Practical Volcano-Independent Recognition of Seismic Events: *VULCAN.ears* Project

Guillermo Cortés[1]*, Roberto Carniel[1], Philippe Lesage[2], M. Ángeles Mendoza[3] and Ivo Della Lucia[1]

[1]Dipartimento Politecnico di Ingegneria e Architettura (DPIA), Università degli Studi di Udine, Udine, Italy, [2]Université Grenoble Alpes, Université Savoie Mont Blanc, CNRS, IRD, IFSTTAR, ISTerre, Grenoble, France, [3]Departamento de Física Teórica y del Cosmos, Universidad de Granada, Granada, Spain

Recognizing the mechanisms underlying seismic activity and tracking temporal and spatial patterns of earthquakes represent primary inputs to monitor active volcanoes and forecast eruptions. To quantify this seismicity, catalogs are established to summarize the history of the observed types and number of volcano-seismic events. In volcano observatories the detection and posterior classification or labeling of the events is manually performed by technicians, often suffering a lack of unified criteria and eventually resulting in poorly reliable labeled databases. State-of-the-art automatic Volcano-Seismic Recognition (VSR) systems allow real-time monitoring and consistent catalogs. VSR systems are generally designed to monitor one station of one volcano, decreasing their efficiency when used to recognize events from another station, in a different eruptive scenario or at different volcanoes. We propose a Volcano-Independent VSR (VI.VSR) solution for creating an exportable VSR system, whose aim is to generate labeled catalogs for observatories which do not have the resources for deploying their own systems. VI.VSR trains universal recognition models with data of several volcanoes to obtain portable and robust characteristics. We have designed the *VULCAN.ears* ecosystem to facilitate the VI.VSR application in observatories, including the *pyVERSO* tool to perform VSR tasks in an intuitive way, its graphical interface, *geoStudio*, and *liveVSR* for real-time monitoring. Case studies are presented at Deception, Colima, Popocatépetl and Arenal volcanoes testing VI.VSR models in challenging scenarios, obtaining encouraging recognition results in the 70–80% accuracy range. VI.VSR technology represents a major breakthrough to monitor volcanoes with minimal effort, providing reliable seismic catalogs to characterise real-time changes.

1 INTRODUCTION

Volcanoes have a big impact on the global economy, society and more relevant, in human casualties. It is estimated that about 800 million people live inside the risk area of the 1,500 active volcanoes in the world (Brown et al., 2017). To address this issue, continuous volcano monitoring is performed by volcano observatories feeding eruption forecasting and early warning systems. These systems need to provide a fast response in case of crisis to evaluate the hazards related to an imminent eruption,

playing a crucial role in the decision-making of an eventual evacuation. Current monitoring is mainly performed in two stages: 1) *detection* of Volcano-Seismic (VS) events in continuous data streams received from monitoring stations (Sparks et al., 2012) and 2) *classification* of events according to their spectral and temporal features. These tasks allow to track the seismic activity of some specific event classes considered as 1) *eruption precursors*, such as volcanic tremors and long-period events (Chouet, 1996) or volcano-tectonic earthquakes (White and McCausland, 2016) and, 2) *imminent-risk* classes as collapses, lahars and pyroclastic flows. While the detection of events can be accomplished by automatic systems (Álvarez et al., 2013; García et al., 2020), the classification is generally manually done by experts, encompassing an inherent high level of subjectivity and lack of reliability. Manual classification cannot be achieved fast enough during a major unrest and/or an eruptive episode because of the strong increase of seismic activity before an eruption (Orozco-Alzate et al., 2012). Hence, in order to efficiently detect and classify volcano-seismic events, there is a need of automatic Volcano-Seismic Recognition (VSR) systems (Malfante et al., 2018), able to operate in nearly real-time (McNutt et al., 2015; Boué et al., 2016). However, the deployment of VSR systems in observatories suffers from three major drawbacks:

(1) Designing costs: *supervised VSR* systems need to be trained using previously labeled catalogs of events gathered in a database (DB) to characterize the volcano-seismic classes. This so-called *training stage* requires considerable human resources and time (Langer et al., 2019), not always affordable for modest observatories. *Unsupervised VSR (U.VSR)* does not need this training step, but achieves a lower recognition accuracy being relegated mainly to data mining purposes.

(2) Lack of robustness: observed seismicity patterns and event types, thus, the seismic features and volcano-seismic catalogs on which supervised VSR is based, vary according to the current state of the volcano (Carniel, 2014), to the environmental noise (Lecocq et al., 2020) and to the type and location of the stations. This variability can decrease the efficiency of the VSR systems designed to model a fixed configuration of networks, classes and patterns.

(3) Poor usability and integrability: installing a VSR system in a monitoring platform requires expert staff. Complex VSR software implies the training of technicians, making the interoperability with standardized protocols and services difficult.
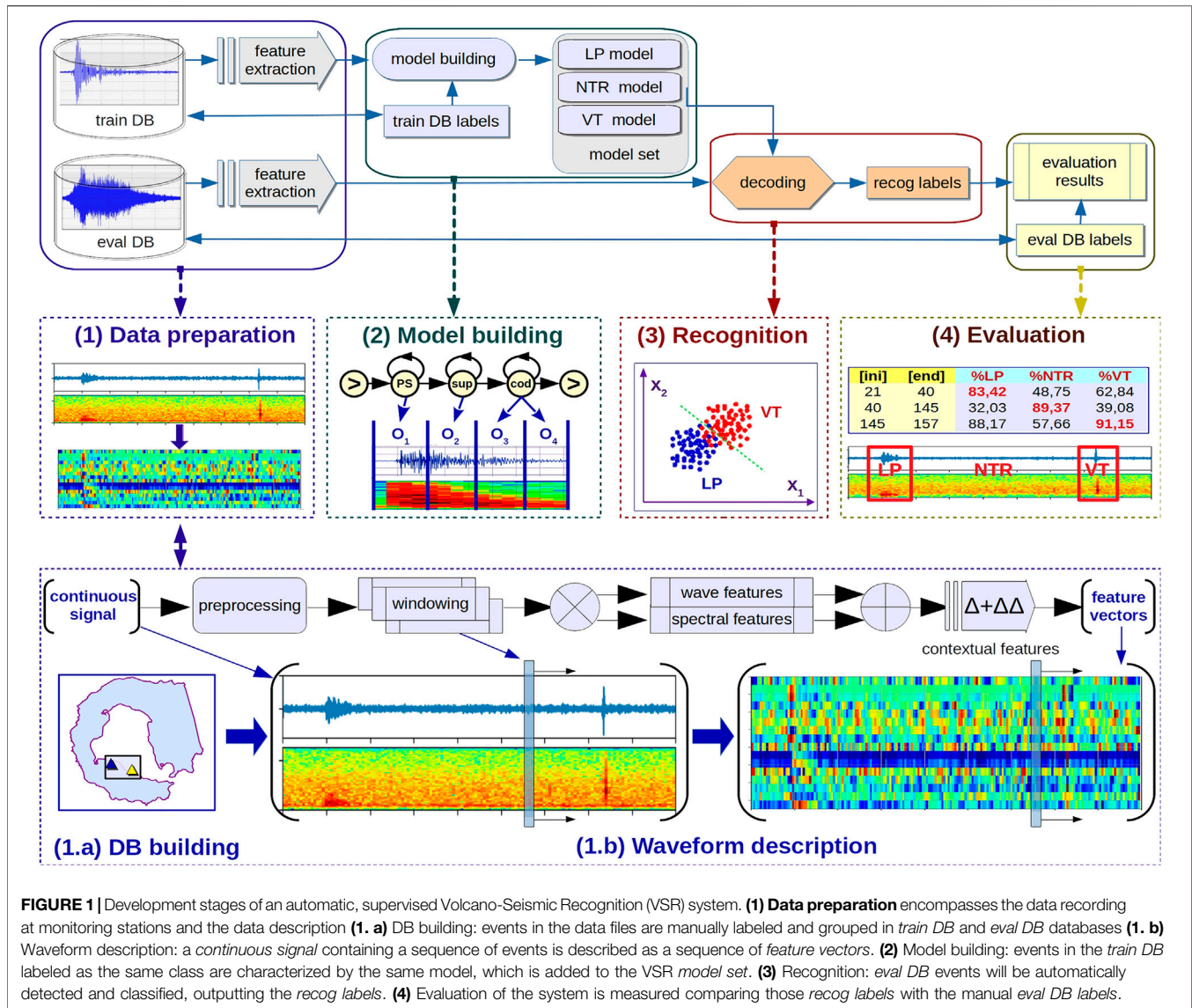
At present, only few volcano observatories have operational VSR systems running in real-time (Cortés et al., 2009b; Maggi et al., 2017). This work presents a *Volcano-Independent Seismic Recognition* (VI.VSR) approach as the solution to the above VSR issues. VI.VSR (Cortés et al., 2017; 2019a) represents a promising trade-off between classic VSR and U.VSR paradigms: it can recognize events from a local volcano without any previous information about it, achieving an acceptable efficiency without implementing the whole system from scratch. It relies on prebuilt, universal VSR models describing universal databases of labeled events recorded on several volcanoes. Hence, our proposal provides a *volcano-portable*, *operational* and *robust VSR platform*, developed under the EU-funded *VULCAN.ears* project, which includes tools to build local VSR models or alternatively, to use prebuilt VI.VSR universal ones. *VULCAN.ears* deployed VI.VSR systems in several observatories, partnering with 18 international institutions to create the *VSR-ALUE* collaboration group. As a result, data from 21 volcanoes have been gathered and currently *VSR-ALUE* continues the support and development of the volcano-independent approach and application to real-case scenarios.

2 VOLCANO-INDEPENDENT SEISMIC RECOGNITION (VI.VSR)

VSR field has intensively grown in the last two decades boosted by the evolution of Machine Learning (Bergen et al., 2019) and by the need of modern observatories of having reliable and robust VSR systems (Langer et al., 2019). A myriad of classifiers have been tested, being Artificial Neuronal Networks (Falsaperla et al., 1996; Langer et al., 2003) and Support Vector Machines (Masotti et al., 2006; Curilem et al., 2014) the most popular ones in early 2000s, followed by advanced Probabilistic Graphical Models (Ohrnberger, 2001; Benítez et al., 2007; Trujillo-Castrillón et al., 2018) popularized in 2007 and Deep Learning approaches since 2017 (Titos et al., 2018, 2019a; Bueno et al., 2019a). Starting with the simple task of classifying events already detected in the data flow in the so-called *isolated VSR*, three major breakthroughs have been deployed in the VSR area (Malfante et al., 2018):

(1) **Continuous VSR** is able to detect and classify volcano-seismic events appearing in a continuous data stream. It can be subdivided in:

   a. *One-step VSR* implements detection and classification in the same stage. Actually, only few recognizers can achieve this, mostly structured graphical models as Hidden Markov Models (HMMs). (Benítez et al., 2007; Beyreuther et al., 2008; Ibáñez et al., 2009).

   b. Two-step VSR requires an extra processing stage to isolate the events. Bayesian Networks (Riggelsen et al., 2007) and Recurrent Networks (Titos et al., 2019b) handle continuous input data but need additional algorithms to delimit events. Most isolated VSR systems can be converted into continuous adding some detection technique such as classic signal triggers or advanced phase picking methods (Álvarez et al., 2013; Bueno et al., 2019b; García et al., 2020), which segment a continuous data stream into a sequence of time-delimited events.

(2) Robust VSR gathers information from different stations monitoring the same volcano, even in different epochs (Cortés et al., 2019a). This yields robust systems than can recognize events in any station of the network, in noisy

FIGURE 1 | Development stages of an automatic, supervised Volcano-Seismic Recognition (VSR) system. **(1) Data preparation** encompasses the data recording at monitoring stations and the data description **(1. a)** DB building: events in the data files are manually labeled and grouped in *train DB* and *eval DB* databases **(1. b)** Waveform description: a *continuous signal* containing a sequence of events is described as a sequence of *feature vectors*. **(2)** Model building: events in the *train DB* labeled as the same class are characterized by the same model, which is added to the VSR *model set*. **(3)** Recognition: *eval DB* events will be automatically detected and classified, outputting the *recog labels*. **(4)** Evaluation of the system is measured comparing those *recog labels* with the manual *eval DB labels*.

conditions and with different types of seismic activity without a noticeable decrease in its efficiency (Maggi et al., 2017; Journeau et al., 2020).

(3) Unsupervised VSR (U.VSR) unlike the two approaches above, does not require labeled data neither modeling, saving resources. Self Organized Maps (Köhler et al., 2010; Carniel et al., 2013a, b) are the standard to cluster volcano-seismic patterns even though recent deep learning technologies (Cannavo' et al., 2020) are obtaining interesting insights related with precursors.

These VSR types have their own application scopes and are not directly comparable. Most classifiers overpass the 90% of recognition accuracy when processing isolated VSR (Cortés, 2015). However, real-time, *online VSR* applied on continuous data ranges a 80–90% of accuracy. U.VSR schemes hardly surpass 70%. Nevertheless, due to the cost required to design supervised systems and their drop of effectiveness when the

feature patterns of the events highly vary reflecting changes in the eruptive cycle, the current trend (Khan et al., 2019; Langer et al., 2019) is to use simpler U.VSR. Despite their low classification scores, U.VSR properly responds to the inherent variability of the seismic activity (Peltier et al., 2018). Conceived as a logical evolution of the robust VSR (Cortés et al., 2009a), the **hybrid VI.VSR** technology aims to be the future state-of-the-art joining supervised VSR scores within U.VSR goals to reach promising recognition results in the 70–80% band (Cortés et al., 2019a). In the following **Sections 2.1** and **2.2** we present our VSR baseline system and its improvements towards the *volcano-independent* VSR platform.

2.1 Automatic Volcano-Seismic Recognition (VSR)

A classic supervised VSR operation is divided in two stages shown in **Figure 1**: 1) the *training* step, including the data preparation,

description and classes characterization and, 2) the system *evaluation*, encompassing the automatic recognition of volcano-seismic events and measuring the performance. The system design is structured in:

(1) Data preparation consisting of:

    a. DB building: expert technicians manually detect and classify seismic events to prepare a labeled catalog which is split into the train DB and eval DB databases to train and evaluate the system, respectively. Their related train DB labels and eval DB labels detail the duration and type of the events appearing in each database.

    b. Waveform description consists of extracting relevant information from the data to be learnt by the system. To perform real-time VSR the continuous waveform is parametrized as a sequence of signal segments, each one described by a feature vector, resulting in a sequence $O = \{o_t\} = \{o_1, o_2, \ldots, o_t\}$ of observable vectors. An adequate scheme description of the data increases the robustness, exportability and recognition scores (Álvarez et al., 2012; Soto et al., 2018). Hybrid features describing waveform, geophysical and spectral information combined with their contextual, time derivatives components, provide an optimal scheme according to Cortés et al. (2016) and Maggi et al. (2017).

(2) Model building or learning phase to characterize the feature space projected by the $\{o_t\}$ feature vectors. Discriminative classifiers as neuronal networks and most deep learning structures delimit the space in clusters assigning each one to a volcano-seismic class of the train DB. Generative classifiers as graphical models independently model each class c estimating its joint probability $P(c, \{o_t\})$ to quantify the relationship between a vector and a class.

(3) Recognition of the events existing in each data file of the eval DB. Given a waveform file described as a sequence $\{o_v\}$ of V feature vectors, the recognition algorithm will uncover its corresponding sequence $\{c_r\}$ of R detected and classified events, mapping $\{o_v\} \rightarrow \{c_r\}$. The type and temporal limits of the recognized R events are outputted in the automatically generated recog labels catalog. In isolated VSR $R = 1$, thus, only substitution errors when an event is wrongly cataloged can be committed. Normally, in continuous VSR $R > 1$, hence, events not previously tagged by experts in the eval DB labels can be mistakenly detected (inserted) by the system which also can not detect (delete) other events actually registered.

(4) Evaluation measuring the similarity between the recog labels outputted by the system and the eval DB labels manually tagged. Precision, recall and F-score measures are common in machine learning literature but more natural comparisons are done counting the event insertion (I), deletion (D) and substitution (S) errors defined by the accuracy (%Acc) score:

$$\%\text{Acc} = \text{mean}_l \{\%\text{Acc}(c_l)\} = \text{mean}_l \left\{ 100 \left[1 - \frac{E(c_l)}{N(c_l)} \right] \right\}, \quad (1)$$
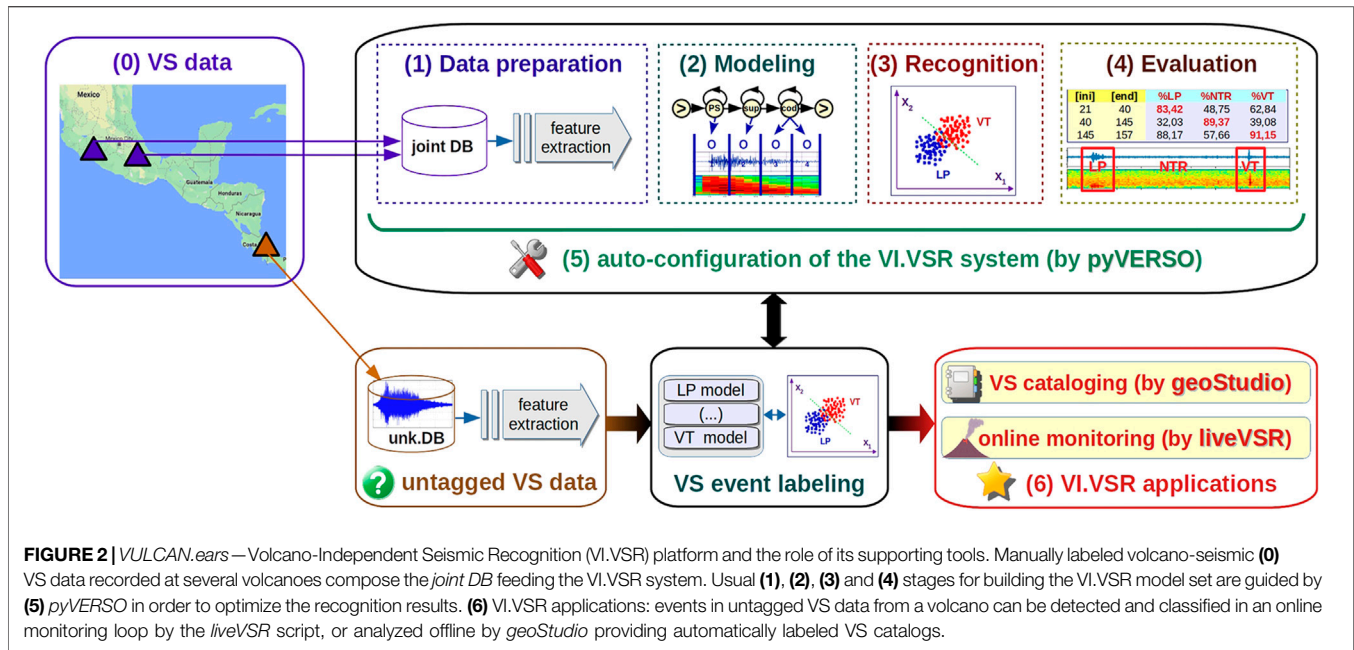
with $N(c_l)$ the number of class c_l events in the *eval DB*, $E(c_l)$ the recognition errors and l an integer ranging from one to L, being L the number of evaluation classes. In continuous VSR $E = D + S + I$ while in isolated VSR $D = I = 0$, which explains its higher scores. The average $\overline{R}$ number of events in the evaluation data files has a large impact when comparing isolated vs. continuous tasks, as the accuracy exponentially decreases by the $\overline{R}$ factor.

2.2 Proposed Volcano-independent Seismic Recognition (VI.VSR) Framework

2.2.1 VI.VSR Underlying Technology

Given the similarities between speech and seismic events (Ohrnberger, 2001), former VI.VSR was inspired by the classic speech recognition area which successfully accomplishes speaker-independent tasks gathering multi-speaker databases modeled with HMMs (Rabiner and Schafer, 2007). HMMs characterize structured patterns, as volcano-seismic events, modeling not only the pattern waveforms, but also the relationship among patterns given by the temporal distribution of the HMM states. Each state represents a pattern observed in the $\{o_t\}$ sequence of feature vectors, such as *P*, S and superficial phase arrivals (**Figure 1**). HMMs are suitable for one-step continuous VSR in real-time, outperforming classic seismic detection algorithms even in noisy scenarios (Beyreuther and Wassermann, 2008). For boosting the system robustness and portability, VI.VSR has pushed beyond the state-of-the-art these innovative concepts:

- Universal databases: gathering data from different types of volcanoes increases the number and variety of patterns found in events of the same seismic class, improving the completeness and robustness of the given class model (Cortés et al., 2009a).
- Standardization of the seismic waveform based on unsupervised, data-driven decomposition and posterior selective reconstruction of a signal. Standardized events are less noisy and, hence, easier to recognize (Cortés et al., 2019b).
- Efficient data description: the extended Discriminative Feature Selection algorithm extracts the most relevant information of a seismogram when selecting the most efficient components of a feature vector to describe it (Álvarez et al., 2012; Cortés et al., 2016). In each iterative step, the worst feature according to a loss function is removed from the original vector, keeping only the most valuable components. This encompasses a better description of the seismic classes, simplifying their models and enhancing the system portability.
- Dedicated parallel VSR channels for each class: they are complete VSR systems specialized and customized in the detection and classification of just one type of seismic events (Cortés et al., 2014). A system unifying the output of these independent, class-focused, recognition channels surpasses the classic serial architecture depicted in **Figure 1** whereas all the classes share the same system configuration (Cortés et al., 2016).

FIGURE 2 | *VULCAN.ears* — Volcano-Independent Seismic Recognition (VI.VSR) platform and the role of its supporting tools. Manually labeled volcano-seismic **(0)** VS data recorded at several volcanoes compose the *joint DB* feeding the VI.VSR system. Usual **(1)**, **(2)**, **(3)** and **(4)** stages for building the VI.VSR model set are guided by **(5)** *pyVERSO* in order to optimize the recognition results. **(6)** VI.VSR applications: events in untagged VS data from a volcano can be detected and classified in an online monitoring loop by the *liveVSR* script, or analyzed offline by *geoStudio* providing automatically labeled VS catalogs.

Figure 2 depicts the structure of the developed VI.VSR platform and its utilization steps. Basically, it encloses the usual (1)-(4) VSR stages detailed in **Section 2.1** to deploy an improved VSR system but fed with labeled events from several volcanoes composing the universal *joint DB*. An extra 5) *auto-configuration* stage optimizes the system to maximize its efficiency by performing iterative train-evaluation tests to select the best data description scheme and modeling setups. The (1)–(5) stages are guided by suitable *pyVERSO* scripts accomplished to obtain robust VI.VSR models. The automatic recognition of events embedded in the seismic records of the *untagged VS data* can be carried out in two different manners:

- Offline cataloging via geoStudio: data stored in files are loaded into a graphical interface and their events are recognized by VSR models selected from a prebuilt set. Then, the tagged data can be plotted for inspection and the resulting catalog is stored for further analysis.
- Online monitoring via *liveVSR* to perform a continuous, real-time monitoring of an active volcano. The *liveVSR* script is able to connect to any available FDSN data server plotting the recognized events and generating volcano-seismic catalogs. Several instances of *liveVSR* can be run concurrently receiving data from several stations or volcanoes.

The resulting labeled catalogs are the input to posterior 6) *VI.VSR applications*.

2.2.2 Building VI.VSR Models With *pyVERSO*
pyVERSO is a collection of Python scripts designed to perform VSR tasks from the command line. It has libraries to prepare and describe data, including many time-domain, cepstral and hybrid parametrization schemes and advanced feature selection routines.

It is highly configurable and easy to use. Taking as input a labeled database and a configuration file it can implement serial or parallel architectures based on HMMs, Gaussian Mixture Models and Conditional Random Fields. *pyVERSO* main aim is to build own, local VSR systems of an active volcano. Once the system is optimized, its models can be exported to be used on online monitoring via *liveVSR* or offline analysis by *geoStudio*. *pyVERSO* is highly integrated within the Python scientific ecosystem and, currently, relies on the *HMM Toolkit* (Young et al., 2006) when using HMMs.
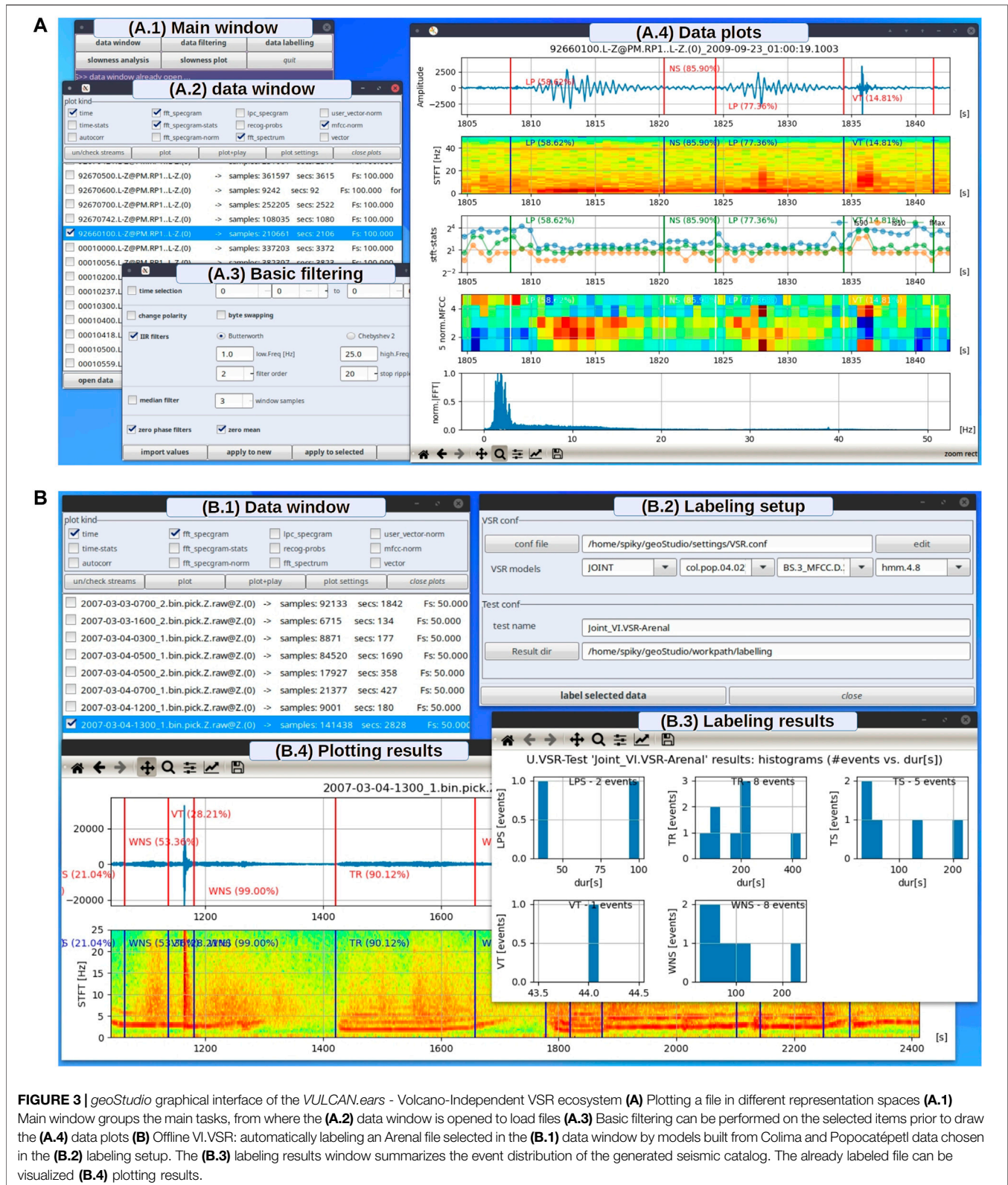
2.2.3 Graphical VI.VSR With *geoStudio*
geoStudio is the graphical frontend of *pyVERSO* developed to simplify seismic analysis (Carmona et al., 2014) and recognition tasks (**Figure 3**). It provides the following complementary functionalities:

- Loading and saving of data supporting several seismological formats. Also it can handle *NumPy* arrays and *HMM Toolkit* encoded files (**Figure 3A 2**).
- Data filtering (**Figure 3A 3**) and advanced seismogram visualization of any custom description scheme defined by *pyVERSO* (**Figure 3A 4**).
- 2D seismic source location by slowness maps via zero-lag cross-correlation (Almendros et al., 1999).

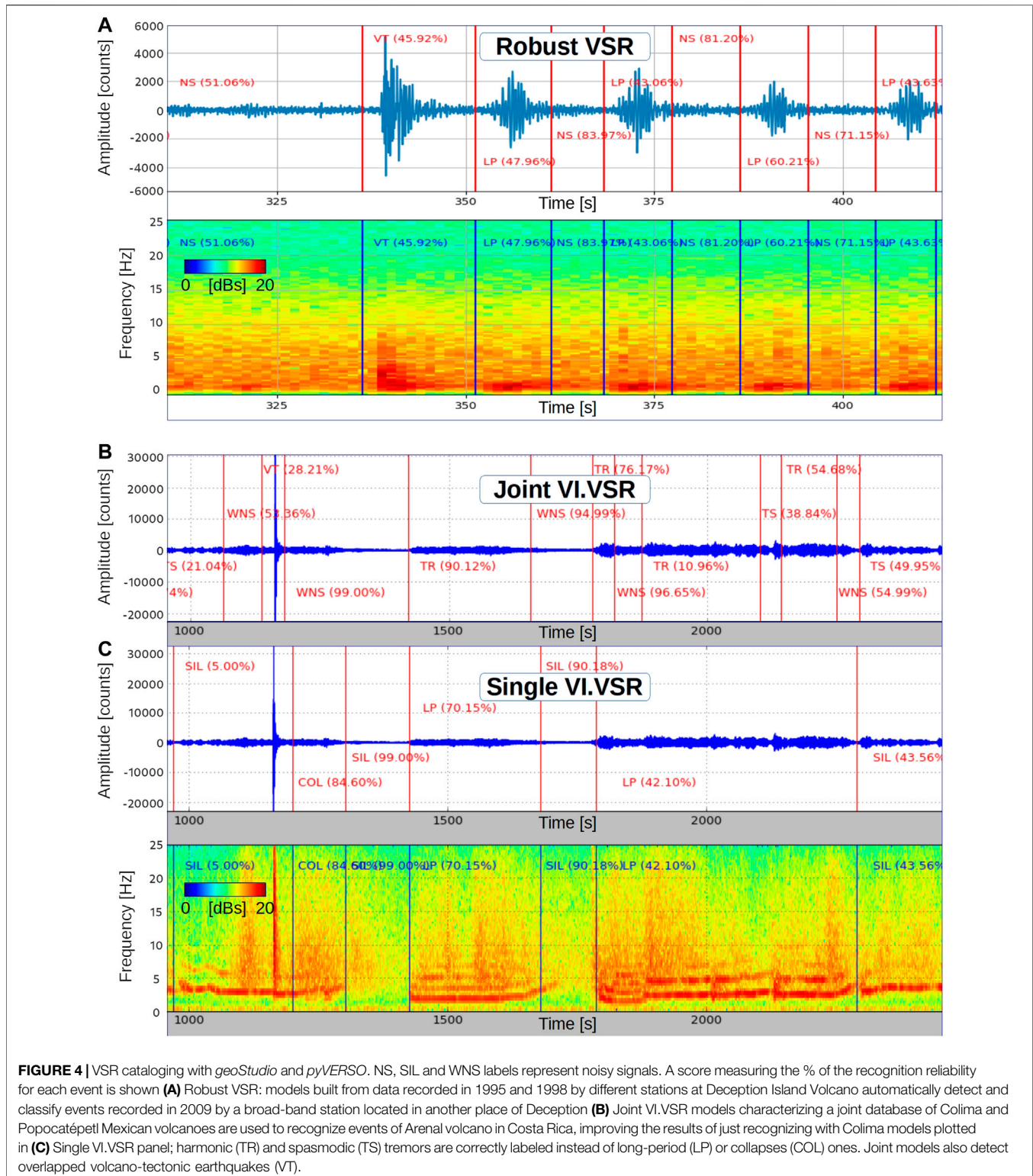A Volcano-Independent VSR test in *geoStudio* can be easily carried out as follows (**Figure 3B**):

(1) *Load data files* (**Figure 3B 1**). Several formats are directly supported, including those readable by *ObsPy*.
(2) *Select VSR models* suitable to your data among a built-in set of models trained with 10 databases from 10 different volcanoes respectively, six joint databases mixing two

FIGURE 3 | *geoStudio* graphical interface of the *VULCAN.ears* - Volcano-Independent VSR ecosystem **(A)** Plotting a file in different representation spaces **(A.1)** Main window groups the main tasks, from where the **(A.2)** data window is opened to load files **(A.3)** Basic filtering can be performed on the selected items prior to draw the **(A.4)** data plots **(B)** Offline VI.VSR: automatically labeling an Arenal file selected in the **(B.1)** data window by models built from Colima and Popocatépetl data chosen in the **(B.2)** labeling setup. The **(B.3)** labeling results window summarizes the event distribution of the generated seismic catalog. The already labeled file can be visualized **(B.4)** plotting results.

volcanoes and two universal volcano-independent solutions with data of several volcanoes (**Figure 3B 2**). Custom models created by *pyVERSO* can also be selected.

(3) *Launch the labeling task* and *examine the results* plotting the event distribution of the generated catalog (**Figure 3B 3**) and the files automatically labeled (**Figure 3B 4**).

FIGURE 4 | VSR cataloging with *geoStudio* and *pyVERSO*. NS, SIL and WNS labels represent noisy signals. A score measuring the % of the recognition reliability for each event is shown **(A)** Robust VSR: models built from data recorded in 1995 and 1998 by different stations at Deception Island Volcano automatically detect and classify events recorded in 2009 by a broad-band station located in another place of Deception **(B)** Joint VI.VSR models characterizing a joint database of Colima and Popocatépetl Mexican volcanoes are used to recognize events of Arenal volcano in Costa Rica, improving the results of just recognizing with Colima models plotted in **(C)** Single VI.VSR panel; harmonic (TR) and spasmodic (TS) tremors are correctly labeled instead of long-period (LP) or collapses (COL) ones. Joint models also detect overlapped volcano-tectonic earthquakes (VT).

3 CASE STUDIES AND RESULTS

This section presents some baseline results obtained with *pyVERSO* and *geoStudio* as application of the VI.VSR innovations detailed in

Section 2.2.1. We start with robust VSR examples evolving to volcano-independent cases, evaluated in typical monitoring scenarios as a volcanic island with noisy recordings or highly active stratovolcanoes with recent eruptive episodes. Demo videos

of geoStudio running these case studies are supplied as **Supplementary Material**.

3.1 Robust VSR at Deception Island

Deception Island Volcano has been a relevant VSR laboratory to evaluate new algorithms since 2006 (Benítez et al., 2007, 2009; Cortés et al., 2014; Titos et al., 2019b). Robust VSR experiments have been deployed involving noisy scenarios and data from different stations installed at different locations and years as short-period clean recordings in 1995, short-period noisy signals acquired in 1998 close to a hydrothermal area and broad-band noisy data from a station in 2009 (Carmona et al., 2014):

- A Multi-station VSR system was trained with 4,011 events from the 1995 and 1998 stations. It was able to automatically catalog long period (LP) events, volcano-tectonic (VT) and hybrid earthquakes, noisy (NS) and tremor segments in the continuous recordings of the whole 2009 years. The test only took 20 h of a modest 2.5 GHz, 2-threads computer (Cortés et al., 2017). **Figure 4A** shows how LP swarms are correctly recognized even in noisy conditions.
- Parallel vs. serial (classic) architecture in continuous VSR: pyVERSO auto-configuration improved the precision from 52% up to 72% of a classic VSR system by using dedicated parallel recognition channels. 40 LPs, 113 VTs and 292 noise-tremor events of the 1998 station were automatically labeled by models trained with 58 LPs, 41 VTs and 189 noise-tremor signals manually cataloged in 1995 (Cortés et al., 2019a).
- Waveform standardization increased from 66–76% the precision in continuous VSR recognizing 204 LPs, 467 noise-tremors and 36 VTs in the noisy 2009 recordings (Cortés et al., 2019b). The models were built describing 58 LPs, 189 noise-tremors and 41 VTs gathered in 1995.

3.2 Volcano-independent VSR at Colima, Popocatépetl and Arenal Volcanoes

The proposed volcano-independent recognition approach has been tested in some of the most active American volcanoes thanks to the collaboration of the *VULCAN.ears* partners. Signals of the andesitic Colima, Popocatépetl and Arenal stratovolcanoes monitored in 2002, 2004, and 2007, respectively, were labeled to evaluate our system (Cortés et al., 2009b).

37 hours of continuous data recorded by a short-period station at Popocatépetl were analyzed by models trained with 17 h of labeled events acquired at Colima by broad-band and short-period sensors. 282 LPs, 184 VTs, 50 regional tectonic earthquakes (REG), 164 noisy segments, 75 harmonic (TR) and 59 spasmodic (TS) tremors were detected and classified with an efficiency of 59%, raised up to 65% after auto-configuring the most convenient waveform description and standardization schemes. 52 h of Mexican Colima and Popocatépetl events were combined to deploy a volcano-independent system for classifying Arenal events detected by a

broad-band sensor in Costa Rica. It was able to discriminate 46 TSs, 46 TRs, 53 pulsant tremors, 46 explosions, 26 REGs, 50 VTs and 285 noises, increasing the precision from 50% to 71% after auto-selecting the best hybrid features to describe the seismograms (Cortés et al., 2019b). Precisely, **Figure 4** presents an example of the accuracy improvements when the training database is enlarged with events of a new volcano: *Joint VI.VSR* models (**Figure 4B**) trained with Colima and Popocatépetl databases perform better recognizing Arenal events than *Single VI.VSR* models (**Figure 4C**) built only with Colima.

4 DISCUSSION AND CONCLUSION

4.1 Discussion

Several alternatives can be found to facilitate the integration and use of VSR systems at volcano observatories but most of them are designed for a determined volcano or focused on pattern discovery (Carniel et al., 2006; Langer et al., 2009; Messina and Langer, 2011). Some graphical programs are useful to characterize events guiding the manual labelling of data (Lesage, 2009). Bueno et al. (2020) designed an interesting interface based on two-step VSR recognition able to cluster events in basic categories. *VULCAN.ears* framework offers a complete volcano-independent VSR solution to monitor volcanoes providing built-in models for *geoStudio* and *liveVSR* alongside the *pyVERSO* ecosystem to build customized VSR systems focused on low-level integration. The proposed approach is an *ongoing project* and the current universal models are only trained with events from three volcanoes. A universal database gathering events of, at least, 10 volcanoes is in preparation in strict cooperation with our partners including catalogs from Copahue, Etna, Flegrei, Merapi, Piton de la Fournaise, San Miguel, Soufrière of Guadeloupe, Stromboli, Turrialba, Ubinas and others publicly available as Llaima and Cotopaxi. *geoStudio* only wraps few functionalities of its *pyVERSO* backend, having a huge potential to be accomplished.

The innovative techniques specified in **Section 2.2.1** and designed to achieve the volcano-independent milestone are functional and their baseline results encouraging, scoring an efficiency in the 70–80% interval in challenging scenarios as continuous recognition under noisy conditions in different stations and epochs, achieving a 76%, or volcano-independent classification and recognition reaching 71 and 65%, respectively. These are promising outcomes, specially, recognizing more than five classes in continuous records from different volcanoes. In addition, experts rarely agree on an 80% when labeling the same data and classic evaluation metrics of most VSR literature do not properly account insertion and deletion errors, providing overrated values compared to the defined in **Equation (1)**. The building of different types of universal recognition models is an attractive option to raise the efficiency, i.e., models of open vent volcanoes vs. closed vent ones, or specific universal models for island volcanoes with oceanic noises. There are still open issues to solve, as the strong influence that the quality of the manual labeling and the data description scheme have on the VSR

scores. In any case, having a robust volcano-independent solution allows to recognize events which have not previously appeared in the volcano recordings. This provides a valuable input to early warning systems monitoring dormant volcanoes and to properly characterize the current volcano state in its eruptive cycle.

4.2 Concluding Remarks and Future Work

This work presents the *Volcano-Independent Seismic Recognition* as a solution to classic issues when implementing automatic *Volcano-Seismic Recognition* systems in volcano observatories. Current monitoring centers usually have limited resources to develop their own systems. They still detect and classify manually, which restrains their response in case of volcanic unrest. The authors have deployed a platform to develop portable recognition systems providing several tools to easily integrate and use the framework in observatories and to build applications for cataloging volcano-seismic events: *pyVERSO* to design recognition systems adapted to a given volcano, *geoStudio* to graphically detect and classify events in offline interactive operations, and *liveVSR* to continuously recognize in real-time events from remote or local data servers. Even though these programs are still in development, their application examples and baseline results point out the proposed approach as an exciting breakthrough in the volcano monitoring area.

Next efforts will be directed to increase the number of prebuilt volcano-independent models for enhancing the system robustness and to extend *geoStudio* capabilities with an interface to manually label data and with a guided-process to deploy customized recognition systems.

DATA AVAILABILITY STATEMENT

Following the recommendations of the European Commission regarding Open Science in EU-funded programmes, several material related with VULCAN.ears project are available, free of charge, in the Zenodo repository, including conference talks and posters, datasets and software. Specifically, the software preview versions and manuals of geoStudio, liveVSR and pyVERSO documented with demo videos of geoStudio running the case studies presented in **Section 3**. VULCAN.ears main webpage can be found at ResearchGate.

REFERENCES

Almendros, J., Ibáñez, J. M., Alguacil, G., and Del Pezzo, E. (1999). Array analysis using circular-wave-front geometry: an application to locate the nearby seismo-volcanic source. *Geophys. J. Int.* 136, 159–170. doi:10.1046/j.1365-246X.1999.00699.x

Álvarez, I., García, L., Cortés, G., Benítez, C., and De la Torre, Á. (2012). Discriminative feature selection for automatic classification of volcano-seismic signals. *Geosci. Rem. Sens. Lett. IEEE* 9, 151–155. doi:10.1109/LGRS.2011.2162815

Álvarez, I., Garcia, L., Mota, S., Cortés, G., Benitez, C., and De la Torre, A. (2013). An automatic P-phase picking algorithm based on adaptive multiband processing. *Geosci. Rem. Sens. Lett. IEEE* 10, 1488–1492. doi:10.1109/LGRS.2013.2260720

AUTHOR CONTRIBUTIONS

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

Benítez, C., Ibáñez, J., García, L., Cortés, G., and Álvarez, I. (2009). "Analysis of volcanic seismicity at Deception Island, Stromboli volcano and Mt. Etna using an automatic CHMM-based recognition method," in *Volume project: VOLcanoes, understanding subsurface mass MoveMEnt, eds. C. B. A. Bean and E. C. P. 6th framework (school of geological sciences, university college dublin*: C. J. Bean, A. K. Braiden, I. Lokmer, F. Martini, and G. S. O'Brien, 140–149.

Benítez, C., Ramírez, J., Segura, J. C., Ibáñez, J., Almendros, J., García-Yeguas, A., et al. (2007). Continuous HMM-based seismic-event classification at deception island, Antarctica, *IEEE Transactions on Geoscience and remote sensing*. NY, United States: IEEE, 45, 138–146. doi:10.1109/TGRS.2006.882264

Bergen, K. J., Johnson, P. A., de Hoop, M. V., and Beroza, G. C. (2019). Machine learning for data-driven discovery in solid earth geoscience. *Science* 363, eaau0323. doi:10.1126/science.aau0323

Beyreuther, M., Carniel, R., and Wassermann, J. (2008). Continuous hidden Markov models: application to automatic earthquake detection and classification at las cañadas caldera. *Tenerife. J. Volcanol. Geoth. Res.* 176, 513–518. doi:10.1016/j.jvolgeores.2008.04.021

Beyreuther, M., and Wassermann, J. (2008). Continuous earthquake detection and classification using discrete Hidden Markov Models. *Geophys. J. Int.* 175, 1055–1066. doi:10.1111/j.1365-246X.2008.03921.x

Boué, A., Lesage, P., Cortés, G., Valette, B., Reyes-Dávila, G., Arámbula-Mendoza, R., et al. (2016). Performance of the "material Failure Forecast Method" in real-time situations: a Bayesian approach applied on effusive and explosive eruptions. *J. Volcanol. Geoth. Res.* 327, 622–633. doi:10.1016/j.jvolgeores.2016.10.002

Brown, S. K., Jenkins, S. F., Sparks, R. S. J., Odbert, H., and Auker, M. R. (2017). Volcanic fatalities database: analysis of volcanic threat with distance and victim classification. *J. of Appl. Volcan.* 6, 1–20. doi:10.1186/s13617-017-0067-4

Bueno, A., Benítez, C., De Angelis, S., Díaz Moreno, A., and Ibáñez, J. M. (2019a). volcano-seismic transfer learning and uncertainty quantification with bayesian neural networks. *IEEE Trans. Geosci. Rem. Sens.*, 1–11. doi:10.1109/TGRS.2019.2941494

Bueno, A., Zuccarello, L., Díaz-Moreno, A., Woollam, J., Titos, M., Benitez, C., et al. (2020). Picoss: Python interface for the classification of seismic signals. *Comput. Geosci.* 142, 104531. doi:10.1016/j.cageo.2020.104531

Bueno, Á., Díaz-Moreno, A., De Angelis, S., Benítez, C., and Ibáñez, J. M. (2019b). Recursive entropy method of segmentation for seismic signals. *Seismol Res. Lett.* 90, 1670–1677. doi:10.1785/0220180317

Cannavo', F., Cannata, A., Palazzo, S., Spampinato, C., Faraci, D., Castagnolo, G., et al. (2020). Unsupervised deep learning on seismic data to detect volcanic unrest, *EGU general assembly conference abstracts.* Vienna, Austria: 22nd EGU General Assembly, 18631. doi:10.5194/egusphere-egu2020-18631

Carmona, E., Almendros, J., Martín, R., Cortés, G., Alguacil, G., Moreno, J., et al. (2014). Advances in seismic monitoring at deception island Volcano (Antarctica) since the international polar year. *Ann. Geophys.* 57. doi:10.4401/ag-6378

Carniel, R., Barbui, L., and Jolly, A. (2013a). Detecting dynamical regimes by self-organizing map (som) analysis: an example from the march 2006 phreatic eruption at raoul island, New Zealand kermadec arc. *Boll. Geofis. Teor. Appl.* 54, 39–52. doi:10.4430/bgta0077

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

Carniel, R., Di Cecca, M., and Jaquet, O. (2006). A user-friendly, dynamic web environment for remote data browsing and analysis of multiparametric geophysical data within the MULTIMO project. *J. Volcanol. Geoth. Res.* 153, 80–96. doi:10.1016/j.jvolgeores.2005.08.005

Carniel, R., Jolly, A. D., and Barbui, L. (2013b). Analysis of phreatic events at Ruapehu volcano, New Zealand using a new SOM approach. *J. Volcanol. Geoth. Res.* 254, 69–79. doi:10.1016/j.jvolgeores.2012.12.026

Chouet, B. A. (1996). Long-period volcano seismicity: its source and use in eruption forecasting. *Nature* 380, 309–316. doi:10.1038/380309a0

Cortés, G., Arambula, R., Gutiérrez, L., Benítez, C., Ibáñez, J., Lesage, P., et al. (2009a). Evaluating robustness of a HMM-based classification system of volcano-seismic events at Colima and Popocatépetl volcanoes, *IEEE International, IGARSS 2009 Geoscience and remote sensing symposium, 2009.* (NY, United States: IEEE), 2, 1012–1015. doi:10.1109/Igarss.2009.5418275

Cortés, G., Arámbula, R., Álvarez, I., Benítez, C., Ibáñez, J., Lesage, P., et al. (2009b). "Analysis of Colima, Popocatépetl and arenal volcanic seismicity using an automatic continuous hidden Markov models-based recognition system," in *VOLUME project: VOLcanoes, understanding subsurface mass MoveMEnt. 6th framework (Dublin: School of Geological Sciences, University College Dublin)*: eds. C.J. Bean, A.K. Braiden, I. Lokmer, C. B. A. E. C. P. Bean, F. Martini, and G. S. O'Brien, 150–160.

Cortés, G., García, L., Álvarez, I., Benítez, C., de la Torre, Á., and Ibáñez, J. (2014). Parallel System Architecture (PSA): an efficient approach for automatic recognition of volcano-seismic events. *J. Volcanol. Geoth. Res.* 271, 1–10. doi:10.1016/j.jvolgeores.2013.07.004

Cortés, G. (2015). Reconocimiento de señales sismo-volcánicas mediante canales específicos basados en modelos ocultos de Markov. PhD. thesis. Granada (Spain): Universidad de Granada.

Cortés, G., Benítez, M. C., García, L., Álvarez, I., and Ibáñez, J. M. (2016). A comparative study of dimensionality reduction algorithms applied to volcano-seismic signals. *IEEE J. Sel. Top. Appl.* 9, 253–263. doi:10.1109/JSTARS.2015.2479300

Cortés, G., Lesage, P., Carniel, R., Benítez, C., Almendros, J., and Arámbula-Mendoza, R. (2017). Unsupervised volcano-seismic event recognition as a tool for real-time monitoring and eruption forecasting: the VULCAN.ears project. Portland, OR: Scientific assembly, IAVCEI, 224.

Cortés, G., Carniel, R., Lesage, P., Ángeles Mendoza, M., and Della Lucia, I. (2019a). Volcano-Independent Seismic Recognition: detecting and classifying events of a given volcano using data from others, EGU general assembly conference abstracts. Vienna, Austria, April, 2019, 11386. doi:10.1002/essoar.10500900.1

Cortés, G., Carniel, R., Mendoza, M., and Lesage, P. (2019b). Standardization of noisy volcanoseismic waveforms as a key step toward station- independent, robust automatic recognition. *Seismol Res. Lett.* 90, 581–590. doi:10.1785/0220180334

Curilem, M., Vergara, J., San Martin, C., Fuentealba, G., Cardona, C., Huenupan, F., et al. (2014). Pattern recognition applied to seismic signals of the Llaima volcano (Chile): an analysis of the events' features. *J. Volcanol. Geoth. Res.* 282, 134–147. doi:10.1016/j.jvolgeores.2014.06.004

Falsaperla, S., Graziani, S., Nunnari, G., and Spampinato, S. (1996). Automatic classification of volcanic earthquakes by using multi-layered neural networks. *Nat. Hazards* 13, 205–228. doi:10.1007/BF00215816

García, L., Alguacil, G., Titos, M., Cocina, O., De la Torre, Á., and Benítez, C. (2020). Automatic S-phase picking for volcano-tectonic earthquakes using spectral dissimilarity analysis. *Geosci. Rem. Sens. Lett. IEEE* 17, 874–878. doi:10.1109/LGRS.2019.2934220

Ibáñez, J., Benítez, C., Gutiérrez, L., Cortés, G., García-Yeguas, A., and Alguacil, G. (2009). The classification of seismo-volcanic signals using Hidden Markov Models as applied to the Stromboli and Etna volcanoes. *J. Volcanol. Geoth. Res.* 187, 218–226. doi:10.1016/j.jvolgeores.2009.09.002

Journeau, C., Shapiro, N. M., Seydoux, L., Soubestre, J., Ferrazzini, V., and Peltier, A. (2020). Detection, classification, and location of seismovolcanic signals with multi-component seismic data, example from the Piton de la Fournaise volcano (La Réunion, France). *J. Geophys. Res.: Solid Earth* 125, e2019JB019333. doi:10.1029/2019JB019333

Khan, M. S., Curilem, M., Huenupan, F., Khan, M. F., and Becerra Yoma, N. (2019). A signal processing perspective of monitoring active volcanoes [applications corner]. *IEEE Signal Process. Mag.* 36, 125–163. doi:10.1109/MSP.2019.2930427

Köhler, A., Ohrnberger, M., and Scherbaum, F. (2010). Unsupervised pattern recognition in continuous seismic wavefield records using Self-Organizing Maps. *Geophys. J. Int.* 182, 1619–1630. doi:10.1111/j.1365-246X.2010.04709.x

Langer, H., Falsaperla, S., and Hammer, C. (2019). *Advantages and pitfalls of pattern recognition: selected cases in geophysics.* 1st Edn. Amsterdam, Netherlands: ISSN Elsevier Science. doi:10.1016/c2016-0-00254-1

Langer, H., Falsaperla, S., Masotti, M., Campanini, R., Spampinato, S., and Messina, A. (2009). Synopsis of supervised and unsupervised pattern classification techniques applied to volcanic tremor data at Mt Etna. *Italy. Geophys. J. Int.* 178, 1132–1144. doi:10.1111/j.1365-246X.2009.04179.x

Langer, H., Falsaperla, S., and Thompson, G. (2003). Application of artificial neural networks for the classification of the seismic transients at Soufrière hills volcano, Montserrat. *Geophys. Res. Lett.* 30, 2090. doi:10.1029/2003GL018082

Lecocq, T., Hicks, S. P., Van Noten, K., van Wijk, K., Koelemeijer, P., De Plaen, R. S. M., et al. (2020). Global quieting of high-frequency seismic noise due to covid-19 pandemic lockdown measures. *Science* 369 (6509), 1338–1343. doi:10.1126/science.abd2438

Lesage, P. (2009). Interactive Matlab software for the analysis of seismic volcanic signals. *Comput. Geosci.* 35, 2137–2144. doi:10.1016/j.cageo.2009.01.010

Maggi, A., Ferrazzini, V., Hibert, C., Beauducel, F., Boissier, P., and Amemoutou, A. (2017). Implementation of a multistation approach for automated event classification at Piton de la Fournaise volcano. *Seismol. Res. Lett.* 88, 878–891. doi:10.1785/0220160189

Malfante, M., Dalla Mura, M., Métaxian, J.-P., Mars, J. I., Macedo, O., and Inza, A. (2018). Machine learning for volcano-seismic signals: challenges and perspectives. *IEEE Signal Process. Mag.* 35, 20–30. doi:10.1109/MSP.2017.2779166

Masotti, M., Falsaperla, S., Langer, H., Spampinato, S., and Campanini, R. (2006). Application of support vector machine to the classification of volcanic tremor at Etna. *Italy. Geophys. Res. Lett.* 33, L20304. doi:10.1029/2006GL027441

McNutt, S. R., Thompson, G., Johnson, J., Angelis, S. D., and Fee, D. (2015). "Chapter 63—seismic and infrasonic monitoring," in *The encyclopedia of volcanoes*. Editor H. Sigurdsson 2nd edn (Amsterdam, Netherlands: Academic Press), 1071–1099. doi:10.1016/B978-0-12-385938-9.00063-8

Messina, A., and Langer, H. (2011). Pattern recognition of volcanic tremor data on Mt. Etna (Italy) with KKAnalysis—A software program for unsupervised classification. *Comput. Geosci.* 37, 953–961. doi:10.1016/j.cageo.2011.03.015

Ohrnberger, M. (2001). *Continuous automatic classification of seismic signals of volcanic origin at Mt. Merapi, Java, Indonesia.* PhD. thesis. Potsdam, Germany: Universität Potsdam

Orozco-Alzate, M., Acosta-Muñoz, C., and Londoño-Bonilla, J. M. (2012). "The automated identification of volcanic earthquakes: concepts, applications and challenges," in *Earthquake research and analysis*. Editors S. D'Amico and Rijeka (Rijeka, Croatia: IntechOpen), Chap. 19. 345–370. doi:10.5772/27508

Peltier, A., Villeneuve, N., Ferrazzini, V., Testud, S., Hassen Ali, T., Boissier, P., et al. (2018). Changes in the long-term geophysical eruptive precursors at Piton de la Fournaise: implications for the response management. *Front. Earth Sci.* 6, 104. doi:10.3389/feart.2018.00104

Rabiner, L. R., and Schafer, R. W. (2007). Introduction to digital speech processing. *Found. Trends Signal Process* 1, 1–194. doi:10.1561/2000000001

Riggelsen, C., Ohrnberger, M., and Scherbaum, F. (2007). Dynamic bayesian networks for real-time classification of seismic signals, *Knowledge discovery in databases: PKDD 2007*. Berlin, Germany: Springer, 4702, 565–572. doi:10.1007/2F978-3-540-74976-9_59

Soto, R., Huenupan, F., Meza, P., Curilem, M., and Franco, L. (2018). Spectro-temporal features applied to the automatic classification of volcanic seismic events. *J. Volcanol. Geoth. Res.* 358, 194–206. doi:10.1016/j.jvolgeores.2018.04.025

Sparks, R. S., Biggs, J., and Neuberg, J. W. (2012). Geophysics. Monitoring volcanoes. *Science* 335, 1310–1311. doi:10.1126/science.1219485

Titos, M., Bueno, A., García, L., Benítez, C., and Segura, J. C. (2019a). Classification of isolated volcano-seismic events based on inductive transfer learning. *IEEE Geoscience and Remote Sensing Letters*, NY, United States: IEEE, 17, 869-873. doi:10.1109/LGRS.2019.2931063

Titos, M., Bueno, A., García, L., and Benítez, C. (2018). A deep neural networks approach to automatic recognition systems for volcano-seismic events. *IEEE J. Sel. Top. Appl.* 11, 1533–1544. doi:10.1109/JSTARS.2018.2803198

Titos, M., Bueno, A., García, L., Benítez, M. C., and Ibáñez, J. (2019b). Detection and classification of continuous volcano-seismic signals with recurrent neural networks. *IEEE Trans. Geosci. Rem. Sens.* 57, 1936–1948. doi:10.1109/TGRS.2018.2870202

Trujillo-Castrillón, N., Valdés-González, C. M., Arámbula-Mendoza, R., and Santacoloma-Salguero, C. C. (2018). Initial processing of volcanic seismic signals using Hidden Markov Models: nevado del Huila, Colombia. *J. Volcanol. Geoth. Res.* 364, 107-120. doi:10.1016/j.jvolgeores.2018.09.008

White, R., and McCausland, W. (2016). Volcano-tectonic earthquakes: a new tool for estimating intrusive volumes and forecasting eruptions. *J. Volcanol. Geoth. Res.* 309, 139–155. doi:10.1016/j.jvolgeores.2015.10.020

Young, S. J., Evermann, G., Gales, M. J. F., Hain, T., Kershaw, D., Moore, G., et al. (2006). *The HTK book*. Cambridge, United Kingdom: Cambridge University Engineering Department). version 3.4.