



HAL
open science

Automatic classification of volcano seismic signatures

Marielle Malfante, Mauro Dalla Mura, Jerome I. Mars, Jean-Philippe
Métaxian, Orlando Macedo, Adolfo Inza

► **To cite this version:**

Marielle Malfante, Mauro Dalla Mura, Jerome I. Mars, Jean-Philippe Métaxian, Orlando Macedo, et al.. Automatic classification of volcano seismic signatures. *Journal of Geophysical Research*, 2018, 123 (12), pp.10,645-10,658. 10.1029/2018jb015470 . hal-01949302

HAL Id: hal-01949302

<https://hal.science/hal-01949302>

Submitted on 21 Dec 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

RESEARCH ARTICLE

10.1029/2018JB015470

Key Points:

- We propose a supervised process for the automatic classification of volcano seismic signals
- We test and validate the proposed architecture on 6 years of continuous seismic recordings from a short-period seismic station
- The automatic classification scheme performs better than human analysts in a crisis situation

Correspondence to:

J.-P. Métaxian,
jean-philippe.metaxian@ird.fr

Citation:

Malfante, M., Dalla Mura, M., Mars, J. I., Métaxian, J.-P., Macedo, O., & Inza, A. (2018). Automatic classification of volcano seismic signatures. *Journal of Geophysical Research: Solid Earth*, 123. <https://doi.org/10.1029/2018JB015470>

Received 30 JAN 2018

Accepted 21 SEP 2018

Accepted article online 26 SEP 2018

Automatic Classification of Volcano Seismic Signatures

Marielle Malfante¹ , Mauro Dalla Mura¹ , Jerome I. Mars¹ , Jean-Philippe Métaxian^{2,3} , Orlando Macedo^{4,5} , and Adolfo Inza⁴ 

¹Univ. Grenoble Alpes, CNRS, Grenoble INP, GIPSA-lab, Grenoble, France, ²Univ. Grenoble Alpes, Univ. Savoie Mont Blanc, CNRS, IRD, ISTERRE, Grenoble, France, ³Institut de Physique du Globe de Paris, Université Sorbonne-Paris-Cité, CNRS, Paris, France, ⁴Instituto Geofísico del Perú, Lima, Peru, ⁵Facultad de Geología, Geofísica y Minas, Universidad Nacional de San Agustín de Arequipa, Arequipa, Peru

Abstract The prediction of volcanic eruptions and the evaluation of associated risks remain a timely and unresolved issue. This paper presents a method to automatically classify seismic events linked to volcanic activity. As increased seismic activity is an indicator of volcanic unrest, automatic classification of volcano seismic events is of major interest for volcano monitoring. The proposed architecture is based on supervised classification, whereby a prediction model is built from an extensive data set of labeled observations. Relevant events should then be detected. Three steps are involved in the building of the prediction model: (i) signals preprocessing, (ii) representation of the signals in the feature space, and (iii) use of an automatic classifier to train the model. Our main contribution lies in the feature space where the seismic observations are represented by 102 features gathered from both acoustic and seismic fields. Ideally, observations are separable in the feature space, depending on their class. The architecture is tested on 109,609 seismic events that were recorded between June 2006 and September 2011 at Ubinas Volcano, Peru. Six main classes of signals are considered: long-period events, volcanic tremors, volcano tectonic events, explosions, hybrid events, and tornillos. Our model reaches $93.5\% \pm 0.50\%$ accuracy, thereby validating the presented architecture and the features used. Furthermore, we illustrate the limited influence of the learning algorithm used (i.e., random forest and support vector machines) by showing that the results remain accurate regardless of the algorithm selected for the training stage. The model is then used to analyze 6 years of data.

1. Introduction

1.1. Volcano Monitoring

Monitoring volcanoes and forecasting their related hazards has been a concern and a challenge for the scientific community for many years. With the potential for dramatic consequences on populations and the economy, volcanic hazards are among the most threatening, and thus, the interest in reliable forecasting and prevention tools is obvious. A common approach for volcano monitoring relies on tracking the evolution of several parameters, including seismicity, deformation, and gas flux and gas composition (McNutt, 1996). Usually, the tracking of these parameters and of their evolution is still essentially carried out manually. Over the last decade, the technical evolution of geophysical instruments and their cost reduction have led volcano observatories to install an increasing number of sensors. The ever increasing amounts of recorded data that have to be processed makes manual approaches inappropriate for carrying out comprehensive analyses. This represents another example of the *big data* phenomenon. As in various domains, machine learning methods are today considered to automate analyses and to help to improve decision making, which can lead to health and safety recommendations.

In this study, we propose a novel approach to automatically classify volcano seismic events that is based on machine learning. The success of these methods relies on signal characterization with a large set (102) of descriptors.

1.2. Machine Learning

We give here a short introduction to supervised machine learning that is aimed at defining the terminology and the framework of the proposed technique. Machine learning is a field of artificial intelligence that is aimed at automatically assigning objects (from a potentially large set) to categories that are associated to semantic information. This automatic decision process relies on a prediction model that is learned on a set of examples where the categories (or classes) are known beforehand. More specifically, the analysis is carried out on a data

set of *observations*. In this study, we refer to observations as the recorded seismic signal that corresponds to a volcano seismic event (e.g., an explosion or a tremor). However, in the literature and in many studies related to machine learning, observations can also be referred to as signatures, samples, examples, data, or simply as signals. The data set is said to be *labeled* if each observation has previously been categorized and associated to a semantic class, that is, to one of the classes of interest (in the present study, classes correspond to the type of volcanic event that is associated to the observation). Supervised machine learning algorithms provide models that can learn to predict the class of a new observation, where the label is not known. Many supervised machine learning algorithms are being used today. We will introduce here two of the most effective ones (random forest and support vector machines-SVMs), which were used to automatically classify volcano seismic observations in this study. Machine learning algorithms are today very effective and are widely used in image processing, speech processing, medical imaging, finance, and robotics, among others (Friedman et al., 2001).

As in many signal processing applications, a key point in machine learning processes is the representations used for the observations. Traditionally, these are represented by *features* that relate to measurements made on the observations that summarize and characterize their properties. In particular, properties that discriminate observations into their respective classes should be highlighted in the feature space. In machine learning, the use of a feature space with respect to the original recording space is also motivated by the curse of dimensionality (Bellman, 1956). More specifically, the number of observations, N , needed to train a model increases with the dimension, d , of the input space, thereby representing the data in a space of low dimension.

Various methods can be used to pass from the original space of representation to a feature space, which include learning features or designed features, which we will briefly explain (Friedman et al., 2001). However, using one or the other method of defining features gives no guarantee of finding the optimum space of representation for the data. Finding an appropriate feature space is always the key point when using machine learning algorithms, for it greatly impacts on the output performance and strongly modifies the results. Learning representations are often referred to as *dictionary learning*, and they include algorithms such as principal component analysis, independent component analysis, singular values decomposition, nonnegative matrix factorization, and convolutional neural networks. Their main idea is to learn a dictionary of elements that can be used to reconstruct the data. The main advantage of these algorithms is the final representation that is adapted to best discriminate the observations into their classes. However, such algorithms often require a large and labeled data set to learn the representation, and they can be very costly. Furthermore, a learning representation cannot be related to a physical meaning and cannot be generalized to other data.

In the present study, as in many studies that have dealt with natural or environmental data, we use designed features. In particular, we introduce $D = 102$ descriptors to extract features that represent and quantify the physical values of the observations. The main challenge of this method is the difficulty of finding features that effectively separate the observations into their classes. This step requires deep knowledge of the data and, especially, why a given observation belongs to a specific class. One advantage of these feature sets is that they can be generalized: They can be designed for one application, but they can be relevant to other fields. Moreover, this technique is less constraining on computation time and allows the observation to be represented by physical quantities that can be associated to a physical meaning of the phenomenon. In signal processing, classical features would be mean values or standard deviation, for example. Yet another option to extract features would be to apply dictionary learning over designed features, which would lead to an even more compressed representation of the signals, although once again, this loses the physical meaning of the features.

We will now briefly introduce the two learning algorithms that will be used throughout the present study. The random forest algorithm (Breiman, 2001) is built upon an ensemble of binary decision trees (Quinlan, 1986), which are combined by majority voting. A binary decision tree is a discriminative algorithm that carries out iterative tests on features in a hierarchical manner, until a classification rule is defined. Potentially, the main advantage of the algorithm lies in its simplicity and its explanatory power. However, binary decision trees tend to overfit (i.e., adapt too closely) the examples from the learning phase, and thereby the learned decision rules might not be general enough to correctly discriminate newly recorded observations. Random forest was proposed to overcome this limitation by using bagging: An ensemble of many binary trees is considered, with each trained on a random subset of the original data set (e.g., typically, a random 10% of the data set is left for each tree). Furthermore, each node of a tree takes decisions on a random subset of the features

used to represent the observations (e.g., typically, \sqrt{d} features are considered, where d is the feature vector dimension). This second approach aims to further decrease the correlation among the binary trees.

The second machine learning approach we use, SVM algorithms were presented by Cortes and Vapnik (1995). The main idea is to build a linear hyperplane in the feature space that maximizes the margin (i.e., the distance between the hyperplane and the closest data). This simple linear classifier works well on data that are linearly separable in the feature space, which is rarely the case in real-case scenarios. Two variants have been proposed to improve the algorithm performances. The first one uses a kernel function to project the data from the original to another feature space of higher dimension, in which the data would be linearly separable. This is known as the *kernel trick* (Friedman et al., 2001). Several kernel functions have been shown to be effective. In the case where a radial basis function is used as the kernel, the final feature space has an infinite dimension, and the data are therefore necessarily linearly separable. The second variant to the original algorithm is to use a *soft margin*, that is, to introduce a cost parameter, C_{SVM} , that allows some training observations to be on the wrong side of the margin. Using a soft margin prevents overfitting, thus provides a model that has better generalization potential. In this study, we use a SVM implementing both improvements over the original algorithm.

1.3. Related Studies

In this section, we review the studies that are related to the present study, in terms of the automatic classification of volcano seismic signals and also the choice of signal representations for machine learning purposes. We especially review several fields, including acoustic data classification. Indeed, automatic analysis of acoustic data is today much more advanced than tools currently used for analysis of seismic data. The interest in such adjusting tools that were developed for other purposes is strong, especially given the similarities between the data involved (i.e., the transient nature of the signals). In the following, we focus our attention on the features that vary significantly from one application to another. The learning algorithm has less impact on the results, and for most studies this was chosen as random forest (Hibert et al., 2017; Maggi et al., 2017), SVMs (Apolloni et al., 2009), neural networks (Apolloni et al., 2009; Langer et al., 2006), or hidden Markov models (Benítez et al., 2007; Cortés et al., 2009; Gutiérrez et al., 2009; Hammer et al., 2012; Ibáñez et al., 2009; Ohrnberger, 2003).

Automatic classification systems for volcano seismic waves are not yet widespread, although several studies on this subject have started to appear. In particular, Hibert et al. (2017), Langet (2014), and Maggi et al. (2017) classified seismic observations represented by features that included duration, skewness, kurtosis, statistics ratios, and centroids. The study of Maggi et al. (2017) had a similar approach to our architecture, where the main difference is that they obtained their best results when using several seismic stations simultaneously, while in the present study, we obtain the same results from a single station. The approach and testing methods are, however, very similar. The approach presented in Hibert et al. (2017) is also similar but on a binary classification problem. Other studies propose alternative features, such as used by Langer et al. (2006), where features based on energetic criteria were considered, such as the signal energy in various frequency bands. Cannata et al. (2011) used three features: peak frequency, quality factor, and amplitude. Apolloni et al. (2009) used linear predictive coding as a normalized difference between maximum and minimum amplitudes. Esposito et al. (2008) propose to use unsupervised method but on the observation of waveforms directly. Hibert et al. (2014) propose to use five features, including the signal duration, duration of increasing and decreasing phases, kurtosis of the envelope, and signal maximum over mean ratio. Those features are used to discriminate rockfall from all other signals. Benítez et al. (2007), Cortés et al. (2009), Gutiérrez et al. (2009), and Ibáñez et al. (2009) use modified mel frequency cepstral coefficients, which were originally designed for speech processing and measure the signals energetic repartition in frequency bands. Physical features, such as polarization and spectral attributes, can also be used (Hammer et al., 2012; Ohrnberger, 2003).

Acoustic signals of various origin can be targeted for automatic classification, including environmental sounds, music, or speech. For automatic classification of environmental sounds (i.e., from nature and animals, or human induced), we can mention Tucker and Brown (2005), who classified transient sonar sounds. More than 20 features were gathered to describe the signals, including duration, peak power, average power, time of peak power, mean skewness, mean kurtosis, power standard deviation, rate of attack, and rate of decay. Similar features were used by Zaugg et al. (2010) to distinguish boat sounds from whale sounds. Fagerlund (2007) also gathered signal descriptors to identify bird species from their calls, which included spectral centroid, bandwidth, and spectral flatness and duration. Spectral centroid, bandwidth, and threshold crossing rates were also considered by Huang et al. (2009) for a frog sound identification system. For the same purpose,

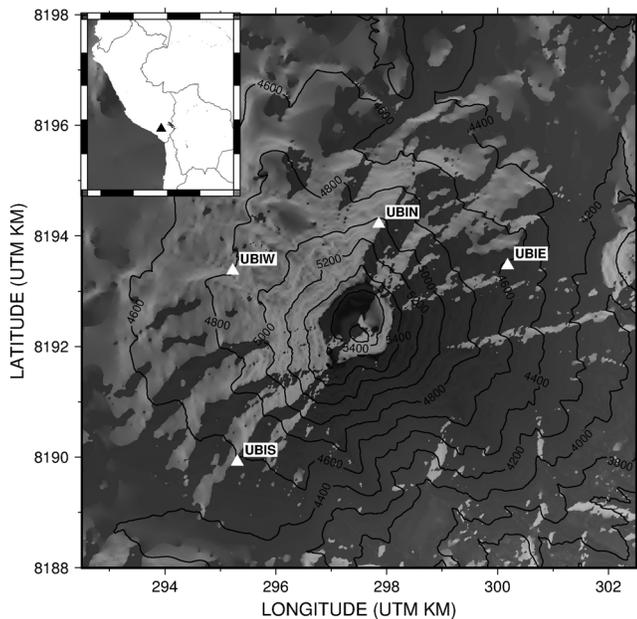


Figure 1. Map of Ubina Volcano with the locations of the permanent IGP seismic network indicated (white triangles). The data used in this study were recorded at UBIW station. Inset, top left: location of Ubina Volcano (black triangle) within Peru.

Han et al. (2011) also used Shannon and Rényi entropies. Similar features have been used for music processing; see, for example, Fujinaga and MacMillan (2000), Eronen and Klapuri (2000), and Esmaili et al. (2004).

For the present study, our main contribution lies in the definition of the feature space in which the signals are defined. Namely, we propose the use of 102 features to obtain a comprehensive description of the observations. This paper is organized as follows: Section 2 presents the volcano seismic data set used in this study, which was recorded at Ubina Volcano (Peru). We present the automatic classification architecture in section 3, and in particular, the feature set is described in section 3. The results are then presented and discussed in section 4. Finally, the prospects and conclusions surrounding this study are summarized in section 5.

2. Experimental Material

2.1. Seismic Monitoring of Ubina Volcano

Ubina Volcano is an andesitic stratovolcano in southern Peru, (16° 22'S, 70° 54'W; altitude, 5,672 m). It is considered to be the most active volcano in Peru, and it is closely monitored by the *Instituto Geofísico del Perú* (IGP). After nearly 40 years of quiescence, Ubina Volcano erupted in 2006. Three eruptions have occurred since 2006, from 2006 to 2011, from 2013 to 2014, and in 2016. Ubina Volcano has been monitored seismically by the IGP since 2006 (Macedo et al., 2009), with the cooperation of the VOLUME project (funded by the European Commission 6th Framework Program) and the *Institut de Recherche pour le Développement* (France). The first per-

manent telemetered station (i.e., UBIW) was equipped with a short-period vertical 1-Hz sensor that was installed in May 2006 on the northwest flank of Ubina Volcano (Macedo et al., 2009). Three additional stations were added in 2007 (i.e., UBIN, UBIE, and UBIS). UBIN was equipped with a broadband vertical sensor, and the other stations had short-period sensors. In addition, UBIN and UBIS were equipped with biaxial tiltmeters with 0.1- μ rad resolution (Inza et al., 2014). These four stations have been working permanently since 2007 (Figure 1). The data are recorded continuously with a sampling rate of 100 Hz, and they are then transmitted in real time to Cayma Volcanological Observatory in Arequipa (Peru). In this paper, our analysis is based on seismic data from the vertical component of UBIW station.

2.2. Seismic Signatures Overview

The data used for the present study came from a catalog of $N = 109,609$ seismic observations of volcanic events that were recorded between May 2006 and October 2011. Each observation was manually identified and extracted by the IGP in Arequipa Observatory. Six very heterogeneous classes of signals were defined by

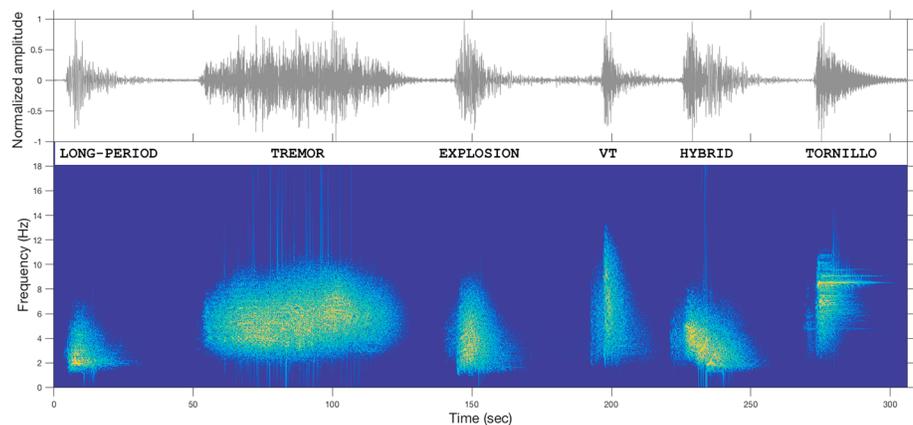


Figure 2. Waveform and spectrogram (Gaussian window of 512 samples width) of volcano seismic signals recorded at Ubina Volcano. Six observations are shown, as examples of long-period (LP) events, tremors (TRs), explosions (EXPs), volcano-tectonic (VT) events, hybrid (HYB) events, and Tornillos (TORs). The amplitude is linear and has been normalized.

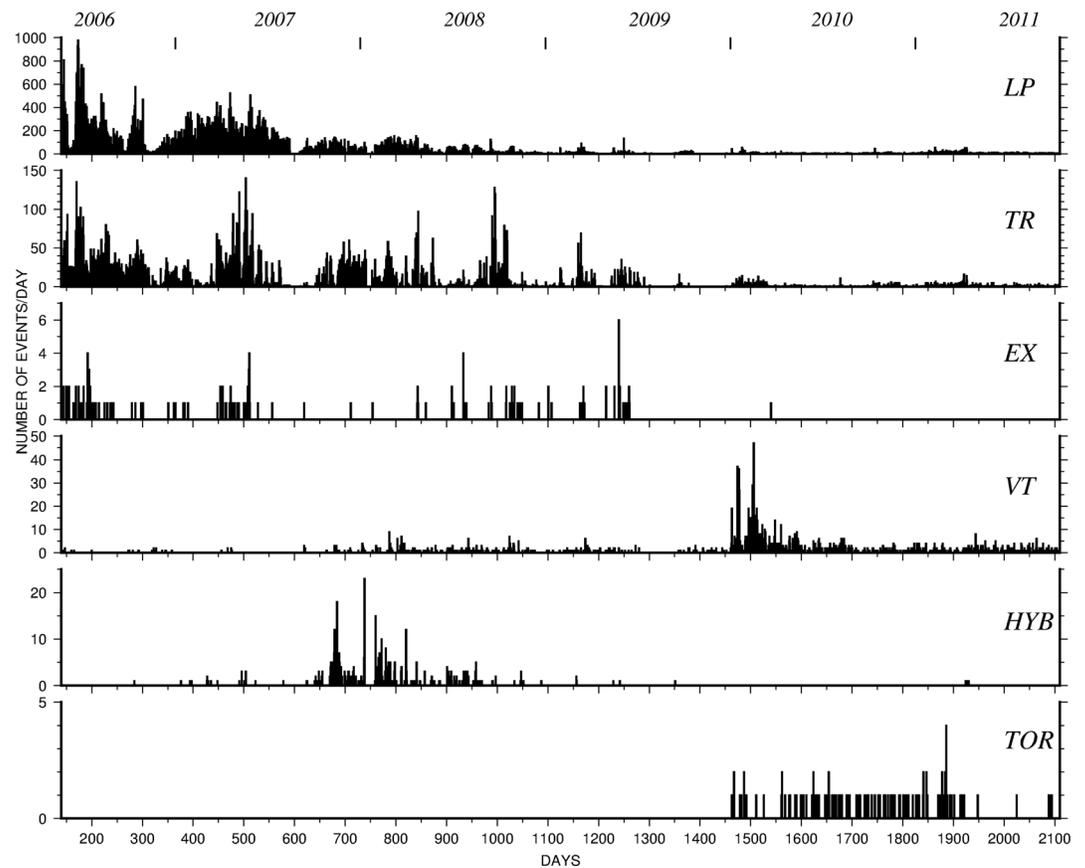


Figure 3. Number of events per day for the six different types of volcano seismic signals recorded at Ubinas Volcano. The date is indicated in cumulated days starting from 1 January 2006.

the IGP after 10 years of observations of Ubinas Volcano. Each class is associated to a physical state or activity of the volcano. These types of signals are observed for other volcanoes, and have been described for many years in the literature, in particular, by (Chouet, 1996; Lahr et al., 1994; Neuberg et al., 2000). They are listed in the following:

1. Long-period (LP) events (95,243 observations): These originate from fluid processes. They are interpreted as a time-localized pressure excitation mechanism, followed by the response of a fluid-filled resonator (Chouet & Matoza, 2013). Different models have been developed to explain the resonance observed for LP events, including in particular the fluid-filled crack model (Chouet, 1986) and the fluid-filled conduit model (Neuberg et al., 2000). A wide variety of volcanic processes can produce the excitation mechanism that triggers crack or conduit resonance, as particularly for lava dome growth for andesitic volcanoes (Chouet & Matoza, 2013; Morgan et al., 2008; Neuberg et al., 2000).
2. Tremors (TRs; 12,309 observations): These are defined by a sustained amplitude that can last from tens of seconds to days, and they occur over a frequency range from 1 to 9 Hz (McNutt, 1992). This author reported that many characteristics of LP events, and in some cases also of explosion quakes (see below), are commonly associated with TRs. Virtually all eruptions are accompanied by TRs (McNutt, 1992). Visual observations at Ubinas Volcano suggest that TRs are associated with magma extrusion and sporadic or continuous gas and ash emissions (Macedo et al., 2009).
3. Explosions (EXPs; 159 observations): These are associated to sudden magma extrusion, and ash and gas emission. Physically, they are related to fragmentation processes in the conduit, as has been observed for many andesitic volcanoes (Druitt et al., 2002; Iguchi et al., 2008; Ohminato, 2006). Inza et al. (2011) noted that for Ubinas Volcano, EXPs are related to destruction of the magmatic plug. Traversa et al. (2011) showed that there is an overall acceleration of the LP events rate above the background level over the 2 to 3 hr before vulcanian EXPs for Ubinas Volcano. This occurrence of a large number of LP events before EXPs is consistent with the dome growth process.

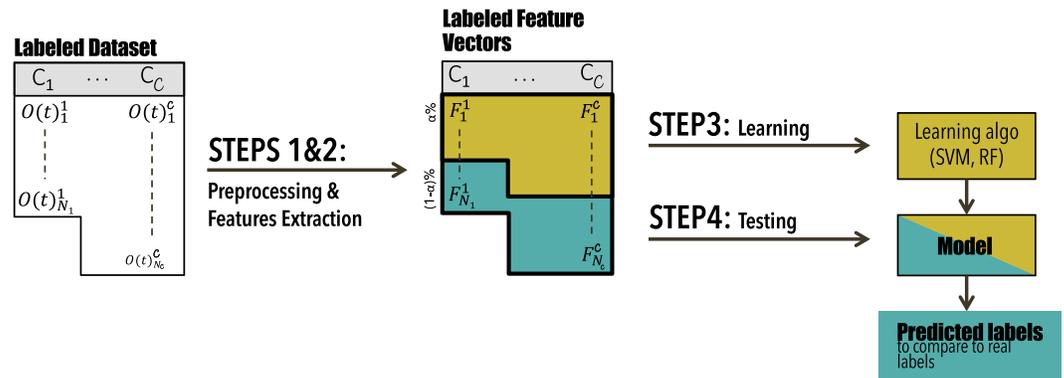


Figure 4. Illustration of the architecture used to build a model (light green), and the cross-validation process used to estimate its performance (dark green). Steps 1, 2, and 3 are used to train a model from a labeled data set of observations, $[O(t)]$. The observation $O(t)_j^i$ is the j -est observation of class c_i , with $1 \leq i \leq C$, $1 \leq j \leq N_i$, and N_i as the number of observations considered for class c_i . F_i^j is its corresponding feature vector. A fraction α with $0 < \alpha \leq 1$ of the data set is used for learning, and the remaining data are used for testing. The process is repeated 50 times, with random trials of learning and testing observations. Step 1: preprocessing, Step 2: features extraction, Step 3: learning, and Step 4: testing (or using) the model. SVM = support vector machine; RF = random forest.

4. Volcano-tectonic (VT) events (1,315 observations): These are brittle-failure events. They are associated to stress changes that are induced by magma movement (Chouet & Matoza, 2013). There are relatively few VT events at Ubinas Volcano, compared to LP events, but their number increased from 2006 to 2011. VT events are more numerous at the end of an eruption period.
5. Hybrid (HYB) events (474 observations): These have characteristics of both VT and LP events, with high frequency of onset followed by low frequencies. HYB events were introduced by Lahr et al. (1994) to describe events observed at Redoubt Volcano. They have also been observed for Soufrière Hills Volcano, Montserrat Volcano (White et al., 1998), and Mount St. Helens Volcano (Harrington & Brodsky, 2007), where they are related to dome growth.
6. Tornillos (TORs; also known as *screw* events; 109 observations): These have a very limited distribution of frequencies and a very slowly decaying coda. They are related to resonating fluid-filled conduits or cavities. TORs can be considered as a specific type of LP event with a long-duration coda that is composed of harmonic oscillations. They were observed at Galeras Volcano before several eruptions in 1992 and 1993 (Narváez et al., 1997). TORs are rare for Ubinas Volcano, but they appear to be more common at the beginning and end of eruptive periods.

See Figure 2 for an example of the waveforms and spectrograms associated to each of these classes. The number of events of different types per day throughout the time period of the study is showed in Figure 3.

There is no class of rockfalls in our list. Generally speaking, rockfalls do not constitute a significant activity at Ubinas, perhaps because this part of the Andes is very dry. During the period 2006–2011, Ubinas was in eruption with episodes of dome growth, explosions, and very regular ash emission generating seismic events of higher amplitude compare to any kind of rockfalls. It is probable that if rockfalls have occurred, the generated signals were masked by the magmatic activity or not selected due to low amplitudes compared to eruptive signals.

3. Methods

We here detail the classification architecture proposed. From a large data set of observations of volcano seismic events, a machine learning algorithm is used to train a model to classify new events into one of the six classes considered (Figure 2). The system architecture is described in Figure 4. The data considered for the input were a labeled data set, as described in section 2.2. $C = 6$ different classes of observations are considered. For each class c_i (with $1 \leq i \leq C$), N_i observations were labeled and are taken into account. All of the observations do not necessarily have the same length. We detail Steps 1 and 2 below that transform the labeled data set of observations into a labeled data set of feature vectors. Step 3 describes the learning process where the prediction model is actually built. Finally, a fourth step of testing can be included to estimate the model performances, as is explained in section 4.

Step 1: Preprocessing. A first step of signal preprocessing is needed to standardize the observations. Typically, observations that were recorded from the same station but with different sensors need to be harmonized (e.g., due to change of sensor following technical failure). In particular, a filtering above 1 Hz is performed. The instrumental response was not removed from the original signals. Preprocessing also includes the automatic or manual detection of relevant seismic observations of volcanic events that will constitute the data set. Observations considered in this catalog were detected either manually by the observatory or automatically using the short-term average/long-term average method. Each observation is also normalized in terms of its energy. The signal energy is computed and normalized to 1. This is done to build a model that is available for all observations, regardless of their amplitude level. It is worth noting that most state-of-the-art methods are based on the energy levels (Langer et al., 2006) and therefore cannot be used to detect observations of low energy. Finally, a data set of $N = 109,609$ observations was considered. We also fixed the maximum of 5 min per observation, in order to build a model that could permit near real-time analysis.

Step 2: Feature extraction. As specified in sections 1.2 and 1.3, a model sees the data exclusively in the feature space; that is, observations are only considered through their associated feature vectors. The choice of features used to represent an observation is therefore one of the major and crucial issues when designing a classification architecture. In the present study, we reviewed, gathered, and selected $D = 102$ features that have been used in many classification tasks from various domains (see section 1.3). The use of a large number of features helps to represent the observations in as thorough a way as possible, to be able to describe their separating properties. We organized these features into three categories, namely, statistical, entropy-based, and shape descriptors. The definitions of these features are given in Table 1.

1. *Statistical features* are valued for their immediate interpretation in terms of the signal shapes. We used means and standard deviations and also higher moments, such as skewness and kurtosis, which measure the asymmetry and flatness of the signal, respectively, compared to the Gaussian distribution.
2. *Entropy features* come from information theory, and these are used to measure the information content of a signal. We used Shannon entropy and Rényi measurements (Esmaili et al., 2004; Huang et al., 2009).
3. *Shape descriptors* can be used, for instance, maximum over average values. These help to describe the shape of a signal and therefore its physical properties. Rates of initiation and decay are particularly useful for this purpose.

Furthermore, we propose to extract these features from three different representations of the observations. Specifically, all of the features were extracted from three representations:

1. *Time domain.* Features extracted from the observation $x[t]$ are used to describe the waveform shape and its specificities (Time).
2. *Frequency domain.* A Fourier transform ($\mathcal{F}\{\cdot\}$) of the discrete temporal signal $x[t]$ leads to $X[f] = \mathcal{F}\{x[t]\}$, which represents the spectral content of the observation (Frequency).
3. *"Cepstral" domain.* This domain is commonly used in speech processing to describe harmonic properties of the signal, that is, the periodic properties of the spectrum. The features are then extracted from $\cdot[q] = \cdot\{|\mathcal{X}[f]|\}$ (Cepstral).

Eventually, each observation is represented by 34 features (i.e., 9 statistical, 9 entropy, and 16 shape descriptor features) that are extracted in three different domains, which leads to a feature vector of dimension $D = 34 \times 3 = 102$. The original data dimension is thereby reduced while maintaining the information content of the signals (as a comparison, an original 30-s-long observation has a dimension of 3,000). Using general signal shape descriptors leads to a precise description of the signal properties. Extraction of these descriptors from the three different domains (i.e., temporal, spectral, and cepstral) underlines different properties of the original signal and eventually leads to a thorough description of each observation. Of note, the feature extraction process used for these data can also be applied to other transient signals.

The labeled data set of feature vectors is then considered for the next step.

Step 3: Learning. This final step can be referred to as the model *learning* or *training*. Here a prediction model is built from a learning algorithm and a data set of labeled feature vectors. Several learning algorithms can be used for this task (leading, in many cases, to similar results). The choice of the optimal learning algorithm for a given application is empirical. However, if the feature space has been correctly chosen, the learning

Table 1
List of Features

Statistic features			
Feature	Definition	Used in	Ref.
Length	$n = \text{length}(s)$	Tucker and Brown (2005)	1
Mean	$\mu_s = \frac{1}{n} \sum_i s[i]$	Tucker and Brown (2005)	2
Standard deviation	$\sigma_s = \sqrt{\frac{1}{(n-1)} \sum_i (s[i] - \mu_s)^2}$		3
Skewness	$\frac{1}{n} \sum_i \left(\frac{s[i] - \mu_s}{\sigma_s} \right)^3$	Langet (2014) and Hibert et al. (2014)	4
Kurtosis	$\frac{1}{n} \sum_i \left(\frac{s[i] - \mu_s}{\sigma_s} \right)^4$	Langet (2014) and Hibert et al. (2014)	5
i of central energy	$\bar{i} = \frac{1}{E} \cdot \sum_i E_i \cdot i$	(Tucker & Brown, 2005)	6
RMS bandwidth	$B_i = \sqrt{\frac{1}{E} \sum_i i^2 \cdot E_i - \bar{i}^2}$	Tucker and Brown (2005)	7
Mean skewness	$\sqrt{\frac{\sum_i (i - \bar{i})^3 E_i}{E \cdot B_i^3}}$	Tucker and Brown (2005)	8
Mean kurtosis	$\sqrt{\frac{\sum_i (i - \bar{i})^4 E_i}{E \cdot B_i^4}}$	Tucker and Brown (2005)	9
Entropy features (with $p(s_j)$ the probability of amplitude level s_j)			
Feature	Definition	Ref.	
Shannon entropy ^a	$-\sum_j p(s_j) \log_2(p(s_j))$	Esmaili et al. (2004) and Han et al. (2011)	10 to 12
Rényi entropy ^b	$\frac{1}{1-\alpha} \cdot \log_2 \left(\sum_j p(s_j)^\alpha \right)$	Han et al. (2011)	13 to 18
Shape descriptor features			
Feature	Definition	Ref.	
Rate of attack	$\max_j \left(\frac{s[j] - s[j-1]}{n} \right)$	Tucker and Brown (2005)	19
Rate of decay	$\min_j \left(\frac{s[j] - s[j+1]}{n} \right)$	Tucker and Brown (2005)	20
Ratios	min/mean and max/mean	Langet (2014) and Hibert et al. (2014)	21 to 22
Energy descriptors	Signal energy, maximum, average, standard deviation, skewness, and kurtosis	Tucker and Brown (2005)	23 to 28
Specific values	min, max, i of min, i of max, threshold crossing rate, and silence ratio	Tucker and Brown (2005)	29 to 34

Note. Features computed for a signal $s[i]_{i=1}^n$ (in which i might correspond to a temporal, frequency, or cepstral sample). $E = \sum_{i=1}^n s[i]^2$ and $E_i = s[i]^2$ describe the signal energy and the energy at sample i , respectively. Some features have a dimension greater than others; e.g., entropy measurements are made on three different estimations of the amplitude probability (i.e., different histogram bin numbers).

^aBin numbers for probability estimation: 5, 30, and 500. ^bBin numbers for probability estimation: 5, 30, 500, $\alpha = 2$, and inf.

algorithm choice should not have a major influence on the results. In this study, we chose to use random forest and SVMs, as both of these are state-of-the-art techniques in machine learning.

Step 4: Testing. An extra step of testing can also be considered to evaluate the performance of a model and thus to validate the classification architecture and, in particular, the feature choice. In this study, we used cross validation, the concept behind which is illustrated in Figure 4 and is explained, in section 4.1.

4. Results and Their Analysis

The code used for this study is based on the Automatic Analysis Architecture that we developed and is available at Malfante (2018). It uses Python 3, and all tests were conducted on a MacBook Pro laptop. All results are measured in term of accuracy and precision. Accuracy describes the good classification rate within a class (see equation (1)), and precision describes the rate of observations belonging to the predicted class among all the prediction of the class (see equation (2)).

$$\text{Accuracy } c_i = \frac{\#(\text{Observations of } c_i \text{ predicted as } c_i)}{\#(\text{Observations of } c_i)} \quad (1)$$

$$\text{Precision } c_i = \frac{\#(\text{Observations of } c_i \text{ predicted as } c_i)}{\#(\text{Observations predicted as } c_i)}. \quad (2)$$

Table 2
Classification of the General Results

Features	All	Most valuable features	Valuable features	Time	Frequency	Cepstral
Dimension	102	3	13	34	34	34
SVM	92.1 ± 0.54%	84.4 ± 0.50%	86.9 ± 0.60%	83.9 ± 0.89%	93.5 ± 0.50%	78.4 ± 1.0%
RF	92.5 ± 0.45%	84.2 ± 0.75%	90.3 ± 0.52%	87.2 ± 0.57%	91.3 ± 0.45%	79.3 ± 0.76%

Note. Comparisons of the accuracies of the results using random forest versus support vector machine, for the different feature sets.

4.1. First Test: Method Validation

To validate the proposed classification scheme, we perform cross-validation analysis on the first year of the data set. To consider relatively balanced classes, we gather 1 year of observations for classes that are highly frequent (i.e., LP events, TRS, and VT events), and all of the observations available for the less frequent classes (i.e., HYB events, EXPs, and TORs). For each class, a fraction α of the observations is used to train the model. Because all the learning data need to be loaded in the computer memory at the same time, the number of training observation per class is also limited to $N_{\max} = 800$ for computational reasons. The remaining data are used for the test stage. Formally, if $0 < \alpha \leq 1$ stands for the learning rate, N_i for the number of observations considered for the class c_i with $1 \leq i \leq C$ and C the number of classes, then $\min(N_i \cdot \alpha, N_{\max})$ observations are therefore considered for each class for the training step. The remaining $\max(N_i \cdot (1 - \alpha), N_i - N_{\max})$ observations per class are used to test the model that is produced. The random selection of learning and testing data among all the available observations is repeated 50 times to consider statistically valid results. In the following, the results are expressed as means and standard deviation over the 50 trials. The cross-validation results for the comparison of the two learning algorithms (i.e., random forest and SVMs) and the four feature sets (i.e., Time, Frequency, Cepstral, and All together) are given in Table 2.

Using the All features, the accuracy of the results reaches $92.5\% \pm 0.54\%$ (i.e., a good classification rate), thereby showing the effectiveness of the proposed architecture and the feature choice. The results also validate the limited influence of the learning algorithm (i.e., random forest vs. SVMs), which was expected ($92.5\% \pm 0.45\%$ vs. $92.1\% \pm 0.54\%$, respectively, when considering All features). The influence of the features used for the automatic classification, however, is much more important. These results vary from $78.4\% \pm 1.0\%$ when using cepstral features and to $93.5\% \pm 0.50\%$ when using frequency features. It is particularly interesting to note that for these data and this application, the frequency features appear to be particularly discriminative and, in particular, that manual classification is often based on the frequency content of the observations.

The confusion matrix is given in Table 3, and it provides more details about the class-by-class results. For instance, the best classified class is for the LP events, with 58,363 correctly classified observations compared to 62,030 considered (94.1%), followed closely by the VT events, with an accuracy of 92.2%. The confusion matrix can also be used to analyze the model limitations. Most of the prediction errors are divided up between: (i) LP events mixed with TRS, (ii) HYB events mixed with VT and LP events, and (iii) EXPs mixed with HYB

Table 3
Confusion Matrix

		True Class (ground truth)						Precision
		LP	TR	VT	EXP	HYB	TOR	
Predicted Class	LP	58,363	627	8	0	5	1	98.9%
	TR	3,000	4,584	0	1	2	0	60.4%
	VT	478	11	475	5	11	3	48.3%
	EXP	15	16	2	29	0	0	47.8%
	HYB	131	3	28	13	125	0	41.7%
	TOR	43	4	3	0	0	28	35.9%
Accuracy		94.1%	87.4%	92.2%	59.8%	87.1%	84.6%	

Note. Learning rate $\alpha = 70\%$; the model was trained using support vector machine (RBF kernel, $C_{\text{RBF}} = 10$, $\gamma = 0.01$) and feature set Frequency, with cross validation with 50 trials. Overall accuracy: 93.5%. Bold values mean number of good classifications.

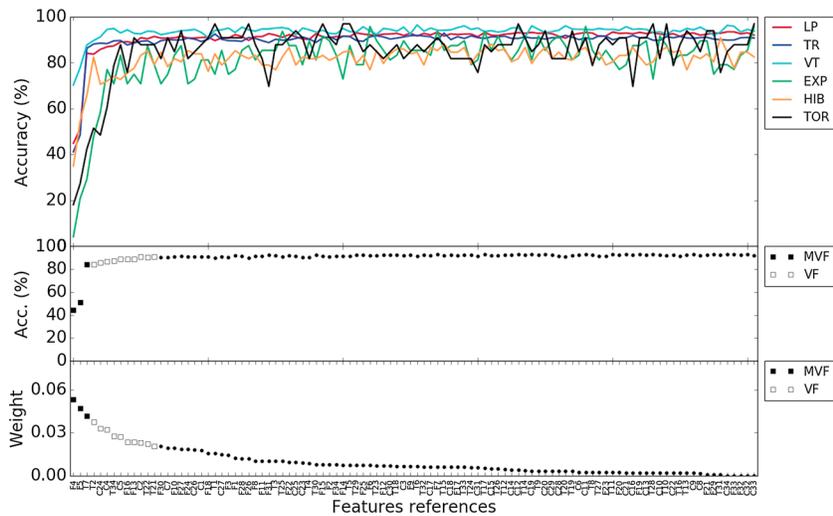


Figure 5. Importance and selection of features. Bottom subplot: Feature weights (mean weights on 1-year cross-validation models, trained with random forest on all features; 100 trees; $\alpha = 0.7$). Middle subplot: mean accuracies (models trained with random forest; 100 trees; on the d -st most important features with $1 \leq d \leq D$, $D = 102$). Top subplot: mean accuracies for each class c_i , $1 \leq i \leq C$. The three most valuable features (MVF; filled squares) and the 13 valuable features (VF; filled and empty squares) are indicated. Features are referenced with T , F , or C depending on their computation domain (time, frequency, and cepstral, respectively), and their reference numbers are as given in Table 1.

events. Those mistakes are seen through the precision rates, which are related to the false detection rates, and are low for some classes. Physical interpretation of these results is valuable, as all of the errors made by the model translate into physical similarities between the signals. For example, to parallel the main three prediction errors above: (i) LP events and TRs are in the same frequency range and can overlap, and typically, a LP event can be found within a TR. This will confuse the model, which predicts a single class at a given time. Macedo et al. (2009) also showed that on some occasions, LP events occur in a repeated way and can be separated at the beginning by some tens of seconds, before merging into a TR (e.g., before an EXP). (ii) HYB events have characteristics of both LP and VT events, and they can belong to one class or the other. Finally, (iii) EXPs and HYB events have similar frequency contents, and EXPs can produce both low and high frequencies. The analysis of this error is thus particularly valuable, as these results can help volcanologists to better analyze the seismicity, the relations between classes of events, and their evolution with time. The ability of these methods to process very large data sets of recordings is essential for volcanic observatories.

4.2. Second Test: Features Selection

As previously shown and explained, the feature choice is decisive to obtain good results. We here investigate the possibility to reduce the feature dimension through the selection of the relevant features from among the whole feature set proposed here. Considering that all of the proposed features might not be optimal, the feature vectors dimension should be kept as low as possible to avoid accuracy losses due to the curse of dimensionality (see section 1.2). In particular, features that are strongly correlated can be left out. We intentionally did not use compression algorithms (e.g., principal component analysis), so as to maintain the physical meanings of the features. Various approaches can be used to select d important features among a feature set of dimension D , which include forward selection or backward elimination (Langley et al., 1994). However, the exhaustive search for the d -st optimal features in terms of gain in accuracy is not feasible practically. As a first approach here, we use a forward selection method, which ranks the features according to their weights given in a random forest trained model. More specifically, random forest can rank features depending on their impurity scores, whereby a feature that leads to a high decrease in impurity in the trees will have a high ranking. Nevertheless, the features that are not selected using this method are not necessarily meaningless. If two features are strongly correlated, then only one will have a high score. Thus, while the random forest ranking tool is a good approach for features selection, the nonselected features should not be considered as unnecessary when interpreting the data. For more details on this subject, the reader is referred to Quinlan (1986) and Breiman (2001). Figure 5 shows the features ranked by importance using the impurity measurement and

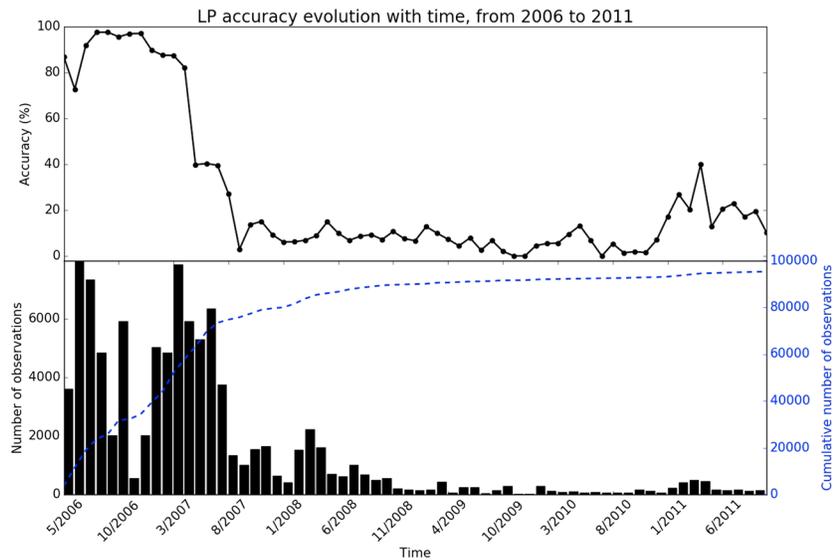


Figure 6. Monthly accuracy evolution for the LP events, with the model trained on the first 800 observations of each class (all recorded in May 2006 for the LP events). LP = long-period.

shows the accuracy of the results when the observations are represented by feature vectors of dimension d with $1 \leq d \leq D$, composed of the d most important features.

From Figure 5, we can extract the two subsets of the most important features, as the three most valuable features (Figure 5, MVF, filled squares) and the 13 valuable features (Figure 5, VF, filled and empty squares, which include the most valuable features). These features were selected as subsets, which led to a notable gain in accuracy with a reasonably small dimension. With just the most valuable features, the accuracy reaches $84.4\% \pm 0.50\%$. This result is interesting, as it shows that good classification can be obtained with a very restrictive number of features. In particular, this is of importance for real-time or embedded systems and applications. Indeed, the most valuable features contain the three most important features (3% of the total features), which are (i) F4: the spectrum skewness, which measures the spectrum asymmetry; (ii) F5: the spectrum kurtosis, which measures the spectrum peakness; and (iii) T7: the RMS bandwidth in time, which measures the mean length of the signal around which the energy is accumulated. The physical analysis of these features is, as such, a good indicator of the signals for the experts. In particular, while the features rank based on random forest can be used to select such relevant features, these features are themselves interesting to analyze, as they gather and embody the information-discriminating signals across the classes. Similarly, the 13 most important features were extracted (13% of the features) for the feature set of the valuable features (which includes the most valuable features). This feature set is particularly interesting as it provides a performance that is close to that of the whole feature set, at 90.3% when using random forest. It is also interesting to note that the valuable features are features from the three computation domains, as time, frequency, and cepstral, thereby confirming the interest in considering these three different domains for signal representation.

4.3. Third Test: Analysis Over 6 Years of Recordings

In this section, we propose to build a model using the $N_{\max} = 800$ first observations of each class recorded from May 2006 and to test it over the 6 years of recording. In particular, we focus our attention on the LP events, as these were numerous throughout the recordings (95,243 observations over 6 years). The monthly evolution of the classification results is shown in Figure 6. Specifically, the analysis reveals three different phases that start in May 2006, May 2007, and December 2010, where the mean accuracies are very different from one phase to the other. In particular, the accuracy collapses in May 2007, which reveals a significant change in the signal shapes compared to the training observations (May 2006). This period also corresponds to a sharp decrease in the number of LP events.

This was discussed with the Ubina Volcano Observatory, in Arequipa. A manual revision of the seismic signals for the period of May–July 2007 was performed to determine whether the observatory analysis might have confused LP events with other signals. Indeed, the classification criteria had been improved since the

beginning of the creation of the catalog in 2006, through the experience acquired during the other eruptions of Ubinas Volcano, as well as through the signal classification of Ticsani and Sabancaya Volcanoes. The criteria used for the reclassification are those that are now used for all of the Peruvian volcanoes. This revision analysis thus showed a difference between the two classification systems starting from 25 May 2007. Some of the LP events were indeed mislabeled, essentially as VT events or TRS, or to a lesser extent, as HYB events. Part of the accuracy collapse observed from 25 May 2007 can thus be explained by this confusion between the classes.

At the same time, the new classification showed a similar decrease in the number of LP events, which would instead suggest a change in activity, and therefore a change in the characteristics of these signals. Macedo et al. (2009) observed strong temporal variations of the degassing and seismic activity in a period starting in November 2006. A downward migration of magma in the crater has been observed on images taken in December 2006, compared to previous images taken in April and October 2006 (Figure 5 in Macedo et al., 2009). The receding of magma in the conduit has been observed over several months. Other images of the crater taken by IGP Arequipa Observatory respectively on 17 April, 8 June, and 26 August 2007 show a clear drop of the magma level between April and June 2007. The backward migration of magma in the conduit implies modifications of seismic source positions and possibly mechanisms due to the complexity of the geometry of the conduit and time modifications of the coupling between magma and the conduit walls. Magma migration can also affect local stress conditions or conduit cavity properties. This can obviously explain modifications of the characteristics of the LP events. The decrease of LP accuracy is starting in January 2007 and its strengthening between April and May 2007, which is coincident with observations of volcanic activity. The interesting point is that these modifications were not detected by manual classification, while the automatic classification perfectly identified them. This result is particularly important, as the use of automatic methods of classification has revealed inconsistencies in the original manual classification. This thereby validates the use of such methods to help with the monitoring task carried out by volcano observatories. The automatic analysis revealed a modification of interpretation of the signals for the LP events, which was not detected in the manual analysis.

5. Conclusion and Prospects

The present study is related to the classification of volcanic events based on an analysis of the seismic signals. Evaluation of the volcanic risk remains a timely and open issue, which triggers strong interest in the international community for approaches to automatically process large databases of recordings. Automatic classification tools for signals that might be either precursors or indicators of volcanic active phases are particularly needed. In this paper, we present our classification architecture that is based on supervised machine learning. The main idea is to use a labeled data set of targeted signals to build a prediction model. This prediction model can then be used to automatically analyze newly recorded signals. Our main contribution consists of the gathering of the large number of features that are used to represent the data. We considered 102 features that we tested on 109,609 seismic observations acquired for Ubinas Volcano, which is the most active volcano in Peru. The features give a general and precise description of transient signals and can be used for other applications. Our model reaches an accuracy of $93.5\% \pm 0.50\%$ over the labeled set of data, thereby validating the effectiveness of the proposed classification scheme and the chosen features for signal description. Furthermore, the analysis of the whole data set used for our method revealed a clear change in the LP event observations linked with a significant evolution of the magmatic activity, which had not been identified by the manual classification. The learned model actually led to a more precise analysis. Further analysis showed that the accuracy of the results can be maintained when only a subset of the most important features is used, as obtained through random forest features ranking. Comparisons of the two learning algorithms (i.e., random forest and SVM) also confirmed the limited influence these have if the feature space has been well chosen.

The prospects for this study indicate the implementation of such tools in volcano observatories to be run in real time, the development of a confidence index, and the inclusion of a detection stage to analyze continuous recordings. More investigation on the feature set could also be conducted, typically features computed from spectrogram representation of the events could be considered, and deep neural network could be used to represent the data. Another perspective would be to use semisupervised or unsupervised learning algorithms to relax the dependency on a labeled data set, which today limits the use of supervised models. The use of machine learning models can also be used to process continuous signals instead of discrete events. This last prospect is currently under investigation for future work.

Acknowledgments

This work was supported by a grant from Labex OSUG@2020 (*Investissements d'avenir*-ANR10 LABX56) and DGA/MRIS. GIPSA-lab SIGMAPHY is part of Labex OSUG@2020 (ANR10 LABX56). The authors thank the IDEX for funding a travel grant (M. Malfante). The code used in this work is available at <https://github.com/malfante/AAA>. The catalogue of data can be found on IGP website <http://ovs.igpp.gov.pe/catalogos-vulcanologicos>. Raw data can be obtained by requesting them to one of the coauthors from IGP, Orlando Macedo or Adolfo Inza. This work was partially supported by the project VOSICA in the framework of the Grenoble Alpes Data Institute (ANR-15-IDEX-02).

References

- Apolloni, B., Bassis, S., & Morabito, F. (2009). *Neural Nets WIRN09-Proceedings of the 19th Italian Workshop on Neural Nets* (pp. 116–123). Vietri sul Mare, Salerno, Italy.
- Bellman, R. (1956). Dynamic programming and Lagrange multipliers. *Proceedings of the National Academy of Sciences of the United States of America*, 42(10), 767–769. <https://doi.org/10.1073/pnas.42.10.767>
- Benítez, M. C., Ramírez, J., Segura, J. C., Ibáñez, J. M., Almendros, J., García-Yeguas, A., & Cortés, G. (2007). Continuous HMM-based seismic-event classification at Deception Island, Antarctica. *IEEE Transactions on Geoscience and Remote Sensing*, 45, 138–146. <https://doi.org/10.1109/TGRS.2006.882264>
- Breiman, L. (2001). Random forests. *Machine Learning*, 45, 5–32. <https://doi.org/10.1023/A:1010933404324>
- Cannata, A., Montalto, P., Aliotta, M., Cassisi, C., Pulvirenti, A., Privitera, E., & Patan, D. (2011). Clustering and classification of infrasonic events at Mount Etna using pattern recognition techniques. *Geophysical Journal International*, 185, 253–264. <https://doi.org/10.1111/j.1365-246X.2011.04951.x>
- Chouet, B. (1986). Dynamics of a fluid-driven crack in three dimensions by the finite difference method. *Journal of Geophysical Research*, 91(B14), 13,967–13,992. <https://doi.org/10.1029/JB091iB14p13967>
- Chouet, B. A. (1996). Long-period volcano seismicity: Its source and use in eruption forecasting. *Nature*, 6572, 309–316. <https://doi.org/10.1038/380309a0>
- Chouet, B. A., & Matoza, R. S. (2013). A multi-decadal view of seismic methods for detecting precursors of magma movement and eruption. *Journal of Volcanology and Geothermal Research*, 252(Supplement C), 108–175. Retrieved from <http://www.sciencedirect.com/science/article/pii/S0377027312003435>. <https://doi.org/10.1016/j.jvolgeoes.2012.11.013>
- Cortés, G., Arámbula, R., Gutiérrez, L., Benítez, C., Ibáñez, J., Lesage, P., & García, L. (2009). Evaluating robustness of a HMM-based classification system of volcano-seismic events at Colima and Popocatepetl volcanoes. In *Geoscience and Remote Sensing Symposium, 2009 IEEE International, IGARSS 2009* (Vol. 2, pp. II-1012).
- Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3), 273–297. <https://doi.org/10.1007/BF00994018>
- Druitt, T. H., Young, S. R., Bapiste, B., Bonadonna, C., Calder, E. S., Clarke, A. B., & Voight, B. (2002). Episodes of cyclic Vulcanian explosive activity with fountain collapse at Soufrière Hills Volcano, Montserrat. *Geological Society, London, Memoirs*, 21(1), 281–306. <https://doi.org/10.1144/GSL.MEM.2002.021.01.13>
- Eronen, A., & Klapuri, A. (2000). Musical instrument recognition using cepstral coefficients and temporal features. In *Acoustics, Speech, and Signal Processing, 2000. ICASSP'00. Proceedings. 2000 IEEE International Conference on* (Vol. 2, pp. II753–II756).
- Esmaili, S., Krishnan, S., & Raahemifar, K. (2004). Content based audio classification and retrieval using joint time-frequency analysis. In *Icassp-2004, IEEE international conference on acoustics, speech, and signal processing* (Vol. 5, pp. 9–12).
- Esposito, A., Giudicepietro, F., D'Auria, L., Scarpetta, S., Martini, M., Coltelli, M., & Marinaro, M. (2008). Unsupervised neural analysis of very-long-period events at Stromboli volcano using the self-organizing maps. *Bulletin of the Seismological Society of America*, 98(5), 2449–2459.
- Fagerlund, S. (2007). Bird species recognition using support vector machines. *EURASIP Journal on Advances in Signal Processing*, 2007(1), 639–645.
- Friedman, J., Hastie, T., & Tibshirani, R. (2001). *The Elements of Statistical Learning* (Vol. 1). Berlin: Springer series in statistics Springer.
- Fujinaga, I., & MacMillan, K. (2000). Realtime recognition of orchestral instruments. In *Proceedings of the International Computer Music Conference (ICMC2000)* (Vol. 141, pp. 141–143).
- Gutiérrez, L., Ibáñez, J., Cortés, G., Ramírez, J., Benítez, C., Tenorio, V., & Isaac, Á. (2009). Volcano-seismic signal detection and classification processing using hidden Markov models. Application to San Cristóbal volcano, Nicaragua. In *Geoscience and Remote Sensing Symposium, 2009 IEEE International, igarss 2009* (Vol. 4, pp. IV–522).
- Hammer, C., Beyreuther, M., & Ohrnberger, M. (2012). A seismic-event spotting system for volcano fast-response systems. *Bulletin of the Seismological Society of America*, 102(3), 948–960.
- Han, N. C., Muniandy, S. V., & Dayou, J. (2011). Acoustic classification of Australian anurans based on hybrid spectral-entropy approach. *Applied Acoustics*, 72(9), 639–645.
- Harrington, R. M., & Brodsky, E. E. (2007). Volcanic hybrid earthquakes that are brittle-failure events. *Geophysical Research Letters*, 34, L06308. <https://doi.org/10.1029/2006GL028714>
- Hibert, C., Mangeney, A., Grandjean, G., Baillard, C., Rivet, D., Shapiro, N. M., et al. (2014). Automated identification, location, and volume estimation of rockfalls at Piton de la Fournaise volcano. *Journal of Geophysical Research: Earth Surface*, 119, 1082–1105. <https://doi.org/10.1002/2013JF002970>
- Hibert, C., Provost, F., Malet, J. P., Maggi, A., Stumpf, A., & Ferrazzini, V. (2017). Automatic identification of rockfalls and volcano-tectonic earthquakes at the Piton de la Fournaise volcano using a random forest algorithm. *Journal of Volcanology and Geothermal Research*, 340, 130–142.
- Huang, C. J., Yang, Y. J., Yang, D. X., & Chen, Y. J. (2009). Frog classification using machine learning techniques. *Expert Systems with Applications*, 36(2), 3737–3743.
- Ibáñez, J. M., Benítez, C., Gutiérrez, L. A., Cortés, G., García-yeguas, A., & Alguacil, G. (2009). The classification of seismo-volcanic signals using Hidden Markov Models as applied to the Stromboli and Etna volcanoes. *Journal of Volcanology and Geothermal Research*, 187(3–4), 218–226. <https://doi.org/10.1016/j.jvolgeoes.2009.09.002>
- Iguchi, M., Yakiwara, H., Tameguri, T., Hendrasto, M., & Ichi Hirabayashi, J. (2008). Mechanism of explosive eruption revealed by geophysical observations at the Sakurajima, Suwanosejima and Semeru volcanoes. *Journal of Volcanology and Geothermal Research*, 178(1), 1–9. <https://doi.org/10.1016/j.jvolgeoes.2007.10.010> (Dynamics of volcanic explosions: Field observations, experimental constraints and integrated modelling of volcanic explosions: Field observations, experimental constraints and integrated modelling)
- Inza, L. A., Mars, J. I., Métaxian, J. P., O'Brien, G. S., & Macedo, O. (2011). Seismo-volcano source localization with triaxial broad-band seismic array. *Geophysical Journal International*, 187(1), 371–384.
- Inza, L., Métaxian, J. P., Mars, J., Bean, C., O'Brien, G. S., Macedo, O., & Zandomeneghi, D. (2014). Analysis of dynamics of vulcanian activity of Ubina volcano, using multicomponent seismic antennas. *Journal of Volcanology and Geothermal Research*, 270, 35–52.
- Lahr, J., Chouet, B., Stephens, C., Power, J., & Page, R. (1994). Earthquake classification, location, and error analysis in a volcanic environment: Implications for the magmatic system of the 1989–1990 eruptions at Redoubt Volcano, Alaska. *Journal of Volcanology and Geothermal Research*, 62(1–4), 137–151.
- Langer, H., Falsaperla, S., Powell, T., & Thompson, G. (2006). Automatic classification and a-posteriori analysis of seismic event identification at Soufrière Hills volcano, Montserrat. *Journal of Volcanology and Geothermal Research*, 153(1–2 SPEC. ISS.), 1–10. <https://doi.org/10.1016/j.jvolgeoes.2005.08.012>

- Langet, N. (2014). Détection et caractérisation massives de phénomènes sismologiques pour la surveillance d'événements traditionnels et la recherche systématique de phénomènes rares (Doctoral dissertation). Ecole doctorale des Sciences de la Terre et de l'environnement, Université de Strasbourg. Retrieved from <http://www.theses.fr/2014STRAH017>
- Langley, P. A. T., Flamingo, L., & Edu, S. (1994). Selection of relevant features in machine learning, 127–131.
- Macedo, O., Métaxian, J. P., Taipe, E., Ramos, D., & Inza, A. (2009). Seismicity associated with the 2006-2008 eruption, Ubinas volcano, 262–270.
- 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. *Seismological Research Letters*, 88(3), 878–891. <https://doi.org/10.1785/0220160189>
- Malfante, M. (2018). Automatic analysis architecture. Retrieved from blabla.
- McNutt, S. (1992). Volcanic tremor. *Encyclopedia of Earth System Science*, 4, 417–425.
- 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* (pp. 99–146). Berlin, Heidelberg: Springer Berlin Heidelberg.
- Morgan, S. C., Malone, S. D., Qamar, A. I., Thelen, W. A., Wright, A. K., & Caplan-Auerbach, J. (2008). Seismicity associated with renewed dome building at Mount St. Helens, 2004-2005: Chapter 2 in A volcano rekindled: The renewed eruption of Mount St. Helens, 2004-2006 (Tech. Rep.) Reston, VA: U.S. Geological Survey. from <http://pubs.er.usgs.gov/publication/pp17502> (Report).
- Narváez, L., Torres, C. R. A., Gómez, M. D. M., Cortés, J. G. P., Cepeda, H., & Stix, J. (1997). 'Tornillo'-type seismic signals at Galeras volcano, Colombia, 1992–1993. *Journal of Volcanology and Geothermal Research*, 77(1), 159–171. [https://doi.org/10.1016/S0377-0273\(96\)00092-3](https://doi.org/10.1016/S0377-0273(96)00092-3)
- Neuberg, J., Lockett, R., Baptie, B., & Olsen, K. (2000). Models of tremor and low-frequency earthquake swarms on Montserrat. *Journal of Volcanology and Geothermal Research*, 101(1), 83–104. [https://doi.org/10.1016/S0377-0273\(00\)00169-4](https://doi.org/10.1016/S0377-0273(00)00169-4)
- Ohminato, T. (2006). Characteristics and source modeling of broadband seismic signals associated with the hydrothermal system at Satsumaïwojima volcano, Japan. *Journal of Volcanology and Geothermal Research*, 158(3), 467–490. <http://www.sciencedirect.com/science/article/pii/S0377027306003325>
- Ohnberger, M. (2003). Continuous automatic classification of seismic signals of volcanic origin at Mt. Merapi, Java, Indonesia continuous automatic classification of seismic signals of volcanic origin at Mt. Merapi, Java, Indonesia. (July 2001).
- Quinlan, J. R. (1986). Induction of decision trees. *Machine Learning*, 1(1), 81–106. <https://doi.org/10.1007/BF00116251>
- Traversa, P., Lengliné, O., Macedo, O., Métaxian, J. P., Grasso, J. R., Inza, A., & Taipe, E. (2011). Short term forecasting of explosions at Ubinas volcano, Perú. *Journal of Geophysical Research*, 116, B1130. <https://doi.org/10.1029/2010JB008180>
- Tucker, S., & Brown, G. J. (2005). Classification of transient sonar sounds using perceptually motivated features. *IEEE Journal of Oceanic Engineering*, 30(3), 588–600.
- White, R. A., Miller, A. D., Lynch, L., & Power, J. (1998). Observations of hybrid seismic events at Soufriere Hills Volcano, Montserrat: July 1995 to September 1996. *Geophysical Research Letters*, 25(19), 3657–3660. Retrieved from <https://doi.org/10.1029/98GL02427>
- Zaugg, S., Van Der Schaar, M., Houégnigan, L., Gervaise, C., & André, M. (2010). Real-time acoustic classification of sperm whale clicks and shipping impulses from deep-sea observatories. *Applied Acoustics*, 71(11), 1011–1019.