Journal of Volcanology and Geothermal Research 431 (2022) 107648

Contents lists available at ScienceDirect

Journal of Volcanology and Geothermal Research

journal homepage: www.journals.elsevier.com/journal-of-volcanology-and-geothermal-research

A new perspective on eruption data completeness: insights from the First Recorded EruptionS in the Holocene (FRESH) database

Vanesa Burgos [a,b,*], Susanna F. Jenkins [a,b], Mark Bebbington [c], Chris Newhall [d], Benoit Taisne [a,b]

[a] Earth Observatory of Singapore, Singapore
[b] Asian School of the Environment, Nanyang Technological University, Singapore
[c] School of Agriculture and Environment, Massey University, Palmerston North, New Zealand
[d] Mirisbiris Garden and Nature Center, Philippines

ARTICLE INFO

Keywords:
Change points
Data completeness
First recorded eruptions
GVP
Recording rate
Regional records
VOTW database

ABSTRACT

Identifying the most complete (best recorded) portion of an eruption record is essential before estimating eruption recurrence and probability. This is typically achieved by plotting cumulative eruptions through time. Here, we evaluate eruption data completeness from a new perspective, by compiling the first dated Holocene eruption from each volcano in the Volcanoes of the World (VOTW) database (i.e., First Recorded EruptionS in the Holocene (FRESH)). In our first analysis, we compared the subregional distribution of FRESH with time using Kolmogorov-Smirnov (K—S) test. We found that the eruption record was best categorised into 31 regions containing subregions with similar degrees of completeness. This opened the way to define new Relative Completeness Date(s) (RCD) as a function of eruption size, volcanic characteristics, and region, by identifying multiple points in the record where the root-mean-square (RMS) level changes abruptly, corresponding to a gap, a decrease or increase in the FRESH rate. Regional RCDs in the Common Era (CE) range from as recently as 1964 CE in the Indian Ocean (southern) to 200 CE in Middle East and Western Indian Ocean. In contrast, some regions like Kamchatka and Mainland Asia have near-constant rates of FRESH over the last 12,000 years, making RCDs impossible to assign. We present and make available our FRESH database, and describe and implement an automatic approach to detect RCDs across our newly defined volcanic regions. We suggest that the different degrees of completeness observed at a regional scale can be explained by: socio-historical events, access to geological studies, submarine volcanism, and/or remoteness. The FRESH database, together with the new regions and proposed RCDs can be used in future studies to estimate eruption probabilities at volcanoes without Holocene records and identify which subregions are most likely to produce a FRESH in the future.

1. Introduction

Eruption under-recording remains one of the main obstacles when working with global catalogues. Despite the rising number of recorded eruptions in the Smithsonian's Global Volcanism Program (GVP) eruption catalogue (Siebert et al., 2011), nearly 40% of the volcanoes (n = 566) in the Volcanoes of the World (VOTW) database lack dated Holocene eruptions (GVP, 2013). Moreover, the percentage of missing eruptions increases significantly as we go back in time (Rougier et al., 2018), meaning that eruption recurrence and probability are likely underestimated if we consider the whole Holocene catalogue (Jenkins et al., 2012; Mead and Magill, 2014). One solution is to rely on the most complete portion of the catalogue. However, a number of questions arise: i) given that the degree of completeness varies geographically and across characteristics (e.g., Volcanic Explosivity Index (VEI)), what portion of the catalogue can be considered as the most complete?; ii) how do we best identify the most complete portion?

Several approaches have been proposed to assess the completeness of eruption records. The most straightforward approach, known as 'break-in-slope', visually identifies the point at which the cumulative number of eruptions increases most rapidly with time (Jenkins et al., 2012). This approach works for records with a well-defined breakpoint, but when

Abbreviations: BCE, Before Common Era; CE, Common Era; ECDF, Empirical Cumulative Distribution Function; FRESH, First Recorded EruptionS in the Holocene; GVP, Global Volcanism Program; K-S, Kolmogorov-Smirnov; RCD, Relative Completeness Date; RMS, root-mean-square; VEI, Volcanic Explosivity Index; VOTW, Volcanoes of the World.

* Corresponding author at: 50 Nanyang Ave, Block N2-01b-30, 639798, Singapore.
E-mail address: burg0001@e.ntu.edu.sg (V. Burgos).

https://doi.org/10.1016/j.jvolgeores.2022.107648
Received 13 April 2022; Received in revised form 12 August 2022; Accepted 16 August 2022
Available online 19 August 2022

the break-in-slope is unclear, the selection is more subjective. Another method of identifying break-in-slope was proposed by Mulargia et al. (1987), who modified a version of the Kolmogorov-Smirnov (K—S) test and combined it with Monte Carlo simulations to identify breakpoints in Mount Etna's records corresponding to different eruptive regimes. Mulargia et al.'s method was also implemented by Garcia-Aristizabal et al. (2012) to analyse the completeness of Miyakejima's eruption record.

Rougier et al. (2016) and Rougier et al. (2018) proposed a non-parametric approach to represent the global recording rate of large explosive eruptions. A parametric alternative to the 'break-in-slope' approach is modelling the probability of recording an eruption at a given time. Guttorp and Thompson (1991) originally proposed a time-smoothed probability function, and Coles and Sparks (2006) later developed a method based on the theory of extreme values to tackle recording bias through modelling large-magnitude eruptions as a Poisson point process. This method, also used by Deligne et al. (2010), is based on a predefined function shape, and assumes that the probability of recording an eruption increases when recent and of large magnitude. A similar method was proposed by Kiyosugi et al. (2015) to model the recording rate of large explosive eruptions with an exponential decay function. To avoid choosing a specific function shape, Furlan (2010) and Mead and Magill (2014) combined the theory of extreme values with a step function, and used the Metropolis-Hastings Markov chain Monte Carlo (MCMC) to obtain a distribution of the model parameters. In contrast with other approaches, Wang et al. (2020) proposed using Bayesian inference methods, to fit a logistic function to the observed intensity function of a marked point process which allowed them to assess the models' goodness-of-fit.

Proposed completeness dates (i.e., date after which it is estimated that 100% of the eruptions are reported and captured in the records) in the literature vary across data sets, eruption sizes, and regions. For example, Guttorp and Thompson (1991) modelled the probability of observing an eruption of any given size using the Volcano Reference File from Simkin et al. (1981), which could be considered the first version of the VOTW database, and assuming that this probability was 100% after 1980, while Wang et al. (2020) reported 1985 as the completeness date for the VOTW database. Other studies identified a global improvement in eruption recording between 500 and 200 years ago (Mead and Magill, 2014; Siebert et al., 2011). When Newhall and Self (1982) introduced the VEI scale, it was already noted that the completeness of the eruption record ranged from a few decades for small eruptions of VEI ≤ 3 to a few centuries for large eruptions of VEI ≥ 4. More recently, Rougier et al. (2016) assumed 1980 as the completeness date for magnitude (M) ≥ 4 eruptions, as described in Mason et al. (2004) 's magnitude scale, in the LaMEVE database from Crosweller et al. (2012). Rougier et al. (2016) reported an improvement in the global recording rate of $M \geq 4$ eruptions after the 15th century and found that global recording probability was lower than 20% and 50% before 1100 CE and 1600, respectively. Variable completeness of records has also been reported for regions and single volcanoes (Jenkins et al., 2012; Kiyosugi et al., 2015; Mead and Magill, 2014; Rougier et al., 2018), underscoring the importance of using an appropriate spatial scale when estimating eruption recurrence and probability.

The completeness dates proposed by the studies reviewed above were based on eruption records from one or multiple volcanoes. These completeness dates can then be used to estimate eruption recurrence and likelihood for volcanoes with recorded eruptions. However, it has not been established yet if these completeness dates are applicable to the hundreds of volcanoes without Holocene dated eruptions that have the potential of erupting again. Before being able to estimate how often we can expect an eruption from a volcano without records, we have to assess eruption data completeness from a new perspective: instead of considering the entire record for each volcano, we look at the first dated eruption from each volcano in the VOTW database with confirmed eruptions. These eruptions have been compiled into the First Recorded EruptionS in the Holocene (FRESH) database. Henceforth, we refer to each of these eruptions as a FRESH. Assuming that the occurrence of FRESH per region and volcanic characteristics (e.g., VEI) can be described as a homogenous Poisson process (De la Cruz-Reyna, 1991), the two main research questions addressed in this study are:

(1) Do the GVP regions contain subregions with significant differences in the distribution of FRESH with time? Statistically distinct subregions should be considered as individual regions for the eruption data completeness analysis.
(2) Does the FRESH rate vary over the Holocene as a function of VEI, primary volcano type, major rock composition, tectonic setting, and region? An increase in the rate would indicate an improvement in eruption recording.

To answer the first question, we analysed the subregional distributions of FRESH with time using the two-sample Kolmogorov-Smirnov (K—S) test. Based on our results, we modify several GVP regions and create new regions containing subregions with similar degrees of completeness. To answer the second question, we followed the most abrupt change point approach implemented by the *findchangepts* function (Mathworks, 2021) that detects significant changes in the root-mean-square (RMS) level of the time between consecutive FRESH start dates. With this straightforward approach, we can detect multiple change points in the FRESH rate automatically and define the Relative Completeness Date (RCD) (i.e., date after which there is a relative improvement in eruption recording) for different subsets of the FRESH database. Another significant advantage is that this method also detects gaps in the record or decreases in the FRESH rate. Future studies can benefit from the FRESH database, the proposed new regions and RCDs, when estimating the regional probabilities of having an eruption at volcanoes without Holocene records. Furthermore, the FRESH database can provide insight into the occurrence of eruptions at apparently quiet volcanoes without Holocene records.

2. Methodology

2.1. Data

The FRESH database was compiled from the VOTW database, the most comprehensive global database of Holocene volcanic eruptions (GVP, 2013). Of the 1431 Holocene volcanoes in the VOTW database, 865 have had at least one confirmed eruption since the beginning of the Holocene until December 2018 (v. 4.7.5, 21 December 2018). We extracted the details of the first dated Holocene eruption (e.g., VEI and start date) as well as general information on each of these volcanoes (e.g., tectonic setting and region), and saved all the data in the FRESH database (Table S1 in Supplementary Material). For the evidence dating method, we follow the GVP nomenclature for the category *Historical Observations* and refer to the other methods (e.g., tephrochronology, radiocarbon, and Ar/Ar, among other techniques (Siebert et al., 2011) as non-historical methods. Note that each FRESH might not be the first-ever eruption in the Holocene or the eruption that first formed a specific volcano. Rather a FRESH is the first Holocene eruption recorded in the VOTW database for a volcano that has a reported age. We consider a volcano to follow the terminology used in the GVP to describe the primary volcano type (e.g., a volcanic field might have multiple volcano-forming monogenetic eruptions, but we consider the first recorded Holocene eruption for the whole volcanic field).

While we acknowledge that start dates in the GVP may be subject to uncertainty, we do not consider dating uncertainty in our study due to lack of data: 629 out of 865 FRESH (72.7%) do not have an error assigned to the start year of the eruption in the VOTW database. This is in line with most studies on eruption under-recording and data completeness, which similarly do not consider dating errors because of a lack of data (Guttorp and Thompson, 1991; Jenkins et al., 2012;

Kiyosugi et al., 2015; Mead and Magill, 2014; Rougier et al., 2018, 2016; Wang et al., 2020). Volcano types with similar morphologies were grouped into general categories to minimise the effect of small sample sizes on the statistical analysis. We transformed the 29 primary volcano types (Siebert et al., 2011) into the following types: caldera ($n = 65$), stratovolcano ($n = 462$), complex (complex, and compound) ($n = 43$), fissure vent (crater rows, and fissure vent) ($n = 11$), explosion crater (maar, tuff ring, and explosion crater) (n = 11), shield (shield and pyroclastic shield) ($n = 86$), submarine ($n = 74$), lava dome ($n = 26$), subglacial ($n = 5$), volcanic field ($n = 33$), and cone (pyroclastic cone, tuff cone, lava cone, and cone) ($n = 49$).

2.2. Regional (in)homogeneity

Volcanoes in the GVP are grouped into 19 regions that were adopted from the Catalogue of Active Volcanoes of the World (CAVW) (Neumann van Padang, 1951) and the Data Sheets of Post-Miocene Volcanoes (IAVCEI, 1973) (Siebert et al., 2011), and modified over time throughout the VOTW database versions. The delimitation of the original regions was partly due to the presence of the experts that contributed to compiling the data in each area (e.g., CAVW: Kurile Islands by Gorshkov (1958) or CAVW: New Zealand by Cole and Nairn (1975)). Later on, IAVCEI (1973) defined the regions in the Data Sheets of Post-Miocene Volcanoes, also known as sector maps, based on areas of particular density of volcanoes. As a result, the GVP regions, further divided into subregions, often cover large geographical areas with multiple tectonic settings and countries. Additionally, as Mead and Magill (2014) suggested, eruption data completeness analysis can be improved when volcanoes are grouped into regions based on their observation history. To test whether the subregions within each GVP region are characterised by similar degrees of completeness, we compared the Empirical Cumulative Distribution Function (ECDF) from the FRESH start dates across subregions using the two-sample K—S test. One advantage of this non-parametric test is that it does not require knowledge of the underlying FRESH distribution. The null hypothesis (H_0) states that the FRESH data from two subregions come from the same continuous distribution. If the *P*-value is lower than the significance level $\alpha = 0.05$, we reject H_0 and accept the alternative hypothesis (H_1) that the two subregions have significantly different ECDF (Massey, 1951). We iteratively compared the ECDF from one subregion against the ECDF from all the remaining subregions, following a leave-one-out cross-validation approach (e.g., we tested the FRESH ECDF from Sumatra against the FRESH ECDF from the whole of Indonesia except Sumatra). To account for the multiple comparison problem, we applied the Bonferroni correction to the P-value ($P\text{-}value_B$) based on the number of statistical tests performed (n_{tests}), which corresponds to the number of subregions within each region (1). This conservative method reduces the chances of obtaining false positives (Shaffer, 1995). Lastly, each subregion with a significantly different ECDF was considered as a new region, while the other subregions remained grouped in the same region.

$$P - value_B < P - value \times n_{tests} \tag{1}$$

2.3. Abrupt change points

In this study, we propose applying a straightforward approach to automatically identify multiple change points (k) followed by either a gap in the record, a decrease or an increase in the slope of the cumulative number of FRESH over time (i.e., FRESH rate). These change points were identified using the *findchangepts* function implemented in MATLAB, which detects automatically the point(s) where a statistical property of a given signal changes most significantly (Killick et al., 2012; Mathworks, 2021). The signal was defined as the time between consecutive FRESH start dates, and the RMS level was chosen as the statistical property instead of the mean because we are seeking periods of constant rate.

The automatic steps in the *findchangepts* function as described in Mathworks (2021) are:

1. The signal (x) (i.e., time between consecutive FRESH start dates) is partitioned into $k + 1$ subsets at arbitrary data points.
2. The RMS level (xRMS) of the n-by-1 signal, where n is the number of FRESH, is calculated for each subset (2). This statistic gives a smoothed level for the subset against which variability can be measured.

$$x\text{RMS} = \sqrt{\frac{1}{n}\sum_{r=1}^{n}|x_r|^2} \tag{2}$$

3. For a given signal data point in a subset, the deviation of the estimated RMS level of the subset from that signal data point is calculated as the log weighted dispersion. Then, the log weighted dispersion from all the data points in a given subset is summed to obtain the total log weighted dispersion (i.e., total deviation) in a subset:

$$(n-m+1)log\left(\frac{1}{n-m+1}\sum_{r=m}^{n}x_r^2\right)$$

(Mathworks, 2021)

4. The total log weighted dispersions from all the subsets are summed to obtain the total log weighted dispersion of the signal (x).
5. The location of the partition data point (k) is changed repeatedly until the total log weighted dispersion of the signal is minimised by using an exhaustive algorithm based on dynamic programming with early abandonment (Mathworks, 2021).

To find the optimal number of change points for a maximum $k = 50$, we used the elbow method (Fig. 1A), which is commonly applied in clustering problems (e.g., Antunes et al. (2018)). The elbow point represents the inflexion in the curve where the total log weighted dispersion improvement obtained from adding more k declines the most. We identified the elbow by minimising the sum of errors derived from fitting two lines at each side of each data point. Additionally, we forced $k = 1$ to find the most abrupt change point in the signal, which we define as the RCD if it is followed by a sharp increase in the FRESH rate indicating a significant improvement in eruption recording relative to the previous portion of the database (Fig. 1B). In other words, we identify the time window between the RCD and the present as the *most* complete (but not necessarily 100% complete) portion of the FRESH database. We searched for the change points in the FRESH database as a function of VEI, primary volcano type, tectonic setting, major rock composition, and region.

With this approach, we assumed that the occurrence of FRESH is a stationary process and that the global/regional FRESH database fits a Poisson distribution (De la Cruz-Reyna, 1991; Papale, 2018; Papale et al., 2022). We attribute the changes in the slope to variations in the completeness since i) Newhall and Self (1982) confirmed that global and regional eruption records are more strongly influenced by under-recording than by non-stationary processes, and b) we cannot capture variations in regional or single volcano eruption rates from looking at one eruption per volcano.

3. Results

3.1. FRESH database

Here, we provide our key findings and discuss potential reasons and implications in the Discussion section. Fig. 2 shows that nearly half of

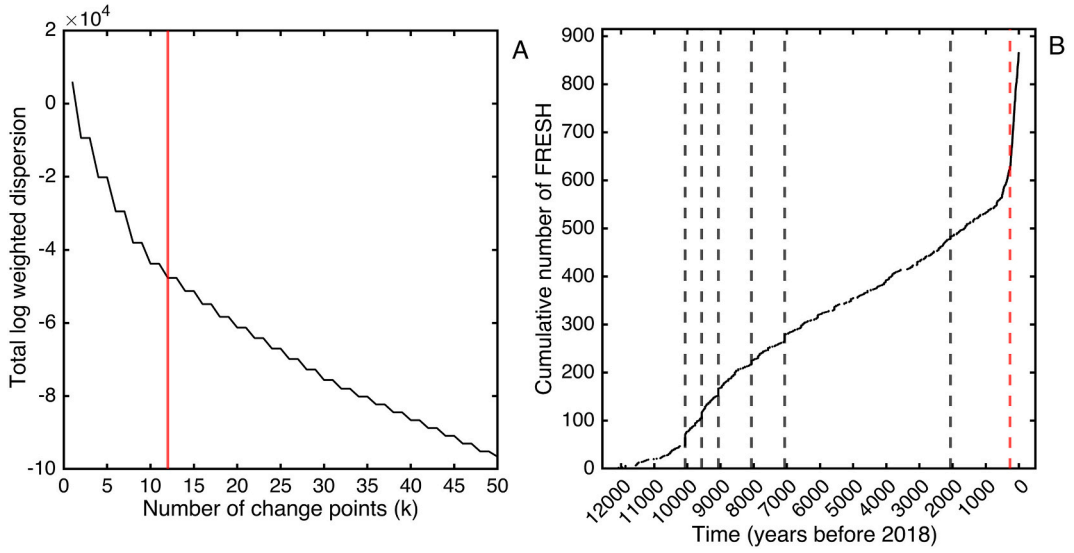

Fig. 1. Example of the abrupt change point approach using the RMS level to identify multiple change points in the global FRESH database. A: Elbow point plot and selection of the optimal number of change points ($k = 12$); B: Cumulative number of FRESH over time with the change points indicated by dashed black lines and the most abrupt change point (i.e., RCD = 1747 CE) indicated by a red dashed line. Fig. B has 7 change points instead of 12, because multiple FRESH were reported in the same year, and therefore, two or more change points were assigned to that year (e.g., around 10 ka). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

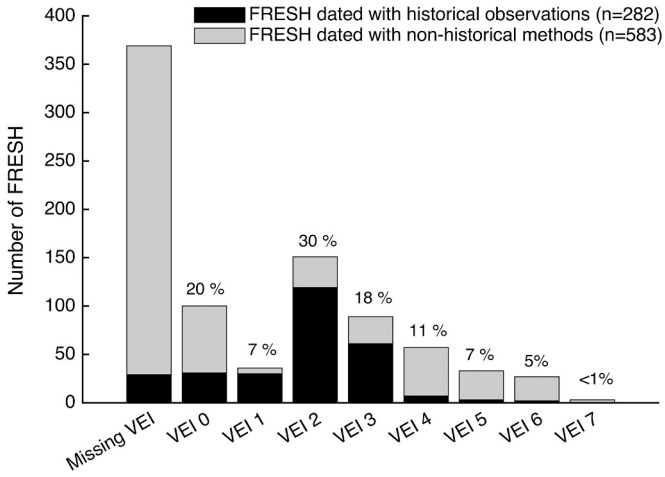

Fig. 2. Histogram of FRESH as a function of VEI and as a function of evidence method reported in the VOTW database. The percentage of FRESH per VEI is calculated from the total number of FRESH with assigned VEIs ($n = 496$).

the global FRESH database lacks a VEI (43%, $n = 369$), out of which 340 are FRESH dated with non-historical methods (e.g., tephrochronology, radiocarbon, and Ar/Ar, among other techniques (Siebert et al., 2011)). The remaining data show that the most common eruption size is VEI 2, followed by VEI 0 and VEI 3. Meanwhile, VEI ≥ 4 accounts for 24% of the FRESH with a known eruption size. The evidence method (i.e., dated with historical observations or non-historical methods) shows that the proportion of historical FRESH (i.e., dated with historical observations) decreases as the VEI increases. The exception is VEI 0, which has a high proportion of FRESH dated with non-historical methods.

The categories of tectonic setting, primary volcano type, and major rock composition with the highest percentage of historical FRESH are rift intermediate (82%, $n = 9$), submarine (55%, $n = 41$), and foidite (43%, n = 3) (Table 1). In comparison, rift continental (12%, $n = 6$), lava dome (15%, n = 4), and rhyolite (11%, n = 4) have the lowest percentages of historical FRESH.

The heatmap of FRESH as a function of tectonic setting and eruption size shows that VEI 0 FRESH are predominant in rift oceanic and intraplate oceanic settings (Fig. 3A). On the other hand, most of the largest FRESH (VEI 6 and 7) were recorded at subduction continental setting volcanoes. Submarine volcanoes contribute to most VEI 0 FRESH, followed by stratovolcanoes and shield volcanoes (Fig. 3B), while calderas and stratovolcanoes produce the most VEI 6 FRESH. Fig. 3C shows that stratovolcanoes are the most abundant volcano type in most tectonic settings. The exceptions are rift oceanic; intraplate oceanic, and rift intermediate; and intraplate intermediate; where submarine, shield, and cones, respectively, are the most frequent morphology. Note that the results shown in Fig. 3C are independent of the FRESH occurrence, given that tectonic setting is an intrinsic characteristic of each volcano, and the primary volcano type corresponds to the largest morphological feature (Siebert et al., 2011).

3.2. Redefining volcanic regions

Based on the $P\text{-}value_B$ of the K—S test results (Table S2 in Supplementary Material), we expanded the 19 GVP regions to 31 regions, and used them to assess eruption data completeness at a regional level (Fig. 4). Two different outcomes are possible from this analysis:

a) None of the GVP subregions within a given GVP region have a significantly different ECDF (i.e., $P\text{-}value_B \geq 0.05$). Therefore, we do not modify the GVP region (e.g., Region 1 (Alaska) in Fig. 4).
b) At least one GVP subregion within a given GVP region has a significantly different ECDF (i.e., $P\text{-}value_B < 0.05$). Therefore, we modify the GVP region as follows:

i) Each GVP subregion with $P\text{-}value_B < 0.05$ becomes a new region (e.g., the two GVP subregions of Tonga Islands and New Zealand, originally grouped in the GVP region of New Zealand to Fiji, are now the new regions with numbers 4 and 31, respectively, in Fig. 4).
ii) The other GVP subregions with $P\text{-}value_B \geq 0.05$ remain grouped, becoming another new region (e.g., Fiji Islands; Wallis and Samoan Islands; and Kermadec Islands (Region 3 in Fig. 4), originally grouped in the GVP region of New Zealand to Fiji).

Fig. 5 shows the significant differences in the FRESH distributions of

Table 1
Categories of tectonic setting, primary volcano type, and major rock composition; number of FRESH; percentage of historical FRESH; and most abrupt change point in the cumulative number of FRESH. The categories are ordered from highest to lowest percentage of historical FRESH. The most abrupt change points in italics represent the RCD, which indicates a relative improvement in recording as described in Section 3.3, while the non-italicised most abrupt change points indicate that a decrease in the FRESH rate or a gap follows the most abrupt change point. NA: RMS level was not applicable due to the small sample size.

Category		Number of FRESH	Historical FRESH (%)	Most abrupt change point (calendar year)
Tectonic setting	Rift Intermediate	11	81.8	*1906 CE*
	Subduction Unknown	51	70.6	*1543 CE*
	Subduction Oceanic	88	60.2	*1510 CE*
	Subduction Intermediate	51	58.8	*1713 CE*
	Intraplate Intermediate	3	33.3	NA
	Rift Oceanic	75	30.7	*1562 CE*
	Intraplate Oceanic	23	26.1	*1966 CE*
	Subduction Continental	447	23.7	*1680 CE*
	Intraplate Continental	64	18.8	*1883 CE*
	Rift Continental	52	11.5	1075 CE
	Submarine	74	55.4	*1773 CE*
	Complex	43	44.2	*1512 CE*
	Stratovolcano	462	36.4	*1500 CE*
	Fissure	11	36.4	7050 BCE
	Caldera	65	21.5	*1712 CE*
Primary volcano type	Volcanic Field	33	21.2	200 CE
	Subglacial	5	20.0	7050 BCE
	Explosion Crater	11	18.2	1050 BCE
	Shield	86	16.3	*1814 CE*
	Cone	49	16.3	*1842 CE*
	Lava Dome	26	15.4	1695 CE
	Foidite	7	42.9	2750 BCE
	Dacite	21	38.1	*1773 CE*
	Trachyandesite/Basaltic Trachyandesite	8	37.5	1891 CE
	Trachybasalt/Tephrite Basanite	11	36.4	850 CE
Major rock composition	Andesite/Basaltic Andesite	557	34.5	*1510 CE*
	Basalt/Picro-Basalt	204	24.5	*1750 CE*
	Trachyte/Trachydacite	9	22.2	6540 BCE
	Phonolite	11	18.2	6940 BCE
	Rhyolite	18	11.1	4450 BCE

the new regions that were originally grouped in the same GVP region. For example, Hawaiian Islands vs. Pacific Ocean; Honshu vs. Izu, Volcano, and Mariana Islands; and New Zealand vs. Tonga Islands. Regions located in the left-hand side of Fig. 5 have records that go further back in time, with most FRESH dated with non-historical methods (Table 2), while those on the right-hand side of Fig. 5, have predominantly historical records. The contrasting FRESH records found among subregions that were grouped in the GVP demonstrate the importance of assessing eruption data completeness at an appropriate scale to avoid aggregating different eruption probability distributions.

3.3. Dates of improved completeness

This section will present the most abrupt change points and RCDs for the global FRESH database and as a function of VEI, volcanic characteristics, and regions in turn. The remaining change points are presented in Table S3 in the Supplementary Material. We identified the most abrupt change point as the RCD if it was followed by a sharp increase in the FRESH rate. Therefore, we assume that from this date, the FRESH database is more complete than the previous portion of the catalogue. For the global FRESH database, the results show that the recording of FRESH improved after 1747 CE (Fig. 1B).

3.3.1. VEI

The RCD for VEI 1 (1730 CE) and VEI 2 (1748 CE) FRESH is close to the global RCD, while the RCD for VEI 0 (1951 CE), VEI 3 (1510 CE), and VEI 4 (1600 CE) differ widely from the global RCD. The RCD cannot be defined for larger FRESH since the record follows a near-constant rate across the Holocene, and the most abrupt change point for VEI 5 (6650 BCE) and VEI 6 FRESH (50 CE) is not followed by an increase in the FRESH rate. The RMS level was not applied to VEI 7 FRESH ($n = 3$) because of the small sample size.

3.3.2. Volcanic characteristics

RCDs in Table 1 range from the 16th century through to the 18th, 19th or 20th century when FRESH are categorised by major rock composition, primary volcano type or tectonic setting, respectively. Several categories, such as subduction oceanic, complex volcano, stratovolcano, and andesite/basaltic andesite, have similar RCDs ranging between 1500 and 1512 CE. In contrast, the most recent RCDs correspond to Rift Intermediate (1906 CE), which contains only volcanoes from Africa (northeastern) and Red Sea, and Intraplate Oceanic (1966 CE), which contains mostly volcanoes from the three regions of Hawaiian Islands, Atlantic Ocean, and Pacific Ocean.

3.3.3. Regions

Table 2 shows the most abrupt change points and RCDs at a regional scale. 21 out of 31 regions have been assigned a RCD. For 16 of these regions, the RCD falls within the last 500 years. Two regions have the RCD in the 16th century, two in the 17th century, five in the 18th century, five in the 19th century, and two in the 20th century. The Indian Ocean (southern) region has the most recent RCD (1980 CE), but the database has only five FRESH in this region. The Pacific Ocean, with a larger sample size ($n = 21$), has a similar RCD in 1964 CE. The Middle East and Western Indian Ocean region has the earliest RCD in the Common Era (200 CE), while the Mexico, Guatemala, Nicaragua, Costa Rica, and Panama region has the oldest RCD Before the Common Era (7930 BCE).

Additionally, the characteristics of several regions prevented us from defining RCDs. First, we found that several regions have a near-constant rate of FRESH over the Holocene (e.g., Kamchatka and Mainland Asia). Second, in regions like Iceland and the Arctic Ocean, Africa (Eastern), and Honshu, the most abrupt change point is followed by a gap in the FRESH record or a decrease in the FRESH rate. Third, several regions have small sample sizes (e.g., Hawaiian Islands ($n = 6$), El Salvador and Honduras ($n = 5$), and Greece ($n = 4$)), which makes identification of

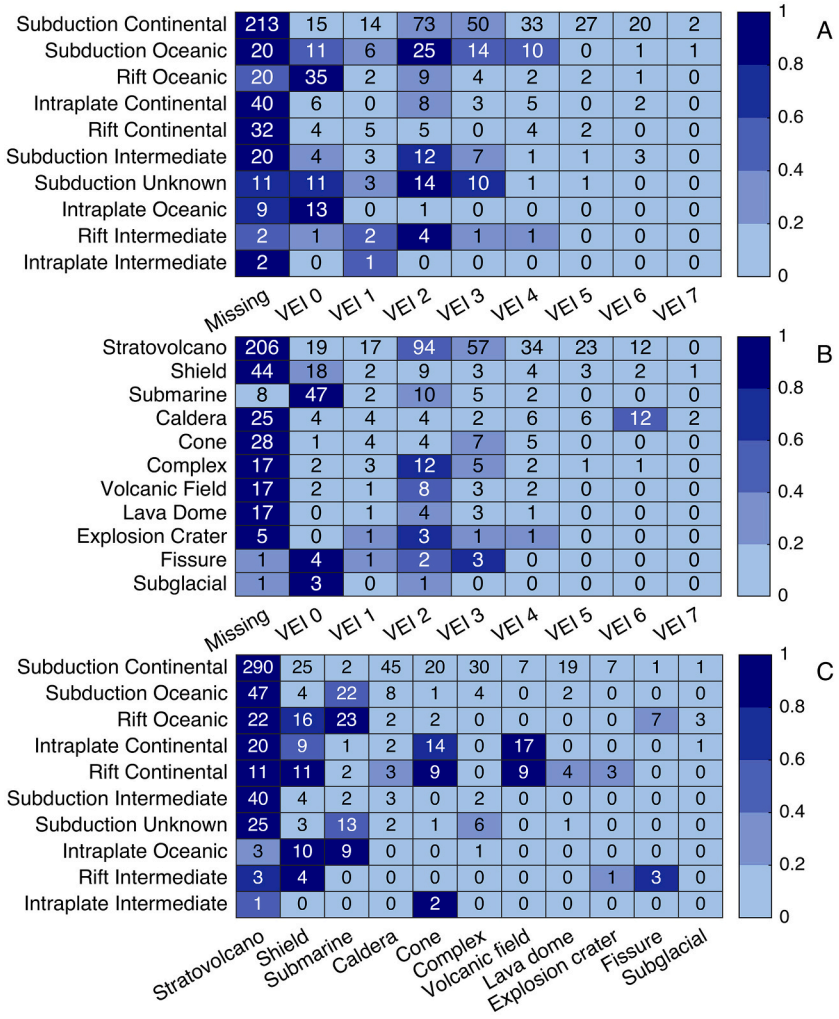

Fig. 3. Contingency table represented with a heat map for A: FRESH count as a function of tectonic setting and VEI; B: FRESH count as a function of primary volcano type and VEI; C: FRESH count as a function of tectonic setting and primary volcano type. The colour scale is normalised along each row with the smallest value of FRESH corresponding to light blue and the largest to dark blue. Except for VEI, categories are listed in descending order of abundance from top to bottom and left to right. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

any temporal changes challenging. For more details on the temporal distribution, refer to Fig. S1 in the Supplementary Material, which contains the regional plots of the cumulative number of FRESH.

4. Discussion

4.1. Eruption size distribution

Despite the FRESH database being a subset of the VOTW database, there are more differences than similarities among both records when looking at the eruption size distribution. Of the similarities, both the VOTW database and FRESH databases have the most common eruption size as VEI 2, after which the number of eruptions decreases with increasing eruption size (Siebert et al., 2011) (Fig. 2). The predominance of VEI 2 FRESH corresponds with this value having previously been assigned as the default VEI to explosive eruptions without detailed descriptions (Newhall and Self, 1982; Siebert et al., 2011).

Compared to the VOTW database, the FRESH database differs in three important aspects. The first difference is the high percentage of FRESH without a VEI (43%), which is almost double the percentage of eruptions without a VEI in the VOTW database (22%) (GVP, 2013). This is a result of the FRESH database considering the oldest dated Holocene eruption at each volcano. Although most FRESH without a VEI have been dated with non-historical methods, determining eruption volume and intensity from old eroded deposits can be challenging, even for large-magnitude eruptions (Brown et al., 2014).

The second difference is the higher percentage of VEI 0 eruptions in the subset of FRESH with assigned VEI (20%) (Fig. 2) than in the VOTW database with assigned VEI (11%) (GVP, 2013). Most VEI 0 FRESH are reported in oceanic settings at submarine volcanoes, or at stratovolcanoes or shield volcanoes (which may or may not be in oceanic settings). Therefore, the higher percentage could reflect the reporting of effusive submarine eruptions and dating of well-preserved old lava flows from stratovolcanoes or shield volcanoes.

The third, and most important, difference is that 24% of the FRESH with an assigned VEI have a VEI ≥ 4 (Fig. 2), while <10% of the VOTW database with an available VEI have a VEI ≥ 4 (GVP, 2013). Two reasons could explain this difference: i) the FRESH database contains a higher proportion of old eruptions, as reflected in the higher percentage of eruptions dated with non-historical methods (67%) than in the VOTW database (35%) (GVP, 2013), and deposits from large-magnitude eruptions are better preserved in the geological records than deposits from small eruptions (Brown et al., 2014; Rougier et al., 2018, 2016); and ii) many volcanoes without dated eruptions prior to their recording in the VOTW database might be long-dormant, and these volcanoes are more likely to reawaken with large explosive eruptions (Bebbington, 2014; Wadge, 1982).

4.2. Patterns across volcanic characteristics

There seems to be an influence of the tectonic setting in the occurrence of FRESH. For example, the low number of historical FRESH in continental settings could be due to the control of the crustal thickness and regional stress on eruption rates, resulting in relatively low annual

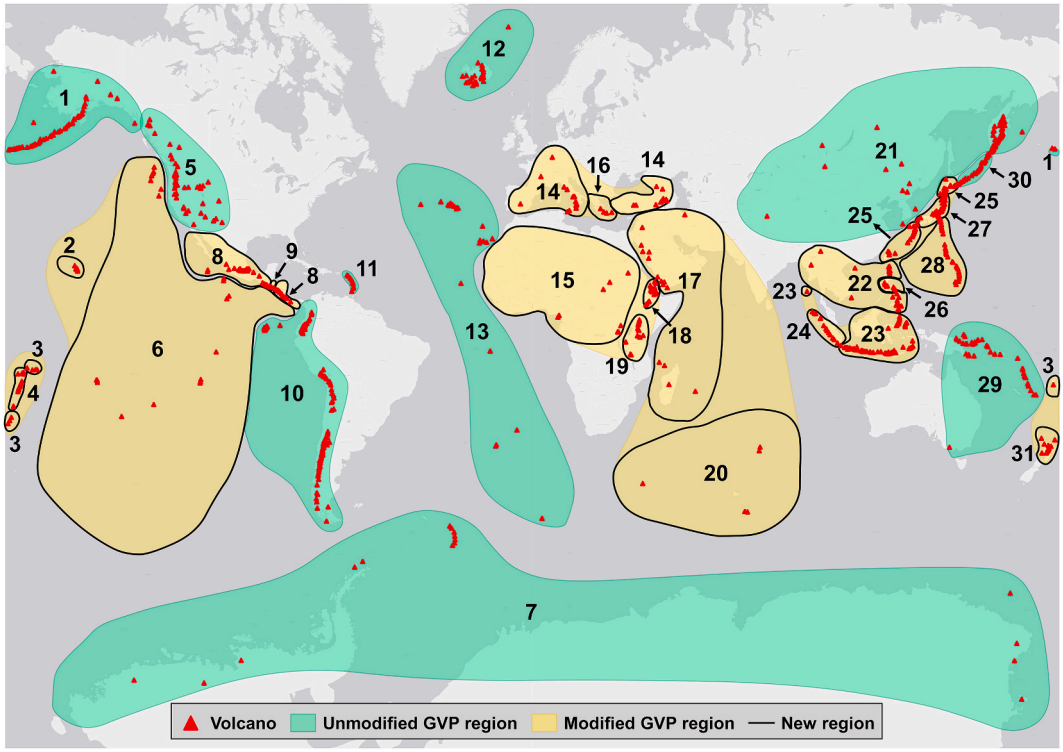

Fig. 4. Map showing unmodified GVP regions (scenario a), modified GVP regions (scenario b), and the new regions defined in this study (solid black line). The numbers in the map correspond to the region ID from Table 2. Of the 19 regions defined in the GVP, 10 kept the original delimitation, and 9 were divided into 21 new regions. Basemap: ESRI grey (light).

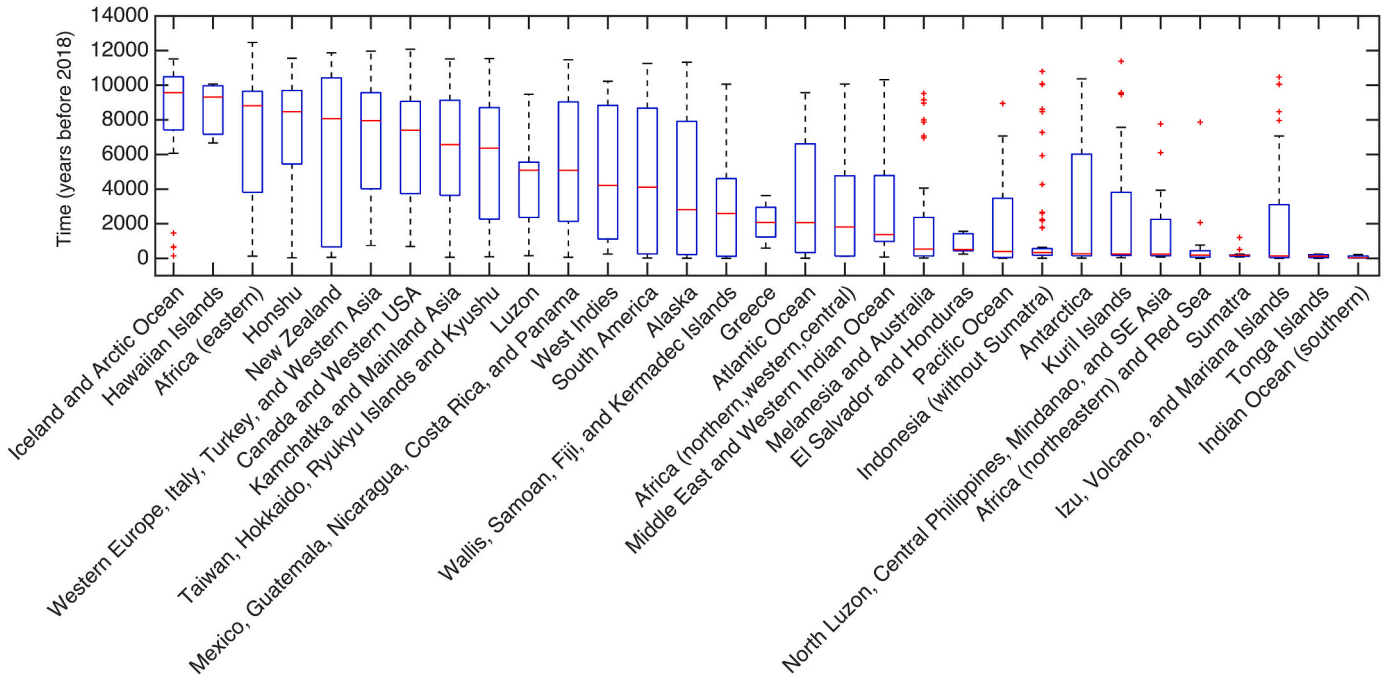

Fig. 5. Boxplot of the statistical distribution of FRESH start dates across regions ordered by a decreasing median represented by a horizontal red line. The box and the whiskers represent the interquartile range, and minimum and maximum values, respectively, excluding outliers, represented by the red crosses. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

magma budgets and larger ratios of intruded and extruded magma volumes in these settings (Crisp, 1984; Wadge, 1982; White et al., 2006). Similarly, the predominance of FRESH with a very large VEI in subduction continental crust settings could be explained by the thick continental crust favouring magma accumulation and differentiation (Mason et al., 2004).

The proportion of historical FRESH (i.e., dated with historical observations) is also highly variable among primary volcano types

Table 2
Regional identification number (ID), region, number of FRESH, date of the last FRESH, percentage of historical FRESH, and the most abrupt change point in the cumulative number of FRESH. The most abrupt change points in italics represent the RCD, while the non-italicised most abrupt change points indicate that a decrease in the FRESH rate or a gap follows the most abrupt change point. NA: RMS level was not applicable due to the small sample size.

ID	Region	Number of FRESH	Last FRESH (calendar year)	Historical FRESH (%)	Most abrupt change point (calendar year)
1	Alaska	59	2006 CE	37.3	*1760 CE*
2	Hawaiian Islands	6	4650 BCE	0	7050 BCE
3	Wallis, Samoan, Fiji, and Kermadec Islands	9	2012 CE	22.2	*1866 CE*
4	Tonga Islands	16	2008 CE	93.8	1791 CE
5	Canada and Western USA	48	1330 CE	0	4940 BCE
6	Pacific Ocean	21	2003 CE	19	*1964 CE*
7	Antarctica	16	2001 CE	43.8	*1819 CE*
8	Mexico, Guatemala, Nicaragua, Costa Rica and Panama	44	1952 CE	11.4	*7930 BCE*
9	El Salvador and Honduras	5	1770 CE	60	1510 CE
10	South America	115	1995 CE	27.8	*1748 CE*
11	West Indies	12	1766 CE	16.7	*790 CE*
12	Iceland and Arctic Ocean	25	1867 CE	12	5850 BCE
13	Atlantic Ocean	25	2004 CE	36	*1500 CE*
14	Western Europe, Italy, Turkey, and Western Asia	26	1282 CE	11.6	5820 CE
15	Africa (northern, western,central)	10	1891 CE	30	*1884 CE*
16	Greece	4	1422 CE	50	NA
17	Middle East and Western Indian Ocean	15	1937 CE	46.7	*200 CE*
18	Africa (northeastern) and Red Sea	17	2011 CE	76.5	*1631 CE*
19	Africa (eastern)	16	1888 CE	12.5	5850 BCE
20	Indian Ocean (southern)	5	1995 CE	80	*1980 CE*
21	Kamchatka and Mainland Asia	69	1951 CE	2.9	8050 BCE
22	North Luzon, Central Philippines, Mindanao, and SE Asia	19	1939 CE	63.2	*1765 CE*
23	Indonesia (without Sumatra)	66	2004 CE	78.8	*1510 CE*
24	Sumatra	14	1993 CE	92.9	*1770 CE*
25	Taiwan, Hokkaido, Ryukyu Islands and Kyushu	37	1924 CE	16.2	*1600 CE*
26	Luzon	8	1860 CE	12.5	*3500 BCE*
27	Honshu	39	1979 CE	5.1	2850 BCE
28	Izu, Volcano, and Mariana Islands	33	2013 CE	54.5	*1864 CE*
29	Melanesia, and Australia	42	1996 CE	42.9	*1835 CE*
30	Kuril Islands	31	1972 CE	61.3	*1712 CE*
31	New Zealand	13	1958 CE	7.7	*1180 CE*

(Table 1). Morphologies such as lava domes, cones, and shields are often the result of effusive activity that generates less-erodible deposits (de Silva and Lindsay, 2015), making it easier to date them with geological techniques, as reflected in the high percentage of FRESH dated with non-historical methods. On the other hand, the high percentage of submarine FRESH dated from historical observations can be partly attributed to advances in the last century in ocean exploration, leading to more eruptions reported in submarine volcanoes (Hammond et al., 2015).

Interpreting the results by composition is difficult since there is no clear relationship between the rock types and the percentage of historical FRESH. This is likely the result of the major rock composition corresponding to the most abundant rock type in terms of volume or the composition from younger edifices in long-living systems (Siebert et al., 2011). Therefore, we cannot be certain that this composition is the same one produced by the FRESH. Additionally, the sample size is highly variable across rock types, ranging from 7 FRESH for Foidite to 557 FRESH for Andesite//Basaltic Andesite.

4.3. Global eruption data completeness

The RCD obtained for the global database of FRESH (1747 CE) coincides with the sharp increase in the estimated probability of recording an eruption reported in Guttorp and Thompson (1991) (their Fig. 1). This probability was obtained by fitting a linear function to the yearly eruption count from the Volcano Reference File (Simkin et al., 1981) and arbitrarily assuming that after 1980, the probability of recording an eruption was 100%. Rougier et al. (2016) also assumed 1980 as the completeness date for Magnitude≥4 eruptions in the LaMEVE database, while Papale (2018) found that the eruption rate for all VEIs remains constant after 1975, and Wang et al. (2020) proposed that the VOTW database is complete after 1985. Other studies have identified an increase in the recording rate between 500 and 200 years ago, coinciding with the start of the European expeditions to the Americas and Asia, and followed by the invention of the printing press, the Industrial Revolution, and globalisation (Mead and Magill, 2014; Siebert et al., 2011).

Two factors can explain why we obtain an earlier global relative completeness date in comparison with previous studies: i) We are looking at a *relative* improvement while completeness dates imply a 100% probability of recording an eruption, and ii) in contrast with previous research on eruption data completeness, we include in the analysis eruptions lacking a VEI since they account for nearly half of the FRESH database and around 20% (n = 2179) of the eruptions in the VOTW database (GVP, 2013).

The relatively recent RCDs obtained for VEI 1–4 FRESH reflect the poor preservation of eruption deposits from small to moderate explosive eruptions in geological records due to erosive and alteration processes (Siebert et al., 2011), which are highly variable among environmental conditions and volcanoes (Blong et al., 2017). The sharp increase in the FRESH rate following the RCD for VEI 1–4 also captures the improvement in eruption records observed in global catalogues (Mead and Magill, 2014; Siebert et al., 2011). The FRESH RCDs could be considered early dates compared to global completeness dates found in other studies. For example, the global completeness dates proposed by Papale (2018) range from 1840 for VEI 4 to 1950 for VEI 1. The difference with our results can be partly explained by the method applied by Papale (2018), which identifies absolute completeness dates from the best linear fit of a unified database of LaMEVE and VOTW databases. In contrast with these results, the completeness date obtained by Papale (2018) for VEI 0 eruptions (1950) is almost identical to the RCD for VEI 0 FRESH (1951 CE) (Fig. 6). This coincidence could be partially attributed to the development of the hydrophone during World War II and its application to detect submarine volcanism, which likely resulted in a global improvement (Ewing et al., 1946). Neither the under-recording of large explosive eruptions (Deligne et al., 2010) nor the improvement on M ≥ 4 eruption recording observed after the 15th century in global records (Rougier et al., 2016) are reflected in the near-constant rate of VEI 5, 6, and 7 FRESH. This difference could be due to the smaller sample size and shorter time scale used in our study. The challenge of identifying change points for the scarce records of large eruptions was also encountered by Mead and Magill (2014), resulting in greater

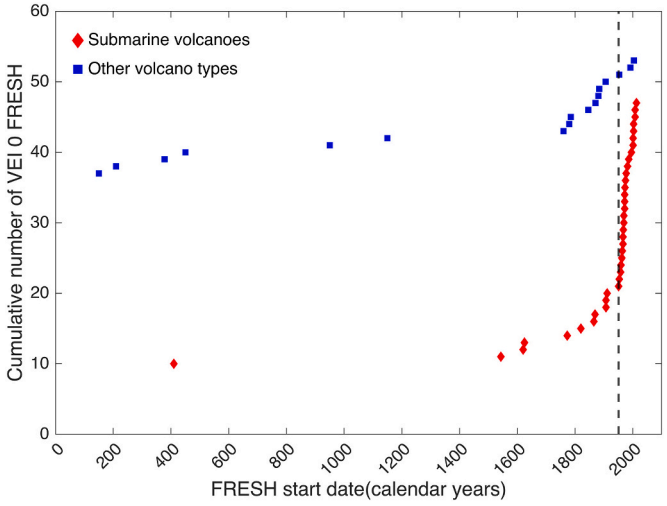

Fig. 6. Cumulative number of VEI 0 FRESH from submarine volcanoes and other volcano types reported since the start of the Common Era. The RCD for all VEI 0 FRESH (1951 CE) is indicated by a black dashed line, which coincides with the application of hydrophones to detect submarine volcanism (Ewing et al., 1946).

uncertainties around the completeness dates.

4.4. Regional eruption data completeness

We propose four factors that explain the different degrees of completeness in the FRESH database across regions and subregions. The completeness of the eruption record in a given region can be influenced by one or more of the following factors. The first factor impacting eruption data completeness, widely acknowledged in previous studies (e.g., Mead and Magill, 2014; Rougier et al., 2018; Siebert et al., 2011), is the occurrence of socio-historical events, such as the European Age of Discovery. This effect is also observed in the sharp increase in the FRESH rate in several regions (Fig. 7A). For example, the RCD in Indonesia (without Sumatra) (1510 CE) seems to be influenced by the presence of the Europeans in the 16th century and the later establishment of the Dutch East Indian Company in the early 17th century (Rougier et al., 2018; Siebert et al., 2011). In contrast, the presence of the Europeans in Sumatra is not reflected instantly as an improvement in the FRESH database. Although the Dutch and British established trading companies in west Sumatra in the second half of the 17th century (Colombijn, 2005), the improvement in eruption recording is not observed until more than a hundred years later (1770 CE). The difference in the RCDs between Sumatra and the rest of Indonesia could be explained by the contrasting type of volcanism. While the volcanism in Sumatra is explosive and infrequent, often coming from long-dormant calderas, the volcanism in the rest of Indonesia, especially in Java, is less explosive and more frequent (Bouvet de Maisonneuve and Bergal-Kuvikas, 2020). These examples suggest that regional eruption frequency could also be a factor when comparing eruption data completeness at different regions.

A delay between socio-historical events and an improvement in the recording is also apparent within the region of North Luzon, Central Philippines, Mindanao and SE Asia, where the RCD, 1765 CE, does not seem to be linked to the Spanish settlements in the early 16th Century (Siebert et al., 2011). This delay could also be reflecting the regional eruption frequency. However, the influence of the regional eruption frequency in the FRESH rate is difficult to establish due to the complex geological setting of the Philippines (Bouvet de Maisonneuve and Bergal-Kuvikas, 2020). One factor that could have contributed to more FRESH reported after 1765 is the opening to international trade towards the first half of the 19th century leading to a higher presence of, for example, European merchants in the region (Seekins, 1993) who often recorded eruptions in ship logs that were later conserved and published. In contrast, the FRESH database in Luzon (i.e., excluding North Luzon,

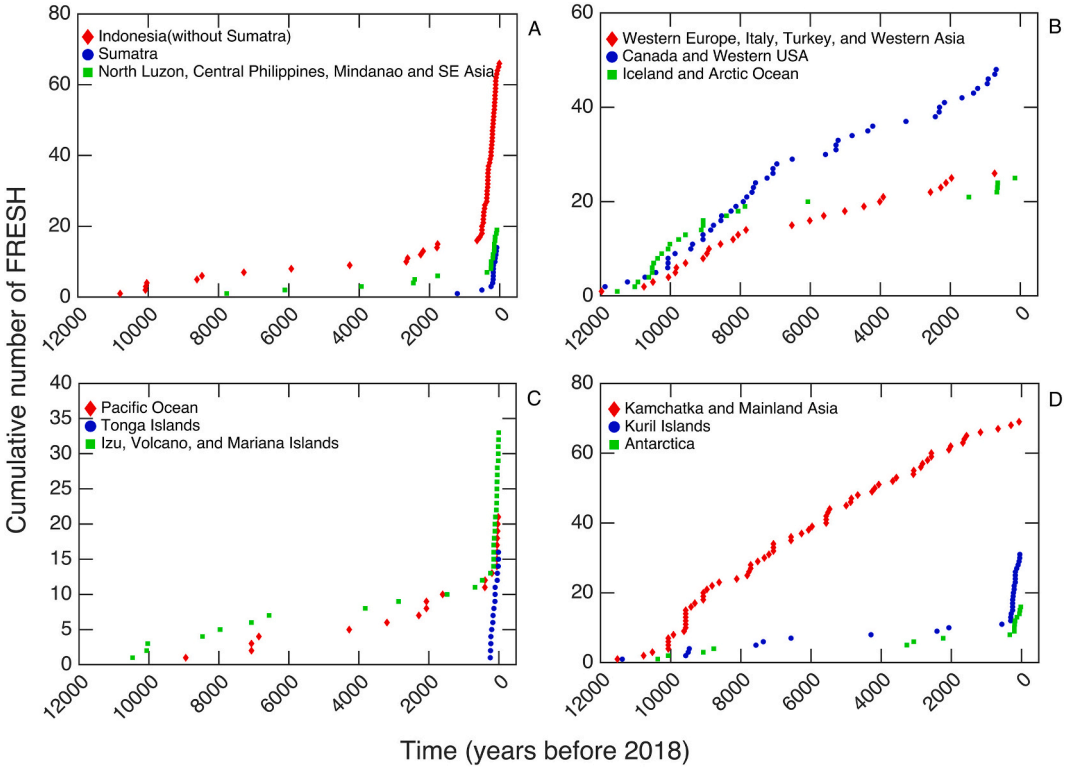

Fig. 7. Cumulative number of FRESH against start date for a selection of regions with a record dominated by FRESH dated with historical observations (A), regions with a record dominated by FRESH dated with non-historical methods (B), regions dominated by submarine volcanism (C), and remote regions (D).

which is part of the abovementioned region) is dominated by eruptions dated with non-historical methods. This difference could be because some of the volcanoes in Luzon have produced large explosive eruptions in the past (e.g., Pinatubo and Taal) or erupt frequently (e.g., Mayon) (Aspinall et al., 2011). Additionally, the capital city, Manila, and other major cities in Luzon are surrounded by volcanoes, resulting in a high Population Exposure Index (PEI $\geq$ 5) at most volcanoes in Luzon (Brown et al., 2014b). These two reasons might have attracted the attention of volcanologists and the local authorities, leading to more geological studies in this region.

The second factor, which has also been discussed previously, is the number of geological studies conducted in different regions as reflected in the VOTW database. Several regions show a comprehensive geological record of FRESH (Fig. 7B), which in some cases results in a near-constant FRESH rate throughout the Holocene (e.g., Canada and Western USA). As a result, most of these regions lack an RCD since the most abrupt change point often corresponds to a deceleration in the FRESH rate (e.g., Western Europe, Italy, Turkey, and Western Asia) or a gap in the record (e.g., Iceland and Arctic Ocean). In these cases, having multiple change points can be helpful to overcome the uncertainty in regions with extensive geological studies and long historical records, as noted by Mead and Magill (2014). The comprehensive records of some of these regions can be attributed to specific research projects focused on studying the volcanism in certain areas (e.g., Kamchatka and Mainland Asia (IVS, 2021)). An interesting example is the two nearby regions, Africa (northeastern) and Red Sea, and Africa (eastern), with contrasting degrees of completeness (Fig. 5). In the former, most of the FRESH are historical, while in the latter most FRESH are dated with non-historical methods. In this case, the comprehensive record of Africa (eastern) could be partly attributed to the numerous geological studies conducted by the local researchers and international collaborators at the East African Rift System (EARS) (e.g., a study of the geology from the northern sector of the Kenyan Rift Valley (Dunkley et al., 1993) or a detailed stratigraphy of the Holocene explosive volcanism at Rungwe Volcanic Province (Tanzania) (Fontijn et al., 2010)). Recent projects, such as RiftVolc (https://www.riftvolc.geos.ed.ac.uk) (2014–2021) (BGS, 2021), which is focused on Ethiopian volcanism, will likely result in an increase in the number of older FRESH dated with non-historical methods in the region of Africa (northeastern) and Red Sea. Additionally, the volcanic record from this region might also benefit from the recent increased interest in the geothermal energy potential at the EARS and the associated research required for its exploitation (Kombe and Muguthu, 2018).

The third factor is the presence of submarine volcanism. In regions with frequent submarine activity, such as the Pacific Ocean; Tonga Islands; and Izu, Volcano and Mariana Islands (Fig. 7C), the majority of FRESH are dated from historical observations, and as a result the RCDs are relatively recent (Table 2). The improvement in the number of reported FRESH in these regions can be in part attributed to the advances in seismic and acoustic detection techniques since the mid-20th century (Ewing et al., 1946) (e.g., first correlation between submarine explosions recorded by the Sofar network and the 1954 submarine eruption at Myojin (Japan) (Dietz and Sheeny, 1964)). Ocean exploration programs, such as the National Oceanic and Atmospheric Administration (NOAA) Vent Program launched in 1983, which contributed to almost half of the submarine vents compiled in the InterRidge Vents database (Beaulieu et al., 2013), have also played an important role in this improvement. For example, the first direct visual observations of a submarine explosive eruption were made at NWRota-1 (Mariana Arc) by remotely operated vehicles used during three diving expeditions conducted between 2004 and 2006 (Chadwick Jr et al., 2008). This type of observation would not have been possible in the mid-late 20th century, especially for deep-sea volcanism (i.e., >500 m depth) (see Rubin et al. (2012) for a detailed chronology of submarine volcanism discoveries).

The fourth factor behind the different degrees of completeness is remoteness (Fig. 7D), which is also a common factor in regions with submarine volcanism. The difficult access to remote continental volcanoes and volcanic islands often results in scarce geological records. Additionally, remote regions are largely uninhabited, and geological studies tend to focus on volcanoes that pose a risk to the population living nearby. An apparent exception of a well-studied remote region is Kamchatka and Mainland Asia, which has comprehensive geological studies (e.g., Ponomareva et al., 2007) thanks to the fieldwork conducted mainly by the Institute of Volcanology and Seismology (IVS, 2021). Interest in studying these volcanoes is motivated mainly by the risk posed to aviation and the local population (Eichelberger et al., 2013). The low population density might also result in scarce historical accounts as we go back in time, with some regions showing an apparent improvement in eruption recording with the arrival of people. For example, the inaccessible Kuril Islands region, which lacks geological studies, shows an increase in reported FRESH after 1712 CE. This increase could be attributed to the arrival of the Cossacks in 1711 CE and the multiple mapping expeditions that followed (Gorshkov, 1970). Lastly, eruption records from remote regions have improved recently thanks to remote sensing techniques that detect eruptive plumes, gas emissions, and thermal signals from subaerial eruptions (e.g., FRESH at Montagu Island (Antarctica) in 2001 CE; Patrick et al. (2005)).

As mentioned above, some of these factors have been previously discussed in other studies looking at regional eruption completeness. However, we do not compare our regional RCDs with completeness dates proposed by other studies (e.g., Jenkins et al. (2012) and Mead and Magill (2014)) because we: i) analyse different subsets of eruptions (i.e., FRESH vs multiple eruptions per volcano); ii) include eruptions lacking an assigned VEI; and iii) use a different spatial scale.

5. Conclusions

The First Recorded EruptionS in the Holocene (FRESH) database has been compiled from the Volcanoes of the World (VOTW) database to look at eruption data completeness from a new perspective. We applied a straightforward method to identify changes in the completeness of eruption records based on the most abrupt change point approach implemented in the *findchangepts* function (Mathworks, 2021). This function can identify multiple change points in the eruption record by looking at abrupt changes in the RMS level of the time-lapse between consecutive FRESH start dates. With this approach, we identify gaps in the record, a decrease, or an increase in the recording rate. Furthermore, we propose new relative completeness date(s), RCDs, indicating an improvement in eruption records identified from the most abrupt change point in the FRESH rate. The completeness of the FRESH database was analysed as a function of VEI, tectonic setting, morphology, rock composition, and region. Results show significant variability across eruption size and volcanic characteristics (e.g., 1951 CE for VEI 0 FRESH vs 1600 CE for VEI 4 FRESH or 1951 CE for intraplate oceanic volcanoes vs 1510 CE for subduction oceanic volcanoes).

Since a regional variability in the degree of under-recording could also be expected, we applied the K—S test to identify which subregions within the GVP regions have significantly different FRESH distributions. As a result, we have expanded the number of regions from 19 to 31, separating subregions within regions with different degrees of completeness (e.g., Hawaiian Islands with an extensive geological record vs Pacific Ocean with a record dominated by historical FRESH) and considering them as new regions. These results confirm that some GVP regions are not suitable for analysing eruption data completeness at a regional level and highlight the importance of using an appropriate geographic scale to reduce uncertainties in eruption recurrence and probability estimations. Furthermore, we discourage assessing eruption data completeness by country since single countries frequently include regions with different degrees of completeness.

We propose four key factors that explain the different degrees of completeness observed in the FRESH database at a regional scale: i) socio-historical events, frequently linked to the European Age of

Discovery, which contributed to having more historical observations (e.g., Indonesia (without Sumatra)); ii) number of geological studies conducted, resulting in regions with comprehensive eruption records (e.g., Canada and Western USA); iii) a dominant presence of submarine volcanism, where a recent improvement in eruption recording is attributable to the advances in submarine eruption detection and monitoring, and the expansion of ocean exploration (e.g., Izu, Volcano, and the Mariana Islands); and iv) remoteness, which usually implies low population exposure or difficult access to the field, leading to scarce geological records and historical accounts (e.g., Antarctica). One or more of these factors can explain the different degrees of completeness in a given region. Therefore, it is important to consider eruption under-recording as a multifactorial problem when assessing data completeness.

Since the FRESH database represents a portion of the VOTW database, the RCDs presented here should only be used for estimations derived from the FRESH database or when looking at volcanoes without Holocene records (i.e., their first known eruption will count as a FRESH). We suggest that new RCDs should be defined when estimating eruption frequency and probability from other datasets and assessing eruption data completeness for different temporal and spatial scales.

This study provides a new look into eruption data completeness by applying a straightforward method that identifies multiple change points in the recording rate while bringing insight into the occurrence of the first dated eruption from Holocene volcanoes. A follow-up project will draw on the FRESH database and the new proposed RCDs to estimate the recurrence interval and probabilities of having a FRESH at volcanoes lacking records of Holocene activity.

Funding

This research was supported by the Earth Observatory of Singapore via its funding from the National Research Foundation Singapore and the Singapore Ministry of Education under the Research Centres of Excellence initiative. This work comprises EOS contribution number 470. Mark Bebbington was supported by the Resilience to Nature's Challenges Volcano program, New Zealand (contract number GNS-RNC047).

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The Supplementary Materials are deposited in the NTU open access research data repository DR-NTU (Data).

Table S1 (SM1): https://doi.org/10.21979/N9/VTNTGW
Table S2 (SM2): https://doi.org/10.21979/N9/S9SICE
Table S3 (SM3): https://doi.org/10.21979/N9/NSVP5W
Figure S1 (SM4): https://doi.org/10.21979/N9/J8PLXZ

Acknowledgements

We are grateful to the anonymous reviewers for their thoughtful and constructive comments which improved the manuscript, and to Sonia Calvari for his editorial handling of the manuscript. We would like to thank Pavel Adamek for language editing and proofreading the manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jvolgeores.2022.107648.

References

Antunes, M., Gomes, D., Aguiar, R.L., 2018. Knee/Elbow estimation based on first derivative threshold. In: Proc. - IEEE 4th Int. Conf. Big Data Comput. Serv. Appl. BigDataService 2018, pp. 237–240. https://doi.org/10.1109/BigDataService.2018.00042.

Aspinall, W., Auker, M., Hincks, T., Mahony, S., Nadim, F., Pooley, J., Sparks, S., 2011. Volcano Hazard and Exposure in Track II Countries and Risk Mitigation Measures-GFDRR Volcano Risk Study. The World Bank.

Beaulieu, S.E., Baker, E.T., German, C.R., Maffei, A., 2013. An authoritative global database for active submarine hydrothermal vent fields. Geochemistry. Geophys. Geosyst. 14, 4892–4905. https://doi.org/10.1002/2013GC004998.

Bebbington, M.S., 2014. Long-term forecasting of volcanic explosivity. Geophys. J. Int. 197, 1500–1515. https://doi.org/10.1093/gji/ggu078.

BGS, 2021. Rift Volcanism: Past, Present and Future. Br. Geol. Surv. https://www.bgs.ac.uk/geology-projects/volcanoes/riftvolc/.

Blong, R., Enright, N., Grasso, P., 2017. Preservation of thin tephra. J. Appl. Volcanol. 6 https://doi.org/10.1186/s13617-017-0059-4.

Bouvet de Maisonneuve, C., Bergal-Kuvikas, O., 2020. Timing, magnitude and geochemistry of major Southeast Asian volcanic eruptions: identifying tephrochronologic markers. J. Quat. Sci. 35, 272–287. https://doi.org/10.1002/jqs.3181.

Brown, S.K., Crosweller, H.S., Sparks, R.S.J., Cottrell, E., Deligne, N.I., Guerrero, N.O., Hobbs, L., Kiyosugi, K., Loughlin, S.C., Siebert, L., Takarada, S., 2014. Characterisation of the quaternary eruption record: analysis of the large Magnitude Explosive Volcanic Eruptions (LaMEVE) database. J. Appl. Volcanol. 3, 1–22. https://doi.org/10.1186/2191-5040-3-5.

Brown, S.K., Sparks, R.S.J., Mee, K., Vye-Brown, C., Ilyinskaya, E., Jenkins, S.F., Loughlin, S.C., 2014b. Regional and country profiles of volcanic hazard and risk. Report IV of the GVM/IAVCEI contribution to the Global Assessment Report on Disaster Risk Reduction. In: United Nations Office for Disaster Risk Reduction 2015. http://www.preventionweb.net/english/hyogo/gar/2015/en/bgdocs/risk-section/GVMd.%20Global%20Volcanic%20Hazards%20and%20Risk%20Country%20volcanic%20hazard%20and%20risk%20profiles.pdf.

Chadwick Jr., W.W., Cashman, K.V., Embley, R.W., Matsumoto, H., Dziak, R.P., De Ronde, C.E.J., Lau, T.K., Deardorff, N.D., Merle, S.G., 2008. Direct video and hydrophone observations of submarine explosive eruptions at NW Rota-1 volcano, Mariana arc. J. Geophys. Res. Solid Earth 113. https://doi.org/10.1029/2007JB005215.

Cole, J.W., Nairn, I.A., 1975. Catalogue of the Active Volcanoes of the World Including Solfatara Fields. Part XXII New Zealand. International Association of Volcanology and Chemistry of the Earth's Interior, Rome.

Coles, S., Sparks, R.S.J., 2006. In: Mader, H.M., Coles, S.G., Connor, C.B., Connor, L.J. (Eds.), Extreme Value Methods for Modelling Historical Series of Large Volcanic Magnitudes, in Statistics in Volcanology. The Geological Society, London, pp. 47–56.

Colombijn, F., 2005. A moving history of Middle Sumatra, 1600-1870. Mod. Asian Stud. 39, 1–38. https://www.jstor.org/stable/3876505.

Crisp, J.A., 1984. Rates of magma emplacement and volcanic output. J. Volcanol. Geotherm. Res. 20, 177–211. https://doi.org/10.1016/0377-0273(84)90039-8.

Crosweller, H.S., Arora, B., Brown, S.K., Cottrell, E., Deligne, N.I., Guerrero, N.O., Hobbs, L., Kiyosugi, K., Loughlin, S.C., Lowndes, J., Nayembil, M., Siebert, L., Sparks, R.S.J., Takarada, S., Venzke, E., 2012. Global database on large magnitude explosive volcanic eruptions (LaMEVE). J. Appl. Volcanol. 1, 1–13. https://doi.org/10.1186/2191-5040-1-4.

De la Cruz-Reyna, S., 1991. Poisson-distributed patterns of explosive eruptive activity. Bull. Volcanol. 54, 57–67. https://doi.org/10.1007/BF00278206.

Deligne, N.I., Coles, S.G., Sparks, R.S.J., 2010. Recurrence rates of large explosive volcanic eruptions. J. Geophys. Res. Solid Earth 115, 1–16. https://doi.org/10.1029/2009JB006554.

Dietz, R.S., Sheeny, M.J., 1964. Transpacific detection of Myojin volcanic explosions by underwater sound. Bull. Geol. Soc. Am. 65, 941–956. https://doi.org/10.1130/0016-7606(1954)65[941:TDOMVE]2.0.CO;2.

Dunkley, P.N., Smith, M., Allen, D.J., Darling, W.G., 1993. The Geothermal Activity and Geology of the Northern Sector of the Kenya Rift Valley. Brithish Geol. Surv. Res. Rep. http://nora.nerc.ac.uk/id/eprint/507920.

Eichelberger, J., Gordeev, E., Izbekov, P., Kasahara, M., Lees, J., 2013. Volcanism and Subduction: The Kamchatka Region. John Wiley & Sons.

Ewing, M., Wolland, G.P., Vine, A.C., Worzel, J.L., 1946. Recent results in submarine geophysics. Bull. Geol. Soc. Am. 57, 909–934. https://doi.org/10.1130/0016-7606(1946)57[909:RRISG]2.0.CO;2.

Fontijn, K., Ernst, G.G.J., Elburg, M.A., Williamson, D., Abdallah, E., Kwelwa, S., Mbede, E., Jacobs, P., 2010. Holocene explosive eruptions in the Rungwe Volcanic Province, Tanzania. J. Volcanol. Geotherm. Res. 196, 91–110. https://doi.org/10.1016/j.jvolgeores.2010.07.021.

Furlan, C., 2010. Extreme value methods for modelling historical series of large volcanic magnitudes. Stat. Model. 10, 113–132. https://doi.org/10.1177/1471082X0801000201.

Garcia-Aristizabal, A., Marzocchi, W., Fujita, E., 2012. A Brownian model for recurrent volcanic eruptions: an application to Miyakejima volcano (Japan). Bull. Volcanol. 74, 545–558. https://doi.org/10.1007/s00445-011-0542-4.

Global Volcanism Program (GVP), 2013. In: Venzke, E. (Ed.), Volcanoes of the World, v. 4.7.5 (21 Dec 2018). Smithson. Inst. https://doi.org/10.5479/si.GVP.VOTW4-2013

Gorshkov, G.S., 1958. Catalogue of the Active Volcanoes of the World Including Solfatara Fields. Part VII-Kurile Islands. International Association of Volcanology and Chemistry of the Earth's Interior, Rome.

Gorshkov, G.S., 1970. History of investigation of the volcanoes: Investigations in the Kurile Island Arc. In: Volcanism and the Upper Mantle. Springer Science & Business Media, pp. 7–19.
Guttorp, P., Thompson, M. Lou, 1991. Estimating second-order parameters of volcanicity from historical data. J. Am. Stat. Assoc. 86, 578–583. https://doi.org/10.1080/01621459.1991.10475082.
Hammond, S.R., Embley, R.W., Baker, E.T., 2015. The NOAA vents program 1983 to 2013: Thirty years of ocean exploration and research. Oceanography 28, 160–173. https://doi.org/10.5670/oceanog.2015.17.
IAVCEI, 1973. Data Sheets of the Post-Miocene Volcanoes of the World with Index Maps. D'ANTIMI, Rome.
IVS, 2021. Institute of Volcanology and Seismology-far Eastern Branch of Russian Academy of Sciences. URL. http://www.kscnet.ru/ivs/eng/index.php.
Jenkins, S., Magill, C., McAneney, J., Blong, R., 2012. Regional ash fall hazard I: a probabilistic assessment methodology. Bull. Volcanol. 74, 1699–1712. https://doi.org/10.1007/s00445-012-0627-8.
Killick, R., Fearnhead, P., Eckley, I.A., 2012. Optimal detection of changepoints with a linear computational cost. J. Am. Stat. Assoc. 107 (500), 1590–1598.
Kiyosugi, K., Connor, C., Sparks, R.S.J., Crosweller, H.S., Brown, S.K., Siebert, L., Wang, T., Takarada, S., 2015. How many explosive eruptions are missing from the geologic record? Analysis of the quaternary record of large magnitude explosive eruptions in Japan. J. Appl. Volcanol. 4 https://doi.org/10.1186/s13617-015-0035-9.
Kombe, E.Y., Muguthu, J., 2018. Geothermal energy development in East Africa: barriers and strategies. J. Energy Res. Rev. 2, 1–6. https://doi.org/10.9734/jenrr/2019/v2i129722.
Mason, B.G., Pyle, D.M., Oppenheimer, C., 2004. The size and frequency of the largest explosive eruptions on Earth. Bull. Volcanol. 66, 735–748. https://doi.org/10.1007/s00445-004-0355-9.
Massey, F.J., 1951. The Kolmogorov-Smirnov test for goodness of fit. J. Am. Stat. Assoc. 46, 68–78. https://doi.org/10.1080/01621459.1951.10500769.
Mathworks, 2021. Signal Processing Toolbox ™ Reference (R2021a). Retrieved July 16, 2020, from. https://www.mathworks.com/help/pdf_doc/signal/signal_ref.pdf.
Mead, S., Magill, C., 2014. Determining change points in data completeness for the Holocene eruption record. Bull. Volcanol. 76 https://doi.org/10.1007/s00445-014-0874-y.
Mulargia, F., Gasperini, P., Tinti, S., 1987. Identifying different regimes in eruptive activity: an application to Etna volcano. J. Volcanol. Geotherm. Res. 34, 89–106. https://doi.org/10.1016/0377-0273(87)90095-3.
Neumann van Padang, M., 1951. Catalogue of the Active Volcanoes of the World Including Solfatara Fields. International volcanological association.
Newhall, C.G., Self, S., 1982. The volcanic explosivity index (VEI): an estimate of explosive magnitude for historical volcanism. J. Geophys. Res. 87, 123–1238. https://doi.org/10.1029/jc087ic02p01231.
Papale, P., 2018. Global time-size distribution of volcanic eruptions on Earth. Sci. Rep. 8, 1–11. https://doi.org/10.1038/s41598-018-25286-y.
Papale, P., Garg, D., Marzocchi, W., 2022. Global rates of subaerial volcanism on earth. Front. Earth Sci. 10, 1–11. https://doi.org/10.3389/feart.2022.922160.
Patrick, M.R., Smellie, J.L., Harris, A.J.L., Wright, R., Dean, K., Izbekov, P., Garbeil, H., Pilger, E., 2005. First recorded eruption of Mount Belinda volcano (Montagu Island), South Sandwich Islands. Bull. Volcanol. 67, 415–422. https://doi.org/10.1007/s00445-004-0382-6.
Ponomareva, V., Melekestsev, I., Braitseva, O., Churikova, T., Pevzner, M., Sulerzhitsky, L., 2007. Late pleistocene-holocene volcanism on the Kamchatka Peninsula, Northwest Pacific region. Geophys. Monogr. Ser. 172, 165–198. https://doi.org/10.1029/172GM15.
Rougier, J., Sparks, S.R., Cashman, K.V., 2016. Global recording rates for large eruptions. J. Appl. Volcanol. 5, 1–10. https://doi.org/10.1186/s13617-016-0051-4.
Rougier, J., Sparks, R.S.J., Cashman, K.V., 2018. Regional and global under-recording of large explosive eruptions in the last 1000 years. J. Appl. Volcanol. 7, 1–10. https://doi.org/10.1186/s13617-017-0070-9.
Rubin, K.H., Soule, S.A., Chadwick, W.W., Fornari, D.J., Clague, D.A., Embley, R.W., Baker, E.T., Perfit, M.R., Caress, D.W., Dziak, R.P., 2012. Volcanic eruptions in the deep sea. Oceanography 25, 143–157. https://doi.org/10.5670/oceanog.2012.12.
Seekins, D.M., 1993. Chapter 1. Historical Setting. In: Rolan, R.D. (Ed.), Philippines: a country study, 4th. Federal Research Division, Library of Congress, pp. 1–56.
Shaffer, J.P., 1995. Multiple hypothesis testing. Annu. Rev. Psychol. 46, 561–584. https://doi.org/10.1146/annurev.ps.46.020195.003021.
Siebert, L., Simkin, T., Kimberly, P., 2011. Volcanoes of the World, 3rd ed. Univ of California Press.
de Silva, S., Lindsay, J.M., 2015. Primary volcanic landforms. In: The Encyclopedia of Volcanoes, Second ed. Elsevier Inc. https://doi.org/10.1016/b978-0-12-385938-9.00015-8.
Simkin, T., Siebert, L., McClelland, L., Bridge, D., Newhall, C., Latter, J.H., 1981. Volcanoes of the World: A Regional Directory, Gazetteer, and Chronology of Volcanism during the Last 10,000 Years, 1st ed. Hutchinson Ross Publishing.
Wadge, G., 1982. Steady State Volcanism- Evidence from Eruption Histories of Polygenetic Volcanoes, 87, pp. 4035–4049. https://doi.org/10.1029/JB087iB05p04035.
Wang, T., Schofield, M., Bebbington, M., Kiyosugi, K., 2020. Bayesian modelling of marked point processes with incomplete records: volcanic eruptions. J. R. Stat. Soc.: Ser. C: Appl. Stat. 69, 109–130. https://doi.org/10.1111/rssc.12380.
White, S.M., Crisp, J.A., Spera, F.J., 2006. Long-term volumetric eruption rates and magma budgets. Geochem. Geophys. Geosyst. 7. https://doi.org/10.1029/2005GC001002.