GSA Data Repository 2017028

Regional variability in the frequency and magnitude of large explosive volcanic eruptions

Tom Sheldrake[1], Luca Caricchi[1]

[1]Department of Earth Sciences, University of Geneva, rue des Maraîchers 13, 1205 Geneva, Switzerland

SUPPLEMENTARY MATERIAL

Below is a summary of the assumptions, method and results for the (a) homogeneity test and (b) change-point analysis; and detailed summary of the statistical methods used for (c) the statistical model to estimate the frequency-Magnitude relationship, and (d) calculation of the global recurrence rate of large-magnitude eruptions.

(a) Chi-square test for homogeneity based on a contingency table

A chi-square test for homogeneity is calculated using for the sum of M4 events and sum of M5-M7 events, across different values of t_{unique}. The rationale for reducing the homogeneity test to two populations is that is reduces the degrees of freedom by half so the test statistic is more informative. The number of M4 events is chosen as a population by itself as it represents the largest variation in the record. When $X^2 < df$ the hypothesis of homogeneous populations for different values of t_{unique} is accepted.

Results:

tstart	X^2	df	pvalue
50 ka	1060.2	500	2.2e-16

11.7 ka	36.991	468	1

(b) Change-point analysis of Magnitude 4 events in the Holocene dataset

A change point for the number of magnitude 4 events is calculated using the segmented package in R[1], which uses a dummy variable to identify a change point in a linear regression by maximum likelihood fitting.

[1] (https://cran.r-project.org/web/packages/segmented/segmented.pdf)

(c) Hierarchical Bayesian Analysis

To statistically characterise the frequency-Magnitude (f-M) relationship for volcanic eruptions we use the methodology set out in Sheldrake (2014), which is based on analysing the proportion of different events. A hierarchical Bayesian approach is employed, which quantifies the common distribution of eruptions for a group of analogous volcanoes, whilst recognising that each volcano has a unique record. Each of the individual eruptive records is considered exchangeable, and so each volcano is assumed to be able to produce an eruption between M4 and M7. In terms of magmatic processes, this assumption is akin to saying that there is a common process determining the frequency of eruptions of various magnitudes globally, but at an individual volcano this common process may only manifest in a particular sub-set of the state space.

The statistical model has three hierarchies:

(a) Eruptive records, or data, which represent the likelihood of each eruption magnitude (*j*) at each volcano (*i*), and is characterised as a multivariate dataset (i.e. mutually exclusive events):

$$x_{i,j} \sim Multi(\theta_{i,j}, n_j).$$

(b) Prior distribution, which characterises the common processes associated with the accumulation and eruption of eruptible magma that are responsible for the recurrence rate of volcanic eruptions. The prior is modelled using a Dirichlet distribution, as we characterise eruption magnitude as a continuous multivariate dataset where the probability of the different magnitudes (θ_j) adds to unity at each volcano. The Dirichlet distribution is parameterised by a series of alpha parameters (α_j), which is advantageous as it does not put any restrictions on the shape of the distribution, allowing different behaviours to be identified for different groups of volcanoes:

$$\theta_{i,j} \sim Dir(\alpha_j).$$

(c) Hyperprior distributions, which allow the prior distribution to be uninformative, and thus only determined by only the data in the model and not by subjective judgement. The hyperprior distributions (φ_j, ψ) are rearranged in terms of the alpha parameters of the Dirichlet distribution (α_j):

$$\alpha_j = \frac{\varphi(\exp(\phi_j)}{J-1+\exp(\phi_j)},$$

where J is the total number of eruption scenarios (i.e. number of eruption magnitudes = 4). Each hyperprior is chosen so that the before observing the data each magnitude is equally likely.

The fist hyperprior characterises the variability in the data between each of the volcanoes, and so is a distribution on α_0, which is the sum of all the α_j parameters:

$$\psi = \alpha_0 = \sum \alpha_j,$$

In the case where the model is uninformative each a_j=1 and so the minimum value of a_0 is the sum of these parameters (in the case here this is the number of Magnitude states = 4), and where the data is fully informative the value of a_0 is equal to the total number of eruptions or observations in the analysis (K, which in the case here is

1,766). Hence, the first hyperprior is parameterised as a uniform distribution between these two values:

$$\psi \sim Unif(4, 1766).$$

The second hyperprior is a distribution of the logit of the mean probability for each magnitude multiplied by (J – 1):

$$\phi_j = \log\left(\frac{(K-1)\alpha_j}{\alpha_0-\alpha_j}\right).$$

In the case where each magnitude is equally likely this will equal zero, and so the second hyperprior is parameterised as a diffuse distribution centred on zero:

$$\phi_j \sim Normal(0,1000).$$

There are two outputs of the statistical model, the prior and posterior distributions. The prior distribution is calculated based on a combination of the total number of events for each magnitude, the variation in the proportions of each magnitude at individual volcanoes, and the total number of events observed at each volcano. Once the prior distribution is calculated, the posterior distribution that is unique to each individual volcano can be calculated. To characterise the behaviour of a group of volcanoes, or to compare the behaviour of different volcanoes, we characterise a group of posterior probabilities for different eruption magnitudes (m) using a power-law distribution:

$$\Pr(M = m) \sim \gamma^m.$$

For Magnitude 4 -7, this becomes with the appropriate normalisation to unit mass:

$$\Pr(M = m) = \frac{\gamma^{m-4}}{1+\gamma+\gamma^2+\gamma^3} = \frac{(1-\gamma)\gamma^{m-4}}{1-\gamma^4}\ .$$

To fit the power law we use a non-hierarchical version of the Bayesian method in Bebbington (2014), with the reference distribution for γ is a diffuse log normal distribution:

$$log(\gamma) \sim N(0, 10^3).$$

To perform the statistical analysis we use a Markov Chain Monte Carlo analysis using RStan[1].

[1]Carpenter, B., et al., 2016, Stan: A probabilistic programming language: Journal of Statistical Software (in press).

(d) Calculation of the recurrence rate of large-magnitude eruptions during the Holocene

To estimate the global recurrence rate of eruptions of different magnitude, we fit a power-law using the assumptions stated in the main text. We solve for the under-recording parameter λ by using the value of γ from the analysis of the global record and rearranging the following equations:

(1) the proportion of eruptions that are Magnitude 4:

$$\theta_4 = \frac{(1-\lambda\gamma)(\lambda\gamma)^0}{1-(\lambda\gamma)^4};$$

(2) the proportion of eruptions that are Magnitude 7:

$$\theta_7 = \frac{(1-\lambda\gamma)(\lambda\gamma)^3}{1-(\lambda\gamma)^4};$$

(3) the expected number of Magnitude 4 events in the Holocene, where $X_{4:1961}$ is the number of Magnitude 4 events observed globally at arc volcanoes between 1961 – 2000 (based on 95% confidence that a change point in under-recording occurred after this date; Furlan, 2010) and normalised to the duration of the Holocene:

$$N_4 = X_{4:1961} \cdot \frac{11{,}700}{2000-1961};$$

(4) the expected number of Magnitude 7 events in the Holocene:

$$N_7 = \frac{N_4}{\theta_4/\theta_7};$$

(5) the level of completeness for Magnitude 7 events in the Holocene, which is estimated to be 70% (Brown et al., 2014):

$$\frac{N_7}{X_7} = 0.7.$$

2017028_Data Summary and Results.zip