Journal of Volcanology and Geothermal Research 407 (2020) 107092

Contents lists available at ScienceDirect

Journal of Volcanology and Geothermal Research

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

# On finding possible frequencies for recognizing microearthquakes at Cotopaxi volcano: A machine learning based approach

Román Lara-Cueva [a,*], Julio C. Larco [a], Diego S. Benítez [b], Noel Pérez [b], Felipe Grijalva [c], Mario Ruiz [d]

[a] Grupo de Investigación en Sistemas Inteligentes (WiCOM-Energy), Centro de Investigaciones de Redes Ad-Hoc (CIRAD), Departamento de Eléctrica, Electrónica y Telecomunicaciones, Universidad de las Fuerzas Armadas ESPE, Sangolquí 171103, Ecuador
[b] Colegio de Ciencias e Ingenierías "El Politécnico", Universidad San Francisco de Quito USFQ, Quito 170157, Ecuador
[c] Departamento de Electrónica, Telecomunicaciones y Redes de Información, Escuela Politécnica Nacional, Quito 170109, Ecuador
[d] Instituto Geofísico, Escuela Politécnica Nacional, Quito 170109, Ecuador

ARTICLE INFO

*Article history:*
Received 1 July 2020
Received in revised form 8 October 2020
Accepted 14 October 2020
Available online 21 October 2020

*Keywords:*
Volcanic events
Microearthquakes classification
Support vector machines
Decision trees
Feature extraction
Feature selection
Power spectrum density

ABSTRACT

Adequate detection and classification of seismic events are crucial for understanding the internal status of a Volcano. Machine learning-based classifiers use different features from the time, frequency, and scale domains related to seismic events. Regarding power spectrum-based features, several methods can be used to compute such features. However, the more suitable method for analyzing volcanic activity is undetermined. This paper presents a study about the main frequency bands, which allows maximizing the performance metrics of an automated classifier for long-period (LP) and volcano-tectonic (VT) events based on parametric (Yule-Walker and Burg) and non-parametric (Welch and Multitaper) power spectrum density estimation methods. Feature selection using embedded (pruning) and wrapper (recursive feature elimination) methods was applied to select the main frequencies that maximize the balanced error rate of suitable classification algorithms, such as decision trees (DT) and support vector machines (SVM). Bootstrapping was used to estimate a confidence interval for the frequencies of the microearthquakes. An amplitude threshold difference of at least 3 dB was used to guarantee that possible frequency features that characterize each type of event do not overlap between classes. The method who achieved the worst overall performance was not considered by the voting strategy. A Dataset from Cotopaxi volcano was used to test the proposed classification schema. The best results show for DT classifier a total of 10 key frequencies, while for SVM classifier 39 key frequencies grouped in three main frequency bands, as main features to distinguish LP events from VT earthquakes. The best classification results were achieved by the Welch method with the DT and by the Multitaper method with the SVM classifiers. Furthermore, the study confirms that there is a frequency band above 40 Hz, which seems like a critical feature for the detection and classification of stages.



## 1. Introduction

In the last decade, many works appearing in the literature have used machine learning classifiers (MLC) such as artificial neural networks (ANN), decision trees (DT), support vector machines (SVM), evolutionary algorithms, among others, for classifying microearthquakes originated from volcanoes (Malfante et al., 2018; Laasri et al., 2015; Curilem et al., 2014). SVM-based approaches have also been proposed for microearthquakes detection (Lara-Cueva et al., 2016a; Ruano et al., 2014). In this context, the use of automatic recognition systems based on MLCs can help experts improve the time-consuming task of detecting and classifying the microearthquakes. In their feature extraction stages, most of the previous works deal with a well-known set of features (quantities derived from inputs (Bishop et al., 2006)) related to the microearthquakes in both time and frequency domains (Cárdenas-Peña et al., 2013; Álvarez et al., 2012). Moreover, there are also studies using together features from time, frequency, and scale domains (Soto et al., 2018; Lara-Cueva et al., 2016b; Lara-Cueva et al., 2015; Duque et al., 2020), and recently, the intensity statistics, shape, and texture features computed from the seismic event pattern represented in the grey-level spectrogram image (Pérez et al., 2020a).

Spectrum analysis based on the power spectral density (PSD) has also been used to analyze and classify microearthquakes (Carniel, 2014), and is a common practice for event classification at Instituto Geofísico of Escuela Politécnica Nacional (IGEPN) and other volcano observatories around the world (Chouet et al., 1994; Wassermann, 2012; Ruiz et al., 1998), defining a different frequency band ($f_b$) for each type of microearthquakes. To simplify the notation of frequency bands throughout the text, we will denote $f_b$ as an element belonging ($\in$) to

* Corresponding author.
*E-mail address:* ralara@espe.edu.ec (R. Lara-Cueva).

https://doi.org/10.1016/j.jvolgeores.2020.107092
0377-0273/

the range between the lower ($F_L$) and upper ($F_H$) frequencies of a given interval, that is, $f_b \in (F_L, F_H)$ Hz. The volcano-tectonic (VT) events, for example, are earthquakes taking place in a volcanic environment with a variable duration, usually of less than 30*s*. Their spectral content is wide and shows a $f_b \in (3, 14)$ Hz (McNutt, 2005). Whereas, the long-period (LP) events, also known as low-frequency events, show a lack of distinguishable phases and a typical duration time below 90*s*, while their main spectral content is limited at narrow $f_b \in (2, 7)$ Hz (Chouet, 1996). The hybrid (HYB) earthquakes share the features of VT and LP events, and signals at high frequencies characterize them, generally with a wide spectral bandwidth above 10 Hz. The tremors (TRE), on the other hand, are characterized by their constant amplitude and long duration, and they are the most distinctive signals generated by volcanoes. They are widely considered the most complex ones, with a period that can range from a few minutes to several days (Lahr et al., 1994; Ibáñez and Carmona, 2000).

However, various pitfalls found in spectral analysis of time series associated with the sampling rate and aliasing, scalloping loss or picket-fence effects, finite data length, trend removal, spectral leakage, and spectral smearing, should be avoided (Ifeachor and Jervis, 2002). Although various parametric and non-parametric spectrum estimation techniques are available in the literature for computing the PSD, to our knowledge, no previous studies have reported a detailed analysis related to the main features that can be extracted in the frequency domain from microearthquakes by using different PSD estimator methods, and which could be used to distinguish among different microearthquakes by using MLCs.

In such a sense, the aim of this work is twofold. First, to identify the main frequency bands, which allow maximizing the performance metrics of an automated classifier for LP and VT events based on PSD features. Second, to compare the performance of such classifiers using parametric and non-parametric power spectrum estimation methods. The intention behind the proposed approach is to apply a three-stage solution, consisting of (1) feature extraction by using parametric and non-parametric power spectrum estimation methods, (2) feature selection by using embedded and wrapper methods and (3) a definition stage which selects the main frequencies that maximize the balanced error rate (BER) of the classifiers, based on a resampling method for estimating their confidence intervals (CI). We focus our analysis on a data set of microearthquakes obtained from the Cotopaxi volcano, located in Ecuador.

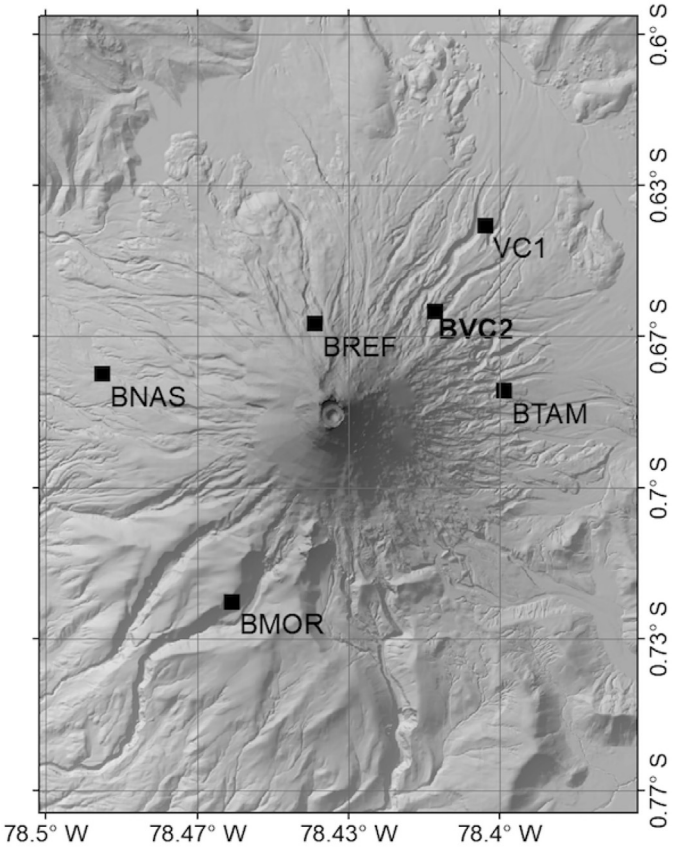


**Fig. 1.** Geographical location of stations deployed by IGEPN.

## 2. Proposed methodology

In this work, we focus on developing an automatic classifier to distinguish LP events from VT earthquakes by considering features obtained by the spectral analysis of the seismic signal with parametric and non-parametric power spectrum estimators. Such a tool could then be used by experts to quantify seismic activity and alert institutions in charge of emergency situations. The proposed system consists of the following stages:

1) A pre-processing subsystem initially treats the acquired microearthquakes.
2) A feature extraction subsystem is used to extract features. For this subsystem, PSD was computed using well-known parametric and non-parametric power spectrum estimators over the time-domain signals, such as the Yule-Walker and Burg methods, and the Welch and Multitaper methods, respectively.
3) Next, a feature selection stage is applied using an embedded method with a DT classifier, and a wrapper method with an SVM classifier.
4) Finally, an analysis of the main frequency bands with a CI of 97.5% is performed using the bootstrap method (Efron, 1981; Efron and Tibshirani, 1994), which could permit identifying frequency related features able to discriminate LP events from VT earthquakes.

### 2.1. Experimental dataset

We focus our analysis on a dataset of microearthquakes (*MicSigV1*) obtained from the Cotopaxi volcano, located in Ecuador (latitude 0° 41′05” S and longitude 78°25′54.8” W). This dataset is available from the ESeismic[1] repository (Pérez et al., 2020b), which is the first annotated Ecuadorian volcano seismic public repository.

The *MicSigV1* contains a total of 1187 independent microearthquakes records. These were recorded at two different seismic stations (BREF and VC1) installed on flanks of the volcano, as showed in Fig. 1, and sampled at 50 Hz and 100 Hz, respectively. These particular recordings have been extracted, identified, and labeled by experts at the IGEPN from continuous monitoring seismogram recordings. The dataset contains samples distributed in five classes corresponding to LP, VT, regional (REG), HYB, and icequakes (ICE). Due to the small number of samples from REG ($N_{REG}$=27), HYB ($N_{HYB}$=8), and ICE ($N_{ICE}$=7) events available within the *MicSigV1* dataset, we only considered LP and VT events classes for this study, where the number of records for such classes was $N_{LP}$=1044 and $N_{VT}$=101, respectively, giving a total of $N = N_{LP} + N_{VT} = 1145$ records. Fig. 2 shows some examples of the time signals available within the *MicSigV1* and their respective PSD. Microearthquakes at Cotopaxi volcano have a typical duration below 60s.

### 2.2. Pre-processing subsystem

Each available record contains a single microearthquake, and records sampled at 50 Hz were upsampled to 100 Hz to standardize

[1] ESeismic repository was provided by courtesy of the IGEPN and collaborators, and it is available at http://www.igepn.edu.ec/eseismic_web_site/index.php, where you should register and fill a disclaimer agreement in order to obtain the data.

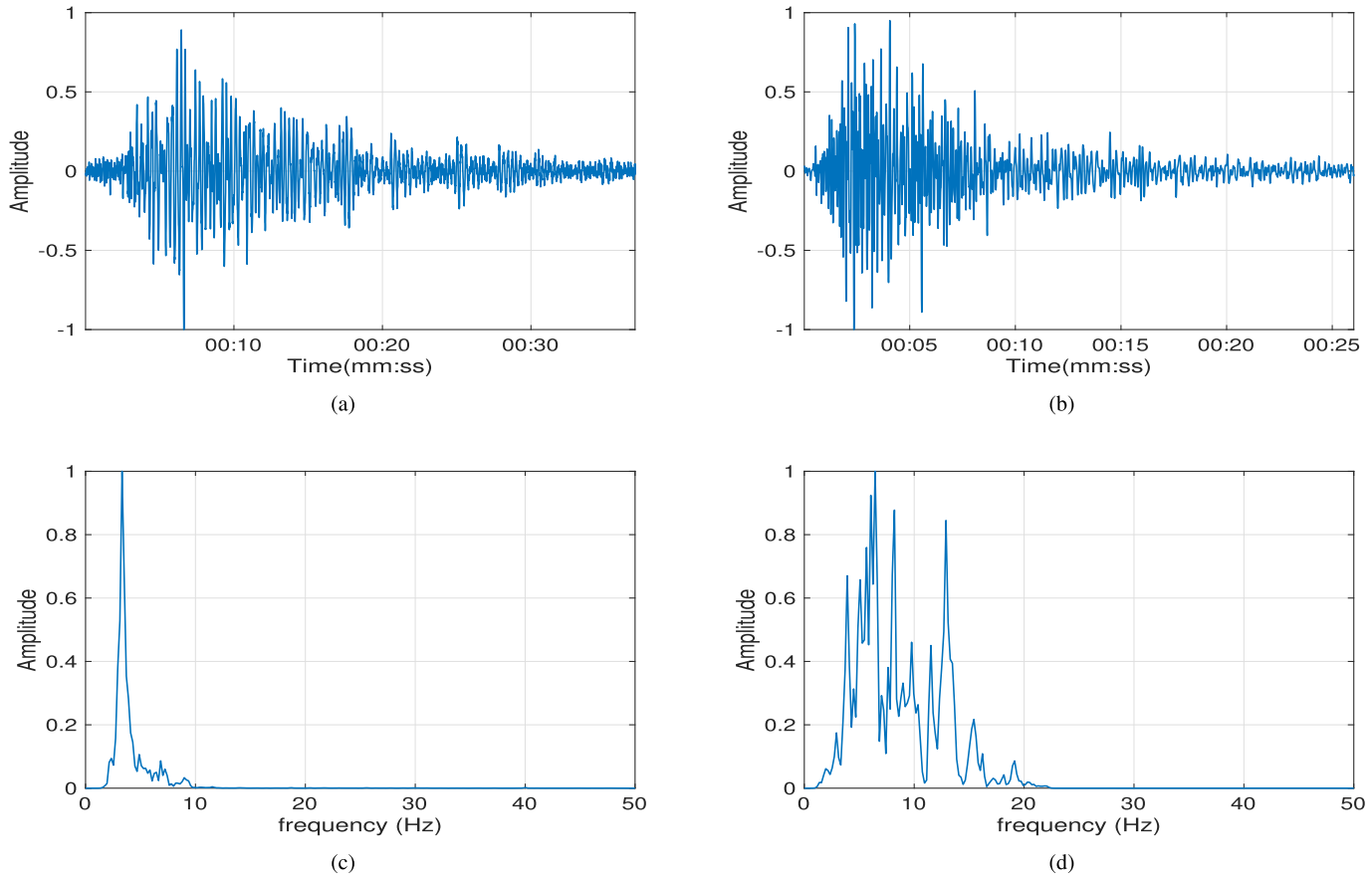


**Fig. 2.** Examples of LP and VT microearthquakes originated at Cotopaxi volcano and contained in *MicSigV1* database: (a) LP event and (b) VT earthquake in the time domain. (c) and (d) corresponding normalized PSD with Welch method for the LP and VT events, respectively.

their sampling rate. Each signal was pre-processed with a 128-th order zero-phase noncausal high pass FIR filter with a cutoff frequency of 1 Hz to remove the undesired sea microearthquakes found around 0.2 Hz (Kenneth, 2001).

### 2.3. Feature extraction stage

The purpose of PSD is to estimate the power spectral density of a stochastic process from a finite set of observations. Spectral estimation can be divided into parametric and non-parametric methods. The main difference between then is that in parametric PSD estimation, the stochastic process is assumed to have some analytical form that depends on a set of parameters (Castanié, 2013). In this paper, we consider that the analytical form is an nth-order autoregressive (AR) model. Two common parametric approaches are the Yule-Walker and the Burg PSD estimators.

On the other hand, non-parametric spectral methods estimate the PSD of the stochastic process without assuming a particular structure or analytical form. Among these methods, two popular approaches are the Welch and Multitaper PSD estimators.

It is worth noting that the stochastic nature of volcano seismic signals is the main reason to use PSD estimate methods, instead of the conventional FFT, which is only defined for deterministic signals. These methods try to find a consistent PSD estimate aiming at reducing bias and variability with respect to the true PSD. The nonparametric methods used in this paper (i.e., Welch and Multitaper) are based on the periodogram, which in turn is numerically calculated from the squared FFT magnitude. However, note that in contrast to a simple FFT, the periodogram ponderates the time-domain signal with a window function prior to the FFT calculation to mitigate the bias with respect to the true PSD. Moreover, Welch and Multitaper methods average several periodograms to reduce variability and produce a more consistent spectral estimate. Regarding the parametric methods used in this work (i.e., Yule-Walker and Burg), their goal is the same, i.e., to estimate the true PSD of a stochastic signal. But, in contrast to nonparametric techniques, they assume an underlying analytical AR process.

Concretely, for each microearthquake record $\mathbf{s}_i$ for $i = 1 \ldots N$, a set of relevant PSD-based features were computed by using two Parametric (i.e., Yule-Walker and Burg) and two Non-Parametric (i.e., Welch and Multitaper) spectral estimators over the time-domain pre-processed signals. Thus, a feature matrix for each spectral estimator was obtained. For all PSD estimators, we evaluate a $n_f = 512$-point FFT.

The Welch method first divides the time-domain signal into segments (which may overlap). Then, from each segment, a windowed periodogram is calculated, and finally, the periodograms are averaged to obtain the PSD estimate (Alessio, 2015). The Windowing and averaging aims at reducing bias and variance on the PDS estimate, respectively.

The PSD was estimated by applying the Welch method to each $\mathbf{s}_i$, yielding feature vectors $\mathbf{w}_i = g_w(\mathbf{s}_i)$, where $g_w(\cdot)$ is the operator producing the Welch-PSD. Each record was windowed with a $n_f = 512$-point Kaiser window, with $\beta$=4.53, and a side-lobe attenuation of 50 dB, yielding a Welch-feature matrix, $\mathbf{W}$, as:

$$\mathbf{W} = \left[\mathbf{w}_1^T, \mathbf{w}_2^T, \ldots, \mathbf{w}_N^T\right]^T.$$

The multitaper method averages periodograms weighted by a set of mutually orthogonal windows (i.e., tapers) aiming at reducing bias and variance on the PSD with respect to the periodogram (Babadi and

Brown, 2014). Moreover, in contrast to Welch, it does not require to divide the signal into segments, which avoids to lose frequency resolution and, at the same time, keeps a low variance. The multitaper approach has been successfully applied to Electroencephalogram (EEG) signals, which are highly susceptible to a variety of noises (Babadi and Brown, 2014). In this paper, for this method the resolution bandwidth of estimation was set to 4, yielding a Multitaper-feature matrix, **M**, as:

$$\mathbf{M} = \left[\mathbf{m}_1^T, \mathbf{m}_2^T, \ldots, \mathbf{m}_N^T\right]^T.$$

The Yule-Walker method fits an nth-order AR model by minimizing the forward prediction error in the least-squares sense (Friedlander and Porat, 1984). Similarly, the PSD was computed by applying the Yule-Walker method to each $\mathbf{s}_i$, yielding feature vectors $\mathbf{k}_i = g_k(\mathbf{s}_i)$, where $g_k(\cdot)$ is the operator yielding the Yule-Walker-PSD.

The PSD was also obtained with the Burg method, which fits an AR model to the time-domain signal, minimizing the forward and backward prediction error. In general, Burg offers a higher frequency resolution than Yule-Walker.

This method is applied to each $\mathbf{s}_i$, yielding feature vectors $\mathbf{b}_i = g_b(\mathbf{s}_i)$, where $g_b(\cdot)$ is the operator yielding the Burg-PSD. For both methods, an 8th order autoregressive model was used to produce the PSD estimate, yielding a Yule-Walker, **K**, and Burg, **B**, feature matrices respectively, as follows:

$$\mathbf{K} = \left[\mathbf{k}_1^T, \mathbf{k}_2^T, \ldots, \mathbf{k}_N^T\right]^T,$$

$$\mathbf{B} = \left[\mathbf{b}_1^T, \mathbf{b}_2^T, \ldots, \mathbf{b}_N^T\right]^T.$$

These $N \times n$ matrices have $N$ instances (1044 LP + 101 VT), whilst the initial number features is $n = \frac{n_f}{2} + 1$, thus, providing 257 features. Note that at this point, four types of feature matrices in the frequency domain have been defined with a different estimator.

### 2.4. Feature selection stage and classifiers

Several methods have been deployed to select the main features, typically there exist three categories, namely: filter methods, embedded methods, and wrapper methods. Filter methods select features according to their scores evaluated via univariate statistics, independent of the classifier performance. Meanwhile, Embedded methods are used to optimize the performance of a learning model; in this sense, the pipeline by considering the feature selection and classification stages (learning block) can not be separated. Finally, Wrapper methods use the learning block to measure the quality of subsets of features; in this context, these methods measure the relevance of features based on the classifier performance (Luo, 2020). For these reasons, we benchmarked two commonly used feature selection techniques: embedded, and wrapper methods, to obtain matrices $\mathbf{W}'$, $\mathbf{M}'$, $\mathbf{K}'$, and $\mathbf{B}'$, which contained most of the discriminative information to classify the events while avoiding overfitting and redundancy, related to the performance of the learning model. For the embedded method, we have selected an algorithm, which as the main criterion uses Mutual Information (MI) between feature $x$ and the output $y$, as follows:

$$MI(x;y) = H(y) - H(y|x),$$

where $H(y)$ is the marginal entropy, and $H(y|x)$ is the conditional entropy between output $y$ and feature set $x$, and iteratively builds by splitting data according to its importance for the classification task, in this sense a DT based classifier was used (Bishop et al., 2006).

DT is a non-parametric supervised learning method used for classification and regression and one of the most widely used methods for inductive inference. The goal of this method is to create a model that predicts the value of a target variable by learning simple decision rules inferred from the data features. It is robust to noisy data, and searches in a completely expressive hypothesis space and, hence avoiding the difficulties of restricted hypothesis spaces. Its preferences are for small trees than large trees, which promotes classifiers with good generalization capabilities. Currently, many algorithms have been developed for learning DT, which obtain a top-down, greedy search in the feature space. The depth or *leafiness* of the tree is the free parameter for this Machine Learning technique. It has to be adjusted for maximizing the classification performance while avoiding overfitting to the training set (Esposito et al., 1997). The measures developed for this function are based on the impurity of the child nodes. These measures include the average amount of information contained in each event (Entropy), and a measure of statistical dispersion intended to represent the measure of inequality (Gini Index) as follows:

$$E(t) = -\sum_{i=0}^{c-1} p(i|t)\,\log_2 p(i|t),$$

$$G(t) = 1 - \sum_{i=0}^{c-1} [p(i|t)]^2,$$

where $p(i|t)$ denotes the fraction of records belonging to class $i$ at a given node $t$, and $c$ is the number of classes.

Meanwhile, for a wrapper method, we chose Recursive Feature Elimination (RFE) (Kohavi and Sommerfield, 1995), which is based on a backward elimination method. It operates by iteratively removing features from data, trying to choose the features which lead to the largest margin of class separation by using SVM as classifier. For our case $\nu$-SVM was used, this method considers just $\nu$ as a free parameter, which controls the number of support vectors. The $\nu$-SVM algorithm is defined in summary as follows, the interested reader can refer to (Schölkopf and Smola, 2002) for details. Given a labeled training data set $\{\boldsymbol{x}_i, y_i\}_{i=1}^n$, where $\boldsymbol{x}_i \in \mathbb{R}^N$ and $y_i \in \{-1, +1\}$, and given a nonlinear mapping $\phi(\cdot)$, the $\nu$-SVM method solves:

$$\min_{\boldsymbol{w}, \xi_i, b, \rho} \left\{ \frac{1}{2}\|\boldsymbol{w}\|^2 + \nu\rho + \frac{1}{n}\sum_{i=1}^{n} \xi_i \right\} \tag{1}$$

subject to

$$y_i(\langle \phi(\boldsymbol{x}_i)\boldsymbol{w} > + b) \geqslant \rho - \xi_i \qquad \forall i = 1, \ldots, n \tag{2}$$

$$\rho \geqslant 0, \xi_i \leqslant 0 \qquad \forall i = 1, \ldots, n \tag{3}$$

where $\boldsymbol{w}$ and $b$ define a linear classifier in the feature space, and $\xi_i$ are positive slack variables enabling to deal with errors. The appropriate choice of nonlinear mapping $\phi$ guarantees that the transformed samples are more likely to be linearly separable in the (higher dimensional) feature space. In this formulation, variable $\rho$ is controlled with a coefficient of $\nu$, which adds another degree of freedom to the margin, the size of the margin increases linearly with $\rho$. Therefore, the trade-off between the training error and the generalization error is controlled in the $\nu$-SVM formulation by adjusting $\nu$ in the range [0,1], which acts as an upper bound on the fraction of margin errors, and it is also a lower bound on the fraction of support vectors.

Primal problem (1) is solved by using its dual problem counterpart, yielding $\boldsymbol{w} = \sum_{i=1}^{n} y_i \alpha_i \phi(\boldsymbol{x}_i)$ (see (Schölkopf and Smola, 2002) for further theoretical and algorithmic details), and the decision function for any test vector $\boldsymbol{x}_*$ is finally given by

$$f(\boldsymbol{x}_*) = sgn\left(\sum_{i=1}^{n} y_i \alpha_i K(\boldsymbol{x}_i, \boldsymbol{x}_*) + b\right) \tag{4}$$

where $\alpha_i$ are Lagrange multipliers corresponding to constraints in (1), being the support vectors those training samples $\boldsymbol{x}_i$ with non-zero Lagrange multipliers $\alpha_i \neq 0$; and the bias term $b$ is calculated by using the *unbounded* Lagrange multipliers as $b = 1/k\sum_{i=1}^{k}(y_i - \langle \phi(\boldsymbol{x}_i), \boldsymbol{w}\rangle)$,

where $k$ is the number of *unbounded* Lagrange multipliers ($0 < \alpha_i < C$). Note that the particularity of SVM is that decision function $f(\mathbf{x})$ is a function of a small subset of the training examples, which are the support vectors. Those are the examples that are the closest to the decision boundary and lie on the margin and wrong-class cases. The existence of such support vectors is at the origin of the computational properties of SVM and their competitive classification performance.

### 2.5. Strategy to define the most relevant frequency bands of microearthquakes

We developed a strategy to select the main frequencies or frequency bands that could identify LP events from VT earthquakes. First, we identified the main frequencies of each event by using a feature selection stage. Then, we use the bootstrapping technique (Efron and Tibshirani, 1994), which allows estimating the sampling distribution of the mean ($\mu$), standard deviation ($\sigma$), and the CI of the frequencies of the microearthquakes as follows: for each group of microearthquakes, LP and VT, to which we have applied the spectral estimator, we build a new group of $N_r = 1000$ resampled instances for each event class such that $N_{LP}^r = N_{VT}^r = N_r$.

For example by considering the Welch-PSD, we have the initial matrix $\mathbf{W}$ which is formed by $\mathbf{w}_i^{|top}$ elements and $i$ varies from 1 to $N$, the re-sampled matrix $\mathbf{W^r}$ will be conformed of $\mathbf{w}_j^{r|top}$ new elements and $j = 2 \times N_r$, where each $\mathbf{w}_j^{r|top}$ is conformed of $\frac{n_f}{2} + 1$ elements. Each new element is selected randomly, but keeping the order from any element that belongs to the LP events class and the VT events class, respectively, resulting therefore in new 1000 resampled LP events and 1000 resampled VT events. As previously mentioned, we consider the $\mu$ value and construct the $CI = \mu \pm 3 \times \sigma$, to have a 97.5% of confidence. Finally, we consider a safety band between similar frequencies from LP and VT events to guarantee that possible frequency features taken from the PSD of the events do not overlap. The difference in amplitude between similar frequency features must be of at least 3 dB to ensure they are discriminative enough for microearthquake classification.

### 2.6. Performance

The classification performance was measured in terms of the percentage of accuracy (A), precision (P), sensitivity or recall (R), specificity (S) criterion, and the BER score, which are defined as:

$$A = \frac{N_C}{N_T} \times 100,$$
$$P = \frac{N_{TP}}{N_{TP} + N_{FP}} \times 100,$$
$$R = \frac{N_{TP}}{N_{TP} + N_{FN}} \times 100,$$
$$S = \frac{N_{TN}}{N_{TN} + N_{FP}} \times 100,$$
$$BER = 1 - \frac{R + S}{2 \times 100},$$

where $N_C$ is the number of correctly classified events, $N_T$ is the total number of events used to feed the classifier, $N_{TP}$ is the number of true positives, $N_{FN}$ is the number of false negatives, $N_{TN}$ is the number of true negatives, and $N_{FP}$ is the number of false positives. We calculated these performance measures using training and testing by using cross-validation. Note that the BER score is a balanced metric that equally weights errors in $R$ and $S$, in this sense this value represents the misclassification events, for this reason, IGEPN suggest this value have to be less than 0.01, which means that 1 of 100 events has been misclassified. Finally, a voting system based on the three best classification results was considered, the worst outcome for each classifier in each scenario, and in terms of the performance metrics, was discarded. To do that, predictions $y_i$ of the three best classification results, for i.e., could be $[-1, +1, +1]$, by a majority, the voting system elects $+1$; thus, to implement this, if the aggregate sum of the three classifiers is greater than $+1$, then the result would be $+1$ (VT); otherwise, it would be $-1$ (LP).

## 3. Results and discussion

This section presents the results obtained by applying the proposed methodology to our dataset from the Cotopaxi Volcano. To avoid imbalanced class problems, an experimental dataset was constructed with $L = 101$ event samples of each class, i.e., a labeled training dataset $\{\boldsymbol{x}_i, y_i\}_{i=1}^{L}$, where $\boldsymbol{x}_i \in \mathbb{R}^n$ (PSD feature vectors) and $y_i \in \{-1, +1\}$ with $y_i \in \{LP, VT\}$, respectively. 60% instances of LP and VT events of $L$ for each class were used for training, and the rest of the dataset was used for testing. Independence, among all the records in each of these sets, was ensured. The PSD feature vectors were normalized, first by using the classical centering and scaling of the data and then normalized for the highest amplitude value to bring all values into the range [0,1]. The experiments were carried out in Matlab™, on a Core I5 PC with 3.1 GHz and 8GB RAM.

### 3.1. Results using DT classifier

The DT algorithm produced a model for each of the spectral estimator matrixes previously obtained, making it possible to induce a top-down tree, as depicted in Fig. 3 (top row). For example, for the Burg method, the DT algorithm selected four key features $X_m$, where $m$ is the number of the selected feature. In this case, corresponding to the amplitude of the PSD computed by the Burg method in a determined frequency. Starting from the top node with the rule $X_{57} \geq 0.034$ (corresponding to the amplitude value at 11.1 Hz), $X_{80} \geq 0.016$ (corresponding to the amplitude value at 15.6 Hz), $X_{215} \geq 2.07e^{-9}$ (corresponding to the amplitude value at 41.8 Hz), and $X_{10} \geq 8.85e^{-3}$ (corresponding to the amplitude value at 2.0 Hz), which made possible to classify the event into any of the five possible leaves, as depicted in Fig. 3(a).

Similarly, for the Multitaper method, as illustrated in Fig. 3(b), the DT algorithm selected the top node with the rule $X_{69} \geq 0.017$ (corresponding to the amplitude value at 13.4 Hz), then $X_{42} \geq 0.279$ (corresponding to the amplitude value at 8.2 Hz), $X_{14} \geq 0.244$ (corresponding to the amplitude value at 2.7 Hz), $X_{43} \geq 0.114$ (corresponding to the amplitude value at 8.4 Hz), and $X_{31} \geq 0.184$ (corresponding to the amplitude value at 6.0 Hz) which made possible to classify the event into any of the six possible leaves. Finally, we observe in Fig. 3(c), the tree constructed by the DT algorithm for the Welch method, the top node was selected with the rule $X_{69} \geq 0.017$ (corresponding to the amplitude value at 13.4 Hz), and then $X_{49} \geq 0.218$ (corresponding to the amplitude value at 9.5 Hz), and $X_{14} \geq 0.364$ (corresponding to the amplitude value at 2.7 Hz) which made possible to classify the event into any of the four possible leaves. The DT presented the worst results for the Yule-Walker method, and thus we omit to present the resulting tree here since, in the final stage the voting system only considers the best three results.

Table 1 shows the results obtained with the testing set, which containing $N_{Test}$=1023 independent cases, $N_{LP}$=983 and $N_{VT}$=40. In this sense, the best result was obtained with the Welch method and the worst with the Yule-Walker method.

A total of 10 key nonrepeating PSD frequencies were identified by the previously described DTs (2.0 Hz, 2.7 Hz, 6.0 Hz, 8.2 Hz, 8.4 Hz, 9.5 Hz, 11.1 Hz, 13.4 Hz, 15.6 Hz, and 41.8 Hz). Therefore, we then used these ten key features to feed our DT classifier. The classifiers determined five critical features for the Burg (11.1 Hz, 15.6 Hz, 41.8 Hz, 8.4 Hz, and 2.7 Hz), and Multitaper methods (13.4 Hz, 2.7 Hz, 8.2 Hz, 8,4 Hz, and 6.0 Hz), and three key features for the Welch (13.4 Hz, 2.7 Hz, and 6.0 Hz), and Yule-Walker methods (11.1 Hz, 2.7 Hz, and 9.5 Hz). Table 2 shows the results obtained with the testing set. As can be seen, the Welch estimator achieved the best overall result.

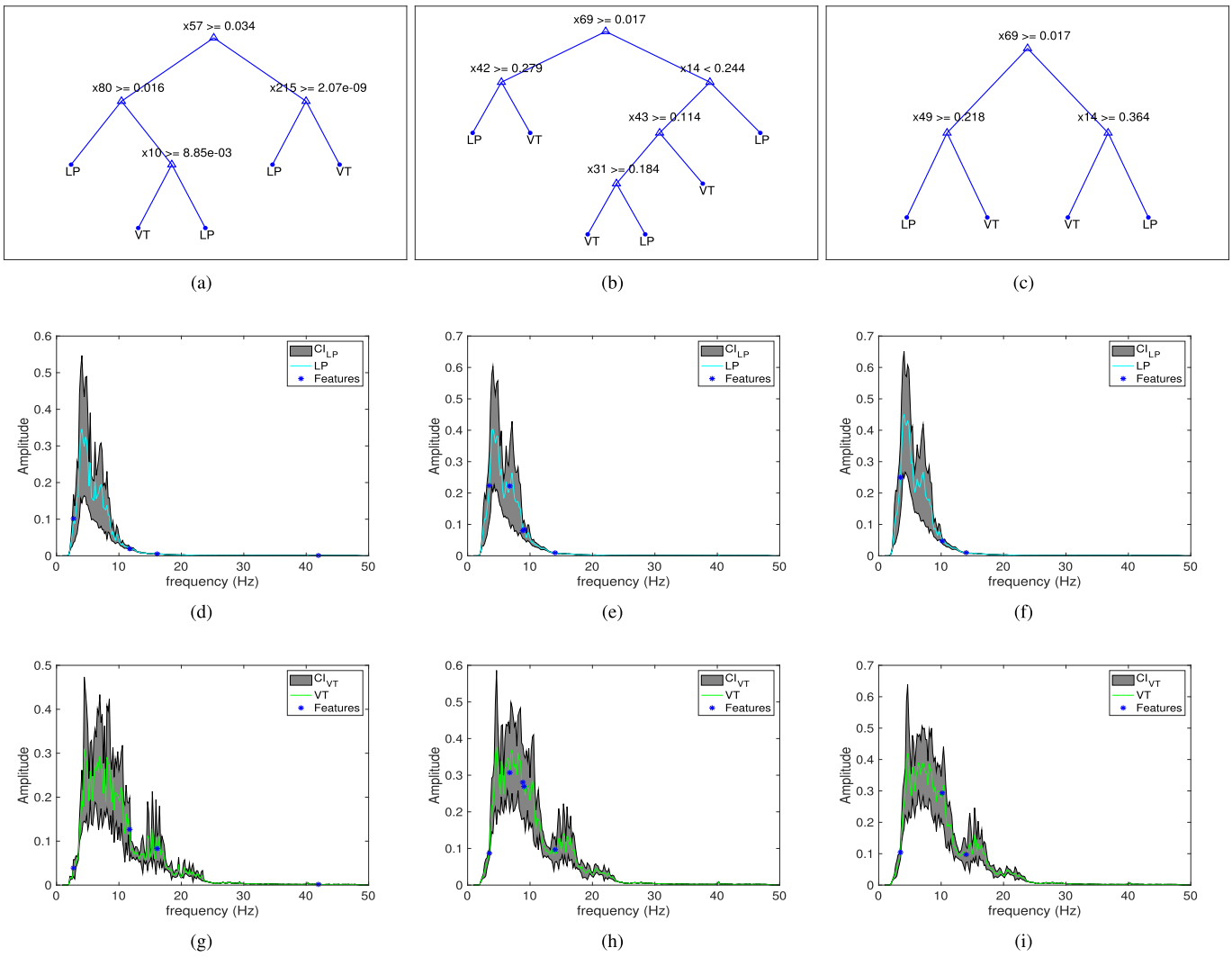


**Fig. 3.** Tree representations obtained for the different PSD method considering main frequency features by the DT classifier: top row: main features selected in the tree by (a) Burg, (b) Multitaper, and (c) Welch methods; middle row: mean of LP events (cyan lines) for (d) Burg, (e) Multitaper, and (f) Welch methods; bottom row: mean of VT events (green lines) for (g) Burg, (h) Multitaper, and (i) Welch methods, respectively. Retrieved features in each case are represented by the blue points which are within the 97.5% of CI (grey zones).

**Table 1**
Experimental classification performance obtained using the DT classifier. The confusion matrix for each PSD matrix and the number of main features retrieved after applying the feature selection stage are presented.

| Matrix | Label | Estimated Label | | N. Features | *A* (%) | *P* (%) | *R* (%) | *S* (%) | *BER* |
|---|---|---|---|---|---|---|---|---|---|
| | | VT | LP | | | | | | |
| B | VT | 32 | 73 | 4 | 92 | 30 | 80 | 95 | 0.14 |
| | LP | 8 | 910 | | | | | | |
| M | VT | 37 | 84 | 5 | 91 | 31 | 93 | 92 | 0.08 |
| | LP | 3 | 899 | | | | | | |
| W | VT | 39 | 86 | 3 | 92 | 31 | 98 | 91 | 0.05 |
| | LP | 1 | 897 | | | | | | |
| Y | VT | 38 | 192 | 6 | 81 | 17 | 95 | 81 | 0.12 |
| | LP | 2 | 791 | | | | | | |
| Voting | VT | 36 | 28 | | 97 | 56 | 90 | 97 | 0.06 |
| | LP | 4 | 955 | | | | | | |

**The Yule-Walker method (Y) achieved the worst average performance, and therefore it was not considered by the voting strategy.

**Table 2**
Experimental seismic event classification performance obtained using the DT classifier with ten key features. The confusion matrix for each PSD matrix and the number of main features retrieved after applying the feature selection stage are presented.

| Matrix | Label | Estimated Label | | N. Features | *A* (%) | *P* (%) | *R* (%) | *S* (%) | *BER* |
|---|---|---|---|---|---|---|---|---|---|
| | | VT | LP | | | | | | |
| B | VT | 34 | 109 | 5 | 89 | 24 | 85 | 89 | 0.13 |
| | LP | 6 | 874 | | | | | | |
| M | VT | 38 | 103 | 5 | 90 | 27 | 95 | 90 | 0.07 |
| | LP | 2 | 880 | | | | | | |
| W | VT | 40 | 99 | 3 | 90 | 29 | 100 | 90 | 0.05 |
| | LP | 0 | 884 | | | | | | |
| Y | VT | 38 | 192 | 3 | 81 | 17 | 95 | 81 | 0.12 |
| | LP | 2 | 791 | | | | | | |
| Voting | VT | 38 | 42 | | 96 | 48 | 95 | 96 | 0.05 |
| | LP | 2 | 941 | | | | | | |

**The Yule-Walker method (Y) achieved the worst average performance, and therefore it was not considered by the voting strategy.

**Table 3**
Experimental seismic event classification performance obtained using the DT classifier with seven key features selected considering at least 3 dB of difference between the amplitudes of similar features. The confusion matrix for each PSD matrix and the number of main features retrieved after applying the feature selection stage are presented.

| Matrix | Label | Estimated Label | | N. Features | *A* (%) | *P* (%) | *R* (%) | *S* (%) | *BER* |
|---|---|---|---|---|---|---|---|---|---|
| | | VT | LP | | | | | | |
| B | VT | 35 | 113 | 5 | 89 | 24 | 88 | 89 | 0.12 |
| | LP | 5 | 870 | | | | | | |
| M | VT | 38 | 127 | 5 | 87 | 23 | 95 | 87 | 0.09 |
| | LP | 2 | 856 | | | | | | |
| W | VT | 39 | 78 | 4 | 92 | 33 | 98 | 92 | 0.05 |
| | LP | 1 | 905 | | | | | | |
| Y | VT | 37 | 142 | 3 | 86 | 21 | 93 | 86 | 0.11 |
| | LP | 3 | 841 | | | | | | |
| Voting | VT | 37 | 43 | | 96 | 46 | 93 | 96 | 0.06 |
| | LP | 3 | 940 | | | | | | |

**The Yule-Walker method (Y) achieved the worst average performance, and therefore it was not considered by the voting strategy.

**Table 4**
Experimental seismic event classification performance obtained using the SVM classifier. The confusion matrix for each PSD matrix and the number of main features retrieved after applying the feature selection stage are presented.

| Matrix | Label | Estimated Label | | N. Features | *A* (%) | *P* (%) | *R* (%) | *S* (%) | *BER* |
|---|---|---|---|---|---|---|---|---|---|
| | | VT | LP | | | | | | |
| B | VT | 36 | 62 | 25 | 94 | 37 | 90 | 94 | 0.08 |
| | LP | 4 | 921 | | | | | | |
| M | VT | 38 | 71 | 25 | 93 | 35 | 95 | 93 | 0.06 |
| | LP | 2 | 912 | | | | | | |
| W | VT | 38 | 77 | 25 | 93 | 37 | 95 | 93 | 0.06 |
| | LP | 2 | 906 | | | | | | |
| Y | VT | 35 | 70 | 25 | 94 | 39 | 95 | 94 | 0.06 |
| | LP | 5 | 913 | | | | | | |
| Voting | VT | 38 | 36 | | 96 | 51 | 95 | 96 | 0.05 |
| | LP | 2 | 947 | | | | | | |

**The Burg method (B) achieved the worst average performance, and therefore it was not considered by the voting strategy.

The number of features was further reduced by applying our strategy of considering at least 3 dB of difference between the amplitudes within a given band and the CI for each event in the PSD. Thus, the key features were reduced to the amplitudes of only seven frequencies (8.2 Hz, 8.4 Hz, 9.53 Hz, 11.1 Hz, 13.4 Hz, 15.6 Hz, and 41.8 Hz). Besides, we then used these seven key features from each PSD estimator. The classifiers determined then five critical features for the Burg and Multitaper methods (8.2 Hz, 11.1 Hz, 13.4 Hz, 15.6 Hz, and 41.8 Hz), and three key features for the Welch and Yule-Walker methods (8.4 Hz, 9.5 Hz, and 13.4 Hz). Table 3 shows the results obtained with the testing set, for this strategy. The Welch estimator achieved the best result.

### 3.2. Results using SVM classifier

SVM is often sensitive to class imbalance, and for this reason, we first determined the training percentage for our database, which maximized the system performance, this value was set to 60%. We then set the free parameter $\nu$ to 0.05 by making an iterative process in order to maximize the system performance (Lara-Cueva et al., 2016b).

The RFE algorithm as wrapper method in conjunction with SVM, recognized the relevance of 25 key features for each PSD matrix, which their weights represent at least the 10% of the relevance according with RFE and the performance parameters. In this sense, these frequencies corresponding to 2.9 Hz; $f_b \in (4.5, 4.9)$ Hz; $f_b \in (5.6, 6.0)$ Hz; $f_b \in (15.4, 18.5)$ Hz; and $f_b \in (40.0, 42.6)$ Hz for the Yule-Walker method.

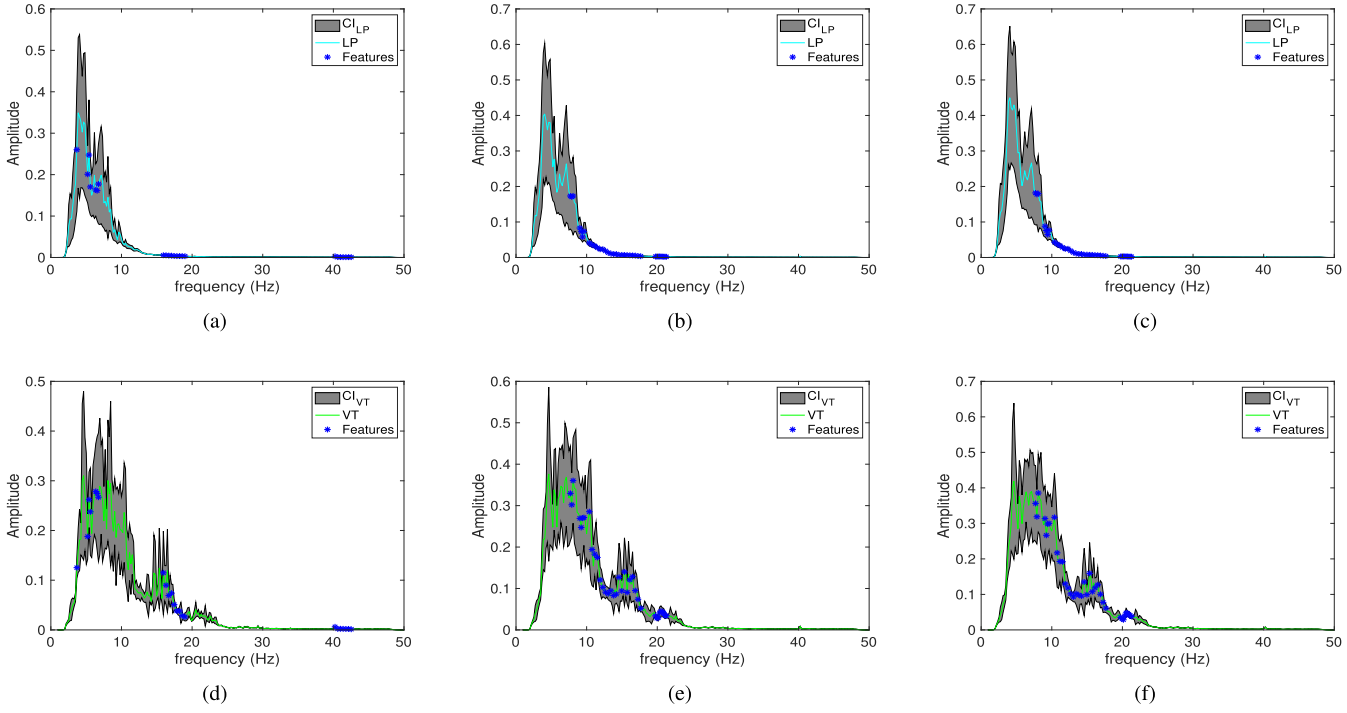


**Fig. 4.** Main frequency features selected by the SVM classifier: top row: mean of LP events (cyan lines) for (a) Yule-Walker, (b) Multitaper, and (c) Welch methods; bottom row: mean of VT events (green lines) for (d) Yule-Walker, (e) Multitaper, and (f) Welch methods, respectively. Retrieved features in each case are represented by the blue points which are within the 97.5% of CI (grey zones).

**Table 5**
Experimental classification performance obtained using the SVM classifier after selecting 53 key features. The confusion matrix for each PSD matrix is presented.

| Matrix | Label | Estimated Label | | N. Features | *A* (%) | *P* (%) | *R* (%) | *S* (%) | *BER* |
|---|---|---|---|---|---|---|---|---|---|
| | | VT | LP | | | | | | |
| B | VT | 36 | 47 | 53 | 95 | 44 | 90 | 95 | 0.07 |
| | LP | 4 | 936 | | | | | | |
| M | VT | 34 | 50 | 53 | 95 | 40 | 85 | 95 | 0.10 |
| | LP | 6 | 933 | | | | | | |
| W | VT | 36 | 38 | 53 | 96 | 49 | 90 | 96 | 0.07 |
| | LP | 4 | 945 | | | | | | |
| Y | VT | 36 | 56 | 53 | 94 | 39 | 90 | 94 | 0.08 |
| | LP | 4 | 927 | | | | | | |
| Voting | VT | 34 | 24 | | 97 | 59 | 85 | 98 | 0.08 |
| | LP | 6 | 959 | | | | | | |

**The Burg method (B) achieved the worst average performance, and therefore it was not considered by the voting strategy.

Meanwhile, for the Welch and Multitaper methods the frequencies corresponding to $f_b \in (7.0, 7.4)$ Hz; $f_b \in (8.4, 8.9)$ Hz; $f_b \in (9.7, 17.1)$ Hz; $f_b \in (19.3, 20.8)$ Hz, as shown in Fig. 4.

The results are summarized in Table 4. The best results were achieved with the Welch method, while the Burg method achieved the worst results.

From all the features identified in the process described above, 53 key features were selected, considering only those features that did not appear repeated in the different PSD estimators. These critical frequencies were then manually grouped into frequencies such as 2.9 Hz; $f_b \in (4.5, 4.9)$ Hz; $f_b \in (5.6, 6.0)$ Hz; $f_b \in (7.0, 7.4)$ Hz; $f_b \in (8.4, 8.9)$ Hz; $f_b \in (9.7, 18.5)$ Hz; $f_b \in (19.3, 20.8)$ Hz; and $f_b \in (40.0, 42.6)$. Table 5 shows the classification results obtained with the testing set. The Welch estimator achieved the best result, while the Burg method achieved the worst result.

The number of features was further reduced by applying our strategy of considering at least 3 dB amplitude differences within a given band and the CI for each event in their corresponding PSD. Thus, only 39 frequencies were finally selected corresponding to $f_b \in (9.7, 18.5)$ Hz; $f_b \in (19.3, 20.8)$ Hz; and $f_b \in (40.0, 42.6)$. Table 6 shows the results obtained with the testing set. The Burg estimator achieved the best overall result, while the Yule-Walker method achieved the worst result.

As shown in previous Tables, but mainly in Tables 1 and 4, we observed the performance was improved by using the voting strategy. Regarding the frequency features, from Figs. 3 and 4, it seems that for the DT algorithm features are concentrated within the CI and close to traditional frequency bands for LP and VT events, as previously described in section I. However, for the SVM, most of the features used by the algorithm to classify between LP and VT events were about the 10 Hz, and there were key frequency features present in the $f_b \in (40.0, 42.6)$ Hz band.

**Table 6**
Experimental seismic event classification performance obtained using the SVM classifier with 39 key features selected considering at least 3 dB of difference between the amplitudes of similar features. The confusion matrix for each PSD matrix is presented.

| Matrix | Label | Estimated Label | | N. Features | *A* (%) | *P* (%) | *R* (%) | *S* (%) | *BER* |
|---|---|---|---|---|---|---|---|---|---|
| | | VT | LP | | | | | | |
| B | VT | 37 | 48 | 39 | 95 | 44 | 93 | 95 | 0.06 |
| | LP | 3 | 935 | | | | | | |
| M | VT | 37 | 70 | 39 | 93 | 35 | 93 | 93 | 0.07 |
| | LP | 3 | 913 | | | | | | |
| W | VT | 37 | 80 | 39 | 92 | 32 | 93 | 92 | 0.07 |
| | LP | 3 | 903 | | | | | | |
| Y | VT | 33 | 67 | 39 | 93 | 33 | 83 | 93 | 0.12 |
| | LP | 7 | 916 | | | | | | |
| Voting | VT | 36 | 22 | | 98 | 62 | 90 | 98 | 0.06 |
| | LP | 4 | 961 | | | | | | |

**The Yule-Walker method (Y) achieved the worst average performance, and therefore it was not considered by the voting strategy.

**Table 7**
Obtained experimental performance results for seismic event classification using the DT and SVM classifiers without the frequencies above the 40 Hz band. The confusion matrix for each classifier using the voting strategy is presented.

| Classifier | Label | Estimated Label | | *A* (%) | *P* (%) | *R* (%) | *S* (%) | *BER* |
|---|---|---|---|---|---|---|---|---|
| | | VT | LP | | | | | |
| DT | VT | 35 | 27 | 97 | 56 | 88 | 97 | 0.08 |
| | LP | 5 | 956 | | | | | |
| SVM | VT | 37 | 30 | 97 | 55 | 93 | 97 | 0.05 |
| | LP | 3 | 953 | | | | | |

**The Yule-Walker method (Y) achieved the worst average performance, and therefore it was not considered by the voting strategy.

In (Lara-Cueva et al., 2016a), each microearthquake was preprocessed using a 128th order band-pass FIR with bandwidth $f_b \in (1, 15)$ Hz, and even so we found two frequencies peaks around 43 Hz and 46 Hz, which could also be used for detecting microearthquakes. Therefore, in this work, we decided to use a 128th order high-pass FIR with a cutoff frequency of 1 Hz, to let pass the quasi entire spectral components. The DT classifier found a total of 10 key features. There are seven features related to frequencies peaks at 8.2 Hz, 8.4 Hz, 9.53 Hz, 11.1 Hz, 13.4 Hz, 15.6 Hz, and 41.8 Hz. While the SVM classifier also found several frequency bands, in this context, there are features in the $f_b \in (15.4, 18.5)$ Hz, and even at the higher $f_b \in (40.0, 42.6)$ Hz. The peak at 41.8 Hz has appeared in both classifiers, which is in agreement with our previous research work that found a frequency peak close to 43 Hz. Thus, we can observe that there is a frequency band above 40 Hz, which seems like a critical feature for the detection and classification stages.

This statement was corroborated by retraining the classifiers for the best scenarios, which means the DT and SVM classifiers with the voting system but without including the frequency features between the 40.0 and 42.6 Hz band. Table 7 shows the results obtained for this test. For DT, a worse BER of 0.08 (0.06 in Table 1) and a *R* of 88% (90% in Table 1) were achieved, which means a loss of sensitivity. Meanwhile, for SVM, a better BER of 0.05 (0.06 in Table 6) was achieved due to gain in sensitivity, but a worse *P* of 55% (62% in Table 4) was obtained, which means a loss of precision. This test confirmed that these frequencies around the 40 Hz band could be considered as critical features for the classification stage.

## 4. Conclusions and future work

In this paper, a study about different parametric and non-parametric methods for computing the PSD in the context of microearthquake classification was presented. Bootstrapping was used to estimate a CI for the main frequencies of microearthquakes, and an amplitude threshold difference of at least 3 dB was used to ensure that possible frequency features that characterize each type of event do not overlap between classes. Embedded and wrapper methods were then applied to select the main frequencies that maximize the *BER* of the DT and SVM classifiers.

In the classification stage, the DT algorithm reached its best performance metrics when the Welch PSD method was used by considering three key frequency features, an accuracy near to 92% and a *BER* of 0.05 were obtained, these results were improved by considering the voting system achieving an accuracy of 97% and a *BER* of 0.06. Finally, features were further reduced to seven key features by applying the 3 dB difference strategy between frequency peaks of different event types, obtaining an accuracy of 92% and a *BER* of 0.05. Meanwhile, the

SVM algorithm achieved its best performance metrics with the Multitaper PSD method using 25 key frequency features with an accuracy of 93% and a *BER* of 0.06. In this case, the voting system improved the results with an accuracy of 96% and a *BER* of 0.05. The frequency features used belong to both low and high frequencies bands. Finally, by applying the 3 dB difference strategy between similar frequency peaks of different event types, the features were further reduced to 39 key features, obtaining an accuracy of 98% and a *BER* of 0.06. Although we are not interested in evaluating the computational cost at this stage, we have estimated the time processing required by considering the best results. The DT and linear SVM take about 2 ms on average for classifying an event. However, the computational cost derived in time processing refers not just to the classification problem; it also has to consider the detection stage, which means recognition (detection + classification).

Regarding the frequency features, several key frequency bands were identified by the developed strategy with the DT and SVM classifiers. Most of these features are concentrated within the CI and close to traditional frequency for LP and VT events. However, key frequency features were also present in the $f_b \in (40.0, 42.6)$ Hz band. Thus, we can conclude that a frequency band is above 40 Hz, which seems like a critical feature for the detection and classification stages.

According to the results obtained, we recommend using non-parametric methods, instead of parametric methods of spectral estimation. Overall, the Yule-Walker and Burg methods were dismissed for both classifiers, since they achieved the worst performance results. This could be due to that parametric methods are most commonly used to analyze white noise-driven systems, which is present in the entire spectrum, but in this case for this kind of volcano monitoring system, the noise is related to colored noise.

As future work, we are interested in reproducing this methodology using Hilbert and Wavelet Transforms to overcome the resolution frequency-time problem presented at Fourier Transform and for the other type of microearthquakes. However, currently, the dataset does not contain enough samples of such events; in this sense, we are designing a strategy for obtaining a synthetic dataset from microearthquakes at Cotopaxi Volcano using the *MicSigV1* dataset. We are also interested in determining if the noise surrounding the events are a necessary part of the event, or is independent of the events, with this information, would be possible recommend the use of possible filters, to clean the events in the preprocessing stage. Finally, the frequency features obtained in this work should also be combined with other features such as temporal, scale, and spectrogram image-based features to increase the performance of MLCs. We plan to perform a comparative study related to this among the next steps.

## 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.

## Acknowledgment

This work was supported in part by the Universidad de las Fuerzas Armadas ESPE under Grants 2013-PIT-014 and 2016-EXT-038, and in part by the Universidad San Francisco de Quito (Poli-Grants Program) under Grant 10100, Grant 12494 and Grant 16916. The seismic data used in this study was provided by Instituto Geofísico, EPN.

## References

Alessio, S.M., 2015. Digital Signal Processing and Spectral Analysis for Scientists: Concepts and Applications. Springer.

Álvarez, I., García, L., Cortés, G., Benítez, C., De la Torre, A., 2012. Discriminative feature selection for automatic classification of volcano-seismic signals. IEEE Lett. Geosci. Remote Sensing Lett. 9 (2), 151–155.

Babadi, B., Brown, E.N., 2014. A review of multitaper spectral analysis. IEEE Trans. Biomed. Eng. 61 (5), 1555–1564.

Bishop, C.M., et al., 2006. Pattern Recognition and Machine Learning. vol. 1. Springer.

Cárdenas-Peña, D., Orozco-Alzate, M., Castellanos-Domínguez, G., 2013. Selection of time-variant features for earthquake classification at the Nevado del Ruiz Volcano. Comput. Geosci. 51, 293–304.

Carniel, R., 2014. Characterization of volcanic regimes and identification of significant transitions using geophysical data: a review. Bull. Volcanol. 76 (8), 848.

Castanié, F., 2013. Spectral Analysis: Parametric and Non-parametric Digital Methods. John Wiley & Sons.

Chouet, B.A., 1996. Long-period volcano seismicity: its source and use in eruption forecasting. Nature 380 (6572), 309–316.

Chouet, B.A., Page, R.A., Stephens, C.D., Lahr, J.C., Power, J.A., 1994. Precursory swarms of long-period events at redoubt volcano (1989–1990), Alaska: their origin and use as a forecasting tool. J. Volcanol. Geotherm. Res. 62 (1–4), 95–135.

Curilem, M., Vergara, J., San Martín, C., Fuentealba, G., Cardona, C., Huenupan, F., Chacón, M., Khan, M.S., Hussein, W., Yoma, N.B., 2014. Pattern recognition applied to seismic signals of the Llaima Volcano (Chile): an analysis of the events' features. J. Volcanol. Geotherm. Res. 282, 134–147.

Duque, A., González, K., Pérez, N., Bentez, D., Grijalva, F., Lara-Cueva, R., Ruiz, M., 2020. Exploring the unsupervised classification of seismic events of cotopaxi volcano. J. Volcanol. Geotherm. Res. 107009.

Efron, B., 1981. Nonparametric standard errors and confidence intervals. Can. J. Stat. 9 (2), 139–158.

Efron, B., Tibshirani, R.J., 1994. An Introduction to the Bootstrap. CRC press.

Esposito, F., Malerba, D., Semeraro, G., Kay, J., 1997. A comparative analysis of methods for pruning decision trees. IEEE Trans. Pattern Anal. Mach. Intell. 19 (5), 476–491.

Friedlander, B., Porat, B., 1984. The modified yule-walker method of arma spectral estimation. IEEE Transactions on Aerospace and Electronic Systems. 2, pp. 158–173.

Ibáñez, J., Carmona, E., 2000. Sismicidad volcánica. *Curso Internacional de Volcanologa y Geofsica Volcánica, Serie Casa de los Volcanes*, no. 7, pp. 269–282.

Ifeachor, E.C., Jervis, B.W., 2002. Digital Signal Processing: A Practical Approach. 2nd ed. Pearson Education.

Kenneth, B., 2001. The Seismic Wavefield, Volume I: Introduction and Theoretical Development. Cambridge Univ. Press, USA.

Kohavi, R., Sommerfield, D., 1995. Feature subset selection using the wrapper method: Overfitting and dynamic search space topology. KDD, pp. 192–197.

Laasri, E.H.A., Akhouayri, E.-S., Agliz, D., Zonta, D., Atmani, A., 2015. A fuzzy expert system for automatic seismic signal classification. Expert Syst. Appl. 42 (3), 1013–1027.

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. J. Volcanol. Geotherm. Res. 62 (1), 137–151.

Lara-Cueva, R., Bernal, P., Saltos, G., Benítez, D., Rojo-Álvarez, J.L., 2015. Time and frequency feature selection for seismic events from Cotopaxi Volcano. Asia-Pacific Conference on Computer Aided System Engineering (APCASE), 2014, pp. 1–6.

Lara-Cueva, R., Benítez, D., Carrera, E., Ruiz, M., Rojo-Álvarez, J., 2016a. Feature selection of seismic waveforms for long period event detection at cotopaxi volcano. J. Volcanol. Geotherm. Res. 316, 34–49.

Lara-Cueva, R.A., Benítez, D.S., Carrera, E.V., Ruiz, M., Rojo-Álvarez, J.L., 2016b. Automatic recognition of long period events from volcano tectonic earthquakes at cotopaxi volcano. IEEE Trans. Geosci. Remote Sens. 54 (9), 5247–5257.

Luo, F.-L., 2020. Machine Learning for Future Wireless Communications. John Wiley & Sons.

Malfante, M., Mura, M.D., Metaxian, J., Mars, J.I., Macedo, O., Inza, A., 2018. Machine learning for volcano-seismic signals: challenges and perspectives. IEEE Signal Process. Mag. 35 (2), 20–30.

McNutt, S.R., 2005. Volcanic seismology. Annu. Rev. Earth Planet. Sci. 32a, 461–491.

Pérez, N., Venegas, P., Benítez, D., Lara-Cueva, R., Ruiz, M., 2020a. A new volcanic seismic signal descriptor and its application to a data set from the Cotopaxi volcano. IEEE Trans. Geosci. Remote Sens. 1–11.

Pérez, N., Benítez, D., Grijalva, F., Lara-Cueva, R., Ruiz, M., 2020b. ESeismic: towards an ecuadorian volcano seismic repository. J. Volcanol. Geotherm. Res. 396 106855.

Ruano, A., Madureira, G., Barros, O., Khosravani, H., Ruano, M., Ferreira, P., 2014. Seismic detection using support vector machines. J. Neurocomputing 135 (0), 273–283.

Ruiz, M., Guillier, B., Chatelain, J.-L., Yepes, H., Hall, M., Ramon, P., 1998. Possible causes for the seismic activity observed in cotopaxi volcano, ecuador. Geophys. Res. Lett. 25 (13), 2305–2308.

Schölkopf, B., Smola, A., 2002. *Learning with Kernels*, 1st Ed. MIT Press, Cambridge, MA.

Soto, R., Huenupan, F., Meza, P., Curilem, M., Franco, L., 2018. Spectro-temporal features applied to the automatic classification of volcanic seismic events. J. Volcanol. Geotherm. Res. 358, 194–206.

Wassermann, J., 2012. Volcano seismology. New Manual of Seismological Observatory Practice 2 (NMSOP-2). Deutsches GeoForschungsZentrum GFZ, pp. 1–77.