ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/341645163

# Understanding the timing of eruption end using a machine learning approach to classification of seismic time series

**Article** *in* Journal of Volcanology and Geothermal Research · May 2020

DOI: 10.1016/j.jvolgeores.2020.106917

CITATIONS
0

READS
144

**9 authors**, including:

David M Pyle
University of Oxford
**368** PUBLICATIONS **10,846** CITATIONS
SEE PROFILE

Mel Rodgers
University of South Florida
**33** PUBLICATIONS **110** CITATIONS
SEE PROFILE

Glenn Thompson
University of South Florida
**52** PUBLICATIONS **830** CITATIONS
SEE PROFILE

John Makario Londoño-Bonilla
Servicio Geológico Colombiano
**51** PUBLICATIONS **240** CITATIONS
SEE PROFILE

**Some of the authors of this publication are also working on these related projects:**

Project GISMO: A seismology toolbox for MATLAB View project

Project Santorini View project

All content following this page was uploaded by Diana C Roman on 12 September 2020.

The user has requested enhancement of the downloaded file.

Journal of Volcanology and Geothermal Research 401 (2020) 106917

Contents lists available at ScienceDirect

Journal of Volcanology and Geothermal Research

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

# Understanding the timing of eruption end using a machine learning approach to classification of seismic time series

Grace F. Manley [a,*], David M. Pyle [a], Tamsin A. Mather [a], Mel Rodgers [b], David A. Clifton [c], Benjamin G. Stokell [d], Glenn Thompson [b], John Makario Londoño [e], Diana C. Roman [f]

[a] Department of Earth Sciences, University of Oxford, UK
[b] School of Geosciences, University of South Florida, USA
[c] Department of Engineering Science, University of Oxford, UK
[d] Department of Mathematics and Mathematical Statistics, University of Cambridge, UK
[e] Servicio Geológico Colombiano, Observatorio Vulcanológico y Sismológico de Manizales, Colombia
[f] Department of Terrestrial Magnetism, Carnegie Institution for Science, USA

ARTICLE INFO

*Article history:*
Received 14 November 2019
Received in revised form 14 May 2020
Accepted 15 May 2020
Available online 25 May 2020

ABSTRACT

The timing and processes that govern the end of volcanic eruptions are not yet fully understood, and there currently exists no systematic definition for the end of a volcanic eruption. Currently, end of eruption is established either by generic criteria (typically 90 days after the end of visual signals of eruption) or criteria specific to a given volcano. We explore the application of supervised machine learning classification methods: Support Vector Machine, Logistic Regression, Random Forest and Gaussian Process Classifiers and define a decisiveness index D to evaluate the consistency of the classifications obtained by these models. We apply these methods to seismic time series from two volcanoes chosen because they display contrasting styles of eruption: Telica (Nicaragua) and Nevado del Ruiz (Colombia). We find that, for both volcanic systems, the end-date we obtain by classification of seismic data is 2–4 months later than end-dates defined by the last occurrence of visual eruption (such as ash emission). This finding is in agreement with previous, general definitions of eruption end and is consistent across models. Our classifications have a higher correspondence of eruptive activity with visual activity than with database records of eruption start and end. We analyze the relative importance of the different features of seismic activity used in our models (e.g. peak event amplitude, daily event counts) and find little consistency between the two volcanic systems in terms of the most important features which determine whether activity is eruptive or non-eruptive. These initial results look promising and our approach may offer a robust tool to help determine when an eruption has ended in the absence of visual confirmation.

© 2020 Elsevier B.V. All rights reserved.

## 1. Introduction

The processes which govern large-scale change in volcanic systems are not yet fully understood. Volcanic systems are dominated by complex and non-linear processes. This complexity has implications for both understanding and forecasting the onset of volcanic activity (e.g. Sparks, 2003), and for managing transitions in behaviour during prolonged eruptions (e.g. Sparks and Aspinall, 2004; Hicks and Few, 2015; Barclay et al., 2019). While much attention has been focussed on forecasting the timing of eruption onset, and the timing of alerts, warnings and calls for evacuation in the run-up to eruption, or as the eruption escalates (e.g. Marzocchi and Woo, 2007; Winson et al., 2014; Cameron et al., 2018), less attention has been focussed on the ends of volcanic eruptions (Bonny and Wright, 2017).

There is no widely-accepted systematic definition for the end of an eruptive period (e.g. Phillipson et al., 2013) and, as a result, end-dates are often poorly reported. Although some global compilations of volcanism contain a field for eruption end-date, the end-dates are often not recorded. The Smithsonian Global Volcanism Program (GVP) highlight that, of the 10,415 eruptions in the Volcanoes of the World (VOTW) database at the time of writing, there were no available termination dates for 59% of the entries (Siebert et al., 2011). This lack of data was attributed to the gradual nature of eruption endings, which made assigning a discrete date difficult. Phillipson et al. (2013) suggest that end-dates in the GVP database have a temporal uncertainty on the order of days, but inspection of the slow decline in observable activity at some systems (such as Mont Pelée after 1905; or Soufrière Hills Volcano Montserrat since 2010; Lacroix, 1908; Wadge et al., 2014) suggests that the uncertainty could be much larger in some cases - not least in the absence of a definition of what constitutes the 'end of eruption'. Even in well-monitored cases eruption end dates are hard to determine. For example,

* Corresponding author.
*E-mail address:* grace.manley@earth.ox.ac.uk (G.F. Manley).

https://doi.org/10.1016/j.jvolgeores.2020.106917
0377-0273/© 2020 Elsevier B.V. All rights reserved.

even though the activity at Mt. St Helens from 2004 to 2008 was closely monitored with networks of instruments (including seismic, tilt and gas measurements), determining the end of the eruption was impeded by poor weather conditions throughout the month of December 2007. It could not be conclusively determined that small-scale extrusion was finished until visual observations were made in January 2008 (Dzurisin et al., 2015).

The lack of a specific definition for eruption end has implications for volcanic hazard. Tilling (2009) cites the mistaken identification of decreased volcanic activity as the end of eruption as one of the primary reasons for major loss of life during the 1982 El Chichón eruption. De la Cruz-Reyna et al. (2017) identify the definition of eruption end as a particular difficulty during sustained periods of activity, such as at Popocatépetl in Mexico. Popocatépetl has been in continuous eruption since 1994, exhibiting both effusive and explosive activity, but with lulls in activity which have led to uncertainty over whether or not they represent the end. Similar stop-start activity has characterised the long-lived and ongoing dome-forming eruptions of Santaguito, Guatemala (1922 to present), and Soufrière Hills Volcano, Montserrat (1995 to present). Better understanding of the timing of eruption end could have implications for the allocation of resources and management of populations living adjacent to volcanoes during both acute and sustained volcanic eruptions.

Obtaining an operational definition for the end of an eruption relies on piecing together various measurements of volcanic activity to determine when a break in volcanic activity represents the end of the eruptive period. Definitions for eruption end in a monitoring context come under two main categories:

(i) Generic rules on the end of eruptions: For example, Simkin and Siebert (1994) used a generic 90-day (or 3-month) rule for the end of eruption, i.e., that if a volcano displays no visible signs of eruption for 90 days, then the eruption can be considered over;
(ii) Definitions that are volcano-specific. For example, eruptions at Piton de la Fournaise, Réunion, are defined solely on increases or decreases in seismic tremor amplitude (Battaglia and Aki, 2003).

Volcanic systems undergo periods of activity and repose, on varying timescales (e.g. Barmin et al., 2002; Lamb et al., 2014). Identifying the critical thresholds which govern when large-scale changes in volcanic behaviour occur is acknowledged as one of the fundamental research questions associated with understanding the beginning, evolution and termination of volcanic activity (NAS, 2017). Development of models for these critical thresholds is necessary to understand the processes which drive large-scale change in volcanic behaviour, but this in turn requires knowing when these changes occur in the timeline of an eruption.

To understand the timing of large-scale change in the system, it is important to define the various states of volcanic behaviour. Siebert et al. (2015) define eruption according to the observation of the following: explosive ejection of either fragmented new magma or older solid material, and/or the effusion of liquid lava. Activity outside eruption is defined as unrest, and a volcano in no state of eruption or unrest is said to be in repose (unless extinct). Phillipson et al. (2013) define two further categories of unrest: pre-eruptive and non-eruptive unrest, which are based on the presence of observable parameters such as deformation or seismicity. Neither of these categories can be assigned in real time, it being necessary to wait either until the crisis has subsided or an eruption has begun to determine whether the unrest was pre-eruptive or not.

Volcanic state has been previously characterised in models of seismic evolution over the eruptive period. McNutt (1996) developed the generic earthquake swarm model, which describes the evolution of seismic activity over an eruptive cycle. In this model, it is suggested that the rate and type of seismicity observed over an eruption can be generalised, and that the physical processes governing the type of seismicity at each eruptive stage may be inferred. Carniel (2014) describe how time series which have undergone a process of data reduction can be used to identify and infer the timing of transitions between different states of volcanism.

Machine learning is the process by which computers learn without being explicitly programmed. In fields such as healthcare, jet engine monitoring or economics, the use of machine learning methods for both data analysis and real-time monitoring is already established. Volcanic systems have conceptual parallels with these systems: they can be described as a “high-integrity” system (Clifton et al., 2014) in which observation of failure (i.e., eruption) is rare in comparison with stable behaviour, and the number of failure modes are not known or not well characterised. The use of machine learning techniques in volcanology is an emerging field. Pattern recognition techniques have previously been applied to volcano-seismic data, with a particular focus on detection and classification of seismic event types from raw waveform data (e.g., Langer et al., 2006; Curilem et al., 2009; Apolloni, 2009; Bicego et al., 2013; Maggi et al., 2017; Malfante et al., 2018). Machine learning has also been applied to satellite data, in order to detect signs of unrest in large numbers of acquisitions: Anantrasirichai et al. (2018) used deep learning to detect ground deformation in Sentinel-1 data, and Flower et al. (2016) used logistic regression to detect volcanic eruptions in global daily observations of $SO_2$ measured using the Ozone Mapping Instrument. Ren et al. (2020) used multi-station seismic tremor measurements to classify behaviour at Piton de la Fournaise volcano and identify fundamental frequencies of the tremor.

In this paper we use classification machine learning models (see Section 2.1 for a full description) (i) to classify eruptive and non-eruptive patterns in volcanic time series data, and (ii) to observe how these patterns differ from inferences based on visual observation or conventional monitoring techniques. Similar classification techniques have been previously successfully applied in a healthcare context to classify patient state (e.g., Clifton et al., 2014) and therefore have potential for characterising volcanic state. Our approach is distinct from previous work in volcanology (discussed above) as we classify overall volcanic state as eruptive or non-eruptive, as opposed to aiming to detect distinct change in one observable. We present a proof-of-concept study in classifying seismic time series for two volcanoes selected to cover a range of eruptive styles.

## 2. Methods

### *2.1. Machine learning methods used*

Fig. 1 illustrates the four multi-class methods used for the analysis in this paper. The methods used are all supervised methods, in which we select days from time series data (e.g., seismic event counts) for training that include both eruptive and non-eruptive examples to train the model. During training, some training points are held back and used to test the trained model to increase the model accuracy (a process known as model validation). Once a model has been trained, it is then tested using days from the time series which were not presented during the training period.

For this preliminary work, we use machine learning models where features are calculated and chosen as inputs to the model, as opposed to other methods such as deep learning wherein features are calculated and chosen within the model. Choosing features as model input is preferred so that we can use features derived from the seismic data that are similar to those used in current monitoring practices. These features, such as event rate or peak signal frequencies, have had widespread success in a monitoring context as a basis for distinguishing between states of eruption (Carniel, 2014). The established use of these features in a monitoring context means that results such as the relative importance of a given feature in the models is directly applicable to current observatory practices.

Each machine learning model we apply is distinct from the others in its method of determining the boundary between non-eruptive and

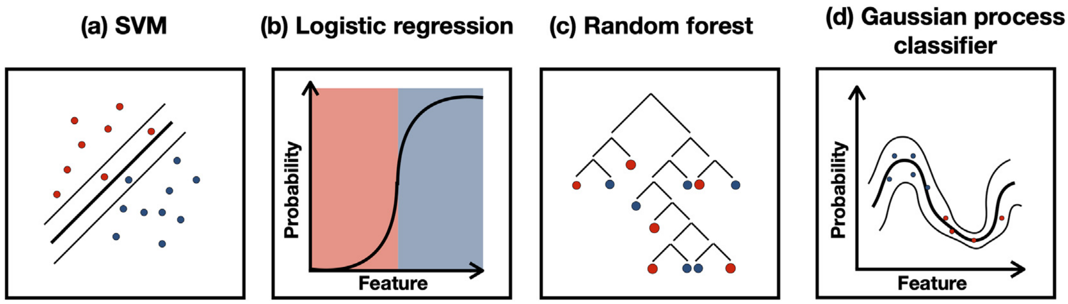


**Fig. 1.** Visualisation of how boundaries are defined within the four machine learning methods used in this paper: (a) Support Vector Machine (SVM) (b) Logistic regression (c) Random forest decision tree (d) Gaussian process classifier with uncertainty bounds marked. Red and blue dots denote two classes of data, which in this study represent non-eruptive and eruptive data. The y-axis in (b) and (d) represents the probability that a given day belongs to the red class or blue class.

eruptive data. We choose to apply multiple methods which have the same training period, to observe whether the classification of eruptive and non-eruptive behaviour is consistent with several distinct methods. Each method determines the boundary between classes in a different way, and thus the methods have their own advantages and disadvantages. Support Vector Machine models (SVMs, Section 2.1.1) are good at handling non-linear relationships between data. Logistic regression (LR, Section 2.1.2) is a straightforward model, and thus rarely overfits data, while offering a fully interpretable approach whereby its parameters are informative of the relative contribution to the classification of the input variables. Random forest models (RFs, Section 2.1.3) are generally associated with high classification accuracy and involve taking an ensemble of individual decision trees (where the latter are weak classifiers). Gaussian Process models (GPs, Section 2.1.4) directly capture the uncertainty associated with the prediction and offer a principled approach to dealing with artefact in time-series data.

*2.1.1. Support vector machine (SVMs)*

SVMs (Fig. 1a) involve finding the hyperplane between two classes of data which maximises the margin of the classification, where the margin is defined as the perpendicular distance between the decision boundary and the closest data points (Bishop, 2006). SVMs use the “kernel trick” to transform the data to a higher dimensional space, in which potential non-linearities in the original data can be separated (which would not be possible for logistic regression and other generalised linear models, for example). The choice of kernel depends on the properties of the dataset, such as non-linearity of the data. SVMs have been widely used in the field of seismic detection (e.g., Ruano et al., 2014), and they are well suited to general models even with few training examples (Mountrakis et al., 2011). We use the LibSVM libraries (Chang and Lin, 2011) to formulate models using both Radial Basis Function (RBF) and even-order polynomial kernels. The values of the hyperparameters for these models are selected using standard 5-fold cross-validation (Hastie et al., 2001).

*2.1.2. Logistic regression (LR)*

LR models (Fig. 1b) are a form of generalised linear model: this means that the classification linearly depends on the features, where each feature has a coefficient in the linear model. Logistic regression models the posterior probability of a given day being eruptive as a continuous (sigmoid, or S-shaped) function of a linear expression of the features. The probabilistic output of these models means that for each day of results, we can infer the certainty of a classification on that day. A full discussion of logistic regression is included in McCullagh and Nelder (1989) and Hastie et al. (2001). In the Earth Sciences, LR models have previously been applied in estimation of landslide hazard (Pradhan and Lee, 2010).

*2.1.3. Random forest (RF)*

RF classification (Fig. 1c) involves the averaging of an ensemble of decision trees: each decision tree comprises a series of operations that consecutively compare available features in the input data to randomly-selected thresholds on those features (Hastie et al., 2001). Many possible decision trees are combined in random forest classification: the result of each tree contributes a vote towards the final classification. The hierarchical nature of decision trees means that these methods can be used to determine the relative importance of the features, where features that appear towards the top of the decision tree contribute more to the classification and are therefore associated with greater importance (Breiman et al., 1984). Random forest models have been applied extensively in remote sensing, due to the high accuracy of classification obtained and their ability to identify important variables (discussed more in Section 4.4; and Belgiu and Drăguţ, 2016).

*2.1.4. Gaussian process classification (GPs)*

Gaussian processes (GPs) (Fig. 1d) fit stochastic models to obtain a probability that a given data point is in a given class (Bishop, 2006). The classification that results from a Gaussian process classification is therefore associated with a given uncertainty. Like SVMs, GP classification involves a choice of kernel function to train the model. We use an Automatic Relevance Determination (ARD) kernel for training models, which allows the importance of each feature input into the model to be evaluated (Williams and Rasmussen, 2006).

*2.2. Model assumptions*

The models used in this paper require the fundamental assumption that the input variables are Independent and Identically Distributed (IID). Assuming data which are IID implies that each day of features is independent from other days of features, and that the features on each day are drawn from the same underlying statistical distribution (Cover and Thomas, 2006). While data are rarely perfectly IID in practice, the assumption typically holds to the degree that models involving such approaches yield satisfactory results. To determine whether or not this hypothesis is appropriate for our data, and therefore whether or not our models can perform well using previously-unseen data, we train and test (using data held-out during the training process) separate models for each volcanic system.

Though we are considering a physical system, which may not provide perfectly IID data, we hypothesise that this assumption is valid to a first order approximation. Other machine learning models exist which can take into account non-IID behaviour; however, these methods are beyond the scope of this paper.

*2.3. Nevado del Ruiz and Telica volcanoes*

We analyze single-station seismic datasets from two volcanic systems: Nevado del Ruiz Volcano, Colombia and Telica Volcano, Nicaragua. We choose to apply models to the datasets from these two volcanic systems for our test study as they represent contrasting styles of activity. Telica displays near-continuous levels of seismic activity, whereas Nevado del Ruiz represents more punctuated volcanic activity. Therefore, successful classification of these differing styles is a useful proof-of-concept that these methods have the potential to be extended to a range of volcanic situations.

Nevado del Ruiz is a stratovolcano in Colombia which primarily erupts products of andesite-basaltic andesite composition (Cuellar-Rodriguez and Ramirez-Lopez, 1987; Londono, 2016). The dataset is from 21st March 2007 to 25th February 2015 and covers two eruptive periods, both recorded in GVP. The first phase has a start-date of 22nd February 2012 and end-date of 12th July 2013, and the second, much shorter, phase has a start-date of 15th December 2014 and end-date of 7th January 2015. An increase in seismic activity began in September 2010 and ash emissions from Nevado del Ruiz were observed from early 2012 onwards (Global Volcanism Program, 2017). The 2012 ash emissions of Nevado del Ruiz were the first emission of ash since the VEI 3 eruption of 1985, which led to the lahar inundation of Armero and >25,000 fatalities (Lowe et al., 1986; Naranjo et al., 1986). The end-date of the first phase is a day later than the last recorded ash emission and coincides with the last advisory of the Washington Volcanic Ash Advisory Centre (VAAC, 2013). It is unclear how the dates for the second phase are chosen, as ash emission was observed both in November 2014 and later in January 2015.

Telica volcano is a persistently restless volcano in Nicaragua which undergoes small (VEI 1–2) eruptions every few years (Geirsson et al., 2014; Rodgers et al., 2013; Rodgers et al., 2015a). Persistently restless volcanic systems are characterised by high and variable rates of seismicity and degassing, with frequent explosive activity (Rodgers et al., 2015a; Geirsson et al., 2014; Roman et al., 2019; Stix, 2007). The Telica data used for this analysis were obtained from the TESAND network, for the period 1st April 2010 to 18th March 2013. This data period contains one VEI 2 eruption, recorded in GVP with a start-date of 7th March 2011 and an end-date of 14th June 2011. The end-date is 3 days after the last ash-and-gas explosion sequence (comprising 17 explosions) was observed on 14th June 2011 (Geirsson et al., 2014).

### 2.4. Classification scheme

Fig. 2 illustrates the process of training and testing the multi-class methods introduced in Section 2.1. We use supervised machine learning algorithms: these models are trained on labelled data and subsequently tested on unseen data. Data are labelled in “eruptive” and “non-eruptive” classes according to the eruption dates recorded by the GVP database. Non-eruptive data comprises all of the data which does not fall under the dates recorded in GVP for the eruption. We select training periods to represent non-eruptive and eruptive data as input to the models. These training periods are selected such that they do not overlap with the GVP start and end dates, because we want to independently constrain the timing of transitions between eruptive and non-eruptive activity. The eruptive and non-eruptive training periods are chosen to represent times in which the presence or absence of visual eruption was confirmed from activity reports, archived by the Servicio Geológico Colombiano (SGC) and Instituto Nicaragüense de Estudios Territoriales (INETER).

For Nevado del Ruiz, the non-eruptive period we select to train the model is from 15th June 2009 to 30th September 2011. This non-eruptive period is approximately 4 months before the weekly reports mark the first possible ash emission, clear deformation signal and increase in $SO_2$ emission (Global Volcanism Program, 2012a, 2012b). The eruptive training periods selected are 23rd March 2012-26th February 2013 and 9th April 2013–25th April 2013. These periods are selected because they coincide with ash emissions confirmed by Manizales observatory (Appendix 2; SGC). We use 938 days of the Nevado del Ruiz daily time series for training the models and 1364 days of the time series for subsequently testing the models.

For Telica, we choose non-eruptive training periods from 29th June 2010–8th January 2011 (before eruption) and 1st September 2012–20th November 2012 (a year after the eruption). The second half of the training period is selected as it is over a year after the end of eruption. The eruptive training period is from 28th March 2011–1st June 2011, starting after ash emission had been confirmed by visual observation (Global Volcanism Program, 2011; Geirsson et al., 2014). We use 333 days of the Telica daily time series for training the models and 730 days of the time series for subsequently testing the models.

### 2.5. Feature extraction

Feature extraction is the process of selecting variables which will be used as inputs into machine learning models. Fig. 3 describes the process of feature extraction from raw seismic data. The inputs to machine learning models are time series derived from raw seismic data. To produce these derived time series, several categories of features are selected from the seismic data (see AP 1.1 and 1.2). For all features apart from event rates, the data are taken from a single seismic station. Telica data is from the TBTN station of the TESAND network, located approximately 1 km east of the active vent (Roman, 2009). Nevado del Ruiz data is from the BISZ station, located approximately 2 km west of Arenas

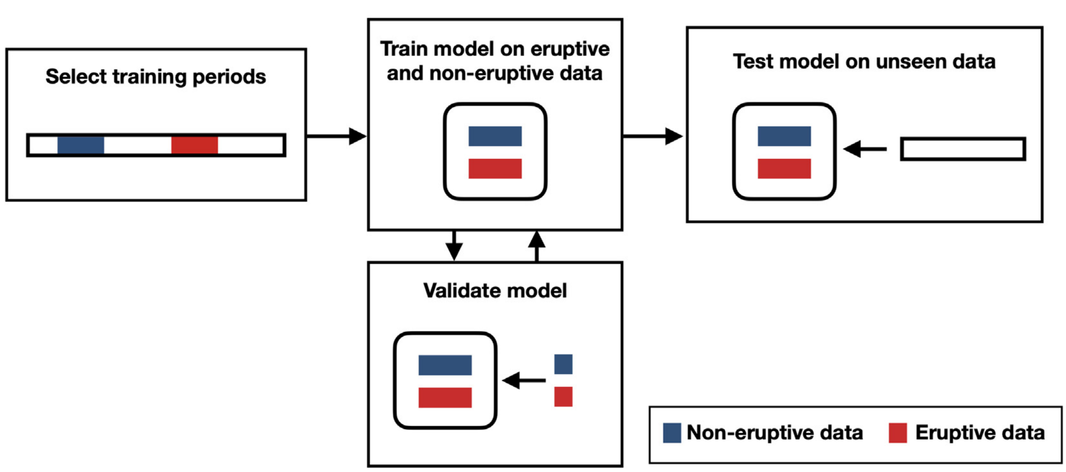


**Fig. 2.** The framework of training and testing supervised multi-class classification models. Training periods which include non-eruptive and eruptive data are selected. The model is trained on a subset of these training data, and concurrently validated using the rest of the training data. After a model has been trained, the model can be run again with testing data. Blue represents data labelled as non-eruptive and red represents data labelled as eruptive.

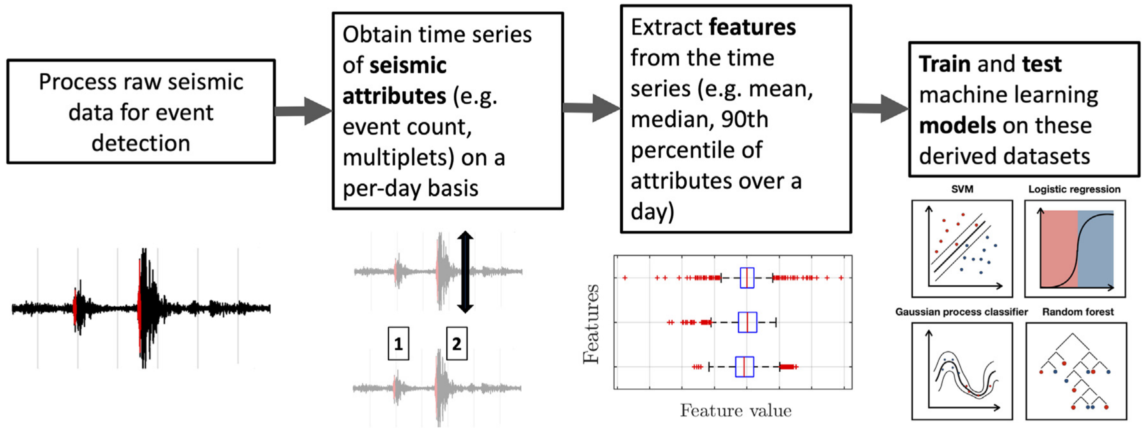


**Fig. 3.** Diagram of the framework for extracting features from raw seismic waveform data. Events are detected from raw waveform data. From each event waveform we extract features including the peak amplitudes and band ratio. We then calculate features including the mean and variance from all of the waveforms in a given day. The resulting time series are used as input for the machine learning model.

crater (Global Volcanism Program, 2012a, 2012b). For Telica, there are 45 features (AP 1.2) and for Nevado del Ruiz (AP 1.1) there are 36 features in total. There are a greater number of features for Telica due to the inclusion of features from individual event classifications and RSAM data.

Total event rates per day from network detections are used for both volcanoes. For the Telica data, two additional features derive from the automatic spectral classification of Low Frequency (LF) and High Frequency (HF) as defined by Rodgers et al. (2015a). Band ratio is defined as the base 2 log of the ratio of high-frequency to low-frequency energy (Rodgers et al., 2015a; c.f. Buurman and West, 2010). The distinction between high- and low-frequency bands is dependent on the typical frequencies of the volcanic system: for Telica, low-frequency activity is defined as 1–6 Hz, and high-frequency activity is defined as 6–11 Hz (Rodgers et al., 2015a). Dominant frequencies are obtained by recording the 5 peak frequencies from each event during the day. Peak amplitude is calculated from the maximum peak-peak amplitude of each event. Waveform standard deviation is obtained by calculating the width of the largest peak of the spectra for each event during the day. RSAM measurements are calculated hourly during the day for the Telica dataset. Multiplet information is the number of waveform families active on a given day, obtained by waveform cross-correlation using Peakmatch (Rodgers et al., 2015b).

From the categories of observations described above, features are calculated on a per-day basis by taking the mean, median, variance, minimum, maximum, 10th percentile, 90th percentile and change in mean from the previous day. For multiplets and event rates, only the per-day value and change in value from the previous day is calculated. The RSAM features are mean and variance of the per-hour readings and change in mean from the previous day. For a full list of features see Appendix 1.

For days in the time series with zero events, the whole day is omitted from the time series as no features can be extracted for this day. The gaps in the dataset could be filled using a method such as imputation in which missing data is replaced by a substitute, such as the mean of the whole dataset (Schafer and Graham, 2002). However, given that the days which have no associated data represent a small proportion of the dataset, we choose to leave these gaps within the time series.

Models can be limited by large quantities of features. High-dimensional systems (those with many features) are not ideal to work with: as the number of dimensions of data increases, the number of training examples required to train a consistent model grows exponentially (Bishop, 2006). This phenomenon is known as the “curse of dimensionality” (Bellman, 1961). For generalised linear models such as logistic regression, high-dimensional systems are especially poor to work with (Johnstone and Titterington, 2009). For this reason, we apply regularisation to the logistic regression model, to reduce the number of dimensions as input to the model. We use a technique known as the Least Absolute Shrinkage and Selection Operator (LASSO) to reduce the number of dimensions as input to the logistic regression models. The LASSO acts to penalise large coefficients in linear models, so that a smaller subset of the full set of features is chosen to model on for each dataset. A full discussion of the LASSO formulation is included in Hastie et al. (2001).

We use features derived from single-station seismic data, hence do not include derived event parameters such as location or depth. Seismic data is the only type of data which is input to the model. We do not use other observables, such as gas or deformation data. The end-date obtained by classification therefore corresponds to the end of seismicity associated with the eruption, and therefore represents the seismic end-date. Seismicity can often continue longer than the end of visible eruption, as it reflects the processes occurring at depth within the volcanic system. Seismic data is one of the most ubiquitous monitoring datasets collected at volcanoes and understanding the path to cessation of processes at depth is crucial in terms of understanding the end of eruptions, supporting the value of focussing only on seismic data for this preliminary study.

The features which we use, with the exception of those associated with RSAM, are derived from detected seismic events. The advantage of using these discrete features is that the method could be easily extended to seismic catalog data from other systems with waveforms attached. However, depending on the seismic characteristics of a given volcano, the inclusion of more features derived from continuous data (such as dominant tremor frequency) may be necessary.

### *2.6. Eruption classification*

Each day in the time series is classified independently from the other days. To pick out the large-scale changes in classification, we define a rolling threshold filter criterion for a day to be classified as eruptive, which is based on a moving average of the classifications. For a day to be classified as eruptive, the day itself and the seven days preceding that day must also be classified as eruptive. By applying this filter to the model output, classification of observations as eruptive is more conservative than if the results are left unfiltered. The choice to require 7 days of eruptive classification is made as the timescale of a week corresponds to typical timescales on which observations of volcanic activity are communicated to the public where the eruption circumstances are ongoing or chronic, for example in weekly reports of activity.

### 2.7. Quality assessment: decisiveness index

Our aim when classifying volcanic state is not necessarily to achieve maximum accuracy relative to GVP labels, due to the issue with GVP definition of volcanic state (discussed in Section 1). We therefore define an alternative index to model accuracy to evaluate our models. This index is a measure of how consistent or decisive the model classification is over the whole dataset, expressed as a percentage of the total number of days containing data. As we are looking to classify overall patterns of eruptive or non-eruptive activity, the decisiveness index favours classifications with less noise.

To define the decisiveness index D, we take the number of transitions between classes in our models ($Nt_M$; where a transition can be from non-eruptive to eruptive or eruptive to non-eruptive) and subtract the expected number of transitions corresponding to the number of eruptive periods in the dataset ($Nt_E$; where one eruptive period would have two transitions, at the beginning and end of the eruption) then normalise by the number of days contained within the data period ($N_d$). An index of 0 means that the number of transitions in the model are exactly equal to the number of assumed transitions. A larger index means that there is more inconsistency (i.e. more indecision) in the final model classification.

$$D = \frac{Nt_M - Nt_e}{N_d} * 100$$

The decisiveness index is reported as a percentage. The worst classification which one might produce would alternate between classes every day, and therefore contain a transition for each day. As the number of days is three orders of magnitude greater than the number of transitions in the datasets presented here, the number of transitions would be nearly equal to the total number of days and the decisiveness index will approach 100%. The decisiveness index is a method to evaluate which models make the most consistent classifications of eruptive state.

The definition of the decisiveness index is a method to evaluate which models make the most consistent classifications of eruptive state. This index makes no assumptions about when the transitions occur within the dataset. Therefore, the index is used in conjunction with comparison to visual observation of volcanic state in order to evaluate the success of the models presented within this paper.

## 3. Results

We independently trained 4 different classification models for each volcanic system, with each type of model trained and tested on each volcano separately. The analysis could be extended by training a model on several different seismic datasets, which would be a general classification model. However, a general model would require datasets from a greater variety of volcanic settings to ensure that the non-eruptive and eruptive distributions were well-characterised by the machine learning models.

### 3.1. Nevado del Ruiz

The results from each machine learning method are summarised in Table 1. For all 4 machine learning methods, the end-date of the first phase of eruption at Nevado del Ruiz obtained by classification of the data is later than the end-date contained in the GVP database. Fig. 4 illustrates the results from the SVM classification of Nevado del Ruiz data in a time series plot for the whole data period. Several observations can be made which are consistent features of all of the models summarised in Table 1, though we only plot the SVM model (Fig. 4) as it has the highest model accuracy of 82.6% (Table 1):

1. There is a sustained classification of non-eruptive activity before the beginning of eruptive activity, with only 1 pre-eruptive day erroneously classified as eruptive in 2007 for the SVM model.
2. The eruption end-date is 4–5 months later than the end-date recorded in GVP. Though the eruption end-date was recorded as the 12th July 2013 just the previous day, active ash emission was observed on the 11th July 2013, with further reports of gas and steam emission until November 2013 (Global Volcanism Program, 2017).
3. The second phase of eruption was classified as longer-lived than the GVP start- and end-dates would suggest. Though recorded in GVP as commencing in December 2014 and finishing in January 2015, the SVM classification of eruptive behaviour lasted from July 2014 to February 2015.

Table 1 summarises the results for the decisiveness index when applied to Nevado del Ruiz. It can be seen that the Gaussian process classifier has the best result for the decisiveness index with 3.13%, followed by SVM with 3.65%. Logistic regression and random forest classifications have a poorer score for the decisiveness index of 4.78% and 4.17% respectively. The range of D for the Nevado del Ruiz models is 1.65%.

Fig. 4 also contains information for the days in the Nevado del Ruiz dataset on which there were insufficient data to calculate features. Where there are many data gaps in the sequence, for example, during non-eruptive activity in 2007 or during the eruption in mid-2012, the classification is the same on either side of and during the data gap. From this observation we can determine that the decision to leave data gaps and not to fill them with a method such as imputation is justified (Section 2.5).

All of the models apart from random forest yield >70% accuracy (where the model result is compared to the GVP label of whether a day is eruptive or non-eruptive) after filtering. The greatest accuracy of 82.6% is achieved by using the SVM model. Logistic regression models have the second highest accuracy (79.6%). High model accuracy is therefore consistent over multiple types of classification, including

**Table 1**
Summary of results from all machine learning models applied to the Nevado del Ruiz and Telica dataset.

| Volcano | Method | Start-date for first phase (GVP) | End-date for first phase (GVP) | Start-date for first phase (model) | End-date for first phase (model) | Decisiveness index (%)[a] | Model accuracy (unfiltered) (%) | Model accuracy (filtered) (%) |
|---|---|---|---|---|---|---|---|---|
| Nevado del Ruiz | SVM | 22nd February 2012 | 12th July 2013 | February 2012 | November 2013 | 3.65 | 76.2 | 82.6 |
| | Logistic regression | | | February 2012 | December 2013 | 4.78 | 75.4 | 79.6 |
| | Random forest | | | February 2012 | November 2013 | 4.17 | 62.8 | 66.9 |
| | Gaussian process classifier (GPC) | | | February 2012 | November 2013 | 3.13 | 67.6 | 72.7 |
| Telica | SVM | 7th March 2011 | 14th June 2011 | March 2011 | August 2011 | 3.01 | 83.6 | 87.9 |
| | Logistic regression | | | March 2011 | October 2011 | 3.76 | 65.1 | 71.6 |
| | Random forest | | | March 2011 | August 2011 | 3.39 | 62. | 69.5 |
| | Gaussian process classifier (GPC) | | | March 2011 | August 2011 | 1.69 | 86.2 | 90.5 |

[a] Decisiveness Index is defined in Section 2.7. Lower D scores are better, and D is comparable across different datasets.

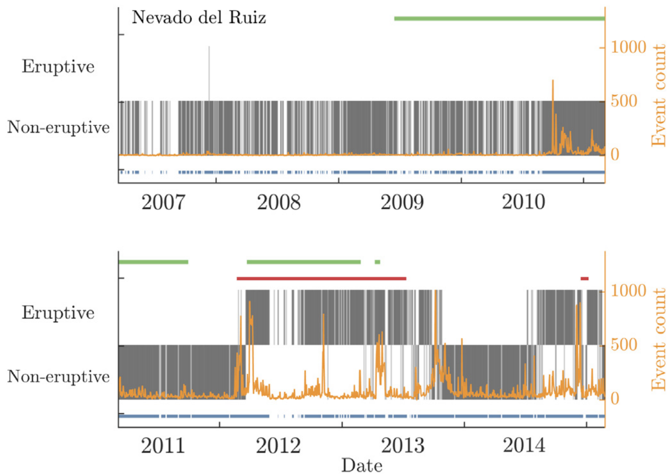


**Fig. 4.** Results from SVM classification on Nevado del Ruiz data over the study period: from 21st March 2007 to 6th March 2011 (top panel) and from 6th March 2011 to 25th February 2015 (bottom panel). Results of the classification are denoted by the grey rectangles where rectangles in the top half of each panel denote a classification of eruptive and rectangles in the bottom half of the panel denote a classification of non-eruptive. As each classification is made independently, consecutive days of the same classification together – such as non-eruptive classification in the top panel – illustrate the decisiveness of the classifier. Above the classification, the green horizontal line at the top of each panel denotes the timing of the training period and the red horizontal line second from the top of each panel denotes the timing of the eruption as recorded by GVP. The right axis and orange line within the plot denote daily event count for all events. Below the classification, the blue horizontal line denotes the days for which we could derive features (listed in AP 1.1). In the top panel, gaps in data were primarily due to low event count, whereas in the bottom panel during the eruption there is a gap which corresponds to instrument failure.

both non-linear and linear models. Random forest and Gaussian process classifiers have a similar accuracy.

Fig. 5 displays the Receiver Operating Characteristic (ROC) curves from all classification models applied to the Nevado del Ruiz dataset. ROC curves plot the false positive rate against the true positive rate for a binary classifier as the threshold of classification is changed. A better classifier will have a higher true positive rate at low false positive rate, as more points will be correctly classified. Better classifiers will also have a greater Area Under the Curve (AUC), which can be seen where the curve is higher than the diagonal line through the origin of the graph. From Fig. 5 it can be seen that logistic regression and random forest have a very similar structure, with an AUC of 0.89. The SVM has a slightly better performance at low false positive rate, but overall has a lower AUC of 0.85. The Gaussian process classifier has a relatively poor performance relative to the SVM, logistic regression and random forest, despite having a similar model accuracy to the random forest models.

Though good accuracy is achieved by the models, it should be highlighted that this represents a comparison of model output for each day compared to the GVP label of whether a day is eruptive or non-eruptive. We anticipate that the GVP labels are not entirely reliable due to uncertainty in the GVP definition of end of eruption (Phillipson et al., 2013). This error in labels leads to a number of false positives after the GVP end-date, where the day has been classified as eruptive by the models though the GVP label is non-eruptive.

Fig. 6 is a summary of the feature importance results for logistic regression, random forest and Gaussian process classification, the three methods for which this analysis is available. From the results we can observe that there is not much consistency between the three methods as to the most important features within the dataset, either in the group of features (e.g. Event rate, dominant frequencies) or the type of features within a group (e.g. Rate, Δ, mean; for a full list of features see Appendix 1). For the logistic regression classifier, multiplet rate per day is the feature that has the greatest importance.

### 3.2. Telica

Table 1 presents the summary of results of models run on Telica Volcano. As with the Nevado del Ruiz data, end-dates obtained by classification of Telica data for successful models (unsuccessful modelling is summarised below) are all approximately 2 months later than the end-date contained within the GVP database.

Gaussian process classification had the best value of decisiveness for data both for models applied to Telica data, and over both Nevado del Ruiz and Telica classifications overall, with a value of 1.69. Gaussian process classification also had the highest accuracy (87.9%) of all the models applied to Telica. SVM had the second-best value of decisiveness index of all the models (3.01%), followed by random forest (3.39%) and logistic regression (3.76%) models. The range of D for models on the Telica dataset is 2.07%.

Fig. 7 summarises the classification from the Gaussian process classifier on the data from Telica, the model which had the highest accuracy and best decisiveness index of any model applied in this study. The attributes of the classification which we identify consistently over all models for the Telica time series are as follows:

1. For all of the machine learning approaches, the end-date inferred from the models was later than the end-date recorded by GVP. The end-date for the eruptive phase at Telica was 14th June 2011, whereas the end-dates obtained by successful classifications were all in August 2011.
2. After the end-date of the eruption inferred by the models, there are 2–3 short periods of elevated event count in November 2011 and March 2012 which correspond to classifications of eruptive activity.

Fig. 8 summarises the ROC curves for all of the methods. Although Gaussian process classification has the best values for decisiveness and accuracy, this model has the lowest Area Under the Curve (AUC) at 0.64. SVM and logistic regression classifiers have similar AUC values at 0.94 and 0.91 respectively. Overall, the AUC values for Telica are less smooth than for Nevado del Ruiz, a consequence of the smaller dataset that we have for Telica.

Fig. 9 is a summary of the feature importance results for the methods applied to Telica data. Event rate features do not have high values for variable importance. This finding can be confirmed by observing the classification in Fig. 7: the overall event rate spans a similar range (from 0 to 500 events per day) during classification of both eruptive

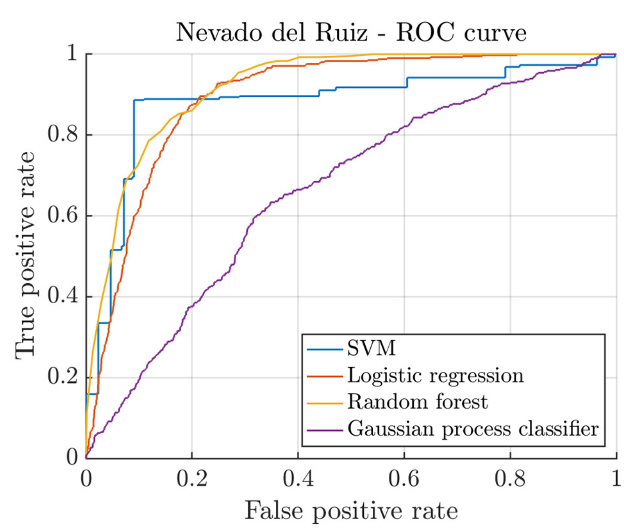


**Fig. 5.** Receiver Operating Characteristic (ROC) curve for all of the methods applied to the Nevado del Ruiz data. The ROC curve plots the true positive rate, false positive rate and AUC value. SVM, logistic regression and random forest have similar performance, compared to Gaussian process classification.

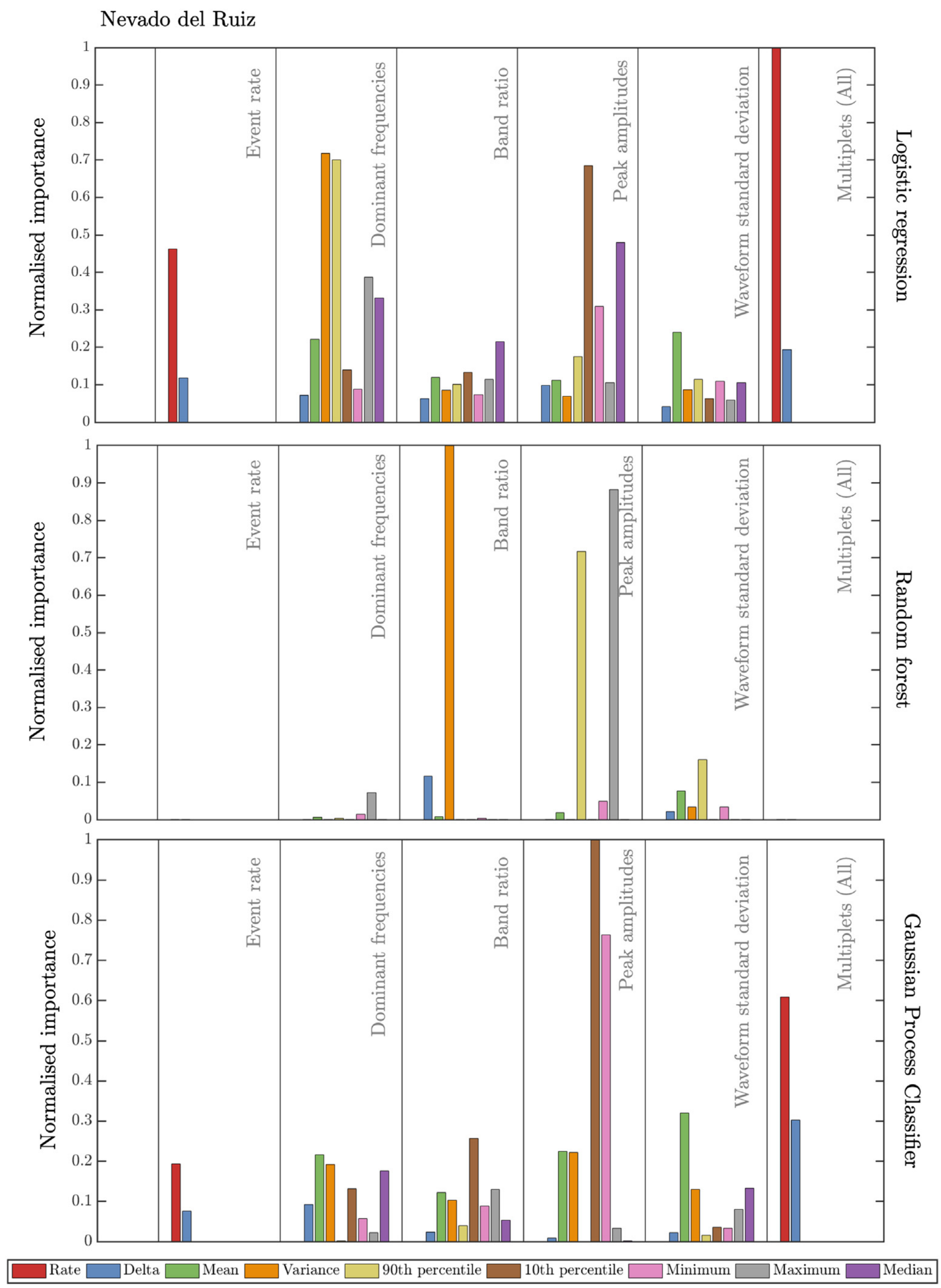


**Fig. 6.** Results from feature importance analysis methods on Nevado del Ruiz data. Feature importance is derived from logistic regression (top), random forest (middle) and Gaussian process classification (bottom). The y-axis denotes absolute importance which varies depending on the models, normalised by the maximum value for the model. There is a much greater range in importance for the Gaussian process classification than for the random forest model. Vertical lines separate the categories of features. For a full list of input features see Appendix 1.

and non-eruptive activity by our models. As seen for the models applied to Nevado del Ruiz, there is not much consistency in the individual features which are associated with a higher importance.

### 3.3. Training models with data from the end of volcanic eruption

The Nevado del Ruiz models are trained using two training periods: a period before the beginning of the first eruption and a period during the first phase of the eruption (Fig. 4). In these models we do not train over any transitions between eruptive and non-eruptive behaviour. We now extend the modelling to train over two extra periods using SVM:

(i) Over the GVP start-date, with training period 23rd June 2009–23rd December 2011 (non-eruptive) and 2nd February 2012–24th March 2013 (GVP start of eruption).
(ii) Over the GVP end-date, with training period 23rd June 2009–23rd September 2011 (non-eruptive) and 29th December 2012–26th September 2013 (GVP eruption end).

We would expect the models to successfully classify non-eruptive and eruptive behaviour if the GVP dates represent reliable labels of the transitions in the dataset.

The results from model (i) are very similar to those presented for a model trained over no transitions in behaviour (Fig. 4). However, for

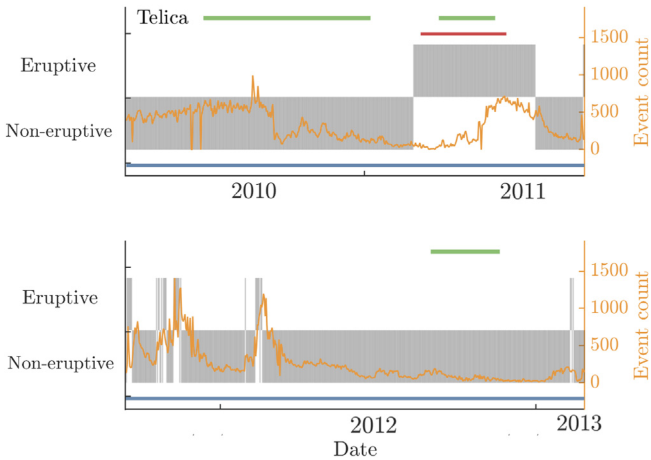


**Fig. 7.** Results from GPC classification on Telica data over the study period: from 1st April 2010 to 6th October 2011 (top panel) and from 7th October 2011 to 18th March 2012 (bottom panel). Results of the classification are denoted by the grey rectangles where rectangles in the top half of each panel denote a classification of eruptive and rectangles in the bottom half of the panel denote a classification of non-eruptive. As each classification is made independently, consecutive days of the same classification together – such as non-eruptive classification in the top panel – illustrate the decisiveness of the classifier. Above the classification, the green horizontal line at the top of each panel denotes the timing of the training period and the red horizontal line second from the top of each panel denotes the timing of the eruption as recorded by GVP. The right axis and yellow line within the plot denote daily event count for all events. Below the classification, the blue horizontal line denotes the days for which we could derive features. There are no significant periods of data shortage throughout the Telica data time period.

model (ii), we obtain a very poor classification: eruptive behaviour is not classified until July 2012 despite visual evidence of eruption from February 2012 onwards (Global Volcanism Program, 2012a, 2012b). Moreover, no activity during the second phase of eruption is classified as eruptive. We conclude from this result that the eruption end-date recorded in GVP does not provide a reliable label for the transition between eruptive and non-eruptive behaviour.

## 4. Discussion

### *4.1. Classification compared to visual observations*

The results presented in Section 3 suggest that in both phases of eruption at Nevado del Ruiz the classification of eruptive activity is more prolonged than the GVP start- and end-dates would suggest. A possible reason for the discrepancy between model classification and GVP eruption duration is that GVP classifications are based on visual observation of volcanic activity, whereas we are running models on the seismicity, with 7-day rolling window filtering. Seismicity can indicate processes occurring at depths of several kilometres within the volcanic system (Moran et al., 2011), which it is reasonable to expect would continue after visual signals of volcanic eruption had ended. The classification of eruption until November 2013 (Table 1) could represent continued or declining seismogenic processes at depth during the declining phase of the eruption. In this respect, the seismic end-dates presented here are hypothesised to represent the most generous bound on the end-date of the eruption.

In Fig. 10 we compare the event rate and model classification to the alert level recorded at Nevado del Ruiz for the duration of the data period, as event rate is a commonly-used parameter for investigating volcanic state. The majority of alert level changes were concentrated in the period leading up to eruption, and the first seven months of eruptive activity. There are no alert level changes following 5th September 2012, on which date the alert level was downgraded from II (Orange) to III (Yellow) (Global Volcanism Program, 2012a, 2012b). The alert level changes are therefore too coarse to provide insights into the processes occurring at the end of the eruption.

Fig. 10 also summarises the event rate and recorded ash emissions according to weekly reports and confirmed visual reports of ash emission (Appendix 2; SGC). The classification of the second phase of eruptive activity precedes the ash emission during July 2014, and continues until further ash emission during November 2014. Our model results show a high correlation of eruptive classification with ash emission during the second phase of volcanic eruption recorded from weekly reports from the Manizales observatory and observatory records (Londono and Galvis, 2018), having trained our model on the seismic signals associated with ash emissions during the first phase of volcanic activity.

The good agreement between classification and observation can also be noted with a comparison to event rate: though the GVP end-date of 11th July 2013 coincides with a consistent low event rate for Nevado del Ruiz, our model continues to classify behaviour as eruptive for 4 further months, which spans a spike in event rate to 1000 events per day in September–October 2011, and culminated in ash emission at the end of November (Global Volcanism Program, 2017).

In the Telica results, we observed deviations between our model classification and that of the GVP in terms of end-dates from August–October 2011, in addition to two periods of elevated event rate from October–November 2011 and February–March 2012. These periods are united by records of “jet-turbine” sounds from the crater initially reported by nearby communities (Global Volcanism Program, 2012a, 2012b), and by crater incandescence and gas emission in September 2011 and February 2012. Overall, comparison to visual observations is more difficult for Telica data as visual reports of activity are released monthly rather than weekly in Colombia, and there is no system of alert level classifications.

For both volcanoes, unfiltered models yielded classification of eruptive activity before the GVP eruption beginning date. However, these classifications were not sufficiently consistent to remain as eruptive after applying the 7-day rolling filter. Further work is required to see how these results could be analysed for the case of pre-emptive classification of eruption to evaluate whether those classifications of eruptive activity are truly eruptive or represent a false positive result.

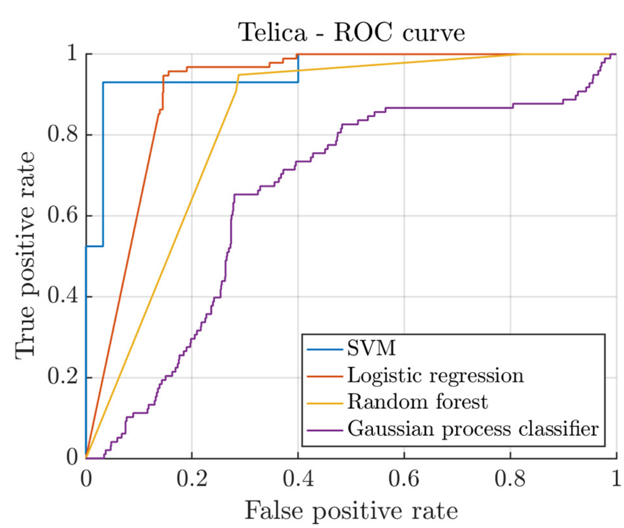


**Fig. 8.** Receiver Operating Characteristic (ROC) curve for all of the methods applied to the Telica data. The ROC curve plots the true positive rate, false positive rate and Area Under Curve (AUC) value. SVM, logistic regression and random forest have similar performance, compared to Gaussian process classification.

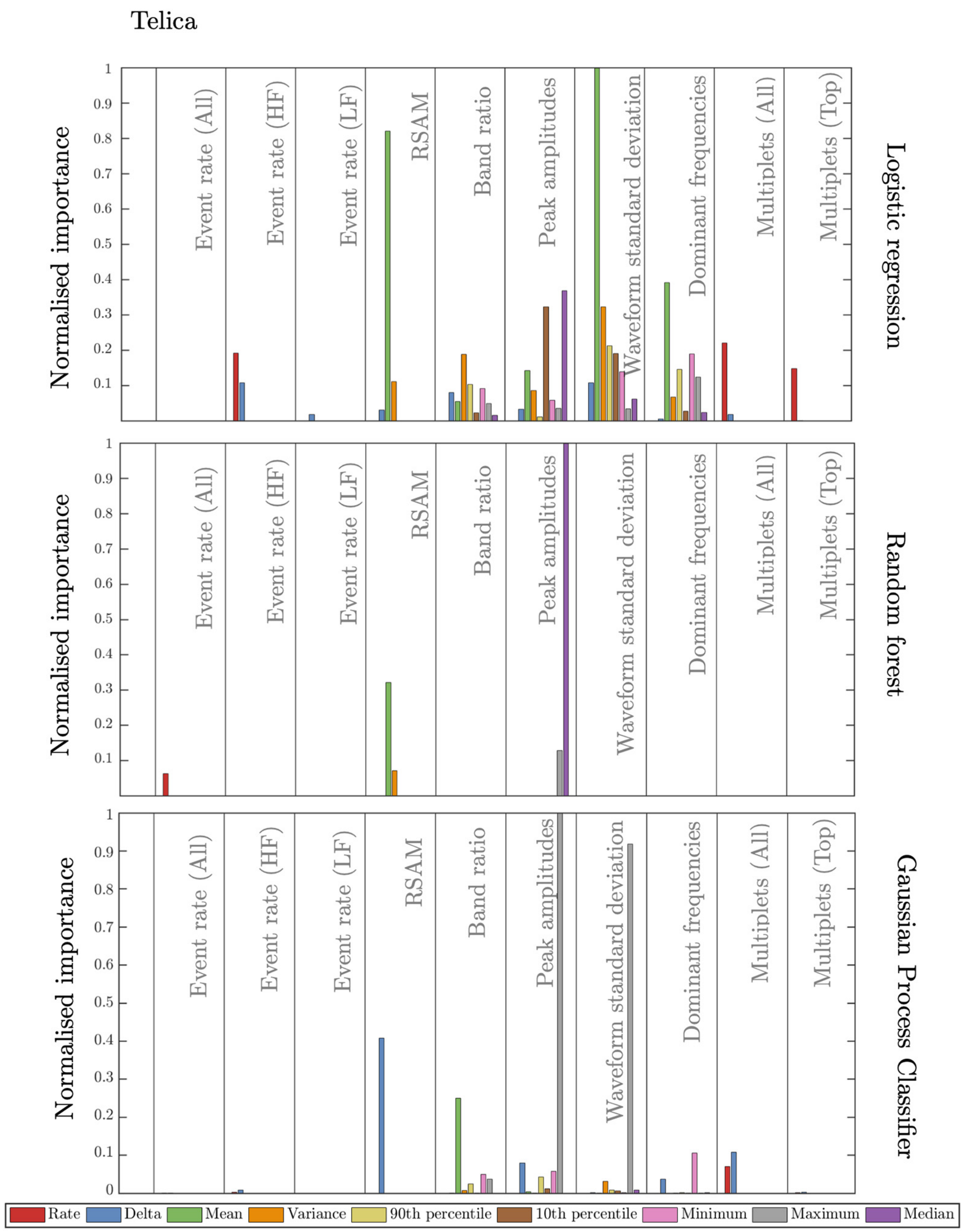


**Fig. 9.** Results from feature importance analysis methods on Telica data. Feature importance is derived from logistic regression (top), random forest (middle) and Gaussian process classification (bottom). The y-axis denotes absolute importance which varies depending on the models, normalised by the maximum value for the model. Vertical lines separate the categories of features.

### *4.2. Possible application of methods and future work*

In this study we have shown that retrospective classification of volcanic activity can yield timing of change with a greater correspondence to heightened activity and ash emission than end-dates denoted by visual activity judged to be the last of an eruptive phase (Fig. 10). The coarseness of alert level changes and the success of the classification model discussed in Section 4.1 presents the possibility to use these machine learning classification methods to identify potential start- and end-dates for seismically monitored volcanoes. Estimates for seismic end of eruption obtained by classification could be combined with other indicators of activity, such as deformation, thermal or gas data, to make a judgement on whether the eruption has ended.

In remote locations, or where conditions are unfavourable for making visual observations, classification of seismic data could yield a consistent method for determining transitions in eruptive state in the absence of other evidence. The work presented here is an example of how to distinguish between eruptive and non-eruptive seismic activity using information from one seismic station alone. For example, though no official visual observations of Telica volcano coincided with our eruptive classifications in late 2011 or early 2012, nearby communities reported jet-like sounds which coincided with these periods. Further applications for this method could include monitoring volcanoes in remote locations where regular visual observations of the volcano are not practical.

The end-dates yielded by the successful methods are between 2 and 4 months later than the end-dates judged to be the last visual indication of eruption (Section 3). This finding is in agreement with the generic 90-day rule discussed in Section 1. The validation of this months-long timescale of eruption cessation, consistent across volcanoes of differing

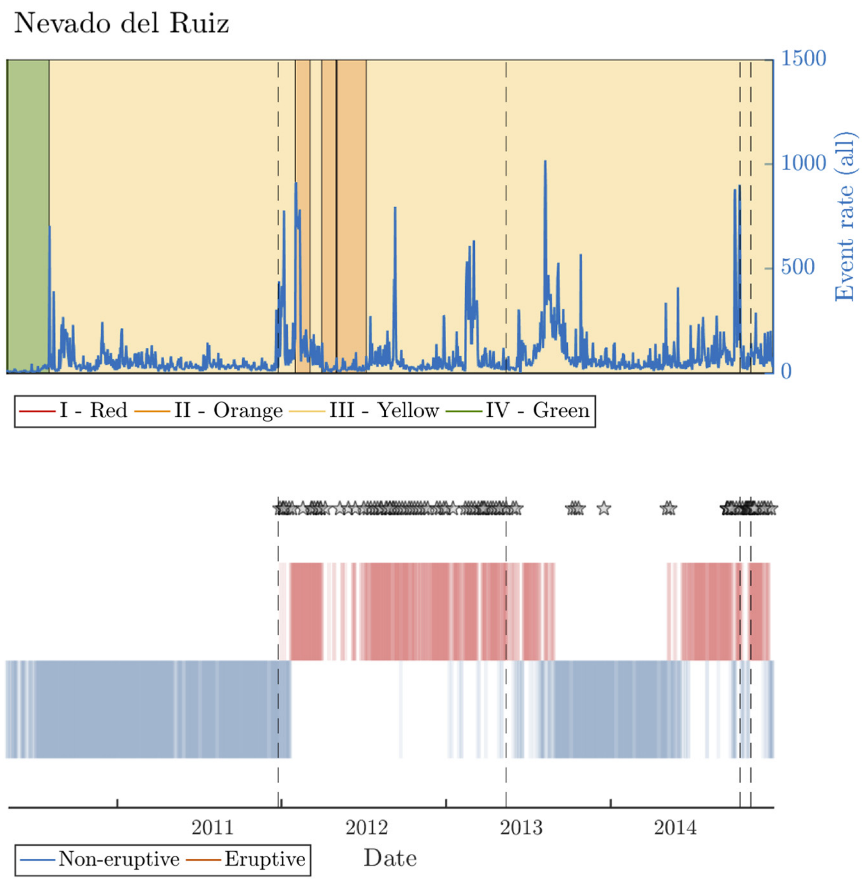


**Fig. 10.** Comparison of the event rate and alert level (top panel) and SVM classification and recorded ash emissions (bottom panel) at Nevado del Ruiz between the dates of 28th June 2010 to 25th February 2015. Vertical dashed lines in each plot indicate the GVP start- and end-dates of the two phases of eruption during the data period. Stars indicate confirmed ash emissions. Rectangles in the bottom panel represent the classification from SVM (as in Fig. 4) where blue rectangles are non-eruptive classification and red rectangles are eruptive classification. We choose to plot the SVM as it was the best-performing classifier for Nevado del Ruiz. The alert level was consistently green from the beginning of the data period to the first alert level change on 30th September 2010.

eruption style, provides new insights on the physical processes which govern changes in seismicity at the end of volcanic eruption. These processes could include magma withdrawal or relaxation, or rheological changes in the magma (which could in turn be due to, for example, increased crystallinity or decreased gas content).

This study is a proof-of-concept of the classification of time series for the detection of large-scale changes in eruptive systems, and further work would be required to make classifications on a real-time basis. In addition, to apply these techniques to a greater number of volcanoes requires representative seismic data during eruptive and non-eruptive periods to train new models. These models also do not give any indication of the type or severity of potential eruptive activity when the classification is eruptive. Further work, incorporating a more diverse range of time-series observations and datasets into the modelling, is needed to investigate the sensitivity of the modelled end-dates determination to the nature and variety of datasets used.

The models presented in this study could be extended by defining 3 classes for model training and classification. Here, the non-eruptive class could be split into two classes: one which represents a background class and one which represents a precursory class to eruption. However, to make this extension it would be necessary to have an independent data stream to reliably distinguish between the background and precursory states, such as a gas time series hence this analysis is not presented here.

### *4.3. Failure of logistic regression*

For both of the datasets, logistic regression performed poorly relative to the other methods applied to the datasets, with end-dates 1–3 months after the other classification models which were all in agreement. As logistic regression involves linear modelling to define the classification, failure to characterise the overall volcanic state indicates that the underlying relationships are non-linear, even when features are removed from the dataset through regularisation. The processes governing volcanic eruption have been previously described as “nonlinear and stochastic” (Sparks, 2003), which could account for the failure of the logistic regression approach here.

### *4.4. Feature importance: feature ranking*

The feature importance results from the 33 methods which yield full feature importance results are summarised in Fig. 11; here the top 10 features ranked as most important in determining the transition between eruptive and non-eruptive for the Nevado del Ruiz and Telica datasets are plotted. As there may be orders of magnitude between the importance score for these two methods, it is better to value the very top features. The groups of features with high ranks for both methods are dominant frequencies, band ratio and waveform standard deviation (Nevado del Ruiz), and peak amplitude and dominant frequencies (Telica). Dominant frequencies rank in the top 5 features for all methods and for both volcanoes.

Though daily event rate is widely used as a parameter for determining changes in volcanic activity, the daily (total) event rate only appears in the top 10 features for one method at Nevado del Ruiz, and change in event rate from the previous day does not appear as a top 10 ranked feature at all. Increases in VT seismicity has been demonstrated as a common precursor to volcanic eruption at closed volcanic systems,

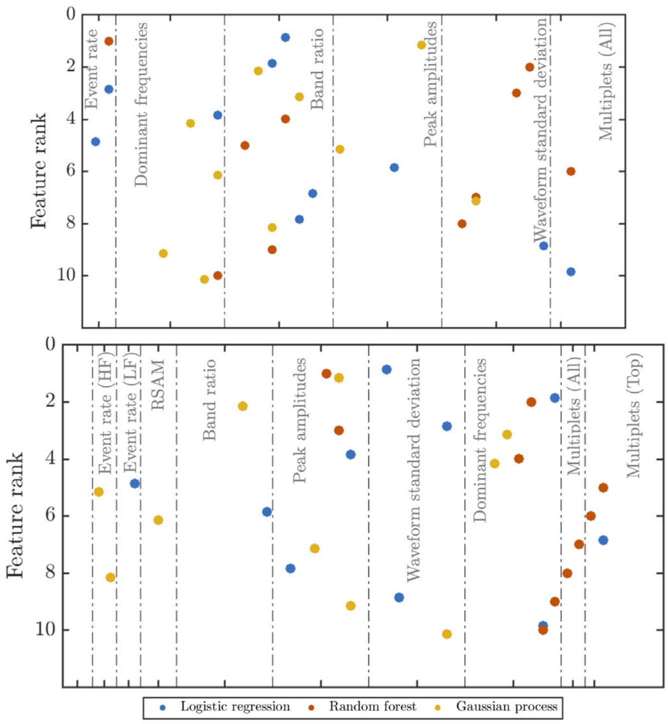


**Fig. 11.** Plot of the top 10 ranked features from each feature importance method on Nevado del Ruiz (top) and Telica (bottom). Vertical dashed lines represent the distinction between groups of features (labelled on the x-axis).

particularly at previously dormant volcanoes (Cameron et al., 2018). However, studies at basaltic systems including Kilauea Volcano, Hawaii (Chastin and Main, 2003) and Piton de la Fournaise Volcano, Réunion Island (Collombet et al., 2003) have found that these precursory increases in VT seismicity are often either not present, or do not always lead to an eruption. Increases in VT seismicity ending in no eruption have also been documented as "failed eruptions" (Moran et al., 2011), representing a challenge in determining whether a phase of unrest will lead to eruption. From the results presented here, we cannot conclusively identify a category of features which distinguishes between eruptive and non-eruptive behaviour for both volcanoes.

### 4.5. Verification of model assumptions

From the successful classification of non-eruptive and eruptive activity presented in Section 4, we conclude that it is at least approximately correct to make the assumption that the data are Independent and Identically Distributed (IID) for the relatively small VEI and short-lived (<3–5 year) eruptions considered in this study. If similar methods are applied on different timescales, this assumption may not be valid. If data were binned on a shorter timescale than daily observations, it may not be appropriate to make this assumption as individual events in certain cases can be quasi-periodic, i.e., not independent from each other (Ignatieva et al., 2018). Following a catastrophic eruption, the assumption that each day is drawn from an identical distribution may not hold. Seismicity has been shown to reflect processes within the conduit (e.g., Jousset et al., 2003), which in turn can be eroded by several processes during eruption including volcanic tremor or wall collapse (Macedonio et al., 1994). Observations of precursory seismic activity at Kelud Volcano, Indonesia preceding the 2007 and 2014 eruptions found significant differences in seismic characteristics before both eruptions, which is consistent with the contrasting eruption dynamics of the two events. (Hidayati et al., 2018). Transition periods between eruptive and non-eruptive data may last on timescales from hours to weeks (Carniel et al., 2003; Ripepe et al., 2002), and behaviour during these transitions may represent a different mode of the volcanic system (Connor et al., 2003; Rodgers et al., 2016). Though the successful models presented here indicate that there are no significant transition periods within the data periods included in this study, for volcanoes with transition periods on longer timescales (such as days – weeks) the transition period may need to be defined as a separate, third class of activity.

## 5. Conclusions

Machine learning methods can successfully classify overall patterns of eruptive and non-eruptive behaviour in seismic time series. This study is the first to apply machine learning techniques to single-station seismic data to classify overall volcanic state as eruptive or non-eruptive. We define a decisiveness index D to evaluate classification of eruptive state based on the consistency of classification, which is comparable across datasets. Our models have a high agreement in terms of eruptive classification with visual indicators of eruption, such as ash emissions. The date of the eruption end is found to be consistently later than the date recorded in GVP, by approximately 60–120 days. This finding is in agreement with previous, non-physical definitions of end of volcanic eruption, such as the 90-day rule for determining the timing of eruption end (Simkin and Siebert, 1994). Classification of eruptive and non-eruptive data could be applied to seismic time series to determine when end of eruption occurred, in the absence of conclusive visual observations. Support Vector Machine and Gaussian Process Classifiers were the most successful classification models applied to Nevado del Ruiz and Telica respectively. Logistic regression, a linear classifier, had lower classification accuracy and decisiveness for both datasets, which could be due to non-linearity in the data. Feature importance methods identified little consistency between the most important seismic features used as model inputs. Work on a larger number and variety of datasets is necessary to determine whether these most important features are consistent between volcanoes, or between volcanoes with similar eruption styles or tectonic settings.

## CRediT authorship contribution statement

**Grace F. Manley:** Conceptualization, Investigation, Formal analysis, Writing - original draft, Writing - review & editing. **David M. Pyle:** Conceptualization, Supervision, Writing - review & editing. **Tamsin A. Mather:** Conceptualization, Supervision, Writing - review & editing. **Mel Rodgers:** Conceptualization, Data curation, Supervision, Writing - review & editing. **David A. Clifton:** Conceptualization, Supervision, Writing - review & editing. **Benjamin G. Stokell:** Data curation, Writing - review & editing. **Glenn Thompson:** Writing - review & editing. **John Makario Londoño:** Data curation, Writing - review & editing. **Diana C. Roman:** Writing - review & editing.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgements and data statement

Telica data was collected by the TESAND network (NSF EAR-0911366 to D. Roman and P. LaFemina). Nevado del Ruiz data was obtained from the Observatorio Vulcanológico y Sismológico de Manizales, Servicio Geológico Colombiano. This research has been supported by a Natural Environment Research Council(NERC) studentship (NE/L002612/1). Contributions to the work were facilitated by a NERC Research Experience Placement to the British Geological Survey. Mather and Pyle acknowledge support from NERC/ESRC grants NE/J020001/1 and NE/J020052/1 (STREVA). Two anonymous reviewers are thanked for their helpful comments on an earlier version of the paper.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jvolgeores.2020.106917.

## References


Anantrasirichai, N., Biggs, J., Albino, F., Hill, P., Bull, D., 2018. Application of machine learning to classification of volcanic deformation in routinely generated InSAR data. J. Geophys. Res. Solid Earth 123 (8), 6592–6606.

Apolloni, B., 2009. Support vector machines and MLP for automatic classification of seismic signals at Stromboli volcano. Neural Nets WIRN09: Proceedings of the 19th Italian Workshop on Neural Nets, Vietri Sul Mare, Salerno, Italy May 28–30 2009. 204. IOS Press, p. 116.

Barclay, J., Few, F., Armijos, M.T., Phillips, J.C., Pyle, D.M., Hicks, A.J., Brown, S.K., Robertson, R.E.A., 2019. Livelihoods, wellbeing and the risk to life during volcanic eruptions. Front. Earth Sci. 7, 205. https://doi.org/10.3389/feart.2019.00205.

Barmin, A., Melnik, O., Sparks, R.S.J., 2002. Periodic behavior in lava dome eruptions. Earth Planet. Sci. Lett. 199 (1), 173–184.

Battaglia, J., Aki, K., 2003. Location of seismic events and eruptive fissures on the Piton de la Fournaise volcano using seismic amplitudes. J. Geophys. Res. Solid Earth 108 (B8), 2364. https://doi.org/10.1029/2002JB002193.

Belgiu, M., Drăguţ, L., 2016. Random forest in remote sensing: a review of applications and future directions. ISPRS J. Photogramm. Remote Sens. 114, 24–31.

Bellman, R.E., 1961. Adaptive Control Processes: A Guided Tour. Princeton University Press, Princeton, New Jersey, USA.

Bicego, M., Acosta-Muñoz, C., Orozco-Alzate, M., 2013. Classification of seismic volcanic signals using hidden-Markov-model-based generative embeddings. IEEE Trans. Geosci. Remote Sens. 51 (6), 3400–3409.

Bishop, C.M., 2006. Pattern Recognition and Machine Learning. Springer, New York.

Bonny, E., Wright, R., 2017. Predicting the end of lava flow-forming eruptions from space. Bull. Volcanol. 79 (7), 52. https://doi.org/10.1007/s00445-017-1134-8.

Breiman, L., Friedman, R.A., Olshen, R.A., Stone, C.G., 1984. Classification and Regression Trees. Wadsworth, Pacific Grove, CA.

Buurman, H., West, M., 2010. Seismic precursors to volcanic explosions during the 2006 eruption of Augustine Volcano. In: Power, J., Coombs, M., Freymueller, J. (Eds.), The 2006 Eruption of Augustine Volcano. U.S. Geological Survey Professional Paper 1769, Alaska (U.S. Geological Survey Professional Paper 1769).

Cameron, C.E., Prejean, S.G., Coombs, M.L., Wallace, K.L., Power, J.A., Roman, D.C., 2018. Alaska volcano observatory alert and forecasting timeliness: 1989–2017. Front. Earth Sci. 6, 86. https://doi.org/10.3389/feart.2018.00086.

Carniel, R., 2014. Characterization of volcanic regimes and identification of significant transitions using geophysical data: a review. Bull. Volcanol. 76 (8), 848. https://doi.org/10.1007/s00445-014-0848-0.

Carniel, R., Di Cecca, M., Rouland, D., 2003. Ambrym, Vanuatu (July–August 2000): spectral and dynamical transitions on the hours-to-days timescale. J. Volcanol. Geotherm. Res. 128 (1–3), 1–13.

Chang, C.C., Lin, C.J., 2011. LIBSVM: a library for support vector machines. ACM Trans. Intell. Syst. Technol. (TIST) 2 (3), 27.

Chastin, S.F., Main, I.G., 2003. Statistical analysis of daily seismic event rate as a precursor to volcanic eruptions. Geophys. Res. Lett. 30 (13). https://doi.org/10.1029/2003GL016900.

Clifton, L., Clifton, D.A., Pimentel, M.A.F., Watkinson, P.J., Tarassenko, L., 2014. Predictive monitoring of mobile patients by combining clinical observations with data from wearable sensors. IEEE J. Biomed. Health Inform. 18 (3), 722–730 2014.

Collombet, M., Grasso, J.R., Ferrazzini, V., 2003. Seismicity rate before eruptions on Piton de la Fournaise volcano: Implications for eruption dynamics. Geophys. Res. Lett. 30 (21), 2099. https://doi.org/10.1029/2003GL017494.

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. Geophys. Res. Lett. 30 (13).

Cover, T.M., Thomas, J.A., 2006. Elements of Information Theory. 2nd edition. Wiley-Interscience, NJ.

Cuellar-Rodriguez, J.V., Ramirez-Lopez, C., 1987. Descripcion de los volcanes Colombianos. Rev CIAF, Bogota, pp. 189–222.

Curilem, G., Vergara, J., Fuentealba, G., Acuña, G., Chacón, M., 2009. Classification of seismic signals at Villarrica volcano (Chile) using neural networks and genetic algorithms. J. Volcanol. Geotherm. Res. 180 (1), 1–8.

De la Cruz-Reyna, S., Tilling, R.I., Valdés-González, C., 2017. Challenges in Responding to a Sustained, Continuing Volcanic Crisis: The Case of Popocatepétl Volcano, Mexico, 1994–Present.

Dzurisin, D., Moran, S.C., Lisowski, M., Schilling, S.P., Anderson, K.R., Werner, C., 2015. The 2004–2008 dome-building eruption at Mount St. Helens, Washington: epilogue. Bull. Volcanol. 77 (17). https://doi.org/10.1007/s00445-015-0973-4.

Flower, V.J., Oommen, T., Carn, S.A., 2016. Improving global detection of volcanic eruptions using the ozone monitoring Instrument (OMI). Atmos. Measurement Techniques 9 (11), 5487–5498.

Geirsson, H., Rodgers, M., LaFemina, P., Witter, M., Roman, D., Muñoz, A., Tenorio, V., Alvarez, J., Jacobo, V.C., Nilsson, D., Galle, B., 2014. Multidisciplinary observations of the 2011 explosive eruption of Telica volcano, Nicaragua: implications for the dynamics of low-explosivity ash eruptions. J. Volcanol. Geotherm. Res. 271, 55–69.

Global Volcanism Program, 2011. Report on Telica (Nicaragua). In: Sennert, S.K. (Ed.), Weekly Volcanic Activity Report, 11 May–17 May 2011. Smithsonian Institution and US Geological Survey.

Global Volcanism Program, 2012a. Report on Nevado del Ruiz (Colombia). In: Sennert, S.K. (Ed.), Weekly Volcanic Activity Report, 7 March–13 March 2012. Smithsonian Institution and US Geological Survey.

Global Volcanism Program, 2012b. Report on Telica (Nicaragua). In: Sennert, S.K. (Ed.), Weekly Volcanic Activity Report, 12 September–18 September 2012. Smithsonian Institution and US Geological Survey.

Global Volcanism Program, 2017. Report on Nevado del Ruiz (Colombia). In: Venzke, E. (Ed.), Bulletin of the Global Volcanism Network. 42. Smithsonian Institution, p. 6.

Hastie, T., Tibshirani, R., Friedman, J., 2001. The Elements of Statistical Learning. Springer, New York. NY.

Hicks, A., Few, R., 2015. Trajectories of social vulnerability during the Soufrière Hills volcanic crisis. J. Appl. Volcanol. 4, 10. https://doi.org/10.1186/s13617-015-0029-7.

Hidayati, S., Triastuty, H., Mulyana, I., Adi, S., Ishihara, K., Basuki, A., Kuswandarto, H., Priyanto, B., Solikhin, A., 2018. Differences in the seismicity preceding the 2007 and 2014 eruptions of Kelud volcano, Indonesia. J. Volcanol. Geotherm. Res. 382, 50–67. https://doi.org/10.1016/j.jvolgeores.2018.10.017.

Ignatieva, A., Bell, A., Worton, B., 2018. Point process models for quasi-periodic volcanic earthquakes. Stat. Volcanol. 4. https://doi.org/10.5038/2163-338X.4.2.

Johnstone, I.M., Titterington, D.M., 2009. Statistical challenges of high-dimensional data. Phil. Trans. R. Soc. A 367. https://doi.org/10.1098/rsta.2009.0159.

Jousset, P., Neuberg, J., Sturton, S., 2003. Modelling the time-dependent frequency content of low-frequency volcanic earthquakes. J. Volcanol. Geotherm. Res. 128 (1–3), 201–223.

Lacroix, A., 1908. La montagne Pelée après ses éruptions, avec observations sur les éruptions du Vésuve en 79 et en 1906. Masson et cie.

Lamb, O.D., Varley, N.R., Mather, T.A., Pyle, D.M., Smith, P.J., Liu, E.J., 2014. Multiple timescales of cyclical behaviour observed at two dome-forming eruptions. J. Volcanol. Geotherm. Res. 284, 106–121.

Langer, H., Falsaperla, S., Powell, T., Thompson, G., 2006. Automatic classification and a-posteriori analysis of seismic event identification at Soufriere Hills volcano, Montserrat. J. Volcanol. Geotherm. Res. 153 (1–2), 1–10.

Londono, J.M., 2016. Evidence of recent deep magmatic activity at Cerro Bravo-Cerro Machín volcanic complex, central Colombia. Implications for future volcanic activity at Nevado del Ruiz, Cerro Machín and other volcanoes. J. Volcanol. Geotherm. Res. 324, 156–168.

Londono, J.M., Galvis, B., 2018. Seismic data, photographic images and physical modeling of volcanic plumes as a tool for monitoring the activity of Nevado del Ruiz Volcano, Colombia. Front. Earth Sci. 6, 162. https://doi.org/10.3389/feart.2018.00162.

Lowe, D.R., Williams, S.N., Leigh, H., Connor, C.B., Gemmell, J.B., Stoiber, R.E., 1986. Lahars initiated by the 13 November 1985 eruption of Nevado del Ruiz, Colombia. Nature 324 (6092), 51.

Macedonio, G., Dobran, F., Neri, A., 1994. Erosion processes in volcanic conduits and application to the AD 79 eruption of Vesuvius. Earth Planet. Sci. Lett. 121 (1–2), 137–152.

Maggi, A., Ferrazzini, V., Hibert, C., Beauducel, F., Boissier, P., Amemoutou, A., 2017. Implementation of a multistation approach for automated event classification at Piton de la Fournaise volcano. Seismol. Res. Lett. 88 (3), 878–891.

Malfante, M., Dalla Mura, M., Mars, J.I., Métaxian, J.P., Macedo, O., Inza, A., 2018. Automatic classification of volcano seismic signatures. J. Geophys. Res. Solid Earth 123 (12), 10–645.

Marzocchi, W., Woo, G., 2007. Probabilistic eruption forecasting and the call for an evacuation. Geophys. Res. Lett. 34, 22.

McCullagh, P., Nelder, J.A., 1989. Generalized Linear Models. 37. CRC press.

McNutt, S.R., 1996. Seismic monitoring and eruption forecasting of volcanoes: a review of the state-of-the-art and case histories. Monitoring and Mitigation of Volcano Hazards. Springer, Berlin, Heidelberg, pp. 99–146.

Moran, S.C., Newhall, C., Roman, D.C., 2011. Failed magmatic eruptions: late-stage cessation of magma ascent. Bull. Volcanol. 73 (2), 115–122.

Mountrakis, G., Im, J., Ogole, C., 2011. Support vector machines in remote sensing: a review. ISPRS J. Photogramm. Remote Sens. 66 (3), 247–259.

Naranjo, J.L., Sigurdsson, H., Carey, S.N., Fritz, W., 1986. Eruption of the Nevado del Ruiz volcano, Colombia, on 13 November 1985: tephra fall and lahars. Science 233 (4767), 961–963.

National Academies of Sciences, Engineering, and Medicine, 2017. Volcanic Eruptions and Their Repose, Unrest, Precursors, and Timing. The National Academies Press, Washington, DC https://doi.org/10.17226/24650.

Phillipson, G., Sobradelo, R., Gottsmann, J., 2013. Global volcanic unrest in the 21st century: an analysis of the first decade. J. Volcanol. Geotherm. Res. 264, 183–196.

Pradhan, B., Lee, S., 2010. Landslide susceptibility assessment and factor effect analysis: backpropagation artificial neural networks and their comparison with frequency ratio and bivariate logistic regression modelling. Environ. Model Softw. 25 (6), 747–759.

Ren, C.X., Peltier, A., Ferrazzini, V., Rouet-Leduc, B., Johnson, P.A., Brenguier, F., 2020. Machine learning reveals the seismic signature of eruptive behavior at piton de la fournaise volcano. Geophys. Res. Lett. 47 (3), e2019GL085523. https://doi.org/10.1029/2019GL085523.

Ripepe, M., Harris, A.J., Carniel, R., 2002. Thermal, seismic and infrasonic evidences of variable degassing rates at Stromboli volcano. J. Volcanol. Geotherm. Res. 118 (3–4), 285–297.

Rodgers, M., Roman, D.C., Geirsson, H., LaFemina, P., Muñoz, A., Guzman, C., Tenorio, V., 2013. Seismicity accompanying the 1999 eruptive episode at Telica Volcano, Nicaragua. J. Volcanol. Geotherm. Res. 265, 39–51.

Rodgers, M., Roman, D.C., Geirsson, H., LaFemina, P., McNutt, S.R., Muñoz, A., Tenorio, V., 2015a. Stable and unstable phases of elevated seismic activity at the persistently restless Telica Volcano, Nicaragua. J. Volcanol. Geotherm. Res. 290, 63–74.

Rodgers, M., Rodgers, S., Roman, D.C., 2015b. Peakmatch: a Java program for multiplet analysis of large seismic datasets. Seismol. Res. Lett. 86 (4), 1208–1218.

Rodgers, M., Smith, P.J., Mather, T.A., Pyle, D.M., 2016. Quiescent-explosive transitions during dome-forming volcanic eruptions: using seismicity to probe the volcanic processes leading to the 29 July 2008 Vulcanian explosion of Soufrière Hills Volcano, Montserrat. J. Geophys. Res. Solid Earth 121 (12), 8453–8471.

Roman, D.C., 2009. Telica Seismic and Deformation Network. International Federation of Digital Seismograph Networks https://doi.org/10.7914/SN/6D_2009.

Roman, D.C., LaFemina, P.C., Bussard, R., Stephens, K., Wauthier, C., Higgins, M., et al., 2019. Mechanisms of unrest and eruption at persistently restless volcanoes: Insights from the 2015 eruption of Telica Volcano, Nicaragua. Geochem. Geophys. Geosyst., 20 https://doi.org/10.1029/2019GC008450.

Ruano, A.E., Madureira, G., Barros, O., Khosravani, H.R., Ruano, M.G., Ferreira, P.M., 2014. Seismic detection using support vector machines. Neurocomputing 135, 273–283.

Schafer, J.L., Graham, J.W., 2002. Missing data: our view of the state of the art. Psychol. Methods 7 (2), 147.

Siebert, L., Simkin, T., Kimberly, P., 2011. Volcanoes of the World. Univ of California Press.

Siebert, L., Cottrell, E., Venzke, E., Andrews, B., 2015. Earth's volcanoes and their eruptions: An overview. The Encyclopedia of Volcanoes. Academic Press, pp. 239–255.

Simkin, T., Siebert, L., 1994. Volcanoes of the World. Geosciences Press, Inc, Tusson.

Sparks, R.S.J., 2003. Forecasting volcanic eruptions. Earth Planet. Sci. Lett. 210 (1–2), 1–15.

Sparks, R.S.J., Aspinall, W., 2004. Volcanic activity: frontiers and challenges in forecasting, prediction and risk assessment. Geophys. Monogr. 150, 359–373.

Stix, J., 2007. Stability and instability of quiescently active volcanoes: the case of Masaya, Nicaragua. Geology 35 (6), 535–538.

Tilling, R.I., 2009. El Chichón's" surprise" eruption in 1982: Lessons for reducing volcano risk. Geofis. Int. 48 (1), 3–19.

VAAC, 2013. Washington VAAC 2013 Volcano Ash Advisory Archive. Satellite Products and Services Division/Office of Satellite and Product Operations, Washington Volcanic Ash Advisory Centre https://www.ssd.noaa.gov/VAAC/ARCH13/archive.html.

Wadge, G., Voight, B., Sparks, R.S.J., Cole, P.D., Loughlin, S.C., Robertson, R.E.A., 2014. An overview of the eruption of Soufriere Hills Volcano, Montserrat from 2000 to 2010. Geol. Soc. Lond. Mem. 39 (1), 1–40.

Williams, C.K., Rasmussen, C.E., 2006. Gaussian Processes for Machine Learning (Vol. 2, No. 3, p. 4). MIT Press, Cambridge, MA.

Winson, A.E., Costa, F., Newhall, C.G., Woo, G., 2014. An analysis of the issuance of volcanic alert levels during volcanic crises. J. Appl. Volcanol. 3 (1), 14.

View publication stats