A Hierarchical Model for Volcanic Repose Times

Un modello gerarchico per tempi di riposo di eruzioni vulcaniche

Claudia Furlan, Stuart Coles
Dipartimento di Scienze Statistiche, Università di Padova
Via C. Battisti 241/243, 35121 Padova
furlan@stat.unipd.it, coles@stat.unipd.it

1. Introduction

One of the most pressing concerns in volcanology is the assessment of risk due to the eruption of Holocene volcanoes (Simkin et al., 2004). Assessment based on current geological measurements is often the most precise form of analysis, but longer-term risk assessments generally have to be made without such precise information. In this case, it is necessary to make assessments based on the known history of relevant volcanic activity. Unfortunately (from a statistical, if not geological, point of view!) most volcanoes erupt only rarely, so there is precious little information with which to provide an analysis. Moreover, distinguishing between volcanoes that are extinct, rather than merely dormant, is a difficult exercise.

In a recent article, Simkin et al. (2004) addressed these issues by modelling the repose times between recorded events at major volcanic sites worldwide. Simkin and Siebert (1994) compiled a list of all known eruptions that have occurred throughout the world during the last 10,000 years. Of the 7886 dated eruptions is the present record, about 113 were major explosive events with a VEI of 5 or greater, for a total of 48 volcanoes (Note, VEI, the volcanic explosive index, is one measure used to record the magnitude of a volcanic event). The “repose interval” derived from the record is actually the number of years from the start of a previous eruption (not necessarily as extreme as 5 VEI) to the start of the “significant” event in question.

In their analysis Simkin et al. (2004) demonstrated that the log-logistic distribution provides a much better fit to the available sample of repose times than an exponential distribution. The distinction between the exponential and log-logistic distributions is important since, as is well known, the hazard rate is constant only in the former case. Thus, if repose times were genuinely log-logistic, there would be some predictive information for time to next event from time to previous event. If the true model were exponential, this would not be the case.

One feature of Simkin et al. (2004)’s analysis that needs to be considered however, is that data are pooled across all 48 volcanoes in the database, to which a single model is fitted. Consequently, their log-logistic model has an interpretation limited to a repose time from a randomly selected volcano, rather than from a specific volcano. Naturally, it would be more informative to make assessments that are volcano-specific, but from a statistical perspective this creates difficulties due...
to the scarcity of data.

We address this problem by formulating and comparing hierarchical models, enabling each individual volcano to have its own distribution of repose times - either exponential or log-logistic - but with a hierarchy across the model distribution parameters that enables some sharing of information. Here we present the first results about the exponential model; therefore, our interest again is to assess whether the log-logistic model performs better than the exponential - but now on a volcano-by-volcano basis - and what can be learnt in terms of distinguishing between dormant and extinct volcanoes.

Figure 1: Log-repose times for 48 volcanoes having generated eruptions with explosivity greater than VEI 5. Observations from the same volcano are connected with a dotted line.

2. The Repose Times process

Let $T_{ij}$ be the $i$th repose time of volcano $j = 1, \ldots, 48$, for $i = 1, \ldots, I_j$. Figure 1 plots the available data on a logarithmic scale, with events from the same volcano plotted vertically. A number of features are immediately evident, most notably the scarcity of data, especially for some volcanoes. Notwithstanding this concern however, there does also appear to be a heterogeneity of behaviour across volcanoes that suggests a simple model that simply pools all data into a single distribution could be misleading.

This presents something of a dichotomy: the heterogeneity implies we should not assume a common distribution of repose times for all volcanoes, while the sparsity of information prohibits an informative analysis on a volcano-by-volcano basis. A natural approach to tackle this dichotomy is through hierarchical models. For example, assuming a basic exponential distribution for
individual volcano repose times, we build the model
\[
T_{ij} \sim \exp(\theta_j) \\
\theta_j \sim \exp(\lambda) \\
\lambda \sim \Gamma(a, b).
\] (1)

The parameters \(a\) and \(b\) are hyper-parameters that determine the extent to which the model shares, or resists the sharing of, information across volcanoes. Replacing the exponential component with a log-logistic gives an alternative model, though it is then necessary to specify a hyper-prior model for both parameters of the log-logistic distribution.

Inference on all the model parameters, and in particular the \(\{\lambda, \theta_j, j = 1, \ldots, 48\}\), is straightforward via MCMC (Gilks et al., 1996). In particular, the conditional conjugate structure (Bernardo and Smith, 2000) of our model, at least for the exponential version, permits Gibbs sampling to be exploited. Specifically, the full conditional distributions at iteration \(w\) are
\[
\begin{align*}
\theta_j|w, T_{ij} & \sim \Gamma \left(1 + I_j, \lambda_{w-1} + \sum_{i=1}^{I_j} T_{ij} \right) \\
\lambda_w|\theta_{jw} & \sim \Gamma \left(a + 48, b + \sum_{j=1}^{48} \theta_{jw} \right).
\end{align*}
\] (2)

Based on choices of parameters \(a\) and \(b\) that seemed to provide a reasonable balance between data fidelity and information transfer, the estimated posterior mean of \(\lambda\) was found to be 286.63 with a 95\% credibility interval of \((179.81, 428.82)\). This provides information on the pattern of variation across volcanoes. The volcano-specific information is provided by the \(\theta_j\) parameters. However, for interpretation it is easier to look at a transform of these parameters:
\[
\delta_j = \log(1/\theta_j); \quad j = 1, \ldots, 48.
\] (3)

Figure 2 (left panel) shows the estimated posterior means of the \(\delta_j\), together with their associated 95\% credibility intervals, for each volcano \(j\). This analysis reinforces the suggestion that it would be a mistake to treat repose timese from different volcanoes as homogeneous processes.

Figure 2: Left panels: Points represent \(\hat{\delta}_j\) (estimated posterior mean of \(\delta_j\)). Vertical lines indicate 95\% credibility intervals. Right panel: Estimated probability to observe a following repose time \(T_{(I_j+1)j}\), for each volcano \(j\).

Though the model exploits information across all volcanoes, it can be used to make probability estimates that are volcano-specific. For example, we can begin to address the problem of identifying
extinct volcanoes. As a first step we can calculate the probability, via the model, of any particular volcano lying dormant from the most recent event to the current time. This is given by

$$P(T_{(I_j+1)j} > t^*) = \int \exp(-\theta_j t^*) f(\theta_j) d\theta_j,$$

where $t^*$ is the time from current day to the most recent recorded event and $f(\theta_j)$ is the marginal posterior density of $\theta_j$; it is easily approximated using the MCMC output. The smaller this probability, the more likely it is that the volcano is now extinct, or at least that its rate of eruption has diminished. Figure 2 (right panel) shows the estimated probabilities for each volcano, except for three (6, 19, 20) whose pattern of data included recent missing values that complicate the analysis in a way that we have so far not addressed. On the basis of this analysis alone, we would conclude that the volcanoes 14 (Llaima, Chile) and 21 (Avachinsky, Kamchatka) with estimated probabilities 0.0169 and 0.0143 respectively, are those that are most likely to be extinct rather than just dormant.

### 3. Conclusion and Discussion

The distinction between a log-logistic and exponential distribution for volcanic repose times is fundamental for risk assessment. Though a log-logistic distribution is apparently necessary to fit the available sample of all repose times, this does not take into account the apparent heterogeneity across volcanoes. We have proposed a hierarchical model that allows some pooling of information, whilst avoiding an assumption of homogeneity in process behaviour. On the basis of this model, we have identified those volcanoes that have remained dormant for unusually long periods given their previous activity. We must stress however, that this is a restricted analysis, using only very limited information. It is obvious that if other geological or statistical information is available, this will also impact upon the determination of the possible extinction of a volcano. However, without additional information, we feel that the use of a hierarchical model within a Bayesian framework is likely to be superior to other statistical approaches to the problem.

**Acknowledgments:** This work was supported by the University of Padova (Italy) grant CPDA037217: “Methods for the analysis of extreme sea levels and for coastal erosion”.

**References**