

# SCIENTIFIC REPORTS



OPEN

## On a Possible Unified Scaling Law for Volcanic Eruption Durations

Flavio Cannavò<sup>1</sup> & Giuseppe Nunnari<sup>2</sup>

Received: 25 August 2015  
Accepted: 11 February 2016  
Published: 01 March 2016

Volcanoes constitute dissipative systems with many degrees of freedom. Their eruptions are the result of complex processes that involve interacting chemical-physical systems. At present, due to the complexity of involved phenomena and to the lack of precise measurements, both analytical and numerical models are unable to simultaneously include the main processes involved in eruptions thus making forecasts of volcanic dynamics rather unreliable. On the other hand, accurate forecasts of some eruption parameters, such as the duration, could be a key factor in natural hazard estimation and mitigation. Analyzing a large database with most of all the known volcanic eruptions, we have determined that the duration of eruptions seems to be described by a universal distribution which characterizes eruption duration dynamics. In particular, this paper presents a plausible global power-law distribution of durations of volcanic eruptions that holds worldwide for different volcanic environments. We also introduce a new, simple and realistic pipe model that can follow the same found empirical distribution. Since the proposed model belongs to the family of the self-organized systems it may support the hypothesis that simple mechanisms can lead naturally to the emergent complexity in volcanic behaviour.

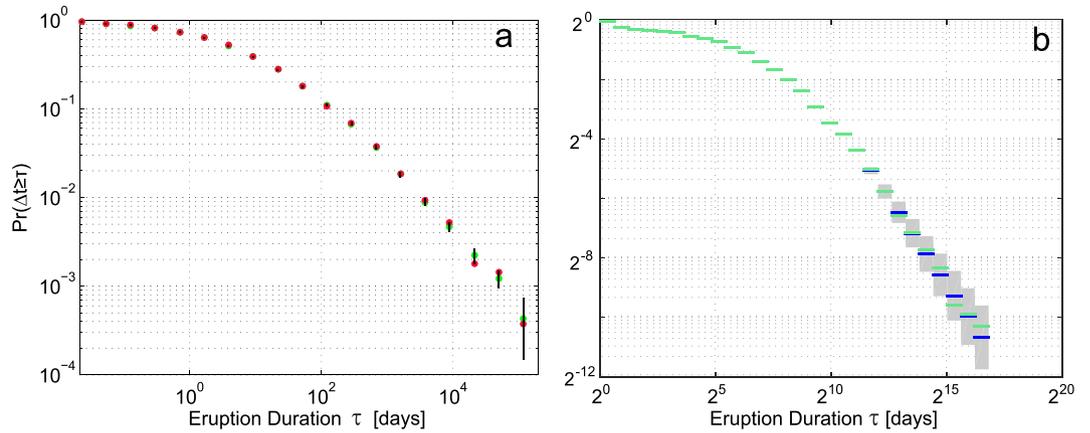
How long a single eruption lasts is one of the most intriguing questions concerning volcano activity. We know that volcanic eruptions can last from less than a day to thousands of years. In 1977, the lava lake at Nyiragongo drained in less than an hour. By contrast, Stromboli has had a low-level of activity since 450 BC (about 2,400 years)<sup>1</sup>. Although the median duration of historic eruptions is ~8 weeks<sup>1</sup>, this value is largely meaningless because of its considerable variance and its non-normal distribution as we demonstrate in the following.

Important factors governing a single eruption are the volume of melt accumulated in the magma reservoir and its degree of over-pressure, but also the possible contribution of new magma from depth. An eruption will normally last until the local melt has been depleted, or until the pressure level of the gas inside the magma reservoir reaches the pre-eruption pressure conditions. Nonetheless, this is a rather simplified overview. The internal plumbing system of an active volcano can be quite intricate, though recent monitoring efforts involving geodetic, seismic and geochemical measurements are helping at improving the estimations of size, depth, and activity of magma chambers under active volcanoes<sup>2-4</sup>.

In recent decades, different researches have been conducted to study volcanism in term of stochastic models<sup>5-8</sup>, while others have tried to characterize the scaling properties and self-similarity of volcanic processes<sup>9-12</sup> such as inter-event times<sup>13,14</sup>, eruption magnitudes<sup>15,16</sup> and tephra produced by eruptions<sup>17</sup>. Some authors have even proposed models based on fluid-rock complex systems to support the hypothesis of emerging self-organization in volcano dynamics<sup>9,18</sup>. Some recent works have also attempted to forecast the duration of volcanic eruption from a probabilistic perspective<sup>19</sup> or from analytical assumptions (as the publicized, but unpublished, example of a forecast of the duration of the recent Bárðarbunga Iceland eruption, see <http://icelandreview.com/news/2014/10/13/volcanologist-predicts-eruptions-end-march> (Date of access: 20/11/2015) where the forecast of duration was based on heuristic considerations about the rate of subsidence of the volcano caldera after the beginning of the eruption). Nevertheless, the forecast of eruption duration is still impaired by the many uncertainties we have in volcanic activity modeling.

The objective of this paper is to present some newly found invariance properties for volcanic activity in terms of eruption time scales. These properties refer to the invariance of the probability distributions under wide changes of temporal scales. In particular, we focus on volcanic eruptions exploring the possibility that their durations follow a power-law distribution, at least for moderate-to-long lasting eruptions.

<sup>1</sup>Istituto Nazionale di Geofisica e Vulcanologia, Osservatorio Etneo, Catania, 95123, Italy. <sup>2</sup>University of Catania, Dipartimento di Ingegneria Elettrica, Elettronica ed Informatica, Catania, 95125, Italy. Correspondence and requests for materials should be addressed to F.C. (email: flavio.cannavo@ingv.it)



**Figure 1. Empirical distribution of eruption durations.** (a) Distribution for eruption durations expressed in days (red dots) and the associated uncertainty (black vertical line). The green dots represent the mean values. (b) Binned distribution of eruption durations (in green) with logarithmic binning scheme of  $2^{0.617}$  and the fitted power-law (in blue) with the associated uncertainty (in gray).

Power-law relationships in natural systems<sup>20</sup> are relevant because they have some intriguing features. They are scale-invariant<sup>21</sup> and thus have the same statistical properties at any scale and are not associated with one characteristic scale. In practise, this means that there is no single correct scale for their analysis because the same principles are valid irrespective of the temporal, spatial or strength scale of analysis<sup>22</sup>, and also makes it technically wrong to apply traditional statistics based on variance and standard deviation (such as regression analysis) when dealing with power-law distributed phenomena. The power-law systems share the concept of “universality”: it has been shown that completely different systems show similar power-law behaviour on approaching a critical state<sup>23</sup>. Moreover, the indication of a power-law behaviour in some data cannot only explain specific kinds of dynamics that might govern the particular natural phenomenon, but, thanks to the universality, may also indicate a deep connection with other, apparently unrelated systems<sup>24</sup>. Several empirical and theoretical investigations have suggested that geophysical systems seem to operate near a critical state, which results in the pervasiveness of power-law behaviour in several descriptors of their dynamics<sup>25,26</sup>. Thus the analysis of power-law and scaling relationships for volcanic eruptions would help us identify general principles that apply across a wide range of scales and types of eruptions, revealing a possible existence of universal principles within the nature of volcanic systems.

In order to sustain our hypothesis of power law for volcanic eruption durations we analyzed all the durations of the known eruptions, avoiding any kind of classification. The data of eruption durations for all the known volcanoes in the world to the best of our knowledge were retrieved from the Smithsonian Institution’s Global Volcanism Program (GVP) database<sup>1</sup>. To date, the GVP catalogue represents the most comprehensive effort in documenting, understanding, and disseminating information about global volcanic activity. It aggregates data from volcanoes and their eruptions for the last 10000 years. For each of the 10415 eruptions, the catalogue reports the starting date to the best of knowledge and for ~40% of them also the ending date. Of the eruptions in the catalogue, 53% are known to the month and 41% to the day<sup>1</sup>. Some ancient eruptions are known only to the century and many only to the year. When possible, the catalogue reports the uncertainty associated with the start and/or end dates (~11% of data).

Discounting the eruptions with unknown duration and any ongoing ones, the database consists of 3841 documented eruptions. The maximum time resolution of the documented eruptions is to the day but about half of the data set shows a larger uncertainty. Because of the time resolution of the GVP catalogue, the eruption duration has been considered as a discrete-valued variable and all the analyses have been carried out considering this constraint.

## Results

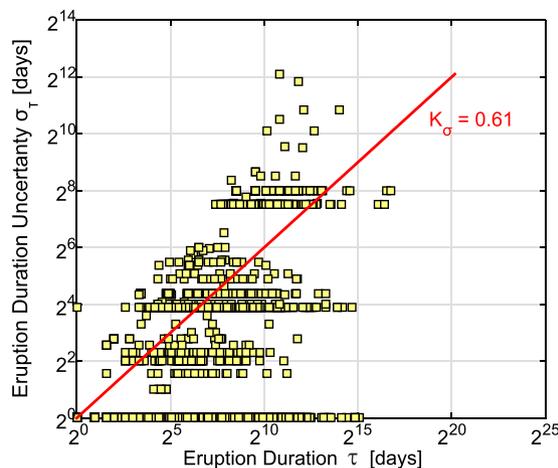
For the purposes of this study, a power-law distribution is described by a probability density function (PDF)  $p(\tau)$  such that

$$p(\tau) \propto \tau^{-\alpha}, \quad (1)$$

where the  $p(\tau)$  represents the punctual probability that an eruption lasts  $\tau > 0$  days and  $\alpha > 1$  is the power-law exponent or scaling factor. For a power-law distributed variable, it is often useful to consider the complementary cumulative distribution function (CCDF) which will still be power-law distributed with:

$$P(\tau) = Pr(\Delta t \geq \tau) \propto \tau^{-\alpha+1} \quad (2)$$

In our analysis we considered the experimental CCDF of durations with its frequency-based approximation. As in most of the natural power-law-distributed systems, in the distribution of eruption durations (Fig. 1a), only the tail of the distribution follows a power-law and the Equations 1 and 2 apply only for values greater than some minimum  $\tau_{min}$ .



**Figure 2. Uncertainty analysis.** Scatter plot in log-log scale of eruption durations and associated uncertainties. The red line represents the best fit for the eruptions with uncertain duration and  $K_\sigma$  its angular coefficient.

$\hat{\alpha}$	$\sigma_{\hat{\alpha}}$	$\tau_{min}$ [days]	$\sigma_{\tau_{min}}$ [days]	Power-law $p$	Log-normal		Exponential		Weibull		$\#\{\tau > \tau_{min}\}$
					$\mathcal{L}$	$p$	$\mathcal{L}$	$p$	$\mathcal{L}$	$p$	
2.16	0.22	1782	581	0.76	-1.36	0.17	17.92	0.00	-0.37	0.71	189

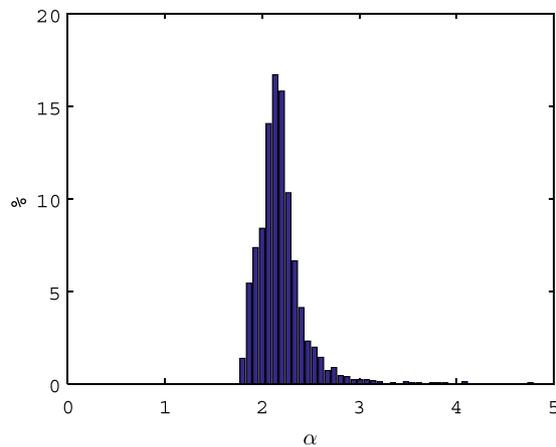
**Table 1. Parameters of the fitted power-law, their  $p$  values and their comparison against the three considered alternatives.**  $\mathcal{L}$  is the normalized log-likelihood ratio against alternatives (positive indicates that the power-law is favored over the alternative) and the  $p$ -value for the significance of the test (the test is considered significant only if  $p < 0.1$  and indicative only with  $\mathcal{L}$  diverging from 0).

Figure 1a shows that, due to the found narrow error band, the distribution of durations seems to follow a power law, even when considering the associated duration uncertainties. In most of the studies of empirical distributions that follow power-laws the slope of the line ( $\alpha$ , from Equation 1) is usually estimated by computing a least-squares linear regression on the log-log space<sup>27,28</sup>. Such graphical analyses can be grossly erroneous<sup>29,30</sup> and thus their results may be unreliable. Following the trace in Clauset *et al.*<sup>30</sup> we embraced the following statistically valid approach for investigating the hypothesis of power-law distribution for volcanic eruption durations: we first fitted the power law with a reliable method based on the maximum likelihood estimator (MLE), then tested the power-law's plausibility and finally we compared the law against alternative distributions.

In order to take into account the uncertainties of the eruption durations, especially for the long and historic eruptions, we binned the data set by converting the values into a series of counts over a set of nonoverlapping ranges in duration size, and applied the methods in Virkar & Clauset<sup>31</sup> for binned distributions to test our hypothesis.

By analyzing the scatter plot of the considered eruption durations against their uncertainties (see Fig. 2) we observed that for most of the eruptions the uncertainty increases with increasing duration and the log-log scale prompted us to choose the logarithmic binning scheme. Bearing in mind that increasing the coarseness of the binning scheme decreases the information available for estimation, and the required sample size increases<sup>31</sup>, we used a fine grained binning scheme (choosing a logarithm base of 2) in order to maximize the statistical accuracy of the analysis. With this choice, the critical sample size required to make the following tests statistically valid is minimum (a few hundreds of samples) as shown in Virkar & Clauset<sup>31</sup>. We roughly quantified the relationship between eruption durations and associated uncertainties with a linear regression in a log-log space, finding an angular coefficient of 0.61 that we used as a step to construct the logarithmic binning scheme. The binned CCDF was calculated by considering the counts of the durations over the bins (see Fig. 1b) thus losing part of the information but indirectly incorporating the uncertainty.

Before estimating the scaling parameter  $\hat{\alpha}$ , we needed to find the minimum duration  $\tau_{min}$  that kept data for which the power-law model was valid. The adopted approach is based on the Kolmogorov-Smirnov (KS) distance<sup>31,32</sup> to quantify the similarity of the estimated distribution and the theoretical power law. We found that for our data set the minimum KS distance is reached for  $\tau_{min} = 1782$  days. Assuming this value as  $\tau_{min}$ , we adopted the maximum-likelihood method for fitting the power-law model to the binned data and identifying the power-law exponent, obtaining  $\hat{\alpha} = 2.16$ . The uncertainties about  $\tau_{min}$  and  $\hat{\alpha}$  were computed by running 1000 Monte Carlo simulations obtaining a  $1\sigma$ -uncertainty of 581 for  $\tau_{min}$  and 0.22 for  $\hat{\alpha}$ . The number of eruptions with duration longer than  $\tau_{min}$  is 189 distributed in 29 bins which are statistically sufficient quantities for empirical data to proceed with the power-law estimation<sup>31</sup>. Figure 1b shows the empirical binned CCDF for the eruption durations and the fitted binned power-law for  $\tau \geq \tau_{min}$  and Table 1 shows the values of the found parameters.



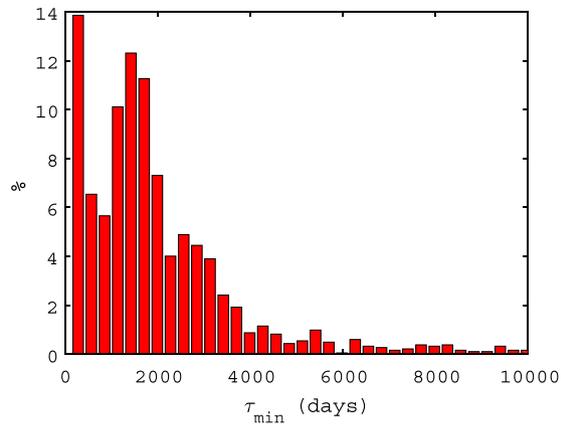
**Figure 3. Normalized distribution of the estimated  $\alpha$  parameter for the power laws obtained from Monte Carlo simulations.**

Given the observed bin counts and a hypothesized power-law distribution from which the counts were drawn, we carried out a goodness-of-fit test in the form of a  $p$ -value to measure the likelihood that the hypothesized model would generate data with a more extreme deviation from the hypothesis than the empirical data<sup>31</sup>. The closer  $p$  is to 0 the more implausible is the model generating the data. Failing to reject the hypothesis is indeed a good indication of the hypothesis itself. As recommended in different works<sup>30,31</sup> we used the conservative choice of ruling out the power-law if  $p < 0.1$ . The  $p$ -value, estimated by the KS statistic approach<sup>31</sup>, results to be 0.76. The rather high value of  $p$  does not necessarily imply the correctness of the power-law hypothesis, but confirms it as a valid hypothesis. In order to support the validity of our conjecture of power-law distribution, we also checked whether another similar kind of alternative heavy-tailed distribution might provide a better fit for our data. Although there are generally an unlimited number of alternative models which can fit the data with any desired precision, only a few are commonly proposed alternatives or correspond to common theoretical mechanisms<sup>30,31,33</sup>: the exponential<sup>27</sup>, the log-normal<sup>34</sup> and the Weibull<sup>35</sup> distributions. We compared the found power-law to the alternative distributions by using a likelihood ratio test for binned data<sup>30,31,36</sup>. The alternative distributions in their binned form were estimated by using an MLE approach<sup>30,31</sup>. Results in Table 1 show that the exponential distribution obtains a positive value of the index measuring the differences of the likelihoods in favour of the power-law distribution,  $\mathcal{L} = 17.92$ , with an associated  $p$ -value of 0.00 which (being  $\leq 0.1$  thus rejecting the null hypothesis of similar likelihood for the compared distributions) allows us to be confident with the value of  $\mathcal{L}$  and (being  $\mathcal{L} > 0$ ) discard the alternative distribution. The comparisons with the log-normal and Weibull distributions give a  $p$ -value of 0.17 and 0.71 respectively, which (being  $> 0.1$ ) indicate that the sign of  $\mathcal{L}$  (although negative) is likely due to chance fluctuations around zero<sup>36</sup> and thus not reliable, and the tests fail to favor one distribution over the other.

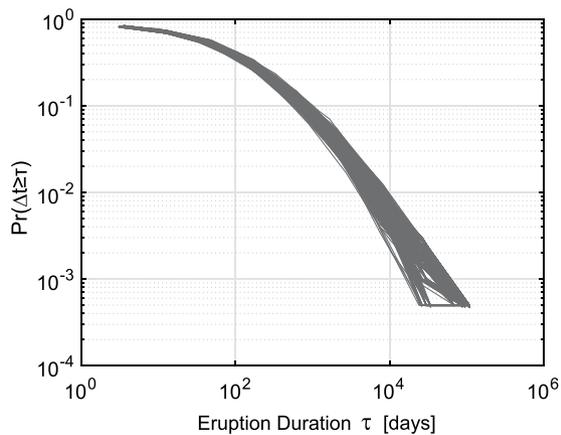
The assessment of  $\tau_{min}$  relies on the quality of the catalogue and its completeness which is defined as the lowest duration at which 100% of the eruptions are likely detected. The high value of  $\tau_{min}$  could be due to the lack of short eruptions in the historical catalogues and hence in a poor completeness of the GPV catalogue. Indeed, it is worth remarking that the GVP database is just a limited sample of recorded eruptions, and an undeterminable number of eruptions may have gone unnoticed or unrecorded entirely. The issue is even more complex since it is also very closely tied to defining the beginning and ending of eruptions. Any volcanic catalogue is clearly the result of observations recorded with complex, spatially and temporally heterogeneous techniques, and processed using a variety of assumptions. Thus, even the best catalogues are heterogeneous and inconsistent in space and time because of the technical limitations in detecting eruptions, and are likely to show as many man-made changes in reporting as natural ones<sup>1</sup>. Choosing to analyze the GVP catalogue we have entirely espoused its assumptions, definitions and consequences. According to the GVP catalogue, an eruption following its predecessor by less than 3 months is considered as a phase of the same eruption, unless careful work has shown it to be distinct<sup>1,37</sup>.

In order to generalize our results we carried out a Monte Carlo simulation by randomly clusterizing a sub-set of eruptions and varying the binning scheme. To increase the result reliability, we selected only the clusters with a significant number of eruptions (i.e.  $\geq 2000$  that is the critical value which allows us to have a set of durations covering at least two decades with a probability of 95% for the used database). For each cluster of eruptions we carried out the analysis previously described by randomly varying the logarithmic binning scheme to check the power law distribution for the durations and to calculate its parameters. After 1000 simulations, the distribution of the found  $\alpha$  parameter shows an expected value of 2.18 with a standard deviation of 0.25 (see Fig. 3) which matches the values estimated for the whole data set. We also analyzed the dependence of  $\alpha$  with the upper duration of the catalogue through a scatterplot between the estimated  $\alpha$  and the maximum duration for all the simulation (see Supplementary Fig. S3). The analysis does not show any first-order correlation between the maximum duration and the scaling factor, at least for large data sets.

The distribution of the found  $\tau_{min}$ , shown in Fig. 4, though more scattered, has an expected value of 1895 days with a standard deviation of 1142, thus sparse but including the range obtained with the whole data set.



**Figure 4.** Normalized distribution of the estimated  $\tau_{min}$  parameter for the power laws obtained from Monte Carlo simulations.



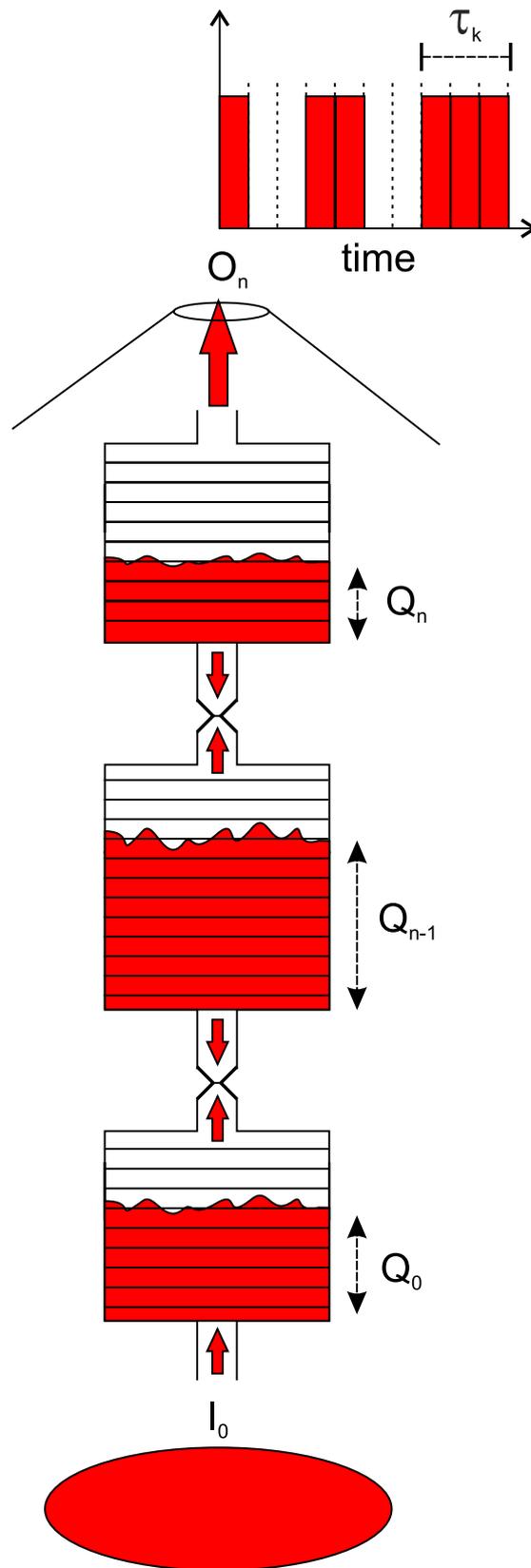
**Figure 5.** Distributions of eruption durations for 1000 Monte Carlo simulations. Each distribution is computed by considering all the eruptions of a randomly chosen set of volcanoes.

To check whether the found power-law in the distribution in Fig. 1 represents a limit of different distributions, or if it is a feature that holds for most volcanoes, we extended the analysis and considered clusterizing the eruptions for volcanoes. We conducted the same Monte Carlo simulations, but this time randomly selecting the volcanoes and taking all their eruptions in order to reach the critical number of eruptions of 2000. Results are comparable with those of the previous analysis thus confirming the previously found values (see Supplementary materials for details). For the same 1000 simulations, we plotted the empirical distributions of durations in Fig. 5, providing an idea of the dispersion of duration distributions around the found power-law and suggesting that the hypothesis of power-law distribution is independent from the volcano.

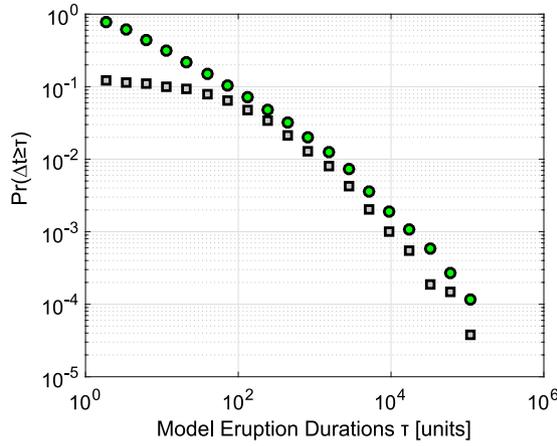
A further analysis performed on the single volcano with the highest number of confirmed eruptions (i.e. Mt. Etna, Italy with 151 eruptions) is presented in Supplementary materials. Although with different coefficients ( $\tau_{min} = 154$  and  $\hat{\alpha} = 1.78 \pm 0.23$ ), it confirms the power-law hypothesis. This allows stressing, that similarly to what happens for the Gutenberg-Richter relation, the power law may represent the universal mechanism which assumes different parameters for different areas.

It is worth noting that this analysis does not exclude that the duration distribution may not follow a power law for other clustering schemes of eruptions<sup>38</sup>. Moreover, a statistical limit of the found parameters is represented by the upper end of the duration scale, the tail of the distribution, that starts to come under the influence of the unavoidable duration truncation, at Holocene age (i.e.  $\sim 4 \times 10^6$  days).

**A simple model of the volcanic process.** In the same way as the block-spring model for the earth crust and earthquakes<sup>39</sup>, we propose a simple model to try to replicate the same behaviour of eruption durations. This is qualitatively inspired on eruption mechanisms, represented by a pipe of  $n$  interconnected vessel cells, as sketched in Fig. 6. Each vessel stores an integer number of magma batches. Because of igneous differentiation<sup>40</sup> and convective motions<sup>41</sup>, magma, despite generally having a net flux directed from bottom to top, depending on the physical conditions within a cell and in neighboring cells, may locally assume a downward motion. Thus, at each time step, part of the stored batches in a cell of the model can migrate to the upper cell or back to the lower cell. For this reason, each cell in the pipe (see Fig. 6), except the first and the last, is modeled as having two outputs: one directed upwards and the other directed downwards. The magma batches issuing from the top vessel, indicated



**Figure 6.** The proposed model to support the hypothesis of power-law distribution for time durations of volcanic eruptions. The fluxes between adjacent vessels are governed by stochastic processes. The uppermost vessel output represents eruptions. The eruption duration is calculated as the number of consecutive batch emissions from the uppermost vessel.



**Figure 7. Power-law of the eruption durations for the simulated model (green circles) with  $n = 10000$  and  $N_T = 10^7$  time steps.** The scaling factor approaches to 2 by increasing  $n$  and  $N_T$ . In gray squares the empirical power-law of eruption durations appropriately vertically translated for a visual comparison.

as  $O_n(t)$  and constrained up to one batch per time step, are considered as erupted. The purpose of the model is to compute the duration of eruptions, measured in time steps. Here, the duration is intended as the number of time steps during which the variable  $O_n(t)$  is continuously positive. It is supposed that the lowest vessel is fed by a constant and continuous flux  $I_0$  of ascending magma batches. Thus the proposed model belongs to the class of slowly driven systems, which also include earthquakes<sup>42</sup>. In order to greatly simplify the model, it is supposed that the magma movement in the  $i_{th}$  cell is decided by local parameters as the local level of magma and a parameter  $\rho_i$ . It is also supposed that  $\rho_i$  is a stochastic process uniformly distributed in the range  $[-1, 1]$  and its strength represents the rate of magma batch transferring from the  $i_{th}$  cell to the neighboring cells. Furthermore, a threshold mechanism ensures that if the accumulated magma in a vessel exceeds the local minimal magma allocation (i.e. 1 batch) then at least one batch must necessarily rise up. The intrinsic nonlinearity in a threshold process should account for the slow-fast dynamics, which is one of the ingredients of self-organized criticality (SOC), as recognized in several natural systems, including earthquakes<sup>42</sup>. Formally, indicating as  $Q_i(t)$  and  $\phi_i(t)$  the number of magma batches accumulated in the  $i_{th}$  vessel and the net flux of batches from the adjacent cells at the time  $t$ , respectively, the process in each model cell can be expressed as in Equation (3).

$$Q_i(t + 1) = Q_i(t) + \phi_i(t) \tag{3}$$

The net flux of batches can be computed as indicated in Equation (4)

$$\phi_i(t) = O_{i-1 \rightarrow i}(t) + O_{i+1 \rightarrow i}(t) - O_{i \rightarrow i-1}(t) - O_{i \rightarrow i+1}(t) \tag{4}$$

where  $O_{i \rightarrow j}$  represents the magma batches that exit from the  $i$ -th vessel towards the  $j$ -th vessel, and are defined as:

$$O_{i \rightarrow i+1}(t) = \begin{cases} 1 + \lfloor \rho_i(t)(Q_i(t) - 1) \rfloor & (Q_i(t) > 1) \& (\rho_i(t) \geq 0) \\ \lfloor \rho_i(t)Q_i(t) \rfloor & (Q_i(t) \leq 1) \& (\rho_i(t) \geq 0) \\ 0 & (Q_i(t) \leq 1) \& (\rho_i(t) < 0) \\ 1 & (Q_i(t) > 1) \& (\rho_i(t) < 0) \end{cases} \tag{5}$$

In expressions above, the symbols  $\lfloor \cdot \rfloor$  represent the rounding operator.

$$O_{i \rightarrow i-1}(t) = \begin{cases} 0 & (Q_i(t) > 1) \& (\rho_i(t) \geq 0) \\ 0 & (Q_i(t) \leq 1) \& (\rho_i(t) \geq 0) \\ -\lfloor \rho_i(t)Q_i(t) \rfloor & (Q_i(t) \leq 1) \& (\rho_i(t) < 0) \\ -\lfloor \rho_i(t)(Q_i(t) - 1) \rfloor & (Q_i(t) > 1) \& (\rho_i(t) < 0) \end{cases} \tag{6}$$

Expressions (5) and (6) clarify that the upward and downwards outputs are mutually exclusive, except in the case  $(Q_i(t) > 1) \& (\rho_i(t) < 0)$ . It can easily be verified, by noting that the mean values of  $\rho_i(t)$  are null in Eqs (5) and (6), that the model output is directed from bottom to top on average. The general mathematical mechanism that rules the proposed model relies to the sandpile model<sup>43</sup> constrained to an anisotropic 1-dimension with a privileged versus (towards the summit). For a sufficiently large number of vessels  $n$  (in our cases  $n \gg 10$ ) and for a sufficiently long time steps, the simulations of such a model show that the eruption durations follow a power-law distribution (see Fig. 7). Despite the extreme simplification, the scaling factor is close to the one obtained empirically from the real data (i.e.  $\alpha \rightarrow 2$ , see Fig. 7 for a visual comparison).

Thus, the proposed model seems to obey the same self-organizing (i.e. it always reaches the same attractor independently of the parameter values) behaviour of several other complex systems, that after the pioneering sandpile model<sup>43</sup>, has been recognized in several natural systems (e.g. earthquakes and forest fires<sup>25</sup>). The similarity of the models allows us to infer same mechanisms observed in other natural systems<sup>44</sup>. This class of systems shows a characteristic slow-fast dynamic which exhibits relatively small variations many times in the output and abrupt large changes a few times (but more than expected) in the output.

## Discussion

The results presented here suggest that the power-law behaviour in volcanic eruption durations seems a reasonably satisfactory representation of the duration-frequency distribution at least for moderate-to-long eruptions. The found power-law together with previous evidences for other volcanic observables<sup>9,45</sup> could be the fingerprints that volcanoes self-organize to the critical state, thus supporting the hypothesis that the mechanisms behind an eruption belong to the class of SOC<sup>26</sup>. However, the induction of SOC nature for systems showing power-law behaviour is controversial, even among experts, and has attracted significant criticism<sup>46–51</sup>. Nevertheless, the origin of scaling law and SOC in natural systems is gaining increasingly interest in the field of statistical geophysics<sup>47,52–56</sup>. Recent studies have shown that the power-law behaviours attributable to SOC can be explained by a version of the central limit theorem applied to a sum of a large number of independent random variables described by a family of statistical heavy-tailed distributions, which need not be power laws<sup>50,56–58</sup>. Thus, the stochastic or SOC origin of the power-law behaviour in natural systems is currently a matter of heated debate. In the hypothesis of SOC origin of the found power-law behaviour, the knowledge of the scaling factor could be used to make inferences about the underlying processes, to test mechanistic models, and to estimate and predict patterns and processes operating beyond the scope of the observed data. The proposed simple model could support the self-organized mechanism of the volcanic eruptions by which the emergence of complexity from simple local interactions can be spontaneous, and therefore plausible as a source of natural complexity. The model also shows that a simple magma transfer/injection in some part of the magmatic system can cause a large variety of eruption durations. The hypothesis of SOC behaviour would lead to some intriguing consequences. First, a significant part of active volcanoes should be close to eruption instability, thus susceptible to erupt, though quiescent at the moment. The quantitative determination of the susceptible magma transfer/injection cannot be ascertained with precision, however the SOC behaviour would allow us to observe that the sensitive part of a volcano can be activated with relatively small perturbations. Second, we should await long eruptions more frequently than expected. This information should be relevant for volcanic hazard mitigation. Third, volcanic systems should show the emergence of invariant quantities. We have found that the volcanic eruption durations seem to show an exponent of the power-law that assumes globally the same value of the  $b$ -coefficient of the well-known Gutenberg-Richter power-law for earthquakes<sup>59,60</sup>. Volcanic systems and earthquakes seem thus to share the same fundamental SOC behaviour whereby the systems naturally and independently of the physical parameters reach a critical state which can abruptly change for stochastic low-powered reasons (thus unpredictable with feasible precision).

In light of the found characteristic of volcanic eruption durations, and due to the enormous quantities of unknowns and the considerable uncertainties in the estimated parameters of the Earth and volcanoes interior, which are modeled as stochastic processes in the proposed model, we believe that the forecasts of eruption durations in terms of statistical expectation have the same reliability as for the earthquake magnitude forecast.

However, we would underline that the proposed analysis is intended as exploratory rather than explanatory. It is clearly tied to the limits of the catalogue, to the long time coverage, thus to indirect and partially inaccurate measurements, as well as the intrinsic heterogeneities in eruption definition.

Notwithstanding, similarly to the seismological literature, the found relation could be a reasonably satisfactory representation of the volcanic duration-frequency distribution of large populations of eruptions in terms of moderate-to-long lasting events, which could break down in the tails of very short or very long cases. Further studies and evidences are needed to confirm or refute the proposed hypothesis.

## Methods

The uncertainties associated to the durations were calculated as the sums of the uncertainties (given explicitly or with date truncation) of the relative start and end dates. For the calculation of the error band in the empirical distribution of eruption durations (in Fig. 1a) we used a bootstrap resampling for propagating the duration uncertainties. In this work, we adopted the reliable method based on the maximum likelihood estimator (MLE) to estimate the power-law exponent. MLE determines the unknown exponent that maximizes the likelihood of the power-law given the observed data.

In order to choose the value of  $\tau_{min}$  that makes the binned CCDF distributions of the best-fit power-law model and the measured data as similar as possible above  $\tau_{min}$ , we used the method described in Virkar & Clauset<sup>31</sup>. To quantify the similarity of the two probability distributions, we used the Kolmogorov-Smirnov (KS) distance<sup>32</sup>, thus estimating the  $\tau_{min}$  as the  $\tau$  that minimizes such a distance for  $\tau \geq \tau_{min}$ . Then we used a nonparametric bootstrap method<sup>61</sup> to estimate the error associated to  $\tau_{min}$ . For the Monte Carlo simulations used to estimate the power-law parameter distributions we varied the logarithmic binning scheme choosing the base randomly between 2 and 10.

We used a goodness-of-fit test generating a  $p$ -value that quantified the likelihood of the dataset, supposing our hypothesis to be true<sup>30</sup>. Such a test is based on measuring the distance between the empirical and the hypothesized distributions. This distance is compared with a set of distances calculated for comparable synthetic data sets obtained from the same distribution, and the  $p$ -value corresponds to the fraction of the synthetic distances that are larger than the empirical distance. In order to obtain accurate estimates of  $p$ -value, we looked for synthetic data that showed a distribution similar to the empirical data below  $\tau_{min}$  but that followed the fitted power-law with scaling parameter  $\hat{\alpha}$  above  $\tau_{min}$ . We fitted each synthetic data set separately to its own power-law model and cal-

culated the KS distance for each one relative to its own model. To generate such datasets, we adopted a semiparametric approach described in Clauset *et al.*<sup>30</sup>. We generated a total of 1000 synthetic data sets.

We evaluated the power-law distribution against another distribution by using the log-likelihood ratio test<sup>33,62</sup> which compares the likelihood of the data under two alternative distributions with PDFs  $p_1(\tau)$  and  $p_2(\tau)$  by calculating the index in (7).

$$\mathcal{L} = \sum_{i=1}^n [\ln p_1(\tau_i) - \ln p_2(\tau_i)] \quad (7)$$

A positive value of  $\mathcal{L}$  suggests a better fit for distribution  $p_1$  against  $p_2$ . The test cannot indicate anything if the expected value of  $\mathcal{L}$  fluctuates close to zero. Thus we look at the probability that the measured log-likelihood ratio is equal or larger than the observed value  $|\mathcal{L}|$  given by<sup>36</sup>

$$p_{\text{value}} = \text{erfc} \left( \frac{|\mathcal{L}|}{\sqrt{2n} \sigma_{\mathcal{L}}} \right), \quad (8)$$

where  $\text{erfc}$  is the complementary Gaussian error function. This  $p$ -value represents the probability that we measure a given value of  $|\mathcal{L}|$  when the true value of  $\mathcal{L}$  is unreliable for discriminating between the distribution (its expected value is close to zero). If this  $p$ -value is small (usually,  $p \leq 0.1$ ), then the calculated  $|\mathcal{L}|$  is unlikely to be a random result and thus its information is statistically valid and it is possible to discriminate between the two distributions<sup>63</sup>.

In the proposed discrete model simulations we used  $I_0 = 1$ , but the model shows the same behaviour for any other parameter values and thus it is justified a vertical translation, in order to compare the two distributions shown in Fig. 7.

## References

- Siebert, L., Simkin, T. & Kimberly, P. *Volcanoes of the World* (Univ of California Press, 2010).
- Lowenstern, J. B., Smith, R. B. & Hill, D. P. Monitoring super-volcanoes: geophysical and geochemical signals at yellowstone and other large caldera systems. *Phil. Trans. R. Soc. A* **364**, 2055–2072 (2006).
- Acocella, V. Great challenges in volcanology: how does the volcano factory work? *Front. Earth Sci.* **2**, doi: 10.3389/feart.2014.00004 (2014).
- Cannavò, F. *et al.* Real time tracking of magmatic intrusions by means of ground deformation modeling during volcanic crises. *Sci. Rep.* **5**, doi: 10.1038/srep10970 (2015).
- Wickman, F. Markov models of repose-period patterns of volcanoes. In *Random Processes in Geology*, 135–161 (Springer, 1976).
- De la Cruz-Reyna, S. Poisson-distributed patterns of explosive eruptive activity. *Bull. Volcanol.* **54**, 57–67 (1991).
- Salvi, F., Scandone, R. & Palma, C. Statistical analysis of the historical activity of mount etna, aimed at the evaluation of volcanic hazard. *J. Volcanol. Geotherm. Res.* **154**, 159–168 (2006).
- Bebbington, M. Trends and clustering in the onsets of volcanic eruptions. *J. Geophys. Res. Solid Earth* **115**, B01203, doi: 10.1029/2009JB006581 (2010).
- Grasso, J. R. & Bachelery, P. Hierarchical organization as a diagnostic approach to volcano mechanics: Validation on piton de la fournaise. *Geophys. Res. Lett.* **22**, 2897–2900 (1995).
- Pelletier, J. D. Statistical self-similarity of magmatism and volcanism. *J. Geophys. Res. Solid Earth* **104**, 15425–15438 (1999).
- Gusev, A. Temporal structure of the global sequence of volcanic eruptions: Order clustering and intermittent discharge rate. *Phys. Earth Planet. In.* **166**, 203–218 (2008).
- Sanchez, L. & Shcherbakov, R. Temporal scaling of volcanic eruptions. *J. Volcanol. Geotherm. Res.* **247**, 115–121 (2012).
- Pyle, D. M. Forecasting sizes and repose times of future extreme volcanic events. *Geology* **26**, 367–370 (1998).
- Marzocchi, W. & Zaccarelli, L. A quantitative model for the time-size distribution of eruptions. *J. Geophys. Res. Solid Earth* **111**, doi: 10.1029/2005JB003709 (2006).
- Simkin, T. Terrestrial volcanism in space and time. *Annu. Rev. Earth Planet. Sci.* **21**, 427–452 (1993).
- Gusev, A., Ponomareva, V., Braitseva, O., Melekestsev, I. & Sulerzhitsky, L. Great explosive eruptions on Kamchatka during the last 10,000 years: Self-similar irregularity of the output of volcanic products. *J. Geophys. Res. Solid Earth* **108**, doi: 10.1029/2001JB000312 (2003).
- Turcotte, D. L. *Fractals and chaos in geology and geophysics* (Cambridge university press, 1997).
- Lahaie, F. & Grasso, J. R. A fluid-rock interaction cellular automaton of volcano mechanics: Application to the piton de la fournaise. *J. Geophys. Res. Solid Earth* **103**, 9637–9649 (1998).
- Gunn, L., Blake, S., Jones, M. & Rymer, H. Forecasting the duration of volcanic eruptions: an empirical probabilistic model. *Bull. Volcanol.* **76**, 1–18 (2014).
- Marquet, P. A. *et al.* Scaling and power-laws in ecological systems. *J. Exp. Biol.* **208**, 1749–1769 (2005).
- Sornette, D. *Critical Phenomena in Natural Sciences: Chaos, Fractals, Selforganization and Disorder: Concepts and Tools* (Springer Series in Synergetics) (Springer, 2006).
- Milne, B. T. Motivation and benefits of complex systems approaches in ecology. *Ecosystems* **1**, 449–456 (1998).
- Binney, J. J., Dowrick, N., Fisher, A. & Newman, M. *The theory of critical phenomena: an introduction to the renormalization group* (Oxford University Press, Inc., 1992).
- Simon, H. A. On a class of skew distribution functions. *Biometrika* **42**, 425–440 (1955).
- Hergarten, S. *Self-organized criticality in earth systems* (Springer, 2002).
- Bak, P. *How nature works: the science of self-organized criticality*, vol. 212 (Copernicus New York, 1996).
- Arnold, B. C. *Pareto distribution* (Wiley Online Library, 1985).
- Newman, M. E. Power laws, pareto distributions and zipf's law. *Contemp. Phys.* **46**, 323–351 (2005).
- Goldstein, M. L., Morris, S. A. & Yen, G. G. Problems with fitting to the power-law distribution. *Eur. Phys. J. B* **41**, 255–258 (2004).
- Clauset, A., Shalizi, C. R. & Newman, M. E. Power-law distributions in empirical data. *SIAM Rev.* **51**, 661–703 (2009).
- Virkar, Y. & Clauset, A. Power-law distributions in binned empirical data. *Ann. Appl. Stat.* **8**, 89–119 (2014).
- Massey, F. J., Jr. The kolmogorov-smirnov test for goodness of fit. *J. Am. Stat. Assoc.* **46**, 68–78 (1951).
- Clauset, A., Young, M. & Gleditsch, K. S. On the frequency of severe terrorist events. *J. Confl. Resolut.* **51**, 58–87 (2007).
- Crow, E. L. *Lognormal Distributions: Theory and Applications* (CRC Press, 1987).

35. Weibull, W. A statistical distribution function of wide applicability. *J. Appl. Mech.* **18**, 293–297 (1951).
36. Vuong, Q. H. Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica* **57**, 307–333 (1989).
37. Venzke, E. *et al.* Global volcanism, 1968 to the present. In *Smithsonian Institution Global Volcanism Program Digital Information Series, GVP-4* (2002).
38. Sparks, R. S. J. & Aspinall, W. P. *Volcanic Activity: Frontiers and Challenges in Forecasting, Prediction and Risk Assessment*, 359–373 (American Geophysical Union, 2004).
39. Burridge, R. & Knopoff, L. Model and theoretical seismicity. *Bull. Seismol. Soc. Am.* **57**, 341–371 (1967).
40. McConnell, R. K., Jr., McClaine, L. A., Lee, D. W., Aronson, J. R. & Allen, R. V. A model for planetary igneous differentiation. *Rev. Geophys.* **5**, 121–172 (1967).
41. Middlemost, E. A. K. *Magma, Rocks and Planetary Development: A Survey of Magma/Igneous Rock Systems* (Taylor & Francis, 2014).
42. Jensen, H. J. *Self-Organized Criticality - Emergent Complex Behavior in Physical and Biological Systems* (Cambridge University Press, 1998).
43. Bak, P., Tang, C. & Wiesenfeld, K. Self-organized criticality: An explanation of the  $1/f$  noise. *Phys. Rev. Lett.* **59**, 381–384 (1987).
44. Turcotte, D. L. Self-organized criticality. *Rep. Prog. Phys.* **62**, 1377–1429 (1999).
45. Sornette, D. Volcanic tremors given mathematical foundation. *Phys. World* **5**, 23–24 (1992).
46. Perković, O., Dahmen, K. & Sethna, J. P. Avalanches, Barkhausen noise, and plain old criticality. *Phys. Rev. Lett.* **75**, 4528–4531 (1995).
47. Carlson, J. M. & Doyle, J. Highly optimized tolerance: A mechanism for power laws in designed systems. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **60**, 1412–1427 (1999).
48. Krommes, J. A. Self-organized criticality, long-time correlations, and the standard transport paradigm. *Phys. Plasmas* **7**, 1752–1758 (2000).
49. Frigg, R. Self-organised criticality - what it is and what it isn't. *Stud. Hist. Philos. Sci. Part A* **34**, 613–632 (2003).
50. Stumpf, M. P. H. & Porter, M. A. Critical truths about power laws. *Science* **335**, 665–666 (2012).
51. Watkins, N. W., Pruessner, G., Chapman, S. C., Crosby, N. B. & Jensen, H. J. 25 years of self-organized criticality: Concepts and controversies. *Space Sci. Rev.* 1–42, doi: 10.1007/s11214-015-0155-x (2015).
52. Lovejoy, S. & Schertzer, D. Scale, scaling and multifractals in geophysics: twenty years on. In *Nonlinear dynamics in geosciences*, 311–337 (Springer, 2007).
53. Schubert, G. (ed.) *Treatise on Geophysics* (Elsevier, 2015), 2 edn.
54. Dmowska, R. (ed.) *Advances in Geophysics* Volume 46 (Academic Press, 2004).
55. Eisler, Z., Bartos, I. & Kertesz, J. Fluctuation scaling in complex systems: Taylor's law and beyond 1. *Adv. Phys.* **57**, 89–142 (2008).
56. Kendal, W. S. Self-organized criticality attributed to a central limit-like convergence effect. *Phys. A* **421**, 141–150 (2015).
57. Willinger, W., Alderson, D., Doyle, J. & Li, L. More "normal" than normal: scaling distributions and complex systems. Paper presented at *Winter Simulation Conference, 2004*, Washington DC, IEEE, doi: 10.1109/WSC.2004.1371310 (2004, Dec. 5–8).
58. Kendal, W. S. & Jørgensen, B. Taylor's power law and fluctuation scaling explained by a central-limit-like convergence. *Phys. Rev. E* **83**, doi: 10.1103/PhysRevE.83.066115 (2011).
59. Bak, P., Christensen, K., Danon, L. & Scanlon, T. Unified scaling law for earthquakes. *Phys. Rev. Lett.* **88**, 178501 1–4 (2002).
60. Marzocchi, W. & Sandri, L. A review and new insights on the estimation of the b-value and its uncertainty. *Ann. Geophys.* **46**, 1271–1282 (2003).
61. Efron, B. & Tibshirani, R. J. *An introduction to the bootstrap*, vol. 57 (CRC press, 1994).
62. Bhattacharya, I. & Getoor, L. Collective entity resolution in relational data. *ACM Trans. Knowl. Discov. Data* **1**, doi: 10.1145/1217299.1217304 (2007).
63. Cramér, H. *Mathematical methods of statistics*, vol. 9 (Princeton university press, 1999).

## Author Contributions

G.N. and F.C. shared the ideas defining the topic of the research. F.C. made the analysis and jointly with G.N. prepared the manuscript.

## Additional Information

**Supplementary information** accompanies this paper at <http://www.nature.com/srep>

**Competing financial interests:** The authors declare no competing financial interests.

**How to cite this article:** Cannavò, F. and Nunnari, G. On a Possible Unified Scaling Law for Volcanic Eruption Durations. *Sci. Rep.* **6**, 22289; doi: 10.1038/srep22289 (2016).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>