Check for updates

AGU ADVANCING EARTH AND SPACE SCIENCE

JGR Solid Earth

RESEARCH ARTICLE

10.1029/2021JB021763

Key Points:

- The continuous distribution of global subaerial volcanism on Earth is modeled and quantified
- The distribution displays an initial (erupted volume < 170 Mm^3) log-normal section, followed by a power law section encompassing several orders of magnitude
- Power law—distributed eruption volumes may rule out the possibility of short-term deterministic forecasts of the size of an impending eruption

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

P. Papale,
paolo.papale@ingv.it

Citation:

Papale, P., Marzocchi, W., & Garg, D. (2021). Global volume distribution for subaerial volcanism on Earth. *Journal of Geophysical Research: Solid Earth*, *126*, e2021JB021763. https://doi.org/10.1029/2021JB021763

Received 29 JAN 2021
Accepted 10 MAY 2021

© 2021. The Authors.
This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

Global Volume Distribution for Subaerial Volcanism on Earth

P. Papale[1], W. Marzocchi[2], and D. Garg[1]

[1]Istituto Nazionale di Geofisica e Vulcanologia, Sezione di Pisa, Pisa, Italy, [2]Dipartimento di Scienze della Terra, dell'Ambiente e delle Risorse, Università di Napoli Federico II, Napoli, Italy

Abstract Knowledge of the global distribution of Earth volcanism is critical in many fields of the Geosciences involving large-scale assessments, such as plate tectonics, global volcanic hazards, and climate change. Recent analysis has revealed that global eruption inter-event times are exponentially distributed, implying that on the global scale volcanic eruptions are Poisson distributed. Here, we employ those findings to calibrate a continuous frequency-volume distribution for subaerial eruptions of any size on Earth from small lava flows to super-eruptions. Obtaining such a continuous global distribution implies considering the existing data and the way they are collected and categorized into databases, as well as extending the available eruption volume data to eruption VEI classes less than 4. The continuous global distribution shows an initial log-normal section up to volumes of about 170 Mm^3, followed by a power-law section, tapered on its extreme right-end side, encompassing about five orders of magnitude of eruption volumes. The potential implications are discussed in terms of short-term eruption forecasts of the size of an impending eruption, critical for volcanic emergency management.

Plain Language Summary The occurrence of volcanic eruptions over the Earth follows apparently complex patterns: while the vast majority of the eruptions are relatively small in size, here and there less frequent large eruptions appear, and even less frequently, cataclysmic eruptions take place menacing vast regions up to the global Earth. Summed up with relatively quick deterioration of the information from the geologic record, especially for small to medium size eruptions, such apparently irregular trends have largely limited our understanding and forecasting capabilities. New databases of volcanic eruptions, and new statistical analyses, allow us to determine the size distribution of volcanic eruptions worldwide, from the smallest lava flows to the largest explosive eruptions known to have occurred on Earth. We find that above a relatively small eruption volume threshold all eruptions distribute according to what is called a power law, which is also known to describe other natural phenomena such as earthquakes, wild fires, and many others. The mechanisms subtending the generation of a power-law distribution for the global subaerial volcanism are not immediately clear. However, as for several other similarly distributed phenomena, the implications may impact our capability to forecast the size of an impending eruption, with relevant consequences for volcanic emergency management.

1. Introduction

Knowledge of the size distribution of Earth volcanism is critical to understand several aspects of the dynamics of our planet. In fact, close relationships between size distribution and intimate process dynamics are described and generally accepted for an increasingly large variety of natural as well as man-made phenomena (e.g., Laherrère & Sornette, 1998; Clauset et al., 2009; Tenreiro Machado & Lopes, 2015). In the case of volcanism, the way the discharged magma distributes among individual eruptions is expected to hold relationships with some of the most relevant processes occurring on Earth, such as mantle dynamics (Huang et al., 1997), Earth crust formation (Ito & Clift, 1998), and plate tectonics (Xu et al., 2009), and to be a factor in global models of climate change (Robock, 2000). Robust estimates of the size distribution of volcanic eruptions on Earth require sufficiently complete global databases of volcanic activity. Such databases are existing and continuously implemented (Crosweller et al., 2012; Global Volcanism Program, 2013), and their analysis led to the development of a statistical model for subaerial volcanism on Earth (Papale, 2018) which provides a basis for the determination of the continuous global eruption volume distribution illustrated here.

Table 1
Rate Parameters and Their Inverse Corresponding to the Average Inter-Event Times for Different VEI Classes of Eruptions

VEI class	Mean rate parameter (day^{-1})	Standard deviation of the mean rate parameter (day^{-1})	Average inter-event time	Catalog completeness (years BP[a])
Any	1.1096×10^{-1}	2.3318×10^{-3}	9.012 days	−25
0	1.6401×10^{-2}	8.2004×10^{-4}	60.97 days	0
1	3.8609×10^{-2}	1.5855×10^{-3}	25.90 days	−25
2	4.1752×10^{-2}	1.3079×10^{-3}	23.95 days	0
3	1.2473×10^{-2}	7.1772×10^{-4}	80.17 days	0
4	1.4904×10^{-3}	1.5538×10^{-4}	671.0 days	110
5	2.0026×10^{-4}	3.1275×10^{-5}	13.68 years	520
6	3.8165×10^{-5}	6.5452×10^{-6}	71.79 years	2,500
7	9.7473×10^{-7}	1.4865×10^{-7}	2,811 years	125,000
8	3.5126×10^{-8}	6.7601×10^{-9}	78,055 years	2,600,000

Abbreviation: VEI, Volcanic Explosivity Index (Newhall & Self, 1982).
[a]BP stands for "Before Present", and refers to the time back from year 1950 CE, taken as positive (e.g., year 1810 CE corresponds to 140 BP, and year 2020 CE to −70 BP).

Extracting a global picture from the existing databases is largely complicated by evident deterioration of the information when getting back in time, by even just a few decades for low VEI eruptions (Crosweller et al., 2012). The key element was represented by understanding that global eruption recurrence times are exponentially distributed, or equivalently, that global volcanic eruptions are Poisson distributed (Papale, 2018). On one hand, that discovery allows a robust definition of catalog completeness (or substantial completeness) for each discrete VEI class of eruptions (VEI: Volcanic Explosivity Index, Newhall & Self, 1982). That is illustrated in Papale (2018) and summarized in Table 1 in terms of parameters of the Poisson distributions and catalog completeness. On the other hand, exponentially distributed eruption inter-event times contribute to explain previous difficulties in obtaining a global picture from the data set. In fact, entropy maximization associated with exponential distributions easily makes time series of events appearing as random sequences with no clear patterns. When summed up with effective holes and gaps in the records and dramatic, size-dependent deterioration of the information with age, the distributions appear as a messy sequence of events, and one would hardly believe they are more than satisfactorily described by a single statistical distribution. That is in fact the case, and once the exponential distribution of inter-event times was recognized, a consistent, simple, robust picture of the time-size distribution of subaerial volcanism on Earth emerged (Table 1).

The analysis in Papale (2018) referred to discrete VEI classes of eruptions. A discussion of the pros and cons in referring to either erupted volumes or masses is reported in the electronic-only Methods section of Papale (2018). In summary, use of masses would be preferable, as this is for countless reasons a much sounder physical quantity with respect to volumes. Unfortunately, use of masses is less straightforward for two main reasons: first, volcanologists typically estimate deposit volumes, by making use of a variety of techniques that are increasingly becoming sophisticated. Such volumes are then converted into masses by assuming some average density or, rarely, some density distribution. As a consequence, mass estimates involve additional uncertainty as they sum up the one from the volume estimates to that deriving from the assumptions on density. Second, such a conversion of volumes into densities is done by the original authors who worked on the deposits and estimated their volumes only in a minor fraction of cases, adding further uncertainty to the final estimates. As a matter of fact, most conversions for dominantly explosive eruptions simply assume an average density of 1,000 kg/m^3 (cfr. the LaMEVE database), establishing a linear relationship between eruption volume and mass which *de facto* makes it equivalent to refer to one or the other. While we stress the relevance of developing and extensively using more reliable methods to estimate erupted masses, we acknowledge that referring to volumes seems currently, as a minimum, equally justified.

Once converted into relative frequency of occurrence, the rate parameters in Table 1 for the discrete VEI classes from 3 to 8 appeared to depict a power-law distribution. These two elements: exponential distribution of eruption inter-event times, and power-law distribution of the frequency of dominantly explosive eruptions, are the fundamental aspects of the global distribution of volcanic eruptions that emerged from the discrete VEI analysis in Papale (2018). This study builds up from those results to determine a continuous eruption volume distribution for global volcanism on Earth, from small lava flows to volcanic super-eruptions. In fact, obtaining a continuous volume distribution requires substantial efforts described in this work, and reveals further aspects of the global distribution that are illustrated and discussed in the following.

2. Data Analysis

We consider first the case of the explosive eruptions with VEI ≥ 4, then the smaller scale eruptions with VEI = 0–3.

2.1. Data Analysis for Explosive Eruptions with VEI ≥ 4

As recalled above, the frequency of the dominantly explosive VEI classes ≥3 appears to distribute according to a power law. It seems reasonable that such a power-law relationship holds also on a continuous volume scale. As a matter of fact, the slope of the power law in the discrete VEI representation was determined on a continuous volume distribution that, once categorized into discrete VEI classes, produces the observed VEI frequency distribution (Papale, 2018). Here, we directly address the problem of the continuous volume distribution, with equal emphasis on the tail of the distribution (the large or very large explosive eruptions) as well as on the bulk distribution describing more frequent, less voluminous eruptions.

A power law distribution is described by the following.

$$PDF = \frac{k-1}{V_{min}}\left(\frac{V}{V_{min}}\right)^{-k} \tag{1}$$

where PDF stands for Probability Density Function, V_{min} is the minimum volume above which the power-law distribution holds, and k is the power-law exponent (also called "slope" or "scaling" parameter). The corresponding CDF (Cumulative Distribution Function) and its complementary cCDF = 1 − CDF (also called the survivor function, here expressed as S) are given by

$$CDF = 1-\left(\frac{V}{V_{min}}\right)^{1-k};\ S = \left(\frac{V}{V_{min}}\right)^{1-k} \tag{2}$$

From Equations 1 and 2, it follows that the PDF and S of a power-law distribution produce a straight line on a log-log plot (while the CDF does not). Such a simple trend is useful, as it can be visually compared with data. In particular, the survivor function S is useful, since constructing it from real data does not require any artificial binning as for the PDF while keeping all of the properties of the distribution. When referring to power-law distributed data within a pre-defined interval, for example, the eruption volumes within a given VEI class or group of VEI classes, there is an additional aspect to take into account. In fact, power laws are such only when they extend to infinity on their right end. When the distribution has an upper limit (e.g., a physical limit to the values that the variable can take, or a bound reflecting a given window under consideration), the power law is truncated, and its survivor function diverges from the rectilinear trend (on a log-log plot) reflecting a more rapid approach to zero (Kagan, 2002, offers a detailed description for earthquake magnitudes). In such a situation, the straight line trend of the survivor only holds for values of the variable sufficiently far from the limit.

The above is relevant for the present situation, since observed eruption volume distributions must be evaluated separately for the different VEI classes. In fact, volcanic eruptions database completeness strongly depends on VEI, and the information rapidly deteriorates with time, much quicker for smaller volume eruptions as a consequence of much lower preservation in the geological record (Crosweller et al., 2012; Papale, 2018). Table 1 provides a reference for catalog completeness in terms of eruption VEI, based on the appearance of statistically significant deviations from the exponential distribution found to describe eruption inter-event times (Papale, 2018). From catalog completeness and individual VEI inter-event times from the same Table, it follows that to manage a complete distribution including a minimum number of large explosive eruptions, the database should give us the volume of any volcanic eruption during at least 1 million years. With the rate parameters in Table 1, that translates into a complete database with about 40 million entries, which is by far beyond presently available databases.

The consequences of the above strong limits in data availability are important, and require the design of a method whereby the eruption volume distribution is reconstructed from the available data. Figure 1 reports the case for eruptions with VEI 4 or larger, for which volume estimates are available from the LaMEVE database (Crossweller et al., 2012).

The data in Figure 1 refer to entries from LaMEVE for which the nominal volume is consistent with the corresponding VEI attribution. For such cases, VEI ≤ $\log\left[V\left(m^3\right)\right] - 4 <$ VEI + 1 (Newhall & Self, 1982; Deligne

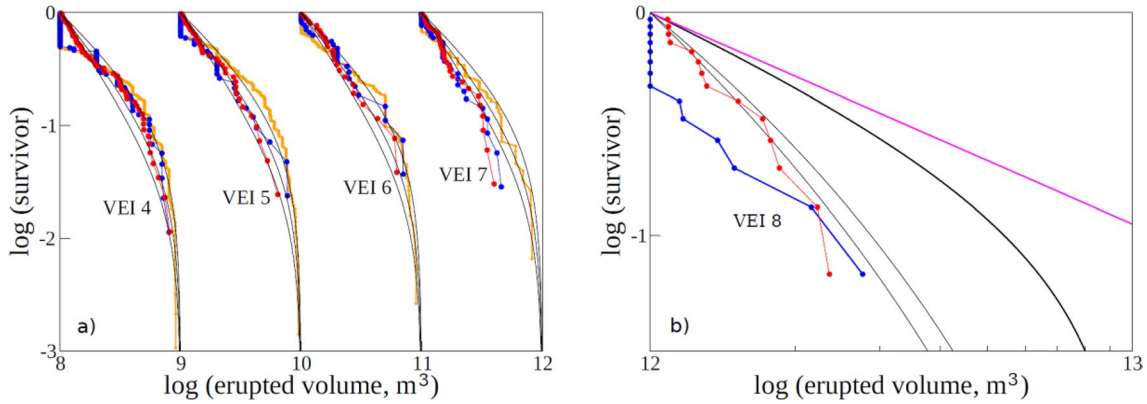

Figure 1. Distribution of eruption volumes for VEI classes ≥4 (data from the LaMEVE database; only data for which VEI ≤ $\log\left[V\left(m^3\right)\right] - 4 < \mathrm{VEI} + 1$). (a) VEI 4–7 eruptions. The blue symbols and lines are volumes from the database up to catalog completeness (Table 1); many data with same volume attribution emerge, especially in correspondence of VEI class limits. Dirtying these data with the procedure described in the text produces smoother distributions, of which the red symbols and lines represent one example outcome. Orange symbols include all data beyond catalog completeness. The thin black lines are theoretical power-law distributions, constrained within the corresponding VEI class (as for the data), with slope parameter k from right to left equal to 1.95, 2.30, and 2.60. (b) VEI 8 eruptions, blue and red symbols as in panel (a). The magenta line shows an open power-law distribution with $k = 1.95$ (as from Papale, 2018), while the thick black line is the same for a truncated power law (truncation at $\log V = 10^{13}$ m^3). The thin black lines are tapered power law distributions (see Equation 3) with $V_C = 2 \times 10^{12}$ m^3 and k from right to left equal to 2.30 and 2.60.

et al., 2010; Pyle, 2015). For a minor subset of data that relationship does not appear to hold in LaMEVE, often resulting in largely inconsistent average density of the volcanic materials required to reconcile the attributed VEI and M scales (an extensive discussion of the two scales is provided in the electronic version of Papale [2018]).

As it is discussed above, a power-law distribution bounded within imposed limits, like the limits of the observational window corresponding to each given VEI class, takes a truncated shape reflecting a more rapid approach to zero than for a power-law distribution. The distributions in Figure 1a clearly illustrate that. The blue symbols are data up to catalog completeness for the corresponding VEI class, whereas the orange symbols extend far beyond catalog completeness for the corresponding VEI class, including all data with consistent volume attribution in the database. The similarities in the overall blue and orange data distributions suggest that loss of information beyond catalog completeness only partially affect the data when the focus is on individual VEI classes, likely as a consequence of the relatively small differences in eruption volumes within each VEI class thus limited differences in terms of preservation in the geologic record. Both the blue and orange symbols show concentration at specific values in particular corresponding to VEI class limits, likely reflecting approximations in eruption volume estimates to the nearest reasonable low-digit number. That results in a sort of data binning, suggesting that volcanologists approach the difficult task of estimating eruption volumes through subsequent approximations: the VEI is order zero (thus it holds lowest uncertainty), whereas the estimated volume is order one, still implicitly binning the data as for the VEI but on a finer mesh. The consequences are substantial for statistical testing, as data binning creates artificial steps, well visible in Figure 1, that do not hold any correspondence in the distribution (e.g., Marzocchi et al., 2020; Virkar & Clauset, 2014). In similar situations a viable way forward is that of "dirtying the data", involving artificial data distribution around their nominal value so to keep the overall distribution while smoothing small-scale artifacts (e.g., van Leeuwen et al., 2019). The red symbols in Figure 1 add to the data a log-normal distributed random error: each nominal volume estimate is taken as the mean value of a log-normal distribution with assumed variance equal to 0.01 (on a log scale, with volumes expressed in cubic meters), by accepting only the part of the distribution, normalized to 1, which sits within the range of volumes consistent with the original VEI attribution, in agreement with lowest uncertainty associated with VEI attribution. Figure 2 graphically illustrates the adopted data dirtying procedure. Further justification to keep the dirtied volumes within the bounds of the original VEI attribution is provided in Section 3.2.

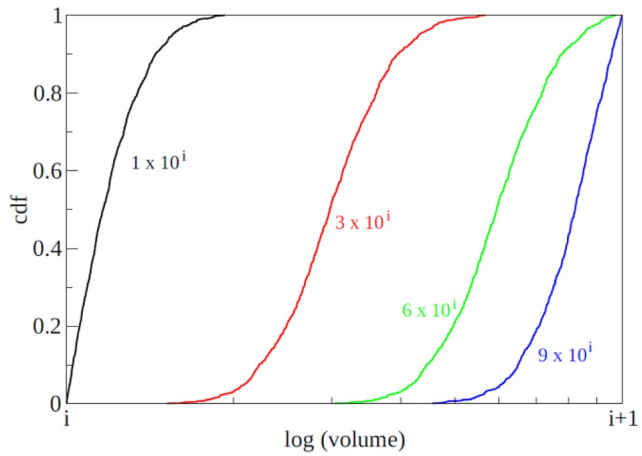

Figure 2. Examples of data dirtying for volumes with nominal value reported besides the corresponding curve.

The distribution of the dirtied red symbols in Figure 1a is similar to that of the blue (and orange) symbols, but the original large steps are smoothed down. For comparison, the three thin black lines superimposed on the distributions for each VEI class are theoretical (truncated) power-law distributions with three different values of the slope parameter corresponding (from right to left) to 1.95, 2.30, and 2.60. Overall, the correspondence between the distributions and the data (either dirtied or not) appears to reinforce the hypothesis that the continuous volume data are in fact power law distributed. Quantitative testing of that hypothesis is left to the analysis below, when a global volume distribution is derived (see Section 3.1 below).

The case for eruptions with VEI 8, considered in panel (b) of Figure 1, is peculiar, as such eruptions are located at the right extreme of the distribution where rare observations due to extremely low frequency, or possible physical limits to further increase in size, or both of them, result in frequencies significantly lower than expected from a power-law distribution. Note that the LaMEVE database (Crossweller et al., 2012) lists 27 VEI 8 eruptions (back to catalog completeness in Table 1), but only 15 of them are reported in association with an eruption volume consistent with that class. In addition to the brief discussion above on inconsistent volume attributions in the database, we acknowledge that there are specific open issues in volume (and mass) estimates for such extreme scale eruptions. As an example, a recent paper (Takarada & Hoshizumi, 2020) reports a detailed re-evaluation of one of such VEI 8 eruptions with inconsistent volume attribution, namely, the 87–89 ka ASO 4 eruption from the Aira caldera, Japan. The new estimates raise the volume from 600 km^3, as from LaMEVE, to 930–1,860 km^3, in line with its original definition as VEI 8.

Despite recent improved volume estimates such as the one for the Aira eruption, our choice is to not modify the database with new or competing volume estimates, rather, refer to it as the current state-of-the-art so to ensure full reproducibility of our results. We are aware that the database can be improved and will likely be in the future. Accordingly, new analyses reflecting major data updates will likely provide more accurate results, as it is always the case in science. We deem reference to a common database as more valuable than questionable improvements from its subjective modifications, and consider further developments of the database as a necessary major accomplishment reducing the still relevant lag of the volcano community with respect to the EU-INSPIRE principles and directive (https://inspire.ec.europa.eu/).

Figure 1b shows in blue symbols the data from the database. The same low-digit approximations and artificial binning described for VEI 4–7 eruptions clearly emerge, with more than 50% of the data corresponding to a volume of exactly 1,000 km^3. Data dirtying with the same procedure as for VEI 4–7 eruptions is shown by the red symbols. While in its very initial portion the dirtied distribution is not dissimilar from those pertaining to lower VEI eruptions, it rapidly departs from open power-law distribution (represented by the magenta line in Figure 1b, with a slope parameter close to 2). As discussed above, such a behavior is not unusual, on the contrary, it characterizes many other natural power law distributed phenomena, including earthquakes, tsunamis, and floods (Geist & Parsons, 2014). In all such cases, deciding if the deviation from the power law is just due to under-sampling of exceedingly rare events, or if it is real, for example, due to the existence of some physical limitations preventing the occurrence of events with larger size, is nontrivial. In the present case, an evaluation is done a-posteriori, requiring the arguments and analysis illustrated in Section 3 below. We anticipate here that an open power-law distribution (i.e., assuming that the quick departure from power-law distribution in Figure 1b is entirely due to under-sampling) would give us the chance to observe one VEI 9 eruption (volume > 10,000 km^3) on average every about 3 million years. Given the presently accessible geological record, and observing that we do not have notice to-date of any VEI 9 eruptions on Earth, it seems plausible that such monstrous eruptions are limited by some physical reason causing the observed deviation from open power-law distribution. Accordingly, we employ here a tapered power-law distribution, also employed for other power law distributed natural phenomena, of the form (Kagan & Schoenberg, 2001; Vere-Jones et al., 2001; Geist & Parsons, 2014):

$$S = \left(\frac{V}{V_{min}}\right)^{1-k} exp\left(\frac{V_{min} - V}{V_C}\right) \quad (3)$$

where V_C is a "corner volume" above which the distribution departs from the log-log linear trend of the power law. With respect to a truncated power law, which assumes an upper limit to the size of the events (e.g., Corral & Gonzàles, 2019; Kagan, 2002), the tapered power law is less strict as it does not prohibit very large event size above the available observations, but it significantly reduces the likelihood of their occurrence. In practical terms, truncated and tapered power laws produce the same distribution over most application cases, but the latter has the conceptual advantage that it does not imply any strong assumption about the existence of a physical size limit. The exponential term in Equation 3 becomes quickly negligible for volumes smaller than the corner volume; accordingly, Equation 3 essentially describes a power-law distribution below V_C, and a more rapidly decaying distribution above it. According to Frank (2009), an exponential term causing the PDF (and survivor) to approach zero more rapidly than for a pure power-law distribution, is the most natural and most likely form for power law-like distributions from the real world. Figure 1b reports, for illustrative purposes only, two tapered power-law distributions (thin black lines) obtained with a corner volume of 2,000 km^3 and slope parameter, from right to left, of 2.3 and 2.6. The thick black line in the same panel is instead a truncated power-law distribution, with truncation volume at 10,000 km^3.

2.2. Data Analysis for Small Scale Eruptions With VEI = 0–3

Obtaining a volume distribution for eruptions with VEI in the 0–3 range is far less straightforward than for larger VEI eruptions, for the two reasons that (i) no clear relationship emerges for the discrete frequency distribution of VEI 0–2 eruptions (Papale, 2018), and most importantly (ii) the volume of VEI 0–3 eruptions is not reported in the reference databases. In fact, the LaMEVE database, which reports volumes, deals with only large explosive eruptions with VEI 4+, while GVP (the Global Volcanism Program database of the Smithsonian Institution, https://volcano.si.edu) reports low VEI eruptions but does not include eruption volumes. The original paper introducing the VEI scale (Newhall & Self, 1982) associates a volume range to each VEI including the smallest ones. However, a quick check in the literature immediately reveals that elements other than eruption volume are largely employed when associating a VEI value to small eruptions. Clear examples include the 2014 Bardarbunga eruption at Holuraun, Iceland, with an estimated volume between 1.2 and 1.75 km^3 (Bonny et al., 2018; Dirscherl & Rossi, 2018; Coppola et al., 2019) as for explosive VEI 5 eruptions, but with assigned VEI 0 in GVP as a reflection of its quiet effusive evolution; many VEI 0–2 eruptions from the Kilauea volcano, Hawaii, with volumes up to >100 Mm^3 (Macdonald et al, 1983); and conversely, other eruptions with small volume but substantial explosive character thus assigned a VEI value of 3. The latter cases include the phreatic eruption at Mount Ontake, Japan, in 2014, with estimated volume barely reaching 1 Mm^3 (Takarada et al., 2016), as well as many dome-collapsing, small volume pyroclastic flow-forming eruptions. In substance, while there is essential (although not complete) correspondence between assigned VEI and erupted volume for explosive eruptions with VEI 4+, for smaller eruptions with VEI 0–3 the situation is much more complex, as VEI assignments appear to include considerations on the effusive versus explosive character and, likely, some degree of subjectivity. As additional examples, the May 2008 to July 2009 eruption of Mount Etna, Italy, produced an estimated volume of 68 Mm^3 (Harris et al., 2011) or 57 Mm^3 (Proietti et al., 2020) and is reported in GVP as VEI 1; while the about 6 months long 1984 eruption at the same volcano, which produced a comparable volume of 46 Mm^3 (Harris et al., 2011), is associated to VEI 3.

To derive a description of the erupted volume distribution for VEI 0–3 eruptions, we have analyzed an ample literature and constructed a database of VEI versus erupted volume (provided as Table S1). Our database is not intended to be exhaustive, what would require far more efforts similar to those that led to the LaMEVE database. Its scope within this study is that of supporting the identification of a viable relationship, if one exists, that may lead to identify on a first approximation a volume distribution for low VEI eruptions. There are many volume estimates available from the literature for low VEI eruptions, but only part of them (roughly 2/3 of those that we found) could be used here and contributed to our database, essentially because establishing a relationship with low VEI eruptions as they are reported in the GVP database is

often not possible. That reflects one other major issue with low VEI eruptions, which requires a discussion on how such eruptions are identified and classified in the database.

Many low VEI eruptions are dominantly effusive and occur at frequently erupting, open system volcanoes. Their dynamics are such that in many cases an objective definition of eruption start and end is not straightforward, implying a degree of subjectivity in even identifying individual eruptions. That is a common situation for volcanoes with activity characterized by events occurring within a sequence lasting for years or decades. An extreme case is that of the Stromboli volcano, Aeolian Islands, Italy (Calvari et al., 2012; Giudicepietro et al., 2020). After a major paroxysm in 1932, the volcano entered a state characterized by small Strombolian events, typically occurring every few to a few tens of minutes, each discharging from a few to a few hundred m^3 of scoria. This activity is punctuated by major explosions with a frequency of a few per year, up to rare paroxysmal events (four recorded in the current century) typically discharging a few tens of thousands m^3 of scoria. Identifying in the database every single explosion from the volcano as an individual eruption would be impractical if not impossible. Accordingly, the GVP database reports the entire period from 1934 to date as one single eruption, associating such a long, spurious event to VEI 2.

The extreme example above is not dissimilar from others at many other volcanoes. The entire period at Mount Etna from September 2013 to date (December 2020) is reported in GVP as just one eruption with VEI 2. However, many individual eruptions from Mount Etna are described and analyzed in the literature (and sometimes, their individual volumes are estimated) in the same period. All of such cases force us to consider the meaning of "eruptions" as they appear in the databases, and consequently, as they can be described by a distribution. The following discussion may also apply, at least partly, to some VEI 4+ eruptions, but it is reported here because it overwhelmingly impacts low VEI eruptions.

Obviously, whatever the database (for low VEI eruptions, the GVP database) means for "eruption," that is necessarily what we refer to, when using that database as a basis for statistical analysis. Accordingly, when we say that the average inter-event time of eruptions with VEI 1 is about 26 days (Table 1) we refer to VEI 1 eruptions as they are identified and cataloged in the GVP database. We do not have inter-event times of minutes like those characterizing subsequent Strombolian events at Stromboli, because those events are not classified as individual eruptions in the database. Accordingly, in associating a volume distribution to low VEI eruptions we only refer to eruptions as they are identified in the GVP database. For all those cases, some of which recalled above, for which we can find a volume estimate but cannot associate it to a defined eruption with assigned VEI in the database, we are forced to reject that volume estimate and not use it in our analysis. In this way, we are entirely consistent: the rate parameters in Table 1 describing the exponential distribution of inter-event times for each VEI class of eruptions, the frequency of the various VEI classes of eruptions, and the global, continuous volume distribution that we seek here, all make reference to the same "eruptions" as they are identified and classified in the databases.

At the end, we have found 151 eruptions with estimated volume and VEI in the 0–3 range, for which an association can be established between the eruption as it is identified in the GVP database and assigned a VEI in-there, and the corresponding volume estimate from the literature. The database, referred to here as the LV (for Low VEI) database to distinguish it from the GVP and LaMEVE databases, is reported in Table S1. It is important to remark that the sampling that led to the LV database can be considered a random one for the scopes of this study, and in particular, it can be considered independent from the eruption volume. In fact, (i) our selection is exclusively based on the availability of a volume estimate and possibility to establish a 1:1 relationship with an event listed in GVP and having a VEI assignment; and (ii) volume estimates for low VEI eruptions that can be found in the literature do not appear to bring in major biases, for example, by overweighting larger eruptions or similar; rather, the production of a volume estimate seems to have been motivated more by the availability of adequate conditions reducing uncertainties thus allowing the application and testing of novel or more refined techniques, and secondarily, by the need of quantifying a few cases which had some significant impact on people and infrastructures (not necessarily associated with particularly large volumes; e.g., the above cited, low volume Mount Ontake eruption in 2014).

Figure 3 shows the cumulative and survival volume distributions for the VEI 0–3 eruptions from the LV database. As from the discussion above, different VEI eruptions overlap to each other in terms of erupted volumes, except for the largest volumes within each VEI class. Black symbols refer to all eruptions taken

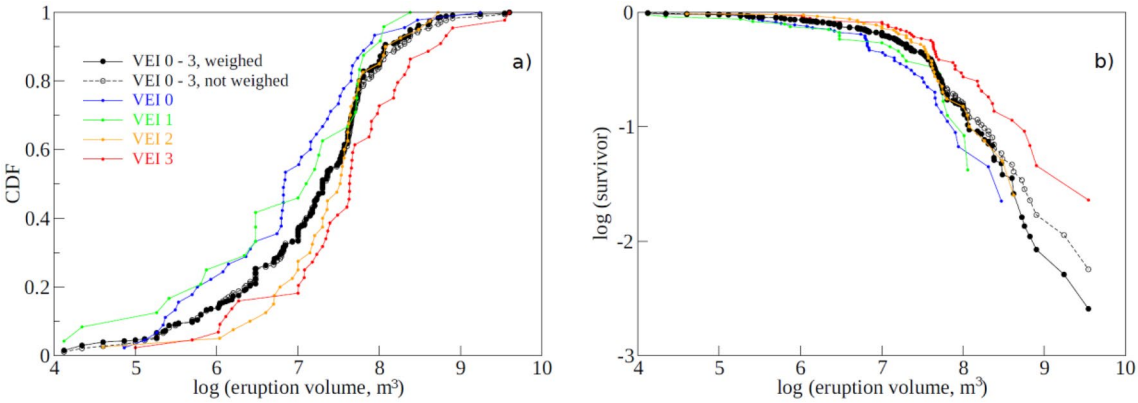

Figure 3. Distribution of eruption volumes for VEI 0–3 eruptions. Data from the database provided in Table S1 (151 data in total, VEI 0: 45, VEI 1: 24, VEI 2: 39, VEI 3: 43). (a) Cumulative distributions. (b) Complementary cumulative (survivor) distributions. Weighting for VEI 0–3 eruptions is done with reference to the corresponding class frequency from Papale (2018).

together, either as they appear in the database (open symbols and dashed line) or by weighing (solid symbols and line) each VEI eruption by the corresponding VEI frequency from Papale (2018). The very similar distributions shown by the open and solid black symbols suggest that differences between VEI frequencies and their frequency of appearance in the LV database, combined with volume distributions for eruptions corresponding to individual VEI classes (also reported in the figure), do not cause major uncertainties in overall VEI 0–3 volume distribution.

The data in Figure 3 reveal some organizations that can be exploited with the aim of defining a representative distribution. At a first sight, the trends in panel (a) (CDF) seem to suggest a distribution close to log-normal. However, the head of the distribution, on the left, is much wider than the tail of the distribution, on the right; in fact, with respect to the median value (CDF = 0.5), the left side of the distribution covers about 3.5 orders of magnitude in volume, whereas the right side covers only about two orders of magnitude, suggesting that the frequencies on the tail of the distribution decay more rapidly than for a log-normal distribution. That is better evidenced by the plot of the survivor in panel (b), showing (on a log-log scale) close to linear trends for volumes above roughly some tens of Mm^3. Such a linear trend may suggest a power-law distribution; however, the coda of log-normally distributed data may also mimic a linear trend, and clear discrimination is often a difficult issue requiring data quality, accuracy, and completeness that our LV database hardly guarantees. For the moment we do not make any assumption on the tail of the distribution, and deserve instead additional evaluations for a later moment, when we attempt to derive, from the data presented so far, an overall distribution for eruption volumes from VEI 0 to 8 (see Section 3 below).

It is worth noting that the separate distributions for VEI 0 and 1 eruptions in Figure 3 appear to be sensibly different from those of the larger eruptions with VEI 2 and 3; in fact, for a given value of the CDF in panel (a), the distributions for VEI 0 and 1 eruptions are systematically shifted toward smaller volumes with respect to the distributions for VEI 2 and 3 eruptions. Although all eruptions largely overlap in terms of volumes, that shift seems to suggest that the erupted volume maintains some relevance when attributing a VEI to also low VEI eruptions. In any case, to determine an overall distribution of eruption volumes, such differences do not require any differential treatment. In fact, the individual VEI distributions sum up in the statistically equivalent distribution provided by the black solid symbols in Figure 3; for the sake of evaluating an overall volume distribution, the four individual ones (associated to their respective frequencies) or the ensemble one (associated to the sum of the individual frequencies) provide the same information and lead to the same results. We have checked that, by repeating the entire procedure described in this paper with either one single or four separate distributions for VEI 0–3 eruptions; as expected, the results are indistinguishable.

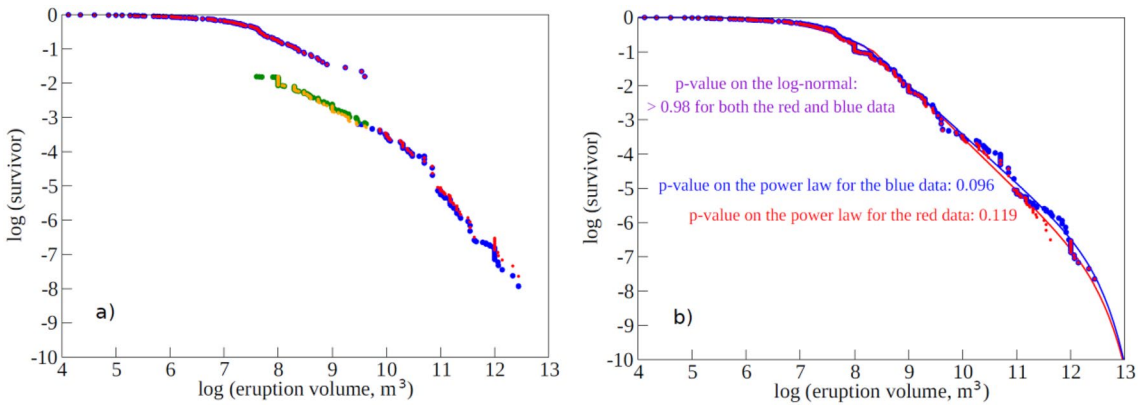

Figure 4. Global distribution of eruption volumes for VEI 0–8 eruptions. Data from the LaMEVE (VEI ≥ 4) and LV (VEI 0–3) databases. (a) Distributions obtained by maintaining separated the eruptions belonging to different VEI classes. The red and orange symbols on one side, and the blue and green symbols on the other side, refer to data with only consistent, or including inconsistent, respectively, volume attribution in the LaMEVE database. The orange symbols belong to the red distribution, with different colors used to evidence the volume range over which eruptions with VEI 4 and 5 overlap with eruptions with VEI 0–3. The green symbols have the same meaning, with reference to the blue distribution. The slopes of the two overlapping branches are significantly lower than for the eruptions with larger volumes. (b) Same as for panel (a), after reconstruction of the global distributions through Equations 5–8. For both the red and blue data sets, after an initial close-to-log-normal portion the distribution displays a unique slope that satisfactorily approximates a (tapered) power-law distribution (*p*-values reported in the figure). Red data points in panel (b) (including the orange ones in panel (a)) refer to 336 eruption cases (151 for VEI 0–3 from the LV database, the rest from the LaMEVE database: 80 for VEI 4; 36 for VEI 5; 24 for VEI 6; 31 for VEI 7; 14 for VEI 8). Blue data points in panel (b) (including the green ones in panel (a)) refer to 373 eruption cases (151 for VEI 0–3, 89 for VEI 4; 40 for VEI 5; 26 for VEI 6; 40 for VEI 7; 27 for VEI 8).

3. Data Processing

As it is illustrated above, the available eruption volume data appear to display an initial part, on the low-volume extreme, characterized by log-normal distribution of volume frequency, followed by a power-law distribution above a certain volume. Close to the large-volume extreme, in correspondence of VEI 8 super-eruptions, the observed frequency decays more rapidly than for a power law, and a tapered power-law distribution (Equation 3) appears to better describe the observations. To complete the description of the global volume distribution of subaerial eruptions, we must determine the parameters describing the distribution. This is the objective of this section.

3.1. Data Merging

The first issue is to determine the actual overall distribution of the volume data in Figures 1 and 3. Those data relate to each other through VEI frequencies determined by the rate parameters in Table 1:

$$f_i = \frac{\lambda_i}{\sum \lambda_i} \tag{4}$$

where λ_i is the rate parameter of the ith VEI class, and the sum is extended to all VEI classes. The distributions in Figure 4a are obtained by starting, for each VEI class, from the value of the cumulative frequency distribution before that class (equal to zero for VEI 0–3 eruptions), and adding a contribution from each data point equal to the frequency of the corresponding class divided by the number of data in that class. The red and orange circles in Figure 4a refer, for the VEI classes 4–8 whose data derive from the LaMEVE database, to only the data with erupted volume attribution consistent with the corresponding VEI class. The blue and green circles in Figure 4a include the data with inconsistent volume attributions (see the discussion above). For both sets of data, the orange and green symbols indicate the subset of VEI 4+ eruptions with volume in a range superimposing to that of VEI 0–3 eruptions. Such a superposition involves all of the VEI 4 and most VEI 5 eruptions. Although an overall trend is already visible for both the blue and green data points on one side, and the red and orange points on the other side, the ample superposition of volumes does not allow direct treatment of the data in Figure 4a. Instead, those data must be recast into a

distribution corresponding to continuously increasing volume, implying that contribute to the CDF from each individual eruption in the figure must be consistently re-determined with reference to volume intervals instead of VEI classes. This is done through Equations 5–8:

Volumes $< \Gamma_1 = 10^8\ \mathrm{m}^3$:

$$w_i = \sum_{j=0}^{3} \frac{n_{\Gamma_1,j}}{N_j} f_j \left[\sum_{j=1}^{3} n_{\Gamma_1,j} \right]^{-1} \tag{5}$$

Volumes $\geq \Gamma_1, < \Gamma_2 = 10^9\ \mathrm{m}^3$:

$$w_i = \left[\sum_{j=1}^{3} \frac{n_{\Gamma_1-\Gamma_2,j}}{N_j} f_j + f_{VEI4} \right] \left[\sum_{j=1}^{3} n_{\Gamma_1-\Gamma_2,j} + N_{VEI4} \right]^{-1} \tag{6}$$

Volumes $\geq \Gamma_2, \leq \Gamma_3 = 5\,x\,10^9\ \mathrm{m}^3$:

$$w_i = \left[\sum_{j=1}^{3} \frac{n_{>\Gamma_2,j}}{N_j} f_j + \frac{n_{VEI5<\Gamma_3}}{N_{VEI5}} f_{VEI5} \right] \left[\sum_{j=1}^{3} n_{>\Gamma_2,j} + n_{VEI5<\Gamma_3} \right]^{-1} \tag{7}$$

Volumes $> \Gamma_3$:

$$w_i = \frac{f_{VEIj}}{N_{VEIj}} \tag{8}$$

In the above equations, w_i is the contribution to the CDF by the ith data point; n is the number of data of the jth VEI class within any given volume interval; N is the total number of data within the jth VEI class; and f is VEI class frequencies from Equation 4. It can be verified that $CDF = \sum_{i=1}^{N_{TOT}} w_i = 1$, where N_{TOT} is the total number of volume data for each one of the red plus orange or blue plus green distributions in Figure 4a.

Once the data are re-organized through Equations 5–8 into a continuously increasing volume sequence, the distributions in Figure 4b are obtained. These distributions do not make any reference to the VEI classes from which they are derived, and they overcome the issue of overlapping volumes in Figure 4a. Accordingly, the complete set of blue + green data of panel (a) is now reported in blue in panel (b), and the reduced set of red + orange data in panel (a) is reported in red in panel (b). The overall volume trend emerges with much increased evidence from Figure 4b. Remarkably, the entire data sets (either red or blue) align close to a unique straight line above a volume of order 100 Mm^3 up to approaching the largest volumes several orders of magnitude higher. That was not visible from the raw data plots in Figure 4a, where instead the eruptions with volume between roughly some tens of Mm^3 and some tens of km^3 appeared to distribute along a significantly lower slope with respect to the higher volume data. A-posteriori, such a remarkably unique slope over so many orders of magnitudes suggests no major flaws in representativeness within the LV database.

It should be stressed that the distributions in Figure 4b involve exclusively (a) knowledge of the rate parameters and catalog completeness for the different VEI classes from Table 1, and b) use of volume data from the LaMEVE plus LV databases (with the GVP database employed in the construction of the LV database). As discussed above when introducing Figure 1a, the LaMEVE data involve some degree of artificial binning, well visible in both panels (a) and (b) of Figure 4 as a vertical sequence of data points corresponding to same eruption volume. Binning is known to be an issue when extracting a distribution from the data, as it results in artificial features which do not have any correspondence in the real distribution. Specific techniques to

test the likelihood of a power-law distribution and determine its parameters have been proposed for binned data sets (e.g., Virkar & Clauset, 2014). However, those techniques strictly require knowledge of the structure of binning, as well as non-overlapping bin classes. In the present case, we can only infer that some artificial binning affects the data as a result of order of magnitude estimates and low-digit approximations when evaluating eruption volumes, but there is not any standard or reference for eruption volume binning emerging from Figures 1a, 4a, and 4b. Instead, that binning reflects complex factors involving unknown levels of accuracy as well as a degree of subjectivity by very many different researchers who estimated eruption volumes by employing a variety of volume estimation techniques, from largely approximated to more sophisticated. The only level of data binning for which the binning structure is defined with no class overlapping, at least to an extent, relates to the VEI scale and large explosive eruptions with VEI 6 or larger. Such a limited range, summed up with failure of the simple power-law distribution for VEI 8 eruptions (Figure 1b), makes the application of the existing techniques to treat binned data distributions poorly effective even at the level of the macroscopic VEI scale classification, and definitely not applicable at the refined level of the present analysis.

The uncertainties in both eruption volume assignment and the tendency to bin the data in correspondence of the lowest digit volume estimate effectively alter, to an uncertain extent, the eruption volume distributions in Figure 4. However, the many orders of magnitude covered by data along an approximated straight line in the log-log plots in Figure 4b reduce the relevance of localized data binning, and allow the application of statistical methods appropriate for continuous data distributions. These include "dirtying" the data so as to statistically redistribute them according to some estimate of their uncertainty, as it is discussed above (see Figure 2), and require an appropriate technique to determine the parameters of the power-law distribution. Such a technique is illustrated in the following.

3.2. Data Fitting

The first relevant consideration to take into account is that establishing whether a given distribution genuinely follows a power law, and extracting the parameters of the power-law distribution, is often a difficult exercise, even with the most accurate and reliable data. As a matter of fact, that is to date an open and very active field of research (e.g., Broido & Clauset, 2019; Clauset et al., 2009). Best-fit methods based on least-squares minimization or similar are known to be inadequate for power-law distributions, as such distributions do not hold any of the required features for their application such as normally distributed errors or uncertainties (Clauset et al., 2009). The classical approach from Kolmogorov-Smirnoff (KS) statistics involves finding the distribution that minimizes the maximum distance D_{KS} between the CDF of the data and that of the distribution:

$$D_{KS} = \max_{x \geq x_{\min}} \left| CDF_{dat}\left(x\right) - CDF_{dis}\left(x\right) \right| \tag{9}$$

where the subscripts *dat* and *dis* refer to the data and the distribution, respectively, and the calculation is extended over the range of $x \geq x_{min}$ for which the power law holds. The KS approach at Equation 9 was employed in Papale (2018) to determine the slope of a power law describing the continuous volume distribution that best approximates the frequency distribution of the discrete VEI classes 3–8. KS statistics provides a very effective, powerful method to extract power law (and others) distribution parameters, often outperforming other empirical distribution function statistics such as Anderson-Darling, Cramer-von Misses, and Watson statistics (Safari et al., 2018). All of these statistics apply to data representing an outcome or observation from some unknown distribution, that the statistics allows us to determine in terms of its likelihood and corresponding parameters. However, the case in this work provides a different exercise. In fact, as it is explained above when introducing Figure 1, managing a true distribution that includes observations of large explosive eruptions is beyond the possibilities offered by the available data. Such a data set should include tens of thousands of observations to include even one single VEI 7 eruptions, and millions to include one VEI 8 eruption (see Table 1). We are several orders of magnitude far from such a database. Still, we can work it out: instead of being the result of direct observations, the distributions in Figure 4 are reconstructed from known rate parameters corresponding to individual VEI classes, as detailed above. Through Equation 4 knowledge of VEI recurrence is converted into VEI frequency, and with Equations 5–8

that is in turn transformed into volume frequencies. The key point here is that the robustness of the initial information does not necessarily deteriorate with size thus with rareness of the events, because the observational window is not the same for all events, rather, it greatly expands with increasing size thus rareness (Table 1). As a consequence, the fundamental assumption implicit in the use of Equation 9 from classical KS statistics, that the information on progressively less frequent observations on the right of the distribution is less informative of the true distribution, does not (necessarily) hold for Figure 4, as it would instead, if Figure 4 was simply a plot of the millions of observations necessary to achieve a minimum number of large eruptions from which to extract a highly uncertain estimate of their frequency.

The concepts above are illustrated in Figure 5. The data points in panel (a) of Figure 5 are obtained from assumed power-law distributions with exponent k equal to 2.50, and three different values of V_{min} (see Equation 2) so to separate the distributions in the graph and visualize them properly. From left to right, we have produced 10,000, 1,000, and 100 power law distributed data. For each distribution, the black line (barely visible close to or below the blue one) represents the true distribution from which the data are extracted. Assume now we only have the data and want to retrieve the unknown parameters of the power-law distribution. The blue lines represent the results obtained from KS statistics. In all cases, the true distribution is retrieved with high accuracy, especially when the data points are many. In fact, those data genuinely represent power law distributed outcomes, and in such a case KS statistics is very effective in quantifying the underlying distribution. It is relevant, for the present discussion, to note how KS works: it gives highest relevance to the central part of the distribution (left portion of the distributions in Figure 5a), while completely neglecting the shape of the low-frequency tail according to its poor representativeness compared to the bulk of the distribution. That is fully appropriate for the case in Figure 5a, where the data are extracted from a distribution, therefore, the low-frequency tail is poorly representative of the distribution itself.

Panels (b) and (c) in Figure 5 illustrate the case for the eruption volume data in Figure 4b. The data points in these panels are two statistically equivalent outcomes from the red set of data in Figure 4b, after dirtying as from Figure 2. The blue lines represent the corresponding distributions obtained through KS statistics (actually, the procedure involves finding consistent distributions for the initial log-normal and subsequent power-law sections of the distribution; that procedure, employed for the fits in Figures 5b and 5c, is illustrated below). Different from the case in panel a, now we want the fit to not give the right side of the distribution minimum relevance; after all, the information on, for example, eruptions with volume larger than 1,000 km^3 summarizes 2.6 million years of reconstructed Earth volcanism, and is not necessarily less robust than the information on eruption volumes between, say, 100 Mm^3 and 1 km^3, that are much more frequent but also much less preserved, thus much less confidently known from the geological record. From the KS fits (blue lines) in panels 5(b and c), it clearly emerges that by giving maximum relevance to the central portion of the distribution, KS fails to adequately describe the tails, ending up in frequencies for the large explosive eruptions that significantly differ, by up to more than one order of magnitude, from those expected by the corresponding rate parameters.

To deal with the distributions in Figure 4b, and obtain appropriate power-law parameters, we must therefore rely on a different approach that equally weighs the entire power-law distribution including its tail. This is done here by referring to a distance different from that used for KS statistics, and corresponding to the relative distance computed along the survivor function:

$$D_{RDS} = \max_{x \geq x_{\min}} \frac{\left|S_{dat}\left(x\right) - S_{dis}\left(x\right)\right|}{S_{dis}\left(x\right)} = \max_{x \geq x_{\min}} \frac{\left|CDF_{dat}\left(x\right) - CDF_{dis}\left(x\right)\right|}{1 - CDF_{dis}\left(x\right)} = \frac{D_{KS}}{1 - CDF_{dis}\left(x\right)} \tag{10}$$

where the subscript RDS stands for "Relative Distance on Survivors," and S indicates the survivor function. By referring to a relative instead of absolute distance, Equation 10 does not weigh off the low-frequency portion of the distribution where absolute distances are orders of magnitude lower than those on its high-frequency portion; and by normalizing the distances on the value of the survivor, very small distances on the tail of the distribution are as important as correspondingly larger distances on the bulk of the distribution. As a matter of fact, by equally weighing the distribution all over its length, Equation 10 provides an optimal criterion for a situation like the one in Figure 4b, where the frequencies corresponding to the different sectors of the distribution do not derive from counting the corresponding observations, but they

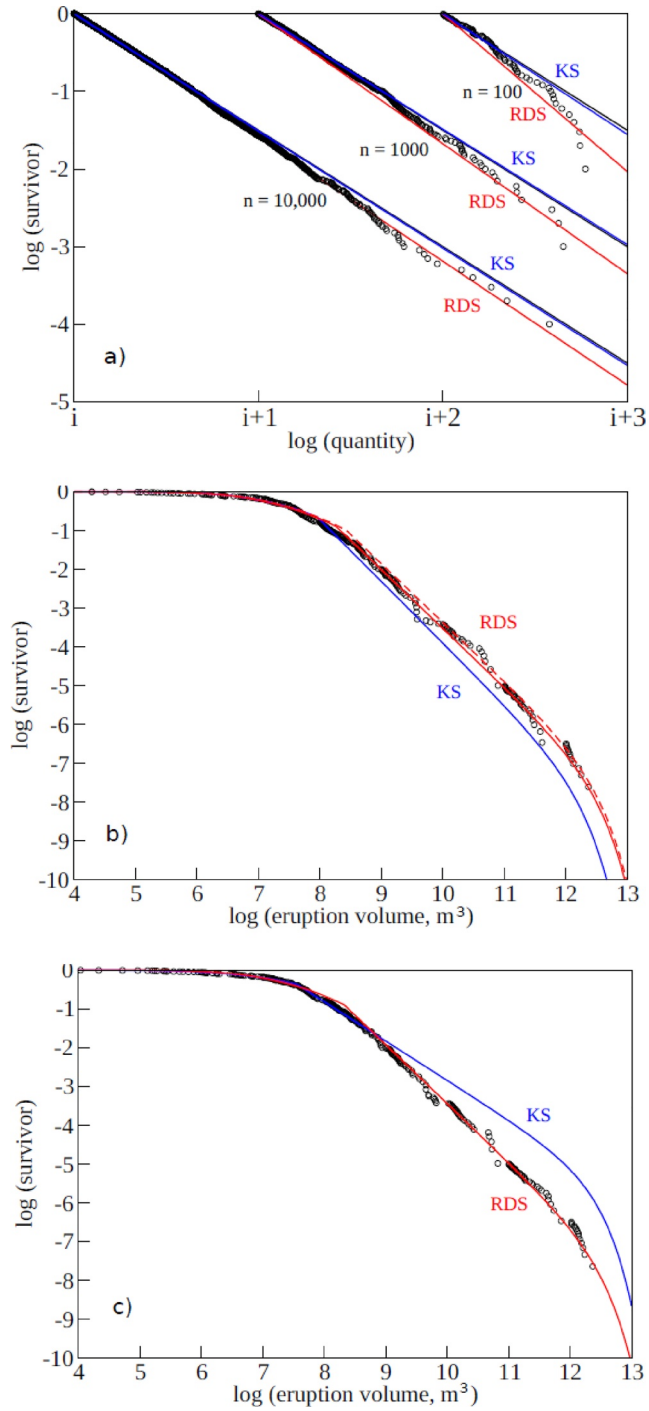

Figure 5. Comparison between the Kolmogorov-Smirnov (KS, Equation 9) and Relative Distance on Survivors (RDS, Equation 10) methods to constrain the parameters of a power-law distribution. (a) The sample represented by the symbols has been obtained from the power-law distribution represented by the black line (nearly overlapped by the blue one). The KS method (blue) reconstructs the power law nearly exactly, whereas the RDS method (red), which equally weighs the bulk and the tail of the distribution, does not. (b) and (c) Two examples of applications of the KS and RDS method to the problem in this work, where the data do not represent individual observations, and the tail of the distribution is as important as the bulk of the distribution. In such a case the KS method results in distributions that largely diverge from the tail of the distribution, whereas the RDS method provides a satisfactory approximation all over the distributions. Panel (b) includes a comparison between the use of D_{RDS} at Equation 10 (dashed red line), and D^*_{RDS} at Equations 11 and 12 (solid red line). The two quantities provide similar results, but D^*_{RDS} provides a more balanced weighting of data points outlying below and above the distribution.

are independently known. From Equations 9 and 10, it can be noted that the fundamental property of D_{KS} to converge to 0 when the number of observations tends to infinity is also a property of D_{RDS}.

The capability of the quantity D_{RDS} to properly deal with the situation represented by the eruption volume data in Figure 4b is illustrated in Figure 5, where the red lines are obtained by D_{RDS} minimization (and the blue lines, as explained above, by D_{KS} minimization). For a problem like the one in panel (a) of Figure 5, where the data do appear with a frequency dictated by the underlying power-law distribution, D_{RDS} is clearly not adequate, since weighing the tail as the bulk of the distribution is not legitimate. In such a case, classical statistics from KS is far superior. On the contrary, for the present case represented in panels (b) and (c), minimization of D_{RDS} adequately accounts for the distribution all over its entire length. A slight modification of D_{RDS} provides more balanced account of data outlying above or below the distribution:

$$D^*_{RDS} = \max_{x \geq x_{min}} \left[\frac{1}{2} \left(\frac{\left| S_{dat}(x) - S_{dis}(x) \right|}{S_{dis}(x)} + \frac{\left| S_{dat}(x) - S_{dis}(x) \right|}{S_{dat}(x)} \right) \right] \tag{11}$$

which corresponds to:

$$D^*_{RDS} = \max_{x \geq x_{min}} \frac{\left| S^2_{dat}(x) - S^2_{dis}(x) \right|}{S_{dis}(x) S_{dat}(x)} \tag{12}$$

The solid red lines in Figures 5b and 5c are obtained by minimization of D^*_{RDS} at Equation 12, while the dashed red line in Figure 5b is obtained by minimizing the quantity D_{RDS} at Equation 10.

With the criterion provided at Equation 12, the volume data in Figure 4b result in the distributions represented by the solid lines in the same figure, with the red line referring to the red data points (reduced data set, see above), and the blue line referring to the blue data points (complete data set). In both cases, the p-value has been evaluated through the statistics represented by the quantity D^*_{RDS}: 1,000 samples each with the same size as the data have been produced from the distribution, with the p-value quantity representing the proportion of samples for which the statistics is at least as bad as for the data. The results are reported in Figure 4b for both the red and blue data and distributions, showing that in both cases the null hypothesis that the data are drawn from the distribution cannot be rejected at usual confidence thresholds (e.g., 5%). That is a particularly relevant result as it is obtained over data that involve artificial, unstructured binning, as it is largely discussed above. On the other hand, the wide range over which the power-law distribution holds reduces the relevance of localized artifacts, as it was anticipated above. The algorithm to extract in a single, consistent procedure the power law and log-normal parameters describing the entire volume distribution in Figure 4b is the following:

1. Refer to a tapered power law with survivor function expressed as:

$$S = \left(\frac{V}{V_0} \right)^{1-k} exp\left(\frac{V_0 - V}{V_C} \right) \tag{13}$$

   where V_0 is the intercept of the power law on the horizontal axis corresponding to a frequency of 1. Note that such a power law does not reflect any real distribution, as any power-law distribution is necessarily truncated on the left thus it never gets to a frequency of 1. In fact, Equation 13 describes the distribution only above V_{min}. However, writing in terms of a purely mathematically defined intercept V_0 ensures that the frequency at V_{min} computed on the log-normal and on the power law is the same, simplifying the mathematical description of the whole distribution. Use of V_0 in the exponential term does not modify the distribution, as the exponential term is only relevant when V approaches V_C which is many orders of magnitudes larger than either V_0 or V_{min}.
2. Find the k, V_0, V_C triplet that minimizes D^*_{RDS} above the minimum volume V_{min} corresponding to the intersection between the power law and log-normal distributions.

3. For each k, V_0, V_C triplet under consideration within the minimization procedure, the log-normal distribution is found for the data with $V < V_{min}$, by scanning a wide range of candidate mean values μ, each time with variance σ^2 from maximum likelihood estimator:

$$\sigma^2 = \frac{\sum_{i=1}^{K}\left(\log V - \mu\right)^2}{K} \tag{14}$$

where the sum is extended to the K data with volume less than current V_{min}. The pair of μ, σ describing the log-normal distribution for the K data with $V < V_{min}$ is determined by KS statistics (minimization of D_{KS} at Equation 9).

The survivor function $S = 1 - CDF$ for the distribution of volcanic eruption volumes is therefore defined by the following:

$$for\ V \le V_{min}: \qquad S = \frac{1}{2}\left[1 - erf\left(\frac{\log V - \mu}{\sigma\sqrt{2}}\right)\right] \tag{15}$$

$$for\ V \ge V_{min}: \qquad Eq.(13) \tag{16}$$

Three-variable (k, V_0, V_C) minimization of D^*_{RDS} at point (2) is carried out through the Powell's direction-set method (Press et al., 1992), appropriate for the present problem as it does not require knowledge of the derivatives of the function being minimized. To prevent convergence toward secondary minima, which are found to exist in the vicinity of any initial guess of V_0, we perform the minimization over a sufficiently ample range of guessed V_0 values, which are scanned with progressively decreasing step until finding the triplet (k, V_0, V_C) which results in the absolute minimum value of D^*_{RDS}.

The procedure above has been employed to determine the distributions corresponding to the red and blue lines in Figure 4b, which refer to the data as they appear in the database, processed with rate parameters corresponding to the mean value for each VEI class (Table 1) to move from the representation in Figure 4a to the one in Figure 4b through the set of Equations 5–8. As it is largely discussed above, that procedure does not remove the effects of data binning, still clearly visible from the distributions of data in Figure 4b. On one side, the results of the p-value analysis reported in Figure 4b suggest that a power-law distribution above a given V_{min} is a statistically robust assumption even in the presence of unstructured, localized data binning. On the other side, the artificial features introduced by data binning, and the uncertainties they imply on eruption volume estimates, are expected to affect the computed distributions, to an extent that must be estimated. To do so, we further elaborate the procedure above:

a) Randomly select a set of λ_i values for use in Equation 4, according to normally distributed uncertainty of rate parameters as from Table 1.
b) "Dirty" the data as in Figure 2.
c) Execute points 1 to 3 above. Store the results for statistical analysis, then repeat from point a) until a sufficiently large number of statistically equivalent distributions is obtained.

The procedure above has been applied to the reduced set of data represented by the red symbols in Figure 4b. In fact, the blue data points, which include volume estimates inconsistent with the assigned VEI class (and often inconsistent with the corresponding eruption mass estimates), bring about substantially increased uncertainties that impact the database to a significant extent. Again, it should be recalled what is the meaning of the data points in Figure 4: they do not just represent individual observations, rather, they are a distribution reconstructed by means of the rate parameters for each VEI class. As a consequence, a few inconsistent data have a significant effect on the distribution. In fact, a few volume data corresponding to, for example, VEI 4 but having a volume 10 times larger than expected for VEI 4 eruptions, result in significantly increased overall frequency of such larger eruptions when the overall distribution is reconstructed through Equations 5–8. The effects are visible when comparing the distributions of the red and the blue data points in Figure 4b, with the latter showing significantly larger oscillations around a common trend.

Table 2
Parameters of the Global Eruption Volume Distribution

Parameter	Mean value	5th percentile	25th percentile	50th percentile	75th percentile	95th percentile
k	2.491	2.416	2.461	2.491	2.522	2.567
V_0, **Mm^3**	46.09	30.43	38.23	45.29	53.96	64.17
$V_{\min}$, **Mm^3**	169.0	87.56	119.9	158.3	208.8	270.1
V_C, **km^3**	1,973.7	1,727.2	2,003.3	2,004.6	2,006.8	2,035.7
μ, **log(m^3)**	7.266	7.229	7.250	7.267	7.282	7.304
σ, **log(m^3)**	0.916	0.897	0.907	0.914	0.924	0.937
$CDF_{V_{\min}}$	0.843	0.772	0.813	0.847	0.873	0.895
$D_{KS,\ln}$	0.402	0.250	0.305	0.364	0.490	0.643
$D^*_{RDS,pl}$	0.730	0.520	0.620	0.721	0.822	0.983

For each parameter, the variance is given by the diagonal of the covariance matrix in Table 3. The quantities that are truly obtained by minimization correspond to the power-law parameters **k**, V_0, and V_C, and the log-normal parameters μ and σ. All other parameters follow thereby, and are reported here to evidence their variability. Similarly, maximum distances on the log-normal and power-law sections of the distribution are reported to evidence their variability.

Although a statistically significant *p*-value is still obtained, the larger overall uncertainties associated with the blue data points are expected to result in less confident final estimates. For that reason, only the more confident, more internally consistent red data points are used for the analysis from here on.

The arguments above provide additional justification to the employed dirtying procedure exemplified in Figure 2, whereby each volume estimate is dirtied without trespassing the volume boundaries of the corresponding VEI class. One alternative possibility would consist in arbitrarily changing the VEI attribution when the dirtying procedure results in a VEI class jump. However, we prefer not to change VEI attributions, according to the view, discussed above, that VEI estimates represent a more robust categorization with lower associated uncertainty, with respect to generally more uncertain volume estimates. Translated in practice, that means to assume a view like the following one: when volcanologists assign, for example, a volume of exactly 10^9 m^3 to an eruption with VEI 5 (as it is often the case, see Figure 1), they are meant to

Table 3
Covariance-Correlation Matrix

Covariance-correlation matrix

	k	V_0, **m^3**	V_{min}, **m^3**	V_C, **m^3**	μ, **log(m^3)**	σ, **log(m^3)**	$CDF_{V_{min}}$	$D_{KS,ln}$	$D^*_{RDS,pl}$
k	2.13×10^{-3}	0.896	0.839	0.018	0.094	-0.015	0.788	0.748	0.143
V_0, **m^3**	4.22×10^{5}	1.04×10^{14}	0.984	-0.045	0.109	0.053	0.893	0.877	-0.003
V_{min}, **m^3**	2.25×10^{6}	5.81×10^{14}	3.36×10^{15}	-0.064	-0.026	0.032	0.923	0.850	-0.029
V_C, **m^3**	7.96×10^{7}	-4.47×10^{16}	-3.55×10^{17}	9.31×10^{21}	0.042	0.153	-0.130	-0.001	-0.017
μ, **log(m^3)**	9.78×10^{-5}	2.48×10^{4}	-3.34×10^{4}	9.15×10^{7}	5.03×10^{-4}	0.299	-0.219	0.238	0.059
σ, **log(m^3)**	-1.07×10^{-5}	8.11×10^{3}	2.75×10^{4}	2.21×10^{8}	1.00×10^{-4}	2.24×10^{-4}	-0.305	0.215	0.205
$CDF_{V_{min}}$	1.62×10^{-3}	4.04×10^{5}	2.37×10^{6}	-5.57×10^{8}	-2.18×10^{-4}	-2.03×10^{-4}	1.97×10^{-3}	0.685	-0.135
$D_{KS,ln}$	4.38×10^{-3}	1.13×10^{6}	6.24×10^{6}	-7.03×10^{6}	6.77×10^{-4}	4.08×10^{-4}	3.85×10^{-3}	1.60×10^{-2}	0.043
$D^*_{RDS,pl}$	9.63×10^{-4}	-3.78×10^{3}	-2.45×10^{5}	-2.45×10^{8}	1.91×10^{-4}	4.45×10^{-4}	-8.69×10^{-4}	7.84×10^{-4}	2.11×10^{-2}

Covariance and correlation are reported below and above, respectively, the diagonal step-wise line. $D_{KS,\ln}$ and $D^*_{RDS,pl}$ are maximum distances at Equations 9 and 12, respectively, for the log-normal (ln) and power law (pl) sections of the distribution. $CDF_{V_{\min}}$ is the value of the CDF at $V_{\min}$. The values on the diagonal of the matrix correspond to variances. The quantities that are truly obtained by minimization correspond to the power-law parameters **k**, V_0, and V_C, and the log-normal parameters μ and σ. All other parameters follow thereby, and are reported here to evidence their variability as well as their mutual dependencies. Similarly, maximum distances on the log-normal and power-law sections of the distribution are reported to evidence their variability as well as their relationships with all other quantities.

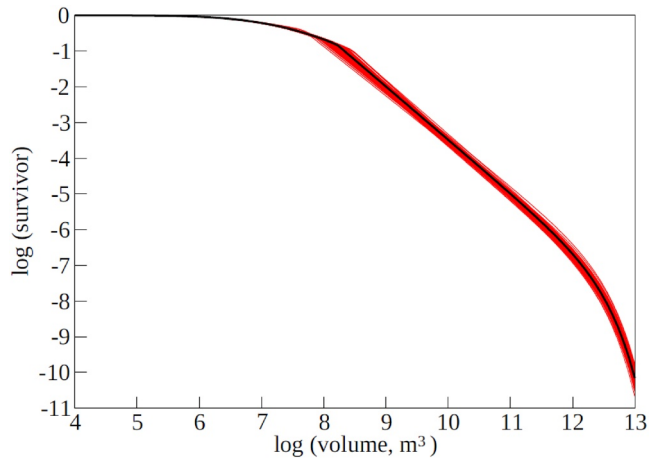

Figure 6. Ensemble of 100 randomly selected global distributions of eruption volumes, out of the 1,000 obtained here (see text). The thick black line is the distribution obtained by employing mean values from Table 2.

refer to *at least* that volume, likely not too much above it; while if instead they assign a volume of 8×10^8 m^3 (VEI 4), they imply *less than* 1 km^3, by some quantity likely not too different from 200 Mm^3. That is, in fact, the ultimate meaning of the dirtying procedure illustrated in Figure 2.

The procedure described at points (a) to (c) above, and embracing the one at points (1)–(3) above, has been applied with 1,000 repetitions so to obtain 1,000 statistically equivalent estimates of the power-law parameters k, V_0 (or V_{min}), V_C, and of the log-normal parameters μ and σ, describing the overall distribution of eruption volumes for subaerial volcanism on Earth. The results are illustrated in the following section.

4. Global Volume Distribution for Subaerial Volcanism

Table 2 reports the mean value and percentiles of the parameters describing the global distribution of eruption volumes, while Table 3 shows the computed covariance and correlation between parameters (note that the true parameters determined through the above minimization procedure are k, V_0, V_C, μ and σ, while the other quantities reported in Tables 2 and 3 are either by-product quantities (V_{min} and $CDF_{V_{min}}$, the latter representing the value of the total cumulative distribution function at V_{min}) or they are the functions being minimized ($D_{KS,ln}$ and $D^*_{RDS,pl}$ for the log-normal and power-law sections of the distribution).

Figure 6 shows a random set of 100 global distributions, out of the 1,000 determined here and used for the statistics reported in Tables 2 and 3. The uncertainty is much more significant along the power-law section than for the log-normal initial section of the distribution, largely reflecting the uncertainty in the identification of V_{min}, that is, the eruption volume threshold separating the log-normal from the power-law sections of the distribution. The scatter is however limited, and the overall distribution is well delineated. Tapering of the power-law distribution results in an expected average frequency of about 7×10^{-11} for eruptions with VEI 9 or larger. Combined with the rate parameters in Table 1, that results in the expectation of one VEI 9+ eruption every about 360 million years, consistent with the to-date lack of any known VEI 9 eruption in the geological record. As it was anticipated when introducing the tapered power law Equation 3, compared to the about 3 million years resulting from the assumption that the deviation from open power law at the extreme right tail of the distribution entirely reflects under-sampling, that provides a-posteriori justification to the assumption of some right-end truncation and use of tapering.

The mean (and median) value of the power-law slope coefficient k, reported in Table 2, turns out to be equal to 2.491, with relatively little variability (the 5th and 95th percentiles are within about 3% of the mean value). Figure 7a shows that the fitted k values are normally distributed (the variance of k as well as of any other quantity in Table 2 is reported in the corresponding value along the diagonal of the covariance matrix in Table 3). The k value determined here for the continuous eruption volume distribution is sensibly higher than the value of 1.95 obtained in Papale (2018) for the distribution of the VEI classes. A look back at Figure 4 explains why. In the unprocessed distribution in Figure 4a, the low VEI classes 3–5 along the power law distribute according to a lower slope than for the higher VEI classes in the same figure or for the processed distribution in Figure 4b. Thus, the KS estimator employed in Papale (2018), which over-weighs the high-frequency classes with respect to the low-frequency ones as largely discussed above, results in low value of k. It is only after processing the volume distribution data through Equations 5–8 that a consistent continuous distribution of eruption volumes emerges (Figure 4b), showing that the low slopes in the volume range roughly between 10^7 and 10^{10} m^3 is an artifact due to superposition of volumes for eruptions within the same volume range but attributed to different VEI classes.

Figure 7b shows that the distribution of V_0, the intercept of the power law with the frequency 1 axis, is also roughly normally distributed, but with a spread larger than for k (in this case, the 5th and 95th percentiles are within 40% of the mean value). k and V_0 are positively correlated (Table 3 and Figure 7c): in fact, a larger

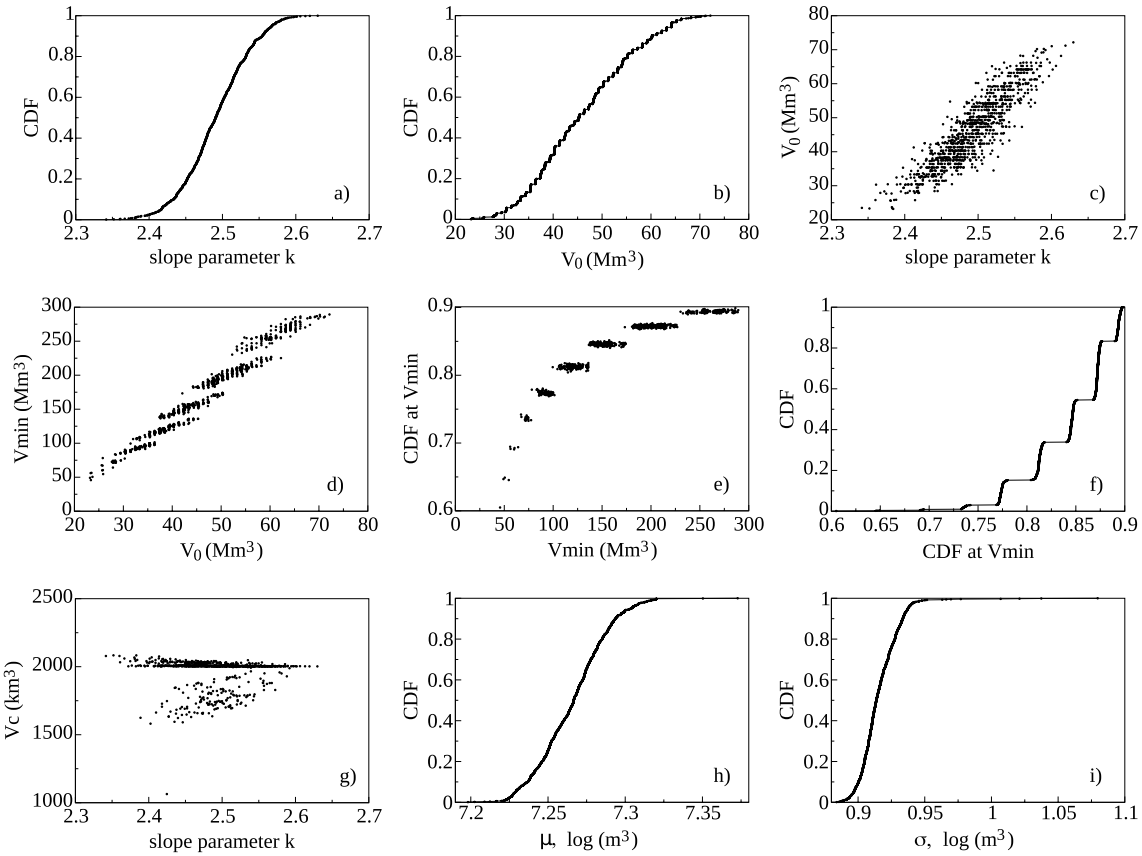

Figure 7. Parameters of the global eruption volume distribution and their relationships. (a) Distribution of the slope parameter k; (b) distribution of V_0, the intercept of the power-law distribution with the frequency 1 axis (Equation 13); (c) relationship between the fitted values of V_0 and k (1,000 data points); (d) relationship between the fitted values of V_{min} and V_0 (1,000 data points); (e) relationships between V_{min} and the CDF of the global distribution computed at V_{min} (1,000 data points); (f) distribution of the CDF of the global distribution computed at V_{min} (1,000 data points); (g) relationship between the fitted values of V_C and k (1,000 data points); (h) distribution of the mean value of the log-normal distribution; (i) distribution of the standard deviation of the log-normal distribution. CDF, cumulative distribution function.

value of V_0 implies a shift to the right of the initial part of the power-law distribution, which is compensated by a higher slope.

The value of V_{min} is of particular relevance, as that represents the lower cutoff for the power-law distribution, therefore, it may embed some physical meaning for the occurrence and dynamics of volcanic eruptions. As for V_0, and basically for the same reason, V_{min} is also positively correlated to k (Table 3). However, the relationship between V_0 and V_{min} is a bit more complex than expected (Figure 7d), and so are the relationship between V_{min} and $CDF_{V_{min}}$, and the distribution of $CDF_{V_{min}}$ (Figures 7e and 7f). In all such cases, the points in the graphs reveal the existence of hiatuses. In particular, the value of the CDF (or its complementary S) at the intersection between the log-normal and power law distributed sections, $CDF_{V_{min}}$, does not take continuous values, rather, it evolves step-wise, with the width of each hiatus increasing with decreasing $CDF_{V_{min}}$. Such peculiar trend appears to relate to the methodology set up to obtain the global distribution parameters. In fact, the fitting procedure described above treats together the power law and log-normal sections of the distribution to maintain consistency in the overall distribution. Within a unique fitting procedure, all of the quantities in Tables 2 and 3 are determined altogether, including the limit between the two sections represented by V_{min} and its corresponding $CDF_{V_{min}}$. Looking at Figure 7, the internal functioning of the global fitting procedure can be better understood: if the distribution of data is such as to require a larger k, a larger V_0 is generally implied (Figure 7c). Further increase of k thus of V_0 also causes progressive shift of V_{min} toward larger values (Figure 7d), thus assigning more data to the log-normal section of the distribution. Minimization of the two quantities $D_{KS,ln}$ and $D^*_{RDS,pl}$ occurs in such a way that discrete pockets of data are

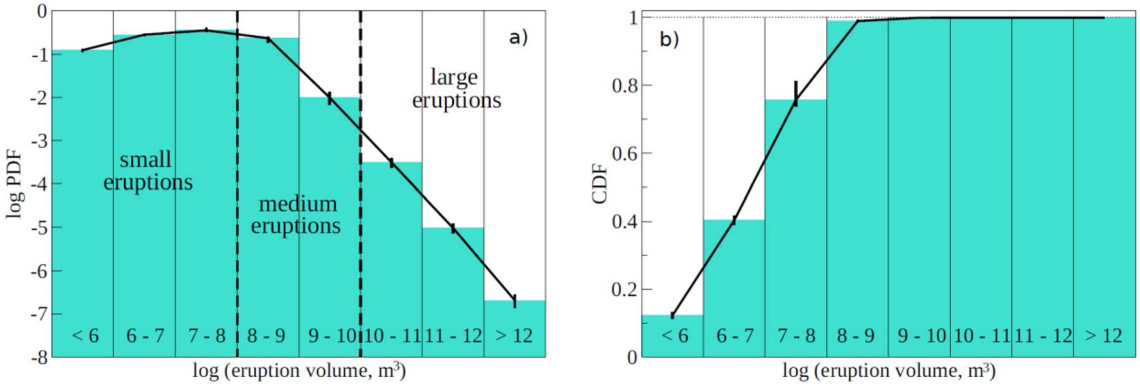

Figure 8. Discrete probabilities for classes of eruption volumes. (a) PDF. (b) CDF. The error bars refer to the 5th and 95th percentiles reported in Table 4. CDF, cumulative distribution function; PDF, probability density function.

assigned to either section of the distribution, producing the discrete jumps in $CDF_{V_{min}}$ in Figures 7e and 7f and the peculiar trend of V_{min} versus V_0 in Figure 7d. Thus, the slope within each one of the discrete batches of points in Figure 7d represents the effective dependence of V_{min} on V_0, rather than the steeper overall trend resulting from the ensemble of all batches.

Different from V_0 (and V_{min}), the corner volume V_C determining the effect of tapering (Equation 3) on the extreme tail of the power-law distribution is largely independent of the slope k as well as from any other distribution parameter (Figure 7g and Table 3). About 90% of the values of V_C are close to 2,000 km^3, with the remaining values defining a cloud extending down to about 1,600 km^3 (with one single value of about 1,060 km^3). The low values likely reflect particular combinations of data from the dirtying procedure, requiring to move back the corner volume so to allow more pronounced curvature of the distribution. The incidence of such cases is however limited, and both the mean and the median of the distribution are close to 2,000 km^3 (Table 2). Noticeably, that is close to the volume of the largest known eruption on Earth (the 74 ka Young Tuff eruption from Toba, Indonesia, 2,800 km^3), with only one other known eruption (the 2.133 Ma Huckleberry Ridge Tuff eruption from Yellowstone, US, 2,160 km^3) overcoming the mean value of V_C in Table 2 (see Figure 1b).

Finally, Figures 7h and 7i show that the mean and standard deviation of the log-normal section of the distribution are about normally distributed with very small variance (Tables 2 and 3), producing very minor variabilities in the ensemble distributions in Figure 6.

It looks relevant to use the ensemble distributions in Figure 6 to determine the probability density function for discrete classes of eruption volume. Such a distribution releases the problems related to volume overlapping when considering eruption categorization into VEI classes as in Papale (2018), largely described above. Figure 8 shows the corresponding discrete PDF and CDF. The discrete PDF values are also reported in Table 4. Note that for the large eruptions with volume ≥ 10 km^3 (VEI ≥ 6), for which the volume classes in Figure 8 correspond to VEI classes, the PDF values in the figure and in Table 4 are fully consistent with the corresponding absolute probabilities of occurrence provided in Papale and Marzocchi (2019) (that can be verified by employing the rate parameters for such high VEI classes from Table 1).

Table 4
Probability Density Function (PDF) for Discrete Eruption Volume Classes

Volume class (log, m^3)	PDF (mean)	PDF (5th percentile)	PDF (95th percentile)
<6	0.1232	0.1137	0.1329
6–7	0.2806	0.2742	0.2863
7–8	0.3537	0.3380	0.4033
8–9	0.2325	0.1796	0.2509
9–10	9.79×10^{-3}	6.65×10^{-3}	1.34×10^{-2}
10–11	3.11×10^{-4}	2.35×10^{-4}	3.89×10^{-4}
11–12	9.70×10^{-6}	7.31×10^{-6}	1.22×10^{-5}
>12	2.05×10^{-7}	1.39×10^{-7}	2.81×10^{-7}

The single most abundant volume class corresponds to volumes between 10 and 100 Mm^3 (35.4% of the total), followed by the 1–10 Mm^3 class (28.1%), the 100 Mm^3–1 km^3 class (23.2%), and the smallest volume class gathering all eruptions with volume < 1 Mm^3 (12.3%). Together, these small-medium eruptions with volume ≤ 1 km^3 constitute about 99% of the subaerial volcanism on Earth. The medium to extremely large eruptions with volumes between 100 Mm^3 to >1000 km^3 still align on a power law when gathered into discrete volume classes. Note that gathering into

VEI classes suggested instead that a power law (with remarkably lower slope) holds from volumes as low as 10 Mm^3 (Papale, 2018). The globally impacting eruptions with VEI ≥ 6 or volume ≥ 10 km^3 constitute only about 0.03% of the total, and the super-eruptions with volume ≥ 1,000 km^3 make up only about 0.00002% of the total subaerial volcanic eruptions on Earth.

5. Discussion and Conclusions

5.1. General Aspects

The analysis presented in this paper rests on the following basic ingredients:

- the data, coming from (a) the LaMEVE database for eruptions with VEI ≥ 4, (b) the GVP database, and (c) our LV database (provided in Table S1) for the volume of eruptions with VEI ≤ 3; and
- the rate parameters for the exponential distribution of inter-event times and catalog completeness for different VEI eruptions, reported in Table 1.

On that basis, we have constrained the shape and parameters of the continuous distribution of eruption volumes from subaerial volcanism on Earth. That distribution consists of an initial log-normal section with mean and standard deviation (expressing volumes in m^3) respectively equal to 7.266 [7.229, 7.304] and 0.916 [0.897, 0.937] (mean values, while numbers within brackets are the 5th and 95th percentiles, respectively; see Table 2), up to an estimated volume of 169 [88, 270] Mm^3 making up 84.3 [77.2, 89.5] % of the volume erupted from global subaerial volcanism, above which the distribution follows a power law with slope of 2.491 [2.416, 2.567] and right end tapering described by a corner volume of 1973.7 [1727.2, 2035.7] km^3.

As this study shows, obtaining a continuous volume distribution was not trivial, but the efforts are more than compensated by the achieved complete analytical description of the global eruption volume distribution at Equations 15 and 16 with parameters in Table 2. The present continuous volume distribution is far superior to the VEI-based distribution in Papale (2018), for the main reasons that:

- the VEI-based distribution is necessarily discrete, thus it involves loss of information with respect to the continuous volume distribution, resulting in less accurate estimate of distribution parameters. As an example, the minimum threshold above which the power-law distribution holds could be best placed at the boundary between the VEI classes 2 and 3 in Papale (2018), nothing comparable with the present estimate involving a mean volume for V_{min} with a variance and quantiles (Tables 2 and 3). Similar loss of information due to discretization also affected the estimate of the slope parameter in the discrete VEI representation, although the elements at the following point are even more critical in establishing the superiority of the present slope estimate. Finally, loss of information due to discretization in the VEI representation did not allow the characterization of the extreme tail of the distribution, where more rapid decay with respect to a power law is now recognized and quantified; and
- the VEI scale mixes up different elements and it involves a degree of subjectivity in its determination, resulting in ample overlapping of physical quantities such as the erupted volume (and mass). As a consequence, the slope of the power-law distribution in correspondence of the overlapping classes appears to decrease as it is abundantly illustrated above (see Figure 4), resulting in a significantly lower slope in the discrete VEI representation (1.91 as from Papale [2018], compared to 2.49 estimated here for the continuous volume distribution).

The discrete volume distribution in Figure 8 is fully consistent with the continuous one in Figure 7, thus it is totally different from the previous VEI-based distribution. Being a discrete representation, it loses information with respect to the continuous one, as it is seen by the missing features at the extreme tail of the distribution. It is however useful, as it provides a more immediate picture of the incidence of volcanic eruptions having different sizes.

5.2. Power-Law Section of the Global Volume Distribution

Power law distributions emerge in complex systems characterized by highly nonlinear dynamics with many degrees of freedom (e.g., Bak et al., 1987; Bak, 1997; Marković & Gros, 2014). Such phenomena invariably have a lower cutoff to the power-law distribution, as power laws diverge when approaching the left bound-

ary of the distribution. In the present case, the cutoff volume corresponds to V_{min}, or the minimum volume above which the power-law distribution holds. Above V_{min} the distribution of the global subaerial volcanism appears therefore to display the properties of power law – distributed phenomena, such as scale invariance and self-similarity: the shape of the distribution is always the same whatever volume interval we examine (as long as that volume interval is sufficiently far from the corner volume V_C; the extreme tail of the distribution is discussed below). Such classes of phenomena are often related to the emergence of self-organized criticality, or SOC, proposed by Bak et al. (1987) and since then invoked for a growing number of natural (largely including geophysical) as well as man-made systems (e.g., Olami et al., 1992; Mantegna & Stanley, 1995; Turcotte, 1999; Yakovlev et al., 2005; and many others). Since the seminal work by Bak et al. in the eighties and nineties, other models have been introduced to explain the emergence of power-law distributions in physical, biological, and man-made systems (Marković & Gros, 2014, provide a model review and analysis). In all such systems, some relevant quantity describing the "size" of the observed phenomenon (e.g., a spatial dimension, a number of connections, a release of mass or energy, etc.) is found to display a power-law distribution, which is a necessary (although not sufficient) element of self-organized critical systems. One relevant aspect of many systems described by power-law distributions is the unfeasibility of deterministic predictions of the size of the next event: the exceedingly large number of degrees of freedom in a system governed by highly nonlinear dynamics makes that endeavor a hopeless one, no matter how well we know the state of the system at a given time (countless experiments with sand piles as the prototype of self-organized critical systems clearly show that, e.g., Yoshioka, 2003). A general model of the Earth as a critical system with respect to the occurrence of volcanic eruptions does not exist yet; nonetheless, global subaerial volcanism displays some of its fundamental ingredients, such as: (i) power law—distributed eruption volumes demonstrated here; (ii) highly nonlinear dynamics well known to govern the volcanic processes (e.g., Sparks, 2003); (iii) scale invariance and self-similarity of processes, at least within the broad classes of explosive eruption on one side, and effusive eruptions on the other side, making progressively larger events to appear essentially as a scaled-up version of smaller ones; and (iv) to-date lack of any general or generally accepted relationship between pre-eruptive observations and size of the consequential eruption, letting open the possibility that forecasts of a next volcanic eruption size can not be defined deterministically (Papale, 2018). Critical systems are generally associated with systems that quickly dissipate energy accumulated over longer periods, as it is the case with potential energy release during sand pile collapse under the action of gravity, mechanical stress release during earthquakes, and pressure release and conversion into mechanical energy during volcanic eruptions. Here, we demonstrate a power-law relationship at global level, for which the equivalent of the sand pile would be the whole Earth, with the eruptions on its surface corresponding to individual landslides. We did not examine the individual volcano scale, for which the range of eruption size is often limited, and the number of available eruptions is usually exceedingly little for the extraction of confident distributions. Individual volcanoes may not display Poissonian behavior as for global volcanism, because the memoryless property of such distributions may not adequately account for their behavior. We do not know if over the individual volcano scale a power-law distribution similar to the one that we describe for global volcanism holds. Certainly, however, different eruptions from individual volcanoes are equally governed by extreme nonlinear processes, they display self-similarity of their events as for the global volcanism, and as long as we know, have escaped so far any systematic attempt to establish a clear, workable relationship between size and precursors. Future investigation should explore (i) the reasons at the basis of the observed global power-law distribution of volcanic eruption volumes, (ii) the relationships between the distributions at the global and individual volcano scale, and (iii) the potential consequences of such distributions for our forecasting capabilities, from the occurrence of a volcanic super-eruption globally to the size of the next event at a given volcano, including the possibility to relate observed episodes of unrests to the order of magnitude of the impending eruption.

5.3. Log-Normal Section of the Global Volume Distribution

Obvious questions follow from the above arguments: what does V_{min} represent in physical terms, and can we count on better predictive capabilities for the vast majority of eruptions with discharged volume below that threshold? Unfortunately, a robust answer does not emerge for none of them. First, it is useful to note that while the vast majority of medium to large explosive eruptions distribute along the power-law section, there is a substantial number of explosive as well as effusive eruptions along both the log-normal and power-law

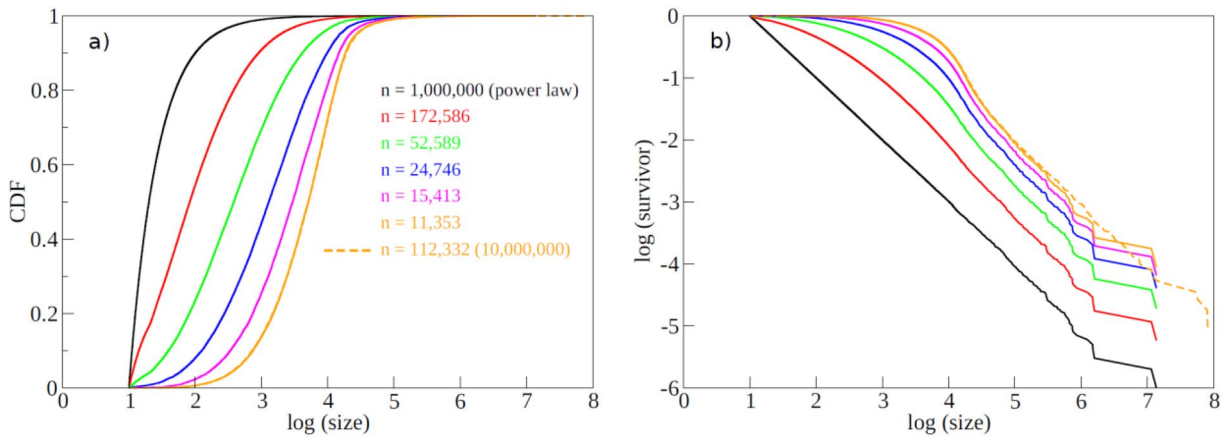

Figure 9. Results of a numerical experiment where an initial power-law distribution (black line) is modified by gathering a progressively larger proportion of small events to other randomly chosen events. (a) CDF and (b) survivor functions. For each colored curve, the reported number with the same color (in panel a) indicates the events resulting from progressive aggregation. The dashed orange curve is similar to the solid orange curve in terms of proportion of aggregated events, but it starts from a power law consisting of 10 instead of 1 million individual events. The CDF of these two curves (panel a) is about superimposed. CDF, cumulative distribution function.

sections of the distribution. Therefore, relating V_{min} to a fundamental difference in the mechanisms of effusive versus explosive eruptions does not seem justified. On the other hand, we have extensively discussed above a number of issues related to progressively smaller eruptions and the way they are identified and categorized in the available databases. In particular, small eruptions from frequently erupting volcanoes tend to lose their distinctiveness in the databases, being instead incorporated into undistinguished longer periods often associated with an average or peak VEI. Although the volume estimates employed here for eruptions with VEI ≤ 3 have been carefully chosen to eliminate or at least minimize that issue, the consequences are still significant for the global distribution. To make the point clear, consider again the case of Stromboli volcano: the current VEI 2 eruption as from the GVP database, lasting from 1934, is actually made of an estimated 10^6–10^7 individual events. Over the same time, the total number of eruptions occurring on Earth, with reference to "eruptions" as they are categorized in the databases, is estimated to be roughly 3.5×10^3 (Table 1). Clearly, the frequency of very small eruptions not individually recorded in the databases is orders of magnitudes larger than that of the recorded eruptions, opening to the possibility that if they were individually recorded, then a power-law distribution may be seen to extend far below the estimated V_{min}. In other words, the value of V_{min} estimated here, as well as the initial log-normal section of the global distribution, may simply emerge as the result of gathering together several small individual eruptions over long periods of volcanic activity, rather than having a true significance in terms of physical mechanisms governing small versus large eruptions. We have checked such a possibility by simulating the process of progressive gathering of small events from an initial power-law distribution (Figure 9). In detail, we associate the smaller events to a higher probability of being aggregated with other randomly chosen events, and repeat the process an increasing number of times so to obtain a series of modified distributions, each corresponding to a different extent of eruption aggregation. The resulting distributions mimic the one observed in the eruption events as they are reported in the employed databases, with an initial close-to-log-normal section followed by a power-law section (cfr. Figures 3 and 4).

The exercise in Figure 9 demonstrates that the artificial process of aggregation of small to very small individual eruptions to form larger ones is by itself capable of modifying a hypothetical initial distribution of the power law type down to the smallest eruption volumes, producing an overall trend similar to the one observed in the data. However, it does not tell us whether that is in fact the correct explanation to the observed trend. In other words, it cannot be excluded, but there is no warranty, that the true distribution would extend the power law down to much lower volumes corresponding to very small individual eruptions. Consequently, the log-normal section characterizing the left side of the volume distribution may not reflect any fundamental difference in the mechanisms of small versus large eruptions. While a definite answer requires a level of knowledge that is not supported by the databases, we must accept those databases as the current reference, and accept that the global distribution that we extract from them cannot be more

informative than the data themselves. We cannot employ the distribution to determine the expectance in terms of individual eruptions below a certain (undefined) small volume threshold. For such small eruptions, the expectance emerging from the distribution refers instead to future database population, assuming the paradigms for database population will remain unchanged.

5.4. Extreme Tail of the Global Volume Distribution

On the opposite extreme of the distribution, things are quite different. Here sit the giant eruptions, the occurrence of which is expected to have global impacts on climate and environment (Robock, 2000; Robock et al., 2009; Self, 2015; Papale & Marzocchi, 2019). Here the global distribution shows negative deviation (i.e., less observations than expected) with respect to an open power law, that we have modeled as a tapered power-law distribution. Similar right-end negative deviations are observed for a large variety of power law–distributed phenomena (e.g., Schoenberg & Patel, 2012; Geist & Parsons, 2014; and references therein). There are at least two good reasons for such negative deviations, already introduced above: (i) the extremely large events are rare, thus they are easily under-sampled; ii) the phenomena under consideration have some physical limits that prevent the power-law distribution to continue indefinitely. Obviously, the latter is necessarily true; for example, no volcanic eruption on Earth may produce more mass than the same mass of the Earth. However, deviations from power-law distributions are commonly observed much below such obvious limits, as in the present case where clear deviations are seen for order 1,000 km^3, or about 1 billionth of the Earth's volume. Statistical methods have been proposed to discriminate between open (observed deviations entirely due to under-sampling) and truncated or tapered (observed deviations representing a property of the distribution) power-law distributions; however, such methods are often poorly stable with respect to addition or subtraction of just one or two observations (Kagan & Schoenberg, 2001). Here we have adopted a tapered power-law distribution, basically on the a-posteriori observation, that we deem as a robust one, that if the effective distribution was open, then we would expect to find in the geological record many more extremely large eruptions than we do, including a significant number of VEI 9 eruptions that would occur with the relatively high frequency of one every 3 million years. In that view, the corner volume in the tapered power-law distribution (V_C in Tables 2 and 3, and in Figure 7g) takes a physical meaning related to a difficulty, or impossibility, for the natural system to increase over certain limits. Exploring the origin of such limits is not among the objectives of this work. Here, we only note that understanding the causes of such eruption volume limits, if they are true, requires evaluating the observed or reconstructed distribution of volumes of igneous intrusions, the relationships between magma accumulation and eruption volume, and the factors that can limit the amount of intra-crustal magma accumulation or the proportion of that magma that can be discharged in a single eruptive event.

The existence of tapering is however relevant, as it provides a mean and variance to the overall distribution whatever the value of the slope parameter k. In fact, an open power-law distribution with $k > 2$ (Table 2) would have a finite mean but infinite variance, corresponding to what is generally known as "black swan" behavior (Newman, 2005), that is, the irregular occurrence of extreme events corresponding to the far right of the distribution. The tapered distribution takes instead a finite variance, ruling out black swan behavior: extreme events far above the observed maxima (e.g., a VEI 9 eruption) are still possible, but their likelihood quickly decreases with increasing size, making their appearance far less likely than that of black swan events. It is noteworthy that $k > 2$ implies that along the power-law section of the distribution, a substantial fraction of the total erupted volume originates from small eruptions. The global eruption rates and the relative contribution from eruptions of different sizes are however not explored here, being the subject of separate work in preparation.

5.5. Final Remarks

This study describes a global distribution for subaerial volcanic eruptions on Earth, taking advantage and inheriting limits from current knowledge as represented in freely available, open access databases, that we have complemented with a database for small eruptions provided in Table S1. Such a global distribution tells us something relevant in terms of forecasting capabilities: on one side, it can be employed to make forecasts of future eruption discharges, providing constraints to global Earth system models, for example, global

tectonic models, models of mantle dynamics, or models of global climate change, as well as providing the basic reference for the occurrence of globally impacting volcanic eruptions. On the other side, it suggests that deterministic short-term forecasts of the size and impacts of an impending eruption may be impossible as a reflection of highly nonlinear dynamics with too many degrees of freedom, translating into a power-law distribution of eruption size as for many other natural (and man-made) phenomena observed on Earth.

Conflict of Interest

The authors declare no conflicts of interest relevant to this study.

Data Availability Statement

The data used in this study are from open access databases managed by third parties, and can be accessed at https://volcano.si.edu/(Global Volcanism Program, 2013) and https://www.bgs.ac.uk/vogripa/view/controller.cfc?method=lameve (Crosweller et al., 2012). Additional data that we have collected from the literature are provided in Table S1.

Acknowledgments
We are grateful to Arnau Folch and one anonymous reviewer for their comments on the originally submitted manuscript, resulting in improved final version. No dedicated funding has supported this study, apart from ordinary support from authors' Organizations.

References

Bak, P. (1997). *How nature works*. Oxford University Press.

Bak, P., Tang, C., & Wiesenfeld, K. (1987). Self-organized criticality: An explanation of the 1/f noise. *Physical Review Letters*, *59*, 381–384. https://doi.org/10.1103/physrevlett.59.381

Bonny, E., Thordarson, T., Wright, R., Höskuldsson, A., & Jónsdóttir, I. (2018). The volume of lava erupted during the 2014 to 2015 eruption at Holuhraun, Iceland: A comparison between satellite- and ground-based measurements. *Journal of Geophysical Research: Solid Earth*, *123*, 5412–5426. https://doi.org/10.1029/2017jb015008

Broido, A. D., & Clauset, A. (2019). Scale-free networks are rare. *Nature Communications*, *10*, 1017. https://doi.org/10.1038/s41467-019-08746-5

Calvari, S., Büttner, R., Cristaldi, A., Dellino, P., Giudicepietro, F., Orazi, M., et al. (2012). The 7 September 2008 Vulcanian explosion at Stromboli volcano: Multiparametric characterization of the event and quantification of the ejecta. *Journal of Geophysical Research, Solid Earth*, *117*, B05201. https://10.1029/2011JB009048

Clauset, A., Shalizi, C. R., & Newman, M. E. J. (2009). Power-law distributions in empirical data. *SIAM Review*, *51*, 661–703. https://doi.org/10.1137/070710111

Coppola, D., Barsotti, S., Cigolini, C., Laiolo, M., Pfeffer, M. A., & Ripepe, M. (2019). Monitoring the time-averaged discharge rates, volumes and emplacement style of large lava flows by using MIROVA system: The case of the 2014-2015 eruption at Holuhraun (Iceland). *Annals of Geophysics*, *62*, 2.

Corral, Á., & González, Á. (2019). Power law size distributions in Geoscience revisited. *Earth and Space Science*, *6*(5), 673–697. https://doi.org/10.1029/2018EA000479

Crosweller, H. S., Arora, B., Brown, S. K., Cottrell, E., Deligne, N. I., Guerrero, N. O., et al. (2012). Global database on large magnitude explosive volcanic eruptions (LaMEVE). *Journal of Applied Volcanology*, *1*, 4. http://www.appliedvolc.com/content/1/1/4

Deligne, N. I., Coles, S. G., & Sparks, R. S. J. (2010). Recurrence rates of large explosive volcanic eruptions. *Journal of Geophysical Research*, *115*. https://doi.org/10.1029/2009JB006554

Dirscherl, M., & Rossi, C. (2018). Geomorphometric analysis of the 2014-2015 Bárðarbunga volcanic eruption, Iceland. *Remote Sensing of Environment*, *204*, 244–259. https://doi.org/10.1016/j.rse.2017.10.027

Frank, S. A. (2009). The common patterns of nature. *Journal of Evolutionary Biology*, *22*, 1563–1585. https://doi.org/10.1111/j.1420-9101.2009.01775.x

Geist, E. L., & Parsons, T. (2014). Undersampling power-law size distributions: Effect on the assessment of extreme natural hazards. *Natural Hazards*, *72*, 565–595. https://doi.org/10.1007/s11069-013-1024-0

Giudicepietro, F., Lòpez, C., Macedonio, G., Alparone, S., Bianco, F., Calvari, S., et al. (2020). Geophysical precursors of the July-August 2019 paroxysmal eruptive phase and their implications for Stromboli volcano (Italy) monitoring. *Scientific Reports*, *10*, 10296. https://doi.org/10.1038/s41598-020-67220-1

Global Volcanism Program. (2013). In E. Venzke (Ed.), *Volcanoes of the world* (pp. 4–8.5). Smithsonian Institution.

Harris, A., Steffke, A., Calvari, S., & Spampinato, L. (2011). Thirty years of satellite-derived lava discharge rates at Etna: Implications for steady volumetric output. *Journal of Geophysical Research*, *116*, B08204. https://doi.org/10.1029/2011JB008237

Huang, Y., Hawkesworth, C., & van Calsteren, P., Smith, I., Black, P. (1997). Melt generation models for the Auckland volcanic field, New Zealand: Constraints from UTh isotopes. *Earth and Planetary Science Letters*, *149*, 67–84. https://doi.org/10.1016/S0012-821X(97)00064-2

Ito, G., & Clift, P. D. (1998). Subsidence and growth of Pacific Cretaceous plateaus. *Earth and Planetary Science Letters*, *161*, 85–100. https://doi.org/10.1016/S0012-821X(98)00139-3

Kagan, Y. Y. (2002). Seismic moment distribution revisited: I. Statistical results. *Geophysical Journal International*, *148*(3), 520–541. https://doi.org/10.1046/j.1365-246x.2002.01594.x

Kagan, Y. Y., & Schoenberg, F. (2001). Estimation of the upper cutoff parameter for the tapered Pareto distribution. *Journal of Applied Probability*, *38*, 158–175. https://doi.org/10.1239/jap/1085496599

Laherrère, J., & Sornette, D. (1998). Stretched exponential distributions in nature and economy: "fat tails" with characteristic scales. *European Physical Journal B*, *2*, 525–539. https://doi.org/10.1007/s100510050276

Macdonald, G. A., Abbott, A. T., & Peterson, F. L. (1983). *Volcanoes in the sea: The geology of Hawaii* (2nd ed.), University of Hawaii Press.

Mantegna, R. N., & Stanley, H. E. (1995). Scaling behaviour in the dynamics of an economic index. *Nature*, *376*, 46–49. https://doi.org/10.1038/376046a0

Marković, D., & Gros, C. (2014). Power laws and self-organized criticality in theory and nature. *Physics Reports*, *536*, 41–74.

Marzocchi, W., Spassiani, I., Stallone, A., & Taroni, M. (2020). How to be fooled searching for significant variations of the b-value. *Geophysical Journal International*, *220*, 1845, 1856. https://doi.org/10.1093/gji/ggz541

Newhall, C. G., & Self, S. (1982). The volcanic explosivity index (VEI) – An estimate of explosive magnitude for historical volcanism. *Journal of Geophysical Research*, *87*, 1231–1238. https://doi.org/10.1029/JC087iC02p01231

Newman, M. (2005). Power laws, Pareto distributions and Zipf's law. *Contemporary Physics*, *46*, 323–351. https://doi.org/10.1080/00107510500052444

Olami, Z., Feder, H. J. S., & Christensen, K. (1992). Self-organized criticality in a continuous, nonconservative cellular automaton modeling earthquakes. *Physical Review Letters*, *68*(8), 1244–1247. https://doi.org/10.1103/physrevlett.68.1244

Papale, P. (2018). Global time-size distribution of volcanic eruptions on Earth. *Scientific Reports*, *8*, 6838. https://10.1038/s41598-018-25286-y

Papale, P., & Marzocchi, W. (2019). Volcanic threats to global society. *Science*, *363*(6433), 1275–1276. https://10.1126/science.aaw7201

Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. (1992). *Numerical recipes in Fortran* (2nd ed.), Cambridge University Press.

Proietti, C., Coltelli, M., Marsella, M., Martino, M., Scifoni, S., & Giannone, F. (2020). Towards a satellite-based approach to measure eruptive volumes at Mt. Etna using Pleiades datasets. *Bulletin of Volcanology*, *82*, 35. https://doi.org/10.1007/s00445-020-01374-8

Pyle, D. M. (2015). Sizes of volcanic eruptions. In H. Sigurdsson (Ed.), *The encyclopedia of volcanoes* (2nd ed., pp. 257–264). Academic Press. https://doi.org/10.1016/b978-0-12-385938-9.00013-4

Robock, A. (2000). Volcanic eruptions and climate. *Reviews of Geophysics*, *38*(2), 191–219. https://8755-1209/00/1998.RG000054

Robock, A., Ammann, C. M., Oman, L., Shindell, D., Levis, S., & Stenchikov, G. (2009). Did the Toba volcanic eruption of ∼74 ka B.P. produce widespread glaciation? *Journal of Geophysical Research*, *114*, D10107. https://10.1029/2008JD011652

Safari, M. A. M., Masseran, N., & Ibrahim, K. (2018). Optimal threshold for Pareto tail modelling in the presence of outliers. *Physica A: Statistical Mechanics and its Applications*, *509*, 169–180. https://doi.org/10.1016/j.physa.2018.06.007

Schoenberg, F. P., & Patel, R. D. (2012). Comparison of Pareto and tapered Pareto distributions for environmental phenomena. *The European Physical Journal - Special Topics*, *205*, 159–166. https://doi.org/10.1140/epjst/e2012-01568-4

Self, S. (2015). Explosive super-eruptions and potential global impacts. In P. Papale (Ed.), *Volcanic hazards, risks, and disasters, hazards and sisasters series* (pp. 399–418). Elsevier. https://doi.org/10.1016/b978-0-12-396453-3.00016-2

Sparks, R. S. J. (2003). Forecasting volcanic eruptions. *Earth and Planetary Science Letters*, *210*, 1–15. https://doi.org/10.1016/s0012-821x(03)00124-9

Takarada, S., & Hoshizumi, H. (2020). Distribution and eruptive volume of Aso-4 pyroclastic density current and tephra fall deposits, Japan: A M8 super-eruption. *Frontiers of Earth Science*, *8*, 170. https://doi.org/10.3389/feart.2020.00170

Takarada, S., Oikawa, R., Furukawa, R., Hoshizumi, H., Itoh, J., Geshi, N., & Miyagi, I. (2016). Estimation of total discharged mass from the phreatic eruption of Ontake Volcano, central Japan, on September 27, 2014. *Earth, Planets and Space*, *68*, 138. https://doi.org/10.1186/s40623-016-0511-4

Tenreiro Machado, J. A., & Lopes, A. M. (2015). Analysis of natural and artificial phenomena using signal processing and fractional calculus. *Fractional Calculus and Applied Analysis*, *18*, 459–478. https://doi.org/10.1515/fca-2015-0029

Turcotte, D. L. (1999). Self-organized criticality. *Reports on Progress in Physics*, *62*, 1377–1429. https://doi.org/10.1088/0034-4885/62/10/201

van Leeuwen, J., Smeets, J. B. J., & Belopolsky, A. V. (2019). Forget binning and get SMART: Getting more out of the time-course of response data. *Attention, Perception & Psychophysics*, *81*(8), 2956–2967. https://doi.org/10.3758/s13414-019-01788-3

Vere-Jones, D., Robinson, R., & Yang, W. (2001). Remarks on the accelerated moment release model: Problems of model formulation, simulation and estimation. *Geophysical Journal International*, *144*, 517–531. https://doi.org/10.1046/j.1365-246x.2001.01348.x

Virkar, Y., & Clauset, A. (2014). Power-law distributions in binned empirical data. *Annals of Applied Statistics*, *8*(1), 89–119. https://doi.org/10.1214/13-AOAS710

Xu, W.-L., Ji, W.-Q., Pei, F.-P., Meng, E., Yu, Y., Yang, D.-B., & Zhang, X. (2009). Triassic volcanism in eastern Heilongjiang and Jilin provinces, NE China: Chronology, geochemistry, and tectonic implications. *Journal of Asian Earth Sciences*, *34*(3), 392–402. https://doi.org/10.1016/j.jseaes.2008.07.001

Yakovlev, G., Newman, W. I., Turcotte, D. L., & Gabrielov, A. (2005). An inverse cascade model for self-organized complexity and natural hazards. *Geophysical Journal International*, *163*, 433–442. https://doi.org/10.1111/j.1365-246x.2005.02717.x

Yoshioka, N. (2003). A sandpile experiment and its implications for self-organized criticality and characteristic earthquake. *Earth, Planets and Space*, *55*, 283–289. https://doi.org/10.1186/bf03351762

References From the Supporting Information

Albino, F., Smets, B., d'Oreye, N., & Kervyn, F. (2015). High-resolution TanDEM-X DEM: An accurate method to estimate lava flow volumes at Nyamulagira volcano (D. R. Congo). *Journal of Geophysical Research: Solid Earth*, *120*, 4189–4207. https://doi.org/10.1002/2015jb011988

Arnold, D. W. D., Biggs, J., Anderson, K., Vallejo Vargas, S., Wadge, G., Ebmeier, S. K., et al. (2017). Decaying lava extrusion rate at El Reventador Volcano, Ecuador, measured using high-resolution satellite radar. *Journal of Geophysical Research: Solid Earth*, *122*, 9966–9988. https://doi.org/10.1002/2017jb014580

Auer, A., Belousov, A., & Belousova, M. (2018). Deposits, petrology and mechanism of the 2010-2013 eruption of Kizimen volcano in Kamchatka, Russia. *Bulletin of Volcanology*, *80*, 33. https://doi.org/10.1007/s00445-018-1199-z

Belousov, A., Voight, B., Belousova, M., & Petukhin, A. (2002). Pyroclastic surges and flows from the 8-10 May 1997 explosive eruption of Bezymianny volcano, Kamchatka, Russia. *Bulletin of Volcanology*, *64*, 455–471. https://doi.org/10.1007/s00445-002-0222-5

Bennington, N., Haney, M., Thurber, C., & Zeng, X. (2018). Inferring magma dynamics at Veniamin of volcano via application of ambient noise. *Geophysical Research Letters*, *45*, 11650–11658. https://doi.org/10.1029/2018gl079909

Branca, S., Azzaro, R., De Beni, E., Chester, D., & Duncan, A. (2015). Impacts of the 1669 eruption and the 1693 earthquakes on the Etna Region (Eastern Sicily, Italy): An example of recovery and response of a small area to extreme events. *Journal of Volcanology and Geothermal Research*, *303*, 25–40. https://doi.org/10.1016/j.jvolgeores.2015.07.020

Branca, S., De Beni, E., Chester, D., Duncan, A., & Lotteri, A. (2017). The 1928 eruption of Mount Etna (Italy): Reconstructing lava flow evolution and the destruction and recovery of the town of Mascali. *Journal of Volcanology and Geothermal Research*, *335*, 54–70. https://doi.org/10.1016/j.jvolgeores.2017.02.002

Castruccio, A., & Contreras, M. A. (2016). The influence of effusion rate and rheology on lava flow dynamics and morphology: A case study from the 1971 and 1988-1990 eruptions at Villarrica and Lonquimay volcanoes, Southern Andes of Chile. *Journal of Volcanology and Geothermal Research*, *327*, 469–483. https://doi.org/10.1016/j.jvolgeores.2016.09.015

Coltelli, M., Marsella, M., Proietti, C., & Scifoni, S. (2012). The case of the 1981 eruption of Mount Etna: An example of very fast moving lava flows. *Geochemistry, Geophysics, Geosystems*, *13*, Q01004. https://doi.org/10.1029/2011GC003876

Coltelli, M., Proietti, C., Branca, S., Marsella, M., Andronico, D., & Lodato, L. (2007). Analysis of the 2001 lava flow eruption of Mt. Etna from three-dimensional mapping. *Journal of Geophysical Research*, *112*. https://doi.org/10.1029/2006JF000598

Coppola, D., Di Muro, A., Peltier, A., Villeneuve, N., Ferrazzini, V., Favalli, M., et al. (2017). Shallow system rejuvenation and magma discharge trends at Piton de la Fournaise volcano (La Réunion Island). *Earth and Planetary Science Letters*, *463*, 13–24. https://doi.org/10.1016/j.epsl.2017.01.024

Dixon, J. P., Cameron, C., McGimsey, R. G., Neal, C. A., & Waythomas, C. (2015). *2013 volcanic activity in Alaska Summary of events and response of the Alaska Volcano Observatory*. U S Geological Survey Scientific Investigations Report. https://doi.org/10.3133/sir20155110

Embley, R., Tamura, Y., Merle, S., Sato, T., Ishizuka, O., Chadwick, W., Jr, et al. (2014). Eruption of South Sarigan Seamount, Northern Mariana Islands: Insights into hazards from submarine volcanic eruptions. *Oceanography*, *27*(2), 24–31. https://doi.org/10.5670/oceanog.2014.37

Favalli, M., Chirico, G. D., Papale, P., Pareschi, M. T., & Boschi, E. (2009). Lava flow hazard at Nyiragongo volcano, D.R.C. *Bulletin of Volcanology*, *71*, 363–374. https://doi.org/10.1007/s00445-008-0233-y

Fitton, J. G., Kilburn, C. R. J., Thirlwall, M. F., & Hughes, D. J. (1983). 1982 eruption of Mount Cameroon, West Africa. *Nature*, *306*, 327–332. https://doi.org/10.1038/306327a0

Gallant, E., Deng, F., Connor, C. B., Dixon, T. H., Xie, S., Saballos, J. A., et al. (2020). Deep and rapid thermos-mechanical erosion by a small-volume lava flow. *Earth and Planetary Science Letters*, *537*, 116163. https://doi.org/10.1016/j.epsl.2020.116163

Geist, D. J., Harpp, K. S., Naumann, T. R., Poland, M., Chadwick, W. W., Hall, M., & Rader, E. (2008). The 2005 eruption of Sierra Negra volcano, Galápagos, Ecuador. *Bulletin of Volcanology*, *70*, 655–673. https://doi.org/10.1007/s00445-007-0160-3

Geshi, N., & Itoh, J. (2018). Pyroclastic density currents associated with the 2015 phreatomagmatic eruption of the Kuchinoerabujima volcano. *Earth, Planets and Space*, *70*, 119. https://doi.org/10.1186/s40623-018-0881-x

Gèze, B. (1953). Les volcans du Cameroun occidental. *Bulletin of Volcanology*, *13*, 63–92. https://doi.org/10.1007/bf02596792

Ghorbanzadeh, D., Durand, P., & Jaupi, L. (2017). Generating the skew normal random variable. *Paper presented at the World Congress on Engineering 2017*. (Vol I). World Congress on Engineering.

Gorbach, N. V., Plechova, A. A., Manevich, T. M., Portnyagin, M. V., Philosofova, T. M., & Samoilenko, S. B. (2018). The composition of volcanic ash and the dynamics of the 2013-2016 Zhupanovsky volcano eruption. *Journal of Volcanology and Seismology*, *12*, 155–171. https://doi.org/10.1134/s0742046318030028

Gordeev, E. I., Girina, O. A., Lupyan, E. A., Sorokin, A. A., Kramareva, L. S., Efremov, V. Y., et al. (2016). The VolSatView information system for monitoring the volcanic activity in Kamchatka and on the Kuril Islands. *Journal of Volcanology and Seismology*, *10*, 382–394. https://doi.org/10.1134/s074204631606004x

Gudmundsson, M. T., Högnadóttir, T., Pàlsson, F., & Björnsson, H. (2000). *Grímsvötn: Eldgosið 1998 og breytingar á botni, rúmmáli og jarðhita 1996–1999. Report RH-03-2000* (in Icelandic).

Gudmundsson, M. T., Sigmundsson, F., Björnsson, H., & Högnadóttir, T. (2004). The 1996 eruption at Gjálp, Vatnajökull ice cap, Iceland: Efficiency of heat transfer, ice deformation and subglacial water pressure. *Bulletin of Volcanology*, *66*, 46–65. https://doi.org/10.1007/s00445-003-0295-9

Hadisantono, R. D., Mulyana, R., Purwoto, & Ridwan, I. (2008). Gejala perubahan pada erupsi G. Soputan, 6 Juni 2008. *Vulkanologi dan Bencana Geologi*. *3*, 6–12 (in Indonesian).

Harris, A., Chevrel, M., Coppola, D., Ramsey, M., Hrysiewicz, A., Thivet, S., et al. (2019). Validation of an integrated satellite-data-driven response to an effusive crisis: The April-May 2018 eruption of Piton de la Fournaise. *Annals of Geophysics*, *61* (2), https://doi.org/10.4401/ag-7972

Houghton, B. F., & Nairn, I. A. (1991). The 1976–1982 Strombolian and phreatomagmatic eruptions of White Island, New Zealand: Eruptive and depositional mechanisms at a 'wet' volcano. *Bulletin of Volcanology*, *54*, 25–49. https://doi.org/10.1007/bf00278204

Jones, C., McConnell, C., Coleman, K., Cox, P., Falloon, P., Jenkinson, D., & Powlson, D. (2005). Global climate change and soil carbon stocks; Predictions from two contrasting models for the turnover of organic carbon in soil. *Global Change Biology*, *11*, 154–166. https://doi.org/10.1111/j.1365-2486.2004.00885.x

Jude-Eton, T. C., Thordarson, T., Gudmundsson, M. T., & Oddsson, B. (2012). Dynamics, stratigraphy and proximal dispersal of supraglacial tephra during the ice-confined 2004 eruption at Grímsvötn Volcano, Iceland. *Bulletin of Volcanology*, *74*, 1057–1082. https://doi.org/10.1007/s00445-012-0583-3

Kaneko, T., Maeno, F., Yasuda, A., Takeo, M., & Takasaki, K. (2019). The 2017 Nishinoshima eruption: Combined analysis using Himawari-8 and multiple high-resolution satellite images. *Earth, Planets and Space*, *71*, 140. https://doi.org/10.1186/s40623-019-1121-8

Kereszturi, G., Németh, K., Moufti, M. R., Cappello, A., Murcia, H., Ganci, G., et al. (2016). Emplacement conditions of the 1256 AD Al-Madinah lava flow field in Harrat Rahat, Kingdom of Saudi Arabia – Insights from surface morphology and lava flow simulations. *Journal of Volcanology and Geothermal Research*, *309*, 14–30. https://doi.org/10.1016/j.jvolgeores.2015.11.002

Kilgour, G., Gates, S., Kennedy, B., Farquhar, A., McSporran, A., & Asher, C. (2019). Phreatic eruption dynamics derived from deposit analysis: A case study from a small, phreatic eruption from Whakari/White Island, New Zealand. *Earth, Planets and Space*, *71*, 36. https://doi.org/10.1186/s40623-019-1008-8

Kushendratno, Pallister, J. S., Kristianto, Bina, F. R., McCausland, W., Carn, S., et al. (2012). Recent explosive eruptions and volcano hazards at Soputan volcano—A basalt stratovolcano in north Sulawesi, Indonesia. *Bulletin of Volcanology*, *74*, 1581–1609. https://doi.org/10.1007/s00445-012-0620-2

Maeno, F., Nakada, S., & Kaneko, T. (2016). Morphological evolution of a new volcanic islet sustained by compound lava flows. *Geology*, *44*, 259–262. https://doi.org/10.1130/g37461.1

Marsella, M., Scifoni, S., Coltelli, M., & Proietti, C. (2011). Quantitative analysis of the 1981 and 2001 Etna flank eruptions: A contribution for future hazard evaluation and mitigation. *Annals of Geophysics*, *54*, https://doi.org/10.4401/ag-5334

Medina, L. N., Arcos, D. F., & Battaglia, M. (2017). Twenty years (1990-2010) of geodetic monitoring of Galeras volcano (Colombia) from continuous tilt measurements. *Journal of Volcanology and Geothermal Research*, *344*, 232–245.

Mendoza-Rosas, A. T., & De la Cruz-Reyna, S. (2008) A statistical method linking geological and historical eruption time series for volcanic hazard estimations: Applications to active polygenetic volcanoes. *Journal of Volcanology and Geothermal Research*, *176*(2), 277–290. https://doi.org/10.1016/j.jvolgeores.2008.04.005

Mendoza-Rosas, T. A., Gomez-Vazquez, A., & de la Cruz-Reyna, S. (2017). Statistical analysis of the sustained lava dome emplacement and destruction processes at Popocatepetl volcano, Central Mexico. *Bulletin of Volcanology*, *79*, 43. https://doi.org/10.1007/s00445-017-1127-7

Métrich, N., & Wallace, P. J. (2008). Volatile abundances in basaltic magmas and their degassing paths tracked by melt inclusions. In K. D. Putirka, & F. J. TepleyIII (Eds.), Minerals, inclusions and volcanic processes, review in mineralogy & geochemistry series (Vol. 69, pp. 363–402). Mineralogical Society of America. https://doi.org/10.1515/9781501508486-011

Morgan, H. A., Harris, A. J. L., & Gurioli, L. (2013). Lava discharge rate estimates from thermal infrared satellite data for Pacaya Volcano during 2004-2010. *Journal of Volcanology and Geothermal Research*, *264*, 1–11. https://doi.org/10.1016/j.jvolgeores.2013.07.008

Mulas, M., Cioni, R., Andronico, D., & Mundula, F. (2016). The explosive activity of the 1669 Monti Rossi eruption at Mt. Etna (Italy). *Journal of Volcanology and Geothermal Research*, *328*, 115–133. https://doi.org/10.1016/j.jvolgeores.2016.10.012

Nakada, S., Nagai, M., Kaneko, T., Nozawa, A., & Suzuki-Kamata, K. (2005). Chronology and products of the 2000 eruption of Miyakejima Volcano, Japan. *Bulletin of Volcanology*, *67*, 205–218. https://doi.org/10.1007/s00445-004-0404-4

Nakamichi, H., Iguchi, M., Triastuty, H., Hendrasto, M., & Mulyana, I. (2019). Differences of precursory seismic energy release for the 2007 effusive dome-forming and 2014 Plinian eruptions at Kelud volcano, Indonesia. *Journal of Volcanology and Geothermal Research*, *382*, 68–80. https://doi.org/10.1016/j.jvolgeores.2017.08.004

Nyome, M. S., Suh, C. E., Sparks, R. S. J., Ayonghe, S. N., & Fitton, J. G. (2008). The Mount Cameroon 1959 compound lava flow field: Morphology, petrography and geochemistry. *Swiss Journal of Geosciences*, *101*, 85–98.

Patefield, M., & Tandy, D. (2000). Fast and accurate calculation of Owen's T function. *Journal of Statistical Software*, *5*(5), 1–25. https://doi.org/10.18637/jss.v005.i05

Patrick, M. R., Dehn, J., Papp, K. R., Lu, Z., Dean, K., Moxey, L., et al. (2003). The 1997 eruption of Okmok Volcano, Alaska: A synthesis of remotely sensed imagery. *Journal of Volcanology and Geothermal Research*, *127*, 87–105. https://doi.org/10.1016/s0377-0273(03)00180-x

Pedersen, G. B. M., Belart, J. M. C., Magnússon, E., Vilmundardóttir, O. K., Kizel, F., Sigurmundsson, F. S., et al. (2018). Hekla volcano, Iceland, in the 20th century: Lava volumes, production rates, and effusion rates. *Geophysical Research Letters*, *45*, 1805–1813. https://doi.org/10.1002/2017gl076887

Peltier, A., Famin, V., Bachèlery, P., Cayol, V., Fukushima, Y., & Staudacher, T. (2008). Cyclic magma storages and transfers at Piton de La Fournaise volcano (La Réunion hotspot) inferred from deformation and geochemical data. *Earth and Planetary Science Letters*, *270*, 180–188. https://doi.org/10.1016/j.epsl.2008.02.042

Pietruszka, A. J., Marske, J. P., Heaton, D. E., Garcia, M. O., & Rhodes, J. M. (2018). An isotopic perspective into the magmatic evolution and architecture of the rift zones of Kīlauea Volcano. *Journal of Petrology*, *59*, 2311–2352. https://doi.org/10.1093/petrology/egy098

Primulyana, S., Bani, P., & Harris, A. (2017). The effusive-explosive transition at Rokatenda 2012-2013: Unloading by extrusion of degassed magma with lateral gas flow. *Bulletin of Volcanology*, *79*, 22. https://doi.org/10.1007/s00445-017-1104-1

Queisser, M., Granieri, D., & Burton, M. (2016). A new Frontier in CO_2 flux measurements using a highly portable DIAL laser system. *Scientific Reports*, *6*, 33834. https://10.1038/srep33834

Reynolds, H. I., Gudmundsson, M. T., Högnadóttir, T., & Pálsson, F. (2018). Thermal power of Grímsvötn, Iceland, from 1998 to 2016: Quantifying the effects of volcanic activity and geothermal anomalies. *Journal of Volcanology and Geothermal Research*, *358*, 184–193. https://doi.org/10.1016/j.jvolgeores.2018.04.019

Richter, N., Favalli, M., de Zeeuw-van Dalfsen, E., Fornaciai, A., da Silva Fernandes, R. M., Pérez, N. M., et al. (2016). Lava flow hazard at Fogo Volcano, Cabo Verde, before and after the 2014-2015 eruption. *Natural Hazards and Earth System Sciences*, *16*, 1925–1951. https://doi.org/10.5194/nhess-16-1925-2016

Rowland, S. K., Harris, A. J. L., Wooster, M. J., Amelung, F., Garbeil, H., Wilson, L., & Mouginis-Mark, P. J. (2003). Volumetric characteristics of lava flows from interferometric radar and multispectral satellite data: The 1995 Fernandina and 1998 Cerro Azul eruptions in the western Galápagos. *Bulletin of Volcanology*, *65*, 311–330. https://doi.org/10.1007/s00445-002-0262-x

Sánchez, E., Gallardo, C., Gaertner, M. A., Arribas, A., & Castro, M. (2004). Future climate extreme events in the Mediterranean simulated by a regional climate model: A first approach. *Global and Planetary Change*, *44*, 163–180. https://doi.org/10.1016/j.gloplacha.2004.06.010

Savin, C., Grasso, J.-R., & Bachelery, P. (2005). Seismic signature of a phreatic explosion: Hydrofracturing damage at Karthala volcano, Grande Comore Island, Indian Ocean. *Bulletin of Volcanology*, *67*, 717–731. https://doi.org/10.1007/s00445-005-0411-0

Schaefer, L. N., Wang, T., Escobar-Wolf, R., Oommen, T., Lu, Z., Kim, J., et al. (2017). Three-dimensional displacements of a large volcano flank movement during the May 2010 eruptions at Pacaya Volcano, Guatemala. *Geophysical Research Letters*, *44*, 135–142. https://doi.org/10.1002/2016gl071402

Stelling, P., Beget, J., Nye, C., Gardner, J., Devine, J., & George, R. (2002). Geology and petrology of ejecta from the 1999 eruption of Shishaldin Volcano, Alaska. *Bulletin of Volcanology*, *64*, 548–561. https://doi.org/10.1007/s00445-002-0226-1

Suh, C. E., Ayonghe, S. N., Sparks, R. S. J., Annen, C., Fitton, J. G., Nana, R., & Luckman, A. (2003). The 1999 and 2000 eruptions of Mount Cameroon: Eruption behaviour and petrochemistry of lava. *Bulletin of Volcanology*, *65*, 267–281. https://doi.org/10.1007/s00445-002-0257-7

Toombs, A., & Wadge, G. (2012). Co-eruptive and inter-eruptive surface deformation measured by satellite radar interferometry at Nyamuragira volcano, D.R. Congo, 1996 to 2010. *Journal of Volcanology and Geothermal Research*, *245–246*, 98–122. https://doi.org/10.1016/j.jvolgeores.2012.07.005

Trusdell, F. A., Moore, R. B., Sako, M., White, R. A., Koyanagi, S. K., Chong, R., & Camacho, J. T. (2005). The 2003 eruption of Anatahan volcano, Commonwealth of the Northern Mariana Islands: Chronology, volcanology, and deformation. *Journal of Volcanology and Geothermal Research*, *146*, 184–207. https://doi.org/10.1016/j.jvolgeores.2004.12.010

Wadge, G., Biggs, J., Lloyd, R., & Kendall, J.-M. (2016). Historical volcanism and the state of stress in the East African Rift system. *Frontiers of Earth Science*, *4*, 86. https://doi.org/10.3389/feart.2016.00086

Wadge, G., McCormick Kilbride, B. T., Edmonds, M., Johnson, R. W., & Johnson, R. W. (2018). Persistent growth of a young andesite lava cone: Bagana volcano, Papua New Guinea. *Journal of Volcanology and Geothermal Research*, *356*, 304–315. https://doi.org/10.1016/j.jvolgeores.2018.03.012

Wallace, K. L., Schaefer, J. R., & Coombs, M. L. (2013). Character, mass, distribution, and origin of tephra-fall deposits from the 2009 eruption of Redoubt Volcano, Alaska-Highlighting the significance of particle aggregation. *Journal of Volcanology and Geothermal Research*, *259*, 145–169. https://doi.org/10.1016/j.jvolgeores.2012.09.015

Xu, W., & Jonsson, S. (2014). The 2007-8 volcanic eruption on Jebel at Tair island (Red Sea) observed by satellite radar and optical images. *Bulletin of Volcanology*, *76*, 795. https://doi.org/10.1007/s00445-014-0795-9