Journal of Volcanology and Geothermal Research 198 (2010) 57–75

Contents lists available at ScienceDirect

Journal of Volcanology and Geothermal Research

journal homepage: www.elsevier.com/locate/jvolgeores

Testing forecasts of a new Bayesian time-predictable model of eruption occurrence

L. Passarelli [a,b,*], B. Sansò [b], L. Sandri [a], W. Marzocchi [c]

[a] Istituto Nazionale di Geofisica e Vulcanologia, sezione di Bologna via M.A. Franceschini, 31 40128 Bologna, Italy
[b] Dept of Applied Mathematics and Statistics, University of California, 1156 High Street, Santa Cruz, CA 95064, USA
[c] Istituto Nazionale di Geofisica e Vulcanologia, sezione di Roma, via di Vigna Murata, 605 00143 Roma, Italy

ARTICLE INFO

Article history:
Received 19 May 2010
Accepted 9 August 2010
Available online 22 September 2010

Keywords:
effusive volcanism
Bayesian hierarchical modeling
Mount Etna
Kilauea
probabilistic forecasting
volcanic hazards

ABSTRACT

In this paper we propose a model to forecast eruptions in a real forward perspective. Specifically, the model provides a forecast of the next eruption after the end of the last one, using only the data available up to that time. We focus our attention on volcanoes with open conduit regime and high eruption frequency. We assume a generalization of the classical time predictable model to describe the eruptive behavior of open conduit volcanoes and we use a Bayesian hierarchical model to make probabilistic forecasts. We apply the model to Kilauea volcano eruptive data and Mount Etna volcano flank eruption data.

The aims of the proposed model are: (1) to test whether or not the Kilauea and Mount Etna volcanoes follow a time predictable behavior; (2) to discuss the volcanological implications of the time predictable model parameters inferred; (3) to compare the forecast capabilities of this model with other models present in literature. The results obtained using the MCMC sampling algorithm show that both volcanoes follow a time predictable behavior. The numerical values inferred for the parameters of the time predictable model suggest that the amount of the erupted volume could change the dynamics of the magma chamber refilling process during the repose period. The probability gain of this model compared with other models already present in literature is appreciably greater than zero. This means that our model provides better forecast than previous models and it could be used in a probabilistic volcanic hazard assessment scheme.

1. Introduction

One of the main goals in modern volcanology is to provide reliable forecast of volcanic eruptions with the aim of mitigating the associated risk. The extreme complexity and non-linearity of a volcanic system make deterministic prediction of the evolution of volcanic processes rather impossible (e.g. Marzocchi, 1996; Sparks, 2003). Volcanic systems are intrinsically stochastic. In general, eruption forecasting involves two different time scales: (1) a *short-term* forecasting, mostly based on monitoring measures observed during an episode of unrest (e.g., Newhall and Hoblitt, 2002; Marzocchi et al., 2008 among others); and (2) a *long-term* forecasting, usually made during a quiet period of the volcano, and mostly related to a statistical description of the past eruptive catalogs (e.g. Klein, 1982; Bebbington and Lai, 1996a among others). Here, we focus our attention only on this second issue.

In a *long-term* eruption forecast perspective we believe that an incisive and useful forecast should be made before the onset of a volcanic eruption, using the data available at that time, with the aim of mitigating the associated volcanic risk. In other words, models implemented with forecast purposes have to allow for the possibility of providing "forward" forecasts and should avoid the idea of a merely "retrospective" fit of the data available. Models for forecasting eruptions should cover a twofold scope: fit the eruption data and incorporate a testable forecast procedure. While the first requirement is mandatory, the latter one is not commonly used in statistical modeling of volcanic eruptions. By carrying out and testing a forecast procedure on data available at the present, one could make enhancement in the forecast matter and reveal the model limitations.

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 for long-term eruption forecast. Klein (1982) and Mulargia et al. (1985) studied the time series of volcanic events looking at the mean rate of occurrence. Bebbington and Lai (1996a,b) used renewal model framework in studying the eruption time series. Sandri et al. (2005) applied a generalized form of time predictable model to Mount Etna eruptions using regression analysis. Marzocchi and Zaccarelli (2006) found different behavior for volcanoes with "open" conduit regime compared to those with "closed" conduit regime. Open conduit volcanoes (Mt Etna and Kilauea volcanoes were tested) seem to follow a so-called *Time Predictable Model*. While closed conduit volcanoes seem to follow a homogeneous Poisson process. 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 the

* Corresponding author. Istituto Nazionale di Geofisica e Vulcanologia, sezione di Bologna via M.A. Franceschini, 31 40128 Bologna, Italy. Tel.: +39 0514151520; fax: +39 0514151498.
E-mail addresses: passarelli@bo.ingv.it (L. Passarelli), bruno@ams.ucsc.edu (B. Sansò), sandri@bo.ingv.it (L. Sandri), warner.marzocchi@ingv.it (W. Marzocchi).

0377-0273/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.jvolgeores.2010.08.011

history of the volcano discharging behavior. In this paper the author studied the time predictability as a particular case of his model with application to Mount Etna, Mauna Loa and Kilauea data series. A different hierarchical approach has 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 changes of the volcanic regime. Salvi et al. (2006) carried out analysis for Mt Etna flank eruption using a non-homogeneous Poisson process with a power law intensity using the model first proposed by Ho, 1991, while Smethurst et al. (2009) applied a non-homogeneous Poisson process with a piecewise linear intensity to Mt Etna flank eruptions.

In a recent paper Passarelli et al. (2010) proposed a Bayesian Hierarchical Model for interevent time-volumes distribution using the time predictable process with application to Kilauea volcano. The model presents a new Bayesian methodology for an open conduit volcano that accounts for uncertainties in observed data. Besides, the authors present and test the forecast ability of the model retrospectively on the data through a forward sequential procedure. While the model seems to produce better forecasts that some other models in the literature, it produces fits to eruption volumes and interevent times that are too large, reducing the forecast performances. This is due to the use of normal distributions for the log-transformed data. This is a restrictive distributional assumption that creates very long tails. Here we propose a more general modeling strategy that allows for more flexible distributions for the interevent times and volumes data.

Using the same framework of Passarelli et al. (2010), we will model the interevent times and volumes data through distributions with exponential decay (Klein, 1982; Mulargia et al., 1985; Bebbington and Lai, 1996a, b; Marzocchi, 1996; Salvi et al., 2006; Bebbington, 2007; Smethurst et al., 2009). This provides a general treatment of the volume and interevent time series, hopefully improving the forecast capability of the model. As eruptive behavior we use the Generalized Time Predictable Model (Sandri et al., 2005; Marzocchi and Zaccarelli, 2006). This model assumes: (1) eruptions occur when the volume of magma in the storage system reaches a threshold value, (2) magma recharging rate of the shallow magma reservoir could be variable and (3) the size of eruptions is a random variable, following some kind of statistical distribution. Under these assumptions, the time to the next eruption is determined by the time required for the magma entering the storage system to reach the eruptive threshold. The more general form for a time-predictable model is a power law between the erupted volume and the interevent time:

$$r_i = cv_i^b \tag{1}$$

where, 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 $b>1$ or $0<b<1$ we have a non-linear relationship implying a longer or shorter interevent time after a large volume eruption compared to a classical time predictable system. The goal of the present work is to infer the parameters of Eq. (1).

In the remainder of this paper, we focus our attention on some specific issues: (1) to discuss the physical meaning and implications of parameters inferred; (2) to verify if the model describes the data satisfactorily; (3) to compare the forecasting capability of the present model with other models previously published in literature using the sequential forward procedure discussed in Passarelli et al. (2010). In the first part of this paper, we will introduce the generality of the model by considering three stages: (1) a model for the observed data; (2) a model for the process and (3) a model for the parameters (Wikle, 2003). Then we will discuss how: (1) to simulate the variables and parameters of the model; (2) to check the model fit; and (3) to use the model to assess probabilistic forecast in comparison with other statistical published models. The last part of the paper contains the application of the model to Kilauea volcano and Mount Etna eruptive data.

2. A Bayesian hierarchical model for time-predictability

In the following sections we present a detailed description of our proposed model. We denote it as Bayesian Hierarchical Time Predictable Model II (BH_TPM II), while the model proposed in Passarelli et al. (2010) is denoted as BH_TPM. In Section 2.1 we discuss the measurement error model. In Section 2.2 we consider a model for the underlying process, which is based on the exponential distribution. In Section 2.3 we discuss the distributions that are placed on the parameters that control the previous two stages of the model. In Section 2.4 we introduce the simulation procedure and in Section 2.5 we consider model assessment and forecasting of volcanic eruptions.

2.1. Data model

The dataset for this model has n pairs of observations: volumes and interevent times denoted as d_{v_i} and d_{r_i} respectively. We assume independence between the measurement errors of interevent times and volumes. This is justified by the fact that these two quantities are measured using separate procedures. Dependence between times and volumes will be handled at the process stage, following the power law in Eq. (1). In addition, we assume that, conditional on the process parameters the interevent times are independent within their group. This is a natural assumption within a hierarchical model framework. It is equivalent to the standard assumption of exchangeability between the times, which implies that all permutations of the array of times will have the same joint distribution. Similar assumptions apply to the volumes. Exchangeability is a weaker assumption than independence, as independence implies exchangeability.

Our measurement error model assumes a multiplicative error for the observations. This follows from BH_TPM where it was assumed that

$$\log\left(d_{r_i}\right) = \log r_i + \log \epsilon_{r_i} \tag{2}$$

with $\log \epsilon_{r_i} \sim N(0, \sigma^2_{D_{ri}})$, where $\sigma^2_{D_{r_i}} = \left(\frac{\Delta d_{r_i}}{d_{r_i}}\right)^2$ (for more details see Passarelli et al., 2010). The analogous assumption $\log(d_{v_i}) = \log v_i + \log \epsilon_{v_i}$ and $\log \epsilon_{v_i} \sim N(0, \sigma^2_{D_{vi}})$, where $\sigma^2_{D_{v_i}} = \left(\frac{\Delta d_{v_i}}{d_{v_i}}\right)^2$, was considered for the volumes. Exponentiating on both sides of Eq. (2) we have

$$d_{r_i} = \epsilon_{r_i} r_i \tag{3}$$

where the observational errors ϵ_{r_i} are multiplicative and require appropriate assumptions for their probability distributions. The data model is built from the distribution of ϵ_{r_i}, as the distribution of the observations is implied by that of the observational errors.

The error term in Eq. (3) follows a probability distribution with positive support. We choose an inverse gamma distribution. This is a flexible distribution defined by two parameters which will provide computational advantages. The inverse gamma distribution for the errors needs to satisfy two requirements: (1) reflect the fact that observation are unbiased measurements of the true values and (2) provide information on the measured relative error associated to the interevent time and volume data. To achieve the previously discussed desiderata we fix the two parameters of the inverse gamma distribution by assuming $E(\epsilon_{r_i}) = 1$, which corresponds to the first requirement, and calculate $\mathrm{var}(\epsilon_{r_i})$ using a delta method approximation as in Passarelli et al. (2010), to deal with the second requirement. Specifically, from