fitted distributions are normalized between a and b, with N number of data, whereas the empirical distributions include all data, with N_{tot} of them, $N_{\text{tot}} \geq N$. Therefore, in order to compare the fits with the empirical distributions, we will plot $Nf(x)/N_{\text{tot}}$ together with $f_e(x_k^*)$.

3.2 Estimation of the cumulative distribution

The estimation of the (complementary) cumulative distribution is much simpler, as bins are not involved. One just needs to sort the data, in ascending order, $x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(N_{\text{tot}}-1)} \leq x_{(N_{\text{tot}})}$; then, the estimated cumulative distribution is

$$S_e(x_{(i)}) = \frac{n_e(x_{(i)})}{N_{\text{tot}}} = \frac{N_{\text{tot}} - i + 1}{N_{\text{tot}}} , \quad (46)$$

for the data points, $S_e(x) = $ constant below these data points, and $S_e(x) = 0$ for $x > x_{(N_{\text{tot}})}$; $n_e(x_{(i)})$ is the number of data with $x \geq x_{(i)}$ in the empirical sample. The formula relating $n_e(x_{(i)})$ with i assumes that repeated values of the variable are not possible, so it would not be valid for a discrete x. We use the case of empirical data as an example, but it is of course the same for simulated data. For the comparison of the empirical distribution with the theoretical fit we need to correct the different number of data in both cases. So, we plot both $[NS(x) + n_e(b)]/N_{\text{tot}}$ and $S_e(x)$, in order to check the accuracy of the fit.

4. DATA ANALYZED AND RESULTS

We have explained how, in order to certify that a dataset is compatible with a simple power-law distribution, many mathematical formulas are required, leading to an astonishingly large number of calculations. Now we check the performance of our method with diverse geophysical data, which were previously analyzed with different, less rigorous or worse-functioning methods. For the peculiarities and challenges of the dataset, we also include the half-lives of unstable nuclides. The parameters of the method are fixed to $N_s = 1000$ Monte Carlo simulations and the values of a and b are found sweeping a fixed number of points per order of magnitude, equally spaced in logarithmic scale. This number is 10 for non-truncated power laws (in which b is fixed to infinity) and 5 for truncated power laws. Three values of p_c are considered: 0.1, 0.2, and 0.5, in order to compare the dependence of the results on this parameter. The results are reported using the Kolmogorov–Smirnov test for goodness-of-fit. If, instead, the Kuiper's test is used, the outcome is not significantly different in most of the cases. In a few cases the fitting range, and therefore the exponent, changes, but without a clear trend, *i.e.*, the fitting range can become smaller or increase. These cases deserve a more in-depth investigation.

4.1 Half-lives of the radioactive elements

Corral *et al.* (2011) studied the statistics of the half-lives of radionuclides (comprising both nuclei in the fundamental and in excited states). Any radionuclide has a constant probability of disintegration per unit time, the decay constant, let us call it λ (Krane 1988). If M is the total amount of radioactive material at time t, this means that

$$-\frac{1}{M}\frac{dM}{dt} = \lambda . \tag{47}$$

This leads to an exponential decay, for which a half-life $t_{1/2}$ or a lifetime τ can be defined, as

$$t_{1/2} = \tau \ln 2 = \frac{\ln 2}{\lambda} . \tag{48}$$

It is well known that the half-lives take disparate values, for example, that of ^{238}U is 4.47 (American) billions of years, whereas for other nuclides it is a very tiny fraction of a second.

It has been recently claimed that these half-lives are power-law distributed (Corral *et al.* 2011). In fact, three power-law regions were identified in the probability density of $t_{1/2}$, roughly,

$$f(t_{1/2}) \propto \begin{cases} 1/t_{1/2}^{0.65} & \text{for } 10^{-6}\,\text{s} \leq t_{1/2} \leq 0.1\,\text{s} \\ 1/t_{1/2}^{1.19} & \text{for } 100\,\text{s} \leq t_{1/2} \leq 10^{10}\,\text{s} \\ 1/t_{1/2}^{1.09} & \text{for } t_{1/2} \geq 10^{8}\,\text{s}. \end{cases} \qquad (49)$$

Notice that there is some overlap between two of the intervals, as reported in the original reference, due to problems in delimiting the transition region. The study used variations of the Clauset *et al.*'s (2009) method of minimization of the KS statistic, introducing and upper cutoff and additional conditions to escape from the global minimum of the KS statistic, which yielded the rejection ($p = 0.000$) of the power-law hypothesis. These additional conditions were of the type of taking either a or b/a greater than a fixed amount.

For comparison, we will apply the method explained in the previous section to this problem. Obviously, our random variable will be $x = t_{1/2}$. The data is exactly the same as in the original reference, coming from the Lund/LBNL Nuclear Data Search web page (Chu *et al.* 1999). Elements whose half-life is only either bounded from below or from above are discarded for the study, which leads to 3002 radionuclides with well-defined half-lives; 2279 of them are in their ground state and the remaining 723 in an exited state. The minimum and maximum half-lives in the dataset are 3×10^{-22} s and 7×10^{31} s, respectively, yielding more than 53 orders of magnitude of logarithmic range. Further details are in Corral *et al.* (2011).

The results of our fitting and testing method are shown in Table 1 and in Fig. 5. The fitting of a non-truncated power law yields results in agreement with Corral *et al.* (2011), with $\alpha = 1.09 \pm 0.01$ and $a = 3 \times 10^{7}$ s, for the three values of p_c analyzed (0.1, 0.2, and 0.5). When fitting a truncated power law, the maximization of the log-range, b/a, yields essentially the same results as for a non-truncated power law, with slightly smaller exponents α due to the finiteness of b (results not shown). In contrast, the maximization of the number of data N yields an exponent $\alpha \simeq 0.95$ between $a \simeq 0.1$ s and $b \simeq 400$ s (with some variations depending on p_c). This result is in disagreement with Corral *et al.* (2011), which yielded a smaller exponent for smaller values of a and b. In fact, as the intervals do not overlap both results are compatible, but it is also likely that a different function would lead to a better fit; for instance, a lognormal between 0.01 s and 10^{5} s was proposed by Corral *et al.* (2011),

Results of the fits for the $N_{\text{tot}} = 3002$ nuclide half-lives data, for different values of p_c

N	a [s]	b [s]	b/a	$\alpha \pm \sigma$	p_c
143	0.316×10^8	∞	∞	1.089 ± 0.007	0.10
143	0.316×10^8	∞	∞	1.089 ± 0.007	0.20
143	0.316×10^8	∞	∞	1.089 ± 0.008	0.50
1596	0.0794	501	6310	0.952 ± 0.010	0.10
1539	0.1259	501	3981	0.959 ± 0.011	0.20
1458	0.1259	316	2512	0.950 ± 0.011	0.50
1311	125.9	0.501×10^{23}	0.398×10^{21}	1.172 ± 0.005	0.10
1309	125.9	0.316×10^{22}	0.251×10^{20}	1.175 ± 0.005	0.20
1303	125.9	0.794×10^{18}	0.631×10^{16}	1.177 ± 0.005	0.50

Notes: We show the cases of a pure or non-truncated power law (with $b = \infty$, fixed) and truncated power law (with b finite, estimated from data), maximizing N. The latter is split into two subcases: exploring the whole range of a (rows 4, 5, and 6) and restricting a to $a > 100$ s (rows 7, 8, and 9).
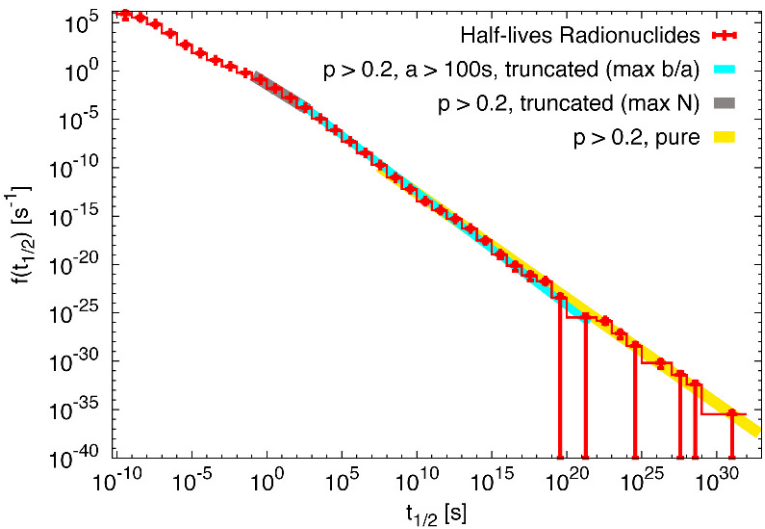

Fig. 5. Estimated probability density of the half-lives of the radionuclides, together with the power-law fits explained in the text. The number of log-bins per order of magnitude is one, which poses a challenge in the correct estimation of the density, as explained in Section 3. Data below 10^{-10} s are not shown. Colour version of this figure is available in electronic edition only.

although the fitting procedure there was not totally reliable. Finally, the intermediate power-law range reported in the original paper (the one with $\alpha = 1.19$) is not found by any of our algorithms working on the entire dataset. It is necessary to cut the dataset, removing data below, for instance, 100 s (which is equivalent to impose $a > 100$ s), in order that the algorithm converges to that solution. So, caution must be taken when applying the algorithm blindly, as important power-law regimes may be hidden by others having either larger N or larger log-range.

4.2 Seismic moment of earthquakes

The statistics of the sizes of earthquakes (Gutenberg and Richter 1944) has been investigated not only since the introduction of the first magnitude scale, by Richter, but even before, in the 1930's, by Wadati (Utsu 1999). From a modern perspective, the most reliable measure of earthquake size is given by the (scalar) seismic moment M, which is the product of the mean final slip, the rupture area, and the rigidity of the fault material (Ben-Zion 2008). It is usually assumed that the energy radiated by an earthquake is proportional to the seismic moment (Kanamori and Brodsky 2004), so, a power-law distribution of the seismic moment implies a power-law distribution of energies, with the same exponent.

The most relevant results for the distribution of seismic moment are those of Kagan for worldwide seismicity (Kagan 2002), who showed that its probability density has a power-law body, with a universal exponent in agreement with $\alpha = 1.63 \simeq 5/3$, but with an extra, non-universal exponential decay (at least in terms of the complementary cumulative distribution). However, Kagan's (2002) analysis, ending in 2000, refers to a period of global seismic "quiescence"; in particular, the large Sumatra–Andaman earthquake of 2004 and the subsequent global increase of seismic activity are not included. Much recently, Main *et al.* (2008) have shown, using a Bayesian information criterion, that the inclusion of the new events leads to the preference of the non-truncated power-law distribution in front of models with a faster large-M decay.

We take the Centroid Moment Tensor (CMT) worldwide catalog analyzed by Kagan (2002) and by Main *et al.* (2008), including now data from January 1977 to December 2010, and apply our statistical method to it. Although the statistical analysis of Kagan is rather complete, his procedure is different to ours. Note also that the dataset does not comprise the recent (2011) Tohoku earthquake in Japan; nevertheless, the qualitative change in the data with respect to Kagan's period of analysis is very remarkable. Following this author, we separate the events by their depth: shallow for depth ≤ 70 km, intermediate for 70 km $<$ depth ≤ 300 km, and deep for depth > 300 km. The number of earthquakes in each category is 26824, 5281, and 1659, respectively.

Second, we also consider the Southern California's Waveform Relocated Earthquake Catalog, from 1 January 1981 to 30 June 2011, covering a rectangular box of coordinates $(122° \text{ W}, 30° \text{ N})$, $(113° \text{ W}, 37.5° \text{ N})$ (Hauksson *et al.* 2012, Shearer *et al.* 2005). This catalog contains 111981 events with $m \geq 2$. As, in contrast with the CMT catalog, this one does not report the seismic moment M, the magnitudes m there are converted into seismic moments, using the formula

$$\log_{10} M = \frac{3}{2}(m + 6.07) \,, \tag{50}$$

where M comes in units of Nm (Newtons times meters); however, this formula is a very rough estimation of seismic moment, as it is only accurate (and exact) when m is the so-called moment magnitude (Kanamori and Brodsky 2004), whereas the magnitudes recorded in the catalog are not moment magnitudes. In any case, our procedure here is equivalent to fit an exponential distribution to the magnitudes reported in the catalog.

Tables 2 and 3 and Fig. 6 summarize the results of analyzing these data with our method, taking $x = M$ as the random variable. Starting with the non-truncated power-law distribution, we always obtain an acceptable (in the sense of non-rejectable) power-law fit, valid for several orders of magnitude. In all cases the exponent α is between 1.61 and 1.71, but for Southern California it is always very close to 1.66. For the worldwide CMT data the largest value of a is 3×10^{18} Nm, corresponding to a magnitude $m = 6.25$ (for shallow depth), and the smallest is $a = 8 \times 10^{16}$ Nm, corresponding to $m = 5.2$ (intermediate depth). If the events are not separated in terms of their depth (not shown), the results are dominated by the shallow case, except for $p_c = 0.5$, which leads to very large values of a and α ($a = 5 \times 10^{20}$ Nm and $\alpha \simeq 2$). The reason is probably the mixture of the different populations, in terms of depth, which is not recommended by Kagan (2002). This is an indication that the inclusion of an upper limit b to the power law may be appropriate, with each depth corresponding to different b's. For Southern California, the largest a found (for $p_c = 0.5$) is 1.6×10^{15} Nm, giving $m = 4$. This value is somewhat higher, in comparison with the completeness magnitude of the catalog; perhaps the reason that the power-law fit is rejected for smaller magnitudes is due to the fact that these magnitudes are not true moment magnitudes, but come from a mixture of different magnitude definitions. If the value of a is increased, the number of data N is decreased and the power-law hypothesis is more difficult to reject, due simply to poorer statistics. When a truncated power law is fitted, using the method of maximizing the number of data leads to similar values of the exponents, although the range of the fit is in some cases moved to smaller values (smaller a, and b smaller than the maximum M on the dataset). The method of maximizing b/a leads to results that are very close to the non-truncated power law.

Results of the non-truncated power-law fit $(b = \infty)$ applied to
the seismic moment of earthquakes in CMT worldwide catalog (separating by depth)
and to the Southern California catalog, for different p_c

| Catalog	N	a [Nm]	$\alpha \pm \sigma$	p_c
CMT deep	1216	0.1259×10^{18}	1.622 ± 0.019	0.10
intermediate	3701	0.7943×10^{17}	1.654 ± 0.011	0.10
shallow	5799	0.5012×10^{18}	1.681 ± 0.009	0.10
CMT deep	898	0.1995×10^{18}	1.608 ± 0.020	0.20
intermediate	3701	0.7943×10^{17}	1.654 ± 0.011	0.20
shallow	5799	0.5012×10^{18}	1.681 ± 0.009	0.20
CMT deep	898	0.1995×10^{18}	1.608 ± 0.021	0.50
intermediate	3701	0.7943×10^{17}	1.654 ± 0.011	0.50
shallow	1689	0.3162×10^{19}	1.706 ± 0.018	0.50
S. California	1327	0.1000×10^{16}	1.660 ± 0.018	0.10
S. California	1327	0.1000×10^{16}	1.660 ± 0.018	0.20
S. California	972	0.1585×10^{16}	1.654 ± 0.021	0.50
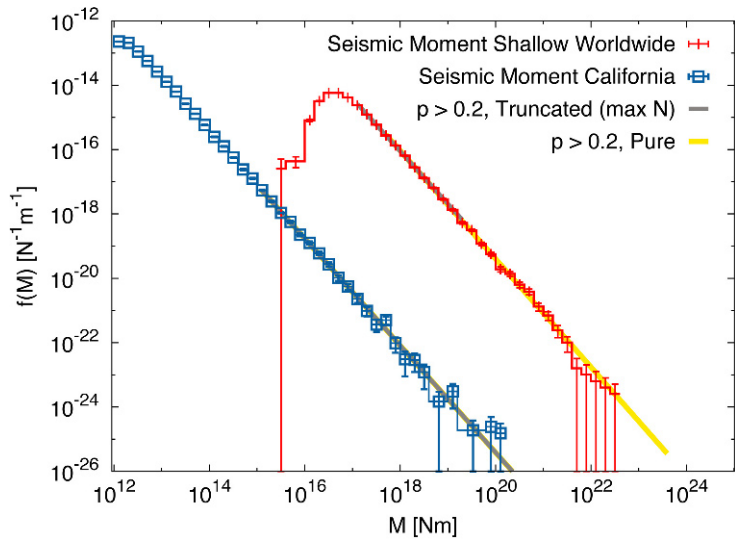
Fig. 6. Estimated probability densities and corresponding power-law fits of the seismic moment M of shallow earthquakes in the worldwide CMT catalog and of the estimated M in the Southern California catalog. Colour version of this figure is available in electronic edition only.

Results of the truncated power-law fit, by maximizing the number of data, applied to the seismic moment of earthquakes in CMT worldwide catalog (separating by depth) and to the Southern California catalog, for the different p_c

Catalog	N	a [Nm]	b [Nm]	b/a	$\alpha \pm \sigma$	p_c
CMT deep	1216	0.1259×10^{18}	0.3162×10^{23}	0.2512×10^{6}	1.621 ± 0.019	0.10
intermediate	3701	0.7943×10^{17}	0.7943×10^{22}	0.1000×10^{6}	1.655 ± 0.011	0.10
shallow	13740	0.1259×10^{18}	0.5012×10^{20}	0.3981×10^{3}	1.642 ± 0.007	0.10
CMT deep	1076	0.1259×10^{18}	0.5012×10^{19}	0.3981×10^{2}	1.674 ± 0.033	0.20
intermediate	3701	0.7943×10^{17}	0.7943×10^{22}	0.1000×10^{6}	1.655 ± 0.011	0.20
shallow	13518	0.1259×10^{18}	0.1995×10^{20}	0.1585×10^{3}	1.636 ± 0.008	0.20
CMT deep	898	0.1995×10^{18}	0.3162×10^{23}	0.1585×10^{6}	1.604 ± 0.021	0.50
intermediate	3701	0.7943×10^{17}	0.7943×10^{22}	0.1000×10^{6}	1.655 ± 0.011	0.50
shallow	11727	0.1259×10^{18}	0.1995×10^{19}	0.1585×10^{2}	1.608 ± 0.012	0.50
S. California	1146	0.1259×10^{16}	0.1259×10^{22}	0.1000×10^{7}	1.663 ± 0.020	0.10
S. California	1146	0.1259×10^{16}	0.1259×10^{22}	0.1000×10^{7}	1.663 ± 0.020	0.20
S. California	344	0.7943×10^{16}	0.1259×10^{22}	0.1585×10^{6}	1.664 ± 0.036	0.50

Note: Except for global shallow seismicity (and for the global deep case with $p_c = 0.20$) the selected b is larger than the maximum value of the variable.

4.3 Energy of tropical cyclones

Tropical cyclones are devastating atmospheric-oceanic phenomena comprising tropical depressions, tropical storms, and hurricanes or typhoons (Emanuel 2005a). Although the counts of events every year have been monitored for a long time, and other measurements to evaluate annual activity have been introduced (see Corral and Turiel (2012) for an overview), little attention has been paid to the statistics of individual tropical cyclones.

In 2005, Emanuel introduced the power-dissipation index (PDI) as a simple way to obtain a rough estimation of the total energy dissipated by all tropical cyclones in a given season and some ocean basin (Emanuel 2005b). But the PDI can also be used to characterize individual events as well, as it was done later by Corral *et al.* (2010). Indeed, the PDI is defined as the sum for all the discrete times t (that comprise the lifetime of a tropical cyclone) of the cube of the maximum sustained wind speed multiplied by the time interval of sampling, Δt. In a formula,

$$\text{PDI} = \sum_{\forall t} v_t^3 \Delta t \,, \tag{51}$$

where v_t is the maximum sustained wind speed. In the so-called best-track records, $\Delta t = 6$ hours; this factor would only be necessary in order to compare

with other data with different Δt (but caution should be present in this case for the possible fractal nature of the speed signal). Although the speeds are reported in knots, they are converted to m/s (using that 1 knot $= 0.514$ m/s), and then we report the PDI in m^3/s^2.

Corral *et al.* (2010) studied the statistics of the PDI (defined for individual events, in contrast to Emanuel's (2005b) work) in 4 different ocean basins for several time periods. The results showed a rapid, perhaps exponential, decay at the tail, but a body of the distribution compatible with a power law, for 1 or 2 orders of magnitude, with exponents close to one. The connection with SOC phenomena was discussed by Corral (2010). The method used was again a variation of the Clauset *et al.*'s (2009) one, introducing an upper cutoff and additional restrictions to the variations of the parameters. Here we revisit this problem, trying to use updated data (whenever it has been possible), and applying the method which is the subject of this paper to $x =$ PDI.

The data has been downloaded from the National Hurricane Center (NHC) of NOAA, for the North Atlantic and the Northeastern Pacific (Jarvinen *et al.* 1988, NHC 2012) and from the Joint Typhoon Warning Center (JTWC) of the US Navy (Chu *et al.* 2002, JTWC Annual Tropical Cyclone Reports 2012) for the Northwestern Pacific, the Southern Hemisphere (comprising the Southern Indian and the Southwestern Pacific), and the Northern Indian Ocean. The abbreviation, time span, and number of events for each basin are: NAtl, 1966-2011, 532; EPac, 1966-2011, 728; WPac, 1986-2011, 690; SHem, 1986-2007 (up to May), 523; NInd, 1986-2007, 110. The latter case was not studied in any of the previous works.

The results for a truncated power law maximizing N, shown in Table 4 and Fig. 7, are in agreement with those of Corral *et al.* (2010). In general, exponents are close but above 1, except for the Northwestern Pacific, where $\alpha \simeq 0.96$, and for the North Indian Ocean, where α is substantially higher than one. We consider that this method performs rather well. It would be interesting to test if universality can nevertheless hold (the high value for the North Indian Ocean is based in much less data than for the rest of basins), or if there is some systematic bias in the value of the exponents (the protocols of the NHC and the JTWC are different, and the satellite coverage of each basin is also different).

If a non-truncated power law is fit to the data, the fits turn out to be rather short, with a high exponent (up to 6) describing the tail of the distribution (except for the Southern Hemisphere, where no such tail is apparent). We do not give any relevance to these results, as other alternatives, as for instance a simple exponential tail, have to be considered (Corral and Turiel 2012, del Castillo *et al.* 2012). Coming back to a truncated power law, but maximizing the log-range, the algorithm sometimes fits the power law in the body of the distribution (with exponent close to 1) and for some other times the algorithm goes to the fast-decaying tail. So the method of maximizing b/a is not useful for this data.
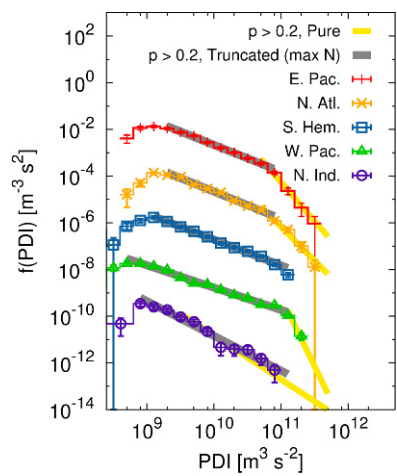
Fig. 7. Estimated probability densities of the PDI of tropical cyclones in 5 ocean basins, together with their power-law fits. The values of the densities are multiplied by 1, 10^2, 10^4, 10^6, and 10^8, for clarity sake. The fits for the non-truncated case are also displayed, although they are not tabulated. Colour version of this figure is available in electronic edition only.

Table 4

Results of the truncated power-law fits, maximizing N, for the PDI in the 5 ocean basins with tropical-cyclone activity, for different values of p_c

Basin	N	a [m³/s²]	b [m³/s²]	b/a	$\alpha \pm \sigma$	p_c
EPac	637	0.1259×10^{10}	0.7943×10^{11}	63	1.094 ± 0.033	0.10
NAtl	417	0.1995×10^{10}	0.7943×10^{11}	40	1.168 ± 0.047	0.10
SHem	523	0.1259×10^{10}	0.1259×10^{12}	100	1.108 ± 0.034	0.10
WPac	637	0.5012×10^{9}	0.1259×10^{12}	251	0.957 ± 0.025	0.10
NInd	102	0.7943×10^{9}	0.1995×10^{12}	251	1.520 ± 0.077	0.10
EPac	571	0.1995×10^{10}	0.7943×10^{11}	40	1.149 ± 0.039	0.20
NAtl	417	0.1995×10^{10}	0.7943×10^{11}	40	1.168 ± 0.047	0.20
SHem	523	0.1259×10^{10}	0.1259×10^{12}	100	1.108 ± 0.033	0.20
WPac	637	0.5012×10^{9}	0.1259×10^{12}	251	0.957 ± 0.025	0.20
NInd	102	0.7943×10^{9}	0.1259×10^{12}	158	1.490 ± 0.077	0.20
EPac	571	0.1995×10^{10}	0.7943×10^{11}	40	1.149 ± 0.040	0.50
NAtl	417	0.1995×10^{10}	0.7943×10^{11}	40	1.168 ± 0.045	0.50
SHem	465	0.1995×10^{10}	0.1259×10^{12}	63	1.132 ± 0.040	0.50
WPac	637	0.5012×10^{9}	0.1259×10^{12}	251	0.957 ± 0.024	0.50
NInd	86	0.7943×10^{9}	0.1259×10^{11}	16	1.323 ± 0.139	0.50

4.4 Area of forest fires

The statistics of the size of forest fires was an intense topic of research since the introduction of the concept of SOC, at the end of the 1980's, but only from the point of view of cellular-automaton models. Real data analysis had to wait several years (Malamud *et al.* 1998, 2005), leading to power-law distributions, more or less in agreement with the models. Here we are particularly interested in a dataset from Italy, for which a power-law distribution of sizes was ruled out (Corral *et al.* 2008). Instead, a lognormal tail was proposed for the fire-size probability density.

   The data considered by Corral *et al.* (2008), and reanalyzed in this study, comes from the Archivio Incendi Boschivi (AIB) fire catalog compiled by the (Italian) Corpo Forestale dello Stato (2012). The subcatalog to which we restrict covers all Italy and spans the 5-year period 1998-2002, containing 36 748 fires. The size of each fire is measured by the burned area A, in hectares, with $1 \, \mathrm{ha} = 10^4 \, \mathrm{m}^2$. In this subsection we analyze the case of $x = A$.

   The results in Table 5 and Fig. 8 show that a pure (non-truncated) power law is only acceptable (in the sense of non-rejectable) for the rightmost part of the tail of the distribution, comprising less than one order of magnitude. It is very indicative that only 51 data are in the possible power-law tail. Therefore, we disregard this power-law behavior as spurious and expect that other distributions can yield a much better fit (not in order of the quality of the fit but regarding the number of data it spans). This seems in agreement with other analyses of forest-fire data (Clauset *et al.* 2009, Newman 2005). If a truncated power-law is considered, fitted by maximizing the number of data, the results are not clearly
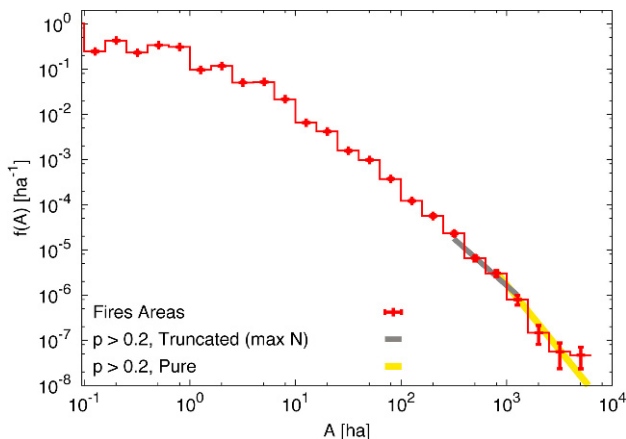

Fig. 8.   Estimated probability density of the area of fires in the Italian catalog, together with the power-law fits. In contrast to the previous datasets analyzed, we consider these power-law fits as irrelevant. Colour version of this figure is available in electronic edition only.

Table 5

Results of the fits for the burned area of the $N_{\text{tot}} = 36\,748$ fires
recorded in the Italian catalog, for different p_c

N	a [ha]	b [ha]	b/a	$\alpha \pm \sigma$	p_c
51	794	∞	∞	2.880 ± 0.276	0.10
51	794	∞	∞	2.880 ± 0.277	0.20
51	794	∞	∞	2.880 ± 0.277	0.50
168	316	7943	25	2.353 ± 0.117	0.10
148	316	1259	4	2.075 ± 0.213	0.20
51	794	79430	100	2.870 ± 0.281	0.50

Note: The cases of a non-truncated power law and a truncated
power law, maximizing N, are shown. In the latter case, for
$p_c = 0.10$ and 0.50 the value of b is larger than the maximun
value of the variable.

better, as seen in the table. Moreover, there is considerable variation with the
value of p_c. So, we do not give any relevance to such power-law fits. Finally,
the method of maximizing b/a yields the same results as for the non-truncated
power law (except by the fact that the exponents are slightly smaller, not shown).
In order to provide some evidence for the better adequacy of the lognormal tail
in front of the power-law tail for these data, it would be interesting to apply an
adaptation of the test explained by del Castillo and Puig (1999).

4.5 Waiting time between earthquakes

The temporal properties of earthquakes have been a subject relatively little stud-
ied (at least in comparison with the size of earthquakes). It is true that the Omori
law has been known for more than 100 years (Utsu 2002, Utsu *et al.* 1995), and
that this is a law extremely important in order to assess the hazard of after-
shocks after a big event, but the Omori law looks at time properties in a very
coarse-grained way, as it only provides the number of events in relatively large
time windows. Thus, no information on the fine time structure of seismicity is
provided, at least directly.

The situation has changed in the last decade, since the seminal study of Bak
et al. (2002), who found a unified scaling law for earthquake waiting-time dis-
tributions. They took Southern California and divided it in different areas, and
computed the time between consecutive earthquakes for each area. So, if t_i^j de-
notes the time of occurrence of the i-th earthquake in area j, the corresponding
waiting time τ_i^j is

$$\tau_i^j = t_i^j - t_{i-1}^j \, . \tag{52}$$

The key point is that all the resulting waiting times were added to the same distribution (and not to a different distribution j for each area). Subsequently, the unified scaling law was found to be valid for other regions of the world (Corral 2004b). The shape of the resulting probability density corresponds to a double power law, one for small waiting times, associated to a mixture of scales of waiting times due to the Omori law, and another for large waiting times, due to spatial heterogeneity arising from the mixture of different areas with different seismic rates (Bak *et al.* 2002, Corral 2003, 2004b, Corral and Christensen 2006). The first exponent was found to be close to 1, whereas the second one was about 2.2; the fits were done by means of the nonlinear least-squares Marquardt–Levenberg algorithm from "gnuplot", applied to the logarithm of the log-binned empirical density. Here we apply our more sophisticated method to updated data for Southern California seismicity, with $x = \tau$.

We use again the relocated Southern California catalog of Hauksson *et al.* (2012), see also Shearer *et al.* (2005), but starting in 1984 and ending in 30 June 2011. This is to avoid some holes in the catalog for the preceding years. As for earthquake sizes, the occurrence takes place in a rectangle of coordinates $(122° \text{W}, 30°\text{N})$, $(113° \text{W}, 37.5°\text{N})$. This rectangle is divided into equal parts both in the West-East axis and in the South-North axis, in such a way that we consider a number of subdivisions of 4×4, 8×8, 16×16, and 32×32. The waiting times for events of magnitude $m \geq 2$ in each of these subdivisions are computed as explained above, resulting in a number of data between $103\,000$ and $104\,000$ in all cases.

For a non-truncated power law, the results are only coherent with the previous reported ones (exponent around 2.2) for the intermediate cases, *i.e.*, 8×8 and 16×16, see Table 6 and Fig. 9. The disagreement for the other cases can easily be explained. For 4×4, the number of resulting subdivisions, 16, seems rather small. As mentioned, in Corral and Christensen (2006) the power-law tail was explained in terms of a power-law mixture of exponentials; so, with only 16 regions it is possible that the asymptotic behavior is still not reached. On the other hand, the effect of the finite duration of the catalog is visible in the 32×32 data. Due to the scaling behavior of the distributions (Corral 2003, 2004b), the possible power-law tail in this case is displaced to larger waiting times; but the time span of the catalog, about 10^{10} s, clearly alters this power law, which starts to bend at about 10^9 s. Thus, we conclude that a power-law exponent of about $\alpha \simeq 2.2$ or 2.3 indeed exists, provided that the number of spatial subdivisions is high enough and the temporal extension of the catalog is large enough.

When a truncated power-law is fitted, using the method of maximizing the number of data N, the other power law emerges, but for a range shorter than what the plot of the densities suggests. The exponent is in a range from 0.95 to 1.04 (except for the 4×4 cases, in which it is a bit smaller). The largest

Results of the fits with a non-truncated power law and a truncated power law,
maximizing N, for earthquake waiting times
calculated for different subdivisions of Southern California

Subdivisions	N	a [s]	b [s]	b/a	$\alpha \pm \sigma$	p_c
4×4	124	0.5012×10^7	∞	∞	1.921 ± 0.085	0.10
8×8	1671	0.3162×10^7	∞	∞	2.198 ± 0.031	0.10
16×16	542	0.3162×10^8	∞	∞	2.324 ± 0.056	0.10
32×32	67	0.3162×10^9	∞	∞	4.404 ± 0.405	0.10
4×4	124	0.5012×10^7	∞	∞	1.921 ± 0.085	0.20
8×8	1671	0.3162×10^7	∞	∞	2.198 ± 0.031	0.20
16×16	542	0.3162×10^8	∞	∞	2.324 ± 0.056	0.20
32×32	67	0.3162×10^9	∞	∞	4.404 ± 0.403	0.20
4×4	77	0.7943×10^7	∞	∞	1.856 ± 0.098	0.50
8×8	322	0.1259×10^8	∞	∞	2.231 ± 0.070	0.50
16×16	24	0.3162×10^9	∞	∞	4.106 ± 0.703	0.50
32×32	67	0.3162×10^9	∞	∞	4.404 ± 0.449	0.50
4×4	38765	1995	0.5012×10^5	25	0.867 ± 0.006	0.10
8×8	39851	316	0.1995×10^5	63	0.987 ± 0.004	0.10
16×16	44178	7943	0.7943×10^6	100	0.956 ± 0.004	0.10
32×32	43512	1259	0.1995×10^6	158	1.029 ± 0.003	0.10
4×4	38765	1995	0.5012×10^5	25	0.867 ± 0.006	0.20
8×8	39851	316	0.1995×10^5	63	0.987 ± 0.004	0.20
16×16	39481	7943	0.5012×10^6	63	0.950 ± 0.005	0.20
32×32	39654	1259	0.1259×10^6	100	1.033 ± 0.004	0.20
4×4	34113	3162	0.5012×10^5	16	0.864 ± 0.007	0.50
8×8	39851	316	0.1995×10^5	63	0.987 ± 0.004	0.50
16×16	39481	7943	0.5012×10^6	63	0.950 ± 0.005	0.50
32×32	39654	1259	0.1259×10^6	100	1.033 ± 0.004	0.50

Note: Different minimum p-values are shown. The total number of data is above
103 000 in any case.

log-range is 100, *i.e.*, two decades. The graphical representation of the density
seems to indicate that the possible power law is influenced by the effect of two
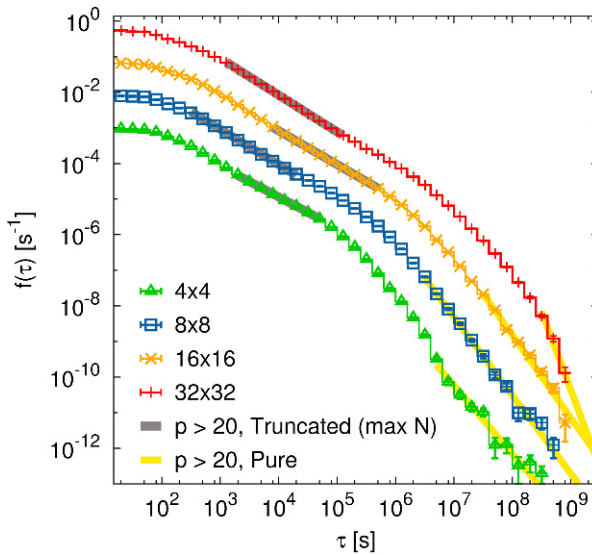crossovers, one for large waiting times, associated to a change in exponent, and

Fig. 9.   Estimated probability densities and corresponding power-law fits for the wait-ing times of $m \geq 2$ in the Southern California catalog, for different spatial subdivi-sions. The values of the density are multiplied by factors 1, 10, 100, and 1000, for clarity sake. Colour version of this figure is available in electronic edition only.

another one for smaller times, where the distribution becomes flat. Finally, the method of fitting which maximizes the log-range leads to results that are similar to the non-truncated power-law case, although sometimes intervals correspond-ing to very small times are selected. The latter results have no physical meaning, as correspond to times below 1 s, *i.e.*, below the error in the determination of the occurrence time.

5. CONCLUSIONS

For power-law distributions, the fitting and the testing of the goodness of the fit is a difficult but very relevant problem in complex-systems science, in general, and in geoscience in particular. The most critical step is to select, automati-cally (without introducing any subjective bias), where the power-law regime starts and where it ends. We have explained in detail a conceptually simple but somewhat laborious procedure in order to overcome some difficulties previ-ously found in the method introduced by Clauset *et al.* (2009); see Corral *et al.* (2011). Our method is summarized in fitting by maximum likelihood estima-tion and testing the goodness of fit by the Kolmogorov–Smirnov statistic, using Monte Carlo simulations. Although these steps are in common with the Clauset *et al.*'s (2009) recipe, the key difference is in the criterion of selection of the

power-law range. Despite the many steps of these procedures, ours can be easily implemented, and the resulting algorithms run very fast in current computers. We also have explained how to estimate properly the probability density of a random variable which has a power law or a fat-tail distribution. This is important to draw graphical representations of the results of the fitting (specially in Fig. 5) but it is not required to perform neither the fits nor the goodness-of-fit tests.

The performance of the method is quite good, as checked in synthetic power-law datasets, and the results of the analysis of previously reported power laws are very consistent. We confirm a very broad power-law tail in the distribution of the half-lives of the radionuclides, with exponent $\alpha = 1.09$, as well as other power-law regimes in the body of the distribution. The results for the power-law exponent of the distribution of seismic moments worldwide and in Southern California are in agreement with previous estimates, but in addition our method provides a reliable way to determining the minimum seismic-moment value for which the Gutenberg–Richter law holds. This can be useful to check systematically for the completeness thresholds of seismic catalogs. For the energy dissipated by tropical cyclones, measured roughly through the PDI, we confirm the power-law behavior in the body of the distribution previously reported, with exponents close to one. We also survey new results for the Southern Indian Ocean, but with a higher power-law exponent. In contrast, for the case of the area affected by forest fires in an Italian catalog, we obtain power-law-distributed behavior only for rather small windows of the burnt area, containing a very few number of fires. Finally, for the waiting times between earthquakes in different subdivisions of Southern California we conclude that the power-law behavior of the tail is very delicate, affected either by a small number of subdivisions, when the size of those is large, or by the finite duration of the record, which introduces a sharp decay of the distribution when the number of subdivisions is high. For the body of the distribution another power law is found, but the range is limited by crossovers below and above it. We conclude that, in general, the method for truncated power laws works better when the number of data in the power-law range is maximized. When the quantity that is maximized is the logarithmic range (b/a), the fitting range can jump between different regimes. Also, the selection of a p-value above 0.50 seems too strict sometimes, and values about 0.10 or 0.20 are more useful. Naturally, the methods studied in this paper can be directly applied to the overwhelming number of fat-tailed distributions reported during the last decades in geoscience.

References

Aban, I.B., M.M. Meerschaert, and A.K. Panorska (2006), Parameter estimation for the truncated Pareto distribution, *J. Am. Stat. Assoc.* **101**, 473, 270-277, DOI: 10.1198/016214505000000411.

Aschwanden, M.J. (2013), SOC systems in astrophysics. **In**: M.J. Aschwanden (ed.), *Self-Organized Criticality Systems*, Open Academic Press, Berlin, 439-478.

Baiesi, M., M. Paczuski, and A.L. Stella (2006), Intensity thresholds and the statistics of the temporal occurrence of solar flares, *Phys. Rev. Lett.* **96**, 5, 051103, DOI: 10.1103/PhysRevLett.96.051103.

Bak, P. (1996), *How Nature Works: The Science of Self-Organized Criticality*, Copernicus, New York.

Bak, P., K. Christensen, L. Danon, and T. Scanlon (2002), Unified scaling law for earthquakes, *Phys. Rev. Lett.* **88**, 17, 178501, DOI: 10.1103/PhysRevLett. 88.178501.

Barndorff-Nielsen, O. (1978), *Information and Exponential Families in Statistical Theory*, John Wiley & Sons Inc., New York, 238 pp.

Baró, J., and E. Vives (2012), Analysis of power-law exponents by maximum-likelihood maps, *Phys. Rev. E* **85**, 6, 066121, DOI: 10.1103/PhysRevE.85.066121.

Bauke, H. (2007), Parameter estimation for power-law distributions by maximum likelihood methods, *Eur. Phys. J. B* **58**, 2, 167-173, DOI: 10.1140/epjb/e2007-00219-y.

Ben-Zion, Y. (2008), Collective behavior of earthquakes and faults: continuum-discrete transitions, progressive evolutionary changes, and different dynamic regimes, *Rev. Geophys.* **46**, 4, RG4006, DOI: 10.1029/2008RG000260.

Boffetta, G., V. Carbone, P. Giuliani, P. Veltri, and A. Vulpiani (1999), Power laws in solar flares: Self-organized criticality or turbulence? *Phys. Rev. Lett.* **83**, 22, 4662-4665, DOI: 10.1103/PhysRevLett.83.4662.

Boguñá, M., and A. Corral (1997), Long-tailed trapping times and Lévy flights in a self-organized critical granular system, *Phys. Rev. Lett.* **78**, 26, 4950-4953, DOI: 10.1103/PhysRevLett.78.4950.

Bouchaud, J.-P., and A. Georges (1990), Anomalous diffusion in disordered media: statistical mechanisms, models and physical applications, *Phys. Rep.* **195**, 4-5, 127-293, DOI: 10.1016/0370-1573(90)90099-N.

Burroughs, S.M., and S.F. Tebbens (2001), Upper-truncated power laws in natural systems, *Pure Appl. Geophys.* **158**, 4, 741-757, DOI: 10.1007/PL00001202.

Burroughs, S.M., and S.F. Tebbens (2005), Power-law scaling and probabilistic forecasting of tsunami runup heights, *Pure Appl. Geophys.* **162**, 2, 331-342, DOI: 10.1007/s00024-004-2603-5.

Carrillo-Menéndez, S., and A. Suárez (2012), Robust quantification of the exposure to operational risk: Bringing economic sense to economic capital, *Comput. Oper. Res.* **39**, 4, 792-804, DOI: 10.1016/j.cor.2010.10.001.

Casella, G., and R.L. Berger (2002), *Statistical Inference*, 2nd ed., Duxbury Advanced Series, Duxbury Thomson Learning, Pacific Grove, 660 pp.

Chicheportiche, R., and J.-P. Bouchaud (2012), Weighted Kolmogorov–Smirnov test: Accounting for the tails, *Phys. Rev. E* **86**, 4, 041115, DOI: 10.1103/PhysRevE.86.041115.

Christensen, K., and N.R. Moloney (2005), *Complexity and Criticality*, Imperial College Press Advanced Physics Texts, Vol. 1, Imperial College Press, London.

Chu, J.-H., C.R. Sampson, A.S. Levine, and E. Fukada (2002), The Joint Typhoon Warning Center tropical cyclone best-tracks, 1945-2000, Naval Research Laboratory, Tech. Rep. NRL/MR/7540-02-16.

Chu, S.Y.F., L.P. Ekström, and R.B. Firestone (1999), The Lund/LBNL Nuclear Data Search, Version 2.

Clauset, A., C.R. Shalizi, and M.E.J. Newman (2009), Power-law distributions in empirical data, *SIAM Rev.* **51**, 4, 661-703, DOI: 10.1137/070710111.

Corpo Forestale dello Stato (2012), http://www.corpoforestale.it.

Corral, A. (2003), Local distributions and rate fluctuations in a unified scaling law for earthquakes, *Phys. Rev. E* **68**, 3, 035102, DOI: 10.1103/PhysRevE.68.035102.

Corral, A. (2004a), Long-term clustering, scaling, and universality in the temporal occurrence of earthquakes, *Phys. Rev. Lett.* **92**, 10, 108501, DOI: 10.1103/PhysRevLett.92.108501.

Corral, A. (2004b), Universal local versus unified global scaling laws in the statistics of seismicity, *Physica A* **340**, 4, 590-597, DOI: 10.1016/j.physa.2004.05.010.

Corral, A. (2005), Comment on "Do earthquakes exhibit self-organized criticality?", *Phys. Rev. Lett.* **95**, 15, 159801, DOI: 10.1103/PhysRevLett.95.159801.

Corral, A. (2006), Universal earthquake-occurrence jumps, correlations with time, and anomalous diffusion, *Phys. Rev. Lett.* **97**, 17, 178501, DOI: 10.1103/PhysRevLett.97.178501.

Corral, A. (2008), Scaling and universality in the dynamics of seismic occurrence and beyond. **In**: A. Carpinteri and G. Lacidogna (eds.), *Acoustic Emission and Critical Phenomena*, Taylor and Francis, London, 225-244.

Corral, A. (2009a), Point-occurrence self-similarity in crackling-noise systems and in other complex systems, *J. Stat. Mech.* P01022, DOI: 10.1088/1742-5468/2009/01/P01022 .

Corral, A. (2009b), Statistical tests for scaling in the inter-event times of earthquakes in California, *Int. J. Mod. Phys. B* **23**, 28-29, 5570-5582, DOI: 10.1142/S0217979209063869.

Corral, A. (2010), Tropical cyclones as a critical phenomenon. **In**: J.B. Elsner, R.E. Hodges, J.C. Malmstadt, and K.N. Scheitlin (eds.), *Hurricanes and Climate Change*, Vol. 2, Springer, Heidelberg, 81-99, DOI: 10.1007/978-90-481-9510-7_5.

Corral, A., and K. Christensen (2006), Comment on "Earthquakes descaled: On waiting time distributions and scaling laws", *Phys. Rev. Lett.* **96**, 10, 109801, DOI: 10.1103/PhysRevLett.96.109801.

Corral, A., and F. Font-Clos (2013), Criticality and self-organization in branching processes: application to natural hazards. **In**: M. Aschwanden (ed.), *Self-Organized Criticality Systems*, Open Academic Press, Berlin, 183-228.

Corral, A., and A. Turiel (2012), Variability of North Atlantic hurricanes: seasonal versus individual-event features. **In**: A.S. Sharma, A. Bunde, V.P. Dimri, and D.N. Baker (eds.), *Extreme Events and Natural Hazards: the Complexity Perspective*, Geopress, Washington, 111-125, DOI: 10.1029/2011GM001069.

Corral, A., L. Telesca, and R. Lasaponara (2008), Scaling and correlations in the dynamics of forest-fire occurrence, *Phys. Rev. E* **77**, 1 016101, DOI: 10.1103/PhysRevE.77.016101.

Corral, A., A. Ossó, and J.E. Llebot (2010), Scaling of tropical-cyclone dissipation, *Nature Phys.* **6**, 693-696, DOI: 10.1038/nphys1725.

Corral, A., F. Font, and J. Camacho (2011), Noncharacteristic half-lives in radioactive decay, *Phys. Rev. E* **83**, 6, 066103, DOI: 10.1103/PhysRevE.83.066103.

Corral, A., A. Deluca, and R. Ferrer-i-Cancho (2012), A practical recipe to fit discrete power-law distributions, arXiv:1209.1270.

Czechowski, Z. (2003), The privilege as the cause of power distributions in geophysics, *Geophys. J. Int.* **154**, 3, 754-766, DOI: 10.1046/j.1365-246X.2003.01994.x.

Davidsen, J., and M. Paczuski (2005), Analysis of the spatial distribution between successive earthquakes, *Phys. Rev. Lett.* **94**, 4, 048501, DOI: 10.1103/PhysRevLett.94.048501.

del Castillo, J. (2013), Exponential models, Lecture notes (unpublished).

del Castillo, J., and P. Puig (1999), The best test of exponentiality against singly truncated normal alternatives, *J. Am. Stat. Assoc.* **94**, 446, 529-532, DOI: 10.1080/01621459.1999.10474147 .

del Castillo, J., J. Daoudi, and I. Serra (2012), The full-tails gamma distribution applied to model extreme values, arXiv:1211.0130.

Devroye, L. (1986), *Non-Uniform Random Variate Generation*, Springer-Verlag, New York.

Dickman, R. (2003), Rain, power laws, and advection, *Phys. Rev. Lett.* **90**, 10, 108701, DOI: 10.1103/PhysRevLett.90.108701 .

Durrett, R. (2010), *Probability: Theory and Examples*, 4th ed., Cambridge University Press, Cambridge.

Emanuel, K. (2005a), *Divine Wind: the History and Science of Hurricanes*, Oxford University Press, New York.

Emanuel, K. (2005b), Increasing destructiveness of tropical cyclones over the past 30 years, *Nature* **436**, 7051, 686-688, DOI: 10.1038/nature03906.

Evans, M., N. Hastings, and B. Peacock (2000), *Statistical Distributions*, 3rd ed., John Wiley & Sons Inc., New York.

Felzer, K.R., and E.E. Brodsky (2006), Decay of aftershock density with distance indicates triggering by dynamic stress, *Nature* **441**, 7094, 735-738, DOI: 10.1038/nature04799 .

Freeman, M.P., and N.W. Watkins (2002), The heavens in a pile of sand, *Science* **298**, 5595, 979-980, DOI: 10.1126/science.1075555.

Geist, E.L., and T. Parsons (2008), Distribution of tsunami interevent times, *Geophys. Res. Lett.* **35**, 2, L02612, DOI: 10.1029/2007GL032690.

Goldstein, M.L., S.A. Morris, and G.G. Yen (2004), Problems with fitting to the power-law distribution, *Eur. Phys. J. B* **41**, 2, 255-258, DOI: 10.1140 /epjb/e2004-00316-5.

Gutenberg, B., and C.F. Richter (1944), Frequency of earthquakes in California, *Bull. Seismol. Soc. Am.* **34**, 4, 185-188.

Hauksson, E., W. Yang, and P. Shearer (2012), Waveform relocated earthquake catalog for southern California (1981 to June 2011), *Bull. Seismol. Soc. Am.* **102**, 5, 2239-2244, DOI: 10.1785/0120120010.

Hergarten, S. (2002), *Self-Organized Criticality in Earth Systems*, Springer, Berlin.

Jarvinen, B.R., C.J. Neumann, and M.A.S. Davis (1988), A tropical cyclone data tape for the North Atlantic basin, 1886-1983: contents, limitations, and uses, NOAA Technical Memorandum NWS NHC 22, National Hurricane Center, Miami, USA, http://www.nhc.noaa.gov/pdf/NWS-NHC-1988-22.pdf .

Jensen, H.J. (1998), *Self-Organized Criticality. Emergent Complex Behavior in Physical and Biological Systems*, Cambridge University Press, Cambridge.

Johnson, N.L., S. Kotz, and N. Balakrishnan (1994), *Continuous Univariate Distributions*. Vol. 1, 2nd ed., John Wiley & Sons Inc., New York.

Johnson, N.L., A.W. Kemp, and S. Kotz (2005), *Univariate Discrete Distributions*, 3rd ed., John Wiley & Sons Inc., Hoboken.

JTWC (2012), Annual tropical cyclone report, Joint Typhoon Warning Center, http://www.usno.navy.mil/NOOC/nmfc-ph/RSS/jtwc/best_tracks.

Kagan, Y.Y. (2002), Seismic moment distribution revisited: I. Statistical results, *Geophys. J. Int.* **148**, 3, 520-541, DOI: 10.1046/j.1365-246x.2002.01594.x.

Kalbfleisch, J.D., and R.L. Prentice (2002), *The Statistical Analysis of Failure Time Data*, 2nd ed., John Wiley & Sons Inc., Hoboken.

Kanamori, H., and E.E. Brodsky (2004), The physics of earthquakes, *Rep. Prog. Phys.* **67**, 8, 1429-1496, DOI: 10.1088/0034-4885/67/8/R03.

Klafter, J., M.F. Shlesinger, and G. Zumofen (1996), Beyond Brownian motion, *Phys. Today* **49**, 2, 33-39, DOI: 10.1063/1.881487.

Kolmogorov, A.N. (1956), *Foundations of the Theory of Probability*, 2nd ed., Chelsea Pub. Co., New York.

Krane, K.S. (1988), *Introductory Nuclear Physics*, John Wiley & Sons Inc., New York.

Lahaie, F., and J.R. Grasso (1998), A fluid-rock interaction cellular automaton of volcano mechanics: Application to the Piton de la Fournaise, *J. Geophys. Res.* **103**, B5, 9637-9650, DOI: 10.1029/98JB00202.

Main, I.G., L. Li, J. McCloskey, and M. Naylor (2008), Effect of the Sumatran megaearthquake on the global magnitude cut-off and event rate, *Nature Geosci.* **1**, 3, 142, DOI: 10.1038/ngeo141.

Malamud, B.D. (2004), Tails of natural hazards, *Phys. World* **17** , 8, 31-35.

Malamud, B.D., G. Morein, and D.L. Turcotte (1998), Forest fires: An example of self-organized critical behavior, *Science* **281**, 5384, 1840-1842, DOI: 10.1126/science.281.5384.1840.

Malamud, B.D., J.D.A. Millington, and G.L.W. Perry (2005), Characterizing wildfire regimes in the United States, *Proc. Natl. Acad. Sci. USA* **102**, 13, 4694-4699, DOI: 10.1073/pnas.0500880102.

Malmgren, R.D., D.B. Stouffer, A.E. Motter, and L.A.N. Amaral (2008), A Poissonian explanation for heavy tails in e-mail communication, *Proc. Natl. Acad. Sci. USA* **105**, 47, 18153-18158, DOI: 10.1073/pnas.0800332105.

McClelland, L., T. Simkin, M. Summers, E. Nielsen, and T.C. Stein (eds.) (1989), *Global Volcanism 1975-1985*, Prentice Hall, Englewood Cliffs.

Mitzenmacher, M. (2004), A brief history of generative models for power law and lognormal distributions, *Internet Math.* **1**, 2, 226-251, DOI: 10.1080/15427951.2004.10129088.

Newman, M.E.J. (2005), Power laws, Pareto distributions and Zipf's law, *Contemp. Phys.* **46**, 5, 323 -351, DOI: 10.1080/00107510500052444.

NHC (2012), National Hurricane Center, http://www.nhc.noaa.gov/pastall.shtml# hurdat.

Paczuski, M., S. Boettcher, and M. Baiesi (2005), Interoccurrence times in the Bak-Tang-Wiesenfeld sandpile model: A comparison with the observed statistics of solar flares, *Phys. Rev. Lett.* **95**, 18, 181102, DOI: 10.1103/PhysRevLett.95.181102.

Peters, O., and K. Christensen (2006), Rain viewed as relaxational events, *J. Hydrol.* **328**, 1-2, 46-55, DOI: 10.1016/j.hydrol.2005.11.045.

Peters, O., and J.D. Neelin (2006), Critical phenomena in atmospheric precipitation, *Nat. Phys.* **2**, 393-396, DOI: 10.1038/nphys314.

Peters, O., C. Hertlein, and K. Christensen (2001), A complexity view of rainfall, *Phys. Rev. Lett.* **88**, 1, 018701, DOI: 10.1103/PhysRevLett.88.018701.

Peters, O., A. Deluca, A. Corral, J.D. Neelin, and C.E. Holloway (2010), Universality of rain event size distributions, *J. Stat. Mech.* **2010**, P11030, DOI: 10.1088/1742-5468/2010/11/P11030.

Press, W.H., S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery (1992), *Numerical Recipes in FORTRAN: The Art of Scientific Computing*, 2nd ed., Cambridge University Press, Cambridge.

Pruessner, G. (2012), *Self-Organised Criticality: Theory, Models and Characterisation*, Cambridge University Press, Cambridge.

Pueyo, S., and R. Jovani (2006), Comment on "A keystone mutualism drives pattern in a power function", Science **313**, 5794, 1739c–1740c, DOI: 10.1126/science.1129595.

Ross, S. (2002), *A First Course in Probability*, 6th ed., Pearson Education, 528 pp.

Saichev, A., and D. Sornette (2006), "Universal" distribution of interearthquake times explained, *Phys. Rev. Lett.* **97**, 7, 078501, DOI: 10.1103/PhysRevLett.97.078501.

Sethna, J.P., K.A. Dahmen, and C.R. Myers (2001), Crackling noise, *Nature* **410**, 6825, 242-250, DOI: 10.1038/35065675.

Shearer, P., E. Hauksson, and G. Lin (2005), Southern California hypocenter relocation with waveform cross-correlation. Part 2: Results using source-specific station terms and cluster analysis, *Bull. Seismol. Soc. Am.* **95**, 3, 904-915, DOI: 10.1785/0120040168.

Shiryaev, A.N. (1996), *Probability*, 2nd ed., Graduate Texts in Mathematics, Springer, New York.

Silverman, B.W. (1986), *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, New York.

Sornette, D. (2004), *Critical Phenomena in Natural Sciences. Chaos, Fractals, Selforganization and Disorder: Concepts and Tools*, 2nd ed., Springer, Berlin.

Takayasu, H. (1990), *Fractals in the Physical Sciences*, Manchester University Press, Manchester.

Utsu, T. (1999), Representation and analysis of the earthquake size distribution: a historical review and some new approaches, *Pure Appl. Geophys.* **155**, 2-4, 509-535, DOI: 10.1007/s000240050276.

Utsu, T. (2002), Statistical features of seismicity. **In**: W.H.K. Lee, H. Kanamori, P.C. Jennings, and C. Kisslinger (eds.), *International Handbook of Earthquake and Engineering Seismology, Part A*, Vol. 81, Academic Press, Amsterdam, 719-732, DOI: 10.1016/S0074-6142(02)80246-7.

Utsu, T., Y. Ogata, and R. Matsu'ura (1995), The centenary of the Omori formula for a decay law of aftershock activity, *J. Phys. Earth* **43**, 1, 1-33, DOI: 10.4294/jpe1952.43.1.

Wanliss, J.A., and J.M. Weygand (2007), Power law burst lifetime distribution of the SYM-H index, *Geophys. Res. Lett.* **34**, 4, L04107, DOI: 10.1029/2006GL028235.

White, E.P., B.J. Enquist, and J.L. Green (2008), On estimating the exponent of power-law frequency distributions, *Ecology* **89**, 4, 905-912, DOI: 10.1890/07-1288.1.