Journal of Volcanology and Geothermal Research 396 (2020) 106855

Contents lists available at ScienceDirect

Journal of Volcanology and Geothermal Research

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

ESeismic: Towards an Ecuadorian volcano seismic repository

Noel Pérez [a], Diego Benítez [a], Felipe Grijalva [b], Román Lara-Cueva [c], Mario Ruiz [d,*], Jorge Aguilar [d]

[a] Colegio de Ciencias e Ingenierías "El Politécnico", Universidad San Francisco de Quito USFQ, Quito 170157, Ecuador
[b] Departamento de Electrónica, Telecomunicaciones y Redes de Información, Escuela Politécnica Nacional, Quito 17012759, Ecuador
[c] 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
[d] Instituto Geofísico, Escuela Politécnica Nacional, Quito 17012759, Ecuador

ARTICLE INFO

Article history:
Received 19 November 2019
Received in revised form 9 March 2020
Accepted 16 March 2020
Available online 23 March 2020

Keywords:
Seismic signal dataset
Seismic benchmark dataset
Volcano seismic repository

ABSTRACT

In this work, we present the development, description, and performance evaluation of two volcano seismological datasets: one containing raw seismic signals (*MicSigV1*) and another containing a set of 84 distinctive attributes or features for bench-marking (*SeisBenchV1*), which was extracted from the *MicSigV1* dataset, recorded at Cotopaxi volcano in Ecuador. These datasets represent the first public Ecuadorian volcano seismicity repository that will update, publish, and support the release of data to the scientific community on a continual basis. Each dataset contains samples of mostly long-period and volcano-tectonic event classes. Other types of events, such as regional, hybrid, and icequakes, were also included despite their minor presence in the datasets. These datasets will serve to provide an important resource for research in signal processing, feature calculation, feature selection, feature space reduction, seismic events detection, and classification areas. The datasets can be downloaded through a copyright agreement that supports the correct use of the released datasets.

1. Introduction

According to worldwide historical records, volcanic eruptions have been responsible for approximately 280 thousand deaths since the year 1500 (Tilling, 1996). Between 1986 and 2019 there were 7670 deaths recorded from direct and indirect volcanic activity worldwide, including the recent Fuego and Anak Krakatau eruptions (Brown et al., 2017). Various highly populated cities are located close enough to active volcanoes to present a significant potential volcanic risk, including: Tokyo (Japan), near Mt. Fuji (recent signs of awakening), Quito (Ecuador) which has suburbs at risk areas from Cotopaxi (last active in 2015), Guagua Pichincha (last active in 2000), and Reventador (last active in 2020) volcanoes; Mexico City (Mexico) near to Popocatepetl volcano (new eruptive phase since 1995); Yogyakarta (Indonesia), near to permanently active Merapi volcano; Naples (Italy) close to Vesuvius (last active in 1944); Seattle (USA) close to Mount Rainier (presently quiet), and Manila (Philippines) close to Taal volcano (last eruption in 2020) (Schmincke, 2004). Currently, about 200 million people globally reside within a 30 km radius and 47 million people within a 5 km radius of approximately 1300 Holocene volcanoes (Chester et al., 2000; Siebert et al., 2011; Phillipson et al., 2013). For this reason, the forecasting of volcanic eruption timing, potential size and style along with identifying eruptive hazards and their extents becomes an indispensable factor in reducing the associated effects of this type of natural phenomena.

The vast majority of volcanic observatories worldwide use seismic methods because it is the most effective tool for monitoring and forecasting eruptions (Chouet, 1996; Schmincke, 2004; McNutt, 1996; Zobin, 2017; Roman and Cashman, 2018; White and McCausland, 2019). One of the fundamental tasks in any volcano observatory is the detection and correct classification of volcanic seismic events (Havskov and Ottemoller, 2010; Bean et al., 2014; Cusano et al., 2015).

Different approaches have been proposed in the literature to advance automatic detection and classification of seismic events. The detection based approaches mainly process the signal in time, frequency and scale domains (Gutiérrez et al., 2011; Rodríguez Cesén and Lara-Cueva, 2018; Lara-Cueva et al., 2016a, 2016d). On the other hand, classification-based approaches calculate some features in time, frequency or scale domains related to the event's signal before employing machine learning classifiers (MLCs) to identify seismic events (Jaramillo et al., 2014; Cortés et al., 2015). Other developed methods like those presented in Lara-Cueva et al. (2016c) have opted to combine detection and classification as a single model to analyze seismic events. In this sense, different seismic events such as LP, VT, hybrid (HB), icequakes (ICE) and even regional events (REG), which are earthquakes with origin at tectonic faults apart from the volcanic edifice (Langer et al., 2003), have been correctly identified.

* Corresponding author.
E-mail address: mruiz@igepn.edu.ec (M. Ruiz).

https://doi.org/10.1016/j.jvolgeores.2020.106855
0377-0273/

The use of these techniques or any combination of them enables the research community to obtain relevant information from different seismic event types that can be identified throughout each volcano study. Usually, this information is stored in digital archives (Almendros et al., 2018b) or in more sophisticated projects that host the data using database management systems for further analysis (Newhall et al., 2017). However, in any case, the data mostly remain under privacy copyright, limiting its use for research purposes around the world.

The aim of this work is to propose the development of the first Ecuadorian volcanic seismic repository (ESeismic), containing seismic signals from Cotopaxi volcano annotated by experts. The ESeismic repository will be publicly accessible to overcome existing data accessibility restrictions and will serve research purposes from both, signal processing and machine learning points of view.

The remainder of the paper is organized as follows: Section 2 "Requirements for a Volcano Database" describes the most important aspects to be considered for creating a volcano database such as database composition and event collection process, relevant information associated with storage records, database organization, maintenance, as well as its distribution; Section 3 "Available volcano seismic signal repositories" overviews previous existing databases; Section 4 provides a detailed description of the developed ESeismic repository from the data acquisition to the released datasets; Section 5 "Performance evaluation" covers several approaches that have been developed using the released datasets; Section 6 "Discussion" analyzes the strengths as well as the weaknesses of the repository. Finally, in the "Conclusions and future work" section, we outline the main achievements of this work and our plans for future work.

2. Requirements for a volcano seismic database

According to the Incorporated Research Institutions for Seismology - IRIS (2019), the National Geophysical Data Center/World Data Service (NGDC/WDS): Significant Earthquake Database (2019), and the Instituto Geofísico de la Escuela Politécnica Nacional – IGEPN (2019), several specific aspects should be considered before creating a volcano seismology database, such as: the database composition and event collection process, relevant associated information (e.g. timestamps, event type), storage format, distribution and maintenance. We will briefly described these requirements below.

2.1. Composition and event collection process

The database should include original signals of all types of seismic events recorded from a given volcano as well as an explanation of how they were collected (e.g. seismometer characteristics). However, in some cases it is necessary to filter them before their inclusion as samples in the database. Filtering the signal as an initial step minimizes the inclusion of seismic noise, which could overlap the desired events. Some of the most popular recorded volcano seismic events are LP, VT, hybrid, tremor and very long-period, explosion events and lahar signals. Others less perturbing non-volcano seismic events are icequakes, lightning or tectonic earthquakes.

2.2. Relevant associated information

The main relevant associated information needed to identify an event consists in the starting and endpoints (timestamps) of the event inside the original signal and the assigned output class label.

Other relevant information to be stored on each volcano seismic database are the network to which the volcano belongs, location, name of the volcano, name of the station, channel or component, start time of recording, number of samples, sampling frequency and size of recording.

2.3. Storage format

There is not a standard procedure referring on how to save the volcanic signal data to be persistent over time. Usually, seismic data are stored in the standard exchange of earthquake data format (SEED) (Ringler and Evans, 2015). Although SEED is considered a completed format for exchanging seismic data since it contains data and metadata in a single file, its use is tricky to handle in machine learning tasks.

2.4. Distribution and maintenance

The database should be available and publicly accessed across the internet under a disclaimer agreement for user registration. The continuous maintenance (e.g. to correct possible labeling errors, add or delete instances) and user support is also mandatory.

3. Available volcano seismic signal databases

Presently, there are both public and restricted volcano seismic databases. However the publically-available ones (Almendros et al., 2018a, 2018b; Koulakov et al., 2019; RESIF, French National Sismological Network, 2014; Newhall et al., 2017) only provide raw seismograms in the time-series data format without any information regarding the location (i.e., starting and endpoints of events) or type of events contained by the signals.

In this regard, it is worth mentioning IRIS (Incorporated Research Institutions for Seismology[1]) that distributes data to support the seismological research community, including time series data, event data and metadata about earthquakes and other seismic events beyond volcano seismic signals. Moreover, institutions such as the United States Geological Survey (USGS) and some initiatives funded by the National Science Foundation provide open access data through IRIS (2018). Other similar earthquake databases are the ANSS Comprehensive Catalogue (ComCAT)[2] and the International Seismological Centre (ISC) catalogue. Note however that IRIS, ANSS ComCAT and ISC catalogue do not offer catalogs of labeled volcano seismic events along with their corresponding raw signals, as we do.

On the other hand, restricted datasets are available only for particular research groups. A brief description of some restricted databases is presented in Table 1; it should be noted that these databases do contain various event types labeled by experts in contrast to the aforementioned public databases. However, since the process of labeling seismic events is manually performed and time-consuming, most of these databases have a small number of records (less than a thousand) and tend to be imbalanced for certain types of events.

4. ESeismic repository description

The repository proposed in this work corresponds to two different datasets: one containing raw seismic signals (*MicSigV1*) and another containing a set of 84 features for benchmarking (*SeisBenchV1*), obtained from the same seismological event records and available at https://www.igepn.edu.ec/eseismic_web_site/index.php. Both datasets contain information about several seismic events recorded at Cotopaxi volcano located in the Andes of Ecuador (latitude 0°41′05″ S and longitude 78°25′54.8″ W), about 50 km south of Quito, Ecuador's capital city. Cotopaxi Volcano is one of the most active volcanoes in the Ecuadorean Andes. Since 1533, this volcano has erupted at least 36 times including major explosive eruptions in 1534, 1742–1744, 1766–1768 and 1877 (Ruiz et al., 1998). The most recent eruption occurred between August and November 2015. Cotopaxi volcano is an active snow-capped volcano continuously monitored by the IGEPN, the institution responsible for monitoring volcanic activity in Ecuador.

[1] https://www.iris.edu/hq/

[2] https://earthquake.usgs.gov/data/comcat/

4.1. Data acquisition and collection

Monitoring the seismic activity of Cotopaxi volcano is based mainly on signals recorded by a high resolution network composed by six broadband (BB) and six short period (SP) seismic stations, permanently deployed on the volcano's flanks, equipped with three-axial seismometers with a sensitivity of 1600 V/m·s-1 and a frequency response range of 0.03–50 Hz for BB stations and 1–50 Hz for SP stations (Ortiz, 2013). These stations are illustrated in Fig. 1. In this work, we specifically collected seismograms from VC1 (SP station) and BREF (BB station) that operate at a sampling rate of 100 Hz and 50 Hz respectively, using a 24-bit analog to digital converter.

Seismic events arrive at different seismological stations with different characteristics, since the internal paths within the volcano are not homogeneous. The acquisition system used the STA/LTA algorithm (Vaezi and Van der Baan, 2015) for detecting seismic events inside the recorded seismograms. Every detected event was separated into a file, which contains a signal extending 10 s before and after the detected event. The purpose of this is to make it easier for experts to estimate the noise level around the possible event to reduce false-positive detected events. Subsequently, each file was used for correct event identification and labeling throughout the signal analysis based on the waveform, spectrogram, and location from all the available stations which recorded the event under analysis (event corroboration), a process made visually by IGEPN specialists.

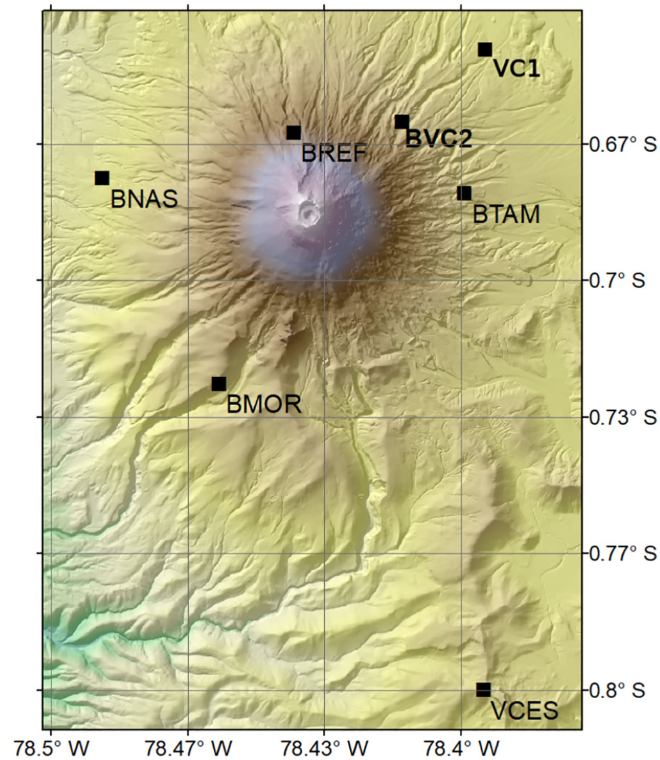

Fig. 1. Some of the stations from the seismological network deployed at Cotopaxi volcano. Data for this study have been recorded at the VC1 and BREF stations.
Image provided by IGEPN.

Table 1
Restricted volcano seismic databases. Legend: TRE (Tremors), LP (Long Period), VLP (Very Long Period), VT (volcano-tectonic), HB (Hybrid), TOR (Tornillo), REG (Regional), ROC (Rockfalls), LP + ROC (LP shortly preceding a rockfall), NS (Noise), LAH (Lahars), EXP (Explosion), LAN (Landslides),MP (Multiphase), FIS (explosions due to bomb fishing), QUA (Quarry blasts), THU (Thunder), ICE (Icequakes), OTH (Other signals).

Total events	Distribution	Volcano
68,008	62,029 LP, 5245 TRE, 512 VT, 48 EXP, 141 HB, 33 TOR	Ubinas, Peru (Malfante et al., 2018)
41,536	35,240 LP, 445 HB, 5461 TRE, 247 VT, 143 EXP	Ubinas, Peru (Traversa et al., 2011)
436	327 LP, 109 VT	Cotopaxi, Ecuador (Lara-Cueva et al., 2016d)
147	147 VLP	Stromboli, Italy (Esposito et al., 2008)
336	44 VT, 78 REG, 30 LP, 62 HB, 84 ROC, 38 LPE + ROC	Soufriere Hills, Montserrat (Langer et al., 2003)
4391	895 VT, 226 REG, 504LP, 589 HB, 2094 ROC, 83 LP + ROC	Soufriere Hills, Montserrat (Langer et al., 2006)
162	27 VT, 27 MP, 27 HB, 27 ROC, 54 NS	Mt. Merapi, Indonesia (Alasonati et al., 2006)
971	75 VT, 765 LP, 54 HB, 77 TRE	Deception Island, Antarctica (Benítez et al., 2006)
475	VT, LP, TRE, ROC, LAH, and EXP.	Colima, Mexico (Alvarez et al., 2011)
1860	311 EXP, 672 NS, 877 TRE	Stromboli and Etna, Italy (Ibáñez et al., 2009)
30	15 VT and 15 FIS	The Phlegraean Fields, Italy (Del Pezzo et al., 2003)
1159	430 EXP, 267 LAN, and 462 TRE	Stromboli, Italy (Apolloni et al., 2009)
331	169 VT, 58 FIS, 81 QUA, 23 THU	Mt. Vesuvius, Italy (Scarpetta et al., 2005)
924	74 REG, 310 NS, 252 VT, 190 LP, 98 TRE	Krakatau, Indonesia (Ibs-von Seht, 2008)
1033	290 LP, 515 TRE, 228 OTH	Villarrica, Chile (Curilem et al., 2009)
505	193 HB, 197 ROC, 115 VT.	Soufrière Hills, Montserrat (Hammer et al., 2012)
71,930	28,013 VT, 151 REG, 42086 ROC, 1680 OTH	Piton de la Fournaise, La Réunion (Maggi et al., 2017)
1845	483 VT, 580 LP, 782 ICE	Nevado del Ruiz, Colombia (Orozco et al., 2006)
400	100 VT, 100 LP, 100 TRE, 100 HB	Galeras, Colombia (Bicego et al., 2012)
1043	98 VT, 120 LP, 244 TRE, 77 ROC, 30 LAH, 45 EXP, 45 REG, NS, 41 HB	Volcán de Fuego, Mexico, and Deception Island, Antarctica (Cortés et al., 2015)

The signals were acquired during 2012, January 2013, January 2014, January 2018, and January to March 2019. Throughout these periods, Cotopaxi showed low levels of activity compared with those exhibited during the 2015 eruptive period (50.2 LP per day on average during the 2015 eruptive period). In 2012, the LP daily average was 6.4, in January 2013 was 6.5, in January 2014 was 13.9, in January 2018 was 20.6, and in 2019 period was 17.9, respectively. An eruptive period was not included in this version of the data set, to focus the analysis only on Cotopaxi's background seismicity. During the experimentation period, a total of 1187 seismic events from BREF and VC1 stations were collected, most of them belonging to the LP class (1044 samples), the VT class (101 samples) and a small number of hybrid (8 samples), regional (27 samples) and icequake (7 samples) classes. It is worth noting that the events collected by BREF are not exactly the same events as those collected by VC1, since one is a broad band and the other is a short period station, respectively. A percentage-based distribution of collected seismic events is shown in Fig. 2. Although the primary station used for manual classification was VC1, in cases when the event has very low frequencies, or when the VC1 record was noisy, analysts used BREF as the reference station.

4.2. Data model and storage format

The recorded signals were manipulated using the Matlab software (MATLAB, 2019 as the main tool for handling data. According to the volcano seismic database requirements, each recorded seismic signal is stored in a Matlab bidimensional structure, containing a set of variables as the seismic signal metadata. Table 2 shows a brief description of each variable. Therefore, the archive that is delivered is primarily a Matlab file (.MAT), which follows this data model. Alternatively, we have provided the data in the open-standard JSON data-interchange format for non Matlab users (e.g. Python users). It should be pointed out that a

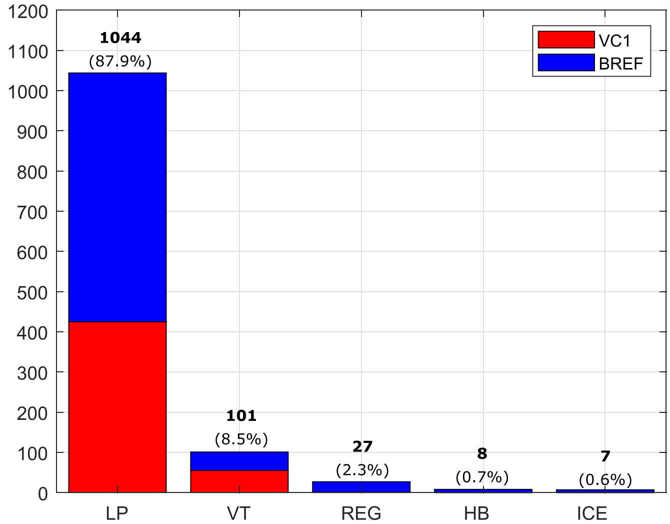

Fig. 2. Distribution of recorded seismic events at VC1 and BREF stations during the experimentation period.

minimal Matlab or Python programming knowledge is required for a better user experience.

4.3. Dataset of volcanic seismic signals

In its first version the volcanic seismic signals (*MicSigV1*) dataset provides a set of discrete seismic signals composed of an array of points with the corresponding relevant associated information, which consists in the timestamps of the starting and ending point (detected by the STA/LTA algorithm (Vaezi and Van der Baan, 2015)) annotations of the current event, as well as its class label according to the classification made manually by the IGEPN specialists.

Before detection, the raw signals were detrended to remove the instrumental effects and then filtered by means of a 128-th order zero-phase noncausal highpass FIR filter with a cutoff frequency of 1 Hz. The goal of this filtering stage aims primarily at removing a predominant peak related to sea microseisms existing at 0.2 Hz, and preserving the main spectral content for microseismic events concentrated from 2 Hz to 20 Hz. Zero-phase noncausal digital filtering processes the input signal, in both the forward and reverse directions. Although it is known that onsets might be distorted with noncausal filters (Rousselet, 2012), observe that given the high order filter employed (i.e. a small transition band) and the cut-off frequency which is far from the 2 Hz–20 Hz microseism event's band, it is safe to use a noncausal filter (please refer to Rousselet (2012) for a complete analysis on this subject). Moreover, FIR filters' linear phase does not produce phase distortion.

Table 2
Variable descriptions of the volcano seismic dataset.

Variable	Description
Network	Ecuadorian Network Code Value: EC
Station	Station Code Possible values: VC1, BREF
SampleRate	Sampling Rate in Hz Possible values: 100 for VC1, 50 for BREF
Component	Seismometer Axe Value: SHZ (vertical axis)
Year	Year in #### format
Month	Month in ## format
Type	Label event. Possible values: VT, LP, ICE, HB, REG
Duration	Duration in seconds
StartPoint	Event starting point in samples with respect to *Data* variable
EndPoint	Event endpoint in samples with respect to *Data* variable
Data	Signal extending 10 s before and after the detected event

The *MicSigV1* dataset was created with a total number of 1187 seismic records according to the distribution depicted in Fig. 2 and were stored in Matlab's.MAT format (see Table 2) and JSON format for non Matlab users. Examples of some seismic events inside this dataset are shown in Fig. 3 and Fig. 4. Fig. 2 shows a typical VT and LP from *MicSigV1* manually labeled by IGEPN experts where the VT waveform and high frequency spectral content are very distinguishable from the LP event. This dataset will serve mainly for providing research challenges in the signal processing, seismic events detection, feature calculation, and seismic events classification.

4.4. Seismic benchmark dataset

On the other hand, the seismic benchmark (*SeisBenchV1*) dataset provides in its first version, a set of feature vectors obtained from the original *MicSigV1* volcanic seismic events, as well as its corresponding class label.

Each vector of features contains a total of 84 descriptors from the time (13 features), frequency (21 features), and scale domains (50 features), which were computed from each recorded seismic signal (see Appendix Table 4). The frequency features were calculated using the periodogram power spectral density estimate, while the scale features were extracted from the application of the Wavelet transform. Time features, on the other hand, belong to the statistical calculation of the signal. The feature set was chosen to be broad and general so that researchers can carry out feature relevance, redundancy, and selection studies based on their approaches.

The *SeisBenchV1* dataset was formed with a total number of 1187 feature vectors, containing event classes according to the distribution shown in Fig. 2. This dataset will serve mainly for providing research challenges in the feature selection, feature space reduction, and seismic events classification. A complete description of the features, as well as more detailed information about their calculation, is shown in Appendix Table 4.

Fig. 5 shows the *SeisBenchV1* two-dimensional embedding using t-Distri-buted Stochastic Neighbor Embedding (t-SNE) (Maaten and Hinton, 2008), which is a widely used nonlinear dimensionality reduction technique for the visualization of high-dimensional datasets. Hence, it allows visualizing the 84-dimensional *SeisBenchV1* dataset to gather some insights about the relevance of computed features. From Fig. 5, the two-dimensional embedding depicts two clusters corresponding to VC1 (circle symbol) and BREF (plus symbol) stations, which demonstrates that features are station-dependent as expected since they are located at different positions. Moreover, inside each station cluster, VT, LP, and regional events tend to form event clusters, which demonstrate that features are highly relevant to discriminate among different types of seismic signals. It should be observed that hybrid events tend to spread out along LP or VT event clusters because they are sharing traits from both type of events. Finally, since the visual classification task is particularly complex even for experts, some overlapping between LP and VT events is expected in the embedded visualization, an issue that constitutes one of the intrinsic challenges when using this dataset.

5. Performance evaluation

Several approaches summarized in Table 3 have been proposed during the last years using earlier versions or subsets of the released datasets in this work, mainly related to seismic events detection, feature subset selection, classification, or any combination of these techniques. A summary of the obtained results is presented below to serve as a benchmark for event detection and classification performance of Eseismic and to facilitate comparisons with other event detection and classification schemes.

In Jaramillo et al. (2014), a new method to detect and characterize volcano seismic events was reported. The method was based on the

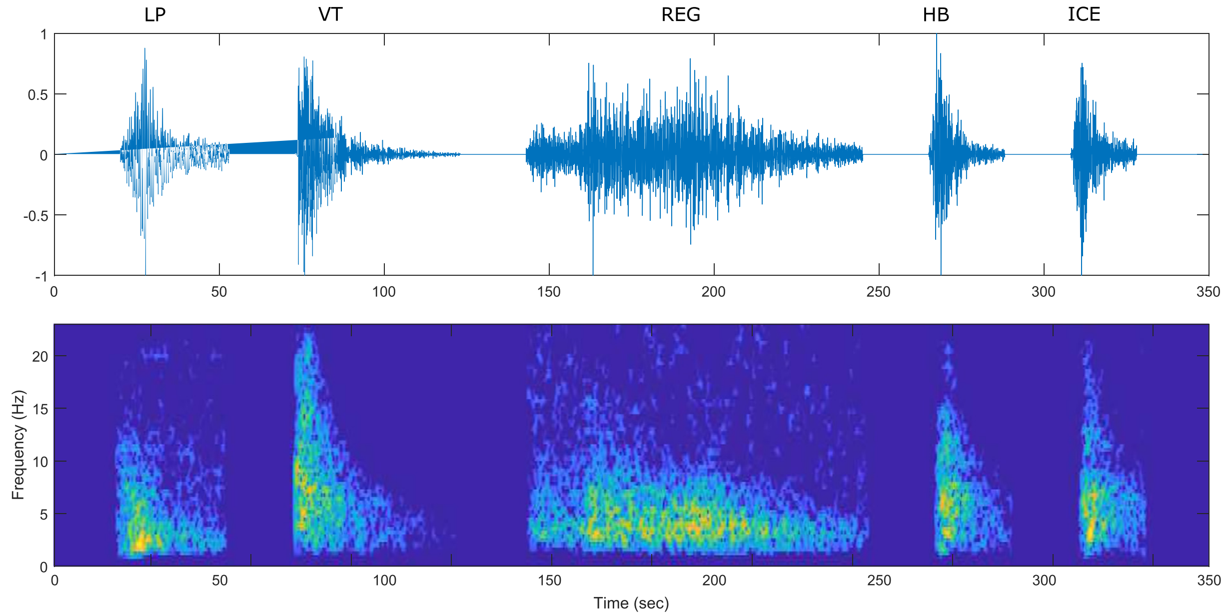

Fig. 3. Example of seismic time-domain signals (top) and their respective spectrogram (bottom) from the *MicSigV1* dataset. The time signals were normalized by their maximum absolute value.

use of classic spectral and maximum entropy estimators to detect the locations in time where some LP or VT seismic events occur. The detector performs in near real-time, maximizing the detection probability while maintaining the false alarm rate constant.

In Lara-Cueva et al. (2016d), a real-time detector of LP and VT seismic events based on voice recognition algorithms was proposed. This detector was able to find the event's start and endpoints (timestamps) in the original signal. The method was validated on a dataset containing 436 seismic events, while the accuracy (ACC) and balanced error rate (BER) scores were considered as the principal assessment metrics. Values of 95.2% and 0.005 were obtained for ACC and BER, respectively, demonstrating therefore that this detector was successful and less complex when compared to other previously developed approaches.

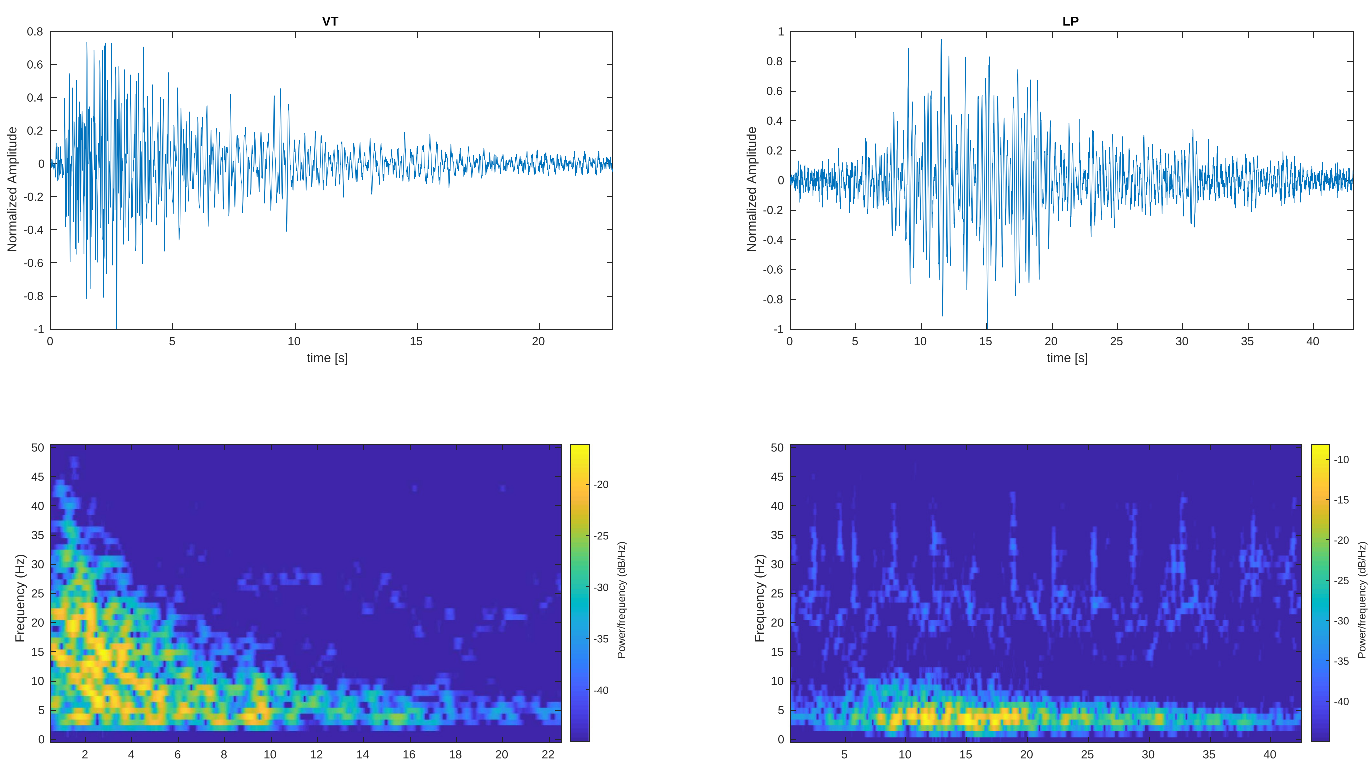

Fig. 4. A typical VT and LP from the *MicSigV1* dataset manually classified by IGEPN experts. The time signals were normalized by their maximum absolute value.

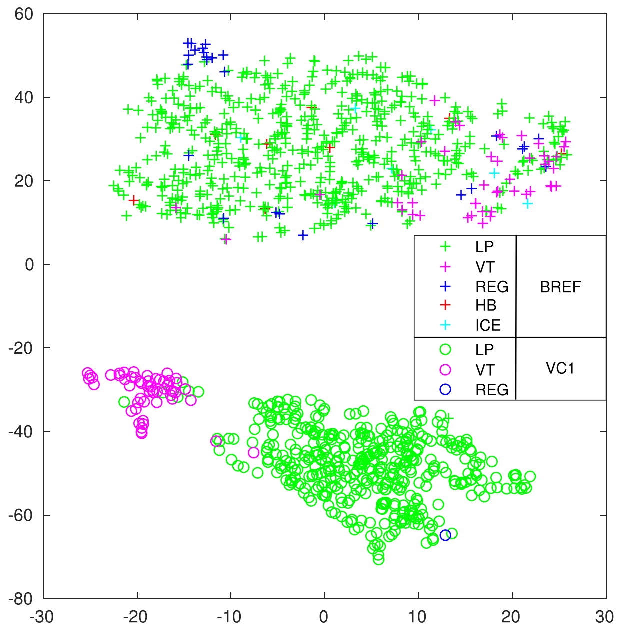

Fig. 5. *SeisBenchV1* two-dimensional embedding using t-SNE. Plus and circle symbols represent events from BREF and VC1 stations, respectively.

Others approaches like those presented in Lara-Cueva et al. (2016a, 2016c) combined feature selection techniques with machine learning classifiers to decrease the size of the feature vectors input to the classifier in order to reduce the processing time required by the classification stage.

In Lara-Cueva et al. (2016a) mutual and statistical information techniques were used as filter methods, and cross-validation and pruning methods were used as an embedded model for selecting the most appropriate subset of features. Subsequently, these subsets were used for feeding a k-nearest neighbor (kNN) and decision tree (DT) classifiers to select the classification model that maximizes the ACC and sensitivity metrics. In a dataset containing a total of 914 instances, a best classification model was composed by using 15 s of segmentation window, a features matrix computed from the frequency domain, and the decision tree classifier, attaining a 99% performance on both evaluated metrics. Similarly, in Lara-Cueva et al. (2016c), a combination of feature selection methods and machine learning classifiers, was used for the detection and classification of LP and VT seismic events in real-time. This work combined in a suitable way the cross-validation and pruning methods as an embedded model and, a recursive feature selection based wrapper method, for selecting the most important subsets of features on a features space containing statistical, temporal, spectral and scale domain-based features. The best result was obtained using the support vector machine classifier on both the detection and classification stages, yielding ACC scores of 99% and 97%, respectively. The most relevant subset of features included time (one), frequency (five), and scale (nine) domains features.

Table 3
Summary of several approaches that have used the released datasets.

Approach	Main results	Ref.
Detection based on classic spectraland maximum entropy estimators	Near real-time detection	Jaramillo et al. (2014)
Detection based on voice recognition	ACC 92.5% BER 0.005	Lara-Cueva et al. (2016d)
Feature selection and classification using kNN and DT	ACC 99% with DT	Lara-Cueva et al. (2016a)
Feature selection, detection and classification using SVM	ACC 99% with SVM	Lara-Cueva et al. (2016c)
Feature selection and classification using kNN, DT and ANN	ACC 98% with kNN	Lara-Cueva et al. (2016b)
Detection and classification using DT and SVM	ACC 98% for detection, 90% for classification	Lara-Cueva et al. (2017)
Filter-based feature selection with a Gaussian mixture classifier	ACC 94.7% BER 14.95%	Venegas et al. (2019b)
Classification using an image-based descriptor	ACC 96%	Pérez et al. (2020)
Exhaustive comparison of 69 classifiers	ACC 95% AUC 93.4% with random forest	Venegas et al. (2019a)
Classification using a semi-supervised approach	ACC 94.7% BER 14.95%	Brusil et al. (2019)

In Lara-Cueva et al. (2016b), a performance comparison of three well-known machine learning classifiers (MLCs) was made: k-nearest neighbors algorithm (kNN), decision tree (DT), and an artificial neural network (ANN) in the classification of LP and VT seismic events. Then, a feature selection process was carried out to determine features with a higher impact in the classification. As a result, it was found that the kNN classifier only needed three features for reaching the highest ACC value of 98%.

In Lara-Cueva et al. (2017), an automatic method for the detection and classification of volcano origin events such as LP, VT, and signals from non-volcano origin earthquakes such as lightning and background noise, was developed. The detection and classification steps used a DT and a multi-class support vector machine (SVM) classifier, respectively. The best-obtained results were an ACC value of 98% for the detection and 90% for the classification.

In Venegas et al. (2019a, 2019b), extensive experimentation using five filter-based feature selection methods in combination with a Gaussian mixture classifier, was made. The reported method was able to determine the best classification scheme for the classification of LP events, VT events, and both events together (as disjoint classes), yielding ACC scores of 95.62%, 96.71% and 96.70%, respectively. Also, it was stated that by using a minimum set of 5 features among the 84 included in the experimental dataset was sufficient for reaching the higher ACC value while decreasing the processing time to the range of 3 to 4 milliseconds.

In Pérez et al. (2020), a new volcano seismic signal descriptor was developed to improve the AUC in the classification of LP and VT seismic events. The method describes volcanic seismic events based on image processing techniques instead of classic seismic signal processing strategies such as frequencies or scale analysis. The intensity statistic of pixels, shape, and texture-based features computed from the event spectogram gray level image were used to form a numeric vector output that was employed to feed five MLCs with different taxonomies. The descriptor was validated on a seismic signal database, containing a total of 637 instances, including LP, VT, and overlapped seismic events (e.g., rockfall or icequakes). The determination of the event patterns reached an ACC value of 96%. In addition, the classification of events using a feed-forward back-propagation (FFBP) artificial neural networks classifier attained AUC values of 0.95 and 0.96, in a dataset with and without events overlapping, respectively.

In Venegas et al. (2019a, 2019b), a method was proposed to exhaustively explore a pool of 69 MLCs derived from five different classifiers: naive Bayes, SVM, kNN, feed-forward back propagation (FFBP) artificial neural network, and random forest. The method attends to find models that provide the higher AUC scores statistically in the classification of LP and VT seismic events. The validation was carried out on an experimental dataset, containing 587 and 81 instances belonging to LP and VT class, respectively. A random forest classifier with 100 tree-based predictors demonstrated to be the best classification model, reaching an ACC and AUC scores of 0.95 and 0.934, respectively. Thus, the proposed method was effective in providing competitive models for the classification of volcano seismic events.

Lately, in Brusil et al. (2019), a semi-supervised approach using the self-training algorithm was proposed to alleviate the need for abundant labeled data. Although the accuracy of using a supervised scheme in a dataset containing 332 instances (50 VT and 282 LP) is still better than

Table 4
Summary of features.

Time-domain		Frequency-domain		Scale-domain	
ID	Feature name	ID	Feature name	ID	Feature name
f1	Mean	f29	Power	f57	Percentage of energy for D1
f2	Standard deviation	f30	Density of peaks above RMS[d]	f58	Percentage of energy for D2
f3	Variance	f31	2nd highest peak value	f59	Percentage of energy for D3
f4	Entropy[a]	f32	Freq. of 2nd highest peak	f60	Percentage of energy for D4
f5	Kurtosis	f33	3rd highest peak value	f61	Percentage of energy for D5
f6	Multiscale entropy (MSE)[b]	f34	Freq. of 3rd highest peak	f62	Percentage of energy for D6
f7	Time to reach the maximum peak	**Scale-domain**		f63	A6 RMS in time-domain
f8	RMS value	f35	A6 Max. peak in freq.-domain	f64	D1 RMS in time-domain
f9	Peak-to-peak value	f36	D1 Max. peak in freq.-domain	f65	D2 RMS in time-domain
f10	Peak-to-RMS ratio[c]	f37	D2 Max. peak in freq.-domain	f66	D3 RMS in time-domain
f11	Energy	f38	D3 Max. peak in freq.-domain	f67	D4 RMS in time-domain
f12	Zero-crossing rate	f39	D4 Max. peak in freq.-domain	f68	D5 RMS in time-domain
f13	Density of peaks above RMS[d]	f40	D5 Max. peak in freq.-domain	f69	D6 RMS in time-domain
Frequency-domain		f41	D6 Max. peak in freq.-domain	f70	A6 Peak-to-peak in time-domain
f14	Frequency of maximum peak	f42	A6 Freq. of max. peak	f71	D1 Peak-to-peak in time-domain
f15	Bandwidth of 90% energy[e]	f43	D2 Freq. of max. peak	f72	D2 Peak-to-peak in time-domain
f16	Entropy[a]	f44	D3 Freq. of max. peak	f73	D3 Peak-to-peak in time-domain
f17	Mean	f45	D4 Freq. of max. peak	f74	D4 Peak-to-peak in time-domain
f18	Standard deviation	f46	D5 Freq. of max. peak	f75	D5 Peak-to-peak in time-domain
f19	Variance	f47	D6 Freq. of max. peak	f76	D6 Peak-to-peak in time-domain
f20	Energy	f48	A6 Mean in freq.-domain	f77	A6 Peak-to-RMS ratio in time-domain[c]
f21	Kurtosis	f49	D1 Mean in freq.-domain	f78	D1 Peak-to-RMS ratio in time-domain
f22	Multiscale entropy[b]	f50	D2 Mean in freq.-domain	f79	D2 Peak-to-RMS ratio in time-domain
f23	Maximum peak in 10–20 Hz band	f51	D3 Mean in freq.-domain	f80	D3 Peak-to-RMS ratio in time-domain
f24	Freq. of max. peak in 10–20 Hz Band	f52	D4 Mean in freq.-domain	f81	D4 Peak-to-RMS ratio in time-domain
f25	Maximum peak in 20–30 Hz band	f53	D5 Mean in freq.-domain	f82	D5 Peak-to-RMS ratio in time-domain
f26	Freq. of max. Peak in 20–30 Hz Band	f54	D6 Mean in freq.-domain	f83	D6 Peak-to-RMS ratio in time-domain
f27	RMS value	f55	Mean energy of components[f]	f84	Mean energy of wavelet coefficients
f28	Peak-to-RMS ratio[c]	f56	Percentage of energy for A6[g]		

[a] Nonnormalized Shannon Entropy (Coifman and Wickerhauser, 1992) calculated using $E(x) = -\sum_i x_i^2 \log_e(x_i^2)$.
[b] MSE at scale 1. MSE measures the degree of complexity of time series signals through Sample Entropy (Costa et al., 2005).
[c] Ratio of largest absolute (i.e L-Infinity norm) to root mean squared value.
[d] Ratio of the number of peaks whose amplitudes are higher than the RMS value to the length of the signal.
[e] Minimum frequency at which the cumulative energy is at least 90% of the total energy.
[f] Mean energy of A6, D1, D2, D3, D4, D5 and D6 wavelet's signal components.
[g] Energy fraction of wavelet's signal component with respect to total energy.

a semi-supervised one, it was demonstrated that allowing a 10% false positive rate, the semi-supervised approach achieved similar performance with respect to supervised techniques with only 50% of labeled data.

6. Discussion

This section discusses the principal advantages and disadvantages of the proposed repository.

Advantages:

- Public access to two seismic datasets of volcano origin. As it was mentioned in previous sections, there is no other currently available database or repository that publicly offer data from both raw signals as well as their associated events. Instead of that, several groups are developing research under a data privatization condition, which is limited to the scientific community until now.
- All the samples were carefully analyzed and labeled by IGEPN specialists to provide the correct associated information (e.g. timestamps, event type) related to each of them. This process facilities the faithfully validation and reproducibility of all employed methodologies and techniques by researchers.
- The released datasets contain the minimum of samples needed for carrying out research regarding the LP and VT seismic events. As it was corroborated by the results described in the performance evaluation section, it is possible to use these datasets for producing satisfactory results in terms of signal processing, feature calculation, feature selection and reduction, seismic events detection, and classification.
- The continuous repository supported over time by the IGEPN specialists, which have the control and authority to perform tasks like updating, changing, and releasing in the future new versions of these volcano seismic datasets, and which might also include previously unseen events (e.g. deep LPs). In this sense, our databases can be seen as dynamic entities.

Disadvantages

- The small number of samples in the classes like hybrid and icequakes and, the unbalanced ratio of instances between LP and VT classes. This is limiting by now the experimentation towards any idea on the use of balanced class instances. However, the unbalanced effect is found in the real-life volcano seismic environment and researchers might deal with this imbalanced problem using, e.g., resampling or ensemble methods (Opitz and Maclin, 1999).

Nonetheless, the proposed repository arises as a novel idea in Ecuador, and it can allow the development of new research in the volcano seismic signal analysis area.

7. Conclusions and future work

This work presents the development, description, and performance evaluation of two volcano seismic event datasets recorded at Cotopaxi volcano, in Ecuador. These datasets represent the first public Ecuadorian volcano seismic origin repository that will update, publish, and support the release of new data to the scientific community over time. Data include raw signals, events labeled by experts, and the event starting and endpoints, which is an important contribution towards the development of new seismic event detection and classification algorithms. The datasets, in its first version, contain 1187 instances of several types of seismic signals, that will serve to provide research challenges in the signal processing, feature calculation, feature selection, feature space reduction, seismic events detection, and classification areas. The repository could be accessed through a disclaimer agreement that mainly helps to the correct citation and use of it. We hope that these datasets can enable the research community to compare the performance of different algorithms in the same way that standardized datasets have supported research in other fields.

As future work, we plan to increase the number of samples per class on each dataset. That means augmenting the examples of existing classes and inserting new ones of previously unseen classes, such as hybrid events, thus making them statistically representative. The continuous release of other data recorded at different experimentation periods, volcanoes and stations, is also expected. Finally, to complement our database, in future dataset versions, we also plan to release the continuous waveforms from which the raw signal events were detected as well as hypocentral information. We expect that with this information, new strategies might be developed to ponderate several predictions from different stations across volcanoes.

CRediT authorship contribution statement

Noel Pérez: Conceptualization, Methodology, Software, Writing - original draft, Writing - review & editing, Investigation. **Diego Benítez:** Conceptualization, Methodology, Writing - original draft, Writing - review & editing, Investigation, Supervision. **Felipe Grijalva:** Visualization, Writing - review & editing, Validation, Data curation, Investigation. **Román Lara-Cueva:** Investigation, Formal analysis. **Mario Ruiz:** Resources, Writing - review & editing, Supervision. **Jorge Aguilar:** Software.

Declaration of competing interest

We have no conflicts of interest to disclose.

Acknowledgment

This work was supported in part by the Universidad San Francisco de Quito (USFQ) under Grants 10100, 12494 and 16916, in part by the Universidad de las Fuerzas Armadas ESPE under Grants 2013-PIT-014 and 2016-EXT-038, and in part by Escuela Politécnica Nacional (EPN) under the Grant PIE-SRASI-IG-2018. The seismic data used for the described datasets were provided by the Instituto Geofísico, EPN, Quito-Ecuador. Authors thank P. Mothes for proofreading the final document.

Appendix A

The time signals were normalized by their maximum absolute value to compute 13 time-domain features (i.e., f1 to f13). Afterward, 21 frequency-domain features (i.e., f14 to f34) were computed from the periodogram power spectral density estimate with a rectangular window using 1024 FFT points. All frequency-domain features were extracted from the log-magnitude power spectrum in dB except for f28 and f30 that were calculated from the magnitude power spectrum. Finally, the scale-domain features were computed from a wavelet decomposition using the db10 Daubechies family up to level 6, forming 50 features (i.e., f35 to f84) from the reconstructed approximation at level 6 (i.e., A6), and the reconstructed detail at levels 1 through 6 (i.e., D1, D2, D3, D4, D5, and D6).

References

Alasonati, P., Wassermann, J., Ohrnberger, M., Mader, H., Connor, C., Coles, S., 2006. Signal classification by wavelet-based hidden markov models: application to seismic signals of volcanic origin. Statistics in Volcanology 1, 161–174. https://doi.org/10.1144/IAVCEI001.13.

Almendros, J., Carmona, E., Agui, J., Moreno, J., Martos, A., Alguacil, G., Martin, J., Jimenez, V., Diaz-Moreno, A., Lorenzo, F., 2018a. Seismic Network at Deception Island Volcano, Antarctica, 2014/2015 Australian Antarctic Data Centre. URL. https://doi.org/10.4225/15/5ac5681e2f528.

Almendros, J., Carmona, E., Agui, J., Moreno, J., Martos, A., Alguacil, G., Martin, J., Jimenez, V., Diaz-Moreno, A., Lorenzo, F., 2018b. Seismic Stations at the Bransfield Strait, Antarctica, 2014–2015 Australian Antarctic Data Centre. URL. https://doi.org/10.4225/15/5ac56ba111460.

Alvarez, I., Garcia, L., Cortes, G., Benitez, C., De la Torre, Á., 2011. Discriminative feature selection for automatic classification of volcano-seismic signals. IEEE Geosci. Remote Sens. Lett. 9, 151–155. https://doi.org/10.1109/LGRS.2011.2162815.

Apolloni, B., et al., 2009. Support vector machines and mlp for automatic classification of seismic signals at stromboli volcano. Neural Nets WIRN09: Proceedings of the 19th Italian Workshop on Neural Nets, Vietri Sul Mare, Salerno, Italy May 28–30 2009. vol. 204. IOS Press, p. 116. https://doi.org/10.3233/978-1-60750-072-8-116.

Bean, C.J., De Barros, L., Lokmer, I., Métaxian, J.-P., O'Brien, G., Murphy, S., 2014. Long-period seismicity in the shallow volcanic edifice formed from slow-rupture earthquakes. Nat. Geosci. 7, 71. https://doi.org/10.1038/ngeo2027.

Benítez, M.C., Ramrez, J., Segura, J.C., Ibanez, J.M., Almendros, J., Garca-Yeguas, A., Cortes, G., 2006. Continuous hmm-based seismic-event classification at deception island, Antarctica. IEEE Trans. Geosci. Remote Sens. 45, 138–146. https://doi.org/10.1109/TGRS.2006.882264.

Bicego, M., Acosta-Muñoz, C., Orozco-Alzate, M., 2012. Classification of seismic volcanic signals using hidden-markov-model-based generative embeddings. IEEE Trans. Geosci. Remote Sens. 51, 3400–3409. https://doi.org/10.1109/TGRS.2012.2220370.

Brown, S.K., Jenkins, S.F., Sparks, R.S.J., Odbert, H., Auker, M.R., 2017. Volcanic fatalities database: analysis of volcanic threat with distance and victim classification. J. Appl. Volcanol. 6, 15. https://doi.org/10.1186/s13617-017-0067-4.

Brusil, C., Grijalva, F., Lara-Cueva, R., Ruiz, M., Acuña, B., 2019, November. A Semi-Supervised Approach for Microseisms Classification from Cotopaxi Volcano. 2019 IEEE Latin American Conference on Computational Intelligence (LA-CCI). IEEE, pp. 1–6. https://doi.org/10.1109/LA-CCI47412.2019.9037033.

Chester, D.K., Degg, M., Duncan, A.M., Guest, J.E., 2000. The increasing exposure of cities to the effects of volcanic eruptions: a global survey. Global Environ Change B Environ Hazard 2, 89–103. https://doi.org/10.3763/ehaz.2000.0214.

Chouet, B.A., 1996. Long-period volcano seismicity: its source and use in eruption forecasting. Nature 380, 309. https://doi.org/10.1038/380309a0.

Coifman, R.R., Wickerhauser, M.V., 1992. Entropy-based algorithms for best basis selection. IEEE Trans. Inf. Theory 38, 713–718. https://doi.org/10.1109/18.119732.

Cortés, G., Benitez, M.C., Garca, L., Álvarez, I., Ibanez, J.M., 2015. A comparative study of dimensionality reduction algorithms applied to volcano-seismic signals. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 9, 253–263. https://doi.org/10.1109/JSTARS.2015.2479300.

Costa, M., Goldberger, A.L., Peng, C.-K., 2005. Multiscale entropy analysis of biological signals. Phys. Rev. E 71, 021906. https://doi.org/10.1103/PhysRevE.71.021906.

Curilem, G., Vergara, J., Fuentealba, G., Acuña, G., Chacón, M., 2009. Classification of seismic signals at Villarrica volcano (Chile) using neural networks and genetic algorithms. J. Volcanol. Geotherm. Res. 180, 1–8. https://doi.org/10.1016/j.jvolgeores.2008.12.002.

Cusano, P., Palo, M., West, M., 2015. Long-period seismicity at shishaldin volcano (Alaska) in 2003–2004: indications of an upward migration of the source before a minor eruption. J. Volcanol. Geotherm. Res. 291, 14–24. https://doi.org/10.1016/j.jvolgeores.2014.12.008.

Del Pezzo, E., Esposito, A., Giudicepietro, F., Marinaro, M., Martini, M., Scarpetta, S., 2003. Discrimination of earthquakes and underwater explosions using neural networks. Bull. Seismol. Soc. Am. 93, 215–223. https://doi.org/10.1785/0120020005.

Esposito, A.M., Giudicepietro, F., D'Auria, L., Scarpetta, S., Martini, M.G., Coltelli, M., Marinaro, M., 2008. Unsupervised neural analysis of very-long-period events at stromboli volcano using the self-organizing maps. Bull. Seismol. Soc. Am. 98, 2449–2459. https://doi.org/10.1785/0120070110.

Gutiérrez, L., Ramrez, J., Ibañez, J., Bentez, C., 2011. Volcano-seismic signal detection and classification processing using hidden markov models-application to San cristóbal and Telica volcanoes, Nicaragua. Hidden Markov Models, Theory and Applications, p. 187. https://doi.org/10.5772/14312.

Hammer, C., Beyreuther, M., Ohrnberger, M., 2012. A seismic-event spotting system for volcano fast-response systems. Bull. Seismol. Soc. Am. 102, 948–960. https://doi.org/10.1785/0120110167.

Havskov, J., Ottemoller, L., 2010. Routine Data Processing in Earthquake Seismology. *Editorial*. Springer https://doi.org/10.1007/978-90-481-8697-6.

Ibáñez, J.M., Bentez, C., Gutiérrez, L.A., Cortés, G., Garca-Yeguas, A., Alguacil, G., 2009. The classification of seismo-volcanic signals using hidden markov models as applied to the stromboli and Etna volcanoes. J. Volcanol. Geotherm. Res. 187, 218–226. https://doi.org/10.1016/j.jvolgeores.2009.09.002.

Ibs-von Seht, M., 2008. Detection and identification of seismic signals recorded at krakatau volcano (Indonesia) using artificial neural networks. J. Volcanol. Geotherm. Res. 176, 448–456. https://doi.org/10.1016/j.jvolgeores.2008.04.015.

Incorporated Research Institutions for Seismology - IRIS, 2019. Waveform data, time-series. URL. https://ds.iris.edu/ds/nodes/dmc/data/types/waveform-data/, Accessed date: 1 June 2019.

Instituto Geofísico de la Escuela Politécnica Nacional – IGEPN, 2019. Red de Observatorios Vulcanológicos (ROVIG). URL. https://www.igepn.edu.ec/red-de-observatorios-vulcanologicos-rovig, Accessed date: 1 June 2019.

IRIS, 2018. NSF Makes 5-Year $93m Award to IRIS to Manage the SAGE Facility. IRIS URL. https://www.iris.edu/hq/news/story/nsf_makes_5_year_93m_award_to_iris_to_manage_the_sage_facility.
Jaramillo, C., León, R., Lara-Cueva, R., Benitez, D.S., Ruiz, M., 2014. A new structure for sequential detection and maximum entropy spectral estimator for characterization of volcanic seismic signals. 2014 IEEE Latin-America Conference on Communications (LATINCOM). IEEE, pp. 1–6. https://doi.org/10.1109/LATINCOM.2014.7041847.
Koulakov, I., Komzeleva, V., Abkadyrov, I., Kugaenko, Y., El Khrepy, S., Al Arifi, N., 2019. Unrest of the udina volcano in Kamchatka inferred from the analysis of seismicity and seismic tomography. J. Volcanol. Geotherm. Res. 379, 45–59. https://doi.org/10.1016/j.jvolgeores.2019.05.006.
Langer, H., Falsaperla, S., Thompson, G., 2003. Application of artificial neural networks for the classification of the seismic transients at Soufriere hills volcano, Montserrat. Geophys. Res. Lett., 30 https://doi.org/10.1029/2003GL018082.
Langer, H., Falsaperla, S., Powell, T., Thompson, G., 2006. Automatic classification and a-posteriori analysis of seismic event identification at Soufriere hills volcano, Montserrat. J. Volcanol. Geotherm. Res. 153, 1–10. https://doi.org/10.1016/j.jvolgeores.2005.08.012.
Lara-Cueva, R., Benitez, D., Carrera, E., Ruiz, M., Rojo-Álvarez, J.L., 2016a. Feature selection of seismic waveforms for long period event detection at Cotopaxi volcano. J. Volcanol. Geotherm. Res. 316, 34–49. https://doi.org/10.1016/j.jvolgeores.2016.02.022.
Lara-Cueva, R., Carrera, E.V., Morejon, J.F., Benítez, D., 2016b. Comparative analysis of automated classifiers applied to volcano event identification. 2016 IEEE Colombian Conference on Communications and Computing (COLCOM), pp. 1–6. https://doi.org/10.1109/ColComCon.2016.7516377.
Lara-Cueva, R., Benítez, D.S., Paillacho, V., Villalva, M., Rojo-Álvarez, J.L., 2017. On the use of multi-class support vector machines for classification of seismic signals at Cotopaxi volcano. 2017 IEEE International Autumn Meeting on Power, Electronics and Computing (ROPEC), pp. 1–6. https://doi.org/10.1109/ROPEC.2017.8261613.
Lara-Cueva, R.A., Benítez, D.S., Carrera, E.V., Ruiz, M., Rojo-Álvarez, J.L., 2016c. Automatic recognition of long period events from volcano tectonic earthquakes at Cotopaxi volcano. IEEE Trans. Geosci. Remote Sens. 54, 5247–5257. https://doi.org/10.1109/TGRS.2016.2559440.
Lara-Cueva, R.A., Moreno, A.S., Larco, J.C., Benítez, D.S., 2016d. Real-time seismic event detection using voice activity detection techniques. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 9, 5533–5542. https://doi.org/10.1109/JSTARS.2016.2605061.
Maaten, L.v.d., Hinton, G., 2008. Visualizing data using t-sne. J. Mach. Learn. Res. 9, 2579–2605.
Maggi, A., Ferrazzini, V., Hibert, C., Beauducel, F., Boissier, P., Amemoutou, A., 2017. Implementation of a multistation approach for automated event classification at piton de la fournaise volcano. Seismol. Res. Lett. 88, 878–891. https://doi.org/10.1785/0220160189.
Malfante, M., Dalla Mura, M., Métaxian, J.-P., Mars, J.I., Macedo, O., Inza, A., 2018. Machine learning for volcano-seismic signals: challenges and perspectives. IEEE Signal Process. Mag. 35, 20–30. https://doi.org/10.1109/MSP.2017.2779166.
MATLAB, 2019. Version 9.6.0.1 (R2019a). The MathWorks Inc, Natick, Massachusetts.
McNutt, S.R., 1996. Seismic monitoring and eruption forecasting of volcanoes: a review of the state-of-the-art and case histories. Monitoring and Mitigation of Volcano Hazards. Springer, pp. 99–146. https://doi.org/10.1007/978-3-642-80087-0_3.
National Geophysical Data Center / /World Data Service (NGDC/WDS), 2019. Significant Earthquake Database. National Geophysical Data Center, NOAA https://doi.org/10.7289/V5TD9V7K (accessed 2019-06-01).
Newhall, C., Costa, F., Ratdomopurbo, A., Venezky, D., Widiwijayanti, C., Win, N.T.Z., Tan, K., Fajiculay, E., 2017. Wovodat–an online, growing library of worldwide volcanic unrest. J. Volcanol. Geotherm. Res. 345, 184–199. https://doi.org/10.1016/j.jvolgeores.2017.08.003.
Opitz, D., Maclin, R., 1999. Popular ensemble methods: an empirical study. J. Artif. Intell. Res. 11, 169–198.
Orozco, M., Garca, M.E., Duin, R.P., Castellanos, C.G., 2006. Dissimilarity-based classification of seismic signals at nevado del ruiz volcano. Earth Sci Res J 10, 57–66. https://doi.org/10.15446/esrj.
Ortiz, H.D.E., 2013. Study of the Site Effects for the Construction of a Seismic Activity Index at the Cotopaxi Volcano - Estudio de los Efectos de Sitio Para la construcción de Un índice de Actividad sísmica en el volcán Cotopaxi. Master's thesis. Escuela Politécnica Nacional, Quito, Ecuador.
Pérez, N., Venegas, P., Benítez, D., Lara-Cueva, R., Ruiz, M., 2020. A new volcanic seismic signal descriptor and its application to a data set from the Cotopaxi volcano. IEEE Trans. Geosci. Remote Sens., 1–11 https://doi.org/10.1109/TGRS.2020.2976896 (Accepted for publication).
Phillipson, G., Sobradelo, R., Gottsmann, J., 2013. Global volcanic unrest in the 21st century: an analysis of the first decade. J. Volcanol. Geotherm. Res. 264, 183–196. https://doi.org/10.1016/j.jvolgeores.2013.08.004.
RESIF, French National Sismological Network, 2014. RESIF Seismic data portal. URL:. http://seismology.resif.fr/.
Ringler, A.T., Evans, J.R., 2015. A quick seed tutorial. Seismol. Res. Lett. 86, 1717–1725. https://doi.org/10.1785/0220150043.
Rodríguez Cesén, M.J., Lara-Cueva, R.A., 2018. A real-time microseisms-detector based on voice activity detection and endpoint detection: an approach to Cotopaxi volcano. URL. http://repositorio.espe.edu.ec/handle/21000/15565.
Roman, D.C., Cashman, K.V., 2018. Top–down precursory volcanic seismicity: Implications for 'stealth'magma ascent and long-term eruption forecasting. Front. Earth Sci. 6, 124.
Rousselet, G.A., 2012. Does filtering preclude us from studying erp time-courses? Front. Psychol. 3, 131.
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, 2305–2308. https://doi.org/10.1029/98GL01689.
Scarpetta, S., Giudicepietro, F., Ezin, E.C., Petrosino, S., Del Pezzo, E., Martini, M., Marinaro, M., 2005. Automatic classification of seismic signals at mt. vesuvius volcano, Italy, using neural networks. Bull. Seismol. Soc. Am. 95, 185–196. https://doi.org/10.1785/0120030075.
Schmincke, H.-U., 2004. Volcanic hazards, volcanic catastrophes, and disaster mitigation. Volcanism. Springer, pp. 229–258. https://doi.org/10.1007/978-3-642-18952-4_13.
Siebert, L., Simkin, T., Kimberly, P., 2011. Volcanoes of the World. Univ of California Press.
Tilling, R.I., 1996. Hazards and climatic impact of subduction-zone volcanism: a Global and Historical Perspective. Washington DC. American Geophysical Union Geophysical Monograph Series 96, 331–335. https://doi.org/10.1029/GM096p0331.
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. J. Geophys. Res. Solid Earth 116. https://doi.org/10.1029/2010JB008180.
Vaezi, Y., Van der Baan, M., 2015. Comparison of the STA/LTA and power spectral density methods for microseismic event detection. Geophys. J. Int. 203, 1896–1908. https://doi.org/10.1093/gji/ggv419.
Venegas, P., Pèrez, N., Benítez, D., Lara-Cueva, R., Ruiz, M., 2019a. Building Machine Learning Models for Long-Period and Volcano-Tectonic Event Classification. 2019 IEEE CHILEAN Conference on Electrical, Electronics Engineering, Information and Communication Technologies (CHILECON), Valparaiso, Chile, pp. 1–6.
Venegas, P., Pérez, N., Benítez, D., Lara-Cueva, R., Ruiz, M., 2019b. Combining filter-based feature selection methods and gaussian mixture model for the classification of seismic events from Cotopaxi volcano. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 12, 1991–2003. https://doi.org/10.1109/JSTARS.2019.2916045.
White, R.A., McCausland, W.A, 2019. A process-based model of pre-eruption seismicity patterns and its use for eruption forecasting at dormant stratovolcanoes. J. Volcanol. Geotherm. Res. 382, 267–297.
Zobin, V.M., 2017. Introduction to volcanic seismology. Science Direct, 3rd edition Elsevier. https://doi.org/10.1016/C2015-0-00304-5.