RESEARCH　　　　　　　　　　　　　　　　　　　　　　　　　　　**Open Access**

Regional and global under-recording of large explosive eruptions in the last 1000 years

Jonathan Rougier[1]* 🄳, R. Stephen J. Sparks[2] and Katharine V. Cashman[2]

Abstract

Recording probabilities for large-magnitude ($M \geq 4$) explosive eruptions are assessed regionally over the last 1000 years, using the LaMEVE database. Although the uncertainty is large, due to the scarcity of large eruptions, it does not swamp differences in recording probabilities across times and regions. Broadly, the results reflect the pattern of European colonial expansion. Iceland presents an interesting anomaly, with a declining recording probability—going back in time—conflicting with its long history of written records. However, this may be explained by the loss of records in the 17th and 18th centuries. Globally, we find that records of roughly 40% of large-magnitude explosive eruptions are missing. There is a marked difference in modern recording probabilities pre- and post-1980, which we attribute to changes in the way that the magnitude of large eruptions is assessed.

Keywords: Recording probability, Large explosive eruptions, Poisson process, LaMEVE

Introduction

This is a follow-up paper to Rougier et al. (2016), to which readers should refer for a wider review of the issue of under-recording. In that paper, we showed that the global recording probability for large explosive eruptions of stratovolcanoes falls rapidly from 100% going back in time, based on the observations in the LaMEVE database (Crosweller et al. 2012; Brown et al. 2014). 'Large' was defined as having a recorded magnitude $M \geq 4.0$, according to the scale of Pyle (2000) and Mason et al. (2004); we denote this as 'M4+' below. The global recording probability was below 50% prior to 1600 CE, and below 20% prior to 1100 CE (with at least 95% confidence); see Fig. 5 below.

Rougier et al. (2016) used a Frequentist approach, partly for simplicity, and partly to avoid the additional structure that would be introduced by a prior distribution. In this paper we estimate recording probabilities for the last millenium, region by region. Our regions are those defined in the Volcano Reference File of the Smithsonian Institution. To a large extent each region is reasonably coherent in terms of tectonic setting, although some involve more than one volcano-tectonic province. We extend our analyis from stratovolcanoes to all volcanoes capable of

producing explosive eruptions. As well as being useful for further analysis at the individual volcano level, differences in the recording probabilities between regions ought to confirm or possibly challenge the narrative evidence about the development of the recording process (Siebert et al. 2010, pp. 31–34).

The cost of this more detailed assessment is mainly in statistical and computational complexity. The statistical model is straightforward, but its implementation within a Bayesian approach requires sophisticated Markov Chain Monte Carlo (MCMC) methods (described in the Appendix). As well as assessing the recording probability as a function of time for each region, we aggregate to produce a gobal recording probability, which largely confirms the findings of Rougier et al. (2016). However, as might be anticipated, we find that recording probabilities vary widely from one region to another, and the uncertainty about recording probabilities is typically high, going back even one hundred years.

Methods

For a specified region and a specified time-interval (taken to be from the start of 1000 to the start of 2015, for every region), recorded eruptions are modelled as a Poisson process with time-varying rate $\lambda \, r(t)$, with units of yr^{-1}, where both the scalar λ and the function r are uncertain. The scalar λ is the rate of $M \geq 4.0$ eruptions, treated as

*Correspondence: j.c.rougier@bristol.ac.uk
[1]School of Mathematics, University of Bristol, University Walk, Bristol BS8 1TW, UK
Full list of author information is available at the end of the article

effectively constant over the last 1000 years. The function r represents the probability that a large eruption at time t is recorded in the LaMEVE database. We impose the conditions that

  (i) The recording probability function r is increasing in time, and
  (ii) it has been 100% since some specified time in the recent past, which we take to be 1980 for all regions (see below, Results).

Rougier et al. (2016) used a simple approach in which the time-period was divided into 10 intervals of 100 years each, from 1050 to 1950, plus two futher intervals of 1950–1979 and 1980–2014, this final interval having a recording probability of 1. The recording function r was then modelled as an average value for each interval. In this paper we take advantage of a Bayesian approach to propose a more flexible representation of r, which is expressed semi-parametrically, as a piecewise linear interpolation of a collection of $m + 3$ knots (discrete points, shown as dots in Figs. 1 and 2), illustrated in Fig. 1 and specified in (4) in the Appendix. Briefly, the recording probability is equal to u at time-point a, the start of the time-interval, and then increases linearly in steps of $(1 - u)/(m + 1)$ at knot locations $v_1, \ldots, v_m$, where $a < v_1 < \cdots < v_m < b_0$, before attaining the value 1 at the time-point b_0, and remaining at 1 until time-point b, at the end of the time-interval. The values (a, b_0, b) are specified, and $(u, v_1, \ldots, v_m)$ are treated as uncertain.

The number of internal knots is set to $m = 11$, to give a set of recording probability functions rich enough to have several different segments, including more than

one segment in which the recording probability increases rapidly. Figure 2 shows how this approach with $m = 11$ approximates the type of recording probability function that might occur in practice.

The Bayesian approach requires a prior distribution for $(\lambda, u, v_1, \ldots, v_m)$. We choose a vague prior distribution in which the three elements are independent (this prior being the same for each region). u and v, which have bounded support, are given uniform marginal distributions (see the Appendix for more details). λ, which has unbounded support ($\lambda > 0$) is given a Gamma distribution with shape parameter 0.5 and rate parameter 20 yr. There being 19 regions, this prior for each region's λ implies a 95% equi-tailed prior credible interval for the global λ of 0.22 to 0.82 yr^{-1}, consistent with our beliefs about the global rate of large explosive volcanism. Equation 6 shows that these prior values are rapidly dominated by the observations. Expectations with respect to the posterior distribution are computed by Markov Chain Monte Carlo (MCMC). Full details of the prior distribution, the model, and the MCMC sampler are given in the Appendix.

Results

The observations are from the LaMEVE database, version 3.1 (Crosweller et al. 2012; Brown et al. 2014), downloaded October 2015. All eruptions in this database are treated as explosive, and the lower bound on large eruptions is $M \geq 4.0$, according to the scale of Pyle (2000) and Mason et al. (2004), or at least 10^8 tonnes of erupted mass. The calculations were performed in the R statistical computing environment (R Core Team 2016). All of the code and the dataset are available from the first author.
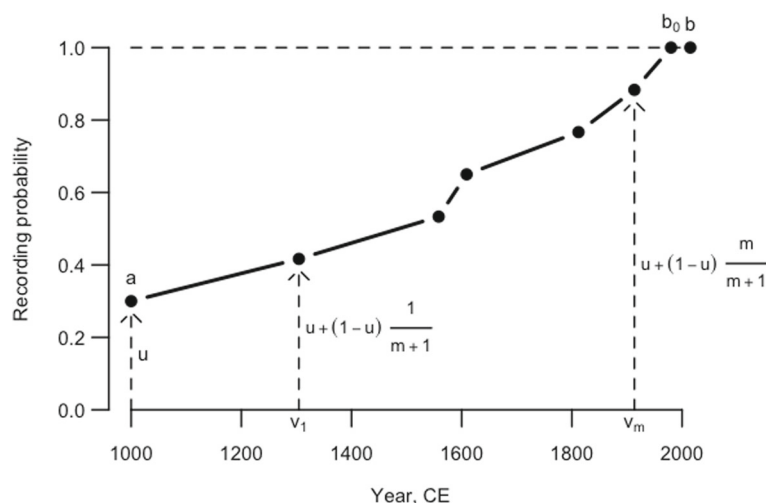

Fig. 1 Schematic of the model for the recording probability $r(t)$, expressed as a deterministic function of (u, v), where $v = (v_1, \ldots, v_m)$. The internal knot locations $v_1, \ldots, v_m$ are treated as uncertain, and augmented with additional knots at known locations $a = 1000$, $b_0 = 1980$, and $b = 2015$. See the "Results" section for our choice of b_0, and (4) in the Appendix for more details

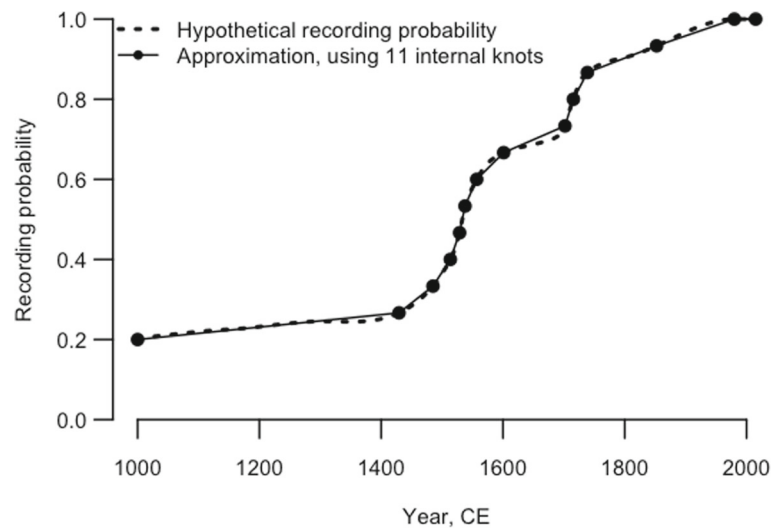Rougier *et al. Journal of Applied Volcanology* (2018) 7:1

Page 3 of 10

Fig. 2 Example of approximating a recording probability function using linear interpolation with $m = 11$ internal knots. The dashed curve is a hypothetical recording probability function. The value $m = 11$ is chosen as being sufficiently large to capture the types of shapes that might appear in the recording probability function across the different regions
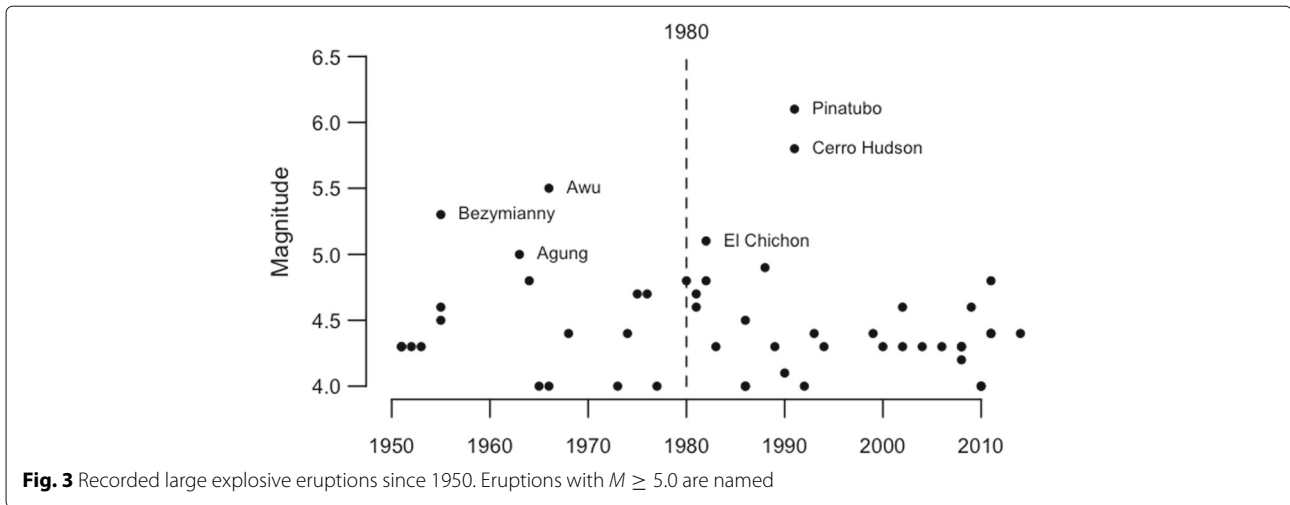
Rougier et al. (2016) used only large explosive eruptions from stratovolcanoes, whereas in this study we use large explosive eruptions from all types of volcano. Although the total number of eruptions in this study is larger (281 versus 188), the globally-aggregated recording probabilities through time are similar to those in Rougier et al. (2016), as shown in Figs. 5 and 6, below.

Choice of time intervals. The semi-parametric model for the recording probability over the time interval (a, b) contains a specified time $b_0 \leq b$ at which the recording probability is effectively one. We take $a = 1000$, $b_0 = 1980$, and $b = 2015$. In Rougier et al. (2016), the value $b_0 = 1980$ was based on a kink in the cumulative number of global large explosive eruptions. Figure 3 shows a marked thinning of M4–M5 eruptions in the period 1950–1980 when compared to 1980–2015, and more magnitude 4.0 eruptions since 1980.

One explanation for the difference between 1950–1980 and more recently is changes in the way that the magnitude of large eruptions is assessed. For eruptions which have occurred since 1980, researchers have increasingly made volume assessments soon after a large eruption, before the distal deposits disappear due to erosion. For earlier eruptions, volume assessments have been mostly based on proximal tephra deposits, usually in conjunction with a simple exponential thinning law (Pyle 2016). As discussed in Bonadonna et al. (1998), this pre-1980 practice can underestimate the volume of eruptions which produce a large amount of ash which is deposited at large distances from the source.

We believe that this underestimation will particularly affect eruptions in the magnitude range M4–M5. The deposits from this smaller end of the spectrum of large-magnitude eruptions will have the thinnest distal deposits which are most easily and quickly eroded making the bias strongest for pre-1980 deposits. Additionally, 1980 coincides with the growth in availability of data from earth-observation satellites; for example, Landsat 1 was launched in 1972. Finally, 1980 is also the year of the Mount St Helens eruptions, which set a new standard for rapid mapping of distal tephra deposits, and for tracking of ash-cloud behaviour using both ground-based radar and satellites.

Recording probabilities by region. The posterior distributions of the recording probabilities are visualised in Fig. 4, for those regions with at least ten large explosive eruptions. The diversity of results across the regions shows that much variation is concealed in a global analysis. A detailed explanation of this Figure is given at the end of the Appendix. Briefly, the grey polygons show pointwise uncertainty (at a specific point in time), while the solid line shows the recording probability function with highest posterior probability density (the *maximum a posteriori* or 'MAP' estimate). Where these two diverge, this indicates that the joint distribution is extremely non-Gaussian. In this case the MAP estimate, while still a helpful point estimate of the recording probability function, should be treated with extra caution, because our usual intuitions concerning uncertainty tend not to accommodate this type of behaviour.

Rougier *et al. Journal of Applied Volcanology*   (2018) 7:1

Page 4 of 10

Fig. 3 Recorded large explosive eruptions since 1950. Eruptions with $M \geq 5.0$ are named

Most of the panels in Fig. 4 show substantial posterior uncertainty in recording probabilities back to 1000 CE, which is a consequence of the low rate of large eruptions at the regional level. But the overall message is clear: in 'Mexico and Central America', and 'Iceland and Arctic', there is a small probability that the recording probability has remained close to 1 back to 1000 CE, but in all of the other regions it has almost certainly fallen substantially. This small probability is an 'empirical' probability, which reflects social, cultural, and technical changes only insofar as they are manifested in the record. The substantial uncertainty suggests that point estimates, including the MAP estimate shown in Fig. 4, are only tentative.

Nevertheless, at a qualitative level these point estimates do a good job of interpreting the recorded eruptions (shown as dots on the horizontal axis), consistently with the narrative evidence. For example, the rise in the recording probability in 'Indonesia' coincides with the establishment and rapid growth of the Dutch East India Company, which was chartered in 1602. At the risk of over-interpreting, we suggest that the more gradual increases in 'Mexico and Central America' and 'South America' reflect not only the marked improvement in recording in the 16th and 17th centuries related to Spanish and Portuguese colonisation, but also the larger amount of modern research in these regions into older eruptions.

'Iceland and Arctic' is an interesting case, where we expected a fairly complete record extending back to the 13th century, reflecting Viking colonisation and a rich written history, notably in monasteries in medieval times. This is not reflected in Fig. 4. However, as explained to us by A. Höskuldsson (pers. comm.), many of the Icelandic written records up to the mid-17th century are missing for a variety of reasons, of which the two

most important are the shipwreck of Hannes Þorleifsson in the late 17th century, and the 1728 fire in Copenhagen. This is a salutory reminder that for an observation to appear in a modern database, the link between the observer and the present day must not be broken.

Global recording probability. Figure 5 reproduces the global recording rate computed in Rougier et al. (2016), which used only stratovolcanoes in the LaMEVE database, and Fig. 6 shows the same algorithm applied to all volcanoes in the LaMEVE database. This shows that the results are similar in the two cases. The analysis in this paper uses all volcanoes.

Figures 5 and 6 represent a top-down approach. In this paper our approach is more 'bottom up', estimating recording probabilities and eruption rates separately for each region. Treating the regions as independent, these can be aggregated into a single global recording probability using the formula given in Rougier et al. (2016, sec. 2)

$$r(t) = \frac{\lambda_1 \, r_1(t) + \cdots + \lambda_k \, r_k(t)}{\lambda_1 + \cdots + \lambda_k}, \tag{1}$$

where there are k regions in total, and (λ_i, r_i) are the large-eruption rate and recording probability of the ith region. Expectations of this function can be derived by combining the MCMC samples for each of the regions, and the result is given in Fig. 7. The 95% credible interval for the recording probability at 1000CE is $(15\%, 34\%)$, with a posterior median of 24%.

In gross terms, the message in Fig. 7 is broadly similar to the Frequentist analysis in Fig. 6: under-recording is a very serious problem in the LaMEVE database, going back 1000 years. The lefthand end of the recording function
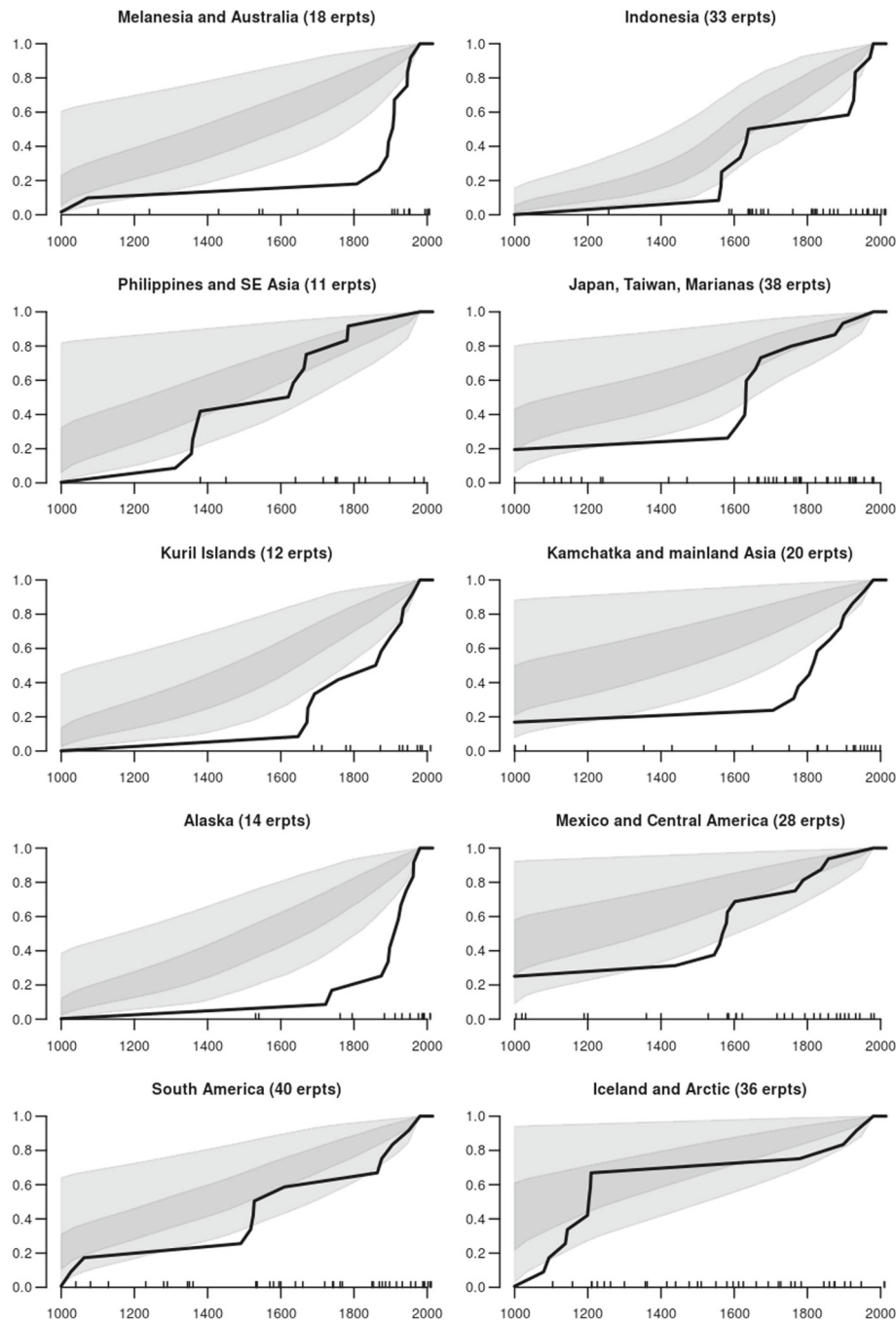
Fig. 4 Recording probabilities by region. The two polygons show pointwise 95 and 50% credible regions, and the solid line shows the *maximum a posteriori* estimate (see the Appendix for details of how this is calculated). The times of recorded large eruptions are shown as ticks on the horizontal axis. Only regions with at least 10 recorded large eruptions are shown

is displaced upwards in the Bayesian regional assessment, compared to the Frequentist global assessment. But there are some differences in the definitions of the two uncertainties. The error bars in Fig. 6 are conservative 95% confidence intervals, so that the coverage of these intervals is likely to exceed 95% in most of the parameter space.

By contrast, the vertical slices through the polygons in Fig. 7 are exact 95% credible intervals.

The two statistical models are also different. The Frequentist analysis uses a parsimonious model which assumes 'Poisson-ness' at the global scale but not below. The Bayesian analysis is less parsimonious (many more

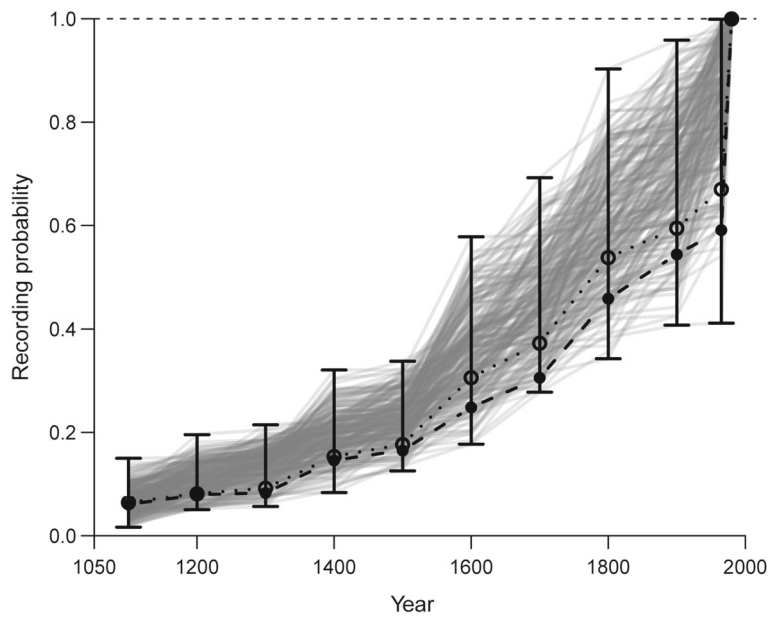Rougier *et al. Journal of Applied Volcanology*   (2018) 7:1

Page 6 of 10

Fig. 5 Reproduced from Rougier et al. (2016), where it was Fig. 4 (p. 7). This Figure shows the global recording probability using only stratovolcanoes in the LaMEVE database. The error bars represent conservative 95% confidence intervals, and the dashed and dotted lines represent alternative point estimates, the Maximum Likelihood estimate and the centroid of the 50% confidence set. Compare with Fig. 6, which uses all volcanoes in the LaMEVE database

parameters), under the stronger modelling assumption of Poisson-ness at the regional scale. More parameters and stronger modelling assumptions will typically result in less uncertainty, but there is no guarantee, under different models, that the uncertainties will nest. This is

particularly the case in extremely non-Gaussian models such as this one; see the Appendix for futher discussion. Broadly speaking, it is extremely difficult to extract the global or regional recording probability functions from 1000 years of data on large explosive eruptions, and this
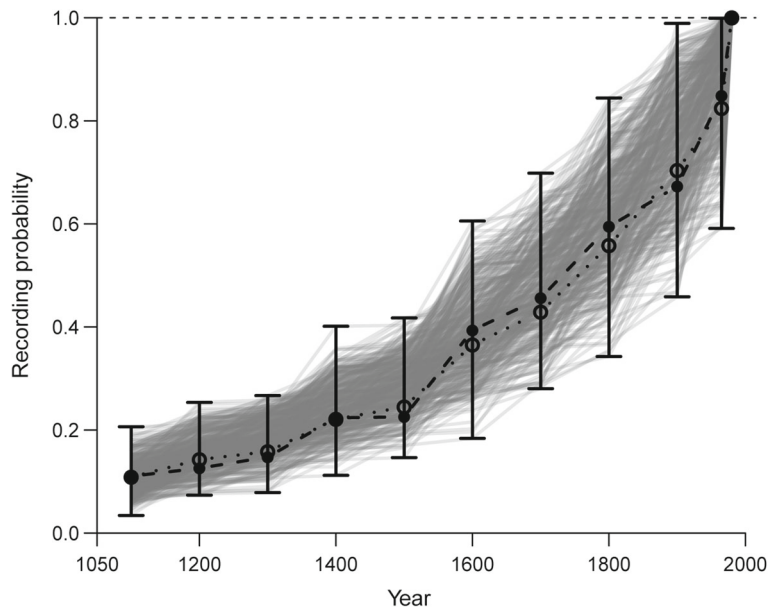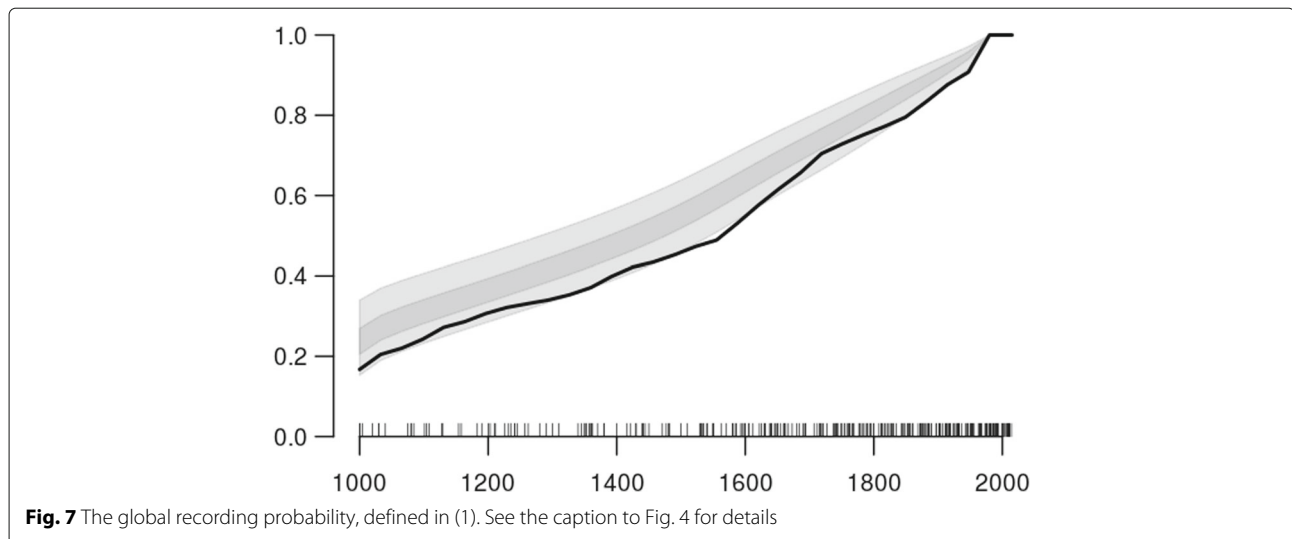

Fig. 6 Similar to Fig. 5, but using all volcanoes in the LaMEVE database. Compare with Fig. 7, which is the result of aggregating a Bayesian treatment at the regional level

Rougier *et al. Journal of Applied Volcanology* (2018) 7:1

Page 7 of 10

Fig. 7 The global recording probability, defined in (1). See the caption to Fig. 4 for details

difficulty extends to quantifying our uncertainty about the functions we extract.

It is helpful to have a single value to summarize the amount of under-recording in the last 1000 years; we use

$$\frac{\text{average rate of recorded large explosive eruptions, 1000–2015}}{\text{average rate were the recording probability to be 1 throughout}},$$

or, in the terms of the Appendix, $R(u, v)/(b - a)$. This is itself a random quantity, depending on (u, v), but we can estimate its distribution from the MCMC samples; it is 59% $\pm$ 6% (mean $\pm$ 2sd). That is to say, we estimate that roughly 40% of the large (M4+) explosive eruptions which happened over the last 1000 years are missing from the LaMEVE database.

Discussion

This paper confirms and extends the global analysis of under-recording in the LaMEVE database over the last millennium, presented in Rougier et al. (2016). Its primary difference is to work 'upwards' from a separate analysis for each region, thus allowing the regions to differ in their recording probabilities and their rates of large explosive eruptions. To support this, a second difference is to use a Bayesian rather than a Frequentist approach, which allows a more flexible representation of the recording probability as a function of time, although at the cost of a more complicated inferential calculation.

The outcome is to confirm the original finding that recording probabilities in the LaMEVE database are surprisingly low for 1000 years ago (best guess about 25%, very probably less than 40%), and that they are very probably less than 100% even in the recent past, say the last

100 years (see Fig. 7). A new summary value arising from this analysis is that about 40% of all large explosive eruptions are missing from the LaMEVE database, going back 1000 years. Clearly, we should be extremely cautious when making risk assessments for large explosive eruptions based on historical records.

The new contribution of this paper is to show that the recording probabilities vary widely from one region to another, but they are largely consistent with the narrative evidence that associates increases in recording probabilities with European colonial expansion, with Iceland being an interesting but resolvable anomaly. The thick lines in Fig. 4 serve as point estimates for regional recording probabilities through time, but these must be used with caution given the large uncertainties and extremely non-Gaussian distributions.

Our quantification and our point estimates are only as defensible as the statistical model we have used to derive them. We assume that the rate of large explosive eruptions in a region is effectively constant over the last 1000 years, and that the occurrence of such eruptions in each region follows a Poisson process (this need not be true for individual volcanoes). This is a very common assumption (see Rougier et al. 2016, for a detailed discussion), and justified mainly in terms of the general scarcity of large explosive eruptions, as a consequence of which we do not have sufficient historical observations to justify fitting a more complex model. An innovation is to use a semi-parametric model for the recording probability as a function of time, imposing only that (i) this function is increasing in time, and (ii) it has been 100% since 1980. This model is capable of adapting to a wide range of shapes (see Fig. 2), some of which are displayed in the regional estimates shown in Fig. 4. For aggregation to the global rate, we treat the processes in the regions as independent.

Rougier *et al. Journal of Applied Volcanology* (2018) 7:1

Page 8 of 10

Again, this assumption is made in the absence of evidence to the contrary.

Appendix: statistical model and MCMC sampler

Let $x = (x_1, \ldots, x_n)$ denote the times of recorded large eruptions in the interval (a, b), measured in years, let λ denote the rate of large eruptions, with units of 'year^{-1}', and let $r : (a, b) \to [0, 1]$ denote the recording probability of large eruptions as a function of time. The recording probability is expressed as a deterministic function of (u, v), where $u = r(a) \in [0, 1]$ and $v = (v_1, \ldots, v_m)$ are knot locations satisfying

$$a < v_1 < \cdots < v_m < b_0 < b;$$

see Fig. 1.

The full model factorizes as

$$p(\lambda, u, v, x) = p(\lambda) p(u) p(v) p\{x \mid \lambda, r(\cdot; u, v)\}, \quad (2)$$

treating λ, u, and v as mutually independent. For the marginal distributions, $p(\lambda) = \text{Gamma}(\alpha_0 = 0.5, \beta_0 = 20)$, $p(u) = U(0, 1)$, and $p(v)$ is the distribution of the order statistics of m IID $U(a, b_0)$ random variables. The likelihood function is that of a Poisson process with rate function $\lambda \, r(\cdot; u, v)$,

$$p(x \mid \lambda, r) = \exp\{-\lambda \, R(u, v)\} \prod_{i=1}^{n} \{\lambda \, r(x_i; u, v)\}, \quad (3)$$

where $R(u, v) = \int_a^b r(t; u, v) dt$; see Davison (2003, sec. 6.5) for the derivation.

The recording probability is modelled as the piecewise linear interpolant of $m + 3$ knots, subject to boundary conditions:

$$\begin{aligned} r(t; u, v) &= u + (1 - u) \, s(t; v) \\ &= s(t; v) + u \{1 - s(t; v)\}, \end{aligned} \quad (4a)$$

where s is the piecewise linear interpolant of

$$(a, 0), (v_1, \Delta), (v_2, 2\Delta), \ldots, (v_m, m\Delta), (b_0, 1), (b, 1) \quad (4b)$$

and $\Delta = 1/(m + 1)$.

The specified quantity b_0 is the time at which all large eruptions are sure to be recorded (taken to be 1980 in this paper). The value of $S(v) = \int_a^b s(t; v) \, dt$ can be evaluated in closed form, from which

$$\begin{aligned} R(u, v) &= u \, (b - a) + (1 - u) \, S(v) \\ &= S(v) + u \{(b - a) - S(v)\}. \end{aligned} \quad (5)$$

In (4a) and (5) both versions of r and R are useful in the factorisations below.

A Markov Chain Monte Carlo (MCMC) sampler is used to estimate expectations with respect to the posterior distribution, $p(\lambda, u, v \mid x)$. Gibbs proposals are used for

each of the three components, although only the first, λ, has a recognisable full conditional distribution. The full conditional for λ is

$$\begin{aligned} p(\lambda \mid u, v, x) &\propto p(\lambda) p(x \mid \lambda, r) \\ &\propto \exp\{-\beta_0 \, \lambda - \lambda \, R(u, v)\} \, \lambda^{\alpha_0 - 1 + n} \end{aligned} \quad (6)$$

which is $\text{Gamma}\{\alpha_0 + n, \beta_0 + R(u, v)\}$. Hence small values of α_0 and β_0 are rapidly dominated by the observations.

The full conditional for u is

$$\begin{aligned} p(u \mid \lambda, v, x) &\propto p(u) p(x \mid \lambda, r) \\ &= \mathbb{1}(0 \leq u \leq 1) \exp\{-\lambda \, R(u, v)\} \prod_{i=1}^{n} \{\lambda \, r(x_i; u, v)\} \\ &\propto \mathbb{1}(0 \leq u \leq 1) \exp\left[-\lambda \, u \{(b - a) - S(v)\}\right] \\ &\quad \times \prod_{i=1}^{n} r(x_i; u, v). \end{aligned}$$

$$(7)$$

This non-standard distribution requires a Metropolis-Hastings (MH) step. We use a slice sampler (Neal 2003, algorithms in Figs. 3 and 5, $K = \infty$, $w = 0.1$).
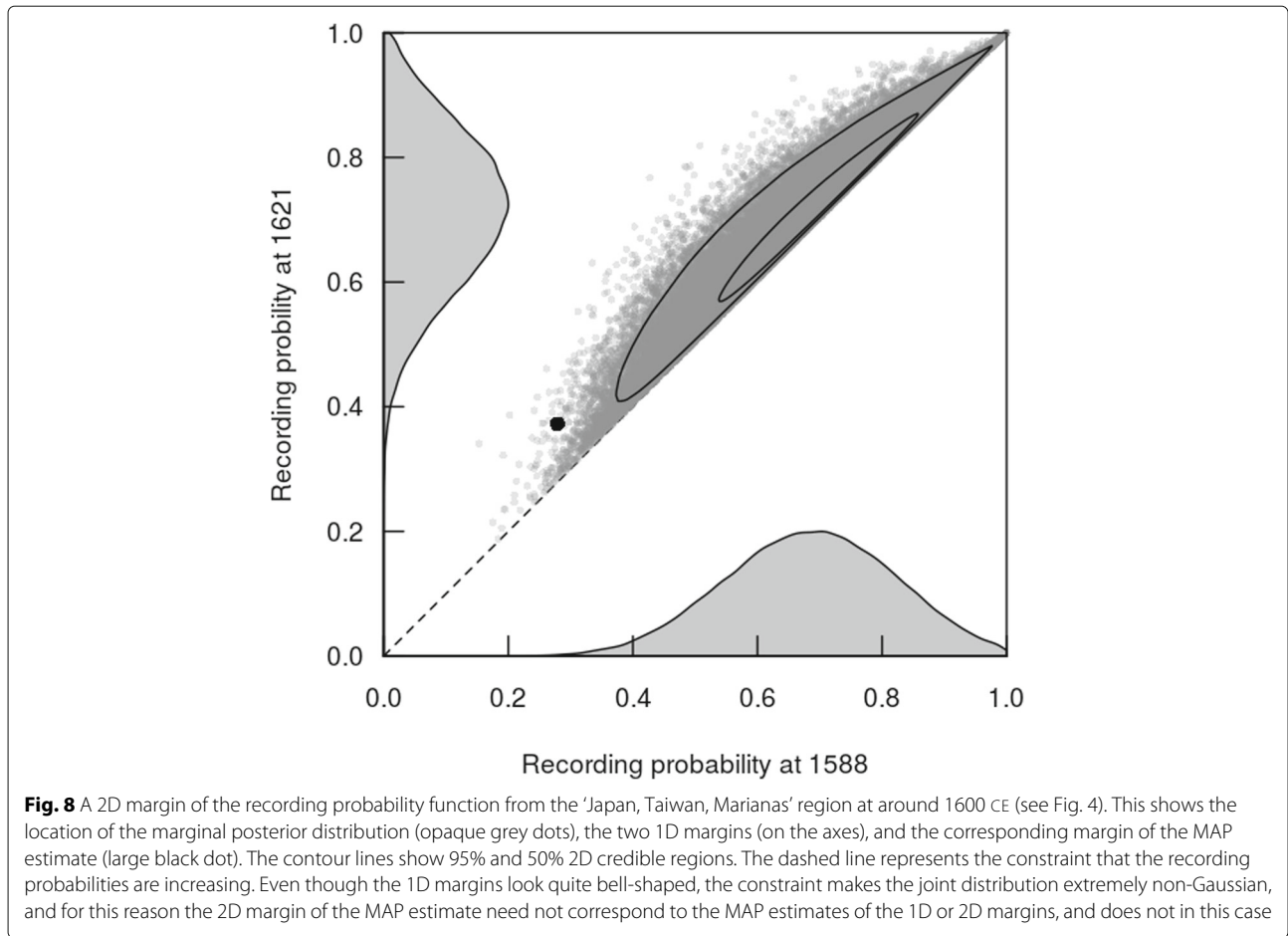
The full conditional for v is

$$\begin{aligned} p(v \mid \lambda, u, x) &\propto p(v) p(x \mid \lambda, r) \\ &\propto \mathbb{1}(a < v_1 \leq \cdots \leq v_m \leq b_0) \\ &\quad \times \exp\{-\lambda \, (1 - u) \, S(v)\} \prod_{i=1}^{n} r(x_i; u, v). \end{aligned}$$

$$(8)$$

This non-standard distribution also requires a MH step. We use a simple 'pick and fling' proposal, in which a knot is picked uniformly at random and then relocated to somewhere else in (a, b_0) uniformly at random.

Burn-in for the MCMC sampler was assessed using the Brooks-Gelman diagnostic; see Brooks and Gelman (1998, sec. 3) and Lunn et al. (2013, sec. 4.4). This was based on 8 parallel chains initialised independently from the prior distribution $p(\lambda, u, v)$. 2000 iterations were sufficient for burn-in, and 5000 were used. The code for the MCMC sampler was verified using the method of Cook et al. (2006), based on 50 samples from the prior distribution. For each region, 8 parallel chains were run for 10^5 iterations each (after burn-in) to collect samples.

Visualization. The recording probability function r is a scalar function $(a, b) \to [0, 1]$, constrained in the prior to be increasing and to attain 1 at time $b_0 \leq b$. We visualize this uncertain function in two ways. First, we show the pointwise uncertainty for a specified time $t \in (a, b)$, i.e. the marginal posterior distribution of $r(t)$. We use 50 and 95% equi-tailed credible intervals (CIs) for each t. For

Rougier *et al. Journal of Applied Volcanology* (2018) 7:1

Page 9 of 10

Fig. 8 A 2D margin of the recording probability function from the 'Japan, Taiwan, Marianas' region at around 1600 CE (see Fig. 4). This shows the location of the marginal posterior distribution (opaque grey dots), the two 1D margins (on the axes), and the corresponding margin of the MAP estimate (large black dot). The contour lines show 95% and 50% 2D credible regions. The dashed line represents the constraint that the recording probabilities are increasing. Even though the 1D margins look quite bell-shaped, the constraint makes the joint distribution extremely non-Gaussian, and for this reason the 2D margin of the MAP estimate need not correspond to the MAP estimates of the 1D or 2D margins, and does not in this case

example, the 50% CI comprises the interval between 25th and 75th percentiles of the posterior distribution of $r(t)$. By applying this calculation for a regular grid of t values spanning (a, b) and joining the results together, we can show the pointwise uncertainty as polygons with the specified coverage.

However, the posterior distribution is extremely non-Gaussian, owing to the monotonicity constraint on r, which constrains the posterior distribution into a space which, in high dimensions, is all corners. Therefore the polygon of pointwise posterior distributions can fail to convey the region of high posterior concentration, as represented by the mode. This is because, in general, the mode does not marginalize—the Gaussian distribution is a very special case where the mode does marginalize. Consider, for example, this tableau of probabilities for (x, y), where each of x and y can take one of three values:

	x_1	x_2	x_3	
y_1	0.2	0.1	0	0.3
y_2	0.1	0.15	0.15	0.4
y_3	0.05	0.15	0.1	0.3
	0.35	0.4	0.25	1.0

The modal value is (x_1, y_1), but the marginal modes are x_2 and y_2. Thus the 1D margins of the mode (x_1 and y_1) do not correspond to the modes of the 1D margins (x_2 and y_2).

We augment each plot with the *maximum a posteriori* (MAP) estimate (i.e. the posterior mode) of the entire recording probability function, estimated by the location of the highest marginal posterior probability (after integrating out λ) in the chains. This modal value need not lie inside the polygons, for the reason given immediately above. To illustrate this, Fig. 8 shows a 2D margin from the 'Japan, Taiwan, Marianas' region at around 1600 CE, plus the corresponding values from the MAP estimate.

Rougier *et al. Journal of Applied Volcanology*   (2018) 7:1

Page 10 of 10

Authors' contributions

JR carried out the statistical analysis, SS and KC provided volcanological expertise. All authors read and approved the final manuscript.

Competing interests

The authors declare they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

[1] School of Mathematics, University of Bristol, University Walk, Bristol BS8 1TW, UK. [2] School of Earth Sciences, University of Bristol, Bristol, UK.

References

Bonadonna C, Ernst G, Sparks R. Thickness variations and volume estimates of tephra fall deposits: the importance of particle Reynolds number. J Volcanol Geotherm Res. 1998;81:173–87.

Brooks S, Gelman A. General methods for monitoring convergence of iterative simulations. J Comput Graph Stat. 1998;7(4):434–55.

Brown S, Crosweller H, Sparks R, Cotterell E, Deligne N, Guerrero N, Hobbs L, Kiyosugi K, Loughlin S, Siebert L, Takarada S. Characterisation of the Quaternary eruption record: Analysis of the Large Magnitude Explosive Volcanic Eruptions (LaMEVE) database. J Appl Volcanol. 2014;3(5):. http://www.appliedvolc.com/content/3/1/5.

Cook S, Gelman A, Rubin D. Validation of software for Bayesian models using posterior quantiles. J Comput Graph Stat. 2006;15(3):675–92.

Crosweller H, Arora B, Brown S, Cotterell E, Deligne N, Guerrero N, Hobbs L, Kiyosugi K, Loughlin S, Lowndes J, Nayembil M. Global database on Large Magnitude Explosive Volcanic Eruptions (LaMEVE). J Appl Volcanol. 2012;1(4):. http://www.appliedvolc.com/content/1/1/4.

Davison A. Statistical Models. Cambridge: Cambridge University Press; 2003.

Lunn D, Jackson C, Best N, Thomas A, Spiegelhalter D. The BUGS Book: A Practical introduction to Bayesian Analysis. Boca Raton: CRC Press; 2013.

Mason B, Pyle D, Oppenheimer C. The size and frequency of the largest explosive eruptions on earth. Bull Volcanol. 2004;66(8):735–748.

Neal R. Slice sampling. Annal Stat. 2003;31(3):705–741.

Pyle, D. Sizes of volcanic eruptions. In: Sigurdsson H, Houghton B, McNutt S, Rymer H, Stix J, editors. Encyclopedia of Volcanoes. London: Academic Press. 2000.

Pyle D. Field observations of tephra fallout deposits. In: Mackie S, Cashman K, Ricketts H, Rust A, Watson M, editors. Volcanic Ash: Hazard Observation. Amsterdam: Elsevier Ltd. 2016. p. 25–37.

R Core Team. R: A Language and Environment for Statistical Computing. Vienna: R Foundation for Statistical Computing; 2016.

Rougier J, Sparks R, Cashman K. Global recording rates for large eruptions. J Appl Volcanol. 2016;5:11. https://doi.org/10.1186/s13617-016-0051-4.

Siebert L, Simkin T, Kimberly P. Volcanoes of the World, 3rd edition. Berkeley and Los Angeles: University of California Press; 2010.