Bayesian Hierarchical Time Predictable Model for eruption occurrence: an application to Kilauea Volcano

Luigi Passarelli,1 Laura Sandri,1 Alessandro Bonazzi1 and Warner Marzocchi2

1Istituto Nazionale di Geofisica e Vulcanologia, Bologna, Italy. E-mail: passarelli@bo.ingv.it
2Istituto Nazionale di Geofisica e Vulcanologia, Roma, Italy

Accepted 2010 March 1. Received 2010 February 9; in original form 2009 May 22

SUMMARY

The physical processes responsible for volcanic eruptions are characterized by a large number of degrees of freedom, often non-linearly coupled. This extreme complexity leads to an intrinsic deterministic unpredictability of such events that can be satisfactorily described by a stochastic process. Here, we address the long-term eruption forecasting of open conduit volcanoes through a Bayesian Hierarchical Modelling information in the catalogue of past eruptions, such as the time of occurrence, the duration, and the erupted volumes. The aim of the model is twofold: (1) to get new insight about the physics of the process, using the model to test some basic physical hypotheses of the eruptive process and (2) to build a stochastic model for long-term eruption forecasting; this is the basic component of Probabilistic Volcanic Hazard Assessment that is used for rational land use planning and to design Emergency plan. We apply the model to Kilauea eruption occurrences and check its feasibility to be included in Probabilistic Volcanic Hazard Assessment.

Key words: Probabilistic forecasting; Effusive volcanism.

1 INTRODUCTION

The extreme complexity, non-linearity, limited knowledge, and the large number of degrees of freedom of a volcanic system make deterministic prediction of the evolution of volcanic processes impossible. Volcanic systems are intrinsically stochastic (e.g. Marzocchi 1996; Sparks 2003), and hazardous volcanic phenomena involve so many uncertainties that a probabilistic approach is practically always needed (e.g. Newhall & Hoblitt 2002; Sparks 2003; Marzocchi et al. 2004).

In general, eruption forecasting can be tackled in two different ways, related to two different timescales: (i) a short-term forecasting, mostly based on monitoring measures observed during an episode of unrest (e.g. Marzocchi et al. 2008); (ii) a long-term forecasting, usually made during a quiet period of the volcano, and mostly related to a statistical description of the past eruptive catalogues. Here, we focus our attention only on this second issue.

In a recent paper, Marzocchi & Zaccarelli (2006) found different behaviour for volcanoes with ‘open’ conduit regime (i.e. volcanoes with high frequency of eruption and repose periods less than few tens of years) compared to those with ‘closed’ conduit regime (i.e. volcanoes with periods of quiescence longer than 30-40 yr). According to that paper, open conduit volcanoes tested there (i.e. Mt Etna, Kilauea volcano) seem to follow a so-called ‘Time Predictable Model’ (TPM), that is, a model where the time to the next eruption depends on the size of the last eruption (De la Cruz-Reyna 1991; Burt et al. 1994). Closed conduit volcanoes, tested by Marzocchi & Zaccarelli (2006), conversely seem to follow mostly a Poisson distribution. These results have been used to build general probabilistic models for volcanic hazard assessment of open and closed conduit systems.

Different methods have been presented in the past years aiming at the identification of possible recurrence or correlation in the volcanic time and/or volume data. Klein (1982) and Bebbington & Lai (1996b) studied the changes in volcanic regimes looking at the mean rate of occurrence of the volcanic events. Sandri et al. (2005) applied a generalized form of time predictable model to Mt Etna eruptions. De la Cruz-Reyna (1991) proposed a load-and-discharge model for eruptions in which the time predictable model could be seen as a particular case. Bebbington (2008) presented a stochastic version of the general load-and-discharge model also including a way to take into account of the history of the volcano discharging behaviour. In this paper, the author studied the time predictability as a particular case of his model with application to Mt Etna and Mauna Loa data series. Finally a different hierarchical approach have been presented by Bebbington (2007) using Hidden Markov Model to study eruption occurrences with application to Mount Etna flank eruptions. This model is able to find any possible underlying volcano activity resulting in volcanic regime changes.

Here, our goal is to improve significantly the modelling of open conduit systems through the implementation of a Bayesian Hierarchical Time Predictable Model (hereafter BH_TPM) for eruption occurrence. The model is a formal generalization of the TPM in a full Bayesian framework. The Bayesian perspective allows accounting for stochastic fluctuations in each parameter of the model and in each recorded measurement (Wikle 2003). In this way, each
2 BAYESIAN HIERARCHICAL MODEL

parameter of the model is described through a probability density function whose posterior distribution is conditioned by the available data. The numerical solution is obtained via MCMC-Gibbs sampling (Gelman et al. 2000). The BH-TPM is then applied to the eruption record of Kilauea Volcano since 1923 published by the Hawaiian Volcano Observatory (see Table 1). The outcomes for model variables and parameters show good convergence properties for all model parameters and errors.

After describing the model in detail, we focus our attention on some specific issues: (1) to discuss the volcanological implications of the model parameters obtained; (2) to verify if the model describes the data satisfactorily and (3) to compare the forecasting capability of BH-TPM with other models in the literature; that is, if X, Y and Z are random variables, we can write the joint distribution in terms of a factorization such as \( p(X, Y, Z) = p(Y|X)p(Z|Y,X)p(X) \). We make use of the bracket notation for probability distribution in which \([Y] \) refers to the distribution of \( Y \) and \( [Y|X] \) refers to the conditional distribution of \( Y \) given \( X \). This simple formula is the basic idea of hierarchical thinking. In general it is easier to specify the distribution of the relevant conditional models than to work with marginal distributions of variables involved in such models. In this case, the product of a series of relatively simple conditional models leads to a joint distribution that can be quite complicated.

In order to build the model, we follow the framework outlined by Wikle (2003; see also references therein). The idea is to approach the problem by breaking it into three primary stages:

(i) Data model: \([data|process, parameters] \)

(ii) Process model: \([process|parameters] \)

(iii) Parameter model: \([parameters] \)

The first stage regards the observational process or data model, which specifies the distribution of the observed data given the process generating them and the parameters describing it. The second stage describes the process, conditional on its parameters. Finally, the third stage accounts for the uncertainty in the parameters. Ultimately, we are interested in the distribution of the process and parameters updated by the data. We obtain the joint posterior distribution for the process and parameters using Bayes’ rule:

\[
[process, parameters|data] \propto [data|process, parameters] \times [process|parameters] \times [parameters].
\]

In order to make inference about the process and parameters governing the occurrence of volcanic eruptions for the case of ‘open’ conduit volcano, we apply this simple approach. In the next subsections we will illustrate each stage that we have performed for our hierarchical model.

2.1 Data model

The data set reported in Table 1 is taken from the Hawaiian Volcano Observatory web site (http://hvo.wr.usgs.gov/kilauea/history/historytable.html). The full catalogue starts from 1823 but only the 42 volcanic events having occurred after 1922 are considered in our analysis, because only this latter part of the catalogue can be considered complete in terms of occurrence time and erupted volume data. Fig. 1 reports the cumulative number of events versus time, where the eruptive rate since 1923 is approximately constant except for a major quiescence period around the 1940s. The catalogue reports the onset of each eruption, the total volume of material ejected.
(lava and tephra) and the interevent time. The volume of the 1924 May 10 event is taken from http://www.volcano.si.edu/ and is only the tephra volume. For more details regarding the definition of interevent times see Klein (1982). Since the interevent time following the last eruption cannot be available, we have 41 pairs of data of interevent time (i.e. the time between the onset of nth and the onset of (n + 1)th eruptions) and volume erupted (in the nth eruption), that from now on we indicate with \( d_i \) and \( v_i \), respectively.

In testing the independence of data via correlation function, the only significant correlation (\( r = 0.06 \)) appears between the volume and the subsequent interevent time. Therefore, we assume that each pair of data \( (d_i, v_i) \) is independent from the other pairs. In a Bayesian framework, the act of measurement does not lead simply to an observed value, but to a state of information described by a distribution where the single measurement is a random realization of this distribution.

In this paper, we assume that the logarithm of the data, made dimensionless by two gauge constant \( R = 1 \) d and \( V = 1 \times 10^9 \) m\(^3\), that is, \( D_i = \ln(d_i/R) \) and \( D_v = \ln(d_v/V) \), are respectively random draws from normal distributions, with the means \( R_i = \ln(r_i/R) \) for the interevent times and \( V_i = \ln(v_i/V) \) for the volumes, where \( r_i \)'s are the interevent time variables and \( v_i \)'s are the volume variables. We test whether or not \( D_0 \) and \( D_v \) are normally distributed using Anderson-Darling test (Anderson & Darling 1952). The null hypothesis is that lognormal of the data comes from a normal distribution. We cannot reject the null hypothesis of normality for \( D_0 \) and \( D_v \) with \( P\)-value = 0.625 and 0.715, respectively (Trujillo-Ortiz et al. 2007). We can conclude that normal distributions fit reasonably the logarithm of the data.

The variables \( r_i \) and \( v_i \), and their natural logarithm, represent the variables of our model. The variances of such normal distributions are the data measurement errors for the interevent times \( \sigma_{D_i}^2 \) and for the volumes \( \sigma_{D_v}^2 \). In this view, each single pair is

\[
D_i \sim N(R_i, \sigma_{D_i}^2) \quad \text{and} \quad D_v \sim N(V_i, \sigma_{D_v}^2) \quad i = 1, \ldots, 41,
\]

where from now on \( N(a, b^2) \) indicate a normal distribution with mean \( a \) and variance \( b^2 \) and the symbol \( \sim \) means ‘is distributed as’.

In order to give appropriate variances for \( R_i \) and \( V_i \) to each distribution, we use the error propagation. We assume two different values for measurement errors on volume data before and after 1960. Such division arises by considering that, after 1960, the measurements were taken by the Hawaiian Volcano Observatory, and we assume that these measurements are more accurate. Systematic and direct measurement of lava flow or modern measurement using satellite techniques should give a more precise estimation of the volume erupted. Indirect measurement on historical lava flow, inferred with geological field methodology probably underestimates the real erupted volume (e.g., Behncke et al. 2005). This is the reason why we assume the relative error \( \Delta v_i/v_i \) equal to 25 per cent for the volumes before 1960 and equal to 15 per cent for more recent data. For the interevent times we choose an error measurement equal to \( \Delta r_i = 1 \) d. Therefore, applying the error propagation rule, we get

\[
\sigma_{D_r_i} = \frac{\partial R_i}{\partial r_i} \Delta r_i = \frac{\Delta r_i}{r_i} i = 1, \ldots, 41
\]

\[
\sigma_{D_v_i} = \frac{\partial V_i}{\partial v_i} \Delta v_i = \frac{\Delta v_i}{v_i} = 0.25 \quad i = 1, \ldots, 13
\]

\[
\sigma_{D_v_j} = \frac{\partial V_j}{\partial v_j} \Delta v_j = \frac{\Delta v_j}{v_j} = 0.15 \quad j = 14, \ldots, 41.
\]

The error \( \sigma_{D_r_i} \) is coincident with the relative error on the interevent time, while \( \sigma_{D_v_i} \) is independent from the data value and error.

At this point we are able to write the joint distributions for the data model, assuming independence among the pairs of data, as

\[
[D_i|R_i, \sigma_{D_r_i}^2] = \prod_{i=1}^{41} N(R_i, \sigma_{D_r_i}^2)
\] (2)
\[
\left[ D_i | V_i, \sigma_{D_i}^2 \right] = \prod_{i=1}^{41} N \left( V_i, \sigma_{D_i}^2 \right).
\]  

(3)

2.2 Process model

Before displaying our statistical considerations for the process model, we have to introduce the underlying physical eruptive process. We use a very simple stochastic process to explain the eruptive dynamic process. It is the Generalized Time-Predictable Model (GTPM, see Sandri et al. 2005) for volcanic eruptions, assuming that eruptions occur when the volume in the storage system reaches a threshold value, given that magma enters in the magma storage system with a variable rate and that the size of eruptions is a random variable, following some kind of statistical distribution. Under these assumptions, we have a generalized time-predictable system with longer/shorter interevent time following large/small volume output eruptions. In fact, for such a model, the time to the next eruption is determined by the time required for the magma entering the storage system to reach the eruptive level. In this view the more general form for a time-predictable model is a power law between the erupted volume and the interevent time

\[ r_i = cv_i^\theta \]  

(4)

that we want to linearize by logarithmic transformation. For this reason we need dimensionless variables and so we introduce two gauge constants (i.e. \( \hat{R} \) and \( \hat{V} \) that are the same of previous section) in order to make \( r_i \) and \( v_i \) dimensionless. Therefore, we choose \( \hat{R} = 1 \mathrm{d} \) and \( \hat{V} = 1 \times 10^6 \mathrm{m}^3 \) and we define

\[ r_i^* = \frac{r_i}{\hat{R}} \quad \text{and} \quad v_i^* = \frac{v_i}{\hat{V}} \]

that we introduce in the previous equation and we obtain

\[ r_i^* = \alpha v_i^{\theta b}, \]

where \( \alpha = (c\hat{V}^b/\hat{R}) \) is a new constant. Now we can take the logarithm of this equation and we have

\[ R_i = K + bV_i, \]

(5)

where \( K = \ln \alpha \) is a constant and \( R_i = \ln r_i^* \) and \( V_i = \ln v_i^* \). This dimensionless transformation does not influence the following numerical solutions, but it is only an algebraic solution to make dimensionless the argument of the logarithms.

In the last equation, if the parameter \( b \) is equal to unity we are in a classical time predictable system (see De la Cruz-Reyna 1991; Burt et al. 1994). If \( b \) is equal to 0 the system is not time predictable. If \( b > 1 \) we have a non-linear relationship implying a longer interevent time after a large volume eruption compared to a classical time predictable system. If \( 0 < b < 1 \) we still have a non-linear relationship but for a big volume eruption it implies a shorter interevent time compared to a classical time predictable system. Assuming this process as a dynamic eruptive behaviour for the volcano, we proceed to show our statistical consideration about this part of the hierarchical model implementation.

In building up the process model, we have to connect the model variables (\( R_i \) and \( V_i \)) with the physical model, that is, with eq. (5). Here, we assume that the \( R_i \)'s are independent and each of them is normally distributed, with mean given by the generalized time-predictable model and unknown variance representing the model error. Hence

\[ R_i \sim N \left( bV_i + K, \sigma_R^2 \right) \quad i = 1, \ldots, 41 \]

and for the all variables \( R \), the resulting joint distribution given the model parameters is

\[ R | V, b, K, \sigma_R^2 \sim \prod_{i=1}^{41} N \left( bV_i + K, \sigma_R^2 \right). \]

(6)

In order to assign the distribution for the volume variables (\( V_i \)) we have to exert a little effort. We do not have information about the real size distribution of Kiluaea eruptions. However, according to the Anderson and Darling test performed in the previous subsection, the set of volume data, that is, \( D_i (i = 1, \ldots, 41) \), is satisfactorily fitted by a log-normal distribution. Because of this goodness-of-fit test, we assume that also the volume variables (i.e. \( V_i \)) in the BHPM have a log-normal distribution. The logarithm of variables, that is, \( V_i \), are therefore normally distributed with unknown mean \( \mu \) and variance \( \sigma_v^2 \), and for each of them we can write:

\[ V_i \sim N \left( \mu_v, \sigma_v^2 \right) \quad i = 1, \ldots, 41 \]

and the joint distribution is

\[ V | \mu_v, \sigma_v^2 \sim \prod_{i=1}^{41} N \left( \mu_v, \sigma_v^2 \right). \]

(7)

In addition, we assume that the parameter \( \mu_v \) has uniform non-informative vague prior distribution. A non-informative prior expresses vague or general information about a variable. Non-informative priors can express objective information (e.g. ‘the variable is positive’) assigning equal probabilities to all possibilities within the defined domain (e.g. for all \( x > 0 \)). The simplest case of non-informative vague prior distribution is the uniform distribution with unlimited domain (e.g. \( -\infty < x < +\infty \)). In this text when we refer to non-informative vague prior distribution, we always use uniform distribution with unlimited domain, see Table 2.

The distribution of variance parameters, that is, \( \sigma_v^2 \) and \( \sigma_R^2 \), are constructed from inverse gamma family, which is the natural conjugate family for the normal distribution (see Gelman et al. 2000). The property of conjugacy is very useful in Bayesian prior to posterior analysis. The conjugacy is formally defined in this way: for a given vector of data \( y = y_1, \ldots, y_s \) and a parameter vector \( \theta = \theta_1, \ldots, \theta_s \), if \( \Phi \) is a class of likelihood \( [y | \theta] \), and \( \Psi \) is a class of prior distribution for \( \theta \), then the class \( \Psi \) is conjugate for \( \Phi \) if \( [\theta | y] \in \Psi \) for all \( [y | \theta] \in \Phi \) and \( \theta \in \Psi \) where data and parameters are linked by Bayes’ theorem, that is, \( [\theta | y] \propto [\theta | y \theta] \). In our case, if we model the prior distribution for variance by an inverse gamma distribution, the likelihood is normal (i.e. eqs 6 and 7), thus the posterior distribution (for the variance) becomes an inverse gamma distribution.

Table 2. Overview of distributions used in BHP.TPM.

<table>
<thead>
<tr>
<th>Data model</th>
<th>Process model</th>
<th>Parameter model</th>
</tr>
</thead>
<tbody>
<tr>
<td>( D_i ) \sim N \left( \mu_{D_i}, \sigma_{D_i}^2 \right)</td>
<td>( R_i \sim N \left( bV_i + K, \sigma_R^2 \right)</td>
<td>b \sim N \left( \mu_b, \sigma_b^2 \right)</td>
</tr>
<tr>
<td>( D_i ) \sim N \left( \mu_{D_i}, \sigma_{D_i}^2 \right)</td>
<td>( V_i \sim N \left( \mu_v, \sigma_v^2 \right)</td>
<td>K \sim N \left( \mu_K, \sigma_K^2 \right)</td>
</tr>
<tr>
<td>( \sigma_{D_i}^2 ) = known</td>
<td>( \mu_v \sim U(\infty, +\infty)^s )</td>
<td>( \mu_k \sim U(\infty, +\infty)^s )</td>
</tr>
<tr>
<td>( \sigma_{D_i}^2 ) = known</td>
<td>( \sigma_{R_i}^2 \sim \Gamma^{-1}(\sigma_{\alpha R}, \sigma_{\alpha R})^s )</td>
<td>( \sigma_{\alpha R} \sim U(0, +\infty)^s )</td>
</tr>
<tr>
<td>( \sigma_{D_i}^2 ) = known</td>
<td>( \sigma_{v}^2 \sim \Gamma^{-1}(\sigma_{\alpha v}, \sigma_{\alpha v})^s )</td>
<td>( \sigma_{\alpha v}^2 \sim U(\infty, +\infty)^s )</td>
</tr>
<tr>
<td>( \sigma_{D_i}^2 ) = known</td>
<td>( \sigma_{k}^2 \sim \Gamma^{-1}(\sigma_{\alpha k}, \sigma_{\alpha k})^s )</td>
<td>( \sigma_{\alpha k}^2 \sim U(\infty, +\infty)^s )</td>
</tr>
</tbody>
</table>

Note: The distributions highlighted with * are prior distributions for the BHP.TPM. We assign for the prior distribution parameters for inverse gamma’s (i.e. \( \Gamma^{-1} \)) initial values equal to 1. U means uniform distribution.
Therefore, the prior distributions for variances are
\[
\sigma_k^2 \sim \Gamma^{-1}(\alpha_{\sigma_k}, \beta_{\sigma_k})
\]
\[
\sigma_i^2 \sim \Gamma^{-1}(\alpha_{\sigma_i}, \beta_{\sigma_i}),
\]
where \(\Gamma^{-1}\) indicates the inverse gamma distribution with mean \(\mu = \beta_{\sigma}/(\alpha_{\sigma} - 1)\) for \(\alpha_{\sigma} > 1\) and variance \(\tau = \beta_{\sigma}^2/(\alpha_{\sigma} - 2)\) for \(\alpha_{\sigma} > 2\), and \(\alpha_{\sigma_k}\) and \(\alpha_{\sigma_i}\) are shape parameters and \(\beta_{\sigma_k}\) and \(\beta_{\sigma_i}\) are scale parameters.

2.3 Parameter model

In a Bayesian perspective, we have to assign a distribution for the parameters \((b, b_0)\) from eq. (5), describing the physical model. From a Bayesian point of view, and for reasons of conjugacy properties of the distributions used, we simply assign a normal distribution to the parameters that we want to make inference on. The means \((\mu_b, \mu_{b_0})\) and variances \((\sigma_b, \sigma_{b_0})\) of those distribution are called ‘hyperparameters’. Hence we have
\[
[b, \mu_b, \sigma_b^2] = N(\mu_b, \sigma_b^2)
\]
\[
[K | \mu_k, \sigma_k^2] = N(\mu_k, \sigma_k^2).
\]
The prior distributions for the hyperparameters are assumed to be independent. We assume non-informative vague uniform prior distributions for the means (see Table 2), and the inverse-gamma prior distributions for the variances; the latter are
\[
[\sigma_b] = \Gamma^{-1}(\alpha_{\sigma_b}, \beta_{\sigma_b})
\]
\[
[\sigma_k] = \Gamma^{-1}(\alpha_{\sigma_k}, \beta_{\sigma_k}),
\]
where \(\alpha_{\sigma_b}\) and \(\alpha_{\sigma_k}\) are shape parameters and \(\beta_{\sigma_b}\) and \(\beta_{\sigma_k}\) are scale parameters (see Section 2.2).

2.4 Posterior distribution for variables and parameters

In Table 2 there is a summary of all the distributions assigned. The last step, now, is to calculate the joint posterior distribution as a product of data model, process model and parameters model. The posterior distribution is the process and parameters distribution updated by the observed data. Remembering eq. (1) and the Bayes’ rule, we have
\[
[R, V, b, K, \mu_b, \mu_{b_0}, \mu_v, \sigma_b, \sigma_b^2, \sigma_k, \sigma_k^2, \sigma_v, \sigma_v^2 | D_v, D_r]
\]
\[
\propto [D_v | R, \sigma_v^{2}] [D_r | V, \sigma_v^{2}] [R | V, b, K, \sigma_k^{2}] [V | \mu_v, \sigma_v^{2}]
\]
\[
\times [b | \mu_b, \sigma_b^2] [K | \mu_k, \sigma_k^2] [\mu_v | \mu_b, \sigma_b^2] [\sigma_v^2 | \sigma_k^2, \sigma_b^2].
\]
(10)
The relevant BH_TPM parameters that we want to simulate from eq. (10) are the parameters of the physical model \(b\) and \(K\), and the error \(\sigma_k^2\). Also, we want to simulate the variables \(R\) and \(V\), in order to compare them with the observations. Finally we simulate \(\mu_v\) and \(\sigma_v^2\) for model check purpose, that will be explained in the following section. In order to simulate these parameters and variables, we have to integrate the joint posterior distribution given by eq. (10).

We use a Monte Carlo integration using Markov Chain (MCMC), where the Markov Chain are constructed using Gibbs sampler (Gilks et al. 1996, and references therein). The Gibbs sampling algorithm generates an instance from the distribution of each variable in turn, conditional on the current values of the other variables/parameters. Therefore, Gibbs sampling works by iteratively drawing samples from the full conditional distribution of each quantity of interest (i.e. variable or parameter); thus we calculate the full conditional distribution for every variable \(R\) and \(V\), and every parameter \(b, K, \sigma_k^2, \mu_v\), and \(\sigma_v^2\). The analytic expression for each full conditional distribution calculated and used for sampling procedure is
\[
[R | \text{rest}] \propto \prod_{i=1}^{41} N(D_{vi} | R_i, \sigma_{D_{vi}}^2)N(R_i | bV_i + K_i, \sigma_k^2)
\]
\[
\times N \left( \sum_{i=1}^{41} \frac{D_{vi}}{\sigma_{D_{vi}}} + \sum_{i=1}^{41} \frac{bV_i + K_i}{\sigma_k^2} , \sum_{i=1}^{41} \left( \frac{1}{\sigma_{D_{vi}}} + \frac{1}{\sigma_k^2} \right)^{-1} \right)
\]
\[
[V | \text{rest}] \propto \prod_{i=1}^{41} N(R_i | bV_i + K_i, \sigma_k^2)N(b | \mu_b, \sigma_b^2)
\]
\[
\times N \left( \sum_{i=1}^{41} \frac{R_i V_i - K_i V_i}{\sigma_k^2} + \frac{\mu_b}{\sigma_b^2} , \sum_{i=1}^{41} \left( \frac{V_i^2}{\sigma_k^2} + \frac{1}{\sigma_b^2} \right)^{-1} \right)
\]
\[
[K | \text{rest}] \propto \prod_{i=1}^{41} N(R_i | bV_i + K, \sigma_k^2)N(K | \mu_k, \sigma_k^2)
\]
\[
\times N \left( \sum_{i=1}^{41} \frac{R_i - bV_i}{\sigma_k^2} + \frac{\mu_k}{\sigma_k^2} , \sum_{i=1}^{41} \left( \frac{1}{\sigma_k^2} + \frac{1}{\sigma_k^2} \right)^{-1} \right)
\]
\[
[\mu_v | \text{rest}] \propto \prod_{i=1}^{41} N(V_i | \mu_v, \sigma_v^2)U(\mu_v | -\infty, +\infty)
\]
\[
\times N \left( \sum_{i=1}^{41} \frac{V_i}{\sigma_v^2} , \frac{41}{\sigma_v^2} \right)^{-1}
\]
\[
[\sigma_k^2 | \text{rest}] \propto \prod_{i=1}^{41} N(R_i | bV_i + K, \sigma_k^2)\Gamma^{-1}(\sigma_k^2 | \alpha_{\sigma_k}, \beta_{\sigma_k})
\]
\[
\times \Gamma^{-1} \left( \frac{41}{2} + \alpha_{\sigma_k}, \frac{1}{\beta_{\sigma_k}} + \sum_{i=1}^{41} (R_i - (bV_i + K_i))^2 \right) \frac{1}{\sigma_k^2}
\]
\[
\times \Gamma^{-1} \left( \frac{41}{2} + \alpha_{\sigma_v}, \frac{1}{\beta_{\sigma_v}} + \sum_{i=1}^{41} (V_i - \mu_v)^2 \right) \frac{1}{\sigma_v^2}.
\]
where the symbol \(U(a, b)\) indicate a uniform distribution on the domain \([a, b]\).

To implement the Gibbs algorithm, we have to set the starting values for each quantity of interest. The numerical solutions, obtained after 11 000 iterations, exhibit good convergence properties for all model variables and parameters. Fig. 2 shows the constructed Markov Chain using Monte Carlo integration for all values of BH_TPM. We discard the first 1000 iterations as the
Figure 2. Constructed Markov Chains for each variable and parameter of the BH_TPM. For $R$ and $V$ we show just one of the 41 chains relative to each variable. In panel (a), each chain (i.e. each subplot) reaches the convergence after few iterations, forgetting the initial guess very quickly. In this case the starting values is chosen to be 10 for all quantities. Iterations in panel (a) represent only the first 100 iteration of the burn-in phase, for more details please refer to the text. The remaining iterations (i.e. from 1001 to 10000), shown in (b) panel, represent the conditional posterior distributions for BH_TPM variables and parameters.
burn-in phase. For readers who are not familiar with MCMC simulations, the burn-in phase is the number of iterations or the time steps needed by chains to reach convergence. After that burn-in phase, the constructed chain can be considered stationary. In Fig. 2 each chain converges after very few iterations. Anyway, for sake of precision, we run each chain longer than needed (i.e. 11 000 time steps), discarding the first 1000 iterations as the burn-in phase (mainly because the longer is the chain, the better is the approximation of the target stationary distribution, and moreover the Gibbs sampler simulation code is very fast to run). Hence the last 10 000 Gibbs sampled time steps set up the posterior distributions for BH-TPM parameters and variables.

In order to ensure the convergence of each chain, regardless of starting values and the number of time steps (i.e. iterations) used, we finally calculate the Gelman Rubin statistics (for a more detailed description of this method see Gelman et al. 2000, 331–332). For this reason we perform a number of different parallel simulations with different starting values, to check that the stationary distributions obtained are not sensitive to the random choice of starting values. We perform 10 parallel simulation of 15 000 runs with different starting values and we monitor the convergence only for parameters $b$, $K$, $\sigma^2_j$, $\mu$, and $\sigma^2_\nu$. We use a higher number of simulations to avoid slow convergence problems related to extreme starting values.

The Gelman Rubin approach is substantially based on comparing different simulated sequences by computing the between-sequence (i.e. $B$) and within-sequence (i.e. $W$) variance (using the same notation present in Gelman et al. 2000). For a general scalar $\phi_i$ with $i = 1, \ldots, n$ and $j = 1, \ldots, J$, where $n$ is the number of the simulations (i.e. 15 000 in our case) and $J$ is the number of parallel sequences (i.e. 10 in our case), we compute

$$B = \frac{n}{J-1} \sum_{j=1}^{J} (\bar{\phi}_j - \bar{\phi})^2,$$

where $\bar{\phi}_j = \frac{1}{n} \sum_{i=1}^{n} \phi_{ij}$ is the mean of the $j$th sequence and $\bar{\phi} = \frac{1}{J} \sum_{j=1}^{J} \bar{\phi}_j$ is the grand mean and

$$W = \frac{1}{J} \sum_{j=1}^{J} s_j^2,$$

where $s_j^2 = \frac{1}{n} \sum_{i=1}^{n} (\phi_{ij} - \bar{\phi}_j)^2$ is the variance within sequence $j$.

Hence we can estimate $\text{var}(\phi|\text{data})$, the marginal posterior variance of the estimate, by a weighted average of $W$ and $B$, namely:

$$\text{var}(\phi|\text{data}) = \frac{n-1}{n} W + \frac{1}{n} B.$$

For a finite number of simulations $n$, the variance $W$ should be an underestimate of $\text{var}(\phi|\text{data})$ because the individual sequences do not cover all the range of target distribution and, as a result, will have less variability; in the limit $n \to \infty$, the expectation of $W$ approaches $\text{var}(\phi|\text{data})$. Therefore, the aim is to monitor convergence by means of the factor

$$R = \sqrt{\frac{\text{var}(\phi|\text{data})}{W}}$$

that has to be close to the unity. For the all parameters in our model, this $R$-factor estimate is in practice equal to the unity. Therefore, we have checked that the constructed Markov chain for BH-TPM parameters are independent of the starting values, then we can use the last 10 000 sampling values as posterior distribution for the parameters $b$, $K$, $\sigma^2_\varepsilon$, $\mu$, and $\sigma^2_\nu$.

### 3 Parameters Estimation and Forecasting

In this section, we examine the results obtained via MCMC–Gibbs Sampling for the model variables and parameters. We explain the physical meaning of the simulated quantities and their reliability to reproduce observational data. We test the forecast capability of this model compared with some appropriate models previously published in the literature.

Before discussing the results obtained for Kilauea Volcano, we test the BH-TPM and its reliability by analysing synthetic data. To this purpose, we generate a sample of 50 synthetic values $v_{\text{syn}}$, from a log-normal distribution with zero mean and unit variance. By definition of log-normal distribution, we have that $V_{\text{syn}} = \log (v_{\text{syn}})$ are normally distributed. This set of 50 $V_{\text{syn}}$ are random draws from a normal distribution and they mimic a synthetic catalogue of volume erupted. These synthetic volume data are substituted into the Time Predictable eq. (5), setting different values for the parameters $b$ and $K$, in order to obtain a ‘purely’ time predictable catalogue of synthetic interevent times $R_{\text{syn}}$. Then, we add a white noise at each synthetic interevent time $R_{\text{syn}}$ using the following equation:

$$R_{\text{syn}} = K + b V_{\text{syn}} + \varepsilon,$$

where $\varepsilon \sim N(0, 1)$ is a stochastic noise term.

Thus we generate three different synthetic data sets of $R_{\text{syn}}$ and $V_{\text{syn}}$, each one with a different value of $b$, that is, $b = 0.5$, $b = 1$ and $b = 1.5$, and the same value of $K = 5$, in order to reproduce three different eruptive regimes achievable with a Time Predictable eq. (4), that is, $0 < b < 1$, $b = 1$ and $b > 1$. With this procedure we build up three synthetic data sets consisting each of 50 pairs of interevent times and volumes. The idea is to use them to test our BH-TPM. Eventually, if the model is robust, we expect to find as outcomes the same $b$ and $K$ values used to obtain the synthetic interevent times $R_{\text{syn}}$ for each data sets. The results of this synthetic test show a good reproducibility of the model respect to the parameters used generating the $R_{\text{syn}}$’s. In Fig. 3 there are the BH-TPM simulations for the parameters $b$, $K$ and $\sigma^2_\varepsilon$; when the synthetic interevent time $R_{\text{syn}}$’s are generated with $b = 0.5$ and $K = 5$. We obtain similar results in the other cases (i.e. $b = 1$ and $b = 1.5$); we do not show them to avoid redundancy. It is even interesting in Fig. 3 that the numerical value of the variance of interevent times distribution of BH-TPM, that is, $\sigma^2_\varepsilon$, is comparable with the noise term $\varepsilon$. Yet, we acknowledge that the three data sets do not contain outliers, so there is a very small variability inside them. Finally, as the model seems to be robust, we apply it to a ‘real’ data set.

### 3.1 Parameters estimation

Using the great flexibility of the implemented Markov Chain, we obtain the numerical values for model variables and parameters in two ways.

1. Using all the first 41 events in the catalogue (Table 1), but discarding the 42nd because it is ongoing, to obtain the distributions of the variables $R$ and $V$ and the parameters $b$, $K$ and $\sigma^2_\varepsilon$, see Figs 4 and 5.

2. Sampling $b$, $K$ and $\sigma^2_\varepsilon$ through a forward procedure. At first, we use only the first event in the catalogue (see Table 1), and we add one pair of volume and interevent time data at a time. Then, we simulate the distribution of each sampled parameter. Therefore, we obtain 41 distributions for the model parameters, each one with
Figure 3. Posterior distributions of relevant parameters of BH_TPM using a synthetic catalogue with $b = 0.5$. The first plot on the left-hand side represents the synthetic data sets (i.e. volumes and interevent times); the other subplots show the parameters inferred by BH_TPM. For more information see the text.

Figure 4. Posterior distribution for relevant parameters simulated using all data in catalogue. In panel (a) it is shown the posterior distribution of parameter $b$; in panel (b) the posterior distribution for parameter $K$ and in panel (c) the posterior distribution for parameter $\sigma^2_R$.

an increasing number of data used (see Fig. 6. In this case, the last distribution in Fig. 6 is the same of the one in Fig. 4.

Note that the second procedure provides interesting information; for example, it allows us to investigate which is the minimum amount of data necessary to have an accurate and informative distribution for model parameters. In other words we can control the amount of data necessary to correctly perform the learning phase for the model. Moreover, this procedure is particularly suitable to mimic a realistic eruption forecasting, since it uses only data available at a specific time to forecast what will happen in the next future.

© 2010 The Authors, GJI, 181, 1525–1538
Journal compilation © 2010 RAS
As it is shown in Fig. 4, the inferred slope parameter $b$ of the GTPM eq. (5) has a well defined distribution. First, we test the null hypothesis $H_0 : b \leq 0$ and we reject it at 5 per cent level of significance, stating that $b$ has a distribution of values significantly greater than zero. Its numerical values are between 0 and 0.5, with mean $\bar{b} = 0.21$ and standard deviation $\sigma_b = 0.10$. This means that GTPM works out for eruptive behaviour at Kilauea Volcano. Moreover its numerical value less than one implies a non-linear
relationship in eq. (4) between interevent times and erupted volumes. Such non-linear relationship implies the possibility of having a non-constant input rate in the magma storage system. Therefore, after a large erupted volume, we expect a shorter interevent time compared with a classic Time Predictable System where the magma input rate is assumed constant in time.

A possible explanation might be represented by an increment in the magma input rate from the depth to the shallow magma storage system after an eruption characterized by a large volume. This might be due to an additional pressure gradient inside the magma chamber owing to magma discharging process, because a large eruption drains the magma chamber and decreases the effective pressure inside it (see Aki & Ferrazzini 2001). This reduction of pressure inside the magma storage system may trigger an increasing of magma buoyancy and, obviously, an increase of the magma input rate. In addition, Takada (1999) shows, as a result of his deterministic model for dyke migrations and stationing in the level of neutral buoyancy, the possibility to have a constant supply rate with oscillations or fluctuations beneath intraplate volcanoes (i.e. Mauna Loa and Kilauea volcanoes).

Another result is reported in Fig. 4, where we show the distribution of intercept $K$ in eq. (5). In terms of its physical meaning, we can consider it as a gauge parameter (see eq. 4) that links together two non-homogeneous quantities, that is, interevent time and erupted volume. As it is shown in Fig. 4, the mean and standard deviation of the parameter $K$ are, respectively, $\overline{K} = 5.27$ and $\overline{\sigma_r} = 0.22$. The main result is that $K$ has a proper finite distribution, that represents the appropriate dimensional constant for eq. (4). The parameter $K$ can also be seen as a function of the average recharge rate: in eq. (5) $K = \ln \alpha$, where $\alpha = (r_i^b R_i)/\bar{R}$, and $\alpha = r_i^b/v_i^b$ from eq. (4). So the dimensionless $\alpha$ parameter, or better the dimensional parameter $c$ is function of the inverse average recharge rate. However, due to the fact that $b$ is different from 1, the term $v_i^b$ makes it difficult to compare physically $1/c$ with the average recharge rate at Kilauea volcano for the period 1923–1983.

The parameter $\sigma^2_{\bar{R}}$ (see Fig. 4) depends on the quantity $R_i - (b V_i + K)$ and it can be seen as a measure of the discrepancy between the simulated interevent times and the Time Predictable equation. This error is a measure of how close the BH$_{\text{TPM}}$ model realizations for $R$ and $V$ fit the data ($D_i$ and $D_r$) when the variables simulated are constrained by the data in the MCMC-Gibbs Sampling. In the process model distribution for interevent times (i.e. eq. 6) errors are additive on the logarithm. After an exponential transformation, this error becomes multiplicative respect to $r_i$. The median of the distribution in Fig. 4 is 1.33, and so an error of about four times the relative interevent time comes out. Nevertheless we cannot consider $\sigma^2_{\bar{R}}$ as a measure of goodness-of-fit for BH$_{\text{TPM}}$ to the data; this aspect is discussed in the next subsection when we simulate and compare synthetic data sets with observational data. This feature of the model in reproducing data with relatively small errors is shown in Fig. 5. The various panels in Fig. 5 represent the simulated volumes and interevent times (blue stars), plotted together with the observed data (red plus) that are always within the simulated distributions.

As mentioned above, Fig. 6 represents the distributions for model parameters $b$, $K$ and $\sigma^2_{\bar{R}}$ using the sampling forward procedure described above at the point 2. Those figures show the learning phase, before the dashed line, and the remaining part used to model checking and forecasting. We choose the first third part of the catalogue, that is, first 14 events, as a learning phase; this means that we test the model on the remaining 27 events. We test again the null hypothesis $H_0 : b \leq 0$ and we can reject it a 5 per cent level of significance for

![Figure 6](https://academic.oup.com/gji/article-abstract/181/3/1525/604166 by guest on 20 March 2019)

---

**Figure 6.** Posterior distributions of: $b$ parameter in panel (a), $K$ parameter in panel (b) and $\sigma^2_{\bar{R}}$ in panel (c), all calculated using the forward procedure discussed in the text. Black dashed line represents the learning phase. Red triangles are the mean of the distributions for $b$ and $K$ and the median for $\sigma^2_{\bar{R}}$. 

---

© 2010 The Authors, *GJI*, 181, 1525–1538

Journal compilation © 2010 RAS
Bayesian Hierarchical Time Predictable Model

Figure 7. Distributions of synthetic interevent times (blue bars) compared with observed values (red line) using descriptive statistic. This goodness-of-fit test (for more detail see the text) shows that our BH_TPM predicts unreasonably long and short interevent times for Kilauea volcano.

3.2 Model checking

The final goal is to check if the model is capable to reproduce satisfactorily the observed data. To this purpose, we follow the approach suggested by Gelman et al. (2000, 161), and we compare the synthetic realizations given by BH_TPM with the real data via descriptive statistics. We choose this approach instead of the classical goodness-of-fit tests, because in this way we can control directly the possible model failures computing the discrepancy between the synthetic realizations (the so called posterior predictive distribution) and data. This is an easy task in Bayesian statistics, because it is always possible to simulate the quantities of interest from their posterior distribution.

In order to compare model realizations and data, we simulate a 10 000 synthetic catalogues from BH_TPM. The first step is to draw a random volume \( V \) from the process eq. (7) using its own mean \( \mu_v \) and variance \( \sigma_v^2 \) already simulated via MCMC–Gibbs sampling using all data. The second step is to simulate an interevent time relative to the simulated volume \( V \) from eq. (6), using the parameters \( b, K \) and \( \sigma_R^2 \) in Fig. 4. We iterate this pattern to replicate the 41 pairs of interevent times and volumes, ending up with a new synthetic catalogue. By replicating this scheme 10 000 times, we obtain 10 000 catalogues each one containing 41 events. The last step is to compare real catalogue (41 observed interevent times) with the 10 000 replicated by BH_TPM, using descriptive statistics. For both real and synthetic catalogues, we calculate the mean number of events (or mean rate of occurrence) \( \lambda \), the maximum, the minimum, the median and the standard deviation of the interevent times.

The results are displayed in Fig. 7, where we show the distributions for the above quantities both for the synthetic realizations (blue bars) and for the real data (red line in figure). The figures suggest that the model generates synthetic data that are reasonably in agreement with real data, even though with some important discrepancies. In particular, the model tends systematically to overestimate the maximum of \( R_i \) and, as a consequence, it tends to underestimate the mean rate of occurrence. Owing to the overestimate of the maximum, the standard deviation is overestimated too. Besides, the minimum is underestimated. In spite of this lack-of-fit, however the median of the distribution shows a better agreement.

A possible explanation of these discrepancies may be linked to the use of log-normal distribution for interevent times and volumes. This choice has been mostly adopted for technical reasons; in fact, the use of conjugate distributions (i.e. normal and inverse gamma distributions) for each level in data model, process model and parameters model, makes the calculations much easier. The log-normal distribution has a fat tail, so when we generate synthetic data by drawing independent samples from such distribution, we obtain large values (both for volumes and interevent times). In this way, in each catalogue generated, there is at least a synthetic eruption with unreasonably large erupted volume. Consequently, there is systematically at least one very large interevent time, implying an overestimate of the maximum. An analogous problem arises for the minimum. We attribute the lack-of-fit for the minimum again to the tail behaviour of the log-normal distribution close to zero. Similarly, the log-normal is not the optimal choice to capture the behaviour of the extreme values of the data in Table 1. Further
developments of the model will drop the assumption of conjugacy in order to improve the model. For now, we argue that these discrepancies do not affect the conclusions about the existence of a time predictable model behind the eruption process. Most important, as we will see in the next section, they do not affect too much the forecasting performances of the model.

3.3 Forecasts

The last check on the reliability of the model consists of comparing the forecasting performances of BH\textsubscript{TPM} against some appropriate models previously published in the literature. We endeavour to compare the forecast capability of BH\textsubscript{TPM} with those of a Poisson model (Klein 1982), Log-Normal model (Bebbington & Lai 1996b) and Generalized Time Predictable Model (GTPM) (Sandri et al. 2005). The test mainly consists of calculating the gain in probability of BH\textsubscript{TPM} with respect to the cited models, under the framework of a probabilistic forecast made on the observed data.

The homogeneous Poisson model is a totally random and memoryless model and it is the simplest model to describe the eruptive process (e.g. Klein 1982; Marzocchi 1996). If the events follow a Poisson distribution, then the interevent times follow an exponential distribution (see Klein 1982; Mulargia et al. 1985; Bebbington & Lai 1996a).

A Log-Normal model has been proposed by Bebbington & Lai (1996b) as a best fit distribution for Kilauea data. According to those authors, a log-normal distribution should take into account the possible eruption cyclicality at Kilauea volcano. The authors test interevent time distribution at Kilauea volcano on ‘all’ data available (i.e. period 1823–1977 AD) trying different possible distributions. The best fit is given by a log-normal one. At the same time, the authors also state that the hypothesis of an exponential interevent time distribution (Poisson process) cannot be rejected when focusing only on data from 1918 to 1977 AD. This latter results is in agreement with Klein (1982). This may mean that for our data set (only eruption form 1923 to 1983, see Table 1) a Poisson model could be preferred. Nevertheless, we compare our model both with Poisson and Log-Normal models.

The GTPM proposed by Sandri et al. (2005) is substantially the non-hierarchical version of the present model (i.e. BH\textsubscript{TPM}). Those authors have applied a regression analysis on the logarithm of the interevent time and volume data at Mt Etna volcano, finding a time predictability for this volcano. However, in that model there is no possibility to use the information given by the volume errors; volume data in GTPM are assumed to be affected only by the scatter around the regression line. Here, we also compare BH\textsubscript{TPM} to GTPM. In this way we point out some justifications for our choice of introducing a hierarchy to better capture the time predictable behaviour, which in turn is mainly due to the necessity of accounting for the volume errors.

To this purpose, following the scheme proposed in seismology by Kagan & Knopff (1987), we calculate the probability gain of BH\textsubscript{TPM} versus Poisson, and Log-Normal and GTPM models as the difference between the log-likelihood of the two models. Because of the complexity of BH\textsubscript{TPM}, we do not have a classical analytical likelihood function, but eq. (6) contains the sample information and the process information, therefore we consider this equation as the likelihood of our model. The probability gain is calculated over the data following the learning phase (see Fig. 6). For each of these eruptions, we calculate the probability of having an event in a time window of 1 month around the observed interevent time. For the BH\textsubscript{TPM} such probability is obtained by eq. (6) with the observed volume datum and parameters estimated from the previous data. For example, in forecasting the 20th interevent time, we use the volume erupted in event number 20 in the catalogue and the parameters inferred from the first 19 events. For the other models, we use the likelihood function to calculate the probability in the same 1 month time window around observed data.

For sake of clarity, this procedure deserves further explanation. To calculate the probability for BH\textsubscript{TPM} we first simulate 10,000 interevent times from the posterior predictive distribution, then we calculate the empirical cumulative distribution function for the simulated interevent times and finally we calculate the probability from the empirical distributions. For Poisson, and Log-Normal models, we instead use the analytical cumulative distribution function. We fit the parameters of those distributions via Maximum Likelihood Estimation using the same forward procedure used for BH\textsubscript{TPM}. For GTPM we first calculate the regression line following the forward procedure described above, then we forecast the interevent time using the regression parameters and the volume datum. The probability here is calculated from the cumulative normal distribution on the logarithm of the data with mean equal to the log-interevent time forecasted and variance equal to the residual mean sum of square. Probability is always calculated as the difference in the 1 month time window around the observed interevent time.

The results are displayed in Fig. 8, where we show the probability gain for each event (the so called ‘punctual probability gain’), and its total value obtained summing up all punctual probability gains. If the probability gain is greater than zero, our model makes better forecast than others. Fig. 8 shows that not all the punctual probability gains are positive, although the total probability gain is positive for all tests. In particular, BH\textsubscript{TPM} does better forecast than all the other models we tested. Our choice of introducing this kind of hierarchy is corroborated by the highest probability gain value, which is obtained against GTPM. In order to check if there are some systematic covariation between the punctual probability gain and the interevent times, we check a possible correlation between these two quantities. We show only for the probability gain against the Poisson process, because this model represents a totally random and memoryless eruptive behaviour for Kilauea. Comparison with a Poisson model allows us to speculate on the physical processes possibly involved in the eruption dynamics.

Fig. 9 shows the relationship between interevent times and punctual probability gains. The inverse linear relationship (the slope is significantly less than zero, \( P\text{-value} \leq 0.01 \)) means that, for very long interevent times, BH\textsubscript{TPM} performs worse than Poisson model. Performing the same analysis for the punctual probability gain against the Log-Normal model, it shows a weaker but still significant, inverse relationship. The slope is less than zero with \( P\text{-value} = 0.0125 \).

There are different possible explanations for the inverse linear relationship: (1) for long interevent times, Kilauea volcano becomes memoryless in its eruptive behaviour (Marzocchi & Zaccarelli 2006); (2) our assumption on the time predictable model as a dynamic eruptive behaviour is too simple to describe events with long repose time; (3) the assumption used to consider eruption as a point event in time without taking into account the eruption duration may become distorting for the model forecast purposes (see Bebbington 2008); (4) with BH\textsubscript{TPM} at Kilauea, we neglect magma intrusions not followed by an eruption (Takada 1999; Dvorak & Dzurisin 1993) and (5) also we neglect possible changes in magma chamber geometry after an eruption (see Guimondson 1986). Further explanations could be derived focusing on the
Figure 8. ‘Punctual probability gain’ of the BH_TPM for each event after the learning phase against: in panel (a) Poisson Model (Klein 1982), in panel (b) Log-Normal Model (Bebbington & Lai 1996b) and in panel (c) Generalized Time Predictable Model (Sandri et al. 2005). Values greater than zero indicate when BH_TPM model performs better forecast than the reference models. The inset in each panel is the total Probability gain, i.e. the sum of the punctual probability gains.

Figure 9. Regression analysis for BH_TPM ‘punctual probability gain’ against Poisson Model versus observed interevent times. The significant inverse linear relationship, whose best-fitting regression coefficients and $R^2$ are given, indicates a systematic negative probability gain for long interevent times. As discussed in the text, this means an additional complexity for long interevent times compared to the time predictable eruptive behaviour. This causes a worse ability of our BH_TPM, compared to Poisson model, to forecast long interevent times.

volumes instead of the interevent times. The volume erupted may change the physical and chemical conditions of the magma chamber and the magma conduit. However performing the same regression analysis as in Fig. 9, but for the volumes instead of interevent times, it does not provide any significative correlation.

4 CONCLUSIONS

In this work, we have developed a time predictable model embedded in a hierarchical Bayesian structure (BH_TPM), to describe the behaviour of eruptive catalogue of open conduit volcanoes. The use
of a Bayesian structure allows to explicitly and formally include in the analysis any kind of uncertainty (relative to data, models, and parameters). We have applied the model to Kilauea eruptive catalogue from 1923 to 1983 AD. The results show that interevent times depend on the previous erupted volume, as in a generalized time predictable model (Sandri et al. 2005; Marzocchi & Zaccarelli 2006). The model shows a reasonable fit with the data observed at Kilauea volcano, although it is not able to capture all the features and variability of the real catalogue. We find also that Kilauea volcano has a weak time-predictability, likely this model could work better when applied to other ‘open’ conduit volcanoes. However, such discrepancies do not seem to affect the forecasting capability of BHT-TPM, that remains superior to the forecasting capability of EF: a probabilistic tool