Statistics in Volcanology
Volume 1
December 11, 2012

Models for temporal volcanic hazard

Mark S. Bebbington
Massey University, New Zealand

Abstract

In estimating hazard from a currently quiescent volcano, the most basic quantity of interest is the likelihood of an eruption in some defined time horizon. Starting with the dichotomy of stationarity (where the average future level of activity is equal to the average level of past activity) or non-stationarity, we outline several classes of stochastic models that can be used to forecast future onsets. Renewal models, including the simple Poisson process and mixtures, are compared with models that incorporate volumes of past eruptions, and models that include a trend in the activity level. The mathematical formulations are supplemented by Matlab programs that fit the models using maximum likelihood. Tests are provided for whether a particular model is consistent with the data, and for identifying the best model from those considered. The philosophy behind assumptions and the limitations of each class of models are discussed, and suggestions for further exploration are given. The models are illustrated on a data set of VEI > 1 eruptions from Mt Ruapehu (New Zealand) since 1860.

Keywords: hazard, volcano, Mt Ruapehu, stochastic model

Correspondence:
MSB: m.bebbington@massey.ac.nz
Volcanic Risk Solutions, Massey University, Private Bag 11222, Palmerston North 4442, New Zealand.

citation:
Bebbington, M. S. (2012) Models for temporal volcanic hazard *Statistics in Volcanology*, **1**:1–24,
URL http://dx.doi.org/10.5038/2163-338X.1.1

Introduction

The quantitative forecasting of hazard is a key aim of volcanology. Resulting models may include the size, location and/or style of eruption, in addition to the timing which is the focus of this paper. Many complex deterministic models exist for various volcanic processes, but provide little information about future activity. This complexity means that volcanic onset data is inherently stochastic in nature, and requires statistical models.

We will avoid questions concerning monitoring and short-term eruption warnings; the reader is directed to a series of papers on Bayesian Event Trees (*Marzocchi et al.*, 2008) and references therein. Medium-long term hazard estimation is easier for volcanoes than for earthquakes; in the case of polygenetic volcanoes, there is no spatial element, and it is often possible to use geology to look backward a long way to see what happened in the past.

During a period of repose, the track record of a potentially active volcano provides the best method of assessing its future volcanic hazards on a long-term basis (*Decker*, 1986). However, this comes with a caveat: given apples, we are unable to forecast oranges or, in other words, we can only forecast the type of event that is present in the data. The data is a record of eruptions, so what constitutes an eruption? Here we immediately encounter a problem, in that eruptions are not point events in time. Further, at any temporal scale, eruptions are a mix of activity and quiet. Hence it is necessary to impose an arbitrary temporal limit on the length of surface quiet required to separate one eruption from the next. *Siebert & Simkin* (2002-) use three months for this purpose. Hence, using data consistent with this definition means that we must be attempting to forecast the onset of such eruption 'episodes', rather than, *e.g.*, short-lived explosive eruptions, of which there may be one or more in a given eruption episode. Moreover, this is compatible with the content of historical records, which generally consist of onset dates, and some measure of eruption size such as VEI (*Newhall & Self*, 1982), or possibly volume. Eruption durations are much less frequently observed in historical data. While there are a number of persistently erupting volcanoes, such as Stromboli, Kilauea and Popocatepetl, for which data is available at a much finer time-scale, the more natural use of such data is in a short-term forecasting framework, *cf. Marzocchi & Bebbington* (2012), using monitoring techniques, rather than in a longer-term framework of forecasting *performed during a prolonged period of repose*, where the primary objective is to estimate when the volcano might become active once more.

We will now give a brief outline of some terminology and notation. The beginning of an eruption is the *onset*, although *event* can be used interchangeably. The *onset times* will be denoted t_i, for $i = 0, \ldots, n$, and considered to occur in an observation window $[S, T]$, where T is typically the present. Usually there is no information about the start of the observation period, so we will adopt the convention that $S = t_0$ in the following, and hence the analysis will be restricted to the n onsets $t_1, \ldots, t_n$. The ith onset at time t_i may be considered to have a *volume* v_i. The *repose* is formally the period during which an eruption is not in progress, but here, as in most papers on this subject, it is considered to be the interval between onsets, that is, the *repose time* (or more correctly, the *inter-onset time*) and is denoted by $r_i = t_i - t_{i-1}$, for $i = 1, \ldots, n$.

We shall see that the two concepts (onset/repose) of time involved underlie two very different families of models. The onset time scale, which we will denote by s, t and u, is *absolute time*, *i.e.*, it is the elapsed time since some fixed origin. We will discuss the inherent problems with identifying a fixed origin later. In contrast, the *relative time* since the last eruption will be denoted by r.

Let the number of events in a time interval (s, t) be denoted $N(s, t)$, where we use the shorthand $N(t)$ if $s = 0$. The *point process intensity* (we shall henceforth use the shorthand 'intensity') of a process is

$$\lambda(t|H_t) = \lim_{\Delta t \downarrow 0} \frac{\Pr(N(t, t + \Delta t) = 1|H_t)}{\Delta t},$$

i.e., in the short period of time $(t, t + \Delta t)$, the probability of an event is approximately $\lambda(t)\Delta t$. The intensity is conditional on H_t, the *history* (*i.e.*, the events) of the process prior to t, and hence is often termed the conditional intensity. The intensity completely characterizes the process in statistical terms.

The paper will be illustrated using data from Mt Ruapehu, sourced from *Siebert & Simkin* (2002-) (cutoff of VEI > 1), with additional volume (in units of 10^6m^3) information from *Latter* (1985) and S. Cronin (personal communication). The data can be found in the accompanying file `ruapehu_year_volume.txt`.

Forecasting and Stationarity

The first question to be examined is whether the observed process is *stationary* in time. Thus, we need to understand what stationarity means. As we are interested in how past occurrences can be used to predict future ones, this is easiest if the underlying process is the same in both cases. To put this into perspective, let us consider the typical question: What is the probability of the next eruption occurring within time u? It is simpler from the algebraic view to examine the complementary probability: $\Pr\left[N(T, T+u) = 0\right]$, where T is the present time. In terms of the intensity, we have

$$\Pr\left[N(T, T+u) = 0\right] = \exp\left[-\int_T^{T+u} \lambda(t)dt\right], \tag{1}$$

and thus we are interested in the quantity $\int_T^{T+u} \lambda(t)dt$. There are two possibilities, deriving from the onset/repose dichotomy above. If the intensity varies with reference to some absolute time origin, then the forecast is different for every T. On the other hand, if the intensity depends only on the time since the last event, then the intensity resets at each event (sometimes called a 'regeneration point') and the forecast following each event is the same. The former we will describe as non-stationary (in time), the latter as stationary. There is also the concept of 'strong stationarity', which is effectively limited in our context to the Poisson process.

As the distribution of $N(t)$ is characterized by the intensity $\lambda(t)$, we can formulate our concept of stationarity as having a constant *expected* level of the conditional intensity, where the expectation is taken over all possible histories of the process. Note that this is purely temporal. *Wadge* (1982) gave a definition in terms of the cumulative eruptive volume which neither implies nor is implied by this. Besides including renewal processes which are stationary from event to event, but not necessarily from interval to interval, this definition includes other processes we shall meet later such as the Generalized Time Predictable Model (*Marzocchi & Zaccarelli*, 2006) and the Volume History Model (*Bebbington*, 2008). The key point is that there is nothing in these models that determines a change in activity level, beyond the history of the process itself. Another way to think of it is to ask, if we simulated the process a large number of times, and examined the average rate of events over time, would it be constant?

A consequence of our definition is that if the process is stationary, the intensity must be able to be rewritten as a function that does not depend on the absolute time t. However, it may include the relative time since the last eruption r, as these differ in each realization.

Having formulated a definition of stationarity, how do we test a volcanic record to see if it is compatible with stationarity? The key is that if the process is stationary, onsets should be statistically random (*i.e.*, according to a uniform distribution) over the observation interval $[S, T]$. There are a variety of tests for this, but given the usual small size of volcanic data sets, a good option is the Kolmogorov-Smirnov test. The nonparametric nature of the test makes it particularly robust. Let

$$F_n(t) = \frac{\#(t_i \le t)}{n}, \quad i = 1, \ldots, n, \quad S < t < T$$

be the empirical distribution of the onset times. The Kolmogorov-Smirnov test statistic is

$$D_n = \max_{t \in [S,T]} \left\{ \left| F_n(t) - \frac{t - S}{T - S} \right| \right\},$$

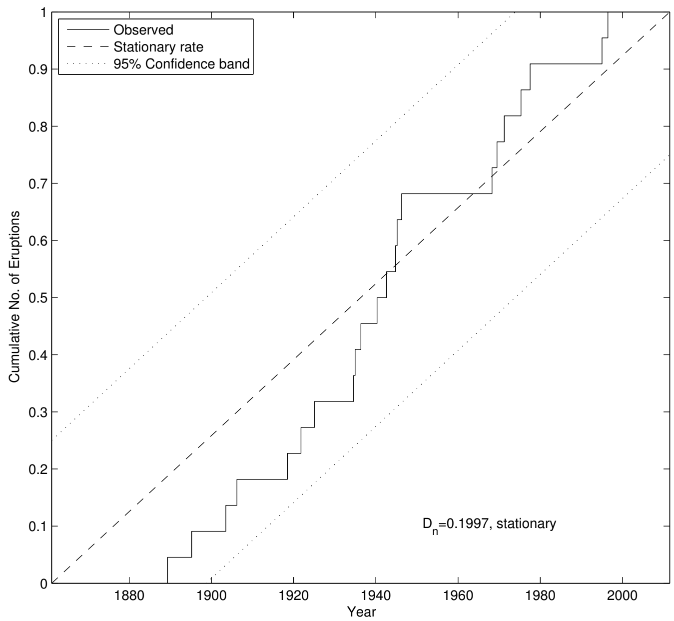

Figure 1: Test of stationarity for the VEI > 1 eruption record of Mt Ruapehu. D_n is the Kolmogorov-Smirnov test statistic.

i.e., the maximum distance between the empirical and theoretical distributions. The former is a step function, and the latter a straight line, and hence the maximum must occur at the top or bottom of a step in $F_n(t)$. Critical values are tabulated in, *e.g.*, *Zwillinger & Kokoska* (2000), but for large enough data sets, the 5% critical value is $1.36/\sqrt{n}$. The accompanying file `stationarity_and_independence_test.m` implements a stationarity test in Matlab. For Mt Ruapehu, with $[S, T] = [1861.1205, 2011.5]$, we see from Figure 1 that there is no evidence that the process is not stationary.

There are many other tests for stationarity, including a running mean of time between onsets (*Klein*, 1982), the theory of change-point problems (*Mulargia et al.*, 1987), the cumulative count of eruptions in a statistical control chart (*Ho*, 1992), and rank order statistics for the size of event (*Pyle*, 1998).

Stationary Models

There are a number of stochastic models which are stationary in the sense of our definition above. The simplest is the homogeneous Poisson process, but as this is a special case of a renewal process, we will begin with the latter.

Weibull Renewal Process

A renewal process is characterized by the intervals between events being independent and identically distributed. Thus,

$$\lambda(t) = \frac{f(t-s;\boldsymbol{\theta})}{1 - F(t-s;\boldsymbol{\theta})}, \quad t > s, \tag{2}$$

where the most recent event occurred at time $s < t$, and $f = F'$ is a density with parameter vector $\boldsymbol{\theta}$. Noting that $t - s$ is the elapsed repose time, we see that we can rewrite (2) as

$$\lambda(t) = \frac{f(r;\boldsymbol{\theta})}{1 - F(r;\boldsymbol{\theta})}. \tag{3}$$

Thus, only the elapsed time since the last eruption controls the time to the next eruption. Previous eruptions exert an influence only through their contribution to the parameter estimates $\hat{\boldsymbol{\theta}}$.

We have already discussed testing for stationarity. In order to satisfy ourselves that a process is a renewal process we must also check that successive intervals are independent (uncorrelated) and that the chosen distribution is a reasonable fit. The former can be checked by calculating the Spearman rank correlation of successive reposes. Let R_i be the rank of the repose length r_i, from shortest ($R = 1$) to longest ($R = n - 1$). Then Spearman's rank correlation is

$$\rho = \frac{\sum_{i=1}^{n-2}\left(R_i - (n-2)^{-1}\sum_{j=1}^{n-2} R_j\right)\left(R_{i+1} - (n-2)^{-1}\sum_{j=2}^{n-1} R_j\right)}{\sqrt{\sum_{i=1}^{n-2}\left(R_i - (n-2)^{-1}\sum_{j=1}^{n-2} R_j\right)^2 \sum_{i=2}^{n-1}\left(R_i - (n-2)^{-1}\sum_{j=2}^{n-1} R_j\right)^2}}, \tag{4}$$

and

$$\rho\sqrt{(n-4)/(1-\rho^2)}$$

is distributed approximately as a Student-t distribution with $n - 4$ degrees of freedom. For the Mt Ruapehu data, we obtain $\rho = -0.035$, with a P-value of 0.8819 (see `stationarity_and_independence_test.m`), and thus there is no significant correlation.

In order to test the distributional assumption, we need a distribution. For a variety of reasons, it is useful to consider a distribution that has the exponential as a special case. One such distribution is the Weibull distribution, suggested for volcanoes by *Bebbington & Lai* (1996*a*). The density is

$$f(r) = \alpha\beta(\beta r)^{\alpha-1}\exp\left(-(\beta r)^{\alpha}\right), \quad r > 0, \tag{5}$$

where α and β are the shape and scale parameters, respectively. Thus, $\alpha = 1$ is the exponential distribution, $\alpha < 1$ corresponds to an 'over-dispersed', or clustering distribution, and $\alpha > 1$ corresponds to a more periodic distribution with a mode at $r = \beta^{-1}(1 - 1/\alpha)^{1/\alpha}$. From (3) the intensity is then

$$\lambda(t) = \alpha\beta(\beta(t-s))^{\alpha-1}, \quad t > s, \tag{6}$$

which is monotonic, and can model either increasing probability of eruption as the repose time increases, or decreasing probability with increasing repose time.

Given a density $f(r;\boldsymbol{\theta})$ and observed inter-onset times r_i, $i = 1, \ldots, n$, the parameters $\boldsymbol{\theta}$ can be estimated by maximum likelihood. That is, the values are chosen, either algebraically or numerically, to maximize the likelihood

$$L(r_1, \ldots, r_n, r^*;\boldsymbol{\theta}) = S(r^*)\prod_{i=1}^{n} f(r_i;\boldsymbol{\theta}), \tag{7}$$

where r^* is the elapsed time since the most recent eruption, and $S(r) = 1 - F(r)$ is the survival function for the reposes, *i.e.*, the probability that a repose lasts longer than r. We will usually consider observation to have started at the first recorded onset, in the typical absence of other information. For the Mt Ruapehu data, we find (see `weibull_renewal_test.m`) that the maximum likelihood estimates (MLEs) are $\hat{\alpha} = 0.8882$, $\hat{\beta} = 0.1546$ per year.

Goodness of fit to a Weibull distribution can approximately be assessed using the Kolmogorov-Smirnov test although in this case, to account for the incomplete repose, we use the Kaplan-Meier Product-Limit estimator (*Lawless*, 2003) of the survival function, $\hat{S}(r)$, and calculate

$$D_n = \max_r \left\{ \left| \hat{S}(r) - \exp\left(-(\hat{\beta} r)^{\hat{\alpha}}\right) \right| \right\}.$$

This will have $n-2$ degrees of freedom, as two parameters are estimated from the data. The result (see `weibull_renewal_test.m`) for Mt Ruapehu, shown in Figure 2B, indicates there is no reason to reject a Weibull distribution.

Homogeneous Poisson Process

A special case of a renewal process is the exponential density $f(r) = \nu e^{-\nu r}$, in this case,

$$\lambda(t) = \frac{\nu e^{-\nu(t-s)}}{1 - \left(1 - e^{-\nu(t-s)}\right)} = \nu, \tag{8}$$

the homogeneous (constant rate) Poisson process. This is a stronger version of stationarity than needed, where the intensity has only to be constant in expectation. The equation (8) is sometimes referred to as the memoryless property, as it says that the time elapsed since the last eruption provides no information about the time of the next eruption. It is easy to show that the log likelihood of the Poisson process is

$$\log L = n \log \nu - \nu(T - S)$$

with MLE $\hat{\nu} = n/(T-S)$. Repeating the analysis for the Poisson process (see `poisson_process_test.m`) for Mt Ruapehu, we obtain $\hat{\nu} = 0.1463$, for the fit shown in Figure 2A.

Obviously the D_n values indicate that the Weibull renewal density is a better fit than the Poisson process, but this is guaranteed by the nesting of the distributions. Is the improvement in fit sufficient to justify the additional parameter? One way to answer this question is to use the Akaike Information Criterion,

$$\text{AIC} = -2k + 2\log L, \tag{9}$$

where k is the number of parameters, and $\log L$ the log likelihood. This compensates for the effect of additional parameters and thus avoids overfitting. Larger AICs indicate better models, with a difference of 1.5 to 2 in AIC generally being considered significant (*Akaike*, 1977). Note that we cannot use the corrected AIC (*Hurvich & Tsai*, 1989), as there is no proof of its validity for point process models (*Claeskens & Hjort*, 2008), unlike for linear regression and autoregressive models. For the Mt Ruapehu data we obtain the results in the first two rows of Table 1. We see that the improved log likelihood is insufficient to justify the additional parameter in going from the Poisson process to the Weibull renewal process, and hence there is no evidence, to this point, that the intensity is not constant over time.

Mixture renewal models

A recent innovation was the use of a renewal density consisting of a mixture of Weibull densities (*Turner et al.*, 2008) to model a multimodal repose distribution. The interpretation is that components correspond to differing conditions set up by the previous eruption, such as the closure of the conduit, or depletion of an upper magnitude chamber. For two components, the density can be written as

$$f(r) = p\alpha_1\beta_1(\beta_1 r)^{\alpha_1 - 1}\exp\left(-(\beta_1 r)^{\alpha_1}\right) + (1-p)\alpha_2\beta_2(\beta_2 r)^{\alpha_2 - 1}\exp\left(-(\beta_2 r)^{\alpha_2}\right), \quad r > 0, \tag{10}$$

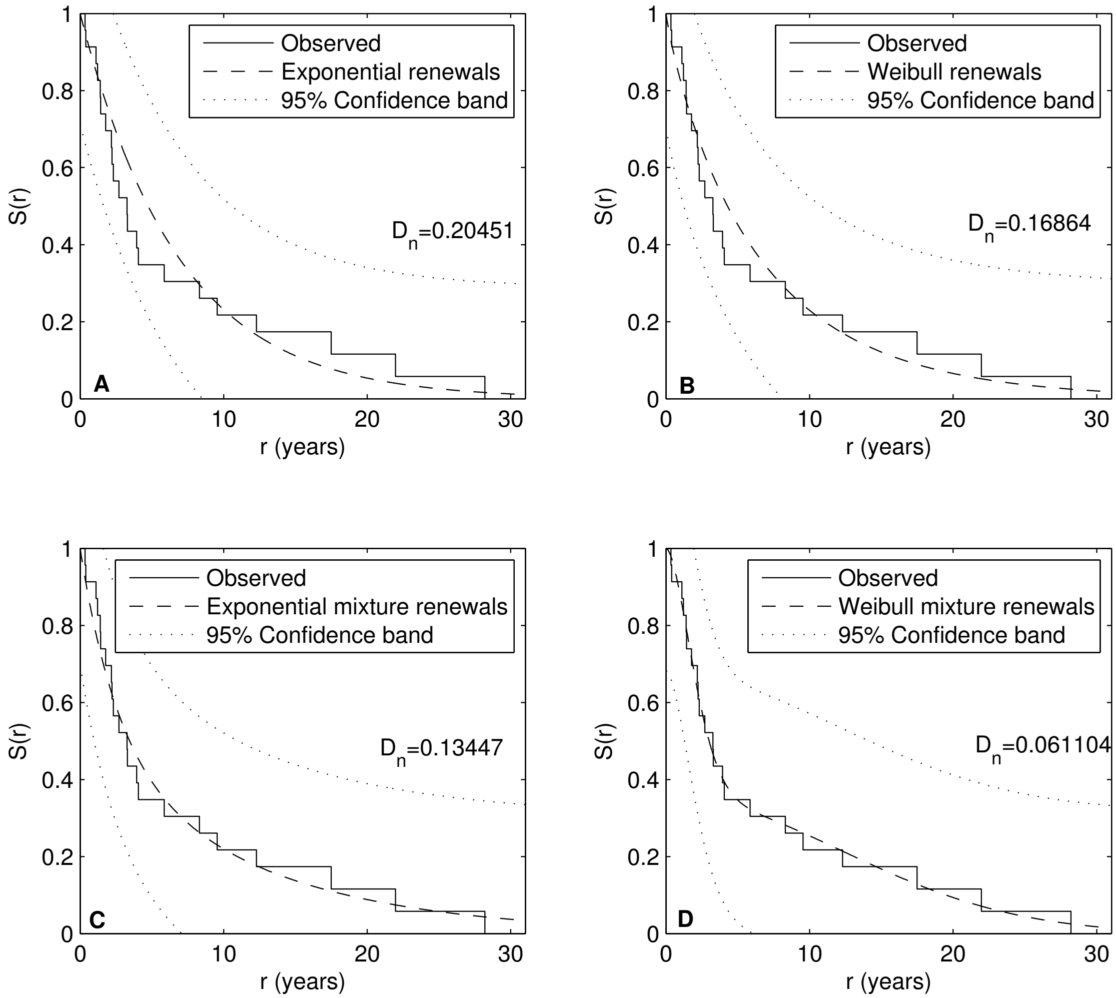

Figure 2: Fitting renewal processes to the VEI > 1 eruption record of Mt Ruapehu. D_n is the Kolmogorov-Smirnov test statistic. $S(r)$ is the probability of a repose being longer than r.

Table 1: Mt Ruapehu: Model fits

Process	k	**LL**	**AIC**
Weibull renewal process	2	-64.01	-132.01
Poisson process	1	-64.29	-130.57
mixture of Weibulls renewal process	5	-60.79	-131.58
mixture of exponentials renewal process	3	-63.12	-132.24
GTPM	3	-62.57	-131.15
volume-history	3	-58.15	-122.31
proportional hazard (Weibull renewal)	3	-64.01	-134.01
Weibull process	2	-62.95	-129.90

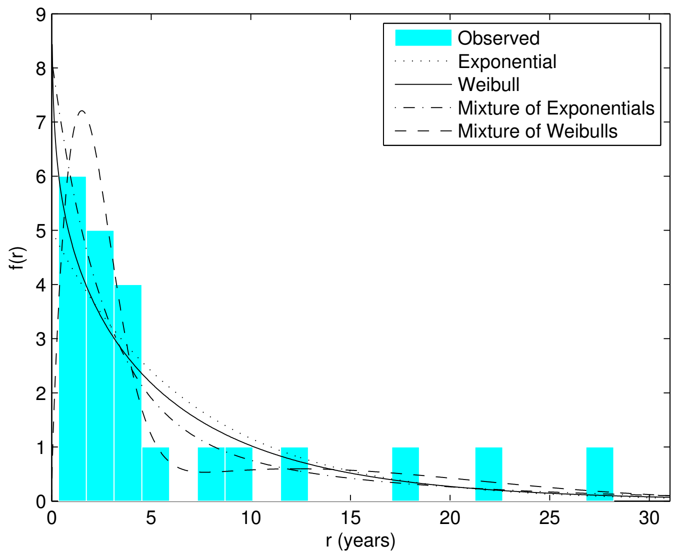

Figure 3: Renewal densities fitted to the VEI > 1 eruption record of Mt Ruapehu. $f(r)$ is the probability density function describing the repose lengths r.

where $0 < p < 1$ is the mixing parameter. The resulting intensity is capable of assuming many shapes (*Jiang & Murthy*, 1998). Applied to the Mt Ruapehu data (see `weibullmix_renewal_test.m`), we obtain Figure 2D. The MLEs are $\hat{\alpha}_1 = 1.7343, \hat{\beta}_1 = 0.4073/\text{year}, \hat{\alpha}_2 = 2.0002, \hat{\beta}_2 = 0.0577/\text{year}, \hat{p} = 0.6462$. This implies two non-zero modes in the repose distribution, one representing approximately two-thirds of the reposes at 1.5 years, and the other at 12.3 years. Although the fitted distribution follows the observed reposes very closely, as indicated by the small value of D_n, the additional 4 parameters, over the Poisson process, are not justified by AIC (Table 1).

A nested model to the mixture of Weibulls is the mixture of exponentials (*Mendoza-Rosas & De la Cruz-Reyna*, 2009), which can be obtained by setting $\alpha_1 = \alpha_2 = 1$ in (10). Fitting this to the Mt Ruapehu data (see `exponentialmix_renewal_test.m`), we obtain MLEs $\hat{\beta}_1 = 0.3918/\text{year}, \hat{\beta}_2 = 0.0867/\text{year}, \hat{p} = 0.5008$ (`exponentialmix_renewal_test.m` uses the notation ν_1 and ν_2 instead of β_1 and β_2 to avoid over-writing) as shown in Figure 2C. However, we see from the AICs in Table 1 that the mixture of exponentials is not as good an explanation of the data as the mixture of Weibulls. In fact it is not as good as the one-component Weibull, either. The reason is that both modes in the mixture of exponentials are at zero. Hence the mixture of exponentials is basically a way of constructing a longer tailed monotonically decreasing density, which is exactly what the Weibull renewal process achieves with a shape parameter $\alpha = 0.8882 < 1$. The densities of the various renewal distributions (see `plot_renewal_models.m`) are shown in Figure 3 from which we see that only the mixture of Weibulls density has positive modes. The intensities (see `plot_renewal_models.m`) are shown in Figure 4. While the exponential (Poisson process) intensity is constant, the Weibull renewal and mixture of exponentials decrease with the length of the repose. However, the mixture of Weibulls renewal starts at zero, rising to a peak at 2.2 years after the previous onset, then declines to a trough at about 7.3 years before rising again. Given that an eruption starting within days or weeks of a previous eruption will be identified as a continuation of the previous eruption, intensities with modes away from zero may be preferred.

The probability of the next eruption occurring within a time r from the present, given that the last eruption occurred a time r^* previously is, writing Q for the (unknown) length of the current repose

$$\Pr(Q < r + r^* | Q > r^*) = 1 - \Pr(Q > r + r^* | Q > r^*) = 1 - \frac{S(r + r^*)}{S(r^*)}.$$

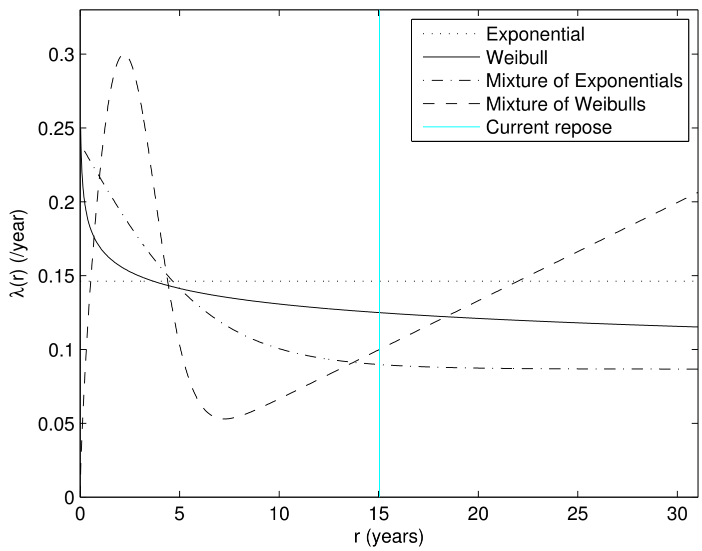

Figure 4: Renewal intensities fitted to the VEI > 1 eruption record of Mt Ruapehu. $\lambda(r)$ is the hazard (per year) at the time r, following the previous onset.

Calculating this for our renewal models (see `plot_renewal_models.m`) produces Figure 5.

Notes and further reading

The Spearman rank correlation (4) is recommended because the repose distribution may be highly skewed. An alternative is to transform the data by taking logarithms, and then use the usual correlation.

Bebbington & Lai (1996*b*) described a method of estimating the parameters in a Weibull renewal process using linear regression. The slope of the regression will be α, and hence standard methods for testing significance of the slope allow for the process to be tested against the Poisson process alternative. However, the method does not extend to more complex models. *Bebbington & Lai* (1996*a*) provided a number of other tests for assessing the applicability of the Weibull distribution, based on the fact that r_i^β should be exponentially distributed.

A number of other renewal densities have been used in volcanology including the gamma (*De la Cruz-Reyna & Carrasco-Nunez*, 2002) and the log-logistic (*Connor et al.*, 2003), which both have unimodal intensities, in contrast to the monotonic intensity (6), and the power-law density (*Pyle*, 1998) with a monotonically decreasing intensity. It is possible to fit mixture models with more than two components (*cf. Mendoza-Rosas & De la Cruz-Reyna* (2009)), but the goodness of fit needs to be carefully examined using AIC or an equivalent method.

The drawback of renewal models is that they commonly fail to explain variations in eruption rate, corresponding to changes in activity level. There are two approaches to this problem. The first option is to model the intensity as a function of absolute time, which we will meet later. The second option is to identify a number of *regimes*. These regimes may represent changes in the eruption mechanism, the mechanism for transport of magma to the surface, or the eruptive style. The estimated properties of such regimes can exclude certain models or mechanisms of volcanic eruption (*Klein*, 1982). In each regime, the volcano exhibits stationarity, but the level of activity differs between regimes. If transitions between regimes are themselves stochastic, and the regimes are recurrent, then the model meets our definition of stationarity. While stationarity tests can identify changes in regime, they can not be used to produce a forecast in the sense required in (1), as there is no model to predict the future regime(s).

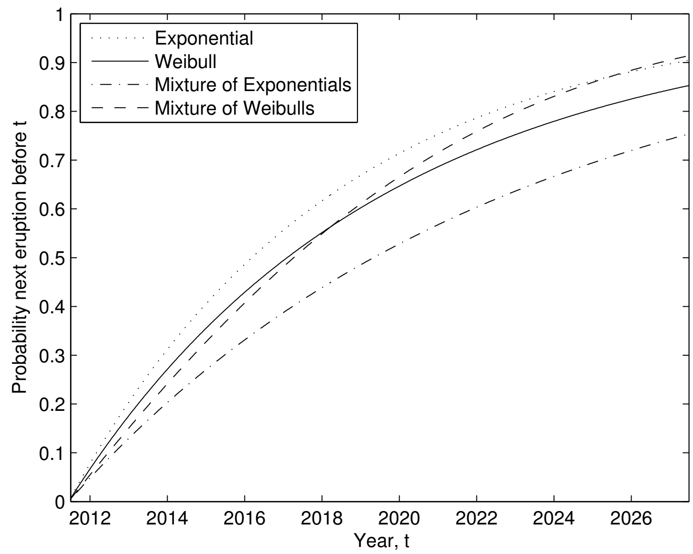

Figure 5: Time t to the next VEI > 1 eruption of Mt Ruapehu in various renewal models.

A mixture model is a trivial example of a regime model, with the regime of each repose being chosen randomly independent of the previous and subsequent regimes. *Cronin et al.* (2001) proposed a variant on the mixture renewal model, where onsets occurred in 'episodes', each of variable numbers of eruptions, with different repose distributions corresponding to inter- and intra-episode reposes.

Structural dependence in the regimes can be introduced by using a hidden Markov model (*Bebbington*, 2007), with the unobserved state representing the regime. Regimes can be statistically identified via the Viterbi algorithm, which finds the most likely path through the hidden states. Moreover, with a stochastic process specified for both the regimes and the activity within a regime, future activity can easily be forecast. The mathematics and programming are beyond the scope of this paper; see *Bebbington* (2007) for details.

In order to model flare ups in activity in the Auckland Volcanic Field, *Bebbington & Cronin* (2011) formulated a *self-exciting model* for the intensity. In this model every eruption increases the likelihood of a subsequent eruption, with the influence decaying over time.

Stationary models with covariates

The most common form of information we have about eruptions, after the onset time, is some measure of size. This could be the volume, the duration, or a more arbitrary figure such as the VEI. So how can this information help us to forecast future onsets?

Time-predictable models

De la Cruz-Reyna (1991) proposed a general *load-and-discharge* model, where the 'energy' of the volcano increases at a constant rate between eruptions. An eruption occurs when this exceeds a threshold H, during which the stored energy drops to a lower threshold L. If the threshold H is constant, the repose will be proportional to the energy of the preceding eruption, which is the *time-predictable model*.

In order to produce a model that can yield forecasts in the sense of (1), we need to build this paradigm into a stochastic model. The *Generalized Time Predictable Model* (*Marzocchi & Zaccarelli*, 2006) does this in the form of a renewal model

$$\lambda(t) = \frac{f(t-s|v)}{1-F(t-s|v)} \tag{11}$$

using a lognormal density, where v is the volume of the most recent eruption,

$$f(r|v) = \frac{1}{r\sigma\sqrt{2\pi}} \exp\left\{-0.5\left[\frac{\log(r)-\mu-\eta v}{\sigma}\right]^2\right\},$$

and

$$F(r|v) = \Phi\left[\frac{\log(r)-\mu-\eta v}{\sigma}\right]$$

where Φ is the standard normal distribution, and μ, η and $\sigma > 0$ are parameters to be estimated.

While this can be fit by maximizing a simple generalization of likelihood (7), and the AIC calculated, the question of whether the model is a good fit is harder to answer, as the reposes are no longer identically distributed. A way around this is to use the point process *compensator*. If a point process with intensity $\lambda(t)$, $S < t < T$ is a satisfactory fit to events at times $S < t_1 < t_2 < \cdots < t_n < T$, then the transformed points defined by

$$\tau_i = \int_S^{t_i} \lambda(t)dt \tag{12}$$

should be indistinguishable from a Poisson process of unit rate; in other words they must be stationary, successive intervals should be independent, and the intervals should have an exponential distribution (*Bebbington & Harte*, 2001). We will denote the transformed reposes by $r_i^\tau = \tau_i - \tau_{i-1}$. Fitting this model to the Mt Ruapehu data (see `gtpm_test.m`) we obtain MLEs of $\hat{\mu} = 1.3244, \hat{\eta} = -0.0144, \hat{\sigma} = 1.2065$, for a $\log L = -62.57$ and AIC $=$ -131.15. Although the fit is satisfactory according to Figure 6 and the P-value of 0.9707 for the correlation, the AIC indicates that the model is not as good as the Poisson process in this case.

Volume-history model

Our next candidate further generalizes the concept of the conditional intensity depending on the size of past eruptions. The *volume-history (dependent) model* (*Bebbington*, 2008) has

$$\lambda(t) = \exp\left\{\alpha + \nu\left[\rho t - V(t)\right]\right\}, \tag{13}$$

where $V(t) = \sum_{k:t_k<t} v_k$ is the cumulative volume erupted prior to time t, and $\alpha, \nu > 0, \rho > 0$ are parameters to be estimated. The term α incorporates the unknown state of the volcano at time S. Note that a regeneration point occurs whenever $\rho t - V(t)$ crosses a given threshold from below, and that between onsets $\rho t - V(t) = \rho r + C$, where C is a constant.

This is clearly not a renewal process, and the parameter estimates must be found by maximizing the point process log likelihood (*Daley & Vere-Jones*, 2003)

$$\log L = \sum_{i=1}^{n} \log\lambda(t_i) - \int_S^T \lambda(t)dt. \tag{14}$$

We see that the likelihood is increased (first term) by a large value of $\lambda(t)$ at the observed onset times $t_1, \ldots, t_n$, and decreased (second term) by large values of $\lambda(t)$ elsewhere.

Fitting the model to Mt Ruapehu data (see `volhist_test.m`) gives MLEs of $\hat{\alpha} = -2.7830, \hat{\nu} = 0.0944, \hat{\rho} =$ 0.4325 and $\log L = -58.15$ and an AIC of -122.31. It is thus the best fitting model we have seen so far, and as shown in Figure 7, the model is not rejected by the data. Although the degree of serial correlation in the

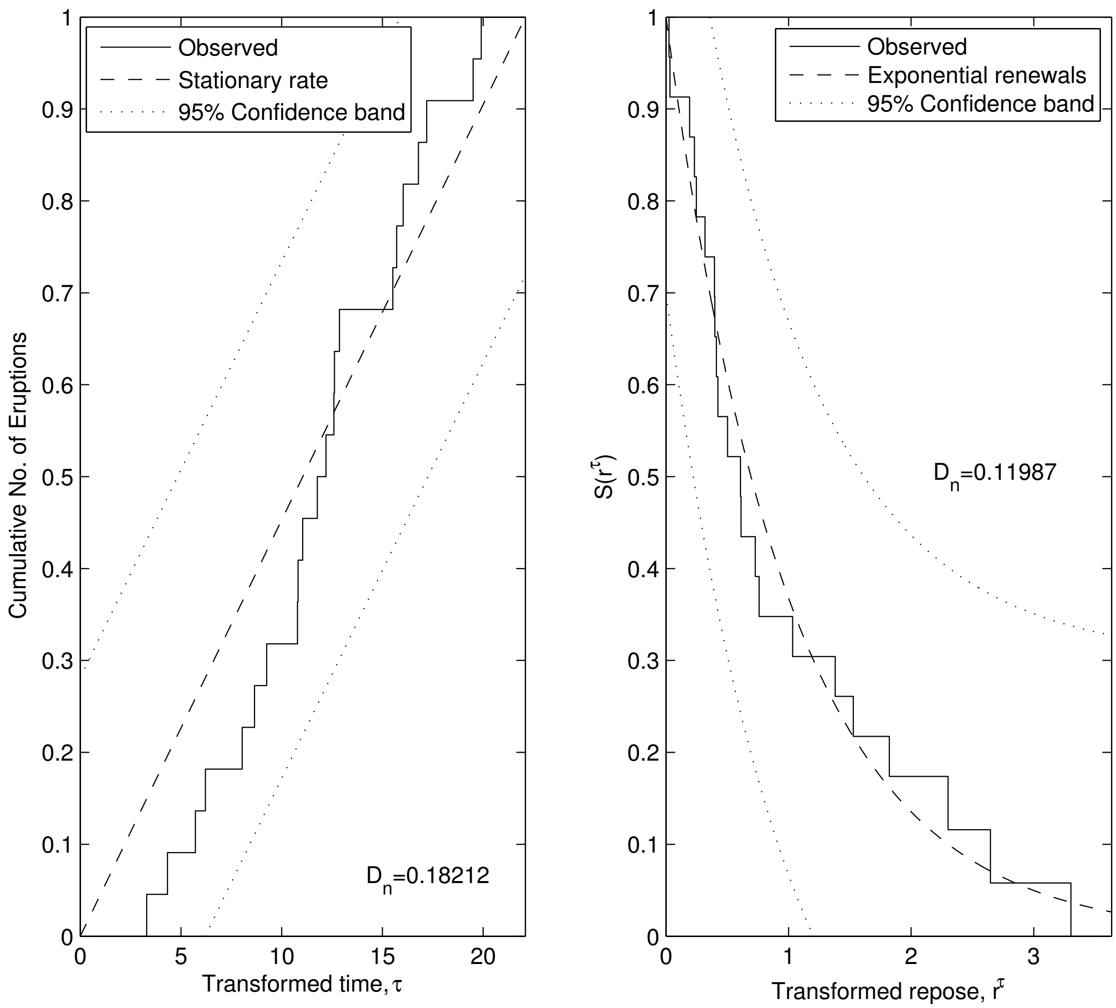

Figure 6: Transformed onsets (left) and intervals (right) for the GTPM fitted to the VEI > 1 eruption record of Mt Ruapehu. D_n is the Kolmogorov-Smirnov test statistic. $S(r^\tau)$ is the probability that the transformed repose is longer than r^τ.

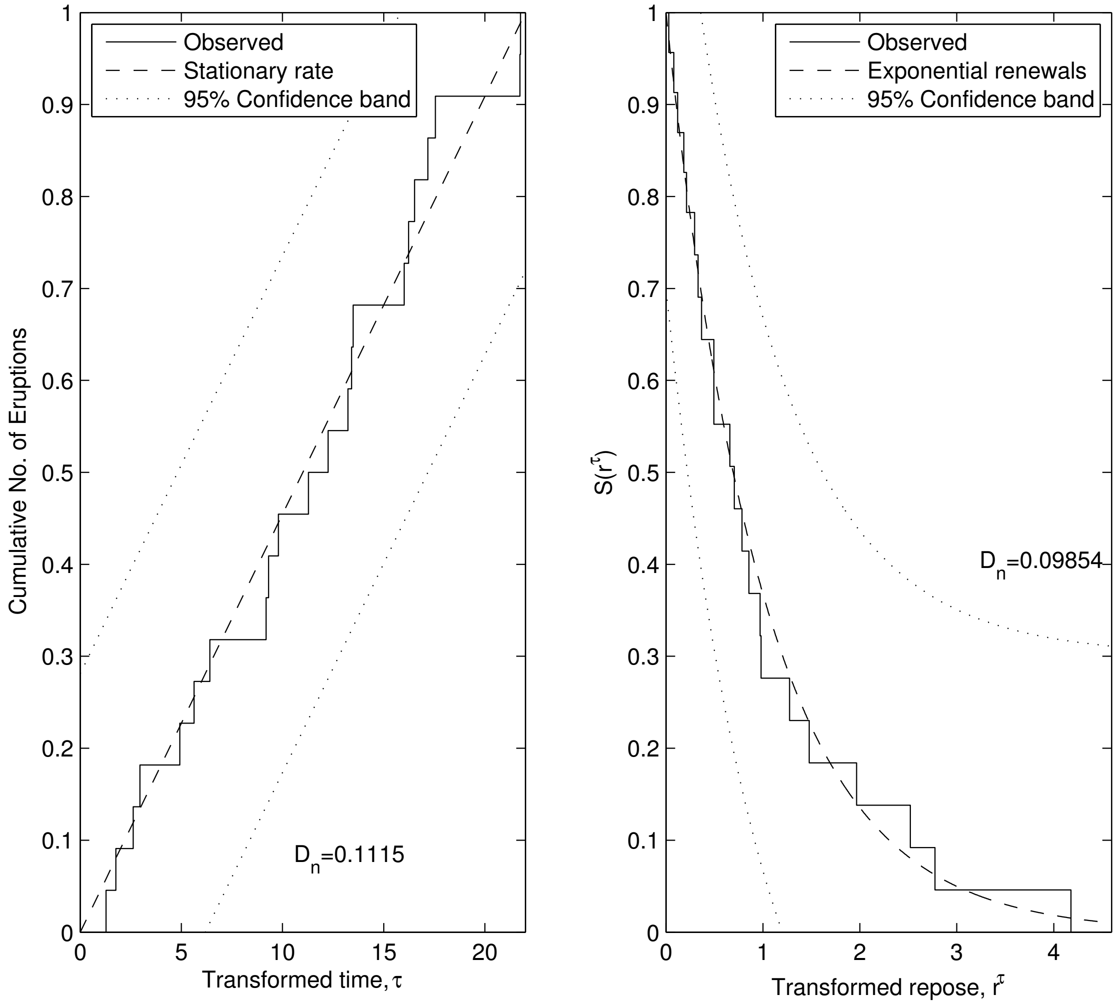

Figure 7: Transformed onsets (left) and intervals (right) for the volume history model fitted to the VEI > 1 eruption record of Mt Ruapehu. D_n is the Kolmogorov-Smirnov test statistic. $S(r^\tau)$ is the probability that the transformed repose is longer than r^τ.

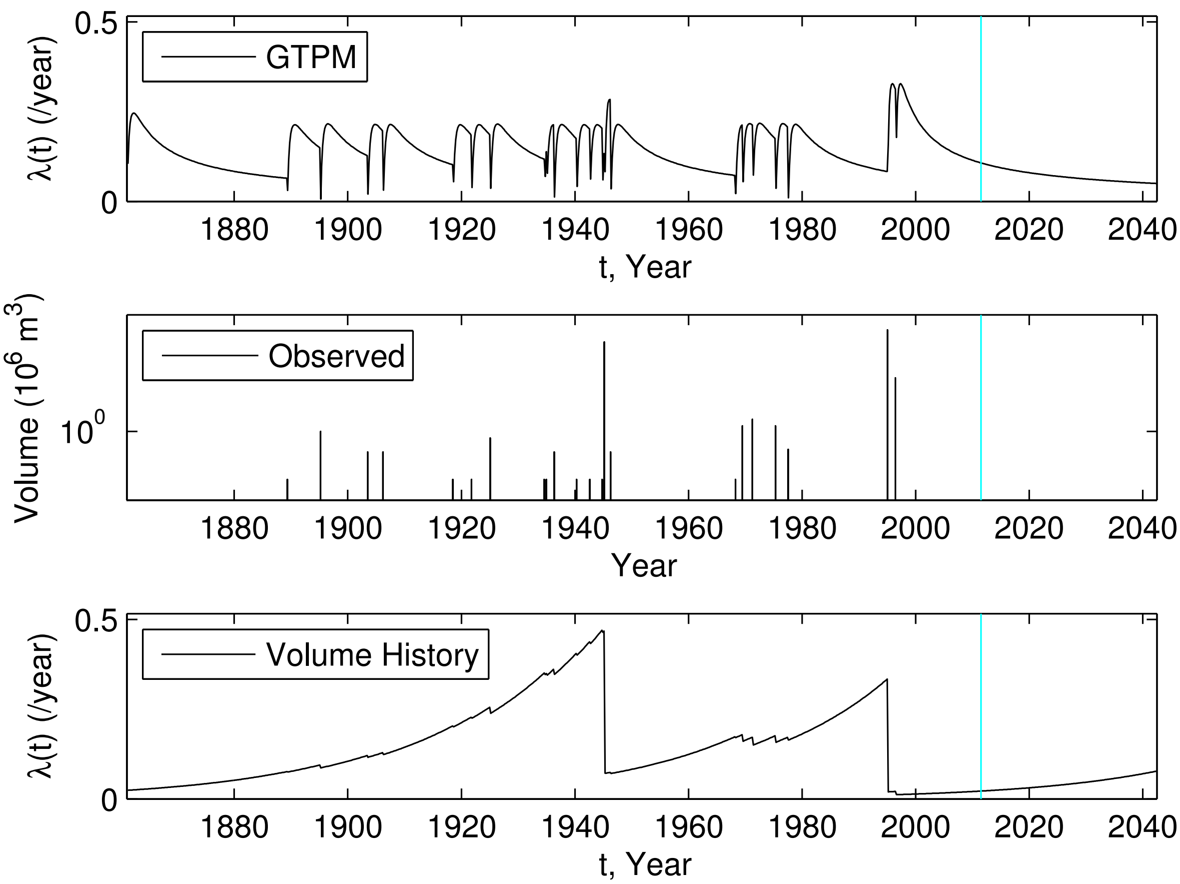

Figure 8: Intensities for the GTPM and volume-history models fitted to the VEI > 1 eruption record of Mt Ruapehu. The observed eruption sizes are shown in the middle panel, and the present day is indicated by the vertical rule. $\lambda(t)$ is the hazard at time t.

transformed data has increased to -0.3429, the P-value of 0.1281 is not significant. The fitted and forecast intensities for the GTPM and volume history models (see `plot_gtpm_volhist_intensity.m`) are shown in Figure 8. We see that both models have lowered intensities after eruptions, the effect persisting longer for larger eruptions. However, the GTPM has an intensity that can begin to decrease some time after an eruption, which is a consequence of the lognormal density used. The volume history model has an intensity that increases monotonically between eruptions. Both cases are supported by records of many volcanoes (*Bebbington & Marzocchi*, 2011).

Proportional hazard models

Given a vector of covariates $c(t)$, a proportional hazard model is obtained from any model with an intensity $\lambda(t)$ by

$$\lambda_{\mathrm{PH}}(t) = \lambda(t) \exp\left(\delta^{\mathrm{T}} c(t)\right), \tag{15}$$

where the vector δ contains the additional parameters to be fitted. For illustrative purposes, we will use the Weibull renewal model (6) as our baseline model, and the volume of the previous eruption as the covariate. This is another way of formulating a time-predictable model. Fitting this model to the Mt Ruapehu data (see `weibull_ph_test.m`) we obtain MLEs of $\hat{\alpha} = 0.8888, \hat{\beta} = 0.1554, \hat{\delta} = -0.0015$, for a $\log L = -64.01$ and AIC $= -134.01$. The fit is satisfactory according to Figure 9 and the P-value of 0.8139 for the correlation. However the $\log L$ is barely better than that of the baseline Weibull renewal model, indicating that the previous eruption volume contains little information about the next repose in formulation (15).

Notes and further reading

Burt et al. (1994) proposed that a volcano could be tested for time-predictability by performing a regression analysis of $\{r_i\}$ on $\{v_i\}$. *Sandri et al.* (2005) generalized the test for time-predictability to a regression analysis of $\{\log r_i\}$ on $\{\log v_i\}$, so that an estimated slope of b significantly different to zero implies a time-predictable relation $r_i \propto v_i^b$. Because both the repose and volume distributions are highly skewed, the logarithmic transformation is also advisable to reduce the high leverage of the tail points.

Bebbington & Marzocchi (2011) fitted a variety of proportional hazard models, using elapsed time, distance and magnitude of earthquakes as covariates, in the Poisson process, Weibull renewal and volume history models. Applied to earthquake and eruption data from Indonesia 1900-2010, only after the entire volume history of the volcano was accounted for using the volume history model (13) was a significant triggering effect observed.

Non-stationary models

We will now move away from models where the past and future are statistically the same. While this allows for trends to be examined, it does open the question of extrapolation (at both ends) of the record: As time is going to be relative to some fixed origin, how it is to be chosen, and does the choice affect the estimated behavior? Secondly, as we want to forecast future behavior, is that behavior going to be physically reasonable?

Nonhomogeneous Poisson process

The Poisson process is stationary in time, in that the distribution of the number of events in an interval depends only on the length of the interval, not its location. In other words, events occur at the same average rate at all times, *i.e.*, there is no trend in the occurrence rate with time. More generally, a process can be a

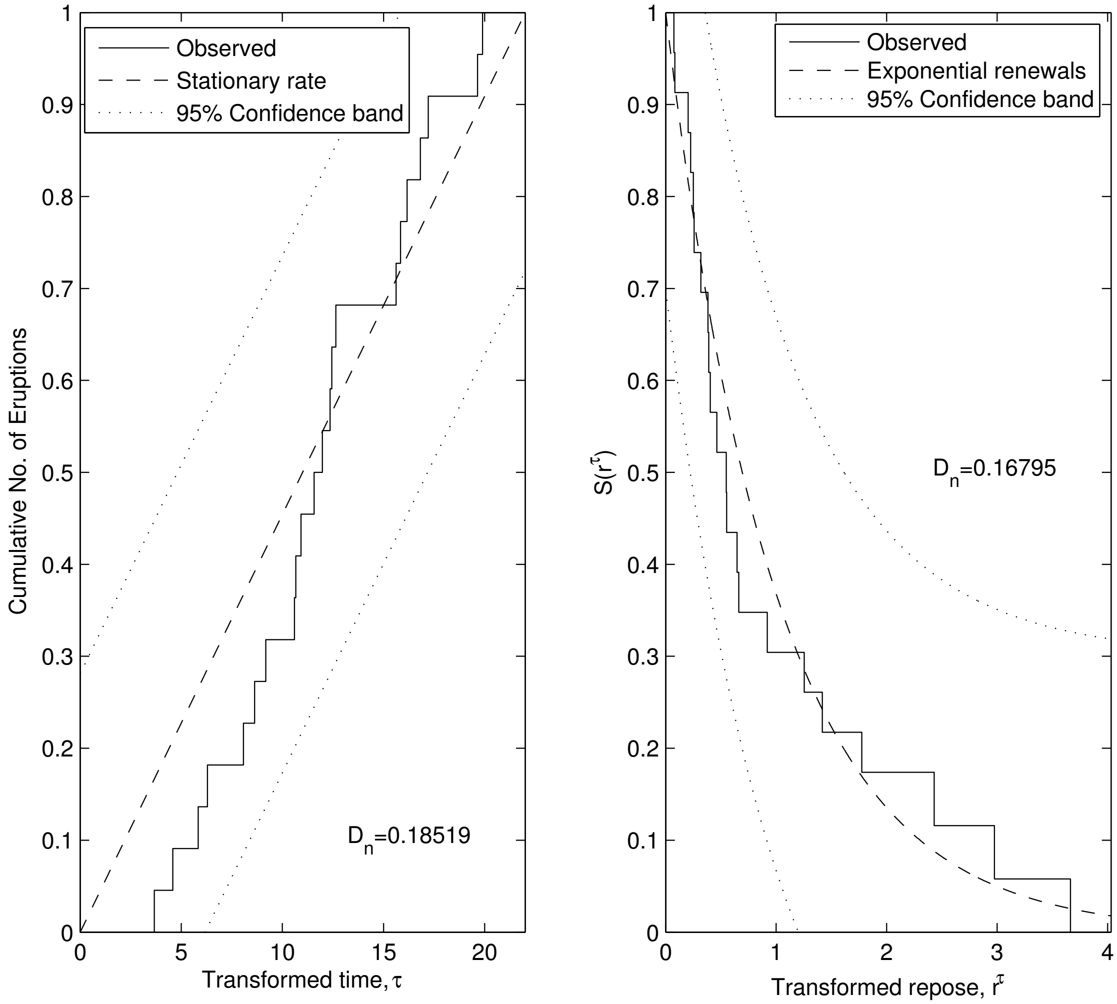

Figure 9: Transformed onsets (left) and intervals (right) for the Weibull proportional hazards model fitted to the VEI > 1 eruption record of Mt Ruapehu. D_n is the Kolmogorov-Smirnov test statistic. $S(r^\tau)$ is the probability that the transformed repose is longer than r^τ.

nonhomogeneous Poisson process, where $N(t)$ has a Poisson distribution with mean $\mu(t) = \int_0^t \lambda(s)ds$. *Ho* (1991) used an example of a nonhomogeneous Poisson process known as the *Weibull Process*, with

$$\lambda(t) = \frac{\kappa}{\theta}\left(\frac{t}{\theta}\right)^{\kappa-1}. \tag{16}$$

This includes the homogeneous Poisson process as a special ($\kappa = 1$) case, while if $\kappa \neq 1$ the process is non-stationary. The intensity (16) is monotonic, and hence can model either an increase or decrease in volcanic activity, but not both.

The parameters can be estimated by maximizing the point process log likelihood (14). Doing so for the Mt Ruapehu data (see `weibull_proc_test.m`) we obtain MLEs of $\hat{\kappa} = 1.4476, \hat{\theta} = 17.7767$, for a $\log L = -62.95$ and AIC $= -129.90$. The fit is satisfactory according to Figure 10 and the P-value of 0.5516 for the correlation. However the AIC is not significantly better than that of the homogeneous Poisson process, indicating that there is no statistically significant trend in the data, in line with Figure 1. In fact, the starting point ($t = 0$) in the Weibull process has something of the character of an additional parameter, and so the AIC is likely to be an overestimate of the degree of fit. Note that the MLE $\hat{\kappa}$ is between 1 and 2, corresponding to an increasing (concave downwards) intensity.

Notes and further reading

The Weibull process can be tested for goodness of fit to a series of data (*Bebbington & Lai*, 1996*a*) either via a standard χ^2 test, or by using the fact that $x_1, \ldots, x_{n-1}$, where $x_i = \ln(t_n/t_{n-i})$, should be a random sample from an exponential distribution of unknown mean.

The MLEs of the Weibull process can be written down explicitly:

$$\hat{\kappa} = \frac{n}{\sum_{i=1}^n \ln\left(\frac{T-S}{t_i-S}\right)} \quad \text{and} \quad \hat{\theta} = \frac{T-S}{n^{1/\hat{\kappa}}}.$$

However, the MLEs are sensitive to the position of the time origin S (*Bebbington & Lai*, 1996*a*). For example, setting $S = 1889$ in `weibull_proc_test.m` results in a MLE of $\hat{\kappa} = 0.8427$. This is now *decreasing* over time, which is a major change from the previously increasing activity. Furthermore, estimates of $\hat{\kappa} > 2$, which are quite feasible, indicate a constantly accelerating, or convex downwards, intensity. Neither of these properties is particularly desirable from a physical viewpoint, and hence the Weibull process is not suitable to model entire volcanic histories. The Weibull process has been used, with the Poisson process, to provide a piecewise intensity model for Etna flank eruptions (*Salvi et al.*, 2006).

Smethurst et al. (2009) fitted both piecewise constant (which can be seen as a stationary regime model) and piecewise linear intensities to the onset of flank eruptions from Mt Etna. Their final piecewise linear model is a constant level of 0.11/year until 1964, after which the intensity rises at 0.016/year per year. Recalling the definition (*Siebert & Simkin*, 2002-), that surface quiet of up to 3 months can be part of a continuing eruption, and that the flank eruptions of Mt Etna have a mean duration of over 200 days (*Tanguy et al.*, 2007), the fitted intensity cannot be used for forecasting much beyond a decade into the future.

The function (16) has been used (*Bebbington*, 2010) as a time-scaling in a *trend renewal process (TRP)*. In the TRP, the independence property of the reposes is preserved, which is both statistically and physically attractive. Models for waxing and waning of activity, and for cyclic patterns can also be placed in this framework (*Bebbington*, 2010).

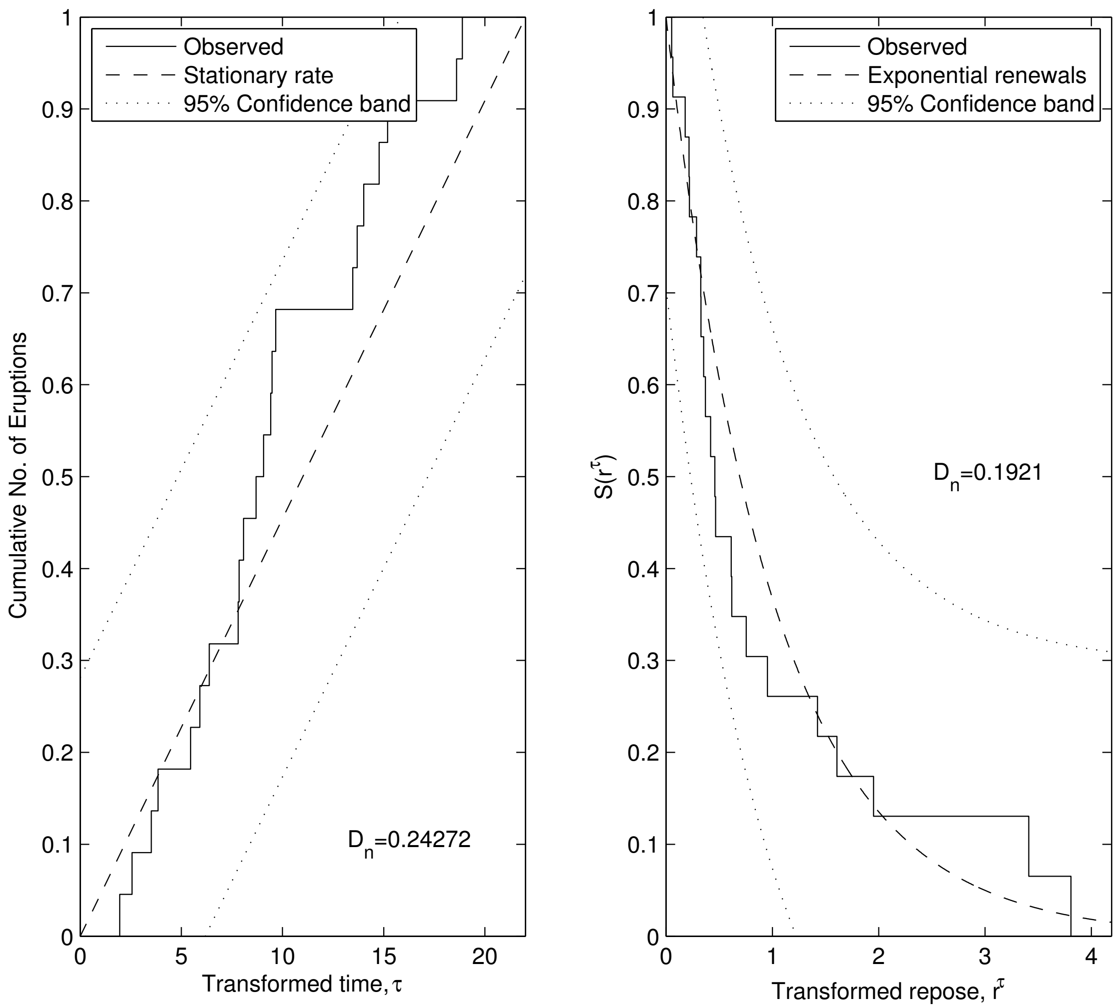

Figure 10: Transformed onsets (left) and intervals (right) for the Weibull process fitted to the VEI > 1 eruption record of Mt Ruapehu. D_n is the Kolmogorov-Smirnov test statistic. $S(r^\tau)$ is the probability that the transformed repose is longer than r^τ.

Conclusion

We have outlined several models for temporal volcanic hazard. The associated Matlab programs provide a means of fitting these models to data using maximum likelihood, and assessing both whether a particular model is consistent with the data, and for identifying the best model from those considered. Stationary models offer the advantage that future forecasts are automatically consistent with past activity, and are hence physically consistent by definition. This does not necessarily apply to non-stationary models, which must thus be used with care. If it is necessary to model differing levels of activity, stationary models incorporating regimes are a safe and physically appealing (*Klein*, 1982; *Mulargia et al.*, 1987; *Bebbington*, 2007) choice; future regimes can be estimated via hidden Markov model techniques (*Bebbington*, 2007). For the Mt Ruapehu (VEI > 1) record considered as an example, the volume-history model appears to be by far the best explanation of those considered. While the non-stationary Weibull process appears statistically to be as reasonable an explanation as any of the renewal processes, it does leave open the question of why the activity of the volcano should be tied to the year AD1861. It should hence be viewed with a degree of scepticism unless there is reason to believe, for example, that the volcano is reawakening from a dormant state (*Wadge*, 1982), or entering an open conduit phase (*Marzocchi & Zaccarelli*, 2006).

References

Akaike, H. (1977) On entropy maximization principle. *In* P. Krishnaiah (ed.) *Applications of Statistics*, 27–41. North-Holland, Amsterdam. 6

Bebbington, M.S. (2007) Identifying volcanic regimes using hidden Markov models. *Geophysical Journal International*, **171**:921–942. URL http://dx.doi.org/10.1111/j.1365-246X.2007.03559.x. 10, 19

Bebbington, M. (2008) Incorporating the eruptive history in a stochastic model for volcanic eruptions. *Journal of Volcanology and Geothermal Research*, **175**:325–333. URL http://dx.doi.org/10.1016/j.jvolgeores.2008.03.013. 3, 11

Bebbington, M. (2010) Trends and clustering in the onsets of volcanic eruptions. *Journal of Geophysical Research*, **115**:B01203. URL http://dx.doi.org/10.1029/2009JB006581. 17

Bebbington, M. & Cronin, S.J. (2011) Spatio-temporal hazard estimation in the Auckland volcanic field, New Zealand, with a new event-order model. *Bulletin of Volcanology*, **73**:55–72. URL http://dx.doi.org/10.1007/s00445-010-0403-6. 10

Bebbington, M. & Harte, D. (2001) On the statistics of the linked stress release process. *Journal of Applied Probability*, **38A**:176–187. URL http://dx.doi.org/10.1239/jap/1085496600. 11

Bebbington, M. & Lai, C.D. (1996*a*) On nonhomogeneous models for volcanic eruptions. *Mathematical Geology*, **28**:585–600. URL http://dx.doi.org/10.1007/BF02066102. 5, 9, 17

Bebbington, M. & Lai, C.D. (1996*b*) Statistical analysis of New Zealand volcanic occurrence data. *Journal of Volcanology and Geothermal Research*, **74**:101–110. URL http://dx.doi.org/10.1016/S0377-0273(96)00050-9. 9

Bebbington, M. & Marzocchi, W. (2011) Stochastic models for earthquake triggering of volcanic eruptions. *Journal of Geophysical Research*, **116**:B05204. URL http://dx.doi.org/10.1029/2010JB008114. 15

Burt, M.L., Wadge, G. & Scott, W.A. (1994) Simple stochastic modelling of the eruption history of a basaltic volcano: Nyamuragira, Zaire. *Bulletin of Volcanology*, **56**:87–97. URL http://dx.doi.org/10.1007/BF00304104. 15

Claeskens, G. & Hjort, N.L. (2008) *Model Selection and Model Averaging.* Cambridge University Press, Cambridge. URL http://dx.doi.org/10.1017/CBO9780511790485. 6

Connor, C.B., Sparks, R.S.J., Mason, R.M., Bonadonna, C. & Young, S.R. (2003) Exploring links between physical and probabilistic models of volcanic eruptions: The Soufriere Hills volcano, Montserrat. *Geophysical Research Letters*, **30**:1701. URL http://dx.doi.org/10.1029/2003GL017384. 9

Cronin, S., Bebbington, M. & Lai, C.D. (2001) A probabilistic assessment of eruption recurrence on Taveuni volcano, Fiji. *Bulletin of Volcanology*, **63**:274–288. URL http://dx.doi.org/10.1007/s004450100144. 10

Daley, D.J. & Vere-Jones, D. (2003) *An Introduction to the Theory of Point Processes*, volume 1. Springer, New York, 2nd edition. 11

De la Cruz-Reyna, S. (1991) Poisson-distributed patterns of explosive eruptive activity. *Bulletin of Volcanology*, **54**:57–67. URL http://dx.doi.org/10.1007/BF00278206. 10

De la Cruz-Reyna, S. & Carrasco-Nunez, G. (2002) Probabilistic hazard analysis of Citlaltepetl (Pico de Orizaba) volcano, eastern Mexican volcanic belt. *Journal of Volcanology and Geothermal Research*, **113**:307–318. URL http://dx.doi.org/10.1016/S0377-0273(01)00263-3. 9

Decker, R.W. (1986) Forecasting volcanic eruptions. *Annual Review of Earth and Planetary Sciences*, **14**:267–291. URL http://dx.doi.org/10.1146/annurev.ea.14.050186.001411. 2

Ho, C.H. (1991) Nonhomogeneous Poisson model for volcanic eruptions. *Mathematical Geology*, **23**:167–173. URL http://dx.doi.org/10.1007/BF02066293. 17

Ho, C.H. (1992) Statistical control chart for regime identification in volcanic time-series. *Mathematical Geology*, **24**:775–787. URL http://dx.doi.org/10.1007/BF00890701. 4

Hurvich, C.M. & Tsai, C.L. (1989) Regression and time series model selection in small samples. *Biometrika*, **76**:297–307. URL http://dx.doi.org/10.1093/biomet/76.2.297. 6

Jiang, R. & Murthy, D.N.P. (1998) Mixture of Weibull distributions - parametric characterization of failure rate function. *Applied Stochastic Models and Data Analysis*, **14**:47–65. URL http://dx.doi.org/10.1002/(SICI)1099-0747(199803)14:1⟨47::AID-ASM306⟩3.0.CO;2-E. 8

Klein, F.W. (1982) Patterns of historical eruptions at Hawaiian volcanoes. *Journal of Volcanology and Geothermal Research*, **12**:1–35. URL http://dx.doi.org/10.1016/0377-0273(82)90002-6. 4, 9, 19

Latter, J.H. (1985) Frequency of eruptions at New Zealand volcanoes. *Bulletin of the New Zealand Society for Earthquake Engineering*, **18**:55–101. 3

Lawless, J.F. (2003) *Statistical Models and Methods for Lifetime Data.* Wiley, Hoboken, 2nd edition. 6

Marzocchi, W. & Bebbington, M. (2012) Probabilistic eruption forecasting at short and long time scales. *Bulletin of Volcanology*, **74**:1777–1805. URL http://dx.doi.org/10.1007/s00445-012-0633-x. 2

Marzocchi, W., Sandri, L. & Selva, J. (2008) BET_EF: A probabilistic tool for long- and short-term forecasting. *Bulletin of Volcanology*, **70**:623–632. URL http://dx.doi.org/10.1007/s00445-007-0157-y. 2

Marzocchi, W. & Zaccarelli, L. (2006) A quantitative model for the time-size distribution of eruptions. *Journal of Geophysical Research*, **111**:B04204. URL http://dx.doi.org/10.1029/2005JB003709. 3, 10, 19

Mendoza-Rosas, A.T. & De la Cruz-Reyna, S. (2009) A mixture of exponentials distribution for simple and precise assessment of the volcanic hazard. *Natural Hazards and Earth System Sciences*, **9**:425–431. URL http://dx.doi.org/10.5194/nhess-9-425-2009. 8, 9

Mulargia, F., Gasperini, P. & Tinti, S. (1987) Identifying different regimes in eruptive activity: An application to Etna volcano. *Journal of Volcanology and Geothermal Research*, **34**:89–106. URL http://dx.doi.org/10.1016/0377-0273(87)90095-3. 4, 19

Newhall, C.G. & Self, S. (1982) The volcanic explosivity index (VEI): An estimate of explosive magnitude for historical volcanism. *Journal of Geophysical Research*, **87**:1231–1238. URL http://dx.doi.org/10.1029/JC087iC02p01231. 2

Pyle, D.M. (1998) Forecasting sizes and repose times of future extreme volcanic events. *Geology*, **26**:367–370. URL http://dx.doi.org/10.1130/0091-7613(1998)026⟨0367:FSARTO⟩2.3.CO;2. 4, 9

Salvi, F., Scandone, R. & Palma, C. (2006) Statistical analysis of the historical activity of Mount Etna, aimed at the evaluation of volcanic hazard. *Journal of Volcanology and Geothermal Research*, **154**:159–168. URL http://dx.doi.org/10.1016/j.jvolgeores.2006.01.002. 17

Sandri, L., Marzocchi, W. & Gasperini, P. (2005) Some insights on the occurrence of recent volcanic eruptions of Mount Etna volcano (Sicily, Italy). *Geophysical Journal International*, **163**:1203–1218. URL http://dx.doi.org/10.1111/j.1365-246X.2005.02757.x. 15

Siebert, L. & Simkin, T. (2002-) Volcanoes of the world: an illustrated catalog of Holocene volcanoes and their eruptions. Smithsonian Institution, Global Volcanism Program Digital Information Series, GVP-3. URL http://www.volcano.si.edu/world/. 2, 3, 17

Smethurst, L., James, M.R., Pinkerton, H. & Tawn, J.A. (2009) A statistical analysis of eruptive activity on Mount Etna, Sicily. *Geophysical Journal International*, **179**:655–666. URL http://dx.doi.org/10.1111/j.1365-246X.2009.04286.x. 17

Tanguy, J.C., Condomines, M., Le Goff, M., Chillemi, V., Le Delfa, S. & Patane, G. (2007) Mount Etna eruptions of the last 2750 years: revised chronology and location through archeomagnetic and ^{226}Ra$-^{230}$Th dating. *Bulletin of Volcanology*, **70**:55–83. URL http://dx.doi.org/10.1007/s00445-007-0121-x. 17

Turner, M., Cronin, S., Bebbington, M. & Platz, T. (2008) Developing a probabilistic eruption forecast for dormant volcanos; a case study from Mt Taranaki, New Zealand. *Bulletin of Volcanology*, **70**:507–515. URL http://dx.doi.org/10.1007/s00445-007-0151-4. 6

Wadge, G. (1982) Steady state volcanism: Evidence from eruption histories of polygenetic volcanoes. *Journal of Geophysical Research*, **87**:4035–4049. URL http://dx.doi.org/10.1029/JB087iB05p04035. 3, 19

Zwillinger, D. & Kokoska, S. (2000) *Standard Probability and Statistics Tables and Formulae*. Chapman and Hall, Boca Raton. 4

Additional Files

The following files are Matlab macros (codes) used to fit the models and produce the output described in this paper. The models are fitted by numerically maximizing the log-likelihood, a simplex search method that is repeated using many (*convno* = 100) random starting points, with the best solution recorded. All solutions correspond to *local* maxima of the log-likelihood, but it is impossible to guarantee that the best solution found is actually the *global* maximum. Mixture models may have the component labels swapped.

The mixture models (`exponentialmix_test.m` and `weibullmix_test.m`) should not be used when the current incomplete repose is longer than any of the other reposes. In this case the numerical procedure will find a solution, but one in which the incomplete repose will be one component of the mixture. As the datum is only a lower bound, the resulting parameter values for that component will be arbitrary.

How to run the macros

1. Create a data file using the format found within the file, `ruapehu_year_volume.txt`.
2. Edit the file, `init_volcano.txt`. Replace `ruapehu_year_volume.txt` with the name of the data file being used, and replace 2011.5 with the present time.
3. (Optional.) In a given macro, change the number of random starting points. The models in this paper use *convno* = 100.
4. Run the macros by typing their name in the Matlab command window.

Code descriptions

The following macro reads in the data and initializes system values.

`init_volcano.m`

Inputs: The present time, plus a text file containing onset times (column 1) and erupted volumes (column 2).
Outputs: None. This macro is called by the `*_test.m` macros.

The following macro tests the data to see if it is consistent with a steady-state process, and if successive repose lengths are independent.

`stationarity_and_independence_test.m`

Inputs: None. The macro calls `init_volcano.m` to read in and set up the data.
Outputs: A plot of cumulative eruptions over time, a window showing the maximum deviation that is not significantly different from steady-state, and a value for Spearman's ρ including the corresponding P-value. A low P-value is evidence of dependence among successive repose lengths.

The following macros fit the specified model to the data, and check whether the model is consistent with the data.

```
poisson_process_test.m
weibull_renewal_test.m
weibullmix_renewal_test.m
exponentialmix_renewal_test.m
gtpm_test.m
weibull_ph_test.m
volhist_test.m
weibull_proc_test.m
```

Inputs: None. The macros call `init_volcano.m` to read in and set up the data.
Outputs: Parameter estimates, log-likelihood and AIC values, a plot showing whether the fitted survival function differs significantly from the Kaplan-Meier product-limit estimate, including the outputs listed under, `stationarity_and_independence_test.m` (above).

The following macro produces Figures 3, 4 and 5.

```
plot_renewal_models.m
```

Inputs: None. The macro calls `init_volcano.m` to read in and set up the data.
Outputs: Figures showing *a)* a histogram of the observed repose lengths and the fitted renewal model densities, *b)* a plot of the fitted renewal hazards, showing how they vary with elapsed time since the last onset, and *c)* the estimated time from the present to the next onset, given as the probability that the next eruption occurs before a given date.

The following macro produces Figure 8.

```
plot_gtpm_volhist_intensity.m
```

Inputs: None. The macro calls `init_volcano.m` to read in and set up the data.
Outputs: A plot showing the observed onset times and erupted volumes, plus the fitted hazards from the GTPM and volume-history models.

Function files

Function files contain bits of code used by the macros listed above. They should not be edited, with the possible exception of editing the values given as initial points in the `fit_*.m` files.

The following functions numerically estimate model parameters by maximizing their log-likelihood.

```
fit_weibull_renewal.m
fit_weibullmix_renewal.m
fit_exponentialmix_renewal.m
fit_gtpm.m
fit_weibull_ph.m
fit_volhist.m
fit_weibull_proc.m
```

The following functions calculate the stationarity limits and the correlation between successive (transformed) reposes.

```
test_stationarity.m
test_correlation.m
```

The following function calculates the value of the fitted volume-history intensity, for plotting purposes only.

```
volhist_lambda.m
```

The following functions calculate the Kaplan-Meier product-limit estimate of the survival function, and the Kolmorgorov-Smirnov limits for various tests of distribution.

```
PLestimate.m
kolmogorov_smirnov2.m
```

Disclaimer: The journal and author make no assertions that these program macros are without errors. Users do so at their own risk. The macros may be used without payment or permission provided the source paper is cited.