Journal of Volcanology and Geothermal Research 264 (2013) 197–209

Contents lists available at ScienceDirect

Journal of Volcanology and Geothermal Research

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

Long-range correlations identified in time-series of volcano seismicity during dome-forming eruptions using detrended fluctuation analysis

Stefan M. Lachowycz [a,*], David M. Pyle [a], Tamsin A. Mather [a], Nick R. Varley [b], Henry M. Odbert [c,d,e], Paul D. Cole [d,e,1], Gabriel A. Reyes-Dávila [f]

[a] Department of Earth Sciences, University of Oxford; South Parks Road, Oxford, OX1 3AN, United Kingdom
[b] Facultad de Ciencias, Universidad de Colima; Av. Bernal Díaz del Castillo #340, Col. Villas San Sebastián, C.P. 28045, Colima, Mexico
[c] School of Earth Sciences, University of Bristol; Wills Memorial Building, Queen's Road, Bristol, BS8 1RJ, United Kingdom
[d] Montserrat Volcano Observatory; P.O. Box 318, Flemmings, Salem, Montserrat
[e] Seismic Research Centre, The University of the West Indies; St Augustine, Trinidad and Tobago
[f] Centro Universitario de Estudios y Investigaciones en Vulcanología, Universidad de Colima; Av. Bernal Díaz del Castillo #340, Col. Villas San Sebastián, C.P. 28045, Colima, Mexico

ARTICLE INFO

Article history:
Received 5 March 2013
Accepted 5 July 2013
Available online 1 August 2013

Keywords:
Volcano seismology
Detrended fluctuation analysis
Lava dome
Eruption dynamics
Soufrière Hills Volcano
Volcán de Colima

ABSTRACT

Understanding the underlying structure of data from volcano monitoring is essential to identify precursors to changes in eruptive activity and to comprehend volcanic processes. However, effective analysis of longer-term trends in these signals is challenging as volcanic data are not necessarily statistically stationary or linear, particularly those from lava dome-forming volcanoes, which are commonly characterised by pulsatory eruptive activity. Here, we use detrended fluctuation analysis (DFA), a statistical technique previously applied to nonstationary data, to identify long-range (slowly decaying, e.g. power-law) correlations in a number of time-series of volcano seismicity recorded during the recent dome-forming eruptions of Volcán de Colima, Mexico, and Soufrière Hills Volcano, Montserrat. For all the time-series analysed, correlation strength varies through time and/or on different timescales; in some cases, this variation is periodic, seasonal, and/or related to activity. These results may provide new insights into eruptive processes and possibly further constrain the generation mechanisms of a number of the volcano-seismic event classes analysed. Furthermore, the correlation properties of real-time seismic measurements are shown (retrospectively) to contain information valuable to real-time volcano monitoring that is not identifiable by conventional analysis techniques. This study therefore demonstrates that long-range correlation analysis may be useful for extracting additional information from monitoring data at dome-forming or similar volcanoes.

1. Introduction

Long-lived lava dome-forming eruptions typically comprise non-linear episodes of extrusive and explosive activity; the eruptive style can switch rapidly, compounding the challenge of modelling and forecasting such eruptions (e.g., Wadge et al., in press). Shifts in activity may be accompanied by significant changes in hazard, as exemplified by recent eruptions of Soufrière Hills Volcano, Montserrat (e.g., in 1997: Voight et al., 1999) and Merapi, Indonesia (e.g., in 2010: Surono et al., 2012). In order to improve resilience to hazards in long-lived dome-forming eruptions, we need to develop better tools to anticipate these changes.

Analysing the signals (whether, for example, seismic, geodetic, or gas-chemical) from observation of complex eruptive behaviour often requires a statistical approach (Mader, 2006; Carniel et al., 2008). Time-series of eruption parameters and monitoring data have been analysed by a wide variety of statistical methods, as summarised in the supplementary table. However, in order to apply the majority of these techniques, one must assume that the data reflect a stochastic process and have at least weak (second-order) stationarity, defined as having time-invariant mean and variance, and autocovariance that is only dependent upon the lag time (Nason, 2006). The application of stationary models is often justified, as they make fewer assumptions about the data or volcanic behaviour, and so are more robust (an incorrect non-stationary model will result in greater bias: Marzocchi and Bebbington, 2012). However, no information about temporal variations in activity are sought or incorporated in using a stationary model (Marzocchi and Bebbington, 2012), and so such models are not appropriate when analysing the temporal evolution of activity at

* Corresponding author. Department of Earth Sciences, University of Oxford, South Parks Road, Oxford, Oxfordshire, OX1 3AN, United Kingdom. Tel.: +44 1865 272000; fax: +44 1865 272072.
E-mail addresses: stefan.lachowycz@earth.ox.ac.uk (S.M. Lachowycz), david.pyle@earth.ox.ac.uk (D.M. Pyle), tamsin.mather@earth.ox.ac.uk (T.A. Mather), nick@ucol.mx (N.R. Varley), h.odbert@bristol.ac.uk (H.M. Odbert), paul.cole@plymouth.ac.uk (P.D. Cole), gard@ucol.mx (G.A. Reyes-Dávila).
[1] Now at: School of Geography, Earth and Environmental Sciences, Plymouth University; Fitzroy, Drake Circus, Plymouth, PL4 8AA, United Kingdom.

0377-0273/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jvolgeores.2013.07.009

volcanoes that show regime changes, periodic behaviour, or trends, which includes many dome-forming systems (Bebbington, 2010).

One group of statistical methods that can inform both forecasting and our understanding of volcanic processes are those that quantify persistence. Persistent (or correlated) behaviour, where similar values are clustered in time, may be one indicator of 'memory' in a system, when the system state at one point in time influences future conditions or events. Case studies of such behaviour by volcanic systems are described by Carniel et al. (2008): for example, Jaquet et al. (2006) use variograms to quantify memory within repose interval and amplitude time-series of a month-long sequence of Vulcanian explosions at Soufrière Hills Volcano, Montserrat. This memory has been attributed to decompression of ascending magma, from which magma ascent rates and conduit geometry could be quantitatively constrained, and later events forecast from the correlations between earlier sequential eruptions (Jaquet et al., 2006). In spite of this potential, similar correlations on longer (monthly to multi-annual) timescales have rarely been investigated quantitatively. Long-range correlations (i.e., correlations that decay slowly, such that the characteristic correlation timescale is indefinable: Kantelhardt, 2009) are commonly exhibited by non-linear dynamical systems far from equilibrium (Peng et al., 1995), so might be expected of active volcanoes.

In this study, we identify such long-range correlations in volcano-seismic time-series from two intensively monitored dome-forming volcanoes, Volcán de Colima, Mexico, and Soufrière Hills Volcano, Montserrat, by detrended fluctuation analysis (DFA) (Peng et al., 1994). This fractal scaling analysis method, which filters any local trends in the time-series, has been used to quantify the correlation properties of non-stationary data in a variety of disciplines (e.g., physiology: Peng et al., 1995, climatology: Livina and Lenton, 2007, and economics: Alvarez-Ramirez and Escarela-Perez, 2010). This technique has also been applied to volcanological data: for example, the hourly variability in geomagnetic signals recorded on Etna (Italy) was (using DFA) found to show persistent behaviour that varies on different length scales and through time, with an abrupt increase in correlation strength being associated with an eruption in October 2002 (Currenti et al., 2005a). Multifractal DFA scaling exponents (Kantelhardt et al., 2002) of these data were shown to be less variable after this eruption than before, further constraining the correlation dynamics of the signal (Currenti et al., 2005b). Similarly, DFA of the daily count of small explosions at Popocatépetl (Mexico) identified quasi-periodic temporal variation of the long-range correlations in this time-series, which varied in step with changes in eruptive activity and slow-slip events at the associated subduction zone (Alvarez-Ramirez et al., 2009, 2011). The 'log-log' plots calculated in DFA (explained in Section 2) have also been used, for example by Del Pin et al. (2008) to detect the presence of tectonic events in segments of noise-contaminated seismic data recorded at Teide (Tenerife, Canary Islands). Hurst rescaled range analysis (Hurst et al., 1965), which calculates an exponent comparable to that from DFA, has also been applied to volcanological data (see the supplementary table), but is only appropriately calculated for stationary data.

2. Detrended fluctuation analysis (DFA)

DFA requires a time-series $u(i)$ of N values (where $i = 1, \ldots, N$) to first be integrated at each point, k, in the series, as follows: $y(k) = \sum_{i=1}^{k}[u(i) - \overline{u}]$ where $\overline{u}$ is the mean of the whole dataset. $y(k)$ is then divided into non-overlapping boxes (time windows) of length n, and the local trend in each box computed by a linear least-squares fit to the constituent data. Higher-order fits may be applied to calculate this trend, but are not routinely used (Little et al., 2006). The trend is removed from each box to leave locally detrended data, $y_d(k)$. The root-mean-square fluctuation $F(n) = \sqrt{\frac{1}{N}\sum_{k=1}^{N}[y_d(k)]^2}$ of the detrended data is then computed, and the whole process repeated for a range of scales of n. In this study, F is calculated for every n-value from 4 to $N/4$, a range comparable to previous applications of DFA (e.g., Currenti et al., 2005a), to give a set of fluctuation values, $F(n)$. A linear relationship between $\log[F(n)]$ and $\log[n]$ indicates self-similarity (scaling). The gradient of this line (calculated by least-squares regression) is the scaling exponent (or self-similarity parameter), $\propto$ (Peng et al., 1995). Changes in $\propto$ with increasing n reflect different scaling properties on different timescales; we identify any break points in the gradient of the $\log[F(n)]$ vs $\log[n]$ plots by inspection (although these could be determined by other means, such as change point analysis (cf. Mulargia et al., 1987)). We calculate $\propto$ using the computationally efficient 'FastDFA' algorithm (Little et al., 2006), which follows the original formulation of DFA (Peng et al., 1994, 1995) summarised above.

In a subsample of a time-series where each value is not correlated with any previous values (e.g., white noise), $\propto \cong 0.5$. Values in the range $0.5 < \propto < 1$ indicate long-range power-law correlation (i.e., persistence), such that a large value (relative to the mean) is more likely to be followed by large values, and vice versa. In contrast, $0 < \propto < 0.5$ signifies anti-persistence, where large and small signal values are more likely to alternate. Strongly persistent, $1/f$-like ('pink') noise would have a value of $\propto \cong 1$. When $\propto > 1$, strong correlations exist, but are not of a power-law form; $\propto \cong 1.5$ would result from Brownian ('red') noise, i.e. random walk-like fluctuations in the signal through time. Thus, $\propto$ may be considered a measure of time-series 'roughness', becoming smoother with increasing $\propto$ (Peng et al., 1995). 'Critical slowing down' (a decreasing rate of recovery from small perturbations of the system) prior to a sudden change in the dynamics of a complex system (i.e., a 'tipping point') may be indicated by an increase in $\propto$ towards $\propto \geq 1$ (Livina and Lenton, 2007).

Power-law scaling can result from either long-range correlations or fat-tailed probability distributions (Mandelbrot and Wallis, 1968). These can be distinguished by removing any correlations in the time-series by randomly shuffling the data: this has no effect on the distribution, so any scaling identified by DFA (i.e. $\propto \neq \sim 0.5$) after shuffling will be due to a fat-tailed distribution (Alvarez-Ramirez et al., 2009). DFA of entire volcanic time-series can be informative for evaluating the general scaling behaviour (e.g., Currenti et al., 2005a); we present the results of such analysis as $\log[F(n)]$ vs $\log[n]$ ('log-log') plots, for the range of box lengths (n-values) for which well-defined scaling (a strong linear relationship between $\log[F(n)]$ and $\log[n]$) is present.

We investigate temporal variation in $\propto$ by DFA of overlapping samples (time 'windows') of the data: i.e., $\propto$ is calculated for a window of a specified constant time length, run incrementally through the time-series. Windows in which >50% of values are zero (due to a gap in recording or absence of seismicity) are not analysed. The window size would ideally be short to minimise lag effects, but longer windows reduce the potential for significant finite-size effects (Ivanova and Ausloos, 1999). The window lengths used were selected to balance these competing factors: smaller window lengths were rejected if they resulted in poorly-defined scaling relationships in a significant number of windows, determined by inspection of the log-log plots of a selection of windows, particularly those for times when the $d\propto/dt$ is comparatively high. The window lengths used are such that the n-values for calculating each exponent are within the range that shows a well-defined scaling relationship in DFA of the whole time-series (specified on the log-log plots in Section 4), but do not necessarily capture the full range of scaling in the time-series as a whole.

The suitability of DFA as an alternative to conventional fluctuation analysis for analysing non-stationary data has been questioned by Bryce and Sprague (2012), on the basis of the impact of bias from finite-size effects. The principal effect is spurious curvature on log-log plots towards the lower limit (i.e., an increasing deviation from the expected linear trend towards the smallest n-values); this causes a bias to the resulting scaling exponent, and implies that fine-scale detrending (to properly address non-stationarities) in DFA can introduce artefacts. However, whole-dataset DFA of each time-series analysed in this study shows that this spurious deviation is only intermittently present for these data when $n < 5$ (i.e., it affects only a couple