**Seasonal modulation of Earth’s volcanic activity**

**Jamie I. Farquharson1, Falk Amelung1**

**1** Rosenstiel School of Marine and Atmospheric Science, 4600 Rickenbacker Causeway, University of Miami, Miami FL 33149.

**Key points:**

* Analysis of the historical eruption record shows that, globally, eruptions are more likely to occur in the boreal summer (July to October)
* This trend is particularly pronounced in regions of the globe associated with high and/ or highly variable rainfall
* We suggest that the hydrological cycle is an important modulator of global volcanic activity

**Introduction**

It has long been suspected that volcanic eruptions occur relatively more frequently at some times of the year than others: several published analyses of the global eruption record (e.g. Kluge [1862], Eggers and Dekker [1969], Hamilton [1973], Belov [1986]) have suggested that, regionally or globally, volcanic eruptions are not uniformly distributed in time. Stothers [1989], on the other hand, asserted that any observed month-to-month variability in eruption frequency was not statistically significant, attributing much of the apparent nonuniformity to seasonal reporting bias, which would serve to amplify the incompleteness of the eruption record and any potential physical correlations between many small eruptions. More recently, Mason et al. [2004] analyzed a larger dataset and, focusing only on small eruptions, concluded that in fact a statistically significant seasonality was evident, not only on a global scale but also in particular volcanic regions (such as the Pacific “Ring of Fire”) and at certain volcanoes: specifically Gunung Semeru (Indonesia) and Sakurajima (Japan).

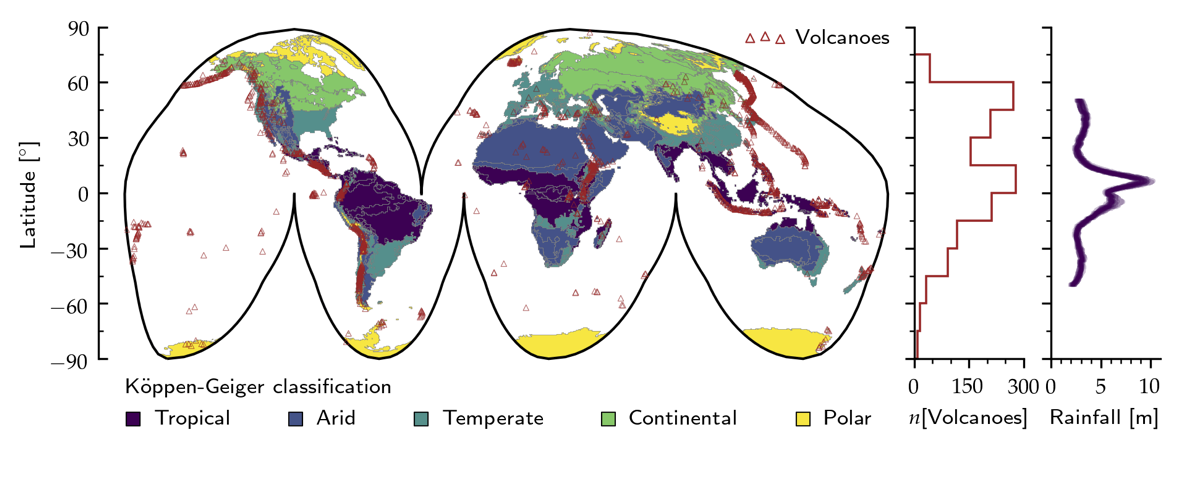
Historically, perceived nonuniformity in eruption time-series at individual volcanoes has been attributed to a range of external factors. Perret [1901], for example, posited that the timing of eruptions of Stromboli (Italy) followed trends dictated by the lunar phase. Eruptions [Jaggar et al. 1924; Stearns and Macdonald 1946] and inter-eruptive behaviour [Wood 1917; Jaggar 1920] at  Kīlauea and Mauna Loa volcanoes (Hawai`i, USA) have both been suggested to be governed by solar declination (i.e. the timing of the equinoxes and solstices). A related—albeit contradictory—hypothesis was proposed to explain seasonality observed at Capelinhos volcano (Faial, Açores, Portugal) by Mendoca Dias [1962]. Although some authors have suggested that nonuniform eruption distributions at certain volcanoes are a function of meltwater pulses which induce seasonal phreatic activity [e.g. Kluge 1862; Day and Allan 1925], the most common mechanism to which observed eruption seasonality is attributed is that of the diurnal and fortnightly body tides (i.e. gravity-induced displacement of the ocean and solid earth). This is true at the global scale (e.g. Mauk and Johnston [1973]; Hamilton [1973]), regionally (Golombek and Carr [1978]), and at individual volcanoes (e.g. Michael and Christoffel [1975]; Dzurizin [1980]; Martin and Rose [1981]; Ramos et al. [1985]; McNutt and Beavan [1987]). However, in a critical review of tidal triggering of volcanic events, Emter [1997] highlights that nearly all such studies suffer from statistical problems (either in terms of small datasets or being overly liberal with significance levels). In support of this, more recent spectral analysis of volcano-seismic time-series data by Neuberg [2000] yielded strong evidence against the hypothesis of the modulation of volcanic activity by either ocean or body tides.

The potential for external modulation of eruptive activity still exists, however. Neuberg [2000] proposes a number of potential factors, including icemelt, variations in barometric pressure, and thermal stress oscillations, that could account for nonuniform eruption distribution through time. A number of studies have also linked volcanism with localized weather patterns [e.g. Mastin et al. 1994; Yamasato et al. 1998; Voight et al. 2000; Matthews et al. 2002, 2009; Matthews and Barclay, 2004; Farquharson and Amelung, 2020]—a variable that is often seasonal by definition—or with seasonal changes in edifice saturation [Violette et al. 2001, Hammond et al. 2019].

In this contribution, we analyze both the global eruption record and satellite-derived SO2 data in order to identify whether statistically significant seasonal volcanism exists at a range of spatial scales. Moreover, we suggest mechanical mechanisms etc.

**Materials and Methods**

The eruption catalogue used throughout this study is based on the Global Volcanism Program (GVP) eruption and volcano databases from the Smithsonian Institution [GVP, 2013]. This GVP database succeeds those published by Simkin et al. [1981], Simkin and Siebert [1994], and Siebert et al. [2010], which have been used in previous analyses (e.g. Mason et al. [2004]). Using the GVP webservices interface, up-to-date volcano and volcanic eruption data (i.e. the “gazetteer” and “chronology”, respectively) are downloaded from the GVP website [<https://doi.org/10.5479/si.GVP.VOTW4-2013>]. These raw data are then subject to a multi-level filter in order to maximize confidence in the dataset and to avoid certain sampling artefacts (full details are given in **Supp. Mat. S1**). The resulting product is a manipulable database listing Holocene volcanoes, dates of their eruptions, and related information such as the “explosivity index” (VEI) attributed to the eruption. [Figure 1](https://docs.google.com/document/d/1mbRNzs_5_l8ij6aZpCJ3UfdOkBcSBBfAyFzN_JmW1ZM/edit#fig_1) shows the global distribution of subaerial Holocene-epoch volcanoes on Earth considered in this study.



Two primary statistical approaches were employed in this study: the chi-squared () test and an information entropy analysis. The former of these tests for uniformity of a dataset, whereas the latter assesses randomness. Reported eruptions from the filtered global dataset were binned according to the month of the eruption start date, and classified according to the reported VEI into “All” (a dataset containing all filtered eruptions, irrespective of explosivity), “VEI 2+” (a subset of “All” where only eruptions with a reported VEI of greater or equal to two are included), and “VEI 3+” (a further subset, including only those eruptions reported as VEI 3 or greater). Separately, data were also classified according to the Köppen-Geiger classification code that intersects with the reported volcano coordinates (see Figure 1). The Köppen-Geiger climate classification is one of the most widely used climate classification schemes [e.g. Cepeda et al. 2010; Rubel and Kottek, 2010; Kottek et al. 2006; Peel et al. Beck et al. 2018], dividing climates into five general groups with further subdivisions in each group according to temperature and precipitation metrics. The top-tier groups are tropical (megathermal), arid, temperate (mesothermal), continental (microthermal), and polar/ alpine. To date, the highest resolution available Köppen-Geiger classification maps are those developed by Beck et al. [2018]; we use the “present day” (1980—2016) classifications [<https://doi.org/10.6084/m9.figshare.6396959>], which are provided at three different spatial resolutions (0.0083°, 0.083°, and 0.05°) alongside corresponding data confidence maps. We note that the highest resolution maps available are not necessarily the most informative in our case. In particular, high-altitude regions are often classified solely based on temperature *T* without accounting for precipitation (“polar tundra” for example, is defined by 0 < *T* ≤10 °C: Beck et al. [2018]). This may be an issue where a volcano—defined by its apical coordinates—is at high elevation but influenced by predominantly tropical or temperate weather systems, for example. To minimize potential misclassification of volcanic systems, we impose a cascading classification based on the confidence level associated with the grid cell. We first impose a threshold of 75 % confidence: if a climate classification for any given coordinate pair is associated with confidence below this value at 0.0083° resolution, then a progressively coarser resolution is tested until the confidence threshold is achieved. If the confidence level is not achieved, the threshold is iteratively decreased (to 50, 25, and 0%), until a classification is derived.

The test is used to determine whether there exists a significant difference between the observed frequencies (*obj*) of a certain event (e.g. a volcanic eruption) given a number of discrete categories (e.g. months of the year) and the expected frequency (*exj*) of that event.  The test statistic is of the form:

(1)

Where the degrees of freedom .  The cumulative distribution function of the chi-squared statistic cdf() = (*s*,*z*)/(*s*), where and , and:

(2)

where *A priori*, we define the expected frequencies as a product of the number of total eruptions *n* and the probability of any given day of the year falling within a particular month. This probability is ℘ ≅ 0.085, 0.082, or 0.077, depending on the number of days in each month (31, 30, or 28.25 respectively) relative to the number of days in a year (taken as 365.25).  We ascribe statistical significance in this case when the cumulative distribution function cdf() lies below 0.05 (an arbitrary but conventional standard for significance testing). For , this threshold corresponds to a critical value of of 19.68.

In order to calculate information entropy, we consider eruption occurrence as a set of nonnegative independent observations , where *n* is the total number of observations (eruptions). In this case, we can further constrain the range of observation values *xj* as *xj* [1, 2, ..., 12] with the values corresponding to the months of the year (January, February, etc.). Given the probability ℘ of an event *xj* we can determine the information entropy H of the dataset:

(3)

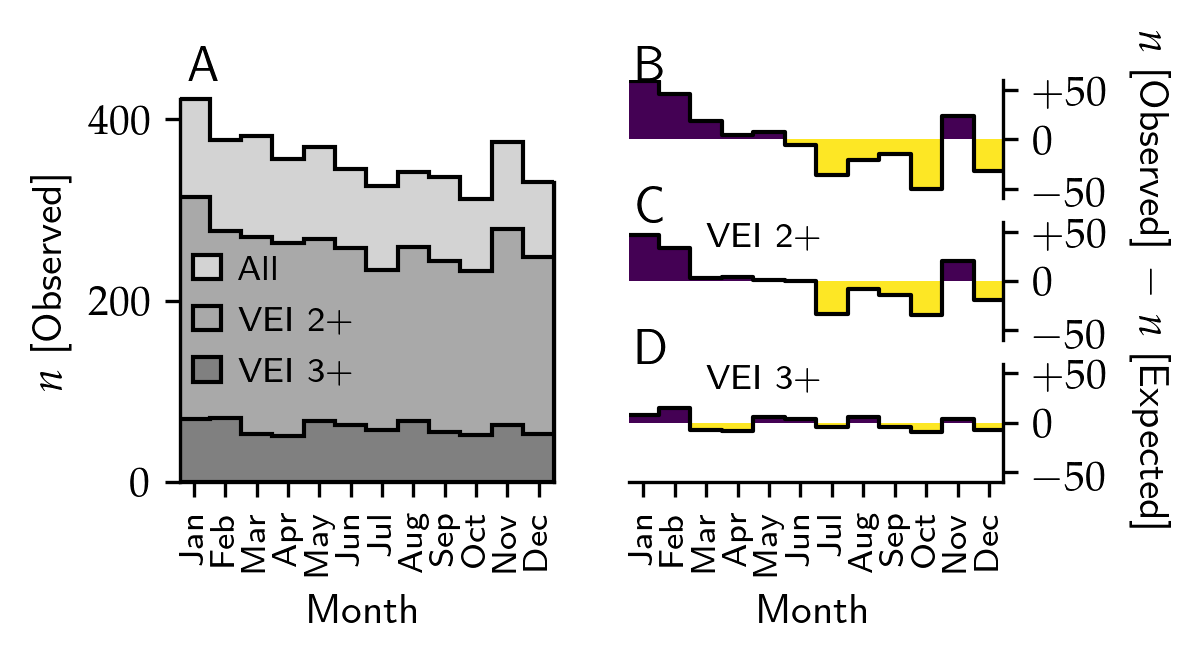
where we take *b* to equal 2 (i.e. the Shannon entropy).

To supplement our analysis of the eruption record, we also analyze satellite-derived SO2­ data. In volcanic environments the flux of SO2 provides a useful proxy for magmatic activity [e.g. Mori and Burton, 2006; Oppenheimer et al. 2017; Carn et al. 2016; Gouhier and Paris, 2019], and affords us detail that might otherwise be missed if discrete volcanic eruptions were to be regarded as the only surface manifestations of volcanic activity. As the satellite SO2 record is more or less continuous over the last decade, we can also perform time-series analyses on these data that are impossible to do with the relatively more sporadic and inherently biased eruption record [see Siebert et al. 2010].

We use NASA’s OMSO2G gridded product from the Ozone Mapping Instrument (OMI), and extract the SO­­­­2 column amounts for the lower tropospheric profile. A time series of SO2 is extracted between 201X and 2020, using the daily mean of a 0.5º × 0.5º grid centered on the volcano of interest. A Fast Fourier Transform (FFT) is performed on these data, allowing us to examine dominant cyclicity [e.g. Flower et al. 2016].

**Results**

After filtering the global eruption database (Supp. Mat. 1), our dataset comprises a total of 4275 eruptions from 407 volcanoes since the year 1583 C.E. (i.e. following the adoption of the Gregorian calendar) up until March 2020 C.E. Of these, 3147 are defined as VEI 2 or greater (VEI 2 is generally used as a default category for explosive eruptions if additional information is missing: Simkin and Siebert [1994]), and 718 are category VEI 3 and higher. Fig. 2A shows the aggregated per-month distribution exhibited by each category.  Fig. 2B illustrates the difference between the observed values and a statistically uniform theoretical dataset, where *n*[Expected] is the total number of observations multiplied by the probability for any given month (a function of the number of days in that month).



Analyzing the per-month distribution of eruptions in the whole dataset as a function of the Köppen-Geiger climate classification we observe that two top-tier classifications—tropical and temperate—exhibit statistically non-uniform distributions (cdf() = 0.0405 and 0.0039, respectively; see [Table 1](https://docs.google.com/document/d/1mbRNzs_5_l8ij6aZpCJ3UfdOkBcSBBfAyFzN_JmW1ZM/edit#tab_01)). If we further assess the data as a function of lower-tier classifications, some of the subgroups also show distributions that are statistically non-uniform; however, we cannot discount the influence of small sample sizes in some of these distributions, as this could result in spurious determinations of significance.

[Table 1](https://docs.google.com/document/d/1mbRNzs_5_l8ij6aZpCJ3UfdOkBcSBBfAyFzN_JmW1ZM/edit#table_1): Results of χ2 analysis for different subsets of the filtered dataset. Tropical, arid, temperate, continental, and polar/ alpine refer to the top-tier Köppen-Geigen classifications A—E.  *n*[Observed] refers to the total number of volcanic eruptions in each subgroup. Results are determined to be significant if cdf(χ2) is less than 0.05 (equivalently, if χ2 is greater than the critical value 19.68).

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Dataset | *n*[Volcanoes] | *n*[Observed] | χ2 | cdf(χ2) | Significant [Y|N] |
| All eruptions | 407 | 4275 | 34.400 | 0.0003 | Y |
| VEI 2+ | 373 | 3147 | 25.493 | 0.0077 | Y |
| VEI 3+ | 232 | 718 | 12.168 | 0.3512 | N |
| Tropical | 110 | 1452 | 19.622 | 0.0508 | Y |
| Arid | 12 | 13 | 11.894 | 0.3717 | N |
| Temperate | 73 | 1148 | 27.389 | 0.0040 | Y |
| Continental | 49 | 330 | 8.208 | 0.6945 | N |
| Polar/ alpine | 47 | 384 | 2.958 | 0.9913 | N |

Use cascade approach on tropical\_volanoess etc. -- > somehow worse?

113 additional “undefined”

Trop:

94, 929, (11.677015718035934, 0.3884039358253447), 3.5762978591198307

Arid:

Trop: 94 929 (11.677015718035934, 0.3884039358253447) 3.5762978591198307 N

Arid: 25 88 (13.064696169517036, 0.289117947226617) 3.4860602997219186 N

Tem: 89 1207 (24.072893609241557, 0.012429364185905345) 3.5723868108323953 Y

Cont: 44 496 (12.817582509922374, 0.305415205326811) 3.5689758593236456

Pol: 146 1504 (26.522405486596593, 0.005422470691043291) 3.573452478877644

For individual systems, chi-squared analysis indicates that 23 volcanoes— almost 6 % of the volcanoes that have erupted since 1583 C.E.—exhibit significant nonuniformity in terms of eruption distribution throughout the year. These 23 volcanoes account for a disproportionate number of the reported eruption record over that timeframe (~10 %). These data are shown in [Figure](https://docs.google.com/document/d/1mbRNzs_5_l8ij6aZpCJ3UfdOkBcSBBfAyFzN_JmW1ZM/edit#fig_5)XXA, highlighting the values that lie below our imposed threshold for statistical significance (cdf() = 0.05 and = 19.68).

[Figure](https://docs.google.com/document/d/1mbRNzs_5_l8ij6aZpCJ3UfdOkBcSBBfAyFzN_JmW1ZM/edit#fig_5) XXB and XXC illustrate the results of the information entropy analysis. The potential range of entropy values (i.e. the number of microstate probabilities 𝜛) increases as the number of observations increases, until the number of observations reaches or exceeds the number of categories, after which point the range of entropy values converges towards a known maximum (in this case, around 3.58).  To demonstrate this, [Figure](https://docs.google.com/document/d/1mbRNzs_5_l8ij6aZpCJ3UfdOkBcSBBfAyFzN_JmW1ZM/edit#fig_5) XXB shows pseudorandom entropy data generated using the same underlying probabilities as nominally random natural data.  The color scale indicates the relative microstate probability of each *n*—H datum. In [Figure](https://docs.google.com/document/d/1mbRNzs_5_l8ij6aZpCJ3UfdOkBcSBBfAyFzN_JmW1ZM/edit#fig_5) XXC, we plot the determined entropy values for all volcanoes in our dataset, where *n* in this case is the true observed number of eruptions. For the real eruption data, values of H range from 0 at *n* = 2, up to a value of 3.58 at *n* = 100. The highlighted data (cdf() <0.05) generally lie at the lower end of the range of H for any given *n*.

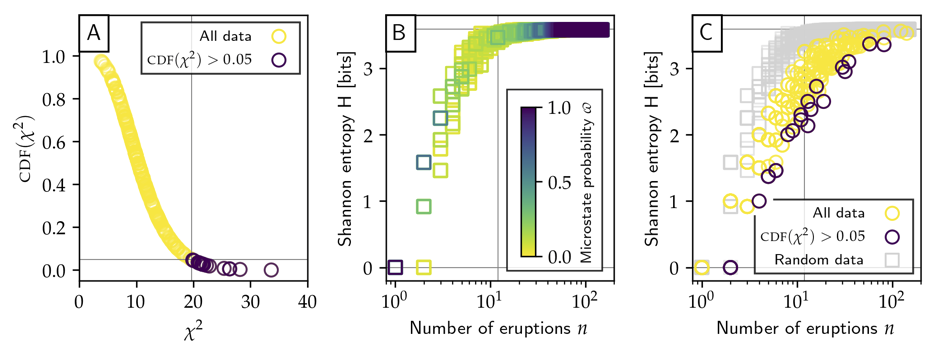
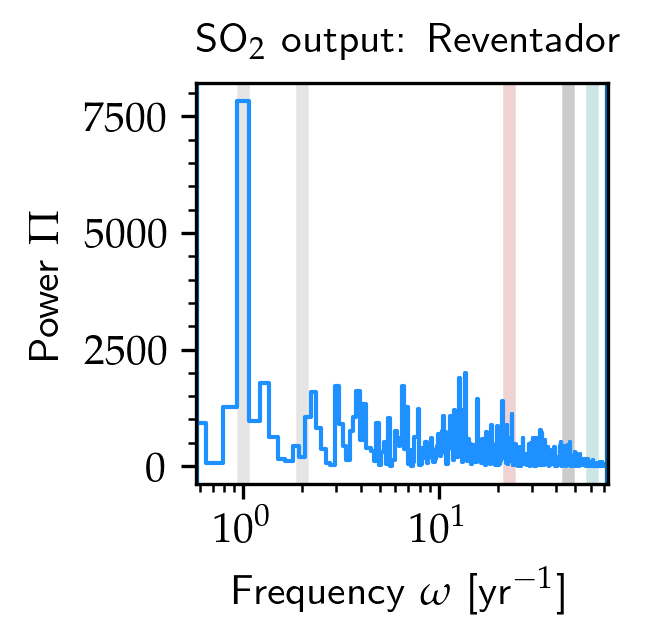


Figure XX shows the power spectrum of Fourier-transformed SO2 time-series signals for selected volcanoes. Dashed vertical lines highlight frequencies associated with the OMI instrument and its orbit [see Flower and Carn, 2015; Flower et al. 2016]. These include a 16 day satellite orbit repeat cycle, a 6 day cyclical divergence of viewing angle from nadir, and short-period peaks associated with the OMI Row Anomaly. The limits of the *x­*-axis ( and ) have been defined such that and , where is the length of one year in days, is the size of the averaging window, and is the total length of the time-series. Periodicities outside of the range [] cannot be robustly identified. Solid vertical lines indicate periodicities of 1 and 2 yr-1­; i.e. annual and biennial signals.



<SO2 data>

SO2 data fig. – Semeru, Sakurajima

**Discussion**

A preliminary assessment of the distribution of raw GVP data highlights that spuriously high numbers of eruptions are reported on the 1st, 15th, and 16th of any given month (for the first of the month, the reported eruption count is as much as a factor of 5 greater than for any other day), and that January and July exhibit much greater eruption counts than other months (see **Supp. Mat. S1**). Moreover, this skew in distribution is predominantly observed for eruptions recorded prior to 1950. We estimate that these patterns are artefacts arising due to reporting bias. Our initial data processing filters eruptions based primarily on the confidence associated with the reported measurement (i.e. whether the eruption was observed, and the uncertainty attributed with the start date: see **Supp. Mat. S1** and Simkin and Siebert [1994] for further details). This approach predominantly strips dates reported in January and July, on the first of the month  before 1950, and the middle of the month post-1950. Filtered data approach a more uniform distribution at the monthly and daily level, although the daily data remain somewhat skewed up until 1950. We suggest that these apparent artefacts may account to some extent for the conclusions of several studies that correlated peaks in eruptive activity with lunar phenomena and fortnightly solid earth tides (e.g. Perret [1901]; Mauk and Johnston [1973]; Hamilton [1973]; Michael and Christoffel [1975]; Golombek [1978]; Dzurizin [1980]; Martin and Rose [1981]; Ramos et al. [1985]; McNutt and Beavan [1987]).

Artefacts aside, we still observe a distinctly non-uniform distribution of global eruptions over time (see [Figure](https://docs.google.com/document/d/1mbRNzs_5_l8ij6aZpCJ3UfdOkBcSBBfAyFzN_JmW1ZM/edit#fig_3) XXA). This is true whether we consider all reported eruptions or eruptions of VEI 2 and greater to a high degree of significance: as shown in [Figure](https://docs.google.com/document/d/1mbRNzs_5_l8ij6aZpCJ3UfdOkBcSBBfAyFzN_JmW1ZM/edit#fig_4) XXB, we tend to see markedly more eruptions in  January and February than would be expected if they were randomly distributed in time. The opposite is true for eruptions reported during the boreal summer (July to October). Chi-squared analysis confirms that when we examine the entire dataset or only those eruptions defined as VEI 2 and higher, eruptions are not uniformly distributed throughout the year. In detail, the  values of cdf() are 0.0003 and 0.0077, respectively, forcing us to reject the null hypothesis that eruptions follow a uniform distribution in time. For VEI 3 and greater however, we do not observe this same trend and the test indicates that the distribution is not significantly different to a uniform distribution: as such we accept the null hypothesis that these larger eruptions are uniformly distributed in time. Ultimately,  this suggests that while some seasonal phenomena may be able to modulate the frequency of relatively small volcanic eruptions, larger eruptions are driven by processes less sensitive to seasonal effects, explaining the lack of periodicity determined by previous studies focused solely on large eruptions [e.g. Palladino and Sottili 2012].

In addition to the global dataset, distribution statistics show that significant eruption periodicity is a feature both regionally and locally (i.e. at individual volcanoes). As highlighted in [Figure](https://docs.google.com/document/d/1mbRNzs_5_l8ij6aZpCJ3UfdOkBcSBBfAyFzN_JmW1ZM/edit#fig_5) XXA, at least 24 volcanoes exhibit significantly non-uniform behavior.  We highlight however that there is clearly no categorical difference between volcanoes that fall slightly above or slightly below the cdf() threshold: rather, the data exist on a continuum between completely uniform (or random) and completely non-uniform (or ordered).

3479 recorded volcanic eruptions fall within regions encompassed by the terrestrial Köppen-Geiger climate classification: approximately 78% of the whole dataset. Of these, 2728 (>78%) occurred within tropical and temperate regions. Month-by-month, volcanoes situated in Köppen-Geiger climates defined as “tropical” and “temperate” exhibit significant nonuniform distribution (CDF(2)> 0.05). Tropical zones are defined by megathermal climates, and encompass more volcanic regions in our dataset than any other climate classification (*n*[Observed] = 1481: see Table 1). While the tropics are generally referred to as aseasonal, this designation refers to the lack of significant monthly differences in daylight hours and mean daily temperature. Seasonality in terms of precipitation and general water availability does exist however, for example due to monsoon systems. Temperate zones, defined by mesothermal climates, are the second most populous in terms of volcanoes in our dataset (*n*[Observed] = 1247). Moreover, if we analyse the dataset based on precipitation subcategories, each of the temperate sub-classifications with precipitation code *f* (seasonal, with significant precipitation in all seasons) exhibit statistical nonuniformity according to our pre-established threshold. Nonuniformity is not detected for the three other top-tier climate classifications: arid (*n*[Observed] = 14), continental (*n*[Observed] = 341), and polar/ alpine (*n*[Observed] = 396). Arid and polar environments are characterized by consistently low precipitation, while continental climates tend to exhibit only  moderate amounts of precipitation (although rainfall seasonality may exist). In short, at the top-tier climate level, we only observe significant nonuniformity in climate zones defined by high and variable rainfall.

<given the coorrelattion between eruptive activity and rainfall at different environments, this could be dignificant: MSH, Soufriere, Perret, Kilauea>

Individual volcanoes tend not to exhibit significant nonuniformity: a function of the relatively low number of observations at any given volcano. However, despite generally small sample sizes, chi-squared analysis indicates that 23 volcanoes—6.0 % of the volcanoes that have erupted since 0 C.E.—exhibit significant nonuniformity in terms of eruption distribution throughout the year. These 24 volcanoes account for a disproportionate number of the reported eruption record (9.4 %). These data are shown in [Figure](https://docs.google.com/document/d/1mbRNzs_5_l8ij6aZpCJ3UfdOkBcSBBfAyFzN_JmW1ZM/edit#fig_5) XXA, highlighting the values that lie below our imposed threshold for statistical significance (CDF(2) =0.05 and 2 = 19.68).

The entropy analysis results are shown in [Figure](https://docs.google.com/document/d/1mbRNzs_5_l8ij6aZpCJ3UfdOkBcSBBfAyFzN_JmW1ZM/edit#fig_5) XXB and XXC. Essentially, the fewer the number of observations there are relative to the number of possible outcomes (or categories: in this case, months of the year), the fewer ways there are to codify this information and therefore the lower the entropy of the dataset. If there is a single observation, then there is only a single way in which this distribution can be described (one month with one eruption, 11 months with zero eruptions) and entropy is therefore zero. If there are two recorded observations, there are multiple ways in which they may be distributed (two eruptions in one month, two eruptions in evenly spaced months, or two eruptions in non-evenly spaced months). As a result, the entropy may be higher. As such, for each value of n the possible values of H have their own associated probabilities based on the number of possible distributions, which we refer to here as microstate probabilities 𝜛. If the number of observations is large enough, there is no entropy distinction between random events and events uniformly distributed in time. This accounts for the distribution of the simulated data ([Figure](https://docs.google.com/document/d/1mbRNzs_5_l8ij6aZpCJ3UfdOkBcSBBfAyFzN_JmW1ZM/edit#fig_5) XXB). For the real data in [Figure](https://docs.google.com/document/d/1mbRNzs_5_l8ij6aZpCJ3UfdOkBcSBBfAyFzN_JmW1ZM/edit#fig_5) XXC, we note that almost all data (with each point representing a different volcano) exhibit values of H lower than the truly random data. Some key conclusions can be drawn from this: firstly, volcanic eruption distributions in time tend to be nonrandom, although they generally tend towards high entropy values (i.e. unpredictability) if the dataset is large enough. The lower the value of H for a given n relative to the simulated data, the more “predictable” the data are. Critically, this does not mean that the volcanic eruptions themselves are predictable. Again, low number statistics are at play: if there are fewer observations than categories (i.e. n< 12 in this case), it is impossible to tell if the data would tend towards randomness (or uniformity) with the addition of new observations. However, those volcanoes exhibiting relatively low H (≲ 3.5), high n ( 12), high 2 ( 19.68), and low CDF(2) ( 0.05), would appear to be good candidates for a statistical assessment of their eruption frequency over time.  Nevertheless, much of this apparent nonrandom behaviour is presumably a reflection of the fact that eruption observations at any given volcano are *not* truly independent: the eruption of a volcano is to some extent dependent on its previous eruptions.  Table 2 highlights volcanoes that meet these criteria. Notably, they are all located in tropical or temperate climate zones.

**3500 = 7 p.u.**

**+4 figs = 11 p.u.**

**+1 table? Or an extra 500 words.**

Conclusions

An analysis of the historical eruption record shows that the frequency of volcanic eruptions globally is not homogeneously distributed throughout the year. Rather, eruptions are more likely to occur in the boreal summer (July to October), a trend which is particularly pronounced in temperate and tropical regions of the globe (i.e. those associated with high and/ or highly variable rainfall). Despite relatively small numbers, nonuniformity can also be discerned for several individual volcanoes, as well as in aggregated regional and global datasets.

We propose that precipitation may be responsible for the observed eruption seasonality in many volcanic systems, both based on the prevalence of individual cases which are affected by rainfall and by the coincidence of seasonal signals in volcanoes found in tropical and temperate climate regions.

There are a series of interrelated mechanisms that can link rainfall and eruption mechanics…

In particular, Okataina (New Zealand), Gunung Semeru (Indonesia), Volcán de Fuego (Guatemala) and Mt St Helens (USA) demonstrate significantly non-uniform eruption distribution in time (characterised by low chi-squared test statistics and low Shannon entropy), as well as exhibiting significantly more eruptions during the rainy season than expected, making them prime targets for further study and the incorporation of meteorological data into their monitoring programs.

Moreover, several other volcanic systems exhibit one or other of these traits (nonuniform eruption distribution or correlation with the rainy season) to a significant degree, including Mt Vesuvius (Italy), Piton de la Fournaise (Réunion), and Kīlauea (USA), warranting further study.

Mid-res

Vesuvius Csa Temperate, dry summer, hot summer

Okataina Cfb Temperate, no dry season, warm summer

Raoul Island Cfa Temperate, no dry season, hot summer

Gaua Af Tropical, rainforest

Kaba Af Tropical, rainforest

Semeru Af Tropical, rainforest

Batur Am Tropical, monsoon

Awu Af Tropical, rainforest

Babuyan Claro Af Tropical, rainforest

Asosan Dfa Cold, no dry season, hot summer

Izu-Torishima 0

Koryaksky ET Polar, tundra

Changbaishan Dwc Cold, dry winter, cold summer

Pavlof ET Polar, tundra

St. Helens Dsb Cold, dry summer, warm summer

Fuego Cwb Temperate, dry winter, warm summer

Santa Ana Aw Tropical, savannah

San Miguel Aw Tropical, savannah

Pilas, Las Aw Tropical, savannah

Irazu Cfb Temperate, no dry season, warm summer

Turrialba Af Tropical, rainforest

Cumbal ET Polar, tundra

Planchon-Peteroa ET Polar, tundra

**——**

**Hi-res**

Vesuvius Csb Temperate, dry summer, warm summer

Okataina Cfb Temperate, no dry season, warm summer

Raoul Island Cfa Temperate, no dry season, hot summer

Gaua Af Tropical, rainforest

Kaba Cfb Temperate, no dry season, warm summer

Semeru ET Polar, tundra

Batur Am Tropical, monsoon

Awu Af Tropical, rainforest

Babuyan Claro Af Tropical, rainforest

Asosan Dfb Cold, no dry season, warm summer

Izu-Torishima 0

Koryaksky EF Polar, frost

Changbaishan Dwc Cold, dry winter, cold summer

Pavlof ET Polar, tundra

St. Helens ET Polar, tundra

Fuego ET Polar, tundra

Santa Ana Cwb Temperate, dry winter, warm summer

San Miguel Cwb Temperate, dry winter, warm summer

Pilas, Las Aw Tropical, savannah

Irazu ET Polar, tundra

Turrialba ET Polar, tundra

Cumbal ET Polar, tundra

Planchon-Peteroa ET Polar, tundra

Lo-res

Vesuvius Csa Temperate, dry summer, hot summer

Okataina Cfb Temperate, no dry season, warm summer

Raoul Island 0

Gaua 0

Kaba Af Tropical, rainforest

Semeru Am Tropical, monsoon

Batur Am Tropical, monsoon

Awu Af Tropical, rainforest

Babuyan Claro 0

Asosan Cfa Temperate, no dry season, hot summer

Izu-Torishima 0

Koryaksky Dfc Cold, no dry season, cold summer

Changbaishan Dwc Cold, dry winter, cold summer

Pavlof Dfc Cold, no dry season, cold summer

St. Helens Dsb Cold, dry summer, warm summer

Fuego Aw Tropical, savannah

Santa Ana Aw Tropical, savannah

San Miguel Aw Tropical, savannah

Pilas, Las Aw Tropical, savannah

Irazu Af Tropical, rainforest

Turrialba Af Tropical, rainforest

Cumbal ET Polar, tundra

Planchon-Peteroa ET Polar, tundra

Lo-res +12

Vesuvius Csa Temperate, dry summer, hot summer

Okataina Cfb Temperate, no dry season, warm summer

Semeru Am Tropical, monsoon

Batur Am Tropical, monsoon

Awu Af Tropical, rainforest

Asosan Cfa Temperate, no dry season, hot summer

Pavlof Dfc Cold, no dry season, cold summer

St. Helens Dsb Cold, dry summer, warm summer

Fuego Aw Tropical, savannah

San Miguel Aw Tropical, savannah

Irazu Af Tropical, rainforest

Planchon-Peteroa ET Polar, tundra

Mid-res +12

Vesuvius Csa Temperate, dry summer, hot summer

Okataina Cfb Temperate, no dry season, warm summer

Semeru Af Tropical, rainforest

Batur Am Tropical, monsoon

Awu Af Tropical, rainforest

Asosan Dfa Cold, no dry season, hot summer

Pavlof ET Polar, tundra

St. Helens Dsb Cold, dry summer, warm summer

Fuego Cwb Temperate, dry winter, warm summer

San Miguel Aw Tropical, savannah

Irazu Cfb Temperate, no dry season, warm summer

Planchon-Peteroa ET Polar, tundra

Hi-res +12

Vesuvius Csb Temperate, dry summer, warm summer

Okataina Cfb Temperate, no dry season, warm summer

Semeru ET Polar, tundra

Batur Am Tropical, monsoon

Awu Af Tropical, rainforest

Asosan Dfb Cold, no dry season, warm summer

Pavlof ET Polar, tundra

St. Helens ET Polar, tundra

Fuego ET Polar, tundra

San Miguel Cwb Temperate, dry winter, warm summer

Irazu ET Polar, tundra

Planchon-Peteroa ET Polar, tundra

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Country | Volcano | *n*[Observed] | χ2 | CDF(χ2) | H (bits) | Climate |
| New Zealand | Okataina | 16 | 19.887 | 0.0469 | 2.727 | Temperate |
| Indonesia | Semeru | 58 | 21.179 | 0.0316 | 3.365 | Tropical |
| Batur | 19 | 26.372 | 0.0057 | 2.500 | Tropical |
| Awu | 13 | 26.332 | 0.0058 | 2.134 | Tropical |
| Japan | Asosan | 130 | 31.609 | 0.0009 | 2.322 | Temperate |
| USA | Pavlof | 30 | 20.821 | 0.0353 | 3.015 | Temperate |
| Mt St Helens | 13 | 20.864 | 0.0348 | 2.500 | Temperate |
| Guatemala | Fuego | 41 | 25.041 | 0.0090 | 3.195 | Tropical |
| Pacaya | 22 | 20.053 | 0.0446 | 2.859 | Tropical |
| El Salvador | San Miguel | 32 | 33.648 | 0.0004 | 2.949 | Tropical |
| Costa Rica | Irazú | 14 | 21.787 | 0.0261 | 2.379 | Temperate |

*Climate classification data*

The Köppen-Geiger climate classification is one of the most widely used climate classification schemes [e.g. Cepeda et al. 2010; Kottek et al. Peel et al. Beck et al. 2018]. It divides climates into five general groups, with subdivisions in each group further refining the classification according to temperature and precipitation metrics. The top-tier groups are tropical (megathermal), arid, temperate (mesothermal), continental (microthermal), and polar/ alpine. Altogether there are 31 sub-classifications [e.g. Kottek et al. 2006; Rubel and Kottek, 2010]. To date, the highest resolution available Köppen-Geiger classification maps are those developed by Beck et al. [2018]; we use the “present day” (1980—2016) classifications [<https://doi.org/10.6084/m9.figshare.6396959>], which are provided at three different spatial resolutions (0.0083°, 0.083°, and 0.05°) alongside data corresponding confidence maps. We note that the highest resolution maps available are not necessarily the most informative in our case. In particular, high-altitude regions are often classified solely based on temperature *T* without accounting for precipitation (“polar tundra” for example, is defined by 0 < T ≤10 °C: Beck et al. [2018]). This may be an issue where a volcano—defined by its apical coordinates—is at high elevation but influenced by predominantly tropical or temperate weather systems, for example. To minimize potential misclassification of volcanic systems, we impose a cascading classification based on the confidence level associated with the grid cell. We impose a threshold of 70 % confidence: if a climate classification for any given coordinate pair is associated with confidence below this value at 0.0083° resolution, then a progressively coarser resolution is tested until the confidence threshold is achieved. Table [XX] shows volcanoes

Over the time period from 1583 to 2019—436 years—we observe 4304 eruptions from 409 volcanoes in our filtered data: a mean eruptive rate of 0.02 yr-1 (or one eruption of each volcano every 41.4 years). While this is a simple average that does not reflect the observed spectrum of eruption dynamics, magnitudes, and types—nor does it account for known reporting biases—it is nevertheless a useful metric against which to crudely compare “sensitive” volcanic systems. The 23 “sensitive” volcanoes are responsible for 422 eruptions in our dataset (0.04 yr-1 or an eruption from each volcano every 23.8 years) and the subset of those eruptions with >12 eruptions on record are responsible for 368 eruptions (0.07 yr-1, or an eruption from each volcano every 14.2 years). This suggests that 1. we are statistically more likely to observe non-uniformity when a volcano’s eruption frequency is high; 2. that volcanoes modulated by seasonal processes are more likely to erupt at a high frequency, or 3. a combination of both of these statistical and phenomenological effects.

**While the 0.0083 degree resolution is nominally better, we note that it tends to classify mountainous volcanic regions as “alpine”, even if they are influenced by predominantly tropical or temperate weather systems.**

**We note that the highest resolution maps available are not necessarily the most informative in our case. In particular, high-altitude regions are often classified solely based on temperature *T* (“polar tundra” for example, is defined by 0 < *T* ≤10 °C: Beck et al. [2018]). This may be an issue where a volcano is at high elevation but influenced by predominantly tropical or temperate weather systems, for example.**

**as polar tundra**

(1980—2016) at a resolution of 0.083°, giving a gridsize of approximately 10 10 km at the equator [<https://doi.org/10.6084/m9.figshare.6396959>].

explicitly corrected for topographic effects, which influences air temperature[26](https://www.nature.com/articles/sdata2018214#ref-CR26) and precipitation[27](https://www.nature.com/articles/sdata2018214#ref-CR27) in mountainous regions.

0.0083°

xv three resolutions (0.0083°, 0.083°, and 0.5°; approximately 1 km, 10 km, and 50 km at the equator,

The classifications are upscaled from 0.0083° to 0.083° and 0.5° using majority resampling and the confidence levels using bilinear averaging.

Polar tundra defined solely in terms of temperature (specifically temperatures 0 < T ≤10 °C [Beck et al. 2018]

We impose a cascading resolution threshold

Confidence threshold

Cascading classification system based on the associated

Characterised by low confidence? This can be checked directly. If confidence lower than threshold, we use lower resolution.

Critical volcanoes responsible for disproportionoately more eruptions than average

Over the time period from 1583 to 2019—436 years—we observe 4304 eruptions from 409 volcanoes in our filtered data: a mean eruptive rate of 0.02 yr-1 (or one eruption of each volcano every 41.4 years). While this is a simple average that does not reflect the observed spectrum of eruption dynamics, magnitudes, and types—nor does it account for known reporting biases—it is nevertheless a useful metric against which to crudely compare “sensitive” volcanic systems. The 23 “sensitive” volcanoes are responsible for 422 eruptions in our dataset (0.04 yr-1 or an eruption from each volcano every 23.8 years) and the subset of those eruptions with >12 eruptions on record are responsible for 368 eruptions (0.07 yr-1, or an eruption from each volcano every 14.2 years). This suggests that 1. we are statistically more likely to observe non-uniformity when a volcano’s eruption frequency is high; 2. that volcanoes modulated by seasonal processes are more likely to erupt at a high frequency, or 3. a combination of both of these statistical and phenomenological effects.

Non-uniform behavior is more likely to be observed at volcanoes with high eruption rates or that

High

0.04208216992421221

0.07033639143730887

41.432156133829

23.76303317535545

14.217391304347826

For the entire dataset, we see 4304 eruptions from 409 volcanoes between 1583 and 2019 (i.e. ~11 eruptions per volcano on average), compared to 422 eruptions from the 23 ‘critical’ volcanoes (~18 eruptions per volcano) and 368 eruptions from the subset of critical volcanoes with at least 12 recorded eruptions (~31 eruptions

4304/409, 368/12, 422/23

10.52322738386308,

30.666666666666668,

18.347826086956523

|  |  |
| --- | --- |
| Polar | Not (B) & *Thot*≤10 |
| T |  | - Tundra | *Thot*>0 |

| **1st** | **2nd** | **3rd** | **Description** | **Criterion a** |
| --- | --- | --- | --- | --- |
| A |  |  | Tropical | Not (B) & *Tcold*≥18 |
| f |  | - Rainforest | *Pdry*≥60 |
| m |  | - Monsoon | Not (Af) & *Pdry*≥100-*MAP*/25 |
| w |  | - Savannah | Not (Af) & *Pdry*<100-*MAP*/25 |
| B |  |  | Arid | *MAP*<10×*Pthreshold* |
| W |  | - Desert | *MAP*<5×*Pthreshold* |
| S |  | - Steppe | *MAP*≥5×*Pthreshold* |
|  | h | - Hot | *MAT≥*18 |
|  | k | - Cold | *MAT<*18 |
| C |  |  | Temperate | Not (B) & *Thot*>10 & 0<*Tcold*<18 |
| s |  | - Dry summer | *Psdry<40* & *Psdry<Pwwet/3* |
| w |  | - Dry winter | *P* *wdry* *<P* *swet* */10* |
| f |  | - Without dry season | Not (Cs) or (Cw) |
|  | a | - Hot summer | *Thot*≥22 |
|  | b | - Warm summer | Not (a) & *Tmon10*≥4 |
|  | c | - Cold summer | Not (a or b) & 1≤*Tmon10*<4 |
| D |  |  | Cold | Not (B) & *Thot*>10 & *Tcold≤*0 |
| s |  | - Dry summer | *Psdry<*40 & *Psdry<Pwwet*/3 |
| w |  | - Dry winter | *Pwdry<Pswet*/10 |
| f |  | - Without dry season | Not (Ds) or (Dw) |
|  | a | - Hot summer | *Thot*≥22 |
|  | b | - Warm summer | Not (a) & *Tmon10*≥4 |
|  | c | - Cold summer | Not (a, b, or d) |
|  | d | - Very cold winter | Not (a or b) & *Tcold*<-38 |
| E |  |  | Polar | Not (B) & *Thot*≤10 |
| T |  | - Tundra | *Thot*>0 |
| F |  | - Frost | *Thot*≤0 |

**Hi-def conf:**

Vesuvius Csb Temperate, dry summer, warm summer 67

Okataina Cfb Temperate, no dry season, warm summer 100

Semeru ET Polar, tundra 67

Batur Am Tropical, monsoon 75

Awu Af Tropical, rainforest 100

Asosan Dfb Cold, no dry season, warm summer 67

Pavlof ET Polar, tundra 67

St. Helens ET Polar, tundra 67

Fuego ET Polar, tundra 67

San Miguel Cwb Temperate, dry winter, warm summer 67

Irazu ET Polar, tundra 67

Planchon-Peteroa ET Polar, tundra 100

**Med-def:**

Vesuvius Csa Temperate, dry summer, hot summer 96

Okataina Cfb Temperate, no dry season, warm summer 100

Semeru Af Tropical, rainforest 64

Batur Am Tropical, monsoon 65

Awu Af Tropical, rainforest 100

Asosan Dfa Cold, no dry season, hot summer 61

Pavlof ET Polar, tundra 96

St. Helens Dsb Cold, dry summer, warm summer 44

Fuego Cwb Temperate, dry winter, warm summer 83

San Miguel Aw Tropical, savannah 100

Irazu Cfb Temperate, no dry season, warm summer 88

Planchon-Peteroa ET Polar, tundra 92

**Low-res:**

Vesuvius Csa Temperate, dry summer, hot summer 96

Okataina Cfb Temperate, no dry season, warm summer 100

Semeru Am Tropical, monsoon 73

Batur Am Tropical, monsoon 76

Awu Af Tropical, rainforest 100

Asosan Cfa Temperate, no dry season, hot summer 80

Pavlof Dfc Cold, no dry season, cold summer 92

St. Helens Dsb Cold, dry summer, warm summer 83

Fuego Aw Tropical, savannah 87

San Miguel Aw Tropical, savannah 100

Irazu Af Tropical, rainforest 78

Planchon-Peteroa ET Polar, tundra 87