University of Granada

DOCTORAL THESIS

Bayesian transfer learning for continuous monitoring of active volcanoes

Author:

Ángel Bueno Rodríguez

Thesis supervisors:

Carmen Benítez Ortúzar

Jesús Miguel Ibáñez Godoy

Doctoral Programme in Information and Communication Technologies
Department of Signal Theory, Telematics and Communications

Granada, June 2021

Editor: Universidad de Granada. Tesis Doctorales
Autor: Ángel Bueno Rodríguez
ISBN: 978-84-1117-025-3
URI: http://hdl.handle.net/10481/70452

To my beloved parents for their devotion and support throughout all my life

Acknowledgments

I am grateful to *Ministerio de Educación y Ciencia de España, (MINECO)* for funding this research through the projects TEC2015-68752 (KNOWAVES) and PID2019-106260GB-I00 (FEMALE).

It is a rare and wonderful thing to find a doctoral advisor who is simultaneously brilliant and kind. I was fortunate enough to find two of them. I would like to thank Prof. Carmen Benítez for her guidance and support through all these years, in an unmatched combination of intellect, intuition, wit, and kindness. Second, I am very grateful to Prof. Jesús M. Ibáñez, who has introduced me to the wonders of volcanology. Through countless and invaluable meetings, he has forged my mind into a seismologist and made me believe that a phenomenal scientist can also be an extraordinary human being.

I would also like to thank Prof Silvio de Angelis for a fruitful collaboration. He had introduced me to the volcanoes studied in this thesis through continuously useful, enjoyable, and challenging discussions. I would like to thank Prof. Maarten de Hoop, who I was fortunate to learn from and collaborate with to produce the scattering network in the last part of this thesis.

I am very grateful to Prof. Ivan Koulakov and Prof. Eduardo Del Pezzo for constructive feedback on this thesis memoir. Further, I would like to thank Prof. Richard Baraniuk for his invaluable feedback on the scattering network.

Also, I would like to thank my collaborator Dr. Alejandro Díaz-Moreno for all the hours he spent helping me with data management or explaining, with great patience, the nuts-and-bolts of applied seismology. Besides, I would like to thank Dr. Randall Balestriero for making collab-

oration work stimulating, challenging, and fun, always open to new ideas no matter how crazy they sound. I am indebted to Dr. Luciano Zuccarello, who has shown tremendous patience in attending to my continuous request for geophysical explanations about Mount Etna.

I would like to thank several collaborators that had helped me to shape my ideas into tangible results. Dr. Isaac Alvarez, who has introduced me to the design of signal processing algorithms, Dr. Manuel Titos for encouraging conversations, and Jack Woollam for our stimulating morning talks about seismology and machine learning.

On a personal note, thanks to my friend Vicente, *el maestro*, who has accompanied me through this adventure with continuous support. Likewise, I thank my friends Alberto, Cristobal, Antonio, and Luis - they kept my erudite conversations only when necessary.

To my parents, Angel and Teresa, for their unconditional support, love, and encouragement throughout my life. This thesis could not be possible without blindly putting faith in my studies, with Italian brewed coffee in the morning or supportive paellas during the tough deadlines.

Last but not least, to Evelyne Mottais. She has kept me alive in the ebb-and-flows of this academic journey, accompanying me through the multiple deadlines, and tolerated my absences and long work hours. In return for this, I have only received love from her.

SUMMARY

This thesis builds a bridge between Bayesian deep learning and seismo-volcanic monitoring systems for waveform recognition and change detection, with short-term volcano forecasting capabilities. We accomplish this task by proposing three scalable, flexible, and universal architectures to perform rapid recognition of earthquake transients while estimating uncertainty to identify volcano dynamics and seismic changes.

The advent of deep learning (DL) has made deep neural networks (DNNs) the preferred choice in the seismological community, boosting seismo-volcano monitoring to unprecedented levels by attaining high prediction quality while reducing a broad range of challenging monitoring tasks to computations within minutes. However, the direct application of DL methods in safety-critical systems is always conditioned to the data available and the flexibility required to interpret the unpredictable real world accurately.

A new research field, known as Bayesian deep learning (BDL), fuses the Bayesian theory with deep learning algorithms. The new BDL approach employs Bayesian neural networks (BNNs) to capture the uncertainty within the learning model. The BNNs allow rapid and robust Bayesian inference of complex high-dimensional data distributions, along with the estimation of data-inherent and model uncertainty. BNNs can be used in their standard form with static waveforms, such as a classifier of acoustic events. However, most acoustic data has a sequential nature. A seismic station collects seismic data streams, and BNNs may not be suitable to model the intra-event and temporal dynamics of the recorded data streams that compose the monitoring task. Thus, the design of robust architectures that can ingest sequential data while providing uncertainty es-

timation is fundamental for safety-critical tasks, such as seismo-volcano monitoring. Further, a set of unique geophysical challenges arise due to the nature of the data streams analyzed, leading to several monitoring drawbacks that impede the exploitation of DL high-capacity models and uncertainty quantification.

The shortage of large datasets correlated with volcanic activity is considered the foremost challenge facing a successful deployment of deep learning methods by volcanic observatories. Unfortunately, such datasets are not often available due to the cost and difficulty in manual labeling. Furthermore, when available, these seismic datasets cover a very short period of a particular eruptive episode. From a machine learning perspective, these brief temporary unrest episodes yield a perpetual problem in which only modest portions of non-uniform data samples are available for training. Moreover, available data catalogs assume that the geophysical variables and conditions in the seismic unrest episode will be later encountered or exportable to other volcanoes. Given all these hindrances, new alternatives have to be studied. We tackle these challenges by invoking a Bayesian approach over the proposed networks.

In the first part of this research work, we investigate a new Bayesian approach to classify seismo-volcanic events from frequency-based characteristics in two geologically related volcanoes. The experimental results of this first approach have shown that the proposed BNN can detect and recognize seismic events with very high performance. The model is even capable of discerning which type of earthquake corresponds to which volcano, with the ability to represent the uncertainty related to the changes in the dynamics of both volcanoes. Next, we have investigated transfer learning and uncertainty quantification as a unified framework to increase monitoring adaptability. We propose the uncertainty as a feature to be considered as a detector or change after an eruption and as a threshold to determine when transfer learning algorithms have to be used.

In the second part of the thesis, we formulate a Bayesian approach for continuous monitoring of seismo-volcanic data streams to elucidate whether the uncertainty changes are due to variations in the tremor background level, changes in the seismic events, or both. We propose a purely convolutional hybrid model to learn and detect the intra-frequency dynamic range of seismic-

volcanic events while performing automatic recognition for three different eruptions of the same volcano. We propose using the estimated uncertainties as a detector of changes with respect to the initial data distribution known by the model, a type of analysis known as data-drift detection. To do this, we divide the total statistical uncertainty into two geophysical terms: the observed variability of the seismogram and the inherent randomness of the monitored seismo-volcanic wave field. We study the temporal evolution of such proposed terms in the uncertainty analysis. From a geophysical point of view, we revealed that the temporal evolution of the waveform uncertainty coincides with higher eruption rates. The discovered data drift in seismo-volcanic datasets produces a degradation in the metric and a significant increase in uncertainty, both very perceptible during eruptions. However, after the main eruption, the stabilization of the studied volcano tends to return the monitoring metric to initial values, but the systems maintain high uncertainty levels. Finally, we checked that the temporal evolution of the uncertainty coincides with seismological bulletins.

We then explore if our proposed framework can exploit the total statistical uncertainty to overcome data scarcity limitations after the data drift has been detected, that is, subsequent to the main eruption. To this end, we blend our uncertainty framework with the temporal convolutional model to perform seismo-volcanic recognition while providing a set of maximally informative data samples to re-train the system. Experimental results have shown that our framework yields compelling performance with minimal dataset samples, with only 6.6% of the total seismic dataset to achieve the same performance as that of the baseline system.

In the last part of the thesis, we demonstrate that BDL methods can also be fine-tuned to learn a direct mapping between the target labels and the raw seismic waveform signals. The scattering network is a recent mathematical tool to parse raw waveforms into a robust, invariant, and stable feature representation via wavelet filterbanks. We propose a novel recurrent deep scattering network to learn the multi-scale temporal dependencies from streaming data and the wavelet filterbanks. We coupled the scattering transform with a carefully designed deep convolutional-LSTM architecture to learn the best intra-event temporal dynamics from the scattering coefficients. The uncertainty is essential in our approach as it identifies data drifts likely associated with the vari-

ations of volcanic processes. We have verified that the temporal evolution of the uncertainty coincides with seismological bulletins, as in our previous works. Therefore, we constructed enough experimental evidence to characterize the data drift in the studied volcanoes. Last, we demonstrate through transfer learning that the implemented architecture can be exported across various volcanoes and eruptive styles. We can reuse the implemented system to switch volcano types with minimal hassle, even if there are new or unknown signals in the target volcano where the transfer is required. Finally, we show that uncertainty can also be exportable between volcanoes as an indicator of eruptions.

RESUMEN

Esta tesis tiene como objetivo el desarrollo de nuevos sistemas de monitoreo sismo-volcánico para el reconocimiento de señales y la detección de cambios en volcanes, con potencial para la predicción de erupciones a corto plazo. Los algoritmos de monitorización volcánicos, a pesar de ser altamente eficientes en la vigilancia sísmica, adolecen de ciertas deficiencias que impiden su completo despliegue como modelos de alta capacidad en un observatorio vulcanológico. Abordamos esta compleja tarea mediante métodos de aprendizaje profundo de tipo bayesiano para desarrollar sistemas avanzados que permitan la identificación escalable, flexible y rápida de la señal sísmica, así como estimaciones de incertidumbre asociadas con los cambios en el campo de onda sísmico monitorizado.

El advenimiento del aprendizaje profundo (DL) ha convertido a las aproximaciones de redes neuronales profundas (DNN) en la opción preferida y por defecto para la comunidad sismológica. Esta adopción de sistemas de aprendizaje profundo ha impulsado el monitoreo sismo-volcánico a niveles sin precedentes, alcanzando tasas de rendimiento muy elevadas mediante cálculos de pocos minutos. Sin embargo, la aplicación directa de métodos DL en sistemas críticos para la seguridad humana siempre está condicionada a los datos disponibles y la flexibilidad requerida para interpretar con precisión el mundo real.

Las DNNs adolecen de una interpretabilidad interna muy limitada, aunque justificable debido a su formulación funcional. Un nuevo campo de investigación, conocido como aprendizaje profundo bayesiano (BDL), emerge para proporcionar capacidades interpretativas a las predicciones de redes DNNs. La nueva metodología BDL emplea redes neuronales bayesianas (BNN)

junto con metodologías aproximativas especializadas para capturar la incertidumbre asociada a un conjunto de datos, a la vez que mantiene todas las ventajas computacionales intrínsecas a los métodos DL. Las redes BNN permiten estimaciones bayesianas para distribuciones de datos altamente complejas y multidimensionales, proporcionando valores estimados de la incertidumbre inherente a los datos y del modelo. Esto algoritmos se pueden utilizar en su forma estándar con formas de onda segmentadas, como un clasificador de eventos sísmicos. Sin embargo, la mayoría de los datos acústicos exhiben una naturaleza secuencial. Una estación sísmica recopila flujos de datos sísmicos secuenciales, lo que implica que las BNN requieren de ciertas modificaciones para poder modelar la dinámica intra-evento y temporal de las secuencias de datos registrados que componen la serie sísmica.

La escasez de grandes conjuntos de datos que reflejen la actividad volcánica es el principal desafío al que se enfrenta un despliegue exitoso de los métodos de aprendizaje profundo por parte de los observatorios vulcanológicos. Desafortunadamente, estos conjuntos de datos no suelen estar disponibles debido al costo y la dificultad del etiquetado manual. Cuando están disponibles, estos conjuntos de datos sísmicos contienen una breve cronología eruptiva, muy específica del volcán monitorizado. Desde una perspectiva de aprendizaje automático, estas breves series sísmica son el origen de un problema perpetuo en el que los conjuntos de datos sísmicos, modestos en su origen, no homogéneos ni estandarizados en el etiquetado, asumen que el episodio sísmico estudiado se corresponderá con situaciones posteriores en el volcán. Dados este conjunto de deficiencias identificadas, junto a lo anteriormente expuesto sobre la necesidad del modelado temporal Bayesiano, debemos estudiar nuevas alternativas.

En la primera parte de la tesis, investigamos un nuevo enfoque bayesiano para la clasificación de eventos sismo-volcánicos a partir de características basadas en frecuencia en dos volcanes relacionados geológicamente. Los resultados experimentales de este primer enfoque han demostrado que una red neuronal profunda BNN puede detectar y reconocer eventos sísmicos con un rendimiento muy alto. El modelo incluso es capaz de discernir el origen volcánico del terremoto analizado, con bajas tasas de error. Este primer modelo ha demostrado la capacidad de proporcionar una representación de la incertidumbre relacionada con los cambios en la dinámica

de ambos volcanes. A continuación, hemos investigado la complementariedad entre técnicas de transfer learning y la cuantificación de la incertidumbre. Además, proponemos la incertidumbre como una característica que debe considerarse precursora de los cambios previos a una erupción y como límite para determinar cuándo deben reentrenarse los algoritmos de monitoreo mediante transfer learning.

En la segunda parte de este trabajo, modificamos el enfoque bayesiano a configuraciones de monitoreo continuo con la finalidad de dilucidar si los cambios de incertidumbre se deben únicamente a variaciones del nivel de fondo de tremor, a los eventos sísmicos, o a ambos. Proponemos un modelo temporal convolucional y Bayesiano para aprender el rango dinámico intrafrecuencia de los eventos sismo-volcánicos a la vez que se realiza el reconocimiento automático de eventos sísmicos, para tres erupciones distintas de un mismo volcán. En este punto de la investigación, empleamos las incertidumbres estimadas como un detector de cambios con respecto a la distribución de datos de entrenamiento aprendida inicialmente. Este tipo de análisis se conoce como *data-drift detection*. Para ello, dividimos la incertidumbre estadística total en dos términos geofísicos: la variabilidad observada del sismograma y la aleatoriedad inherente del campo de ondas sismo-volcánico monitoreado. Estudiamos la evolución temporal de tales términos y confirmamos, desde el punto de vista geofísico, que el data-drift en un volcán produce una degradación en la métrica y un aumento significativo de la incertidumbre, ambos efectos bastantes perceptibles durante las erupciones. Sin embargo, la estabilización del volcán estudiado tras una erupción tiende a retornar la métrica de monitoreo a valores iniciales, pero mantiene elevados niveles de incertidumbre.

En la siguiente parte de esta tesis, exploramos la capacidad del módulo clasificador temporal de nuestra arquitectura convolucional Bayesiana propuesta anteriormente para superar las limitaciones de la escasez de datos tras haber detectado el cambio. Con este fin, combinamos nuestro marco de incertidumbre con el modelo convolucional temporal para realizar tareas de monitorización volcánica al tiempo que muestreamos un conjunto de datos inciertos para volver a entrenar el sistema. Este procedimiento, conocido como aprendizaje activo (active learning), ha demostrado que nuestro marco ofrece un rendimiento prometedor con un número de muestras mín-

imas: el 6,6 % del conjunto de datos sísmicos total ha sido suficiente para lograr un rendimiento similar al sistema de base entrenado con todo el conjunto de datos.

En la parte final de esta investigación, demostramos que los métodos BDL también se pueden ajustar para aprender la tarea de monitorización a partir de la forma de onda sísmica sin procesar. Proponemos una nueva red neuronal profunda para modelar las dependencias temporales de los datos volcánicos. Para ello, empleamos una nueva metodología, basada en un banco de filtros wavelet (scattering network), para aprender una representación de características invariantes a cualquier tipo de deformación o translación matemática, conocida como coeficientes de scattering. A partir de la red de scattering propuesta, construimos una arquitectura convolucional en el tiempo (conv-LSTM) cuidadosamente diseñada para aprender la mejor dinámica temporal intra-evento a partir de los coeficientes de scattering. La incertidumbre es esencial en este enfoque, ya que identifica el data-drift asociado con las variaciones del volcán monitorizado. Hemos comprobamos que la evolución temporal de la incertidumbre coincide con boletines sismológicos y, por tanto, construimos suficiente evidencia experimental para caracterizar el cambio en los volcanes estudiados. Por último, demostramos a través del aprendizaje por transfer-learning que la arquitectura implementada es exportable a otros volcanes y estilos eruptivos. Con un número mínimo de operaciones, podemos reutilizar nuestro sistema de un volcán a otro, incluso si existen nuevas señales que sean desconocidas en el volcán donde es necesario realizar el nuevo aprendizaje. Finalmente, demostramos que la incertidumbre también puede ser exportable entre volcanes como un indicador de erupciones.

List of publications

This thesis covers the work presented in the publications listed below:

1. **A. Bueno**, R. Balestriero, S. De Angelis, C. Benitez, L. Zuccarello, R. Baraniuk, J.M. Ibánez and M. V. de Hoop. Recurrent Scattering Network detects metastable behavior in polyphonic seismo-volcanic signals for volcano eruption forecasting. IEEE Transactions on Geoscience and Remote Sensing (in press).

2. **A. Bueno**, M. Titos, C. Benitez and J.M. Ibánez. Continuous Active Learning for Seismo-Volcanic Monitoring. IEEE Geoscience and Remote Sensing Letters. (in press).

3. **A. Bueno**, L. Zuccarello, M.C. Benitez, S. De Angelis, J.M. Ibánez. Bayesian Monitoring of Seismo Volcanic Dynamics. IEEE Transactions on Geoscience and Remote Sensing. DOI: 10.1109/TGRS.2021.3076012.

4. **A. Bueno**, C. Benítez, S. de Angelis, A. Díaz-Moreno and J. M. Ibánez. Volcano Seismic Transfer Learning and Uncertainty Quantification with Bayesian Neural Networks. IEEE Transactions on Geoscience and Remote Sensing, vol. 58, no. 2, pp. 892-902, Feb. 2020.

5. **A. Bueno**, L. Zuccarello, A. Díaz-Moreno, S. De Angelis, J.Woolam, I. Alvarez, M. Titos, M.C. Benitez, J.M. Ibánez. PICOSS: Python Interface for the Classification of Seismic Signals. Computer & Geosciences. 2020 Sep 1; 142-104531.

6. **A. Bueno**, A. Díaz-Moreno, S. de Angelis, M.C. Benítez and J.M. Ibánez. REMOS: Recursive Entropy Method of Segmentation for Seismic Signals. Seismological Research Letters (SRL), 90, 2019. doi: https://doi.org/10.1785/0220180317.

Throughout this thesis, very fruitful collaborations have yielded a broad range of scientific contributions. Although *we do not explicitly present them as part of this thesis*, these research works have helped to progress towards the objectives of the thesis.

1. A. Diaz Moreno, **A. Bueno**, L. Zuccarello, J. Woolam, M. Titos , M.C. Benítez, J. M. Ibáñez, (2019, December). PICOSS: Python Interface for the Classification of Seismic Signals. In AGU Fall Meeting 2019. AGU.

2. **A. Bueno**, A. Díaz-Moreno, I. Alvarez, A De la Torre, O.D. Lamb, L. Zuccarello, S. De Angelis. VINEDA—Volcanic INfrasound Explosions Detector Algorithm. Frontiers in Earth Science. 2019 Dec 13;7:335.

3. J. Woollam, A. Rietbrock, **A. Bueno**, S. De Angelis. (2019). Convolutional neural network for seismic phase classification, performance demonstration over a local seismic network. Seismological Research Letters, 90(2A), 491-502.

4. **A. Bueno**, M. Titos, A. Díaz-Moreno, S. De Angelis, L. García, M.C. Benitez and J.M. Ibáñez (2018, December). An End-to-End System for Automatic Segmentation and Classification of Volcano-Seismic Waveforms. In AGU Fall Meeting 2018. AGU.

5. **A. Bueno**, M. Titos, L. García, A. Diaz-Moreno, S. de Angelis and J.M. Ibañez. (2018). A Bayesian Neural Network approach for the classification of volcano-seismic events. 20th EGU General Assembly, EGU2018, Proceedings from the conference held 4-13 April, 2018 in Vienna, Austria, p. 9526

6. **A. Bueno**, M. Titos, L. García, I. Álvarez, J.M. Ibañez, M.C. Benitez. (2018). Classification of volcano-seismic signals with Bayesian neural networks. In 2018 26th European Signal Processing Conference (IEEE -EUSIPCO) (pp. 2295-2299).

Preface

This thesis was prepared at the Department of Signal Theory, Telematics, and Communications, University of Granada (TSTC), whilst enrolled at the Doctoral Programme in Information and Communication Technologies of the University of Granada, Spain. This Ph.D. project was financed by *Ministerio de Educación y Ciencia de España, (MINECO)* through its Ph.D. Scholarship BES-2016-078495, research project KNOWAVES, TEC2015-68752 (MINECO/FEDER). This thesis has been supervised by Prof. Dr. Carmen Benítez Ortúzar (TSTC, University of Granada) and by Prof. Dr. Jesús M. Ibáñez Godoy (Department of Theoretical Physics and Cosmology and Instituto Andaluz de Geofísica, University of Granada).

During this Ph.D., I have worked on many different topics related to automatic Bayesian monitoring, change detection in volcanoes, signal processing segmentation algorithms, and data-curation tools, resulting in six peer-reviewed publications. However, this thesis only focuses on Bayesian monitoring and change detection from seismo-volcanic data streams. The thesis is divided into two main parts. The first part introduces the problem and the outline of this work. The second part contains four research papers that describe the architectures and how these demonstrate new venues of probabilistic deep learning in seismic monitoring. Two research papers for geophysical applications are also included.

Granada, 17 June 2021,

Angel Bueno

List of Figures

Contents

1 Introduction

1.1 Motivation

The monitoring of active volcanoes has been characterized as a very challenging data domain. From a machine learning perspective, the seismo-volcanic data domain is scarce as the availability of publicly open-access seismic records that ended in an eruption is minimal. Therefore, the monitoring challenges associated with the evolution of unrest have not received sufficient attention.

The complexity of seismic data streams entails years of study by experts to supply the scientific community with theoretical models that best explain the observables of a particular eruption. Unfortunately, the theoretical foundations of geophysical and eruptive models yield very sparse, non-uniform data taxonomies across volcanoes: theoretical labels valid in one volcano do not have to generalize to other volcanic environments worldwide. Last but not least, economic barriers do condition the available electronic equipment for monitoring: not all volcanoes in the world are monitored the same, or sometimes, stored data is overwritten in favor of periods of seismic interest due to lack of extensive storage facilities. As a result, seismo-volcanic data catalogs remain modest in some regions of the world and are often non-practical from a monitoring perspective.

Unforeseen seismic conditions may become more frequent in a volcanic environment, but the relationship between the algorithm and the monitoring task remains the same. Hence, the primary data challenge associated with the monitoring of volcanoes is the pertinence of the model over time. Unfortunately, this evolving situation can pass unnoticed by the deployed monitoring algorithm, ending in erratic behavior, with a high error rate, much after the eruption has passed.

However, in a safety-critical application such as seismic surveillance of active volcanoes, the monitoring algorithm can output predictions with high accuracy when visual evidence tells otherwise. This is especially concerning for seismo-volcanic monitoring applications, where datasets are typically small, labeling requires domain expertise, and any incorrect forecast may influence decisions that directly impact the human population.

The procedural approaches to recover and deploy monitoring systems after significant eruptions are epitomized in the retrospective analysis of new seismic data streams, a manual selection of events followed by a *fine-tuning* of the monitoring algorithm. The new seismic unrest conditions can outpace the data analysis and lead to insufficient data samples to retrain the algorithm. As an example of this situation, the eruptive periods studied here, with short quiescent periods before the next eruption. Thus, a practical problem is the prompt detection of seismic data stream alterations before they occur. The detection of change is intertwined with forecasting methodologies: knowing when and how the observables of the monitored volcano are changing can help to mitigate volcanic hazards and establish better early warning protocols.

This thesis tackles the challenges of monitoring interpretability and seismic change detection by developing a Bayesian methodology that permits the comprehensive consideration of how the volcano is changing and how it influences the monitoring task. From continuous data, the developed Bayesian procedure discovers, without supervision, seismic data stream alterations, maintaining the exportability of the monitoring system and the methodology proposed at all times. Estimating the uncertainty of the monitoring model conditioned to the available seismic data establishes new venues to communicate volcanic risk. Transfer learning procedures are proposed to transplant waveform information and prediction capabilities across volcanoes, even if no prior knowledge (i.e., labels) are available. However, transfer learning still requires substantial annotation efforts to provide a data catalog to perform learning procedures. Removing redundant information by selecting the optimum set of significant best new seismic signals reduces annotation efforts and increases monitoring adaptability under data shits. In this regard, we explore how our Bayesian algorithms can be incorporated as an uncertainty-based active learning framework. Finally, the uncertainty of the model is proposed as a potential precursor and an exportable

forecasting tool across volcanoes.

1.2 OBJECTIVES OF THE THESIS

The main objective of this thesis is to investigate, build and extend recently proposed advanced machine learning techniques in the context of seismo-volcanic recognition systems. The development of such systems can assist seismological observatories with the accurate forecasting of volcanic eruptions, help develop theoretical models to explain what drives volcanic eruptions, and increase early-warning capabilities while reducing the time required for evacuation protocols. Although the primary field of application of this thesis is seismo-volcanic monitoring, the algorithms introduced in this research are exportable to other acoustic data domains. We introduce an application example of our novel recurrent architecture for spoken digits classification and continuous environmental acoustic monitoring in Appendix A.

The core components of our proposed deep learning frameworks include deep neural networks (DNN), convolutional neural networks (CNN), temporal convolutions (TCNs), and specialized convolutional and recurrent architectures (ConvLSTMs). While utilizing these techniques for monitoring tasks, we also aim to understand how transfer learning can export the embedded knowledge in the hidden states of a deep learning framework when switching across volcanoes. The wide range of monitoring tasks tackled in this thesis, detection, segmentation, and classification, presents a clear idea of the scalability and the robustness of the presented techniques. Moreover, we aim to propose an uncertainty-based change detection in seismo-volcanic data streams, potentially using the uncertainty of the model in early-warning procedures. Lastly, we aim to investigate how these techniques can be integrated into an end-to-end monitoring framework and propose a novel deep learning architecture with a learnable feature representation obtained through a modified scattering transform from the raw signal.

In this thesis, the main research questions that we ask can be summarized as follows. First, we investigate whether Bayesian deep learning techniques could be applicable to seismo-volcanic recognition systems and if robust performance can be achieved. Second, the adaptability of the

model via uncertainty-based active learning is explored to solve the data-scarcity problems. Third, we question the effectiveness of the established feature representation techniques and search for new venues to integrate deep learning with raw seismic data streams in a multi-label setting. Finally, we investigate how the estimated uncertainty can be exploited from a forecasting perspective: if data drifts are detectable, how these affect performance, and how change correlates with the evolution of volcanic unrest.

1.3 Main results

The main results and the contributions of our publications included in this thesis are listed below, as follows:

Publication 1: Volcano Seismic Transfer Learning and Uncertainty Quantification with Bayesian Neural Networks.

In this first publication [14], we propose Bayesian Neural Networks (BNNs) to perform event identification, classification, and uncertainty estimation on seismic data gathered at two active volcanoes, Mount St. Helens (USA) and Bezymianny (Russia). When both datasets are merged, and no additional training information is provided, the proposed BNNs offer a considerable 92.08% performance in discriminating both the type of event and the volcano of origin. Similarly, the BNNs attain monitoring recognition performances above 90% on each volcano, independently. We illustrated frequency content variations with estimated uncertainty for each volcano, thus associating uncertainty to changes in the state of unrest at the studied volcanoes. We propose that the uncertainty gauge whether the learned models could be exported to other eruptive scenarios with transfer learning.

Publication 2: Bayesian Monitoring of Seismo Volcanic Dynamics.

In this second publication, [10], the validity of the claim that seismic variations lead to higher uncertainty levels before main eruptions are investigated from the perspective of continuous moni-

toring. Our strategy blends a segmentation and temporal recognition module to learn a denoised time-frequency feature representation, later used in a dynamic contextual analysis for automatic classification. The experimental results demonstrate efficient signal detection and classification accuracy. Furthermore, the estimated uncertainty provides a measure of the seismic wavefield variations, with the characterization of the data drifts in the volcanic system in the hours preceding eruptive activity.

Publication 3: Recurrent Scattering Network detects metastable behavior in polyphonic seismo-volcanic signals for volcano eruption forecasting.

In the proposed deep neural network method in publication [12], we augment the capability of deep learning architectures with a modified, time-dependent scattering transform to learn a robust feature representation from raw seismic signals. Specifically, we propose morphing the scattering transform into a novel E2E hybrid and recurrent learnable deep scattering network to adapt to the multi-scale temporal dependencies from streaming data. At the same time, with a carefully designed deep convolutional-LSTM architecture, we learn intra-event temporal dynamics from the scattering coefficients or features. Through evolving epistemic uncertainty, invoking a Bayesian network strategy, we detect the seismic change and demonstrate the significance of uncertainty as an indicator for possible forecasting of eruptions. We verify the effectiveness of transfer learning switching between similar and geophysically different volcanoes.

Publication 4: Continuous Active Learning for Seismo-Volcanic Monitoring

In this work, we explore active learning from a deep neural network perspective using seismic data streams [11]. Active learning has remained unexplored in seismic monitoring due to the operational requirements of scalable uncertainty estimation of the data stream and a principled procedure to select uncertain events within the seismic data stream. This work proposes an active learning procedure based on temporal convolutions to estimate a temporal uncertainty map to segment and select the class memberships that need to be reviewed. As a result, we attain a significant improvement in monitoring metrics, with only a fraction of the initial dataset to achieve

a similar performance of 83% for five seismo-volcanic events compared to the baseline system trained on the full dataset.

Publication 5: REMOS: Recursive Entropy Method of Segmentation for Seismic Signals.
In publication [13], we propose an algorithm to detect, extract, and classify volcanic earthquakes starting from raw, continuous, waveform data. This algorithm uses the well-established STA/LTA method to identify trigger times of candidate earthquakes. A minimum entropy criterion is then employed to investigate many triggers and parse events into individual waveforms. Experimental results suggest that REMOS can effectively produce data catalogs, even when the signal-to-noise ratio is poor or event occurrence rates are high.

Publication 6: PICOSS: Python Interface for the Classification of Seismic Signals.
The large seismo-volcanic datasets pose a significant challenge for the manual detection and classification of volcano-seismic signals. This research work [15] presents a novel open-source Python interface designed to support seismo-volcanic data analysis. The high level of modularity within this interface permits the geophysicist to quickly detect seismic signals and label raw seismic events using either automated or manual approaches. Furthermore, implemented modules follow an intuitive design, enhancing the efficiency and accuracy of the essential data-labeling tasks required for large-scale volcano-seismological studies.

1.4 Outline of the thesis

The remainder of this work is organized as follows:

1. The theoretical background about Bayesian deep learning methods, and uncertainty quantification, along with the implemented Bayesian methods, are presented in chapter 2
2. The scientific discipline of volcano seismology, how we tackled the classification of events, and the developed uncertainty framework is presented in chapter 3.

3. Chapter 4 describes the implemented Bayesian deep neural network for waveform classification and uncertainty estimation. The links and temporal evidence between uncertainty and unrest are presented here. Transfer learning is used to export system performance across volcanoes.

4. In chapter 5, we introduce the research work with the hybrid Bayesian monitoring framework to detect and segment seismo-volcanic signals. The connection between Bayesian uncertainty analysis and continuous volcano-seismic monitoring is presented here.

5. We extended the scattering transform to perform learnable temporal modeling and introduced a multi-modular recurrent architecture to perform multi-label detection, segmentation, and classification of seismo-volcanic waveforms. Our architecture and applicability on data from three volcanoes are discussed in chapter 6.

6. In chapter 7, we introduce the active learning approach and cost-effective fine-tuning strategy to perform retraining on the system and adaptability to new conditions when the volcano has changed.

7. Finally, chapter 8 includes the geophysical applications developed as part of this thesis and that support the data curation process and datasets of this research work.

8. Conclusions and discussions for the current and future research on seismic monitoring applications are provided in chapter 9.

In the appendixes of this document, we include additional figures to demonstrate the interpretability of the developed scattering network in the field of seismo-volcanic monitoring, along with benchmark results that have helped us to fine-tune the architectures and additional experimental results of this thesis.

2 | THEORETICAL BACKGROUND

Most problems in Bayesian statistics consist of estimating probability densities that are difficult to calculate. This type of approach includes a set of algorithms to determine unknown quantities. In the case of neural networks, the non-linearity introduced by the multiple hidden layers implies that the exact knowledge of the parameters of the network is not directly computable. However, the knowledge of these parameters is of interest in using neural networks, for example, to make predictions and compute uncertainties. Therefore, it is necessary to make an approximation that must be precise, scalable, and flexible enough for deep neural network architectures working on streaming data. This section summarizes the variational inference framework based on Monte Carlo numerical sampling, its connections with stochastic regularization techniques in deep learning. In addition, we introduce the problem of the data drift and how the uncertainty can be exploited to detect such change. The training workflow, model selection, optimization procedures, and the technical scaling of the architectures are also presented as part of this chapter.

2.1 MATHEMATICAL DEFINITIONS

First, we establish the mathematical notation that we will follow in the formulation of Bayesian modeling for neural networks. We define our dataset (assuming we are working with seismic data) as a set of N points, $D = (\mathbf{X}, \mathbf{Y})$ where $\mathbf{X} = \{x_1, x_2, \ldots, x_N\}$ the matrix whose rows correspond to a set of seismic events (in samples) extracted from a continuous seismic record, and $\mathbf{Y} = \{y_1, y_2, \ldots, y_N\}$ is the matrix whose rows contain the corresponding labels that are assigned to

every seismic event and that are categorized over a set of C classes. It is important to emphasize that the X and Y matrix can be adapted according to the type of monitoring task at hand. In event classification, this matrix $\mathbf{X}$ contains the frequency characteristics extracted from the segmented signal for each of its rows and each $\mathbf{Y}$, the corresponded label in event classification. We refer to the opening problem of temporal event recognition, being $\mathbf{X}$, the matrix whose rows contain data streams with multiple events, and the $\mathbf{Y}$ matrix whose rows contain the labels (or multi-labels) of each event.

We refer as L to the total number of layers in the neural network, being i the sub-index of any of its layers, $i = \{1, \cdots, L\}$. Therefore, we can specify our neural network as a function parameterized by its weights, $\mathbf{y} = f^{\omega}(\mathbf{x})$, where ω represents all weights matrices associated with the hidden layers of the neural network, that is, $\omega = \{W_i\}_{i=1}^{L}$, $\mathbf{x}$ the input feature vector of the network (i.e., a row of the matrix $\mathbf{X}$) and $\mathbf{y}$ the associated vector labels (i.e., a row of the matrix $\mathbf{Y}$). In this section, we refer as θ to the parameters that define the approximate distribution $q_{\theta}(\omega)$.

2.2 Bayesian modeling in neural networks

It is well known that Bayesian methods provide a measure of uncertainty for each input and output of a given model, based on all observed data. In most of the so-called *frequentist approaches*, commonly used in neural networks, the final optimization result is a set of best-fitting parameters. Unlike frequentist methods, the result of a Bayesian fit is a probability distribution of each parameter of the model, called the *posterior distribution*. For a given set of parameters in our neural network, and the dataset D defined, the posterior $p(\omega|\mathbf{X}, \mathbf{Y})$ is determined by using the Bayes Theorem:

$$p(\omega|\mathbf{X}, \mathbf{Y}) = \frac{p(\mathbf{Y}|\mathbf{X}, \omega)p(\omega)}{p(\mathbf{Y}|\mathbf{X})}. \tag{2.1}$$

with $p(\mathbf{Y}|\mathbf{X}, \omega)$ is defined as the *model likelihood distribution*, that is, the knowledge of the model on the data distribution, and therefore, the assignment of probabilities for each $\mathbf{X}$ and $\mathbf{Y}$, given the parameters of the model. The term $p(\omega)$ is known as *prior* and constitutes the initial, known prob-

ability distribution of the parameters of the network. Hence, in a Bayesian neural network (BNN), the prior distribution is specified as a set of probability distributions located on their weights [38]. The denominator of (2.1) corresponds to the *model evidence* or *marginal likelihood*, a normalizing constant that can be obtained by marginalizing the likelihood over the parameters ω:

$$p(\mathbf{Y}|\mathbf{X}) = \int p(\mathbf{Y}|\mathbf{X}, \omega) p(\omega) \, \mathrm{d}\omega. \tag{2.2}$$

Theoretically, the marginalization in (2.2) involves the average with respect to all possible parameters of the model ω, weighted by $p(\omega)$. For complex models, such as BNNs, an approximation is required [30]. Defining all the terms in the numerator and denominator of (2.1), a BNN can predict the outputs y^* for any new input x^* through the predictive function by integration over the parameters of the network ω:

$$p(y^*|x^*, \mathbf{X}, \mathbf{Y}) = \mathbb{E}_{p(\omega|\mathbf{X},\mathbf{Y})} = \int p(y^*|x^*, \omega) p(\omega|\mathbf{X}, \mathbf{Y}) \, \mathrm{d}\omega. \tag{2.3}$$

where $p(y^*|x^*, \omega)$ is the data likelihood for this new point x^*. The prediction of new seismic events for multiple instances in y^* is known as *inference*. However, the exact inference in (2.3) is impossible given that the posterior is part of the integral. The computation of (2.3) with $p(\omega|\mathbf{X}, \mathbf{Y})$ is equivalent to evaluate an infinite number of neural networks with all the possible parameter configurations. This is computationally intractable for neural networks of any size. For this reason, in Bayesian modelling, an *approximate inference* procedure is required. This type of inference entails an optimization conditioned to the training of the architecture, that is, the approximation of this integral. Variational inference methods are used to approximate $p(\omega|\mathbf{X}, \mathbf{Y})$ and therefore the equation (2.3).

2.3 VARIATIONAL INFERENCE IN BNNS

Variational Inference (VI) focuses on obtaining an approximation to $p(\omega|\mathbf{X}, \mathbf{Y})$ by using optimization procedures [9]. Formally, this optimization aims at determining a probability density, $q_\theta(\omega)$, that should be as close as possible to $p(\omega|\mathbf{X}, \mathbf{Y})$. The measure of closeness is given by the Kullback-Leibler (KL) divergence between both distributions:

$$KL(q_\theta(\omega) \,||\, p(\omega|\mathbf{X}, \mathbf{Y})) = \int q_\theta(\omega) \; \log \left\{ \frac{q_\theta(\omega)}{p(\omega|\mathbf{X}, \mathbf{Y})} \right\} d\omega. \tag{2.4}$$

with $q_\theta(\omega)$ known as the *variational distribution*. We minimize the (2.4) by optimizing the variational parameters θ of our variational distribution $q_\theta(\omega)$:

$$\hat{\theta} = \underset{\theta}{\operatorname{argmin}} \; \mathbb{E}_{q_\theta(\omega)}[log \; q_\theta(\omega) - log \; p(\omega|\mathbf{X}, \mathbf{Y})] \tag{2.5}$$

$$\hat{\theta} = \underset{\theta}{\operatorname{argmin}} \; KL(q_\theta(\omega) \,||\, p(\omega|\mathbf{X}, \mathbf{Y})) \tag{2.6}$$

with $\hat{\theta}$ the parameters that results in the minimum KL divergence. Once we obtained our variational approximation $q_\theta(\omega)$, and the KL in (2.4) has been minimized, the predictive distribution is given as:

$$p(y^*|x^*, D) \approx \int p(y^*|x^*, \omega) q_{\hat{\theta}}(\omega) \, \mathrm{d}\omega =: q_{\hat{\theta}}(y^*|x^*). \tag{2.7}$$

However, we can verify that the evaluation of the KL divergence in (2.4) requires the computation of the posterior distribution for our network's parameters, which are precisely the distribution that we want to approximate. To circumvent this, we can minimize a function similar to (2.4) added to a constant term. This function is known as Evidence Lower Bound (ELBO). The mathematical relationship between the ELBO and KL divergence can be derived from (2.4) as follows:

$$
\begin{aligned}
KL(q_\theta(\omega) \,||\, p(\omega|\mathbf{X},\mathbf{Y})) &= \int q_\theta(\omega)\, log \left\{ \frac{q_\theta(\omega)\, p(\mathbf{Y}|\mathbf{X})}{p(\mathbf{Y}|\mathbf{X},\omega)\, p(\omega)} \right\} d\omega \\
&= \int q_\theta(\omega)\, log\, p(\mathbf{Y}|\mathbf{X}) d\omega + \int q_\theta(\omega)\, log \left\{ \frac{q_\theta(\omega)}{p(\mathbf{Y}|\mathbf{X},\omega)\, p(\omega)} \right\} d\omega \\
&= log\, p(\mathbf{Y}|\mathbf{X}) - \underbrace{\int q_\theta(\omega)\; log \left\{ \frac{p(\mathbf{Y}|\mathbf{X},\omega)\, p(\omega)}{q_\theta(\omega)} \right\} d\omega}_{\text{ELBO or } \mathcal{L}_{ELBO}(\theta)}
\end{aligned}
$$

Therefore it is observable from the above equations that the KL divergence is equal to the ELBO ($\mathcal{L}_{ELBO}(\theta)$) and a constant which is given by the marginal log-likelihood of our data. Since the KL divergence is a probabilistic distance and always positive, we can thus write:

$$
log\, p(\mathbf{Y}|\mathbf{X}) \geq \mathcal{L}_{ELBO}(\theta) + KL(q_\theta(\omega) \,||\, p(\omega|\mathbf{X},\mathbf{Y})). \tag{2.8}
$$

with $\mathcal{L}_{ELBO}(\theta)$ becoming the objective of our optimization problem. In addition, minimizing the divergence of KL is also equivalent to maximizing ELBO with respect to the variational parameters of the distribution $q_\theta(\omega)$. We can expand the term $\mathcal{L}_{ELBO}(\theta)$, and obtain its closed-form expression:

$$
\mathcal{L}_{ELBO}(\theta) := -\int q_\theta(\omega)\; log\, p(\mathbf{Y}|\mathbf{X},\omega) d\omega \;+\; KL(q_\theta(\omega) \,||\, p(\omega)). \tag{2.9}
$$

The optimization of the first integral term conditions the Bayesian model to better fit our data. The second KL term acts as a regularizer, keeping $q_\theta(\omega)$ from extreme deviations of $p(\omega)$. This analytical representation can be used to rewrite (2.4) in approximative terms for the parameters of our neural network:

$$
\begin{aligned}
KL(q_\theta(\omega) \,||\, p(\omega|\mathbf{X},\mathbf{Y})) &\propto -\int q_\theta(\omega)\;\; \log\; p(\mathbf{Y}|\mathbf{X},\omega) d\omega + KL(q_\theta(\omega) \,||\, p(\omega)) \\
&= -\sum_{n=1}^{N} \int q_\theta(\omega)\;\; \log\; p(y_n|f^\omega(x_n)) d\omega + KL(q_\theta(\omega) \,||\, p(\omega))
\end{aligned} \tag{2.10}
$$

where $f^\omega(x_n)$ is the output of the neural network for a given arbitrary input x_n, and the summatory term defined as the *expected log likelihood*. Once all the parameters for variational optimization in a BNN have been established, it is necessary to choose the prior and explicitly define the vari-

ational $q_\theta(\omega)$ distribution to optimize in (2.10). In a BNN, $q_\theta(\omega)$ is always conditioned to the distribution given by the matrices of its neural connections. The multiple non-linearities and the evaluation of the first integral in (2.10) with N events of a dataset entails a prohibitive, non-scalable computation. However, we can consider the Monte Carlo sampling estimators and their connections to regularization techniques in deep neural networks. Monte Carlo estimators permit an approximation of the expected log-likelihood for neural network models with multiple hidden layers and their derivatives with respect to the variational parameters θ. The so-called Monte-Carlo dropout (MC-dropout) is thus a variational estimation that connects dropout regularization and standard neural network optimization with the inference procedure in (2.10) [25].

2.4 Monte Carlo Dropout

The dropout technique can be used in a BNN as a Bayesian approximation of the posterior distribution of the network parameters. Initially, the dropout is formulated by [62] as a stochastic regularization technique for deep neural networks, randomly deactivating the parameters of a neural network with a given probability, p_i. A Bernoulli distribution can model this probability p_i, selecting which of the hidden units remain active in the network. The key result for this reasoning is derived by [25] and [24]: the integral and KL terms in (2.10) can be linked to standard dropout training in deep neural networks. This permits scalable and robust inference for large datasets in very complex networks.

Formally, our neural network is composed of a set of weight matrices in all its layers, $\omega = \{W_i\}_{i=1}^{L}$. Each weight matrix has a dimension $K_i \times K_{i-1}$. We define the variational distribution $q_\theta(\omega)$ as the factorization over the weight matrices of all the hidden layers conditioned to the dropout technique:

$$W_i = M_i \cdot \text{diag}([z_{i,j}])_{j=1}^{K_i}, \tag{2.11}$$

$$z_{i,j} \sim \text{Bernoulli}(p_l), \quad i = 1, \ldots, L, \; j = 1, \ldots, K_{l-1}, \tag{2.12}$$

where $z_{i,j}$ represents the dropout masks (matrices of zeros and ones drawn from the Bernoulli distribution) which disable the hidden element j on layer $i-1$. The term M_i is a *mean weight matrix*, whose set $\theta = \{M_i\}_{i=1}^{L}$ are the variational parameters. Finally, having defined the variational distribution, we can use Monte-Carlo estimation to approximate the integral of the expected log-likelihood in (2.10):

$$\begin{aligned} -\int q_\theta(\omega)\ \log\ p(\mathbf{Y}|\mathbf{X},\omega)d\omega &= -\sum_{n=1}^{N}\int q_\theta(\omega)\ \log\ p(y_n|f^\omega(x_n))d\omega \\ &= \frac{1}{N}\sum_{n=1}^{N} -log\ p(y_n|f^{\hat{\omega}}(x_n)) \end{aligned} \tag{2.13}$$

where $\hat{\omega}$ is not a maximum posterior estimate, but multiple realisations of random variables from the Bernoulli distribution, $\hat{\omega} \sim q_\theta(\omega)$. This reasoning is identical to applying successive dropout masks to the network weights. Hence, the averaged sum of $log\ p(y_n|f^{\hat{\omega}}(x_n))$ represents, by definition, the cost function of a neural network.

In order to link the variational inference optimization $\mathcal{L}_{ELBO}(\theta)$ to the optimization objective of standard neural networks with dropout, $\mathcal{L}_{dropout}(\theta)$, it is necessary that the mathematical relation known as KL-condition in the regularizer term is fulfilled. The KL-condition links the derivatives of the optimization objective in (2.4) with standard loss functions in neural networks. In this Chapter, we do not cover the entire proof in detail and refer the reader to the original work by [25]. The KL-condition establishes that the regularizer KL term in (2.4) can be approximated as a standard dropout regularizer weighted by a normalization constant λ. Our objective of variational minimization is defined as:

$$\mathcal{L}_{dropout}(\theta) = \frac{1}{N}\sum_{n=1}^{N} -log\ p(y_n|f^{\hat{\omega}}(x_n)) + \lambda\sum_{l=1}^{L}(\|M_l\|_2^2 + \|b_l\|_2^2) \tag{2.14}$$

Therefore, approximate inference procedures result in an optimization goal identical to that of a neural network using the loss function $\mathcal{L}_{dropout}(\theta)$. This function is defined to optimize the parameters of the neural network and find the best $q_{\hat{\theta}}$ that minimizes the KL divergence term in equation (2.10), $KL(q_\theta(\omega)\ ||\ p(\omega|\mathbf{X},\mathbf{Y}))$. Finally, we can use the approximation learned by our

network to evaluate the predictive function in (2.7), using Monte-Carlo sampling with T sampling steps:

$$q_{\hat{\theta}}(y^*|x^*) = \int p(y^*|f^{\hat{\omega}}(x^*))q_{\hat{\theta}}(\omega)d\omega \approx \frac{1}{T}\sum_{t=1}^{T} p(y^*|f^{\hat{\omega}_t}(x^*)) \tag{2.15}$$

or equivalently $\hat{\omega}_t \sim q_\theta(\omega)$. Therefore, at the time of inference, the dropout layers are applied to the M_i matrices, generating a Monte-Carlo sample from the posterior distribution (see equation 2.3). In practice, the average of these samples can be interpreted as the prediction of the network, although a single estimate is not obtained, as many as T sampling steps are performed. We can use the probabilities obtained by MC-dropout to estimate the uncertainty in the application of seismo-volcanic recognition.

2.5 Training of the architectures

The monitoring models that we have developed as the core algorithms of this thesis fuses deep neural network, seismic datasets, and Bayesian techniques that have to be fine-tuned following the standard practices of machine learning. While the structure of deep learning networks permits rich modeling capacity, the learning of the appropriate parameter values for the models is quite a complex and challenging task. This section begins with a description of the optimization frameworks in which the models and techniques here are implemented. Then, we present the regularization techniques implemented to reduce training time but increase generalization on the models.

2.5.1 Technical implementation

The data processing routines required for reading raw seismic data streams and array operations are based on ObsPy [8], and NumPy [28]. There has been a growing interest in the seismological community to incorporate unified open-source platforms to standardize the variety of seismic data file formats and access to seismic network data. The package ObsPy unifies many geophysical signal processing tasks into a workflow that seamlessly integrates with the deep learning

libraries used in this work. Similarly, the package NumPy is the core array programming package with a very expressive syntax for accessing and operate high-dimension data structures. We integrated these two packages in the geophysical applications research manuscripts presented in this thesis, [13], and [15].

The deep learning models in [10], [12], [11] and [14] have been implemented using these open-source scientific libraries integrated with a state-of-the-art, high-level deep learning research library, TensorFlow [40]. TensorFlow is an industry-wise, well-established Python workflow developed by Google research teams to standardize and make accessible novel research deep learning operations, with built-in support for automatic gradient computations. Also, TensorFlow offers integration at many programming levels, ranging from very explicit routines to already pre-designed networks, with capabilities to extend single-device computations to multiple, heterogeneous computational graphs. The language expressivity of TensorFlow comprises a range of functions, implemented mathematical expressions, and optimization sub-routines that can be used as the core components of our optimization routines.

2.5.2 Training methodology

In figure 3.4; we depict a schematic overview of the machine-learning training methodology used in this work. We divide the algorithmic fine-tuning into two well-differentiated blocks. The upper part of Figure 3.4, or **A**, corresponds to the standard data processing pipeline required to parse raw waveforms into input features and target labels. The lower part of Figure 3.4, or **B**, corresponds to the definition and optimization process of the model. The following operations are performed for all the research works concerning DL architectures unless stated otherwise:

1. **Data preparation steps:** This first step ingests the seismic data catalogs already prepared with our developed algorithms [13], [15]. Then, given the differences of tasks and architectures implemented in this thesis, this block is geared according to the subroutines and the monitoring task at hand. In [14], this pipeline is adapted to read individual, segmented waveforms. Then, the data is band-pass filtered, followed by cepstral features extraction to

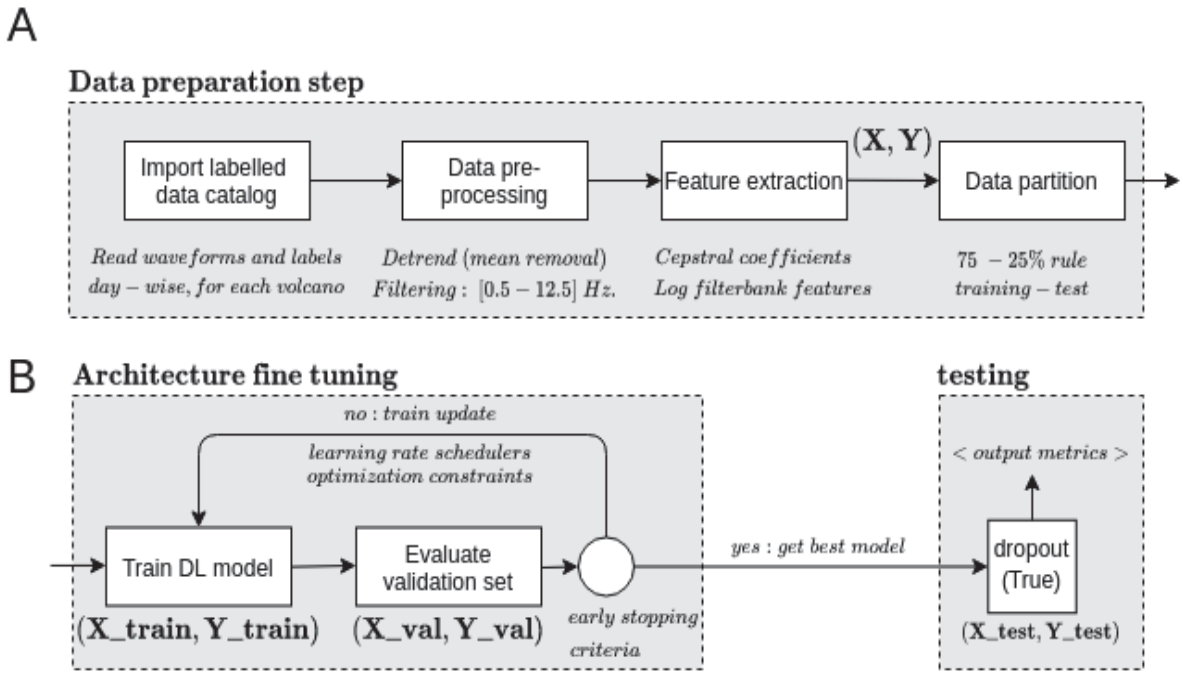


Figure 2.1: Machine-learning training methodology followed during the fine-tuning of the implemented models. Seismo-volcanic signals mandate the data preparation steps to produce a structured dataset, later used in the fine-tuning of the architecture. Regularization techniques, learning rate halving, and early stopping are systematically used to speed up training time while eliminating overfitting

perform seismic event classification. Hence, we can define our dataset $(\mathbf{X}, \mathbf{Y})$, with X the matrix whose rows contain the extracted cepstral features, and $\mathbf{Y}$ the matrix whose rows contains the label. In [10] and [11], we follow a very similar procedure for this block, but considering that our dataset $(\mathbf{X}, \mathbf{Y})$ is given by $\mathbf{X}$, the matrix whose rows contains the spectrogram of 5 minutes, whereas $\mathbf{Y}$ contains the matrix whose rows corresponds to the categorical events within $\mathbf{X}$. In [12], we erased the feature extraction process, substituted it with a learnable wavelet filterbank as part of the scattering neural network component. We refer reader the subsequent chapters for further insight into the data preparation steps for each block.

2. **Architecture fine-tuning:** This block requires the definition of a loss function, the network architecture, and the constraints that condition the training of the architecture. Each data sample in the training set is randomly indexed and gathered into mini-batches that the

model individually processes at each training step (or epoch). Hence, for each X, being an input feature vector or sequential waveforms, along with the corresponding labels in the Y, are randomly indexed and processed by the network. The average loss for each mini-batch is estimated, and the gradients are computed for each mini-batch via back-propagation algorithms.

Gradient-based methods are the most common procedures used to optimize deep learning networks. We first compute the gradient vector of the loss function of choice for the entire training dataset with respect to the parameters of our model for a pre-defined number of epochs. Then, the parameters are updated according to the delta rule, with the *learning rate*; a parameter that scales the magnitude of our weight updates. In this thesis, we selected *Adam* optimizer [34]; an optimization technique that improves traditional stochastic gradient descent (SGD) by computing adaptive learning rates for each parameter conditioned to a regressive approximation of second-order gradients. The step size of *Adam* update rule is robust across epochs, which helps the optimization to navigate through saddle points or ravines in the loss functions. We use the default values proposed in [34], (0.9 for β_1, 0.999 for β_2, and 10^{-8} for ϵ), as these empirically serve well in practice for a broad range of deep learning tasks.

The fine-tuning process of deep learning architectures is iterative. Once the network parameters are updated and a training step is finished, each data sample in the training step is randomly indexed again into mini-batches, and the optimization procedure is repeated.

One can think that optimizing a neural network can continue unbounded by a pre-defined number of epochs. However, that is not the case, as the standard practice in the machine learning community concern the use of regularization techniques and frequent generalization checks to speed up training procedures and increase the robustness of the model. In this research work, we have used the following, unless stated otherwise:

1. **Early stopping:** This technique can be defined as the periodic check with a validation set (sometimes named development set in audio domains) [51], [19]. Early stopping retains

the generalization capabilities of the model for unseen data samples while speeding up training time. Early stopping defines an initial state, $t = 0$, and a threshold named the *patience interval* that controls the dynamics of optimization and the number of steps that can lead to overfitting, halting the training if necessary. When an epoch n has finished and the averaged loss computed, we check the performance with the new model. If the losses and performance attain better metrics than the previous $n - 1$ training epoch, the optimization continues, and the parameters of the model are saved. If the losses and performance yield worse metrics than the previous $n - 1$ training epoch, the state t is incremented by one, and the parameters of the model are not saved. Then, early stopping allows the optimization during a number of epochs equal to the value *patience interval*, incrementing by one on each step if metrics do not fulfill the validation criteria. Suppose the model has navigated the loss function and reached the *patience interval*, the training halts. Otherwise, if before the *patience interval* the model attains a superior metric than the best-stored model, the state t is reset to zero, and optimization continues from this point. In [14] we set 50 epochs, with early-stopping set to a patience interval of 5 epochs. In [10] we set 50 epochs, with early-stopping set to a patience interval of 5 epochs. In [10]; [12] and [11] we used early-stopping with a patience interval of 5 epochs over 300 training epochs.

2. **Learning rate scheduler:** The loss landscape of a neural network is defined as the error obtained for each of the neural network parameters. For each parameter configuration (and even network configurations), the landscape loss can not be smooth, difficult to navigate through the optimization process, and highly complex. This technique is used in combination with early-stopping to refine the optimization of the models further. Thus, we can establish a learning rate schedule to update the learning rate during the training. We define an initial state $l = 0$ to track the metrics with the validation set. The number of epochs that the learning rate scheduler allows the optimization process is constrained to another *patience-interval*, not necessarily the same as in the early-stopping case. When the patience interval is reached, the learning rated is halved, and the state l reset to zero again. This

avoids learning rates that are too high that the network can start to diverge during the training. In this work, we generally used this technique to monitor validation losses, keeping the patience interval a few epochs behind the early-stopping patience interval (3 and 4).

3. **Dropout:** Dropout is a regularization technique that randomly sets to zero the hidden units of a neural network. The dropout regularization technique is the pivotal approach of all our neural networks. The basic idea builds upon training smaller models instead of a larger one that can overfit the network. These smaller models are sub-sampled by the random deactivation of the weights. Hence, as these smaller models share parameters with the original model, training a model with dropout is equivalent to training the entire model with all the parameters at once. Thus, this procedure implicitly provides data augmentation techniques and ensemble training. Furthermore, the statistical averaging of all the sub-sampling models improve the generalization of the optimized architecture. Typically, the statistical averaged network with all the parameters, often named as *mean-network*, is used at prediction time. However, the use of dropout at test time involves a stochastic approximation with Bayesian implications (section 2.4 of this chapter). Therefore, we employ dropout on training and test time. In this thesis (Figure 3.4), to permit dropout at test time, TensorFlow sub-routines flagging the inference phases, along with custom-made TensorFlow layers, are used to invoke dropout at test time. The predictive output is performed via Monte Carlo T sampling steps [10], [12].

2.5.3 Selecting the architecture

It is a common practice to fine-tune deep-learning systems towards best performance in the task at hand. For example, in seismo-volcanic classification, select the best parameters of a neural network that provide the best monitoring metrics. This type of problem, known as *hyper-parameter* optimization, involves a non-trivial, high dimensional, non-convex optimization of parameters towards optimum performance. Finding the best set of hyperparameters in a neural network is a very active field of research, with a wealth of literature and a gamut of approaches, ranging from

nature-based algorithms [45], tree-structure based optimizations, [6], reinforcement learning [71], among many others methodologies [69]. Each one of the hyperparameter optimization algorithms has different setups or requirements. Nonetheless, three primary methodologies are the accepted practice in deep learning for hyper-parameters tuning: grid search, random search, and Bayesian search.

1. **Grid search** is a thorough approach over all the training parameters, but it can be computationally prohibitive. The optimization is performed over an exhaustive search on the hyperparameter space, often as independent trials specified by the users. The main limitation of grid search is the exponential growth of the number of times required to evaluate the model, making it unfeasible.

2. **Random search** serves on smaller architectures but with the risk of missing essential training parameters during the random trials. However, this approach is feasible if some *a-priori* knowledge of the search space is available to the user beforehand.

3. **Bayesian search** tackles hyperparameter optimization in a reasonable amount of time, leveraging random and grid search approaches. In Bayesian search, we implement a framework with a given initial, non-uniform but plausible grid of hyperparameters. Then, the search is performed in this grid by successive optimizations to find the best set towards high performance after standard validation/test procedures. The simulation follows an exploration/exploitation approach, i.e., the optimization explores the entire grid of the given hyper-parameters space and then exploits the surroundings of those most promising hyperparameters to attain high performance. This exploitation step can be performed manually if the search space is too constrained.

We have adopted an optimization procedure based on a Bayesian search of a plausible set of hyperparameters representing the structure of a neural net, with random exploitation of the most promising hyperparameters. The design required by the developed neural architectures for the monitoring tasks incorporates improvements to provide better gradient flows and smooth losses,

tighten the hyperparameter search space to a more confined region. For example, in [10], [12] and [11] incorporate skip connections, defined as bypass links between regions of a neural network. These skip connections smooth out the optimization landscape loss, which yields a higher hyper-parameter tuning. When networks are very deep, the loss landscape exhibit nearly chaotic behavior, making the optimization of the architecture very challenging [36]. The skip connections systematically incorporated in [10], [11] and [12] provide smoother optimization procedures, whilst keeping lower generalization error, and hyperparameter optimization feasible. Further, in [10], the U-net structure requires symmetry between the encoder and the decoder [55], to achieve a balance between the learned embedding features in the encoder and the representational capacity of the decoder. Hence, these theorems, along with the structural neural requirements (i.e., symmetry in the encoder), partly explain the neural network design presented in this thesis.

Training deep architectures requires plenty of computational resources and simulation time. For example, in volcanic monitoring applications, it is necessary to perform hyper-parameter fine-tuning based on efficiency and speed. An early-stopping methodology, previously explained, is a method equivalent to the behavior of an A.I. expert who must maximize the performance of a network for the most promising set of hyperparameters, conditioned on limited computational resources. In this context, early stopping has permitted us to finish the training of the architectures, freeing up computational resources, saving time, and increasing efficiency by discarding spurious architectures.

An initial, starting learning rate of 0.01 is declared the starting point of the optimization in all our networks. Early stopping is used aggressively to prune out non-optimal architectures during the hyper-parameter search, whereas we focus on the most promising ones. Also, we build upon previous knowledge of implemented architectures and the problem at hand, such as in [10] where filters are selected to yield good convergence behavior. Finally, in Appendix B, we add additional results of the best architectures we do not prune during the optimization process.

2.6 SUMMARY

This chapter has presented the mathematical background for our thesis, introducing BNNs and the Monte Carlo dropout approach. In addition, we have provided an overview of the best practices in the machine learning community and established the optimization and fine-tuning procedures that we will follow in this thesis. Then, we have mentioned the open-source software pillars implemented and used in this research, ranging from geophysical data applications, array operations to the core deep learning library. The field of deep learning evolves very quickly, and at the time of writing this dissertation, there may be other regularizations or optimization procedures that support more comprehensive fine-tuning of the architectures already presented. However, the employed techniques and the hyperparameters obtained via exploration-exploitation have complied with a proper convergence behavior and have been demonstrated to generalized across different volcanoes and audio domains.

3 VOLCANO SEISMOLOGY

Volcanic eruptions occur when lava, pyroclastic material, and gas are discharged from a volcanic vent due to energetic fluid migration and other phenomena inside Earth [20] [43]. Volcanic activity generates a variety of seismo-volcanic signals recorded by seismometers placed near volcanoes. The presence of these signals is evidence of multiple sources acting within the volcanic system, most of them responsible for or associated with the occurrence of volcanic eruptions. Thus, there is a marked interest in the seismological community to associate different types of seismo-volcanic signals with physical source models to understand the underlying unrest better. This section presents the volcanoes studied, the type of seismo-volcanic events, and the criteria for seismic labeling signals developed in this thesis. We then propose using the uncertainty as a change detector and to discern different types of uncertainty, likely associated with the variability of the seismogram or the monitored process itself.

3.1 SEISMIC SIGNALS

The main objective of volcano seismology is to identify causal relationships between earthquakes and the evolution and outcomes of volcanic unrest. These signals contain information about many aspects of the volcano, including parameters of the seismic-volcanic source (position, geometry, dynamics, energy, or spatio-temporal evolution) or characteristics of the medium (velocity structure, attenuation, or the spatial heterogeneity distribution) [20]. The theoretical categorization of seismo-volcanic signals and how they associate with magma migration remain an open

question. Each seismic signal is believed to represent a distinct source location and mechanism at a given volcano. Seismologists favor signal categorization schemes with detailed descriptions of every possible earthquake sub-type at a specific volcano. This fine-grained categorization yields seismic data nomenclatures specific to individual volcanoes; that is, what holds for one volcano may not be valid for another one. Regardless of the nomenclature used, identifying seismic signals and their evolution through time comprise the basic foundations of forecasting systems. As part of the daily routine at volcanic observatories, seismologists associate the detected seismic events with a *name or label*, using already established criteria and models of seismic sources. By giving a name to the signal, the observer assumes a volcanic state and infers a potential future evolution of unrest.

As scientific knowledge advances, a paradoxical situation has developed: there is a lack of uniformity in naming the observed seismic signals, and new source models are specific to particular cases. The initial subsets of seismic signals have been extrapolated across volcanoes and broadly modified to incorporate a gamut of names and labels that match the theoretical models of seismic sources. In addition, when a volcano exhibits unrest, several volcanic processes can co-exist in time, thus producing a suite of overlapping signals in the raw seismic record that do not belong to the already established criteria [70]. Traditionally, seismo-volcanic signals are classified based on signal attributes. A summary of earthquake types, their frequency, and time-domain characteristics, and source mechanisms proposed in the literature, is presented in [14], and Chapter 3 of this thesis.

3.1.1 LABELLING CRITERIA

From a machine learning perspective, the sparse data taxonomy that characterizes the seismo-volcanic science leads to seismic data catalogs that do not provide a consistent number of categorical samples. The categorical data scarcity is aggravated when volcanoes generate distinct waveforms with respect to the known initial conditions, a consequence presumably related to transitional physical mechanisms. As a result, very specialized training procedures, often condi-

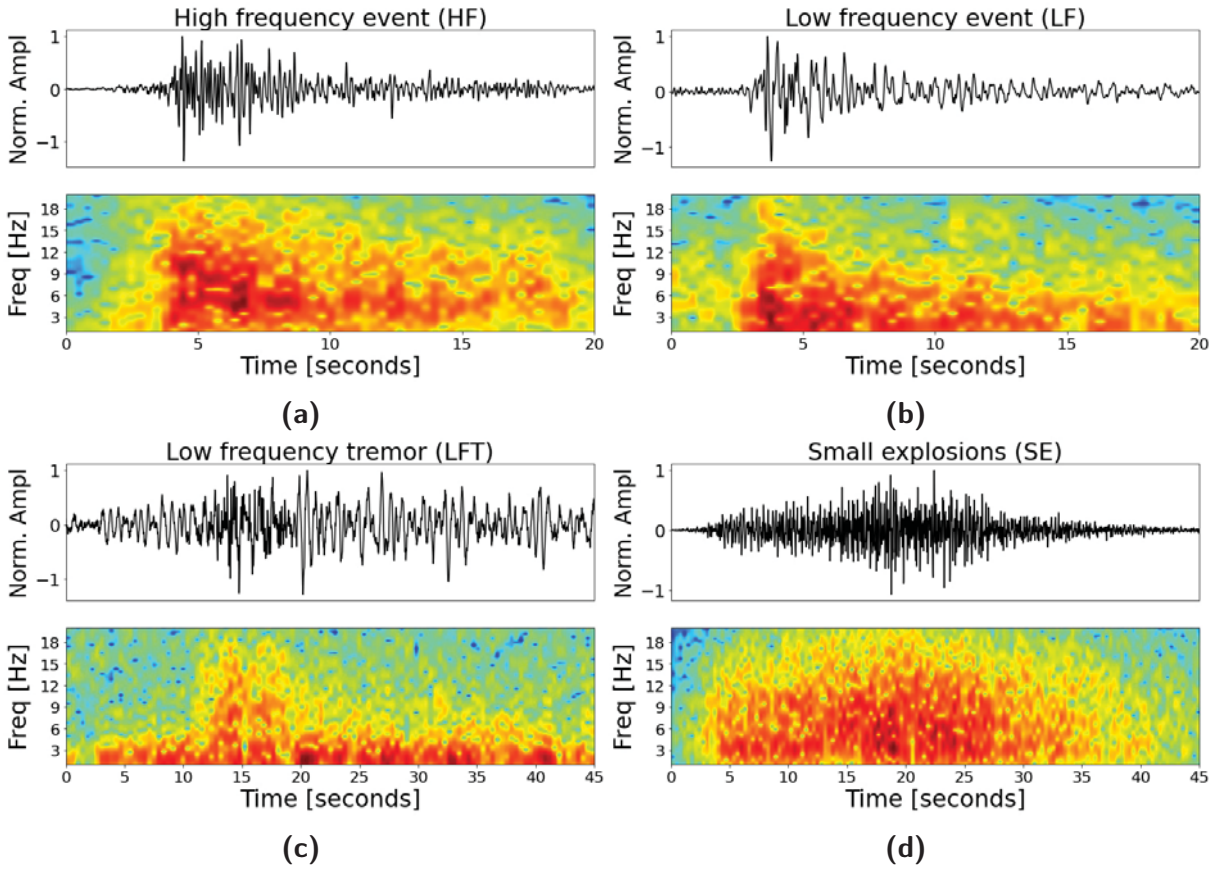


Figure 3.1: Representative examples of common waveforms recorded during the 2007 eruption at Bezymianny volcano. For each waveform type, the normalized waveform (in black) and spectrogram are depicted. (a) High frequency (HF) event (b) Low frequency (LF) event (c) Low-Frequency tremor (LFT) (d) Surficial events (SE) as a small explosion. For visualization purposes, all waveforms have been filtered between 1 and 20 Hz.

tioned with in-depth analysis of the predictions, are the recommended best practices, often implemented as mandatory operational requirements [23]. Over the recent years, seismo-volcanic monitoring has favored adopting higher-level data taxonomies that could benefit near-wavefield, quasi-real-time monitoring without losing volcanic dynamics information. The labeling scheme in this thesis is based on foundation work by [42]. We refer the reader to Chapter 4 for a complete summary of labels, categories, and potential sources associated with these labels.

The data labeling scheme in [42] comprises five of the most common seismo-volcanic categories, based on the frequency range of events. For example, the Alaska Volcano Observatory (AVO), a world-class example of effective information data acquisition related to volcanic activ-

ity, employs this data taxonomy as the distilled recognition information for volcano-monitoring in real-time, for more than 20 volcanoes [43]. Also, this categorical scheme has been adopted to monitor Telica (Nicaragua) [54], Ngauruhoe (New Zealand) [49], or Mount Pinatubo (Philippines) [53]. The data labeling scheme adopted in this thesis has also been employed to study the relationships between magma migrations and earthquake dynamics for the volcanoes considered in this thesis. Previous research work by [63], [68] and [27] summarized the eruptions at Bezymmiany and Mount Saint Helens with these labels, which have helped to build and check the geophysical foundations of this work. Thus, we can categorize our signals as:

1. **High-frequency events (HF)**: Also known as volcano-tectonic earthquakes (VT) or A-type earthquakes. These events are characterized by a broad frequency range (from 1 to up to 30 Hz) and are often associated with brittle failure processes occurring locally within volcanic systems. They include similar mechanisms to those of classical tectonic earthquakes. In general, they provide evidence of changes in the local stress regime during magma migration at depth. Several factors influence the timing, character, and occurrence of HF/VT seismicity, including local tectonics, cooling of magma inside of volcanic edifices, gravitational settling of the volcanic edifice, unstable structures, and pressure changes (associated with volcanic fluids, hydro-thermal interaction, or fluid injection) in the subsurface (that is, hydraulic fracturing processes related to stress propagation due to magma movement).

2. **Low-frequency events (LF)**: This category comprises seismic events named long-period events (LP), long coda events, and B-type earthquakes. This type of earthquake concentrates energy in the 0.5–5 Hz band, with typical durations of approximately 5 to 60 s. These events are thought to be associated with resonances or fluid transport in the volcano plumbing system or fluid-driven cracks or pressurization processes (bubbles), among other processes. Theoretical models consider LF events as precursors to eruptions, given their connection to internal volcanic activity associated with fluids (mostly interactions between magma and underground water).

3. **Hybrid Events (HYB)**: The energy of these earthquakes straddles the spectra of both LF and

HF events. They are often explained by a mixture of processes, including fluids filling rocks and the generation of cracks.

4. **Volcanic Tremor (LFT)**: Volcanic tremor is a complex volcanic signal. Volcanic tremor is a sustained seismic signal with a duration from minutes to days with no clear onset of seismic phases. Its origin and frequency content is diverse and can be associated with the volcanic system's internal or external processes. In many cases, the tremor shares the same frequency band with the LF events, and some geophysical studies have associated TRE with swarms of LF events recorded in a time span of seconds. At Bezymianny, tremor has a low frequency, and thus we adopted the terminology of [63], in which the tremor is renamed a low-frequency tremor (LFT). Although different models have been proposed to explain the source mechanism of this type of tremor, most concur that its origin is the complex interplay between magmatic or other fluids and their host rocks.

5. **Surficial Events (SE)**: Surficial events are so named because the origin of their source seems to be located in the summit zone. According to their waveform and spectral content, these events share the characteristics of small summit explosions, often associated with moderate degassing processes with a high content of water steam and rockfalls. Hence, in this category, we gathered those signals associated with external effects such as pyroclastic flows, rockfalls, and other processes that could offer insights into impending volcanic activity. In Bezymianny volcano, they are present as cigar-shaped rockfalls, with very high-frequency components and long duration (> 50.0 seconds).

3.2 VOLCANOES STUDIED

The number of volcanoes that dot the planet Earth is enormous, forming volcanic belts that span from the ocean to the surface, including vast geographic areas with volcanoes very close to each other [70]. Volcanoes can be categorized or grouped according to related features, for example, similar chemical composition or eruptive style. From the perspective of probabilistic deep

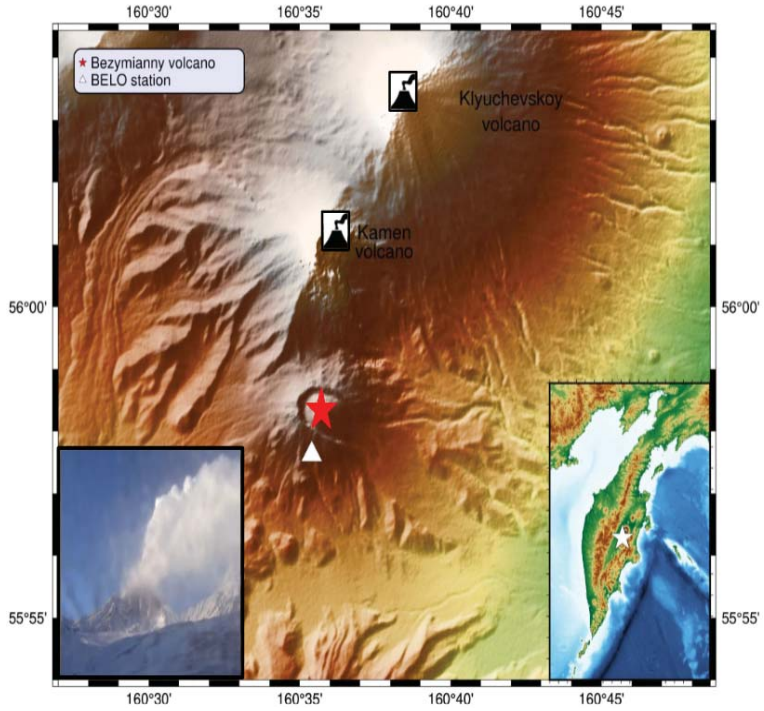


Figure 3.2: Geographical area of Bezymianny volcano, with an ash emission of $\approx$ 15 km above sea level. The red star represents the eruptive center of the volcano. The white triangle corresponds to the coordinates of the BELO station.

learning, we define the entire set of recorded seismic data streams as the known support data distribution for a given volcano and monitoring time period. Therefore, this initial support data distribution is defined and conditioned by the geophysical characteristics of the recorded seismo-volcanic data streams, which contain the events detailed in subsection 3.1.1, explicitly related to the volcano that generates them. In this thesis, we focus our analysis on three volcanoes: Bezymianny (Kamchatka, Russia), Mount Saint Helens (Washington, USA), and Mount Etna (Sicily, Italy).

3.2.1 BEZYMIANNY VOLCANO

The Bezymianny volcano is located on the Kamchatka Peninsula (Russia), within the Klyuchevskoy Volcanic Group (KVG). Figure 3.2 depicts the geological area of the Bezymianny volcano within the KVG group. [1]. The research works in this thesis, [10], [12], [11] are focused on a three well-

[1]Picture reproduced with permission of the Kamchatka Branch of FRC EGS RAS

known 2007 eruptions, over a period of approximately three months [63], [67], [66]. The three eruptions, 25th September, 14th-16th October, and 5th November are brief but energetic, covering increased pre-seismicity rates and various eruption mechanisms. These three significant eruptive episodes were recognized and reported by the Kamchatka Volcanic Eruption Response Team (KVERT) [27], and confirmed by posterior geophysical studies. Of the three eruptive episodes dominated by strong ash explosions and lava emissions, it is the second one that is considered the most energetic and relevant, lasting two full days, with ash plumes reaching 1000 km southeast and a plume height of 10 km [68]. We subdivided our dataset according to the reported eruptive chronology in [27] and [66]. From the deep learning perspective of our framework and drawing parallels to neighbors' acoustic domains, we strictly maintain the chronology of all the seismic data for the definition of the available training data and the temporal estimation of the uncertainty evolution.

3.2.2 Mount Saint Helens

Mount Saint Helens is located in Skamania County, within the Cascade Mountain Range of Washington, United States. For this volcano, we have selected seismic data from the eruptive sequence from 2004-2006 [32]. Previous foundation research has established seismological comparisons between Mount St. Helens and Bezymianny volcano [46]. Preliminary geophysical analysis at these two volcanoes has demonstrated similar seismic patterns before explosive activity, with very similar waveform characteristics in amplitude and duration [68]. Compared with the Bezymianny, this volcano is not significantly different in terms of external geometric shape; however, seismic sequences in this volcano occur with a very high event rate. In the context of our thesis, these similarities translate into a different scenario, with similar data distributions (although conditioned to the geological context of this volcano), to test the generalization and exportability capabilities of the implemented system on another volcano.

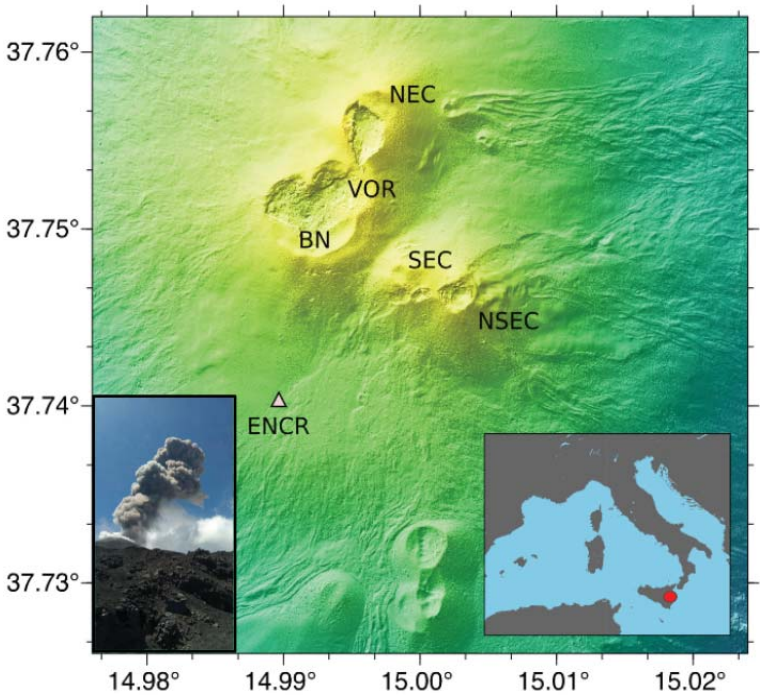


Figure 3.3: Geographical area of Mount Etna volcano, with a photography of the 2019 paroxysm. The white triangle corresponds to the coordinates of the selected station, ENCR.

3.2.3 Mount Etna

The internal morphology of Mount Etna is frequently changed due to intracrater volcanic activity [4], [18], [58], [52], [26], which provides a very challenging scenario for our approach to detecting change with estimated uncertainties. Figure 3.3 depicts the geographical area of Mount Etna, the ENCR station selected in the study of uncertainty exportability [12], and a picture of the eruption [2]. We use data gathered at the southeast of Mount Etna (Bocca Nuova, BN) during the paroxysm recorded from 4th July to 24th July 2019 at the ENCR station. The primary eruption was reported by seismological bulletins and visual observations on 18th July 2019, at 23:09 UTC time. The significant changes in eruptive activity and correlation with other volcanic data are described in [22].

[2]From VOSSIA research project, courtesy and reproduced with permission of Prof. Silvio de Angelis, Dr.Luciano Zuccarello, and Dr. Alejandro Diaz-Moreno

3.3 LABELING PROCEDURE

Once the labeling criteria and the volcanoes selected for this thesis have been introduced, we describe the labeling and the data curation procedures followed in this thesis. This data curation methodology has been applied to the three volcanoes in this thesis, Bezymianny, Mount Saint Helens, and Mount Etna, to extract seismic events from unrest periods in which no primary labeled data were available. The software tools implemented and used for these tasks are included, along with a detailed description, in chapter 8.

At present, a wealth of available seismic data acquired at volcanoes worldwide remain vastly underutilized. Often, the myriad of data formats, differences in the sampling frequencies across streams, scientific controversies in the labeling, or the temporal organization of the data implies that performing the most basic machine learning training procedures can be challenging. These difficulties can be worsened by strict, non-inclusive data policies complied by observatories subject to government rules that share minimal data (or none). Recently, data platforms such as *Incorporated Research Institutions for Seismology* (IRIS) ingests and distribute a broad range of geophysical data, ranging from ground motion to atmospheric data. They provide a set of principles and a range of practices to access and manage seismic data for the global earth science community, free of cost or other access barriers. However, for volcano-seismic data, no machine learning metadata nor geophysical interpretation is provided by IRIS. The continuous data streams recorded in a seismogram need to be organized into a dataset with meaningful labels that correlate with volcanic activity. Here, we supply means to produce, from raw seismic data streams available at IRIS, relevant information on said data.

Remark that before applying these labeling procedures, it is a mandatory step to conduct a literature review to match the IRIS available data with studied and interpreted research works or volcanological observatory bulletins. This thesis has checked that the obtained metadata using the labeling procedures explained here concurs with previous seismological studies. For the selected eruptive period in 2007, Bezymianny volcano, we followed the classification criteria and

the chronological eruptive process reported by the Kamchatka Volcanic Eruption Response Team (KVERT) [27] and research work on this volcano [66], [63]. For Mount Saint Helens volcano, our segmentation is based on the seismological bulletin by [46]. Finally, for Mount Etna, we selected a recent eruption, which is fully described in [22]. We divided the labeling procedure of our datasets into two primary steps:

1. Automatic segmentation and frequency-based categorization with a waveform descriptor, as proposed in [13].

2. A manual inspection of the events confirms and modifies the labels or the segmented times derived from the automatic procedures [15].

Both methodological steps are interchangeable since we can use automatic or manual segmentation depending on the requirements to generate the catalogs. In the following sections, we will rigorously explain each one of them.

3.3.1 AUTOMATIC SEISMIC EVENT SEGMENTATION

The raw data is pre-processed to extract events of interest from the seismic data stream using the REMOS (Recursive Entropy Method of Segmentation) package [13]. The algorithm REMOS is a complete workflow for the detection, segmentation, classification, and visualization of seismic data. Starting from the continuous data streams and the activation times of STA/LTA, REMOS identifies a set of energy and entropy-based criteria to parse data into individual waveforms (segmentation). Then, each segmented waveform is then categorized according to the frequency content, using well-established frequency domain metrics [17]. REMOS also incorporates an exploratory data visualization tool based on t-Distributed Stochastic Neighbor Embedding (t-SNE), and frequency attributes [37]. The steps of the segmentation are as follows:

1. **Preparing the data:** Preliminary processing steps are applied to the continuous data before segmentation. Trends associated with effects such as very long-period instrument drift are removed. A band-pass filter in a user-selected frequency range $[f_{low}\ f_{high}]$ $(Hz.)$ is applied

to enhance the presence of earthquake signals. We observed that the frequency band [1 – 15] Hz is appropriate for most volcanoes during our tests.

2. **Building the activation vector:** The filtered data are then scanned using a recursive STA/LTA algorithm to obtain trigger times of potential events. The parameters for the STA/LTA algorithm are selected by the user according to the specific dataset [65]. For volcano-seismic signals recorded at a relatively close distance from the source ($< 10km$), a short-term window of $0.5 – 2s$ and a long-term window of $8 – 15s$ are, frequently, appropriate. This procedure yields a one-dimensional activation vector, **on** = $[on_1, on_2, on_3, ..., on_n]$ of n earthquake trigger times.

3. **Segmenting the data:** The vector of activation times, **on** is used in combination with the filtered signal to investigate regions within the continuous data stream that contain the signal of interest. REMOS defines two parameters, the maximum search window, W_s, and the minimum duration window, W_d. The maximum search window W_s represents the time window that REMOS explores to detect an event; W_d captures the minimum duration that REMOS uses to calculate signal energy. REMOS, then, considers all activation times in **on**, and extracts segments of data (exploration regions) with duration W_s starting at these times. Each of these regions is then windowed into k frames with duration W_d. A noise reduction procedure is applied to the data stream to mitigate the influence of external noise sources. The energy of the signal is computed according to all exploration regions for each individual frame. The entropy, $H(p)$, is then calculated from the pseudo-probability normalized vector, p, as:

$$H(p) = \sum_{i=1}^{k} -p_i \, log(p_i) \tag{3.1}$$

where p_i is the normalized energy of each frame. If the entropy is below a pre-defined threshold, $\epsilon < 2.5$, then the onset of an event is declared. In this instance, the candidate event is extracted from the exploration region and parsed into an individual waveform according

to the minimum of the energy distribution criteria:

$$\arg\min_{k} E_k \tag{3.2}$$

The end of an event is declared when the energy reaches a minimum. To overcome long segmentations when taking the lowest energy, REMOS computes, in parallel, the ratio R of the energies of the initial and final parts of the segmented signal. This ratio would be close to zero if no earthquake signal were present in the final section of the segmented data. In that case, the original candidate is segmented again. Otherwise, for ratios closer to one, the original candidate is selected, as it could be, for instance, an episode of long-duration tremor. Once the candidate is segmented, results are stored in a matrix. REMOS continues segmenting signals along the main trace until the activation vector *on* is exhausted.

The segmented candidate and the starting and end times are stored in a matrix; row-wise. As a final step, candidates with low SNR can be erased from the final dataset. In REMOS, the SNR is computed as the peak amplitude ratio within a time window containing the surface wave signals to the root-mean-square of the noise trailing the signal arrival window.

3.3.2 Frequency index event categorization

The individual traces extracted from the continuous data are then automatically classified based on a frequency metric. This metric, known as the frequency index (FI), significantly reduces the time required to analyze data, and any mismatch in event classification due to the bias introduced by a human analyst is potentially eliminated [17]. The label of an event can be assigned using the FI, formally defined as a logarithmic interpretation of the spectral frequency ratio (FR), given as:

$$\mathrm{FI} = \log_{10}(\mathrm{FR}) = \log_{10}\left(\frac{A_{high}}{A_{low}}\right) \tag{3.3}$$

where A_{low} and A_{high} are the mean amplitude of high and low spectral bands, respectively.

REMOS defines a set of parameters as adjustable thresholds that control how events are categorized. Note that the logarithmic spectral ratio forces high values of the FI to be associated with higher frequency mechanisms (e.g., brittle fracture). In contrast, lower values correspond to seismic events with narrow lower frequency bands (e.g., soft ruptures [2]). The logarithm is appealing as waveforms with similar amounts of high and low energy (as defined by the frequency range chosen in A_{low} and A_{high}) will have a FI closer to zero. Events are classified as Low-Frequency (LF) if the FI is below a given threshold η_1. Similarly, if the FI is greater than the given threshold η_2, earthquakes are classified as High-frequency (HF). The range of values between η_1 and η_2 could represent hybrid events. Low-frequency tremors (LFT) and surficial events (SE) are discriminated according to their frequency index and duration t. The minimum duration t over which to consider an event, either low frequency or tremor, depends on the user and the volcanic environment [44] or can be selected according to past research studies on the given volcano [63].

We adopted the labeling criteria of [42], explained above in subsection.3.1.1. Unrest at Bezymianny during 2007 and the dataset studied here have been previously investigated by [63] [66], thus providing a reliable benchmark for the performance of REMOS and geophysical insight for the configuration of the REMOS parameters. Each extracted event was labeled according to the FI value ($FR = 0.5$) and duration of high- (HF) and low-frequency (LF) events, seismic background tremor (SBT), and debris processes (DP). The selected A_{low} and A_{high} spectral bands for this volcano are $[1-5]Hz$ and $[6-12]Hz$, respectively.

REMOS stores the segmentation results in a matrix, **S**, whose rows contain, in order: the segmented event, the FI, the *start* time, the *end* time, and the segmentation label. This matrix **S** constitutes the data seed on which recognition systems can be trained. This type of matrix structure obeys, on the one hand, a pragmatism oriented towards seamless integration with deep learning models; that is, waveforms can be extracted quickly. For continuous monitoring systems, start and end times can index the seismic events and allow continuous recognition training. On the other hand, any necessary waveform parameterization can be computed, and the matrix expanded according to operational requirements. For example, the matrix computation allows obtaining parameters not only on the entire raw seismic data stream but by families of events. Finally, for an

eruptive period in a given volcano, the set of all these matrices indexed by the day of the current year (0-365), $\mathbf{S}_{day}$, constitute the dataset that we will use to study the volcanic dynamics.

3.3.3 Refining segmented datasets

The features computed by REMOS are connected with the segmentation boundaries of events. However, due to the statistical nature of seismo-volcanic data streams, the *correct segmentation* is still not guaranteed. In this context, to ensure proper time boundaries, we extended REMOS with a Graphical User Interface (GUI) to refine the $\mathbf{S}_{day}$ matrices.

The custom GUI, named PICOSS (*Python Interface for the Classification of Seismic Signals*) permits the inspection, confirmation, and modification of the *start* and *end* times of events, along with the labels [15]. Further, PICOSS provides a collection of tools to refine seismic meta-data, designed to reduce data labeling fatigue while increasing data classification efficiency. For each day in the dataset, a team of domain experts visually verifies each extracted event and its label. The result is a complete and annotated catalog. Manual inspection of the seismic trace is required on the main eruptions, given the complexity and multiple processes involved. In Chapter 8, we describe the capabilities of each implemented software.

Once the seismic classes and the selected volcanoes have been established, we can formulate the Bayesian monitoring framework for continuous monitoring and drift detection from pre-eruptive seismicity.

3.4 Monitoring of volcanoes

Identification of seismic events is currently already carried out with very efficient supervised Machine Learning techniques [5], [64], [21], [31], [39], [33], [50], [56]. However, identifying signals by itself does not provide information about the dynamics of the processes that lead to eruptions. The proposed approach departs from handcrafted features for detecting change [35], [29], [47], [7] and mitigates that subtle changes in events that are critically important are possibly hidden from

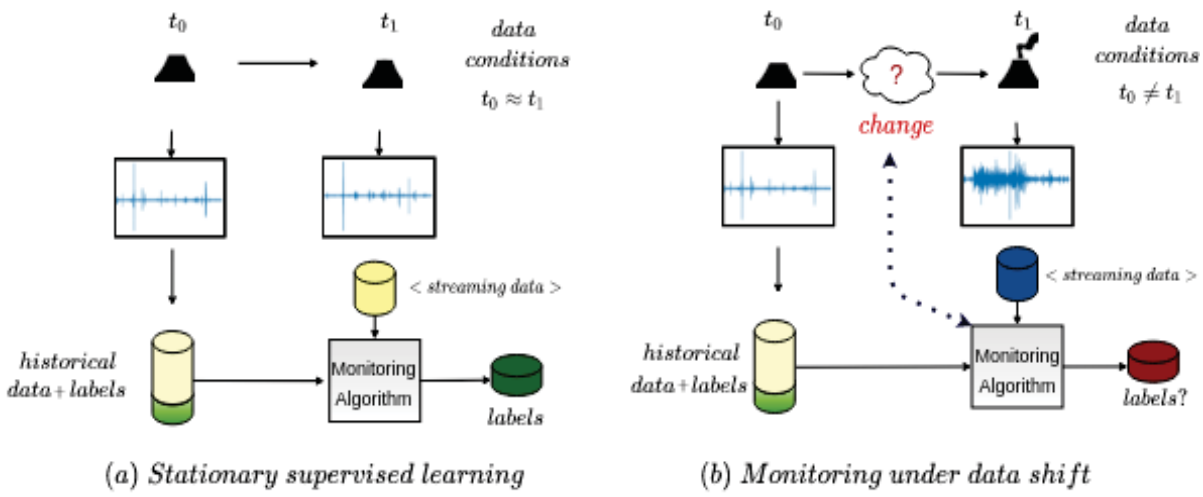


Figure 3.4: The traditional seismo-volcanic monitoring architectures are based on static supervised learning, assuming that the initial situation at t_0 would be similar to the one later encountered in t_1. Hence, in (a), a data catalog of historical seismicity is used for routine monitoring of volcanic unrest. In reality, case (b) highlights when an unforeseen change can induce a data shift that compromises the performance of the model and the monitoring task. The work on this thesis develops the monitoring algorithm to detect such change, from pre-eruptive seismicity, with short-term forecasting applicability.

standard data analysis.

In the monitoring problem, seismic data arrives in real-time as a continuum data stream. Given the space-time evolution of magma ascent and other processes involved during an eruption, it is expected that the recorded data stream evolve, for example, as steady increments in the energy of volcanic tremor or subtle shifts in the frequency components of the signals. Over past years, there has been a growing interest in probabilistic forecasting and uncertainty quantification methods to provide direct knowledge about new seismic anomalies, which extends to a scientific understanding of seismic sources variation [48], [59], [60]. Further, research insights by [61] emphasize the necessity to shift from deterministic to probabilistic approaches to enhance monitoring outcomes.

Figure 3.4 depicts the monitoring problem from a monitoring perspective and highlights the need for a probabilistic machine learning approach. The accepted practice is to assume that the data used to fine-tune monitoring systems (t_0) represents the situations later encountered (t_1). However, in practice, that assumption does not hold.

The analysis of figure 3.4 had summarized the key concepts presented in the research work of

this thesis and highlighted the importance of uncertainty estimation. Monitoring situations such as those in figure 3.4.b can benefit from probabilistic approaches, helping in the systemization of early warning protocols while providing transparent hazard information to local authorities or decision-makers. Beyond interpretability of the monitoring algorithms with estimated uncertainties, one main challenge remains: how we detect the change? Is it possible to capture variations in anticipation of eruptions (dotted line in figure 3.4.b)? We seek to supply our deep learning models with the capabilities to discover unforeseen seismic patterns, the *unknown unknowns*.

3.4.1 UNCERTAINTY AND MONITORING

Following the key concepts introduced in Chapter 2, seismo-volcanic monitoring can be framed as a non-stationary environment with seismic data streams subject to unforeseen yet persistent wavefield perturbations. From a probabilistic perspective, this is equivalent to temporal shifts with respect to the initial support data distribution. For this purpose, the change could be known ahead of time if the monitoring algorithm can detect particular environmental factors that condition the new situation.

Therefore, we can use the uncertainty to control the quality of the physical measurements of a volcanic system and as an indicator of the evolution of these physical measurements over time. We propose that the monitoring algorithms and the seismic data must be used in symbiosis to address unforeseen inconsistencies in raw observations. This distilled information can be used as surrogate knowledge to interpret unforeseen monitoring situations that could not be derived from raw observations otherwise. Formally, we define the joint data distribution at the initial monitoring time t_0, as $p_{t_0}(\mathbf{X}, \mathbf{Y})$, and the conditional predictive distribution $p_{t_0}(y^*|x^*, \mathbf{X}, \mathbf{Y})$ (see equations (2.1) and (2.15)). A data drift between time t_0 and time t_1 can be defined as:

$$\exists \mathbf{X} : p_{t_0}(\mathbf{X_0}, \mathbf{Y}) \neq p_{t_1}(\mathbf{X}, \mathbf{Y}) \tag{3.4}$$

with $p_{t_1}(\mathbf{X}, \mathbf{Y})$ the new distribution. Changes in data can be characterized as changes in the components of this mathematical relation. In this regard, considering the classification scheme of

the seismic signals remains invariant [42], the volcanic changes would be perceived if changes in the underlying distribution of the seismic data, $p(\mathbf{X})$, can be perceived by the parameters of the model. The estimated uncertainty U_t can be used to detect *out-of-distribution* data samples, and thus, quantify the migration from $p_{t_0}(\mathbf{X}, \mathbf{Y})$ to $p_{t_1}(\mathbf{X}, \mathbf{Y})$. Such value U_t can be computed from the Bayesian uncertainty framework derived in Chapter 2. The most striking quality of this framework implies that monitoring volcanoes with uncertainty detection do not entail any supervision, except for the initial fit in which the model is conditioned. Therefore, the detection of changes is an additional parameter that the proposed Bayesian approach provides in an unsupervised way, allowing us to discover patterns of change within the seismic data streams.

In Chapters 4 and 5, we show that a monitoring algorithm can probe the estimated U_t as a proxy to detect if the recorded seismograms depart from $p_{t_0}(\mathbf{X}, \mathbf{Y})$. To known what is driving change, we subdivided the total statistical uncertainty U_t into two terms, the observed seismogram variability and the inherent randomness of the monitored seismic wavefield:

$$U_t = U_{src} + U_{wav} \tag{3.5}$$

where U_{wav} is the uncertainty associated with the seismic wavefield, and U_{src} is the uncertainty linked to the monitoring process. This equation (3.5) links with the concepts of epistemic and aleatoric uncertainties. From a monitoring perspective, we can link the geophysical uncertainties in (3.5) to the statistical ones in equation 3.4. When different seismograms are recorded, increments in U_{src} and U_{wav} must be observed, and thus, lead to the assumption that the volcano has changed.

In chapter 6, we extend these ideas and build a very specialized network to detect the change from the raw waveforms and measure the exportability of the estimated U_t to other volcanoes. Chapter 7 exploits the uncertainty in active learning setups to build an adaptable monitoring system to perform uncertainty-based sampling detection and classification of seismo-volcanic events.

3.5 SUMMARY

This chapter has presented the seismological context of this thesis. First, we have identified the main challenges and how these affect machine-learning-based seismo-volcanic monitoring. Then, we overcome these problems by presenting a self-contained signal processing workflow to perform detection, segmentation, and classification of seismo-volcanic signals with a higher-level data taxonomy. We then present the monitoring uncertainty framework and the association of statistical uncertainties to the raw waveform. Finally, we introduce the concept of data drift and its implications in volcano-monitoring science.

4 | Volcano-seismic transfer learning and uncertainty quantification with Bayesian neural networks

This chapter is devoted to the classification of seismo-volcanic events across eruptions and volcanoes. We also analyze how the epistemic uncertainty can be used as a probability threshold that, if exceeded, justifies the use of transfer learning procedures. The first explorations about uncertainty quantification as a change detector are also illustrated. This article is published and available online at IEEE Transactions on Geoscience and Remote Sensing, with the following journal metrics: (IF), JCR 2020: (5.855). Remote Sensing (Rank 5/32) (Q1). Geochemistry and geophysics (Rank 5/85) (Q1). Imaging Science and photographic technology (Rank 5/27) (Q1). Engineering, electrical and electronic science (Rank 27/266) (Q1). We reproduce the draft accepted by the journal, which can be cited as:

1. **A. Bueno**, C. Benítez, S. de Angelis, A. Díaz-Moreno and J. M. Ibánez. Volcano Seismic Transfer Learning and Uncertainty Quantification with Bayesian Neural Networks. IEEE Transactions on Geoscience and Remote Sensing, vol. 58, no. 2, pp. 892-902, Feb. 2020.

Volcano-Seismic Transfer Learning and Uncertainty Quantification with Bayesian Neural Networks

Angel Bueno*, Carmen Benítez*, Silvio De Angelis†, Alejandro Díaz Moreno†, Jesús M. Ibáñez‡

*Department of Signal Theory, Telematic and Communications, University of Granada, Spain

†Department of Earth, Ocean and Ecological Sciences, University of Liverpool, UK

‡Instituto Andaluz de Geofísica, University of Granada, Spain

Abstract

Over the past few years, deep learning has emerged as an important tool in the fields of volcano and earthquake seismology. However, these methods have been applied without performing thorough analyses of the associated uncertainties. Here we propose a solution to enhance volcano-seismic monitoring systems, through probabilistic Bayesian Deep Learning; we implement and demonstrate a workflow for waveform classification, rapid quantification of the associated uncertainty, and link these uncertainties to changes in volcanic unrest. Specifically, we introduce Bayesian Neural Networks (BNNs) to perform event identification, classification, and their estimate uncertainty on data gathered at two active volcanoes, Mount St. Helens, USA, and Bezymianny, Russia. We demonstrate how BNNs achieve excellent performance (92.08 %) in discriminating both the type of event and its origin when the two datasets are merged together and no additional training information is provided. Finally, we demonstrate that the data representations learned by the BNNs are transferable across different eruptive periods. We also find that the estimated uncertainty is related to changes in the state of unrest at the volcanoes, and propose that it could be used to gauge whether the learned models may be exported to other eruptive scenarios.

I. Introduction

Over the past two decades, the integrated use of methods and techniques from different disciplines including ground deformation, geochemistry, satellite remote-sensing and seismology,

has allowed scientists to identify and track volcanic unrest with increasing confidence. Due to comparatively low costs and the availability of real-time data with high temporal resolution, seismology remains the backbone of most volcano monitoring programmes worldwide. Volcano-seismic signals are frequently classified based on their frequency content, and source type; one of the most widely adopted classification schemes ([1]) includes high-frequency (also known as volcano-tectonic), low-frequency earthquakes, mixed-frequency (or hybrid) earthquakes, volcanic tremor, explosions, and other superficial signals (e.g. rockfalls, lahars, pyroclastic flows). Low-frequency and mixed-frequency earthquakes have been attributed to various mechanisms, which include volumetric sources, magma fracture, stick-slip along the margins of volcanic conduits, and slow-rupture of soft material. It has also been shown that the characteristic lack of high-frequency energy in their waveforms may result from propagation through strongly attenuating volcanic material [2], [3]. Separation of source and path effects remains a challenging task. High-frequency earthquakes are nearly unanimously attributed to brittle failure processes locally within volcanic systems with mechanisms similar to ordinary tectonic events, hence, the frequently used name of volcano-tectonic. Volcanic-tremor is a continuous signal, at times with harmonic frequency spectrum, which is recorded during periods of either eruption and non-eruptive unrest. Explosion earthquakes are high-amplitude, short duration, pulses associated with the sudden and violent ejection of gas and pyroclastic material from volcanic vents into the atmosphere [4], [5], [6], [7]. A summary of earthquake types, their frequency- and time-domain characteristics and source mechanisms proposed in the literature, is presented in Table I.

One of the main goals of volcano seismology and volcanology is to identify causal relations between the occurrence of earthquakes, and the evolution and outcomes of volcanic unrest. Success depends, clearly, on the ability to identify and track the evolution of earthquakes during periods of volcano-seismic unrest. At present, although large amounts of seismic data are continuously gathered at volcanoes worldwide, much of these data remain underutilised. Seismic analysts typically focus on comparatively small subsets of earthquakes. During volcanic crises, seismic networks can record earthquakes at rates of up to multiple events/minute over time periods as long as months, or even years, making manual identification, classification and location an unfeasible task [3]. Earthquake classification is often subjective, based on human experience, and analysis of a small fraction of the available data may result in a partial and biased interpretation of unrest. For instance, it has been shown that empirical methods used to forecast volcanic eruptions, may fail due to the incompleteness of the seismic catalogues [8].

Over the past decade, large computational resources have become more widely available at reasonably low cost, and machine learning and advanced signal processing algorithms have emerged as tools for use in volcano-seismic monitoring [9] [10]. This progress is parallel to advances in the investigation of micro-seismicity [11], earthquake detection [12] and other fields of the Earth Sciences [13]. A wealth of literature exists on automatic detection and classification of volcano-seismic signal, including algorithms based on signal properties [14], dimensionality reduction [15], embedding vectors [16], Gaussian Mixture Models (GMM) [17], Hidden Markov models (HMMs) [18] [19] and Artificial Neural Networks [10]. A large body of research has also explored the application of Deep Learning (DL) algorithms for identification and classification of earthquake signals. In particular, unsupervised training has been shown to be effective for use with Deep Neural Networks (DNN) [20]. Recurrent Neural Networks (RNNs) have also been used for the classification of volcano-seismic data streams [21]. These workflows have, however, been applied without performing any assessment of their uncertainty. Reliability in adverse conditions is an important consideration for real-time applications. Many investigators have focused their efforts on improving earthquake classification in order to increase the accuracy of pattern recognition methods. Quantification of uncertainty, on the other hand, can provide direct knowledge on the diversity of source mechanisms, which ultimately could inform scientific interpretations of volcanic unrest [22]. In this paper, we propose a workflow for the classification of volcanic earthquakes, enhanced by the integration of a Bayesian framework that could provide fast uncertainty quantification at the seismic waveform level. In particular, we explore the use of Bayesian Deep Learning, which combines the flexibility of Bayesian theory with the computational advantages of deep learning, allowing rapid and robust Bayesian inference. This probabilistic framework is applied to seismic data streams from two volcanoes, *Bezymianny* (Kamchatka, Russia) and *Mount St. Helens* (Washington, USA). St. Helens and Bezymianny are two examples of active strato-volcanoes with similar andesitic composition and morphology, located in similar tectonic environments, and characterized by intermediate-to-high explosivity [3]. These volcanoes are excellent candidates to asses whether the patterns detected by Bayesian Neural Networks (BNNs) on one volcano, could be exported to other eruptive scenarios. Our initial experiments are directed to understand the behaviour of BNNs on these two separate volcanoes. After that, we focus our effort into one unified framework that is trained jointly, under label sparsity conditions (different labels for each volcano), but same categorisation scheme. These experiments would help to evaluate how BNNs classify

events that are related in frequency content but distinct in seismic nature. For the new eruptive periods, we perform an uncertainty analysis on the new data ranges to evaluate if uncertainty estimations could be interpreted as a typical feature associated to the lack of specific knowledge about the new volcanological situation. The proposed workflow can capture this information, yielding higher uncertainties estimates and reduced recognition accuracy on later eruptive periods. The framework is enhanced by exploring the complementarity between transfer learning techniques and uncertainty quantification. Specifically, we examine passing prior weights to the new seismic period to assess if further improvements can be obtained by merging knowledge at multiple scales. Our framework allows the seismological community to tackle the problem of data scarceness and demonstrates the robustness of the proposed approach with more extensive seismic catalogues, and on different volcanological conditions.

II. Bayesian Deep Learning

A. Deep Neural Networks

Artificial Neural Networks (ANNs) are mathematical algorithms designed for function approximation. We define $D = \{(\mathbf{X}, \mathbf{Y})\} = (x_i, y_i)_{i=1}^{N}$ as our dataset containing a collection of N recorded seismic signals, x_i, along with their annotated labels y_i. The output of an ANN, noted as y, is computed through a non-linear transformation (hidden layer) of the input data x. ANNs work well on well-defined problems but lack the flexibility of modern deep learning techniques to discover statistical regularities in high-dimensional datasets [23]. Deep neural networks (DNNs) are defined as sets of fully connected hidden layers, $f(\cdot)$, in which the output $y = f^W(x)$ is parameterized by $w = (w_1, w_2, ..., w_n)$, known as weights. On multi-class classification problems, class probabilities p_c are derived from the output layer of the DNN as:

$$p_c(y = i|x, w) = \tilde{f}(x_i; w) \tag{1}$$

with $\tilde{f}$ the output of the *softmax* probability layer. The softmax layer is defined as a normalized exponential function which computes class probabilities $p(c)$ from the last layer output, o:

$$p_c = \frac{exp(o_c)}{\sum_k exp(o_k)} \tag{2}$$

where k is the index over all classes and exp, the exponential function. The training of a deep model typically consists of finding the optimum set of weights that maximizes the likelihood

distribution, $p(y|x, w)$, that best explains our observable data. This weights optimization is computed via the backpropagation algorithm by measuring the discrepancies between the labels and the predicted outputs [24].

B. Bayesian Neural Networks

In a classification problem, the probability output of *softmax* layer alone (equation 2) could lead to over-confident predictions for points out of the data distribution. In this context, Bayesian neural networks (BNNs) are defined as *"artificial neural networks in which a probability distribution is placed over the network weights"* [25]. BNNs do not compute a single estimate of the weights w, but a probabilistic approximation over all of them. This approximation allows a rigorous approach to tackle statistical approximation problems. Given a volcano-seismic dataset, D and the likelihood $p(y|x, w)$, the posterior distribution of the network weights, $p(w|D)$, can be approximated using Bayesian inference:

$$p(w|D) = \frac{p(y|x, w) * p(w)}{p(y|x)} \tag{3}$$

With $p(y|x)$ known as the *evidence* and $p(w)$ the prior distribution over the weights, on a vector space $w \in \Omega$. The predictive distribution is computed as:

$$p(y^*|x^*, D) = \int_{\Omega} p(y^*|x^*, w)p(w|D, w)\ dw \tag{4}$$

With x^* and y^* the new input and output, respectively. The computation of equation 4 requires the evaluation of an intractable integral. First work by [26] described a Bayesian inference framework based on a Laplace approximation of the posterior. Work by [25] introduced Hamiltonian Monte Carlo (HMC), an integration of Markov Chain Monte Carlo (MCMC) and Hamiltonian dynamics to sample from the posterior distribution. However, the amount of time required for computation limited their applicability to volcano-seismic data. Variational inference (VI) algorithms cast the approximation of the posterior distribution as an optimisation problem [27]; first, VI finds a set of variational distributions $Q = \{q_\theta(w)\}$ and select the closest $q_\theta(w)$ to the *true* posterior distribution by computing the Kullback-Leibler (KL) divergence between the two. This workflow permits the optimisation of a cost function and batch learning, thus being suitable to be applied in deep learning.

A large body of literature is devoted to finding fast and accurate estimates of the posteriors

using variational inference [28]. Previous applications in BNNs include Stochastic Gradient Langevin Dynamics (SGLD) [29], Bayes by Backprop [30] and the reparametrisation trick [31]. However, they still suffer from scalability or computational issues [22]. Bayesian Deep Learning arises as the intersection of Bayesian methods with deep learning. They offer principled uncertainty estimates by combining the hierarchical feature learning of deep networks with the flexibility of Bayesian theory. Recent work by [32] links the dropout regularisation technique with variational learning, enabling an efficient posterior approximation by sampling from multiple dropout masks.

C. Variational dropout

Dropout is ANN regularization technique based on random de-activations of the network weights for a given probability p [33]. The randomness of this technique has been associated with VI in BNNs: the variational family $Q = \{q_\theta(w)\}$ can be sampled from a Bernoulli distribution to parametrize the neural network weights, W [32]. Therefore, the cost function of a BNN can be used to approximate the posterior distribution, with p, the drop-out probability. Once the network has been trained, the predictive function can be obtained by running T stochastics sampling steps from the dropout variational distribution. In this case, equation 4 can be approximated as:

$$p(y = c|x) \approx \frac{1}{T}\sum_{i=1}^{T}\tilde{f} \tag{5}$$

With $\tilde{f}$ the probabilistic output of the softmax layer. By randomly dropping weights with probability p at test time, we ensure that an ensemble of neural networks with weight dropout distribution $q(w|\theta)$ can approximate the posterior over the weights $p(w|D)$. This approximation is based on how the dropout strengthen network weights that are essential during the learning process, modelling uncertainty throughout the information dropped by de-activated weights from an ensemble of models [32]. Therefore, deep learning could improve the learned representation of volcano-seismic signals whilst gathering uncertainty estimates under a flexible Bayesian methodology. The prediction of probabilities from deep networks, when plugged into a Bayesian framework, allows the fast computation of uncertainty estimation on real time.

Table I

MOST REPRESENTATIVE VOLCANO-SEISMIC SCIENTIFIC LABELS ASSOCIATED WITH THEIR GEOPHYSICAL INTERPRETATION

Mc. Nutt [1]	**Minakami [34]**	**Other Names [35] [36]**	**Frequency (Hz) [9]**	**Duration (s) [9]**	**Some Source Models [1] [7] [6] [5] [9]**
High Frequency (HF)	*A-Type*	Volcano Tectonic Earthquakes, Tectonic, Short Period Earthquakes	>5.0	20-60	Shear failure or slip on faults, usually as swarms within the volcanic edifice.
Low Frequency (LF)	*B-Type*	Long Period Event, Volcanic, Long Coda Event, Tornillo	1.0 - 5.0	10-60	Fluid driven cracks, pressurization processes (bubbles), or attenuated waves.
Mixed Frequency (MX)	-	Hybrid Event, Medium Frequency	1.0 - 12.0	20-60	Mixture of processes, e.g: cracks and fluids
Explosion Quake (EXP)	Explosion Quake	Explosion, Volcanic Explosion	>10.0	<10.0	Accelerated emission of gas and debris to the atmosphere
Volcanic Tremor (TRE)	Volcanic Tremor	Volcanic Tremor, Harmonic Tremor,	1.0 - 12.0	150	Pressure disturbance, gas emissions, debris processes or pyroclastic flows

III. Volcano-Seismic Uncertainty Monitoring

The wealth of seismic-data recorded during an eruption requires accurate classification. The morphology, tectonic environment and composition of volcanoes contribute to shaping seismic signal, for example, due to attenuation of high-frequency energy. Current machine learning monitoring systems have highlighted topography changes or seismic variations of the medium as a significant influence on detection performance. Concretely [37] identified a substantial change in the physical mechanism of the events recorded at *Piton de La Fournaise* as the primary factor influencing the predictive performance. Similarly, [9] describes seismicity changes over time at *Ubinas* volcano as the main accuracy decay across eruptive periods. From a machine learning perspective, changes within the seismic environment produce probability distributions that are very distinct from the original training data. This leads to oversimplified assumptions that do not reflect the current situation, thus decreasing detection and classification under-performance. As a result, these statistical limitations could undermine the capacity to produce an objective methodology to consistently classify signals with high-levels of confidence, which ultimately can be extended to more refined early-warning methodologies [8].

Mathematically, two types of uncertainties can be defined: epistemic and aleatory. Epistemic uncertainty is associated with the absence of knowledge about the natural process and aleatory uncertainty is connected to the natural variability of volcanic unrest [22]. Quantifying aleatory uncertainty in a volcanic environment can be very challenging, as it is a direct consequence of the inherent non-linearity of volcanic processes. However, epistemic uncertainty could be quantified from the randomness of statistical parameters and can be characterised as the uncertainty linked to the neural network weights θ. Here, we propose to evaluate seismic uncertainty at a wave-form level, using BNNs as stochastic parsers from raw signals into event probabilities. These probabilities are sampled from the approximated variational dropout distribution Q, associating the uncertainty of the statistical parameters with the current dynamic of a volcano, i.e., the interaction of the seismic event with the environment. Thus, epistemic uncertainty for C classes can be computed from the per-class probability vector p_c using the entropy $H(p)$ as generalised measured of uncertainty [38]:

$$H(p) = -\sum_{c=1}^{C} p_c \, log \, p_c. \tag{6}$$

For both models, this probability vector p_c is the result of principled sampling from the

variational distribution Q. Therefore, we do not obtain a single point estimate of the event, but a probabilistic representation associated with the unknown knowledge of the model for the selected event. In a multi-classification setting, these estimations provide not only the annotation (label) of the waveform but also a probabilistic assessment of how far from the original data distribution our estimates are. Information of individual events is not missed, and we can quantify potential seismic variations that are associated to changes in the overall data distributions: BNNs can detect and classify the event while providing high uncertainty, indicating that there is a change in the probability distribution of seismic events. The association of weight uncertainty to seismic changes would lead to more refined seismic catalogs and improved assessment of volcanic hazards: not only the events are processed, but data distribution shifts can be tracked between seismic snapshots.

A. Uncertainty and Transfer Learning

Volcano-seismic monitoring systems based on machine learning are very accurate on selected periods but tend to decrease its performance given the data distribution shifts over time, reflecting the evolving volcanic environments [9][37]. This leads to continuous manual analysis of the new eruptive periods in order to produce datasets large enough to cover the novel range of data distributions. Hence, the insufficient amount of new labelled data, along with the time needed to analyse and retrain monitoring systems are the main factors that limit the exportability across seismic campaigns in volcanological observatories. Transfer Learning could relieve the data scarcity problem and the time needed to react to these changes [39]. Successful applications of transfer learning in a number of disciplines have helped to identify the essential knowledge that needs to be transferred across domains and tasks; great improvements in the performance of these systems have been achieved, for example, in music [40], analyses of electroencephalograms [41] and geophysical image processing [42]. However, given the extremely dynamic nature of volcano-seismic sources, it is advisable to not apply brute-force transfer learning, but a more refined approach to avoid negative transfer learning, i.e, the decrease of accuracy in the new domain [43].

In this context, we link transfer learning and epistemic uncertainty in order to mitigate the generalisation error gap between data distributions whilst detecting at the same time subtle differences in the new seismic data. This would yield more polished monitoring systems, able

Table II
NUMBER OF EVENTS FOR BEZYMIANNY AND ST.HELENS VOLCANOES, COVERING BOTH ERUPTIVE PERIODS

Labels	**St. Helens**		**Bezymianny**	
	2004-2005 *(pre-eruptive)*	2005-2006 *(post-eruptive)*	2007-2008 *(pre-eruptive)*	2008-2009 *(post-eruptive)*
HF	8353	1437	6929	10617
LF	8423	9310	8523	2843
MX	8525	8357	9464	8715
Total	**25301**	**19104**	**24916**	**22175**

to quantify uncertainty, detect statistically meaningful changes and help analysts to build large-scale, high-quality annotated datasets.

IV. SEISMO-VOLCANIC DATASETS

In volcano seismology, there is not a uniform way to classify earthquake signals. Waveforms are classified based on a set of properties measured in the time or frequency domain. Table I shows a summary of earthquake classifications and possible source models that have been traditionally attributed to them. Our Bayesian framework will be focused on the identification and classification of the three most representative classes of earthquakes that are encountered in a volcanic environment, low-, high-, and mixed-frequency events. Figure 1 illustrates the typical frequency content for these three classes of volcano-seismic signals. We will test the performances of our Bayesian workflow on data from two volcanoes, Mt. St. Helens, USA, and Bezymianny, Russia. In Table II, we show the composition and per-class distribution of events for the selected volcanoes. In summary, our database includes:

1) Low frequency events (LF) (Figure 1.a): This type of earthquakes deliver energy mainly in the 0.5-5 Hz band, and have typical durations of $\approx$25.0 seconds.
2) High frequency events (HF) (Figure 1.b): They are characterized by broadband spectra, with significant energy delivered well above $5Hz$, clear P and S waves onsets, and typical durations of less than 25.0 seconds.
3) Mixed frequency events (MF) (Figure 1.c): They are characterized by energy delivered across the spectrum of both LF and HF events, across the 1-20 Hz band.

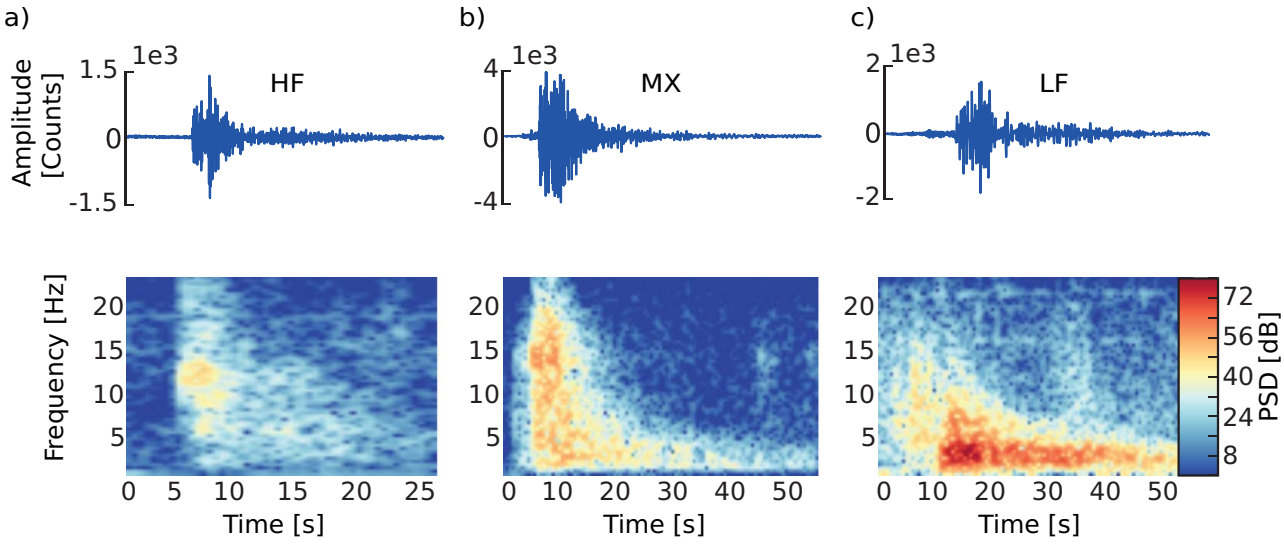


Figure 1. Waveform and spectrograms for three of the most representative volcano-seismic signals [1]. The frequency content of the events shifts from higher (a) to lower (c) frequencies. In the Mixed Frequency (MX) case (b), notice frequencies are mixed between lower and higher part of the frequency bands, with a broad spectra.

This compact classification scheme allows avoiding label sparsity. Other events such as tremor, explosions, rockfalls, ice-quakes or regional earthquakes are not considered in this work.

We focus our study on Mount St. Helens and Bezymianny volcanoes, during the eruptive periods of 2004-2005 and 2007-2008, respectively. Prior waveform analysis at Bezymianny and Mount St. Helens, [44] suggests similar seismic patterns at the two volcanoes prior to explosive activity, and similar waveform characteristics in terms of amplitude, duration and standard deviation from the average signal.

Following previous work of [44] and [3], we used, for initial classification, data from stations $BELO$, $BESA$ and $BERG$ at Bezymianny, and from stations $S02$, $S06$ and $S15$ (dome reactivation) at St. Helens. In order to assess how BNNs could transfer seismic knowledge across eruptive crisis, and if uncertainty can be quantified during unrest periods and post-eruptive crisis, we analysed $BELO$, $BESA$ and $BERG$ stations from 2008-2009 at Bezymianny volcano and $S03$ and $S07$ and $S15$ stations from the 2005-2006 eruption at St. Helens (spine destruction, stations $S02$ and $S06$ were not available during this period).

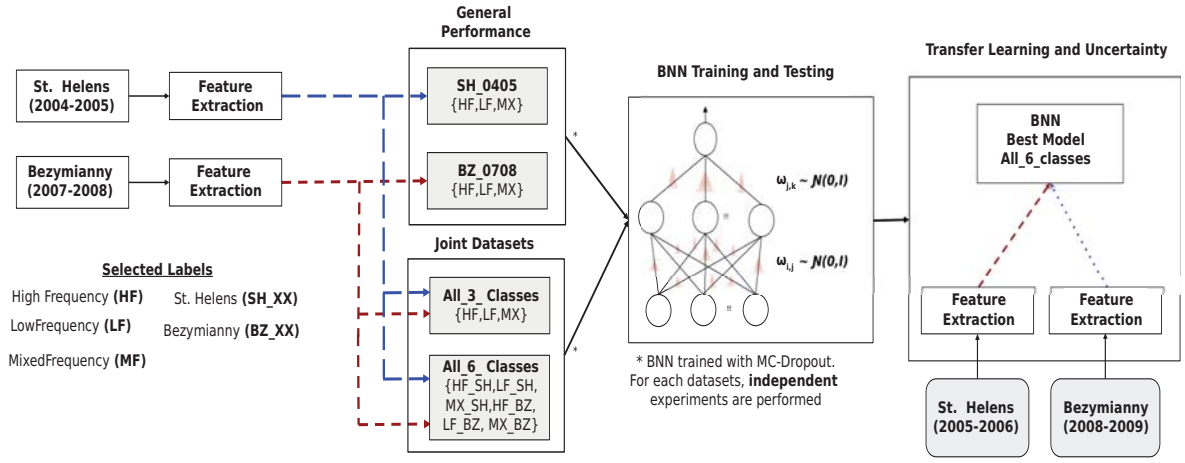


Figure 2. Experimental framework implemented in this work. For each volcano, frequency features are computed from pre-processed raw signals. Four datasets are created: SH (St.Helens), BZ (Bezymianny), similar labels (both_3_classes) and separately (both_6_classes). Best models from joint datasets are tested with different eruptive periods to evaluate transfer learning and uncertainty quantification capabilities of the BNN.

V. Experimental Setup

In this section, we aim to establish if BNNs are suitable for volcano-seismic monitoring and, for each volcano, if epistemic uncertainty could offer insights into volcanic and seismic unrest. Figure 2 summarises the implemented data pre-processing workflow and experiments. We implemented the same probabilistic framework as described in section II with St. Helens and Bezymianny data from Table II. We evaluate transfer learning and uncertainty quantification on a new eruptive period by selecting only the best model, on $both\_6\_classes$ from previous trained periods.

A. Data pre-processing and feature extraction

The raw continuous data streams for each volcano, at each of the selected stations, were pre-processed in order to extract events of interest using the REMOS (Recursive Entropy Method of Segmentation) algorithm [45]. Each extracted event is later characterised using a well-tested set of features already investigated by us in several volcanic-scenarios, including Deception Island [18], Etna [14] and Stromboli [46]. The REMOS algorithm performs data segmentation and semi-supervised categorisation of events into classes based on their frequency index (FI), the

ratio of energy within low- and high- frequency bands:

$$FI = \log_{10}\left(\frac{E_{high}}{E_{low}}\right) \tag{7}$$

where E_{high} and E_{low} are the spectral energy in the, high ($[6-12]Hz$) and low ($[1-5]Hz$) frequency bands. All events detected and classified by REMOS were visually inspected using a custom Python Graphical User Interface (GUI) to confirm or modify the initial label of the event. As a result, complete and annotated catalogues are generated (see Table II). Data pre-processing and feature extraction pipelines are implemented using the signal processing modules in the well-known *Obspy* seismic toolbox [47] on the detected waveform. Following the same procedure as defined in previous work by [14], we derived a set of 13 cepstral coefficients on a logarithmic scale for the data; first, using a Hamming window (4.0s), the spectrum of the seismic signal is computed, and an log-spaced filterbank (16 triangular weighting function, 50% adjacency) is designed to yield an individual average of the spectral frequencies. Cepstral analysis is performed and 13 cepstral coefficients are then derived for each earthquake in the database [48].

B. Evaluation metric

Here we use confusion matrices and accuracy (Acc) to evaluate the performance of the BNNs to classify volcano-seismic events. We compute the Acc as:

$$Acc\,(\%) = \frac{Number\,of\,Correct\,Predictions}{(Total\,Number\,of\,Events)} * 100 \tag{8}$$

The Acc is the standard measure of overall effectiveness for a classifier. Moreover, we compute precision (PR) and recall (RC) metrics as:

$$Precision\,(PR) = \frac{True\,Positives}{(True\,Positives + False\,Positives)} \tag{9}$$

$$Recall\,(RC) = \frac{True\,Positives}{(True\,Positives + False\,Negatives)} \tag{10}$$

For a given model, these metrics can diagnose how many events are correctly detected and classified. In practice, recall measures the proportion of relevant detected seismic events, i.e, how good the model can detect events from a given class. Precision measures the refinement of statistical model, i.e, how good the classifier can discriminate specific instances [49]. Both

metrics offer more explicit information about the number of miss-classified events than accuracy alone and are accepted performance measures in volcano-seismic monitoring [46]. The weighted average of precision and recall, known as $F1$ score, can be computed as:

$$F1 = \frac{2 * (RC * PR)}{(RC + PR)} * 100\% \tag{11}$$

$F1$ score provides an informative trade-off measure between the PR and RC. These metrics are of particular interest in seismology, as RC is related to a sensitivity of the system (how many earthquakes is able to detect), and PR to specificity (how many earthquakes are correctly classified), thus offering more global information than accuracy itself.

C. Model Implementations

For each of the volcano-seismic datasets in Table II, we perform data pre-processing as described in subsection V-B. Once the features are extracted, we divide each dataset into training (80%), and test (20%) sets, and different BNNs are trained independently. Further, joint datasets with events from both volcanoes are used to train two independent BNNs: a mixed ($Both\_3\_dataset$) and sparse ($Both\_6\_dataset$). All BNNs models are initialised with *Glorot Initialization*. Hyper-parameter fine-tuning is based on a Bayesian optimisation towards best configurations, followed by a random search over the most promising hyperparameters [50]. All models are optimised with Adam [51], initial learning rate of 0.01, *ReLU* activation function, mini-batch size set of 32, and dropout probability at ($p = 0.25$). The cross-entropy loss is used as cost function. The training stage is set to 50 epochs, with early-stopping set to a patience interval of 5 epochs in order to mitigate overfitting. MC-Dropout has been implemented as described in Section II. The transfer learning setting follows a similar procedure: Using the best obtained models, volcano-seismic events from 2008 at Bezymianny and 2005 at St. Helens are pre-processed and extracted as described in subsection V-B. Our BNN framework is implemented entirely in Tensorflow, and simulations executed in an NVIDIA Tesla P40 GPU, 24 GB GPU memory and 32 GB RAM.

VI. Results and discussions

Our data analysis is divided into three steps (see Figure 2). Firstly, we analyse each volcano independently, i.e., we test the performance of the BNN to classify the three selected types

of volcano-seismic events (LF, MF and HF) separately for each volcano. In the second step, we explore how the BNN is able to recognise the signals when all labelled signals from both volcanoes scenarios are merged together. Finally, after a second joint training of all events from. both volcanoes, we investigate the associated uncertainties. Since we use data that change over time from the two volcanoes, pre and post an eruption, we assess how these signals changed, and the associated uncertainty in the recognition that can be interpreted as a change in the seismic source mechanisms.

A. General performance of the system

After the pre-processing analysis performed in the previous section, we selected a large, high-quality and balanced dataset for each volcano. For St. Helens we identified 25,301 seismic events, and for Bezymianny 24,916. The size of the two datasets is similar, allowing to generalise our observations.

Table III (a) and (b) presents the averaged *Acc*, confusion matrix, *PR*, *RC* and *F1* metrics for St. Helens and Bezymianny volcano, respectively. The overall mean of the epistemic uncertainty is also reported. Being this our baseline system, notice that all optimised architectures result in high-performance when the datasets are independently studied. The accuracy remains high, with 95.7% for St. Helens (SH_0405) and 94.1% for Bezymianny (BZ_0506). Precision (PR) and recall (RC) remain high for all classes in both datasets, which highlights that the feature vector fed to the neural network carries rich information to exploit. From Table III, the confusion matrix reveals that MX events are the only events that present fluctuations among the classes. Due to their spectral characteristics, LF and HF were never misclassified at both volcanoes, demonstrating the high quality of the characterization process. In general MX events, also known as *hybrid events*, (see Table I) share characteristics of both HF and LF events. We infer that the observed confusion matrix is associated with both attenuation effects and source effects, rather than incorrect initial labelling.

B. Performance on joint datasets

In this section, we explore the exportability of the labelled database and assess whether it is necessary to re-train the systems with new data when it is used at a new volcano. For this purpose, we merge our test datasets and perform again the BNN analysis.

Table III

AVERAGED CONFUSION MATRICES AND PERFORMANCE (PR, RC, F1) FOR ST.HELENS (a) AND BEZYMIANNY (b) VOLCANOES

(a)

Pred. / True	**HF**	**LF**	**MX**	**PR**	**RC**	**F1**
HF	**2025**	0	82	0.97	0.96	0.96
LF	0	**1986**	88	0.98	0.96	0.96
MX	52	48	**2043**	0.92	0.95	0.93

Overall Accuracy: 95.7%, Mean Epistemic Unc.: 0.10

(b)

Pred. / True	**HF**	**LF**	**MX**	**PR**	**RC**	**F1**
HF	**1628**	0	87	0.95	0.95	0.95
LF	0	**2305**	75	0.96	0.97	0.96
MX	93	84	**2214**	0.93	0.93	0.93

Overall Accuracy: 94.5%, Mean Epistemic Unc.: 0.12

Table IV

AVERAGE CONFUSION MATRIX (a) AND PERFORMANCE (PR, RC, F1) (b) FOR BOTH VOLCANOES, SAME LABELS

(a) (b)

Pred. / True	**HF**	**LF**	**MX**	**PR**	**RC**	**F1**
HF	**3681**	0	172	0.94	0.96	0.95
LF	0	**4035**	151	0.97	0.96	0.96
MX	224	130	**4153**	0.93	0.92	0.92

Overall Acc,: 94.6 % Mean Epistemic Unc.: 0.13

In the first stage of this analysis, we merged the two datasets (SH_0405 and BZ_0809) in order to train a unique BNN, independently of the origin of the signal. In Table IV (a) we report the confusion matrix, PR, RC, $F1$, averaged accuracy ($Acc.$) and epistemic uncertainty for both datasets, when labels are unified ($Both\_3\_classes$). It is interesting to note that the previous trend is maintained when labels from both volcanoes are merged together, with a slight decrease in accuracy and an increment in uncertainty: more events are incorporated from distinct sources, forcing the network to learn a more complex data distribution. We note that RC and PR are elevated for HF and LF events, whereas the MX events present a lower RC but higher PR.

Table V

AVERAGED CONFUSION MATRIX a, AND PERFORMANCE (PR, RC, F1) (b) FOR JOINT DATASET, SPARSE LABELS.

	(a)						***(b)***		
Pred. / True	**HF_SH**	**LF_SH**	**MX_SH**	**HF_BZ**	**LF_BZ**	**MX_BZ**	**PR**	**RC**	**F1**
HF_SH	**2003**	0	63	63	0	1	0.94	0.94	0.94
LF_SH	0	**1959**	66	0	18	1	0.95	0.96	0.95
MX_SH	88	66	**1927**	10	2	28	0.90	0.91	0.90
HF_BZ	33	0	25	**1533**	0	157	0.91	0.88	0.89
LF_BZ	0	39	6	0	**1961**	117	0.95	0.93	0.93
MX_BZ	2	2	48	88	69	**2180**	0.88	0.91	0.89

Overall Accuracy: 92.08% Mean Epistemic Unc.: 0.20

These results reveal that only events that are correctly detected are classified with great precision, which can be translated into a decreased number of false positives for the three classes. This idea is highlighted by the LF events, as they can be discriminated with higher RC and PR. Therefore, the first observation is the demonstration that BNN is a powerful tool to discriminate distinct seismic events, even if they have a different origin: the only condition is to perform an accurate pre-processing and data labelling. Hence, we can infer that the system is exportable, i.e. the experience from one volcano is transferable to a new one, if the seed database was generated with high-quality and large number of data.

In order to explore differences between volcanoes, we performed the same test, but differentiating the labelled seismic classes according to their origin, here separating them between the two test volcanoes. Table V (a) and (b) show the recognition performance when labels are separated for each volcano. These results highlight an important property: increasing label sparsity could decrease performance, but provide geophysical insight about the seismic events. From Table V (b), whilst the overall performance of the BNN is good in terms of PR, RC and $F1$, the trend with respect the unified dataset presents subtle differences. First, RC and PR remains high for LF events. Second, MX events have similar recall at both volcanoes, 0.91, but lower precision at Bezymianny volcano. Similarly, the recall and precision of HF events at Bezymianny volcano is lower when compared to St. Helens. This could indicate that this seismic classes share frequency properties across the two volcanoes, as RC magnitudes are influenced by the seismic events, yielding a less sensitive system in the case HF events, but with increased PR. Similarly, PR

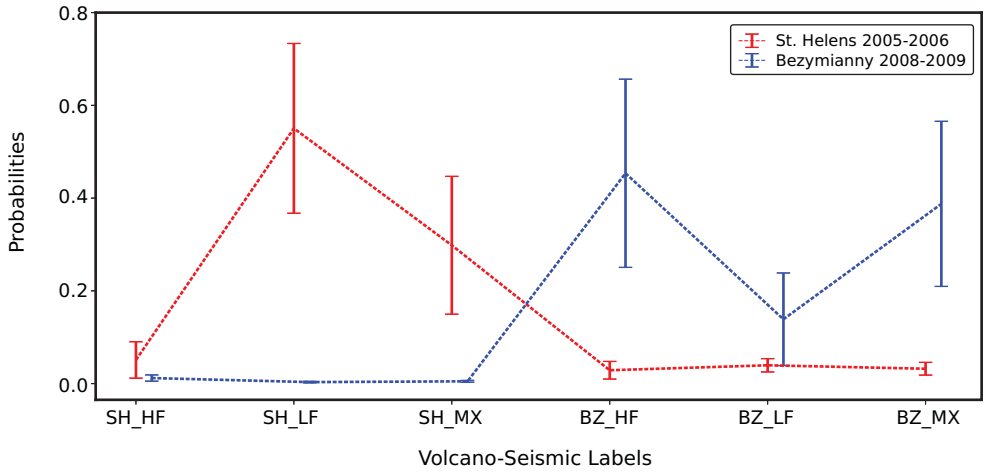


Figure 3. Per-class overall mean and variance within predictions for (a) St.Helens 2005-2006 new eruptive period, blue line (b) Bezymianny 2008-2009 post-eruptive period, red line. The axis order is the same as the assigned categorical labels in our training data. By following the lines, notice that the probability of assignment between volcanoes is very low, but high in the case of the same volcano.

and RC in the mixed frequency events, BZ_MX and SH_MX, are lower when compared to the combined dataset: the system detects less mixed frequency events, but classify them with higher precision, and can discern their origin.

The confusion matrix for the sparse labels ($Both\_6\_classes$) in Table V (a) suggests that even if the BNN was able to merge classes previously when applied separately, is able to determine at which volcano the seismic signal was generated. In general HF events are interpreted as the result of brittle failure as a consequence of stress accumulation. Source depth and its mechanisms, and path effects, influence the final characteristic of the recorded waveform; this allows differentiating HF earthquakes with different sources, even at the same volcano.

C. Epistemic Uncertainty

In this section we introduce how uncertainties can be interpreted as a consequence of similarities and differences between volcanoes, specific seismic source, and the general character of unrest.

As described in the previous sections, we used seismic signals from two similar volcanoes during pre- and post-eruption periods. In the previous analysis, we used for each volcano the whole data set and observed a high degree of success of the BNN to recognise and classify

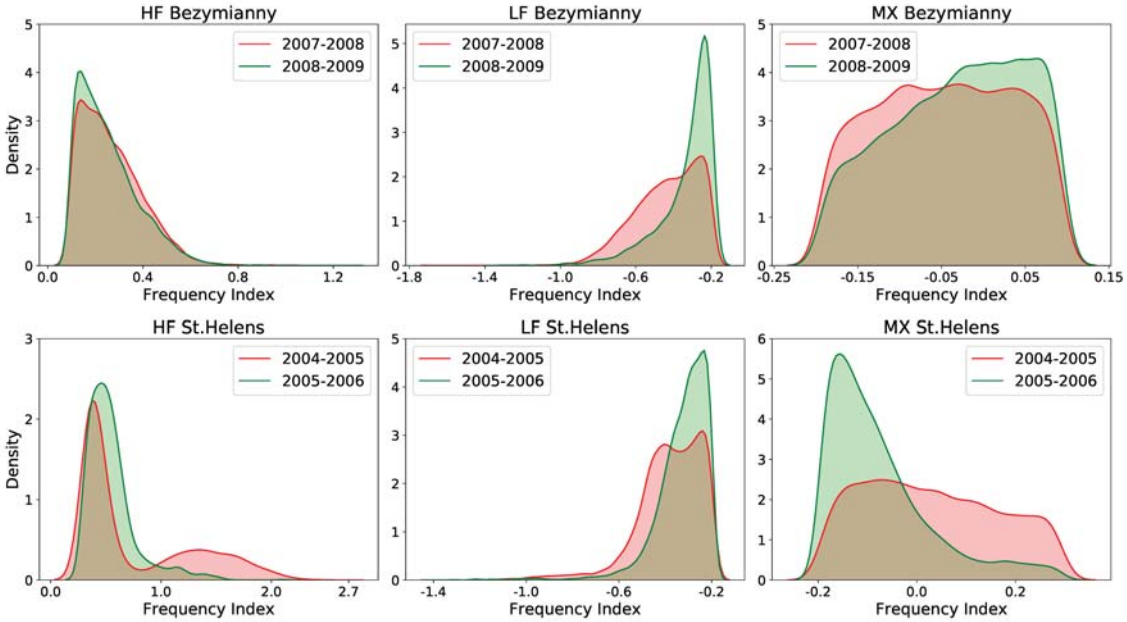


Figure 4. Frequency distribution variations for the two selected eruptive periods on Bezymianny (upper part) and St. Helens (lower part) volcanoes. Notice that, after the main eruptions, the frequency content in all selected events changes, shifting the frequency index distribution.

seismic events, i.e. BNN is a powerful tool both for all data from each volcano, and for exporting knowledge for both and separated labels. A deep and detailed frequency analysis of the seismic signals shows that some characteristics of the seismic signals changed after eruptive activity took place. We focus our analysis on the FI. Figure 3 depicts the per-class overall mean and variance within predictions, and in Figure 4, we plot the frequency index distribution before and after eruption for both volcanoes and the whole data set. In Table VI we report the accuracy, along with the epistemic uncertainty, for the new eruptive periods, before and after the application of the transfer learning methodology. From Figure 4, the differences are obvious; these differences are likely associated with changes within the volcanoes. Thus, the BNN is able to quantify variations in the seismic signals, assigning greater variance in their predictions, even if labels are kept the same. Moreover, despite the recognition results at Table VI, higher accuracy on blind-test in a new volcano-seismic dataset does not imply greater class probability. Therefore, the models are able to correctly classify events, but they do yield per-class greater variances and lower probabilities. Epistemic uncertainty has two roles: quantification of dataset distribution, and its association to volcano-seismic changes, such as those reported in [3] and [44]. In Figure

Table VI

EPISTEMIC UNCERTAINTY AND ACCURACY (%) FOR THE NEW ERUPTIVE DATASETS

Dataset	*Blind Test*		*After Transfer Learning*	
	Acc (%)	**Epistemic**	**Acc (%)**	**Epistemic**
Both_post-eruptive	84.90	0.27	93.39	0.16
Bezy_0809	90.82	0.23	93.88	0.17
St.Helens_0506	80.08	0.32	95.13	0.14

Table VII

AVERAGED CONFUSION MATRIX BEFORE (A) AND AFTER (B) TRANSFER LEARNING, FOR THE POST-ERUPTIVE JOINT DATA TEST DISTRIBUTION (70%)

	(a)						***(b)***					
Pred. / True	**SH_HF**	**SH_LF**	**SH_MX**	**BH_HF**	**BH_LF**	**BH_MX**	**SH_HF**	**SH_LF**	**SH_MX**	**BH_HF**	**BH_LF**	**BH_MX**
SH_HF	**716**	29	38	14	178	4	**810**	0	46	123	0	0
SH_LF	2	**5201**	250	0	979	7	0	**6061**	349	0	29	0
SH_MX	114	2087	**2480**	4	817	226	39	218	**5405**	30	2	34
BH_HF	340	8	47	**3834**	312	2830	56	0	56	**7061**	1	197
BH_LF	1	34	1	0	**1953**	28	0	36	19	0	**1661**	301
BH_MX	24	23	53	88	1620	**4300**	3	0	83	459	76	**5487**

3, we plotted the evolution of the main frequency indices for each class and volcano during both, pre- and post-eruptive stage. The shift in the frequency index according to activity is a clear consequence of a change in the physical properties of the sources and medium.

D. Transfer Learning

In this last experiment, we test the capabilities of the BNN ($Both\_6\_classes$) to learn with data from new eruptive periods, aiming to investigate if by *fine-tuning* the weights of the pre-trained network we could classify new events whilst decreasing the uncertainty of the models. Transfer learning was implemented following procedures similar to those described in Section V. Table VI reports the accuracy and the epistemic uncertainty on both post-eruption data (see Table II), on a blind test prediction, and after the transfer learning procedure. Firstly, we notice that as the frequency distribution of the events has changed, the accuracy drops in our dataset, and uncertainty remains high. When the seismic source changes (see Figure 3), the overall uncertainty increases, for each volcano, and on both datasets. Therefore, given that our feature vector is trained on pure frequency attributes, changes within the frequency bands of the events

can be perceived through the uncertainties associated with the BNN model, resulting in higher uncertainties. The frequency characteristics previously learnt by the probabilistic weights of the BNNs are transferable to the new eruptive period, helping the BNN to adapt itself to the new changes in frequency bands. This yields higher accuracy and lower epistemic uncertainty. Additionally, Table VI (b), the confusion matrix, demonstrates the transfer learning capabilities of the BNN, trained on 25$\%$ of the training data, and 75 $\%$ test partition. This simulates the condition in which only a small subset of the new data is available for re-training, which is often the situation during a new eruptive crisis. Note that the sparsity of the matrix is reduced, and more events are correctly classified. Additionally, the exportability is manifest given that many of the errors are placed on similar type of events (MX, HF, LF), but at different volcanoes. Only the recognized LF events at Bezymianny are lower when compared to the blind-test case, being confused with LF of Mt. St. Helens. This is indicative that some of these events share similar properties at the two volcanoes, which was also reported by [44].

The lower number of miss-classified events, jointly with lower epistemic uncertainty, and higher recognition accuracy highlights an important point: there is no need to train an early-warning system from scratch, but it is possible to export systems that are related in other to simplify the deployment. Considering the dynamics of learning, the selective reuse of the prior model to *unlearn* irrelevant information from the previous datasets help the new model to exploit at least some common structure on the new datasets (new eruptive periods). Therefore, this new *fine-tuning* helps to improve recognition results (see Table VI) and to decrease data uncertainty, and thus, to mitigate issues with volcano-seismic data scarcity. We are proving that BNN are a powerful tool that, allow exporting knowledge from one volcano, or one stage to another and, simultaneously, are capable of tracking how signals evolve over time from a probabilistic perspective, even for mixed, sparse, datasets.

VII. Conclusion and future work

In this work, we investigated a new Bayesian approach for application to volcano-seismic monitoring. We focused our research on finding new ways to exploit uncertainties derived from a Bayesian deep learning framework as a realistic unrest detector. Two different eruptive periods at two volcanoes, Mount St. Helens and Bezymianny, were studied. Results demonstrate that BNNs are able to detect and recognise volcano-seismic signals with outstanding performance for the two volcanoes, separately. Moreover, when the two datasets are combined, the BNN attains

an excellent performance in terms of PR, RC and Accuracy, and is able to classify events from the two volcanoes, based on their frequency characteristics. Additionally, when the datasets are separated according to their volcanic origin, the BNN is able to detect the volcano where signals were generated.

The proposed approach provides uncertainty representation related to changes in the dynamics of both volcanoes. Further, the flexibility of Deep Learning, when viewed through the lens of Bayesian theory, allow us to tackle the problem of data scarcity from monitoring networks with no prior data available. We illustrated frequency content variations during pre- and post-eruptive periods, which are well-sensed by the epistemic uncertainty associated to the BNNs *a-priori* weights. The epistemic uncertainty derived from the BNN weights has two main implications: it stands not only as a feature to be considered as an unrest precursor, but also as a threshold level to determine when transfer learning algorithms should be used.

The exploration of monitoring systems from the perspective of Bayesian theory has highlighted the advantages of their deployment, and how the transfer of learned features with appropriate datasets could mitigate the data scarcity problem, even under intense volcanic activity. Our results can be exported to other volcanoes worldwide.

VIII. Acknowledgements

This work was supported by the Spanish Government TEC2015-68752 (MINECO/FEDER), KNOWAVES project and by NERC Grant NE/P00105X/. We would like to thank all the members at the Seismic Laboratory of University of Liverpool for continuous support and advice on seismic data, along with the IAG (Instituto Andaluz Geofísica) who contributed with helpful analysis and seismic data reviews.

References

[1] S.R. McNutt, G. Thompson, J. Johnson, Silvio S. De Angelis, and D. Fee. “Seismic and infrasonic monitoring”. In: *The Encyclopedia of Volcanoes (Second Edition)*. Elsevier, 2015, pp. 1071–1099.

[2] R. M. Harrington and E. E. Brodsky. “Volcanic hybrid earthquakes that are brittle-failure events”. In: *Geophysical Research Letters* 34.6 (2007). DOI: 10.1029/2006GL028714. URL: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2006GL028714.

[3] SC. Moran, S. Malone, D. Stephen, A. Qamar, W. Thelen, A. Wright, and J.Caplan-Auerbach. “Seismicity associated with renewed dome building at Mount St. Helens, 2004-2005”. In: *US Geological Survey professional paper* 1750 (2008), pp. 27–60.

[4] B. Chouet. “Volcano seismology”. In: *Pure and Applied Geophysics* 160 (2003), pp. 739–788.

[5] R. M. Iverson, D. Dzurisin, C. A. Gardner, T. M. Gerlach, R. G. LaHusen, M. Lisowski, J. Major, S. Malone, J. Messerich, S. C. Moran, J. Pallister, A. Qamar, S. Schilling, and J. Vallance. “Dynamics of seismogenic volcanic extrusion at Mount St Helens in 2004-05”. In: *Nature* 444.7118 (2006), pp. 439–443. ISSN: 1476-4687. DOI: 10.1038/nature05322. URL: https://doi.org/10.1038/nature05322.

[6] J.E. Kendrick, Y. Lavallée, T. Hirose, G. Di Toro, A. J. Hornby, S. De Angelis, and D. B. Dingwell. “Volcanic drumbeat seismicity caused by stick-slip motion and magmatic frictional melting”. In: *Nature Geoscience* 7 (2014), 438 EP –. URL: https://doi.org/10.1038/ngeo2146.

[7] C. J. Bean, L. De Barros, I. Lokmer, J. Métaxian, G. O’Brien, and S. Murphy. “Long-period seismicity in the shallow volcanic edifice formed from slow-rupture earthquakes”. In: *Nature Geoscience* 7 (2013). Article, 71 EP –. URL: https://doi.org/10.1038/ngeo2027.

[8] A. Boue, P. Lesage, G. Cortés, B. Valette, G. Reyes-Dávila, R. Arámbula-Mendoza, and A. Budi-Santoso. “Performance of the Material Failure Forecast Method in real-time situations: A Bayesian approach applied on effusive and explosive eruptions”. In: *Journal of Volcanology and Geothermal Research* 327 (2016), pp. 622–633.

[9] M. Malfante, M. Dalla Mura, J. Metaxian, J. I. Mars, O. Macedo, and A. Inza. “Machine Learning for Volcano-Seismic Signals: Challenges and Perspectives”. In: *IEEE Signal Processing Magazine* 35.2 (2018), pp. 20–30. ISSN: 1053-5888.

[10] Q. Kong, D. Trugman, Z.E Ross, M. Bianco, b.j. Meade, and P. Gerstoft. “Machine learning in seismology: Turning data into insights”. In: *Seismological Research Letters* 90.1 (2018), pp. 3–14.

[11] B. Rouet-Leduc, C. Hulbert, N. Lubbers, K. Barros, C.J. Humphreys, and P.A. Johnson. “Machine Learning Predicts Laboratory Earthquakes”. In: *Geophysical Research Letters* 44.18 (2017), pp. 9276–9282. DOI: 10.1002/2017GL074677. URL: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2017GL074677.

[12] K.J. Bergen, T. Chen, and Z. Li. “Preface to the Focus Section on Machine Learning in Seismology”. In: *Seismological Research Letters* 90.2A (2019), pp. 477–480.

[13] K.J. Bergen, P.A. Johnson, M.V. de Hoop, and G. Beroza. “Machine learning for data-driven discovery in solid Earth geoscience”. In: *Science* 363.6433 (2019). ISSN: 0036-8075. DOI: 10.1126/science.aau0323. eprint: https://science.sciencemag.org/content/363/6433/eaau0323.full.pdf. URL: https://science.sciencemag.org/content/363/6433/eaau0323.

[14] I. Álvarez, L. García, G. Cortés, M. C. Benítez, and Á. De la Torre. “Discriminative Feature Selection for Automatic Classification of Volcano-Seismic Signals”. In: *IEEE Geoscience and Remote Sensing Letters* 9.2 (2012), pp. 151–155. ISSN: 1545-598X.

[15] G. Cortés, M. C. Benítez, L. García, I. Álvarez, and J. M. Ibáñez. “A Comparative Study of Dimensionality Reduction Algorithms Applied to Volcano-Seismic Signals”. In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 9.1 (2016), pp. 253–263. ISSN: 1939-1404. DOI: 10.1109/JSTARS.2015.2479300.

[16] M. Bicego, C. Acosta-Muñoz, and M. Orozco-Alzate. “Classification of Seismic Volcanic Signals Using HMM-based Generative Embeddings”. In: *IEEE Transactions on Geoscience and Remote Sensing* 51.6 (2013), pp. 3400–3409. ISSN: 0196-2892. DOI: 10.1109/TGRS.2012.2220370.

[17] G. Cortés, R. Carniel, M. Ángeles, and P. Lesage. “Standardization of Noisy Volcanoseismic Waveforms as a Key Step toward Station-Independent, Robust Automatic Recognition”. In: *Seismological Research Letters* (2019).

[18] M. C. Benítez, J. Ramírez, J. C. Segura, J . M. Ibáñez, J. Almendros, A. García-Yeguas, and G. Cortés. “Continuous HMM-Based Seismic-Event Classification at Deception Island, Antarctica”. In: *IEEE Transactions on Geoscience and Remote Sensing* 45.1 (2007), pp. 138–146. ISSN: 0196-2892. DOI: 10.1109/TGRS.2006.882264.

[19] L. Gutiérrez, J. M. Ibáñez, G. Cortés, J. Ramírez, C. Benítez, V. Tenorio, and I. Álvarez. “Volcano-seismic signal detection and classification processing using Hidden Markov Models. Application to San Cristóbal volcano, Nicaragua”. In: *2009 IEEE International Geoscience and Remote Sensing Symposium*. Vol. 4. 2009, pp. IV–522–IV–525. DOI: 10.1109/IGARSS.2009.5417428.

[20] M. Titos, A. Bueno, L. García, and M. C. Benítez. “A Deep Neural Networks Approach to Automatic Recognition Systems for Volcano-Seismic Events”. In: *IEEE Journal of*

Selected Topics in Applied Earth Observations and Remote Sensing 11.5 (2018), pp. 1533–1544. ISSN: 1939-1404. DOI: 10.1109/JSTARS.2018.2803198.

[21] M. Titos, A. Bueno, L. García, M. C. Benítez, and J. M. Ibáñez. "Detection and Classification of Continuous Volcano-Seismic Signals With Recurrent Neural Networks". In: *IEEE Transactions on Geoscience and Remote Sensing* (2018), pp. 1–13. ISSN: 0196-2892. DOI: 10.1109/TGRS.2018.2870202.

[22] RSJ. Sparks and WP. Aspinall. "Volcanic activity: frontiers and challenges in forecasting, prediction and risk assessment". In: *The State of the Planet: Frontiers and Challenges in Geophysics* abs/1703.04977 (2004).

[23] Y. LeCun, Y. Bengio, and G. Hinton. "Deep learning". In: *nature* 521.7553 (2015), p. 436.

[24] G. Hinton, L. Deng, D. Yu, G. E. Dahl, A. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, T. N. Sainath, and B. Kingsbury. "Deep Neural Networks for Acoustic Modeling in Speech Recognition: The Shared Views of Four Research Groups". In: *IEEE Signal Processing Magazine* 29.6 (2012), pp. 82–97. ISSN: 1053-5888. DOI: 10.1109/MSP.2012.2205597.

[25] R. Neal. *Bayesian learning for neural networks*. Vol. 118. Springer Science & Business Media, 2012.

[26] D.JC. MacKay. "Probable networks and plausible predictions—a review of practical Bayesian methods for supervised neural networks". In: *Network: computation in neural systems* 6.3 (1995), pp. 469–505.

[27] G. Hinton and D. Van Camp. "Keeping neural networks simple by minimizing the description length of the weights". In: *Proc. of the 6th Ann. ACM Conf. on Computational Learning Theory*. Citeseer. 1993.

[28] A. Graves. "Practical variational inference for neural networks". In: *Advances in neural information processing systems*. 2011, pp. 2348–2356.

[29] M. Welling and Y.W. Teh. "Bayesian learning via stochastic gradient Langevin dynamics". In: *Proceedings of the 28th international conference on machine learning (ICML-11)*. 2011, pp. 681–688.

[30] C. Blundell, J. Cornebise, K. Kavukcuoglu, and D. Wierstra. "Weight uncertainty in neural networks". In: *arXiv preprint arXiv:1505.05424* (2015).

[31] D. Kingma, T. Salimans, and M. Welling. “Variational dropout and the local reparameterization trick”. In: *Advances in Neural Information Processing Systems*. 2015, pp. 2575–2583.

[32] Y. Gal. “Uncertainty in deep learning”. PhD thesis. PhD thesis, University of Cambridge, 2016.

[33] N. Srivastava, G. Hinton, A. Krizhevsky, Ilya I. Sutskever, and R. Salakhutdinov. “Dropout: a simple way to prevent neural networks from overfitting”. In: *The Journal of Machine Learning Research* 15.1 (2014), pp. 1929–1958.

[34] T. Minakami. “Prediction of volcanic eruptions”. In: *Developments in Solid Earth Geophysics*. Vol. 6. Elsevier, 1974, pp. 313–333.

[35] J.M. Ibáñez, E. del Pezzo, J. Almendros, M. La Rocca, G. Alguacil, R. Ortiz, and A. García. “Seismovolcanic signals at Deception Island volcano, Antarctica: Wave field analysis and source modeling”. In: *Journal of Geophysical Research: Solid Earth* 105.B6 (2000), pp. 13905–13931. DOI: 10.1029/2000JB900013. eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2000JB900013. URL: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2000JB900013.

[36] B. Chouet and R.S. Matoza. “A multi-decadal view of seismic methods for detecting precursors of magma movement and eruption”. In: *Journal of Volcanology and Geothermal Research* 252 (2013), pp. 108–175.

[37] C. Hibert, F. Provost, JF. Malet, A. Maggi, A.Stumpf, and V. Ferrazzini. “Automatic identification of rockfalls and volcano-tectonic earthquakes at the Piton de la Fournaise volcano using a Random Forest algorithm”. In: *Journal of Volcanology and Geothermal Research* 340 (2017), pp. 130–142.

[38] A. Kendall and Y. Gal. “What Uncertainties Do We Need in Bayesian Deep Learning for Computer Vision?” In: *NIPS 31th conference* abs/1703.04977 (2017).

[39] S. J. Pan and Q. Yang. “A Survey on Transfer Learning”. In: *IEEE Transactions on Knowledge and Data Engineering* 22.10 (2010), pp. 1345–1359. ISSN: 1041-4347. DOI: 10.1109/TKDE.2009.191.

[40] K. Choi, G. Fazekas, M. Sandler, and K. Cho. “Transfer learning for music classification and regression tasks”. In: *CoRR* abs/1703.09179 (2017). arXiv: 1703.09179. URL: http://arxiv.org/abs/1703.09179.

[41] Y. Jiang, D. Wu, Z. Deng, P. Qian, J. Wang, G. Wang, F. Chung, K. Choi, and S. Wang. "Seizure Classification From EEG Signals Using Transfer Learning, Semi-Supervised Learning and TSK Fuzzy System". In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 25.12 (2017), pp. 2270–2284. ISSN: 1534-4320. DOI: 10.1109/TNSRE.2017.2748388.

[42] E. Lima, X. Sun, J. Dong, H. Wang, Y. Yang, and L. Liu. "Learning and Transferring Convolutional Neural Network Knowledge to Ocean Front Recognition". In: *IEEE Geoscience and Remote Sensing Letters* 14.3 (2017), pp. 354–358. ISSN: 1545-598X. DOI: 10.1109/LGRS.2016.2643000.

[43] J. Yosinski, J.Clune, Y. Bengio, and H. Lipson. "How transferable are features in deep neural networks?" In: *Advances in neural information processing systems*. 2014, pp. 3320–3328.

[44] W. Thelen, M. West, and S. Senyukov. "Seismic characterization of the fall 2007 eruptive sequence at Bezymianny Volcano, Russia". In: *Journal of Volcanology and Geothermal Research* 194.4 (2010), pp. 201 –213. ISSN: 0377-0273. DOI: https://doi.org/10.1016/j.jvolgeores.2010.05.010. URL: http://www.sciencedirect.com/science/article/pii/S0377027310001708.

[45] A. Bueno, A. Díaz-Moreno, S. De-Angelis, C. Benítez, and J. M. Ibáñez. "Recursive Entropy Method of Segmentation". In: *Seismological Research Letters (accepted)* 90.4 (2019), pp. 1670–1677. DOI: https://doi.org/10.1785/0220180317.

[46] J. M. Ibáñez, M. C. Benítez, L. Gutiérrez, G. Cortés, A. García-Yeguas, and G. Alguacil. "The classification of seismo-volcanic signals using Hidden Markov Models as applied to the Stromboli and Etna volcanoes". In: *Journal of Volcanology and Geothermal Research* 187 (2009), pp. 218–226.

[47] M. Beyreuther, R. Barsch, L. Krischer, T. Megies, Y. Behr, and J. Wassermann. "ObsPy: A Python toolbox for seismology". In: *Seismological Research Letters* 81.3 (2010), pp. 530–533.

[48] D. G. Childers, D. P. Skinner, and R. C. Kemerait. "The cepstrum: A guide to processing". In: *Proceedings of the IEEE* 65.10 (1977), pp. 1428–1443. ISSN: 0018-9219. DOI: 10.1109/PROC.1977.10747.

[49] M. Sokolova and G. Lapalme. "A systematic analysis of performance measures for classification tasks". In: *Information Processing & Management* 45.4 (2009), pp. 427–437.

5 | Bayesian Monitoring of seismo-volcanic dynamics

This chapter defines the Bayesian monitoring framework adopted for continuous detection, segmentation, and classification of seismo-volcanic data streams. The definition of how statistical uncertainties intersect with geophysical interpretation is also formulated here. Then, we provide visual means to interpret the uncertainty. Finally, we highlight that changes in the seismicity can be detectable by the uncertainty of the model. This article is published and available online at IEEE Transactions on Geoscience and Remote Sensing, with the following journal metrics: (IF), JCR 2020: (5.855). Remote Sensing (Rank 5/32) (Q1). Geochemistry and geophysics (Rank 5/85) (Q1). Imaging Science and photographic technology (Rank 5/27) (Q1). Engineering, electrical, and electronic science (Rank 27/266) (Q1). We reproduce the draft accepted by the journal, which can be cited as:

1. **A. Bueno**, L. Zuccarello, M.C. Benitez, S. De Angelis, J.M. Ibánez. Bayesian Monitoring of Seismo Volcanic Dynamics. IEEE Transactions on Geoscience and Remote Sensing. DOI: 10.1109/TGRS.2021.3076012. Geoscience and Remote Sensing, vol. 58, no. 2, pp. 892-902, Feb. 2020.

Bayesian Monitoring of Seismo-volcanic Dynamics

Angel Bueno, Carmen Benítez, Luciano Zuccarello, Silvio De Angelis, and Jesús M. Ibáñez

Abstract

Methods for volcano monitoring that are based on analysis of geophysical data often rely on deterministic approaches without considering the complex and dynamic nature of volcanic systems. In order to detect subtle changes within seismic sequences associated with volcanic unrest, specialized workflows for data classification and analysis are required. Here, we present an inference framework based on Bayesian Deep Learning as a probabilistic proxy, which allows monitoring continuous changes in seismic activity at volcanoes. This architecture has been designed and trained to detect and classify individual earthquake transients from continuous seismic data recorded in volcanic environments. We tested this new framework by analyzing seismic data associated with eruptions at Bezymianny Volcano (Russia) during 2007. Our results demonstrate efficient signal detection and classification accuracy, and effective detection of changes in the volcanic system in the hours preceding eruptive activity. This approach can be extended to other volcanoes and earthquake-prone areas, and demonstrates a new application of deep learning in the field of seismic monitoring.

I. Introduction

Forecasting of volcanic eruptions is grounded in the ability to identify changes in metrics derived from the analysis of geophysical time series, and in the successful implementation of

A. Bueno and C.Benítez are with the Department of Signal Theory, Telematic and Communications, University of Granada, Spain. e-mail: angelbueno@ugr.es.

L. Zuccarello is with the Istituto Nazionale di Geofisica e Vulcanologia (INGV), Sezzione di Pisa, Pisa, Italy.

S. De Angelis is with the Department of Earth, Ocean and Ecological Sciences, University of Liverpool, UK.

J. M. Ibáñez is with the Andalusian Institute of Geophysics and Department of Theoretical Physics and Cosmos, University of Granada, Spain

This work was supported by TEC2015-68752 (KNOWAVES), PID2019-106260GB-I00 (FEMALE), NERC Grant NE/P00105X/1, and by the European Union's Horizon 2020 Research and Innovation Programme Under Marie Sklodowska-Curie Grant Agreement No. 798480.

such data analysis frameworks for pattern recognition in real- or quasi-real-time [1]. Volcano seismology remains one of the most important tools for volcano monitoring. Volcanic activity is known to generate a variety seismic signals, which represent evidence of multiple complex processes acting within the volcanic system. Changes in the style of seismicity, and its rates of occurrence and magnitude, are frequently recognized as precursors to eruptions [2]. Recent technological advances, the reduced costs of equipment, and improved open access policies for scientific data have meant that more information has become available. Research efforts in recent years have, thus, focused on improving our ability to process large amounts of seismic data efficiently; in particular, this has promoted the use of deep learning [3], [4], [5], [6], [7], [8], [9]. Pre-trained deep neural networks (DNNs) and recurrent neural networks (RNNs) have been explored in multi-class discriminative frameworks [10], [11], showing good performance in selected periods of seismic unrest (*snapshots*); for time-intervals associated with well-identified volcanic activity, they are used to *fine-tune* monitoring algorithms [12], [13], [14], [15]. These workflows have, however, been implemented without consideration for the assessment of uncertainty, which can affect their performance in certain conditions [16]. These algorithms can experience significant performance drops due to limitations in the training seismic database and procedures that do not account for the non-stationary evolution of volcanic unrest and of its seismic fingerprint. These issues can be partly mitigated through constant updating of the training dataset, a time-consuming task that involves continual manual inspection of a large volume of seismic data and re-training of the neural network. A probabilistic approach to deep learning can partly address these challenges [17]. Recent work by [18] has demonstrated the capabilities of deep learning to operate as a multi-volcano classifier by re-using accumulated seismic knowledge throughout time. The uncertainties derived from the Bayesian formulation in [18] serve as an indicator of changes in the frequency content distribution of the seismic signal, casting differences in eruptive periods as transfer learning scenarios. Although this approach helps to mitigate some past issues it still requires expert-validated, manually segmented training datasets.

In seismo-volcanic monitoring, uncertainty characterizes the dispersion of values that can reasonably be attributed to the monitored process. In this context, aleatory uncertainty refers to the notion of randomness; that is, the variability in the outcome due to inherently random, unforeseen effects. Epistemic uncertainty refers to uncertainty caused by the model´s lack of knowledge about the complete data distribution. In other words, epistemic uncertainty refers

to the reducible part of the (total) uncertainty, whereas aleatory uncertainty refers to the non-reducible part. The application of epistemic and aleatory uncertainty within seismo-volcanic monitoring can be defined as a measure of how much data resembles known conditions, aiding in the discovery of data variations and the temporal evolution of a physical system. Therefore, the concept of uncertainty can be used to control the quality of the physical measurements of a volcanic system and as an indicator of the evolution of these physical measurements over time.

A homogeneous volcanic system generates waveforms that resemble one another and lie within the known support data distribution. When we deploy a trained system for seismo-volcanic monitoring, the uncertainty varies according to whether there is an increase in the homogeneity of the data or, to the contrary, the data cease to resemble known conditions. This evolution is associated with changes in the characteristics of seismo-volcanic signals (e.g., increases in energy, the existence of volcanic tremor, a shift in the frequency of the signals). With an increase in seismic noise, for example, a signal would cease to resemble another of the same class because a random noise level has been added to it. In this case, while the aleatory uncertainty will increase, the growth of epistemic uncertainty will be more evident and more useful for defining the change in the volcanic system. However, the nominal value of the aleatory uncertainty and the difference with respect to the previous period reveals the significance of the produced change; that is, whether epistemic variation is due to new data heterogeneity (i.e., mild increments in the signal-to-noise ratio [SNR]) or to more significant internal variations (e.g., volcanic inflation, change in seismic impedance, etc.). For this reason, randomness in monitoring data can imply a change in data and/or in the environment, always conditioned to the available training data. If the analyst can improve knowledge by taking more refined measurements, it may be sensible to consider variables that exhibit dependence on those measures and that can explain stochasticity in the data.

In this study, we introduce a new Bayesian deep learning (BDL) method for the detection, segmentation, and classification of seismo-volcanic data streams, including uncertainty quantification. We developed a new hybrid architecture that combines segmentation and temporal sequence classification for the simultaneous identification and separation of seismic signals from background noise. We tested and confirmed the capabilities of our model on a well-known, short-lived and high-energy eruption at Bezymianny Volcano in 2007 (Russia) [19], recorded by a near-field seismic station. Finally, we considered the potential future use of Bayesian uncertainty in volcano monitoring.

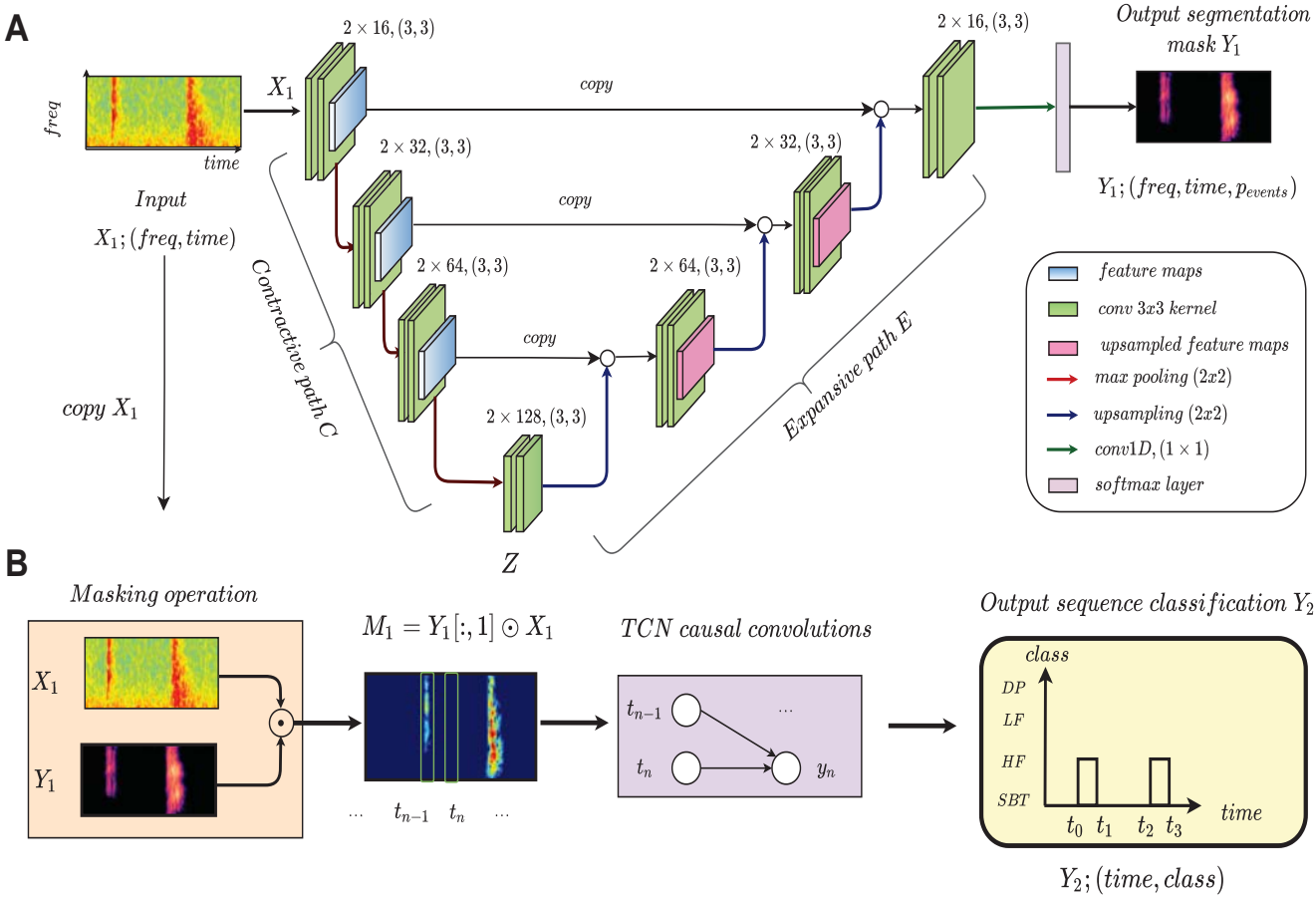


Fig. 1. Diagram of the proposed hybrid architecture to perform seismic event detection (A) and continuous sequence classification (B) of seismo-volcanic events. The input is given by the two-dimensional spectrogram, X_1, of size $(freq, time)$. The outputs are given by the segmentation mask Y_1 of size $(freq, time, p_{event})$ and Y_2 the labeled sequence $(time, class)$. The X_1 matrix is embedded into a latent representation Z, in which the network learns to untangle frequencies from noise. This representation is later upsampled and decoded through the expansive path, *E*, into an event detection mask highlighting the active frequencies of the data stream, X_1. The temporal Convolutional Neural Network (TCN) probes this representation to perform seismo volcanic sequence recognition, producing Y_2, size $(t, n_{classes})$, the per-frame classified data stream.

The remainder of this paper is organized as follows: Section 2 introduces the designed hybrid architecture and the Bayesian methodology. Section 3 connects Bayesian theory and volcano monitoring. Section 4 summarizes the eruptive phases studied. In section 5, we introduce the experimental methodology. Section 6 presents the obtained segmentation and classification results. Finally, section 7 closes the study with the conclusions and future research directions.

II. Bayesian Monitoring

In volcano monitoring, background seismic tremor is a natural and continuous process implicitly related to volcanic unrest [15], [20]. The separation of seismic signals from the background noise remains a challenging task, as they are located in a very narrow frequency band, overlapping many other seismo-volcanic events [21], [22]. However, identifying the level of seismic background tremor can (i) help to perform seismo-volcanic event separation, segmenting only those signals caused by volcanic unrest, even if noise levels prevail over the target events; (ii) bound the classification of seismic events to the retrieved frequencies in the trace; and (iii) use frequency masks to gain direct knowledge of potential volcanic sources and the environment in which the seismograms are recorded.

To this end, our model learns two mapping operations (Figure 1). First, the network parses a time-frequency representation of the seismic data stream (spectrogram), which we note as X_1, into a segmentation mask. The learn mask Y_1 is overlapped with the input spectrogram to generate an enhanced feature map representation M_1; this feature map contains the broad spectrum of the seismic events and the background noise. The second mapping operation performs continuous event recognition with M_1 as the input, generating Y_2, the output sequence of geophysical labels. This Bayesian approach permits quantification of the total uncertainty of the model, including the uncertainty in classification and segmentation tasks, the frequency variations of the data streams, and potential data drift.

A. Segmentation network module

The learning of target frequencies from other background signals has been applied in different contexts, where *background* can have very diverse meanings, including vocals and music [23], speech denoising [24], and, very recently, seismic signals for earthquake seismology [25]. Yet, in volcano seismology, the word *background* might have an implication of rapid magma ascent, among many other interpretations that require end-to-end methodologies that can exploit the learn mask in the monitoring outcome [22]. Our segmentation framework adopts these approaches and proposes an encoder/decoder architecture, designed to retrieve the frequency range of seismo-volcanic events from copious and sustained tremor noise, applying multiple array-wise convolution operators at a given frequency range [26]. We cast the problem of learning the seismic background tremor as a segmentation task; with $S(t, f)$ the seismo-volcanic signal, and $T(t, f)$ the seismic background tremor. The magnitude spectrum of the short-time Fourier transform

(STFT), size $(freq, time)$, is forwarded to a neural network to output the segmentation mask Y_1, size $(freq, time, p_{event})$, with p_{event} the index to the target seismo-volcanic events or noise. As a final step, this learned representation Y_1 is forwarded to a masking operation, that is later used by the classification module.

The training target of this segmentation module is given by the ideal ratio mask (IRM) [27]; defined as the ratio of the given seismo-volcanic monitoring event, $S|(t, f)|^2$ and the mixed noisy representation of the data stream spectrum, $S|(t, f)|^2 + T|(t, f)|^2$. From the ground truth label, the IRM is obtained at the frequencies corresponding to the trace, and later, mixed with noise. The complementary mask yields the noise mask. A final binary operation is applied to compose the Y_1 target mask for both events.

The encoder (Figure 1, contractive path *C*), is based on a CNN with three successive convolutional layers, containing two convolutions each. The number of filters in each convolutional layer is twice that of the previous one: from the first to the last layer in *C*, 16, 32, and 64 filters are used, with a kernel size (3x3). The latent path, noted in this work as *Z*, is a single layer CNN with two convolutions operations, kernel size (3x3), and 128 filters. This *Z* path contains a set of sparse feature maps with refined frequency information from the input feature matrix X_1.

The decoder, noted in Figure 1 as the expansive path, *E*, is built to keep the symmetry with the contractive path *C*. The decoding steps comprises three up-sampling operations and three convolutional layers that transform the latent features Z into a segmentation map, Y_1. The up-sampling operation halves the number of feature maps from the Z path, to permit tensor concatenation of upsampled maps with convoluted feature maps from the encoder via skip connections. These connections transmit the convoluted feature maps in each layer of the encoder as an effective means of providing more informative fine-grained features with the decoding steps. Furthermore, the successive skip connections and up-sampling operations help to reassemble the time and frequency matrix dimensions of the input Y_1, later used by the classification network component.

Each of the three layers in *E* mirrors the number of convolutions and filters from *C*: two convolutional operators with kernel size (3x3), and with 64, 32, and 16 filters, from first to last decoding layer. A 1x1 convolution is applied to assign per-class frequency probability, producing the output segmentation mask Y_1. This output Y_1 is a map of interconnected frequencies of all the seismic events in our continuous trace. The network automatically indexes those frequencies

and performs an overlapping operation with the input feature X_1, resulting in a masked matrix M_1:

$$M_1 = Y_{1_{[:,1]}} \odot X_1 \tag{1}$$

with $[:, 1]$ the indices over the detected frequencies, and $\odot$ the Hadamard product. The distilled map M_1 contains all the original frequencies from X_1 and can be interpreted as an enhanced version of the input spectrogram. The background noise has been eliminated from the trace, and the presence/absence of events in the full seismic frequency range is marked. These feature representation masks, M_1, compose a structured representation of the input data suitable for temporal modeling by the TCN component.

B. Temporal seismo-volcanic classification

The seismo-volcanic sequential module is a two-block temporal Convolutional Neural Network (TCN) with causal constraint dilated convolutions; that is, a convolutional operator that considers past contextual information without extensive computations [28] [29]. The causality property refers to the mathematical property that temporally bounds future information to past frames. Therefore, given the enhanced frequency map $M_1 = (m_{t0}, m_{t1}, ..., m_t)$ as a sequence of temporal frames, causal convolution operators assume that the prediction at any time t for $Y_2 = (y_{t0}, y_{t1}, ..., y_t)$, depends only on previous frames, $(m_{t0}, m_{t1}, ..., m_{t-1}, m_t)$, but not on future inputs, m_{t+1}. The causal property analyses the past sequence linearly, expanding the number of processed past frames with increasing network depth. The dilated convolution operator, $F(t)$, permits exponentially larger receptive fields with broader past frame sequence contextualization. It is defined as:

$$F(t) = \sum_{i=0}^{k-1} f(i) M_{t-d \cdot i} \tag{2}$$

where $f(i)$ is the i^{nth} filter in layer i, k is the filter size, and d is the dilation factor. Note that equation 2 bounds each frame to consider $t - d \cdot i$ past frames in the sequence M_1, for any given dilation d. Hence, per-class prediction at time t depends solely on the number of past frames, and no future frames from the seismic sequence are analyzed. This framework is employed in online seismic sequence classification from the denoised representation M_1. To this end, our convolutions are made causal as in equation 2, with a kernel convolution size 3, 32 filters, and exponential dilation rates of [2, 4, 8] to cover the full segmented sequence. A final 1-D convolutional layer is added before the final *softmax* layer with a small kernel size of 3

to improve prediction smoothness of the sequence results. Dropout is added after each dilated convolution for regularization and learning stabilization.

C. Bayesian deep learning

Bayesian neural networks (BNNs) define its weights as probability distributions, $\omega = (\omega_1, \omega_2, ..., \omega_n)$, also known as *prior* $\omega \sim p(\omega)$. Given seismic dataset D as the set of paired data samples (x, y), with x as the processed input data stream and y as the target labels, this probabilistic approximation allows the computation of the posterior distribution of the network weights, $p(\omega|D)$ as:

$$p(\omega|D) = \frac{p(y|x,\omega) * p(\omega)}{p(y|x)} \tag{3}$$

where $p(y|x)$ is the *evidence*. The predictive distribution for new input data (x^*, y^*) is given as:

$$p(y^*|x^*, D) = \int p(y^*|x^*, \omega) p(\omega|D)\ d\omega \tag{4}$$

However, equation 4 is not analytically calculable, as the second term of the integral requires the evaluation of the posterior $p(\omega|D)$. In this regard, variational inference (VI) has emerged as the preferred choice in Bayesian inference; the posterior approximation is cast as an optimization procedure designed to find the closest tractable distribution $q_\theta(\omega)$ by minimizing the Kullback-Leibler divergence (KL) to the true posterior, $KL(q_\theta(\omega)||p(\omega|D)$. Recent work by [30] connected dropout regularization with VI in any arbitrary convolutional network structure to approximate the posterior distribution of weights. The KL-divergence for approximant $q(\omega)$ is formulated as:

$$KL(q(\omega)||p(\omega|D)) \propto -\int q(\omega) log\ p(y|x,\omega) d\omega + KL(q(\omega)||p(\omega)) \tag{5}$$

Note that the second term of the KL divergence acts to regularize the weights, keeping $q(\omega)$ from extreme deviations of the prior $p(\omega)$ but bounding the approximant towards $p(\omega|D)$. The nexus of $q(\omega)$ to the regularization term in equation 5 is the dropout technique. This approach, formally known as *Monte Carlo dropout* (MC-dropout), formulates $q_\theta(\omega)$ to the network posterior as ω, the set of weight matrices in the l^{th} layer ($\omega = \{W_l\}_{l=1}^{L}$), and θ as the variational approximate in which the optimization has to be performed:

$$q_{V_l}(W_l) = V_l \cdot diag[p_{l,i}]_{i=1}^{K_i} \tag{6}$$

where V_l is the set of variational parameters, and dimension $K_i \times K_{i-1}$, $p_{l,i}$ is the distribution of Bernouilli parameters. This mathematical formulation yields a VI foundation to approximate

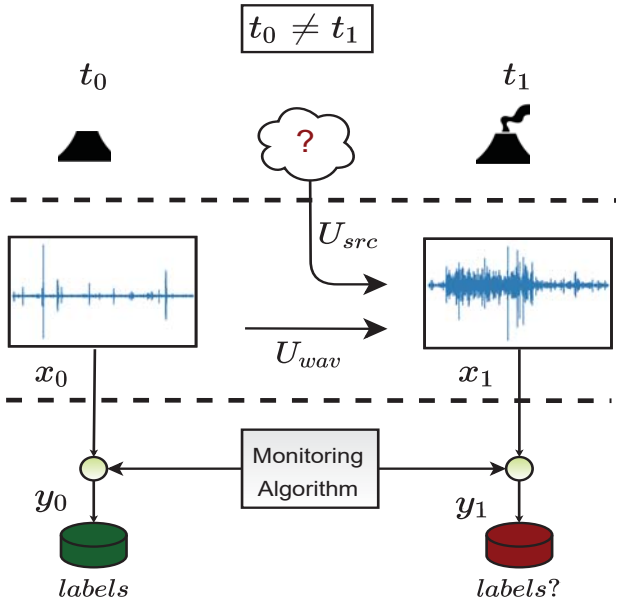


Fig. 2. Standard seismo-volcanic monitoring architectures based on static supervised learning. In practice, when an unforeseen *change* in a volcano alters the seismograms, this translates into a potential data shift that can compromises the performance of monitoring algorithm. Our framework integrates Bayesian theory with CNNs to monitor and perceive data changes due to U_{wav} and U_{src} from seismo-volcanic data streams.

equation 5 with an optimization target function and a regularization term [30]. Therefore, applying dropout in our convolutional architecture can produce an approximant $q(\omega)$ for predictive equation 4 that can be approximated by sampling the parameterized weights at T stochastic forward passes:

$$p(y^*|x^*, D) = \int p(y^*|x^*, \omega)p(\omega|D)\, d\omega \approx \int p(y^*|x^*, W_l)q((W_l)) \approx \frac{1}{T}\sum_{t=1}^{T} p(y^*|x^*, (W_l)_t) \quad (7)$$

where $(W_l)_t \sim q(W_l)$. This approximation leverages the performance prowess of deep learning with a Bayesian uncertainty quantification framework, which has been demonstrated to achieve outstanding performance in earthquake location [17] and isolated seismo-volcanic waveform classification [18].

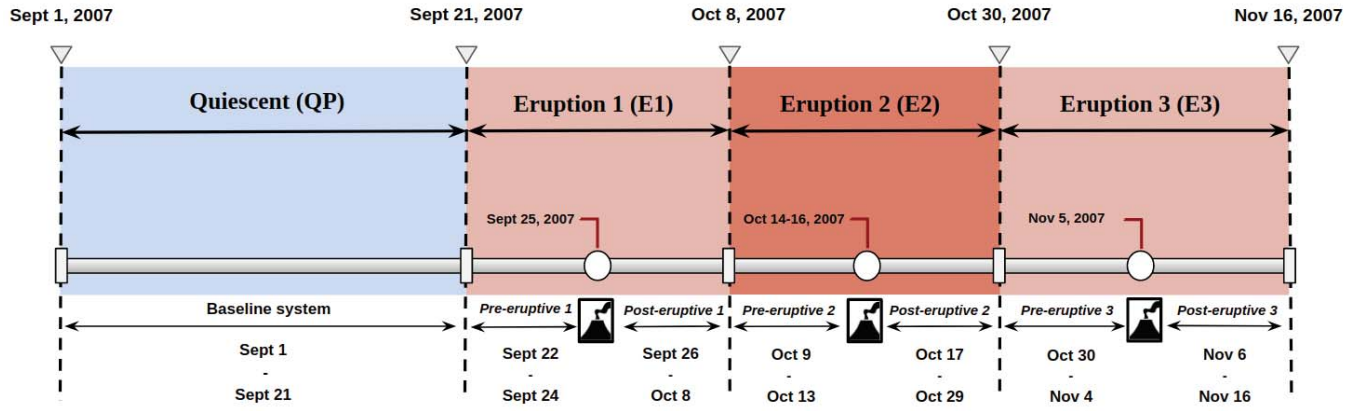


Fig. 3. Eruptive chronology of 2007 Bezymianny Volcano eruption. We subdivided the eruptive sequence into four minor groups: Quiescent Period (QP), Eruption 1 (E1), Eruption 2 (E2), and Eruption 3 (E3). The QP covers a time-span of usual background seismic data. The unrest periods (E1, E2, and E3) are each subdivided into pre- and post-eruptive episodes. This division correlates with past geophysical bulletins and is designed to obtain insight into the statistical links of seismic wavefield evolution and uncertainty quantification.

III. Seismo-volcanic Uncertainties

Recorded seismograms gather a set of informative parameters related to the seismic wavefield and the volcanic environment in which they are generated (i.e., spectral content, P/S waves amplitude, coda length, among many others). Figure 2 depicts the studied problem from a machine learning perspective; the data used to fine-tune seismo-volcanic monitoring systems may not represent the situations later encountered. This situation, defined as data drift, arises if the data samples (seismograms) from an initial data support distribution are altered owing to unforeseen changes in the volcano. Under these new conditions, we can estimate the uncertainty in the segmentation and classification module to determine if the new recorded seismograms depart from the initial distribution of frequencies. This section describes the uncertainty quantification approach for classification tasks in seismo-volcanic monitoring applications.

A. Statistical uncertainties

The sources that could drive changes in a volcano include an unknown number of latent, heterogeneous variables that contribute to the overall alteration on the seismic observable [2], [21]. From statistical standpoint, the uncertainties can be categorized as epistemic or aleatoric, based on whether more data can reduce their estimated values [31]. The Bayesian framework proposed in section II can gauge both uncertainties by exploiting the mathematical relationship

between the covariance and mean vector. This approach, derived by [32], decomposes the model uncertainty into aleatoric and epistemic, offering a probabilistic proxy to uncertainty estimates without additional network parameterization. Hence, given the sampled dropout-masked model weights after T stochastic forward passes $\{\hat{\omega}_t\}_{t=1}^{T}$, and the predictive probability distributions for realization of each sampling step, that is $\tilde{p_t} = p(\hat{\omega}_t)$, the aleatoric and epistemic uncertainty can be computed as:

$$U_t = \underbrace{\frac{1}{T}\sum_{t=1}^{T}\left(diag\left(\tilde{p_t}\right) - \tilde{p_t}^{\otimes 2}\right)}_{\text{Aleatoric}} + \underbrace{\frac{1}{T}\sum_{t=1}^{T}(\tilde{p_t} - \overline{p})^{\otimes 2}}_{\text{Epistemic}} \tag{8}$$

where $\overline{p}$ is the averaged p_t over all T stochastic forward passes from the dropout variational distribution. The sum of both terms, U_t, is the total uncertainty of the model. In a monitoring system, estimates of U_t can be used to infer shifts from the initial data distribution (i.e., changes in the frequency content of the events, or the volcanic medium).

B. Monitoring uncertainties

A monitoring framework can probe the estimated uncertainties as a proxy to detect if the recorded seismograms depart from the initial data distribution. Hence, the total statistical uncertainty U_t of a seismo-volcanic monitoring algorithm comprises two terms, the observed seismogram variability, and the inherent randomness of the monitored seismic wavefield:

$$U_t = U_{src} + U_{wav} \tag{9}$$

where U_{wav} is the uncertainty associated with the seismic wavefield, and U_{src} is the uncertainty linked to the monitoring process. Equation 9 intersects with equation 8 and is based on whether the algorithm can identify reducible sources of uncertainties. In this context, epistemic uncertainty arises if continuous seismograms are not affected by unexpected changes; gathering more data can refine the original data space approximation. The uncertainty linked to the unforeseen changes driving the observed process corresponds to the aleatory uncertainty. In this framework, the aleatory uncertainty estimates a single value that reflects the aggregated contribution of all irreducible sources of uncertainty, ranging from the inherent unpredictability of the eruption itself to how seismic signals interact with the environment [16] [33]. From a monitoring perspective, we can link these uncertainties to data changes; when evident alterations in the seismic wavefield are

recorded, increments in U_{src} and U_{wav} must be observed, which can lead to different seismograms, and thus, to the assumption that the volcano has changed.

IV. Bezymianny Dataset

To study potential changes in uncertainty according to eruptive and non-eruptive stages and to observe possible similar behaviors in the seismic wavefield, we selected the 2007 eruptive sequence at Bezymianny Volcano. This sequence included three well-identified eruptions, those of 25 September, 14–16 October, and 5 November; Figure 3 summarizes the whole eruption chronology. This eruptive crisis represents a non-stationary environment characterized by seismic re-activations, eruptions, structural failures, and generic unrest over a short period. The seismic network at Bezymianny includes numerous seismic stations; among them, we selected data from the BELO station owing to the well-tested quality of its seismic records for the 2007 eruption [19], [34]. The BELO station is ideal for recording potential changes in waveforms, as it is 1.7 km from the eruptive center and can be considered a very near field seismic station with negligible attenuation [34].

Following the volcanological observatory bulletins and previous reports about eruptive time-lines [19], [34], [35], the seismic dataset was divided into four well-differentiated periods: the Quiescent Period (QP), Eruption 1 (E1), Eruption 2 (E2), and Eruption 3 (E3). Figure 3 depicts a detailed overview of the dataset organization, covering unrest periods and volcanic evolution across three different scenarios. This dataset organization manages the sheer volume of seismic streams while offering a well-defined chronological structure for testing, and a chronological uncertainty framework based on a well-known, historical eruption. The QP period corresponds to a segment of 20 days during which ordinary volcanic activity was registered. The definitions of eruptions E1, E2, and E3 were arranged according to volcanological observatory reports, and were used to compare potential changes when a volcanic crisis occurs. Each eruption was further subdivided into pre- and post-eruptive periods to achieve fine-grained precision in uncertainty estimates before and after the main eruptions.

A. Bezymianny seismic events categorization

Based on preliminary geophysical knowledge of this volcano, the whole dataset was curated with precise annotations and temporal onsets at the waveform level. Active volcanic systems generate a full extent of seismo-volcanic signals. They receive different labels according to their

generative volcanic sources, which are not always homogeneous or standard (a comprehensive summary of naming conventions and associated source models are given in [18]). Nevertheless, the past decade has adopted a novel classification scheme to ease source terminology in favor of a unified data taxonomy. This categorization scheme, based on signal duration and frequency content [36], includes high and low-frequency earthquakes, volcanic tremor, and other superficial signals (e.g., rockfalls, lahars, pyroclastic flows). By adopting this geophysical criterion, the label of an event can be assigned using the frequency index (FI), a logarithmic interpretation of the spectral frequency ratio (FR), given as:

$$\text{FI} = \log_{10}(FR) = \log_{10}\left(\frac{A_{high}}{A_{low}}\right) \tag{10}$$

where A_{low} and A_{high} are the mean amplitude of high and low spectral bands, respectively. We adopted the labeling criteria of [19], with A_{low} and A_{high} defined as $[1-5]\ Hz$ and $[6-13.5]\ Hz$. Each extracted event was labeled according to the FI value ($FR = 0.5$) and duration of high- (HF) and low-frequency (LF) events, seismic background tremor (SBT), and debris processes (DP). Note that the logarithmic spectral ratio forces high values of the FI to be associated with higher frequency mechanisms (e.g., brittle fracture). In contrast, lower values correspond to seismic events with narrow lower frequency bands (e.g., soft ruptures [37]). The full eruptive dataset was curated in two main steps. First, semi-supervised segmentation and categorization at the stated FI frequency bands were performed using an entropy-based algorithm, REMOS [38]. This first procedure generates a set of preliminary event onsets and associated classes. Then, all events are translated back into their original sequences and visually inspected and corrected using PICOSS (Python Interface for the Classification of Seismic Signals); a data-curator graphical interface that allows manual confirmation or modification of annotated labels, along with the onset times of detected events [39]. The expert-reviewed data catalog from September to November 2007 was compiled into a sequence-level dataset.

V. Experimental Methodology

We performed two experiments to investigate how the proposed hybrid framework performs in a non-stationary seismo-volcanic monitoring application. With the first set of experiments, we aimed to obtain a broad understanding of the performance of the method. By establishing baselines for each of the chronological periods as independent *data-snapshots*, we simulated the

TABLE I

DATASET ORGANIZATION FOR THE BASELINE SYSTEMS, B-QP, B-E1, B-E2, AND B-E3

(a)

Baseline system	Dev. Set	Training (hrs)	Nevents	Blind test	Testing (hrs)	Nevents
b-QP	QP	480	3844	post-1	312	1941
b-E1	pre-1, post1	384	2422	post-2	288	9060
b-E2	pre-2, post-2	384	10295	post-3	264	1369
b-E3	pre-3, post-3	576	4681	QP	480	3844

(b)

	Dev. set isolated events			Blind-test isolated events		
Baseline system	**HF**	**LF**	**DP**	**HF**	**LF**	**DP**
b-QP	1825	1241	778	303	1130	508
b-E1	415	1306	701	3698	2383	2979
b-E2	3984	2883	3428	138	2254	920
b-E3	768	2839	1074	1825	1241	778

generic situation in which an algorithm, fine-tuned for initial monitoring conditions, is evaluated after major changes have happened at the volcano (eruption). This can help to determine if the error rate decreases as a potential indicator of data drift. Table I (a) depicts the dataset organization with the continuous baseline system, which we defined as b-QP, b-E1, b-E2, and b-E3. The total number of hours of each training baseline period, along with the total number of detected events N_{tot} in the seismic data stream, are also given. In Table I (b), we depict the overall per-class extracted and isolated events from the continuous trace, for each baseline system. The training of each baseline system was performed in isolation from the other baselines, with no data leakage from past or future periods (see Figure 3) during the training stage. Thus, each baseline system was independently trained and tested on each continuous *data-snapshot*, with 75%–25% data cross-validation split. We then performed blind-tests for each post-eruptive period after the next eruption in time (see Figure 3). Hence, for b-QP, b-E1, and b-E2, we blind-tested with the post-1, post-2, and post-3 periods, respectively. For the b-E3 system, the blind test-set corresponded to the entire QP period, where the monitoring conditions differed after the three eruptions [19].

In the second experiment, we aimed to determine if our framework, starting from the QP and extending for the whole eruptive chronology, can detect when a data drift is happening and

TABLE II

OVERALL SEGMENTATION INTERSECTION OVER UNION (IOU) AND CLASSIFICATION METRICS (PR, RC, F1, ACC) FOR THE BASELINE SYSTEM AND BLIND-TEST PERIODS

	Test set					Blind test				
Period	**IoU**	**PR**	**RC**	**F1**	**Acc**	**IoU**	**PR**	**RC**	**F1**	**Acc**
b-QP	0.98	0.97	0.95	0.96	0.97	0.94	0.94	0.92	0.93	0.92
b-E1	0.98	0.96	0.94	0.95	0.95	0.97	**0.86**	**0.84**	**0.85**	**0.85**
b-E2	0.96	0.91	0.87	0.89	0.89	0.98	0.90	0.85	0.86	0.87
b-E3	0.97	0.91	0.86	0.88	0.87	0.98	0.96	0.92	0.94	0.94

asses the severity of the effect on monitoring system performance. To this end, we selected the best model from the QP period, (25% of the whole Bezymianny dataset) and blind-test with the rest of the eruptive period, for each temporal sub-partition of the dataset. This testing strategy was motivated to provide a model with highly granular resolution of data partitions in order to refine the time interval in which data change can be detected. We introduced *monitoring uncertainty maps* as a visual means to provide monitoring interpretability of the connections between waveforms and uncertainty. Finally, regardless of the dataset organized, we evaluated the temporal evolution of the uncertainty for the whole dataset based on a broadly used monitoring tool (energy), in order to link the monitored process with the estimated uncertainty.

A. Feature extraction

For each day in the eruptive chronology, the seismic data streams (100 Hz) are filtered in the range $[1 - 13.5]\ Hz$, and windowed with a 5-min window (30,000 samples), with an overlap of 2.5 min. From these windows, we computed the input magnitude spectra X_1, using windows of 2.5 and 1.5 s overlap. As a final result, each X_1 is characterized by a matrix of dimensions $(freq, time)$, in our case, $(256, 256)$. We created the target mask Y_1 from the estimated IRM as explained in section II. A matrix of size $(freq, time, p_{events})$, with p_{events}; the index for the noise or frequency masks, composes the input Y_1. The per-class target sequence classification Y_2 of size, $(time, event)$ is also given during the training time, with $event$ belonging to any of the classes described in section IV. In this work, we consider SBT as a single class and part of the mask.

B. Optimization procedure

Training was perform using a multitask approach, in which a double loss function $\mathcal{L}_t$ was used: $\mathcal{L}_{seg}$ for segmentation and $\mathcal{L}_{class}$ for classification, with a softmax with cross-entropy. In the case of $\mathcal{L}_{seg}$, the softmax layer was applied over the frequency map of the target Y_1, whereas $\mathcal{L}_{class}$ was applied frame-wise. We selected the *Adam* [40] optimizer with an initial learning rate of 0.01, and *ReLU* activation function and mini-batch size of 64. The dropout probability was set to $p = 0.25$. We used early-stopping with a patience interval of 5 epochs over 300 training epochs to prevent over-fitting. A random search was performed over the most promising hyper-parameters derived from a Bayesian optimization procedure towards best classification and segmentation performance [41]. The Bayesian inference procedure was implemented as described in section II, with T = 20 *MC-dropout* sampling steps. We followed this optimization procedure for all of the system trained on each period, independently. Our entire experimental methodology was simulated using an NVIDIA Tesla P40 GPU (24 GB GPU memory) on a 64 GB RAM computer.

C. Monitoring metrics

The implemented volcano-seismic segmentation framework can be formulated as a multi-class classification and binary segmentation problem. The model is trained to identify the frequency ranges of seismo-volcanic events whilst categorizing seismic sequences. For the classification TCN module, the accuracy (Acc) measures the overall effectiveness. Precision (PR) quantifies the positively classified event rate, whereas the recall (RC) measures the sensibility of the system to recognize correct frames. The $F1$ score is a weighted average between the precision and recall, computed as:

$$F1_{score} = \frac{2 * (RC * PR)}{(RC + PR)} * 100\% \tag{11}$$

Thus, $F1$ represents an informative balance between the refinement of our classifier (PR) and the number of correctly detected events (RC) for any number of specific classes. These are standard metrics in sequence classification [42]. The goodness of a model to correctly identify class-boundaries is based on segmentation metrics. In this work, we report the *IoU* (Intersection Over Union); a standard metric in segmentation to compute the overlap between the target T and predictive mask P:

$$IoU(T, P) = \frac{\|T \cap P\|}{\|T \cup P\|} \tag{12}$$

where $\cap$ is the intersection between the pixels in the target and prediction mask, and $\cup$ is the union of pixels between both masks [26]. An IoU score of 1.0 is a perfect segmentation that fully overlaps the target and predictive masks.

VI. Results and Discussion

A. Segmentation baseline performance

We evaluated the monitoring capabilities of the proposed model to perform continuous seismic event recognition. Table 2 depicts the attained metrics for each selected eruptive period and test set. The implemented systems (b-QP, b-E1, b-E2, and b-E3) present high performance segmentation IoU and classification PR, RC, and F1, with consistent generalization for the blind-test partitions. The segmentation IoU remains above 96% for all of the test sets and 94% for the blind-test partitions. Such IoU generalization capabilities reflect the capacity of the segmentation module to isolate the set of frequencies that composes the seismic events in the continuous data stream. However, the IoU metric was 4% lower for the blind test period of the b-QP and 1% for the b-E1 system. In terms of classification metrics, the performance gap in the b-QP system is less evident, being only 3% less in F1, PR, and RC for the blind-test partition.

Similarly, the b-E2 system experienced performance drops of 1%, 2%, and 3% for PR, RC, and F1, respectively. However, all of the classification metrics dropped 10% for the b-E1 system when blind-tested on post-eruptive 2, which may be a indicator that a change occurred in the volcano. Note that the PR, RC, and F1 metrics were calculated from the enhanced M_1 spectrogram representation; hence, a change in the frequency spectra of the post-eruptive events, despite good segmentation metrics ($> 97\%$ in the b-E2 system), can be perceived by the classification module, which is fine-tuned to the initial distribution of frequencies from the previous eruptive period of the b-E1 system. For example, changes in the physical properties of the medium (i.e., seismic impedance) can shift the frequency spectra of events, compromising the performance of the model. The results of the blind-test for b-E2 suggest that predictions become less accurate after significant eruptions, pointing towards potential data drift.

B. Visualizing predictions

Based on the results, our hybrid framework can segment and classify seismic events from continuous data streams. From top to bottom, Figure 4 displays the raw waveform, STFT magnitude spectrum, retrieved frequencies of events (M_1), and continuous predictions for a set of volcano-seismic events. Overall, segmentation results suggest that the proposed architecture can provide accurate classification and segmentation for each event. The main frequency spectra of each seismo-volcanic event are highlighted within the main trace; for example, in Fig. 4 (a), there is a clear low-frequency band of LF earthquakes, ranging from 1.0 to the 3.5 Hz for the second LF earthquake. Fig. 4 (b) shows the high frequency of spectra of the DP and HF classes. The detection of target frequency components from broad spectra yields a denoised feature representation of the continuous TCN to perform enhanced seismic event segmentation and classification. Our network also generalizes data streams containing seismic events with copious background tremor noise. The good classification ability is evidenced in Fig. 4 (c) and Fig. 4 (d); the segmenter traces the intra-dynamic frequency range of the seismic events from the background, tailoring frequency segmentation towards accurate boundary detection and classification performance. The TCN correctly marks the time boundaries of the seismic signals, even for long events such as the cigar-shaped DP class. This larger contextual predictive power is due to the increased receptive fields and the stack of dilated causal convolutions, yielding a network that *keeps looking* forward for potential longer events, considering a broad time-span that incorporates a larger number of contextual past frames for the prediction at time t.

C. Evaluation of eruption uncertainties

Table 3 presents the performance metrics, along with the epistemic and aleatory uncertainties; the situation in which an observatory would not have time to analyze the sheer volumes of eruptive/post-eruptive data, given the substantial annotation effort required to provide new data for retraining the monitoring system. The estimated root mean square (RMS) of the background tremor, the spectral frequency ratio (FR) in the studied frequency bands ($A_{high} = [6 - 13.5]$ and $A_{low} = [1 - 5.0]\ Hz$, and equation 10) are also reported for each sub-partition of the dataset.

Overall, the segmentation metric IoU remains above 96% for all test-set periods. The classification performance shows high PR and $F1$ scores, with RC being at an adequate level (¿ 80%). For the pre- and post-eruptive periods, the system has a performance gap of 10% after the main October eruption, a seismological situation in which the volcano had suffered a significant

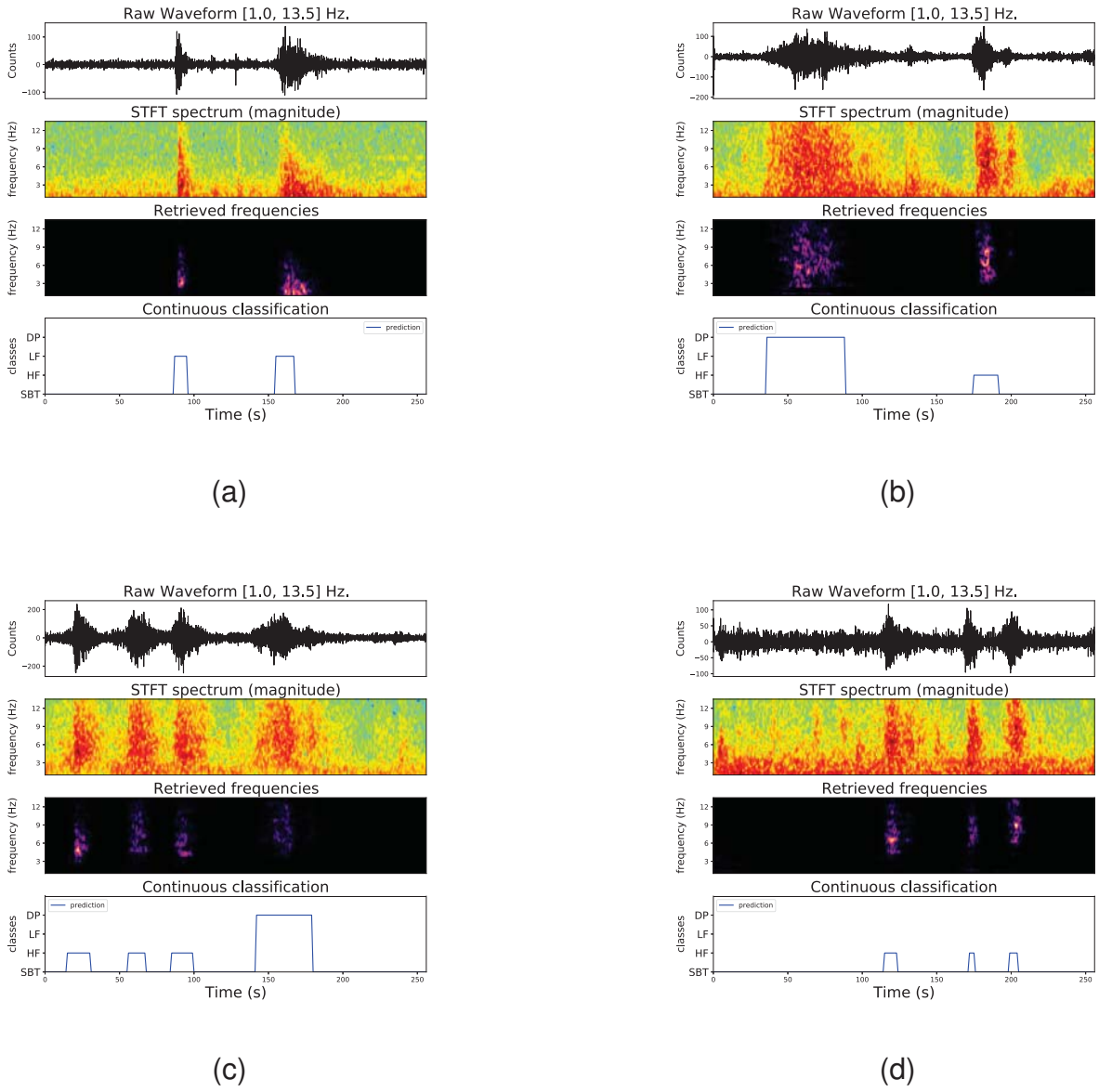


Fig. 4. Normalized waveform, input magnitude spectrum, retrieved frequencies, and continuous recognition of the signal. The network can segment event onsets, and produce a segmentation mask highlighting the active frequencies in the trace. The seismic background noise is learned efficiently, and the TCN can track the events despite copious background noise. Besides, the network boundary segmentation for larger debris processes (DP) is consistent with event and frequency onsets.

TABLE III

CHRONOLOGICAL PERFORMANCE, FREQUENCY RATIOS, NOISE ROOT MEAN SQUARE (RMS), AND OVERALL UNCERTAINTY QUANTIFICATION FOR THE 2007 BEZYMIANNY DATASET

	Geophysical params.		**Monitoring metrics**					**Segmentation**		**Classification**		
Period	**FR**	RMS_{noise}	**IoU**	**PR**	**RC**	**F1**	**Acc**	**Epistemic**	**Aleatory**	**Epistemic**	**Aleatory**	U_t
QP	0.34	10.01	0.98	0.97	0.96	0.96	0.96	0.016	0.00021	0.014	0.0008	0.0310
pre1	0.56	9.95	0.97	0.91	0.89	0.90	0.90	0.021	0.00034	0.016	0.0011	0.0384
E1	**0.55**	**77.16**	**0.95**	**0.83**	**0.81**	**0.82**	**0.81**	**0.034**	**0.00059**	**0.020**	**0.0018**	**0.0564**
post1	0.22	15.17	0.98	0.94	0.92	0.93	0.93	0.021	0.00034	0.017	0.0011	0.0394
pre-2	0.5	97.94	0.96	0.90	0.88	0.89	0.88	0.035	0.00056	0.019	0.0017	0.0563
E2	**0.58**	**405.34**	**0.93**	**0.68**	**0.64**	**0.65**	**0.66**	**0.057**	**0.00110**	**0.026**	**0.0032**	**0.0873**
post-2	0.42	16.34	0.96	0.88	0.85	0.86	0.86	0.039	0.00072	0.022	0.0019	0.0636
pre-3	0.38	14.79	0.97	0.91	0.89	0.90	0.89	0.027	0.00046	0.018	0.0013	0.0467
E3	**0.40**	**143.1**	**0.96**	**0.85**	**0.84**	**0.83**	**0.84**	**0.037**	**0.00073**	**0.021**	**0.0017**	**0.0604**
post-3	0.30	12.11	0.97	0.88	0.87	0.87	0.88	0.027	0.00047	0.018	0.0012	0.0466

explosion and was recovering towards stability [19]. Nonetheless, the system experienced a generalization gap in segmentation and classification metrics for all of the eruptions (E1, E2 and E3), reaching up to 31% in the F1 classification metric for the E2 eruption. Hence, changes in the eruptive periods are evidenced in the lowered performance and can be used as an indirect measurement to indicate the severity of data drift. If learning accuracy has diminished significantly, this can be a potential indicator of change. Yet, the metrics alone do not offer information about what is generating these changes, since the estimated uncertainty provides this information.

The uncertainties U_{ep} and U_{al} in both the segmentation and classification modules act as a coupled mechanism. Higher uncertainty values correspond to lower monitoring performance. The variation of U_t with the eruptive chronology can thus be explained by comparing the variations in frequency and noise levels for each time-lapse. Note that even if the FR, the initial frequency distribution, remains approximately the same, the uncertainty is reflective of the RMS levels of seismic background tremor. For example, post-eruptive period 3 presents a very similar FR but higher RMS than the initial data period QP. Yet, our model emits higher U_t values and lower performance metrics. The most elevated U_t values correspond to the eruptions, with E2 attaining the highest uncertainty in U_t, linked to peak RMS levels and FR changes. The epistemic uncertainty U_{ep} in Table 3 has slightly larger values than the computed aleatory U_{al}, for all of the eruptions, but is significantly greater than its counterpart in the classification module. The

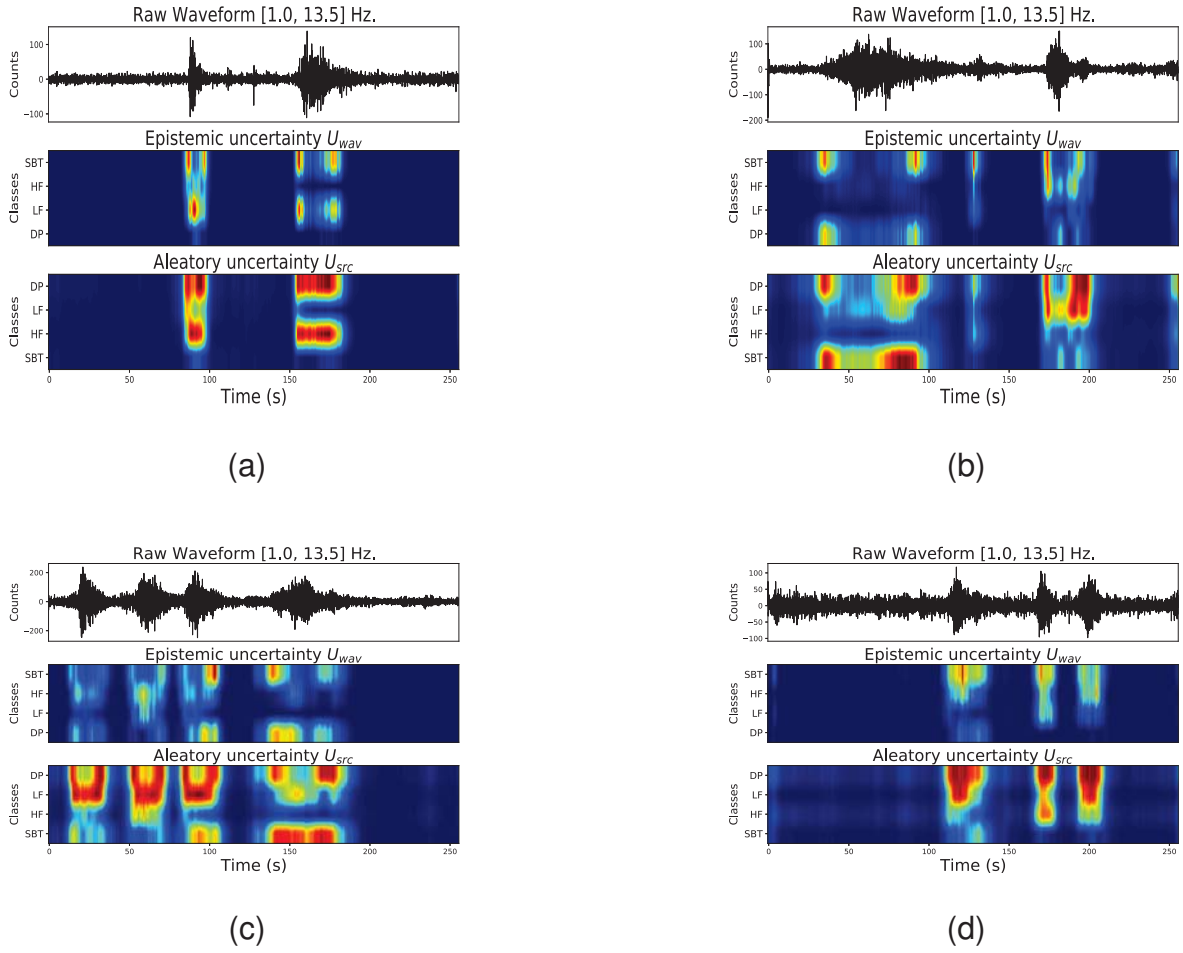


Fig. 5. Raw waveform (up), along with epistemic (middle) and aleatoric (bottom) uncertainty seismic maps. The association uncertainty-energy for each event is clearly visible. All uncertainties are associated with both seismic energy release and signal onset.

larger values in segmentation U_{ep} are due to the learning of the intra-frequency range of the volcanic tremor and seismic signals, on very narrow spectra which can overlap. The evolving wavefield yields seismograms and noise levels dependent on geophysical parameters, in which frequency, energy, or duration could have changed owing to volcanic source evolution.

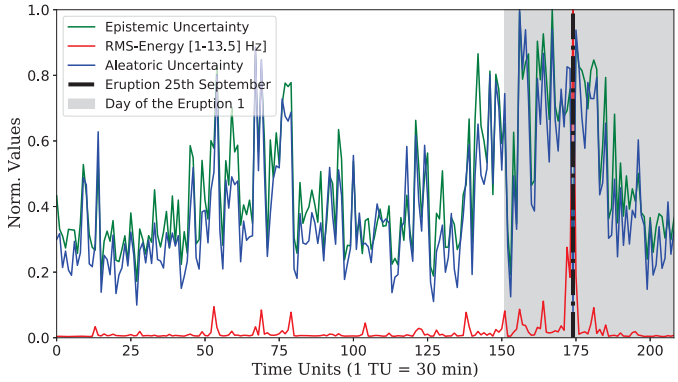
Consequently, changes in the sequences of data streams are connected to metric variations, consistent with the estimated epistemic and aleatory uncertainties. A volcano changes its physical state from one eruptive process to another, which means that the relevance of the seismic data is not applicable to a new eruptive period in a few days. Most seismic events are the consequence of fluid interaction (magma). A change in magma rheology, location, gas content, surface

temperature, and/or a change from open to closed conduit (or vice versa) would all imply a change in the waveform and spectral content, with substantial differences between signals observed within days. The proposed framework demonstrates that the predictive uncertainty estimates are meaningful as they are amenable to emerging power patterns in the seismic background tremor and noise levels. The differences in the power levels of the tremor and noise conditions across eruptive periods are equivalent, by analogy, to the distortion of images by varying the corruption intensities to evaluate the predictive uncertainty. Hence, the uncertainty evolves for each eruptive period, being an indicator of how and where the system evolves.

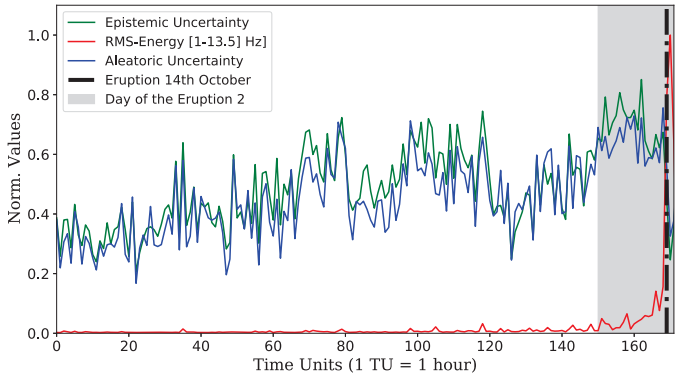
D. Visual uncertainty interpretation

We gathered numerical evidence that volcanic changes imply changes in the data space of the signals. We then decomposed equation 9 and associated waveforms to seismo-volcanic uncertainties. Visual interpretation of the uncertainty can help in *post hoc* seismic analysis and in the intuitive understanding of the monitored situation. Our framework can provide an estimation of the total epistemic and aleatory uncertainty, U_{wav} and U_{src}, to generate a matrix representation of each uncertainty source as part of the monitoring outcome. We call these representations *monitoring uncertainty maps*.

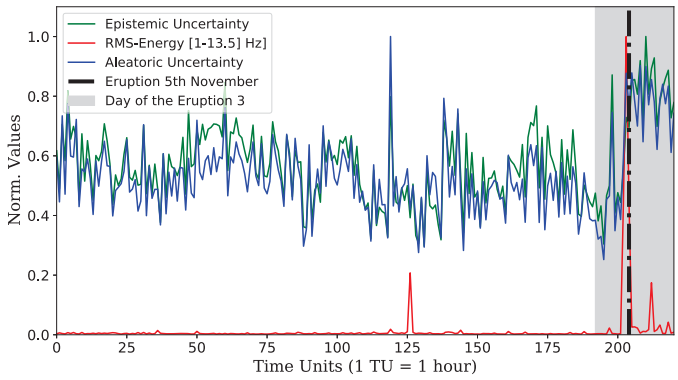
Figure 5 displays four seismo-volcanic events, along with epistemic U_{wav} (middle) and aleatoric U_{src} (bottom) *monitoring uncertainty maps*. First, note that both uncertainties, U_{wav} and U_{src}, are consistently emitted through time. The most significant uncertainty phases are associated with low detected energies, which is expected given the very narrow spectral band carrying the information about the continuous volcanic tremor. For this volcano, U_{wav} and U_{src} present as a coupled, synchronous mechanism, for which the input waveform U_{wav} displays notable uncertainty levels at the abrupt transition from background activity to a seismic event. The uncertainty U_{src} is linked to potentially co-existent sources in the seismic wavefield that the model has learned to detect. Figure 5 (a) shows seismic uncertainty maps at two LF earthquake boundaries. At the onset of the LF earthquakes, epistemic uncertainty is emitted in the SBT and LF classes, and the switches immediately to the LF class until the event offset. In parallel, the aleatoric uncertainty tracks the SBT and LF classes simultaneously. Figure 5 (b) depicts a cigar-shaped rockfall (DP class) and an HF earthquake with similar behaviors in the U_{wav} and U_{src} uncertainty maps.



(a)



(b)



(c)

Fig. 6. Hourly normalized plots of the overall epistemic (green), aleatoric (blue), and RMSE energy (red) for the pre-eruptive and eruptive periods of E1 (a), E2 (b), and E3 (c). There is a direct correspondence across energy and uncertainties, aligned as a coupled mechanism; a rise in both uncertainties is simultaneous to rising seismic energy. Drift is noticeable either suddenly (c) or gradually (a) (b) for each eruption, which demonstrates drift traceability through time.

Figure 5 (c) and (d) depicts the uncertainty maps of co-existent events with copious seismic noise background. In Figure 5 (c), the three first HF earthquakes exhibit a low-frequency component in uncertainty U_{src}; the uncertainties can identify simultaneous frequency components of events, even if the system is trained to recognize single categories, or classes, from the seismic data stream. Similarly, the DP class shows this behavior in the low-frequency SBT. The uncertainties follow the same pattern in Figure 5 (d); that is, the aleatoric uncertainty tracks the high spectral component of the HF earthquakes whilst assigning high uncertainty to the LF class. These maps contrast with Figure 5 (a) and (b), where the amount of background noise is much lower, but the uncertainty exhibits a similar behavior on frequency components, high or low, that are concurrent with the event.

These *monitoring uncertainty maps* serve as interpretable, multi-source activation maps that highlight potential co-existent events at any given time. The maps reflect the intrinsic non-linearity of seismic energy release, providing refined visual information about the behavior of U_{wav} and U_{src} in a seismo-volcanic wavefield. These *seismic monitoring uncertainty maps* complement the segmentation mask M_t and classification outcome, providing a new source of detailed information for seismo-volcanic monitoring.

E. Monitoring seismic wavefields

In previous sections, we have demonstrated how uncertainty can detect drift for each period independently. We also investigated if the estimated uncertainty is correlated to other geophysical parameters, and it evolves through time. Figure 6 depicts the normalized temporal variation of U_{ep} and U_{al}, along with the RMSE (energy). The x-axis of Figure 6 represents time-units (TU); that is, time scales over which the uncertainty has been estimated. For all three main eruptions, the estimated short-term evolution of the uncertainty exhibits similar temporal behavior to that of the RMSE seismic energy. The sustained energy exchange from the volcano with the medium depends on the waveform propagation and signal frequency, and is thus an independent parameter of the estimated uncertainty. In a non-stationary environment, changes in the medium or the seismic sources are associated with variations in seismograms. These waveform distortions translate into an alteration of the monitored variable and thus into higher uncertainties 6. These graphs indicate *when* data drift is happening. The accurate identification of such drifts is essential to improve monitoring adaptability by indicating the requirement to re-train the algorithm with data from the new period. These empirical uncertainty observations are in good agreement with

previous scientific reports in seismological bulletins [19]. E1, a sustained but small eruption, was preceded by minor increases in seismicity. Note that this sustained seismicity is reflected in the gradual data drift. The principal, classic eruption (E2), which had the highest peak for U_t, corresponded to the maximum release of seismic energy during the 2007 eruptive cycle, with high rates of seismic events and a massive, longer explosion. Previous research reported steady increments in the background tremor energy [34]. The seismic activity leading to eruption E3 was shallow, as reflected by the almost constant temporal evolution of the uncertainty. Eruption E3 had a relatively small energy release, and the seismological observatory could not visually verify the event owing to poor atmospheric conditions; the only observations were via thermal camera anomaly detection over the dome and the RMSE measures [19] [35]. Therefore, the continuous trend in the uncertainty rises abruptly, simultaneously with the RMSE, indicating an explosion in the volcano that suddenly changed the monitored conditions, which is consistent with past seismological bulletins.

VII. Conclusions

We applied Bayesian theory to continuous seismo-volcanic recognition to create new connections between non-stationary environments and monitoring uncertainties. The proposed model departs from established approaches based on supervised learning with modest datasets, which are not capable of detecting data drifts as part of the monitored outcome. We performed simulations on three eruptive time-periods to gain insight into the generalization and performance of the model.

The designed convolutional network can learn and detect the full intra-frequency dynamic range of seismo-volcanic events whilst performing seismic event recognition. Moreover, the designed framework is a probabilistic surrogate for estimating the total uncertainty U_t as the summation of the seismic wavefield variations (reducible with more data) and the randomness of the monitored volcanic process. The severity of the drift can be evaluated from the computed metrics, yet this assessment is incomplete without knowing what is driving the change. This formulation has allowed us to introduce *monitoring uncertainty maps* as a supportive tool for illustrating the presence or absence of simultaneous sources. The epistemic uncertainty acts as a complementary onset detector, whereas the aleatory uncertainty proxy multi-source identification greatly enhancing monitoring outcomes. Finally, the principled uncertainty estimates are useful for categorizing changes in seismic signals. The estimated short-term evolution of the uncertainty

is consistent with the real-time seismic energy measurement, a direct waveform parameter independent from the estimated uncertainty.

ACKNOWLEDGMENTS

The authors are grateful to Alejandro Díaz-Moreno for continuous support and advice on Bezymianny data, along with all of the people who contributed to seismic data analysis reviews with the PICOSS software. The facilities of IRIS Data Services, and specifically the IRIS Data Management Center, were used for access to the waveforms, related metadata, and/or derived products used in this study. IRIS Data Services were funded through a Seismological Facilities for the Advancement of Geoscience (SAGE) Award of the National Science Foundation under Cooperative Support Agreement EAR-1851048.

REFERENCES

[1] S. Spampinato, H. Langer, A. Messia, and S. Falsaperla. “Short-term detection of volcanic unrest at Mt. Etna by means of a multi-station warning system”. In: *Scientific Reports* 9.1 (2019), p. 6506.

[2] DE. Dempsey, Cronin SJ, S. Mei, and W. Kempa-Liehr. “Automatic precursor recognition and real-time forecasting of sudden explosive volcanic eruptions at Whakaari, New Zealand”. In: *Nature communications* 11.1 (2020), pp. 1–8.

[3] Q. Kong, D. Trugman, Z.E Ross, M. Bianco, b.j. Meade, and P. Gerstoft. “Machine learning in seismology: Turning data into insights”. In: *Seism.Res.Lett.* 90.1 (2018), pp. 3–14.

[4] K.J. Bergen, P.A. Johnson, M.V. de Hoop, and G. Beroza. “Machine learning for data-driven discovery in solid Earth geoscience”. In: *Science* 363.6433 (2019).

[5] G. Cortés, M. C. Benítez, L. García, I. Álvarez, and J. M. Ibanez. “A Comparative Study of Dimensionality Reduction Algorithms Applied to Volcano-Seismic Signals”. In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 9.1 (2016), pp. 253–263.

[6] M. S. Khan, M. Curilem, F. Huenupan, M. F. Khan, and N. Becerra Yoma. “A Signal Processing Perspective of Monitoring Active Volcanoes [Applications Corner]”. In: *IEEE Signal Processing Magazine* 36.6 (2019), pp. 125–163.

[7] M. Malfante, M. Dalla Mura, J. Metaxian, J. I. Mars, O. Macedo, and A. Inza. “Machine Learning for Volcano-Seismic Signals: Challenges and Perspectives”. In: *IEEE Signal Processing Magazine* 35.2 (2018), pp. 20–30.

[8] S Mostafa Mousavi, Weiqiang Zhu, Yixiao Sheng, and Gregory C Beroza. “CRED: A deep residual network of convolutional and recurrent units for earthquake signal detection”. In: *Scientific reports* 9.1 (2019), pp. 1–14.

[9] S Mostafa Mousavi, Weiqiang Zhu, William Ellsworth, and Gregory Beroza. “Unsupervised clustering of seismic signals using deep convolutional autoencoders”. In: *IEEE Geoscience and Remote Sensing Letters* 16.11 (2019), pp. 1693–1697.

[10] M. Titos, A. Bueno, L. García, and C. Benítez. “A Deep Neural Networks Approach to Automatic Recognition Systems for Volcano-Seismic Events”. In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 11.5 (2018), pp. 1533–1544.

[11] M. Titos, A. Bueno, L. García, M. C. Benítez, and J. Ibañez. “Detection and Classification of Continuous Volcano-Seismic Signals With Recurrent Neural Networks”. In: *IEEE Transactions on Geoscience and Remote Sensing* 57.4 (2019), pp. 1936–1948.

[12] J.M. Ibáñez, E. del Pezzo, J. Almendros, M. La Rocca, G. Alguacil, R. Ortiz, and A. García. "Seismovolcanic signals at Deception Island volcano, Antarctica: Wave field analysis and source modeling". In: *J. Vol. Geoth. Res.* 105.B6 (2000), pp. 13905–13931.

[13] G. Cortés, L. García, I. Álvarez, Ca. Benítez, Á. de la Torre, and J.M.Ibáñez. "Parallel System Architecture (PSA): An efficient approach for automatic recognition of volcano-seismic events". In: *Journal of Volcanology and Geothermal Research* 271 (2014), pp. 1 –10.

[14] M. C. Benítez, J. Ramírez, J. C. Segura, J . M. Ibáñez, J. Almendros, A. García-Yeguas, and G. Cortés. "Continuous HMM-Based Seismic-Event Classification at Deception Island, Antarctica". In: *IEEE Transactions on Geoscience and Remote Sensing* 45.1 (2007), pp. 138–146.

[15] Gilberto Saccorotti and Ivan Lokmer. "A review of seismic methods for monitoring and understanding active volcanoes". In: *Forecasting and Planning for Volcanic Hazards, Risks, and Disasters* (2020), pp. 25–73.

[16] RSJ. Sparks. "Forecasting volcanic eruptions". In: *Earth and Planetary Science Letters* 210.1-2 (2003), pp. 1–15.

[17] S Mostafa Mousavi and Gregory C Beroza. "Bayesian-Deep-Learning Estimation of Earthquake Location From Single-Station Observations". In: *IEEE Transactions on Geoscience and Remote Sensing* (2020).

[18] A. Bueno, C. Benítez, S. De Angelis, A. Díaz Moreno, and J. M. Ibáñez. "Volcano-Seismic Transfer Learning and Uncertainty Quantification With Bayesian Neural Networks". In: *IEEE Transactions on Geoscience and Remote Sensing* 58.2 (2020), pp. 892–902.

[19] W. Thelen, M. West, and S. Senyukov. "Seismic characterization of the fall 2007 eruptive sequence at Bezymianny Volcano, Russia". In: *J. Vol. Geoth. Res.* 194.4 (2010), pp. 201 –213.

[20] S. Mcnutt and Takeshi Nishimura. "Volcanic tremor during eruptions: Temporal characteristics, scaling and constraints on conduit size and processes". In: *Journal of Volcanology and Geothermal Research* 178 (Nov. 2008), pp. 10–18. DOI: 10.1016/j.jvolgeores.2008.03.010.

[21] H. Langer, S. Falsaperla, A. Messina, S. Spampinato, and B. Behncke. "Detecting imminent eruptive activity at Mt Etna, Italy, in 2007–2008 through pattern classification of volcanic tremor data". In: *Journal of Volcanology and Geothermal Research* 200.1-2 (2011), pp. 1–17.

[22] Konstantinos I Konstantinou and Vera Schlindwein. "Nature, wavefield properties and source mechanism of volcanic tremor: a review". In: *Journal of Volcanology and Geothermal Research* 119.1-4 (2003), pp. 161–187.

[23] A. Jansson, E. Humphrey, N. Montecchio, R. Bittner, A. Kumar, and T. Weyde. "Singing voice separation with deep u-net convolutional networks". In: *18th International Society for Music Information Retrieval Conference (ISMIR 2017)*. 2017, pp. 23–27.

[24] N. Takahashi, N. Goswami, and Y. Mitsufuji. "Mmdenselstm: An efficient combination of convolutional and recurrent neural networks for audio source separation". In: *2018 16th International Workshop on Acoustic Signal Enhancement (IWAENC)*. IEEE. 2018, pp. 106–110.

[25] W. Zhu, S. M. Mousavi, and G. C. Beroza. "Seismic Signal Denoising and Decomposition Using Deep Neural Networks". In: *IEEE Transactions on Geoscience and Remote Sensing* 57.11 (2019), pp. 9476–9488. DOI: 10.1109/TGRS.2019.2926772.

[26] O. Ronneberger, P. Fischer, and T. Brox. "U-net: Convolutional networks for biomedical image segmentation". In: *International Conference on Medical image computing and computer-assisted intervention*. Springer. 2015, pp. 234–241.

[27] Yuxuan Wang, Arun Narayanan, and DeLiang Wang. "On training targets for supervised speech separation". In: *IEEE/ACM transactions on audio, speech, and language processing* 22.12 (2014), pp. 1849–1858.

[28] S. Bai, J.Z. Kolter, and V. Koltun. "An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling". In: *CoRR* abs/1803.01271 (2018). eprint: 1803.01271. URL: http://arxiv.org/abs/1803.01271.

[29] A. Pandey and D. Wang. “TCNN: Temporal convolutional neural network for real-time speech enhancement in the time domain”. In: *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE. 2019, pp. 6875–6879.

[30] Y. Gal. “Uncertainty in deep learning”. PhD thesis. PhD thesis, University of Cambridge, 2016.

[31] Armen Der Kiureghian and Ove Ditlevsen. “Aleatory or epistemic? Does it matter?” In: *Structural safety* 31.2 (2009), pp. 105–112.

[32] Y. Kwon, JH. Won, BJ. Kim, and MC. Paik. “Uncertainty quantification using bayesian neural networks in classification: Application to ischemic stroke lesion segmentation”. In: *2018 Medical Imaging with Deep Learning (MIDL)*. 2018.

[33] L. Sandri, P. Tierz, A. Costa, and W. Marzocchi. “Probabilistic hazard from pyroclastic density currents in the Neapolitan area (Southern Italy)”. In: *J. Vol. Geoth. Res* 123.5 (2018), pp. 3474–3500.

[34] M. E. West. “Recent eruptions at Bezymianny volcano: A seismological comparison”. In: *J. Vol. Geoth. Res.* 263 (2013), pp. 42 –57.

[35] Olga A Girina. “Chronology of Bezymianny volcano activity, 1956–2010”. In: *Journal of Volcanology and Geothermal Research* 263 (2013), pp. 22–41.

[36] S.R. McNutt, G. Thompson, J. Johnson, S. De Angelis, and D. Fee. “Seismic and infrasonic monitoring”. In: (2015), pp. 1071–1099.

[37] C. J. Bean, L. De Barros, I. Lokmer, J. Métaxian, G. O’Brien, and S. Murphy. “Long-period seismicity in the shallow volcanic edifice formed from slow-rupture earthquakes”. In: *Nature Geoscience* 7 (2013).

[38] A. Bueno, A. Díaz-Moreno, S. De-Angelis, C. Benítez, and J. M. Ibáñez. “Recursive Entropy Method of Segmentation”. In: *Seism. Res. Lett.* 90.4 (2019), pp. 1670–1677.

[39] A. Bueno, L. Zuccarello, A. Díaz-Moreno, J. Woollam, M. Titos, C. Benítez, I. Álvarez, J. Prudencio, and S. De Angelis. “PICOSS: Python Interface for the Classification of Seismic Signals”. In: *Computers and Geosciences* 142 (2020), p. 104531.

[40] K.P. Diederik and J. Ba. “Adam: A Method for Stochastic Optimization”. In: *CoRR* abs/1412.6980 (2014).

[41] H. Larochelle, D. Erhan, A. Courville, J. Bergstra, and Y. Bengio. “An Empirical Evaluation of Deep Architectures on Problems with Many Factors of Variation”. In: *Proceedings of the 24th International Conference on Machine Learning*. ICML 07. 2007, pp. 473–480.

[42] M. Sokolova and G. Lapalme. “A systematic analysis of performance measures for classification tasks”. In: *Information Processing & Management* 45.4 (2009), pp. 427–437.

6 | Recurrent Scattering Network detects metastable behavior in polyphonic seismo-volcanic signals for volcano eruption forecasting

This chapter presents the recurrent scattering deep neural network to detect seismic variations and demonstrate the significance of our theoretical framework as an indicator for possible forecasting of eruptions. This article is in press at IEEE Transactions on Geoscience and Remote Sensing, with the following journal metrics: (IF), JCR 2020: (5.855). Remote Sensing (Rank 5/32) (Q1). Geochemistry and geophysics (Rank 5/85) (Q1). Imaging Science and photographic technology (Rank 5/27) (Q1). Engineering, electrical, and electronic science (Rank 27/266) (Q1). We reproduce the draft in press by the journal, which can be cited as:

1. **A. Bueno**, R. Balestriero, S. De Angelis, C. Benitez, L. Zuccarello, R. Baraniuk, J.M. Ibánez and M. V. de Hoop. Recurrent Scattering Network detects metastable behavior in polyphonic seismo-volcanic signals for volcano eruption forecasting. IEEE Transactions on Geoscience and Remote Sensing (in press).

Recurrent Scattering Network detects metastable behavior in polyphonic seismo-volcanic signals for volcano eruption forecasting

Angel Bueno, Randall Balestriero, Silvio De Angelis, Carmen Benítez, Luciano Zuccarello, Richard Baraniuk, Jesús M. Ibáñez and Maarten V. de Hoop

Abstract

We introduce an End-to-End (E2E) deep neural network architecture designed to perform seismo-volcanic monitoring focused on detecting change. Due to the complexity of volcanic processes, this requires a polyphonic detection, segmentation and classification approach. Through evolving epistemic uncertainty, invoking a Bayesian network strategy, we detect change and demonstrate its significance as an indicator for possible forecasting of eruptions using data from the Bezymianny and Etna volcanoes. Specifically, we propose morphing the scattering transform from previous work into a novel E2E hybrid and recurrent learnable deep scattering network to adapt to the multi-scale temporal dependencies from streaming data. The time-dependent scattering is in some sense physics informed, namely through time-frequency representation (TFR) of the data. At the same time, with a carefully designed deep

A. Bueno and C. Benítez are with the Department of Signal Theory, Telematic and Communications, University of Granada, 18071, Spain. (e-mail: angelbueno@ugr.es).

R. Balestriero and R. Baraniuk are with the Department of Electrical and Computer Engineering, Rice University, TX 77005, USA.

Maarten V. de Hoop is with the Department of Computational and Applied Mathematics, Rice University, TX 77005, USA.

L. Zuccarello is with the Istituto Nazionale di Geofisica e Vulcanologia (INGV), Sezzione di Pisa, Pisa, Italy.

S. De Angelis is with the Department of Earth, Ocean and Ecological Sciences, University of Liverpool, L69 3GP, UK.

J. M. Ibáñez is with the Andalusian Institute of Geophysics and Department of Theoretical Physics and Cosmos, University of Granada, Spain

This work is supported by TEC2015-68752 (KNOWAVES), PID2019-106260GB-I00 (FEMALE), Department of Energy grant DE-SC0020345, Simons Foundation MATH + X program and the Geo-Mathematical Imaging Group at Rice University.

convolutional-LSTM architecture, we learn intra-event, temporal dynamics from the scattering coefficients or features. We verify the effectiveness of transfer learning switching between volcanoes. Our experimental results set a new norm for semi-supervised seismo-volcanic monitoring.

Index Terms

Volcanoes, Seismology, Uncertainty, Recurrent Neural Networks, Wavelet transforms.

I. Introduction

We introduce an End-to-End deep neural network architecture designed to perform polyphonic seismo-volcanic monitoring on the one hand and detecting change on the other hand. The search for observable eruption precursory signals and their evolution has remained one of the challenges of modern volcanology and is critical for effective monitoring and the forecasting of eruptions. We address this search with a comprehensive deep learning framework.

A volcanic eruption, the emission of magma and gases as well as the exchange of various forms of energy, is the final consequence of a series of energetic physical and chemical processes in Earth's interior. Volcanic eruptions range from fluid lava flows to explosive emissions that inject large volumes of material into the atmosphere. Repose times also vary widely, even for individual volcanoes, extending from minutes to years, suggesting that eruptions are associated with quasi-stable processes [1], [2], complicating forecasting. From a seismic perspective, there are currently two main approaches to forecasting eruptions. The first one is based on the study of changes in the tensional state of the volcanic system when a volcano evolves towards an eruption, as recorded by seismic velocity [3]. This approach assumes that pre-eruptive changes to the edifice shape (volume) and tensional state affect the wave propagation velocity which can be detected with seismic noise recorded by stations on the volcano. Indeed, important progress in the analysis of seismic background noise at volcanoes has been made [4], [5]. The second approach is to detect and identify seismo-volcanic signals associated with eruptive dynamics (i.e., physical processes within the volcano) and to use these as precursors.

Here, we propose a third approach in which deep learning is used to relate changes in the seismic wave field to changes in the state of the volcano (presumably directly related to the underlying physics), including volcanic tremor. The results of this study show that the wave field properties of tremor can reflect the evolution of volcano dynamics and can be used as precursor of volcanic eruptions.

Seismology has played a leading role in monitoring volcanoes and identifying precursors to eruptions [6], [7], [8]. The sources of seismic energy are diverse and include rock rupture through an accumulation of elastic energy, ground resonance phenomena, pressure changes due to the movement of fluids, conduit resonance, and many more. As a consequence, seismic signals exhibit variable durations and spectral contents [9], [10], [11], [12], [13], [14]. One challenge is to identify classes of events causally related to distinct sources, providing an association that can be used as a precursor to what is happening inside a volcano. Identification is currently carried out using standard, supervised machine learning techniques [14], [15], [16], [17], [18], [19]. However, identifying signals does not provide information about the dynamics of the processes that lead to eruptions.

Our approach departs from those using handcrafted features to detect change [20], [21], [22], [23]; we suggest that subtle changes not currently identified in standard data analysis are important. Our deep neural network can identify such changes by dynamically learning a scattering representation of streaming data. The introduction of epistemic uncertainty revealed by Monte Carlo dropout [24] is pivotal in our approach, identifying a drift in uncertainty in the classification of events presumably related to the relevant processes.

State-of-the-art procedures for the detection, segmentation, and classification of signals or events in streaming data are often implemented as separate workflows by combining signal processing (to provide a priori representations) and traditional deep learning strategies (to probe these representations) [9], [16], [17], [18], [25], [26]. The End-to-End approach proposed here addresses the shortcomings of past methods.

New approach: The main contributions of this paper can be summarized as follows:

1) We propose a scattering network that cascades learnable wavelet transforms and complex moduli, in its original form given by [27], to generate features with a learnable spline approximation in both central frequencies and wavelet shape for each layer. The central frequency of the mother wavelet in a given task can be calibrated by letting the dilation factor of the filters, along with the spline knots, be learnable parameters. Knots and filter learnability leads to a mother wavelet capable of shape morphing to better capture signal onsets, even in very challenging environments.
2) We introduce a novel integration of this learnable scattering network in a recurrent architecture. The learnable scattering network component produces a set of multiple-order representations (so-called scatter-grams) that contain a full range of wavelet scales. This

range is converted to a sequence of temporal, structured representations of local scattering variations, with the scatter-grams as complementary channels, in which two stacked convolutional long-short term memories (conv-LSTMs) can extract intra-frequency temporal variation across multiple scales [28], [29]. In parallel, a learnable scattering denoising operation is performed via convolutional skip connections, suppressing background noise and enhancing the scatter-grams from the data stream. The outputs of both components, conv-LSTMs and skip connections, are fused, captured by bidirectional LSTMs [30], and forwarded to a dense layer to output a probabilistic event detection and classification matrix highlighting coexistent seismic signals as *polyphonic* events.

3) We use the uncertainty to study the goodness of a classifying process of seismo-volcanic signals for the purpose of eruption early warning protocols that are exportable across volcanic systems. We define uncertainty as a measure of how little or how much one set of data resembles another and detect the variation or evolution of a physical system. Therefore, the concept of uncertainty can be used to control the quality of the physical measurements of a volcanic system and as an indicator of the evolution of these measurements over time; as such, this approach is suitable for forecasting volcanic eruptions. We employ epistemic uncertainty to detect change and reveal that the power-law drifts towards eruptions. The evolution of the seismic data in a volcano can be potentially associated with the appearance of new seismic signals or a change in the characteristics of those seismic signals. Therefore, an increase (or decrease) in the uncertainty can be used as a good indicator for early warning of volcanic eruptions and exported from one volcano to another.
4) We demonstrate through transfer learning that our implemented architecture can be exported across a range of different volcanoes and eruptive styles. With minimal hassle, we are able to reuse our system from one volcano to another one, even if there exists new or unknown signals in the target volcano. The use of uncertainty in a transfer learning approach permits establishing universal early warning and predictive protocols that are the same regardless of the volcano, thus giving an objective measure that can be communicated during volcanic crises.

We demonstrate the potential of our architecture using data from three well-studied 2007 eruptions of Bezymianny volcano, over a period of approximately 3 months. The three eruptions—those on 25 September, 14 October, and 5 November—were brief but very energetic, and

included various pre-eruptive seismicity rates and eruption mechanisms. Moreover, application to data from Mt. Etna and Mt. St. Helens confirms that our approach can be applied to different volcanic systems. To the best of our knowledge, this is the first deep learning approach of its kind; that is, it is the first to address polyphonic detection, segmentation, and implicit classification of physical processes in an eruptive sequence while detecting change through drift in epistemic uncertainty.

II. Prior work

Machine learning in seismology. Advanced machine learning techniques provide tools beyond human intuition to discover as yet unused signals and patterns, and have been applied to data analysis in the field of seismology [31], [32], [33], [34]. In volcano-seismology, the application of machine-learning methods has focused mostly on automated detection and classification of seismic signals through handcrafted signal properties [16], embedding vectors [35], Hidden Markov Models [26], and standard deep learning methodologies [17]. These works follow archetypal machine-learning pipelines: A set of features related to continuous measurements made on volcanic observations are selected to fine tune monitoring procedures. However, catalogs remain incomplete leading to the underestimation of hazard assessment. This is partly due to the lack of uniformity in labeling events, and the occurrence of multiple sources at the same time in low SNR environments. Moreover, these methods have two main limitations when dealing with seismic data. First, they can only handle non-overlapping, monophonic seismic signals. Second, they cannot be applied to new seismic wave fields without re-training the entire network.

Study of volcano dynamics. Past attempts at the detection of volcanic precursors have concentrated on ideas adapted from engineering applications: the Failure Forecast Method (FFM) carries out a regressive estimation of the time to target (i.e., time to eruption or structural failure) based on handcrafted features from accelerated strain. Recent machine learning methods discern eruptive behavior from a set of handcrafted features. A filter-bank analysis of 6 years of data from Piton de la Fournaise volcano (Reunion, France) was performed to derive 990 features in the spectral band of 0.5–26 Hz, and forwarded to a decision-tree gradient boosting algorithm named XGBoost [36]. The empirical results showed feature evolution through time, with tracking capabilities of the variation in a narrow frequency range (3–5 Hz). A posterior analysis with spectral clustering provided a physical interpretation of the volcano dynamics, highlighting the dominance of tremor and high-frequency components during the eruptive behavior. An analysis

similar to that of [36] but with different classifiers was carried out for Telica (Nicaragua) and Nevado del Ruiz (Colombia), linking eruptive and non-eruptive behavior to the temporal variations of features [37].

Despite the promising performance of integrated machine learning and handcrafted features, two issues arise for the real-time deployment of such systems in an observatory. The tabular set of features reflects a predetermined range based on analysts' experience; thus, unforeseen situations can saturate the classifier while underestimating volcanic unrest. Second, experimental results show that not all of these features contribute to the final prediction. Instead, current, traditional machine learning methods exhibit a predictive bias towards a specific frequency band that is considered essential to categorize all types of seismic transients (from pressure pulses in fluids to the fracturing of rocks) in a volcano. Finally, the consideration of parameterized windows over the seismic data stream can effectively trace its temporal evolution but do not identify which families of events contribute to those changes. The observable variables are often intertwined with a myriad of source-dependent seismic events, such as volcano-tectonic earthquakes or tremor. Given the different associations of these signals to physical mechanisms, distilling such knowledge is essential to understand the eruptive dynamics of a volcano.

Learnable filter banks. With the availability of large scale audio classification datasets, multiple methods have been developed to bring learnability to filter banks [38], [39], [40], [41], [42]. Those methods can be divided into two main categories. The first category [43] performs filter-bank learning by independently adapting the center frequencies and bandwidths of a collection of Morlet wavelets. Another family of methods [44] relies on learning the start and cutoff frequencies of a band-pass sinc filter apodized with a Hamming window. The second category of methods are based on the STFT. For example, [39] propose to learn Mel filters that are applied to a spectrogram (modulus of STFT). Those filters linearly combine adjacent filters in frequency which can be interpreted as learning a linear frequency sub-sampling of the spectrogram. Learning the apodization window used to produce the spectrogram has also been developed [41], [45]. Alternatively, one may select a transform a priori that can be adapted with learnable hyperparameters [46]. In monitoring applications where signals overlap in time, the intra-class separability can be extremely difficult, and using the correct transform can greatly impact predictive performance.

Convolution LSTM, 2D. Recurrent Neural Networks (RNNs) are highly successful in temporal modeling of time sequences. LSTMs and GRUs have been proposed as specialized archi-

tectures to refine temporal modeling. If applied over the scattering features, one of the LSTMs' limitations is that intra-frequency scatter-grams variations information is missing. In this regard, Convolutional-LSTM (ConvLSTM) is proposed as a mathematical modification able to model scales variations through time by explicitly encoding this information into state tensors. The first application of ConvLSTM was for weather forecasting [29]. The network is trained on a forecasting problem with temporal sequences of radar echo maps. ConvLSTMs have been applied in speech recognition [28], gesture detection [47] and, as mentioned, weather forecasting [48], among many other disciplines. Modifications over the original structure have been developed to increase robustness in specific settings [49]. The applicability of ConvLSTMs in seismology remains to be explored. In our proposed framework, we modify the working regime of the ConvLSTM architecture to cope with a multi-scale time-frequency representation, adjusting the network to apply to waveform dynamics in a multi-source or polyphonic setting.

III. Problem description

Seismo-volcanic monitoring can be approached from a sequence prediction perspective. If multiple and simultaneous sources generate waveforms or signals, a polyphonic approach is needed to detect the presence and class-membership of overlapping events. We define our seismic data domain as all of the recorded waveforms and known seismo-volcanic event types for any given monitoring period, T. Within this defined data domain, our training dataset is $D = \{(X, Y)\}$, with $X = (x_0, x_1, \ldots, x_n)$, where $X = (x_0, x_1, \ldots, x_n)$ represents raw, sampled data streams of arbitrary time duration, and $Y = (y_0, y_1, ..., y_n)$ represents the labeled sequence, with $y_i, i = (0, \cdots, n)$ containing one or multiple categorical labels over K classes. From a machine learning perspective, finding the relationships between pairs of multi-source, framewise X and Y can be cast as polyphonic event detection and classification. To achieve this, the Y labels must be converted into a binary temporal matrix whose columns represent time and whose rows the frequencies are present in the trace. This type of formulation allows the ubiquitous identification of seismic events whose sources may vary over time or for which enough data samples are not available for training purposes.

To quantify the performance of our deep neural network, we adopt the polyphonic audio metrics [50], that is, precision (PR), recall (RC), and the F-measure (F1), according to the frame error rate at one second resolution in our monitoring setting. The PR quantifies the rate of

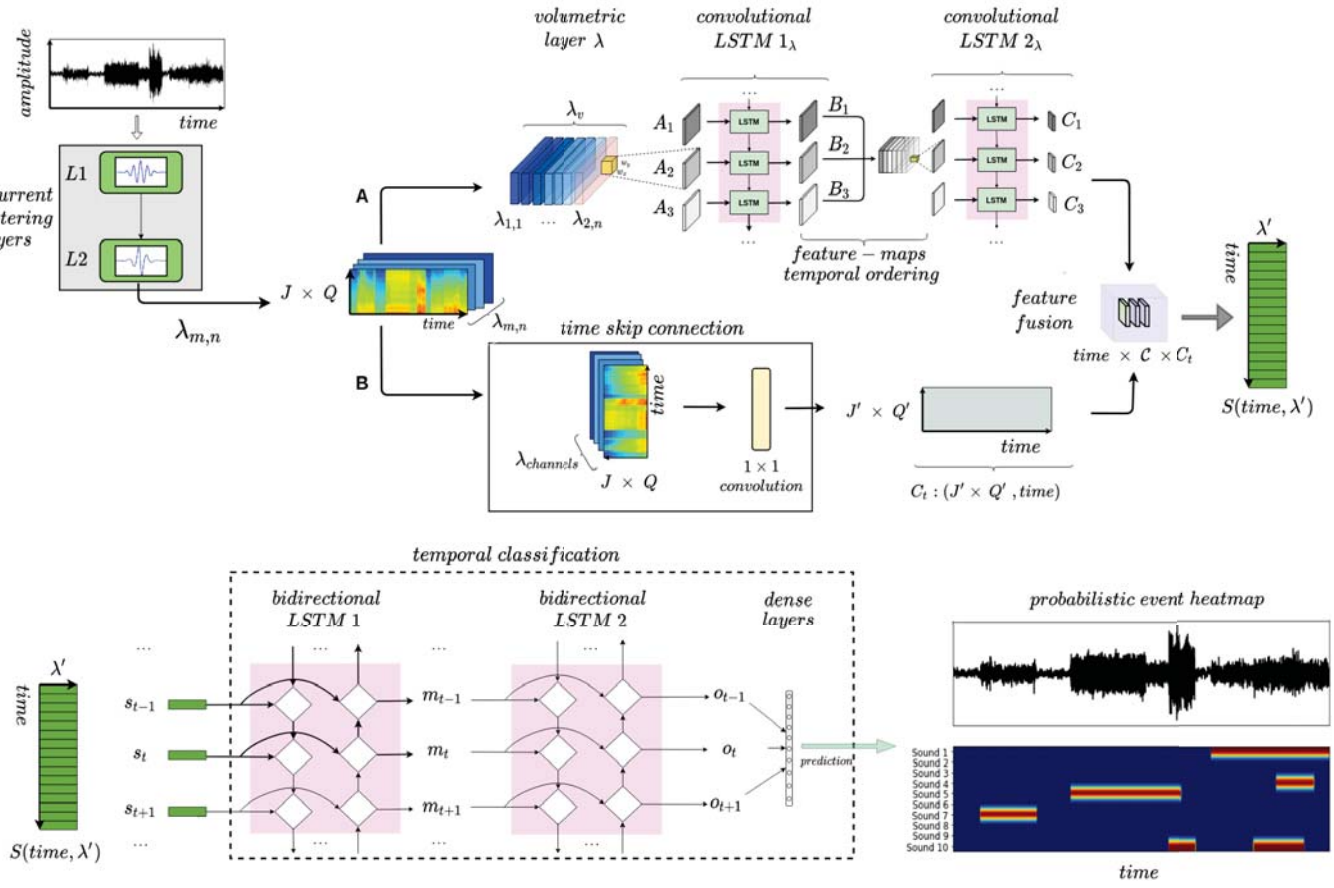


Fig. 1. **Network architecture for polyphonic sound event detection**: the recurrent learnable scattering network permits flexible temporal feature representation, while the joint modeling yields increased adaptability of the mother wavelet to the target sounds. Path A performs dynamic multi-scale convolutions, whereas path B performs dimensionality reduction and scattering compression yielding sensitivity to the onsets of events. After fusion, features are captured by a temporal classification module, with two bidirectional long-short term memory (LSTM) and an Fully Connected (FC) layer that outputs the probabilistic event detection matrix with per-class onsets.

positively classified frames, whereas RC measures the sensitivity of the system to recognize correct frames. The F1 is a weighted average between the PR and RC.

IV. Representation Learning

A. Scattering Network

The deep scattering network or DSN [27], [51], [52], extracts locally invariant robust representations from a raw signal, by systematically applying a cascade of wavelet transforms and modulus operators. Consider the input signal of length N as $x \in \mathbb{R}^N$ where each time sample

is accessed via $x(t), t = 1, \ldots, N$. Each DSN layer indexed by ℓ is parameterized by $J^{(\ell)}$ and $Q^{(\ell)}$, the number of octaves and the number of wavelets per octave, respectively. In fact, the wavelet filter bank is derived as $\{\phi_\lambda, \lambda \in \{2^{i/Q}, i = 1, \ldots, JQ\}\}$ with $\phi_\lambda = \lambda^{-1/2}\phi(\lambda^{-1}t)$ where ϕ is a mother wavelet chosen a priori. This structure is similar to a convolutional neural network architecture. At layer ℓ, the wavelet filter-bank produces the representation $U_\lambda^{(\ell)}$ which corresponds to the application of each filter followed by application of taking the modulus as

$$U^{(\ell)}_{\lambda^{(\ell)},\lambda^{(\ell-1)},\ldots,\lambda^{(1)}}(t) = |(\phi_\lambda \star U^{(\ell-1)}_{\lambda^{(\ell-1)},\ldots,\lambda^{(1)}})(t)|, \tag{1}$$

where $|.|$ denotes the (complex) modulus, and the initialization at the first layer is defined as $U^{(0)} = x$. Note that the wavelet filter bank can vary depending on the layer ℓ. Once the representations are obtained for all the layers, the scattering features or coefficients, which form the representation, are obtained by performing a time averaging. In the case of global average pooling this corresponds to

$$S^{(\ell)}_{\lambda^{(\ell)},\ldots,\lambda^{(1)}} = \frac{1}{N}\sum_{t=1}^{N} U^{(\ell)}_{\lambda^{(\ell)},\ldots,\lambda^{(1)}}(t), \tag{2}$$

with $S^{(\ell)}_{\lambda^{(\ell)},\ldots,\lambda^{(1)}}$ the scattering representation at layer ℓ. The time averaging can be made local based on the desired time invariance, thus defining the time resolution for the successive layers.

B. Learnable Scattering Transform

The scattering network introduced above consists of a succession of wavelet transforms and complex moduli, in a neural-network fashion with per-layer wavelet filter banks. Hermite Cubic Splines provide a learnable parametrization of the mother wavelet from which the wavelet filter bank per layer is derived. Formally, we introduce a partition of a compact support of the mother wavelet into R intervals using $K = R + 1$ knots, $t_r, r = 1, \ldots, K$, so that

$$\begin{aligned} f_{\Theta,\Gamma}(t) = \sum_{r=1}^{K-1} \Bigg(& \gamma_r + u_0\left(\frac{t-t_r}{t_{r+1}-t_r}\right) + \gamma_{r+1} + \\ & + u_1\left(\frac{t-t_r}{t_{r+1}-t_r}\right) + \theta_r + \\ & + v_0\left(\frac{t-t_r}{t_{r+1}-t_r}\right) + \theta_{r+1} + \\ & + v_1\left(\frac{t-t_r}{t_{r+1}-t_r}\right) 1_{\{t\in[t_r,t_{r+1}].\}} \end{aligned} \tag{3}$$

where $\Theta = \{\theta_1, \ldots, \theta_K\}$, $\Gamma = \{\gamma_1, \ldots, \gamma_K\}$ and the basis functions are given by

$$u_0(t) = 2t^3 - 3t^2 + 1$$
$$u_1(t) = -2t^3 + 3t^2$$
$$v_0(t) = t^3 - 2t^2 + t$$
$$v_1(t) = t^3 - 2t^2$$

subject to the following constraints,

$$\gamma_1 = \theta_1 = \theta_K = \gamma_K = 0, \qquad \text{(compact support)}$$

$$\gamma_2 = -\sum_{r \neq 1} \gamma_r, \qquad \text{(zero mean)}$$

$$\max_r |\gamma_r| < \infty, \max_r |\theta_r| < \infty \qquad \text{(boundedness).}$$

The parameters γ and θ allows one to learn the shape while keeping a uniform partition for the spline with knots, t_r, equally spaced in time. In this work, we propose a further adaptivity of the mother wavelet by employing two critical extensions of the above formulation. First, we allow learning of the knots' positions of the mother wavelet. This enables learnability with varying instantaneous frequency. Even chirplets [53] can be learned in this scenario. Second, we allow learning of the scaling factors used to generate the filter bank from dilating the mother wavelet layerwise. This entails learning the center frequencies of the wavelets. Formally, the filter bank is obtained through dilation of the mother wavelet given a collection of scaling factors, Λ, as in

$$\text{FilterBank} = \{\psi_\lambda, \lambda \in \Lambda\} \;\text{ with }\; \psi_\lambda(t) = \frac{1}{\sqrt{\lambda}} f_{\Theta,\Gamma}\left(\frac{t}{\lambda}\right) \tag{4}$$

As a result, the learnable parameters of the scattering transform are $\Theta \cup \Gamma \cup \{t_1, \ldots, t_K\} \cup \Lambda$, where the labeling of the DSN layers by ℓ is suppressed in the notation.

C. Time scattering learning stability

The learnable scattering transform coupled with RNNs generates learning instabilities due to variations of the knots. In fact, a translation in time of the knots results in the well-known phenomenon of *exploding gradients* [54]. We employ *gradient clipping*, an optimization constraint to maintain the gradient values of the knots within an interval. In our case, gradients are clipped between $[-1, 1]$. The optimization of the filters does not need additional constraints nor gradient clipping.

V. Architecture: Scattering Aware Recurrent Convolutional Layers

We now describe the modules of our recurrent architecture taking the scattering coefficients or features as input. These recurrent components comprise the volumetric layer, the pair of modified ConvLSTM 1_λ and ConvLSTM 2_λ, the time-skip connection and the temporal classification module in Figure 1.

A. Volumetric Layer

The recurrent DSN layers, $L1$ and $L2$ say, generate the first- and second-order scattering coefficients. The scattering coefficients generated by both DSN layers can be considered as an image upon identifying time-scale as spatial coordinates; the size of the image is $(J \times Q) \times$ (number of time samples). Thus an image is obtained as an alternative representation of the scattering coefficients of a signal while these vary with time. The number of time samples (and time interval) is determined by the time resolution after the pooling operation at the DSN layers $L1$ and $L2$ (see (2)). The volumetric layer arranges feature vectors of dimension n, say, into a single *volumetric representation*, λ_v. The volumetric representation consists of a multi-channel image with $\lambda_{2,n} + 1$ channels, arranged by scattering order. The first channel consists of the first-order scattering coefficients, and the $\lambda_{2,n}$ following channels constitute the second-order scattering coefficients. Thanks to the learnability of the filter banks in the DSN layers coupled with the subsequent recurrent networks, the multi-channel volume λ_v adjusts its sensitivity to the set of important frequencies, event shapes and transient dynamics.

B. Dynamic Convolutions

In the upper branch, denoted by A, two temporal LSTM networks take the volumetric representation and produces a sequence of tensors, $\mathcal{C} = (C_1, C_2, ..., C_n)$ as output. Instead of learning a unique representation from the scattering layers $L1$ and $L2$, the upper branch of the network learns the full sequence of scattering feature vectors from the volumetric layer. The multi-channel image will be codified into a set of sequential hidden states H and memory cells C that explicitly encode the intra-scale variations of the input feature vectors. To achieve this, the core components of *dynamic convolutions* are Convolutional-LSTMs (ConvLSTMs), a mathematically modified LSTMs units that have replaced its internal structure of standard multiplications with convolutions [29]. The incorporation of internal convolutions in ConvLSTMs remove the limitation of standard LSTMs that operate sequentially over single feature vectors

[55]. Our branch of *dynamic convolutions* follows the governing equations described in [29], but are modified to accept the volumetric representation, λ_v, as input and performing a sequential analysis of n steps equal to the number of channels in λ_v.

The first ConvLSTM, (ConvLSTM 1_λ in figure 1) probes the volumetric representation and performs a sequential analysis with each feature vector. We note the sequence $\mathcal{A} = [A_1, A_2, A_3, \ldots, A_n]$ as the input-to-state transitions from λ_v to the memory cell $\mathbf{C}_n$ in ConvLSTM 1_λ. The memory cell in a ConvLSTM has an embedded *read-and-write* functionality to select and store sequential meaningful information. This functionality is implemented by a set of matrices, also known as gates, that control the flux of information that has to be stored in $\mathbf{C}_n$ from the input feature vector sequence. In our ConvLSTM 1_λ, these gates control and select which time-scale dependencies are written and kept in memory cell $\mathbf{C}_n$. More precisely:

$$\mathbf{C}_n = f_n \circ C_{n-1} + i_n \circ \tanh(W_{\mathcal{A}\,c} * A_n + W_{hc} * H_{n-1} + b_c) \tag{5}$$

$$H_n = o_n \circ \tanh(\mathbf{C}_n) \tag{6}$$

with b_c the bias and H_{n-1}, H_n, the previous and current and hidden states for sequential step n. The symbol $\circ$ denotes the Hadamard product, whereas $*$ represents convolution. The term $W_{\mathcal{A}\,c}$ in (5) corresponds to the parameters of the input-to-state transitions, while W_{hc} corresponds to the internal state-to-state transitions, that is, the internal connections in ConvLSTM 1_λ from the hidden state H_n to the memory cell $\mathbf{C}_n$. The f_n or forget gate controls the amount of information that is erased, whereas i_n is the input gate that controls what must be written in the memory cell. The output gate o_n controls the amount of information that is passed from the memory to the output sequence. All these three gates, f_n, i_n and o_n are 3-dimensional tensors that for each sequential step n reduces to a matrix whose rows and columns are the processed values of A_n that must be stored or erased in the memory cell $\mathbf{C}_n$. As the gates operate over the sequence of time-scale variations, the internal memory preserves the number of time samples for all the elements in $\mathcal{A}$. ConvLSTM 1_λ is designed to learn and select the feature vectors that best inform the subsequent temporal classification modules. The output of ConvLSTM 1_λ is the sequence $\mathcal{B} = [B_1, B_2, \ldots, B_n]$, with each element signifying a new feature vector.

The second ConvLSTM, indicated by ConvLSTM 2_λ, is identical in its design to ConvLSTM 1_λ. The output of this ConvLSTM 2_λ is the feature vector sequence $\mathcal{C} = (C_1, C_2, ..., C_n)$. Its application results in a refined time-scale representation in which the interaction of the scattering

coefficients across multiple scales is manifest. We observe that the physical, frequency transient phenomena are linked with the dynamics of the learnable scattering transform.

C. Time Skip Connection

In parallel to subjecting the volumetric layer to ConvLSTMs, we apply a single convolution. This is a *time-skip connection* (path B), that parses the multi-channel scattering coefficients into an image of size $(J'\ Q') \times (\text{number of time samples})$. The single convolution collapses the multi-channel scattering coefficients into a single output value. We do not apply any pooling to preserve time resolution in the output representation.

D. Feature Fusion

The output of the branches A and B are concatenated in a new multi-channel image as follows. For each time we stack the *columns* from the images generated by branches A and B. The components for each time are labeled by λ'. The result we denote by S, essentially a matrix signifying feature fusion (in green in Figure 1).

E. Temporal Classification

The bidirectional configuration is designed to provide long-term contextual information from the entire input sequence S, one working as a forward and the other as a backward recurrence (see BiLSTM1 and BiLSTM2 in Figure 1). The number of sequential steps for the forward and backward LSTM unfolds is given by the length of the dimension $time$ in S. The bidirectionality of the recurrence allows incorporating contextual information from past and future frames, reinforcing the detection capacity of our network. Finally, the output of the BiLSTM2 is fed into an MLP to produce the prediction, that is, the predictive probability heatmap.

VI. Uncertainty and change quantification

The reliable detection of change in seismo-volcanic data streams remains a complex task given the polyphony associated with multiple sources, tremor background, and overlapping signals. The challenge is to find a universal approach, irrespective of the geological environment, to systematically detect anomalous behavior and change.

The distribution variations of $D = \{(X, Y)\}$ are especially challenging to detect in streaming data [56]. Estimating the uncertainty in a monitoring framework with streaming data requires

the generation of a distribution over the network's predicted outcomes, fitting with a Bayesian statistical framework [57]. The challenge in developing this in deep learning lies in providing an effective approximation to the high-dimensional parameter space of deep networks that is both fast and numerically reliable. Research work by [24] has related dropout neural networks with Bayesian statistics by sampling from multiple dropout masks to infer an approximation to the neural network's posterior distribution (for details, see appendix A). This approach has the advantages of scalability and integration with already well-established deep-learning training methodologies. In seismology, Bayesian deep learning has been explored through a dropout approximation for phase picking, and earthquake localization [25]. Bayesian deep learning has also been applied to the probabilistic classification of events from filter-bank based features, characterizing pre- and post-eruptive periods based on different types of seismic events [57], [58].

A. Bayesian Monitoring

The sources driving change in a volcano are composed by an unknown number of latent, heterogeneous variables that contribute to the overall alteration during monitoring. We gather all sources of uncertainty into the *epistemic uncertainty* of our model at any given monitoring time [59]. Starting from the distilled scattering feature vectors S_t in the *feature fusion* module, MC dropout is invoked in the final two bi-directional LSTMs (BiLSTM1 and BiLSTM2) and the fully connected predictive layers (contained in the dashed square in Figure 1). That is, we fix both the scattering layers and the ConvLSTMs modules. The variational inference framework based on Monte Carlo numerical sampling and its connections with stochastic regularization techniques in deep learning are described in the Appendix A. During the training stage, invoking stochastic dropout on the last layers can simplify the variational procedure while still providing meaningful results from a Bayesian perspective. From a statistical perspective, it can be interpreted as a jointly-learning point estimation followed by a shallow BNN [60], [61]. When new data streams are presented to the neural network, the two bi-directional LSTMs and the fully connected predictive layers analyze the processed scattering feature vectors and compute an uncertainty estimate of the classification and segmentation for each of these processed vectors. This procedure permits quantifying signal variations due to any external mechanism presumably closely related to volcano dynamics.

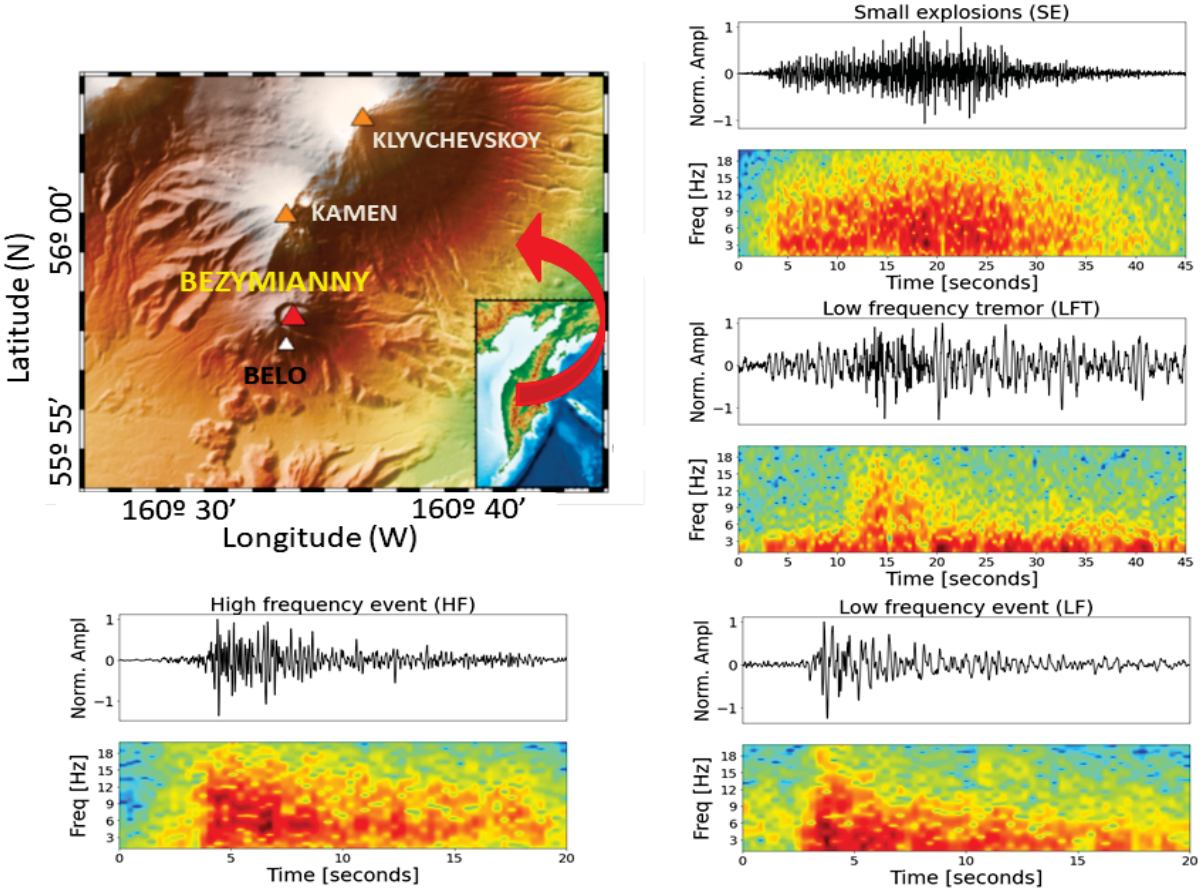


Fig. 2. **Map of Bezymianny volcano (Kamchatka, Russia) and representative examples of common waveforms recorded during the 2007 eruption.** For each waveform type, the normalized waveform (in black) and spectrogram are depicted. For visualization purposes, all waveforms have been filtered between 1 and 20 Hz.

VII. Results

This section has been structured into three main subsections. The first subsection of this study demonstrates the monitoring capabilities of the implemented architecture and multisource (or polyphonic) monitoring tasks. To do this, we select three main eruptions of the Bezymianny volcano during the 2007 eruptive sequence. Then, using the polyphonic system, we present a complete application of how uncertainty and its adjustment with power-laws or exponential functions can predict eruptive processes successfully in Bezymianny volcano. We then study the exportability of the polyphonic system to other volcanic scenarios by doing a blind test with data from Mount Saint Helens (Mt. St. Helens), a volcano similar to Bezymianny.

The second part of this study demonstrates the ability of the system to perform predictive

TABLE I

BEZYMIANNY 2007 POLYPHONIC METRIC RESULTS

System	Eruption 1			Eruption 2			Eruption 3		
	F1 (1s)	**PR (1s)**	**RC (1s)**	**F1 (1s)**	**PR (1s)**	**RC (1s)**	**F1 (1s)**	**PR (1s)**	**RC (1s)**
Vanilla	85.86	87.09	84.66	61.97	69.07	56.19	84.28	86.75	81.96
Filters	85.35	86.72	84.07	61.45	69.24	55.24	84.62	86.23	83.06
Knots + Filters	85.29	86.64	83.98	59.71	65.09	55.15	84.45	86.36	82.62

tasks in a completely different scenario, the Mt. Etna volcano. We first study how the uncertainty behave in a blind test after we switch the volcano type. We then perform a transfer learning approach and extend this study to use the uncertainty as a predictive element of eruptions in the Mt. Etna volcano.

A. Bezymianny Volcano

We selected an eruptive sequence from Bezymianny volcano (Kamchatka, Russia). The data catalog contains 3 months of daily records, from 1 September to 5 November 2007 [62]. During this period, three significant eruptive episodes (those on 25 September, 14 October, and 30 October) were reported by the Kamchatka Volcanic Eruption Response Team (KVERT), and confirmed via posterior geophysical studies [63]. Figure 2 shows the volcanic area and representative seismo-volcanic events recorded during this period and that comprises the studied dataset. Of the three eruptive episodes, which were all dominated by strong ash explosions and lava emissions, the second is considered the most energetic. This event continued for 2 full days, with the plume reaching 10 km in height and extending 1000 km to the southeast. Of the seismic stations monitoring the eruptive activity, we selected BELO, located near the Bezymianny dome and crater; BELO is a near-field station where any attenuation effects on the seismic wave field can be neglected.

We categorized recorded waveforms according to the terminology proposed by [8]. This categorization scheme includes low-frequency (LF), high-frequency (HF), seismic volcanic tremor (SBT), surficial events (SE), and low-frequency tremor (LFT); (see figure 2). These labels presumably distinguish seismic source mechanisms induced by volcanic processes. Moreover, although seismic networks can record several events per minute during volcanic crises (over periods of days to years), there remains a lack of uniformity in giving meaning to the labels

of observed signals. Further complications arise when several processes co-occur, producing a suite of overlapping signals. Our neural network architecture utilizes all of the data, and through polyphonic detection, segmentation, and classification enables insight in the dynamics of the volcanic system.

We subdivided the dataset into time intervals based on periods of seismic unrest. During quiescence periods, the level of seismic activity was very low. During periods of seismic unrest, occurring before each eruptive episode, seismic activity increased in number of events and energy. After each eruption, seismicity subsided, leading to another quiescence period. This dataset division strategy also considered representative numbers of seismic events for training and testing. For the first eruption (on 25 September), the training data covered the period from 1 to 24 September. For the quiescence period between the first and second eruptive events, the training data covered the period from 26 September to 13 October. For the quiescence period between the second and third eruptive events, the training data covered the period from 17 October to 4 November. We tested our approach for each period independently, so as to gain deeper insight into the performance of our neural network as a monitoring tool.

B. Monitoring Results: Bezymianny

We follow the nomenclature vanilla, filters, and knots+filters to distinguish the learnable options of the scattering network module in our E2E approach. The hyper-parameters of our network are explained in Appendix B. We adopted the categorization scheme proposed by [8] for training our network.

Table 1 shows the polyphonic metrics for the test (eruptive) partitions, trained using data from the quiescent periods before each of the respective eruptions, with a frame precision of 1 s. Our approach yielded excellent performance for each of the eruptions. The PR, RC, and F1 metrics attained high values for all of the scattering network options. Notably, our proposed learnable configuration resulted in a high number of correctly detected frames at 1 s precision, along with the effective event categorization of localized scattering information (i.e., high PR). This performance is rooted in scattering learning stability inside the recurrent architecture. It is important to emphasize that the three eruptive processes (i.e., quiescence, unrest, and eruption) occur over a very small interval of time, but each involves different physical processes that produce different seismic sources. In each process we have a first interval (quiescence) where the volcanic system seems to be in an apparent rest. The system quickly enters unrest and shows

an acceleration of all the seismic observable culminating in an eruption. Finally there is a period of return to stability to enter the period of quiescence again.

Our polyphonic approach yielded high RC and PR during all three quiescent periods. The recurrent scattering architecture mitigates one of the main challenges in seismo-volcanic monitoring owing to its ability to adapt. Notably, in the second eruptive period, during which multiple physical mechanisms were active at the same time, the network's performance showed high PR and F1 scores, with RC being at an acceptable level for seismic signal detection and identification. The network successfully dealt with significant variability in the durations of events and with overlapping events. More than 50% of frames were correctly detected and assigned with high precision to multiple and co-existent seismo-volcanic classes. As an aside, we achieve favorable performance metrics for event detection as a single task (see appendix C). These results confirm that our approach has significant advantages over more traditional monitoring protocols [14].

C. Predictive Heatmaps: Bezymianny

We use probabilistic heat-maps as an intuitive representation to verify if the learnable scattering transform has effectively adapted to performing polyphonic event segmentation; that is, these maps provide visual evidence that our neural network can detect, segment, and classify simultaneous seismic waveforms. Figure 3 shows probabilistic heat-map predictions with 1 s resolution for the pre-eruptive seismo-volcanic data stream of 22–24 September 2007. Attaining accurate segmentation is critical for seismo-volcanic monitoring; the duration of a recorded event is directly linked to the type of source mechanism that exchanges energy with its surroundings. For example, low and high-frequency events have short durations (up to 30 s), while volcanic tremor can last from minutes to hours. Isolated events are generated by sudden fluid (water or magma) pressure changes (e.g., LF events) or the sudden release of energy due to rock rupture (e.g., HF events). Meanwhile, volcanic tremor arises from the sustained exchange of energy with the surroundings as the result of multiple pressure pulses within a fluid owing to bubbles, energy transients due to fluid flow, and/or other dynamics. Given their geophysical relation to the source, information inferred from each seismic event is equally important but is only partially encoded in its duration and frequency content. Our recurrent scattering network captures long- and short-term dependencies at 1 s resolution, departing from standard approximations that often require a posteriori analysis. The multi-resolution analysis underlying scatter-gram variation enables the network to learn the set of scattering coefficients with maximum intra-

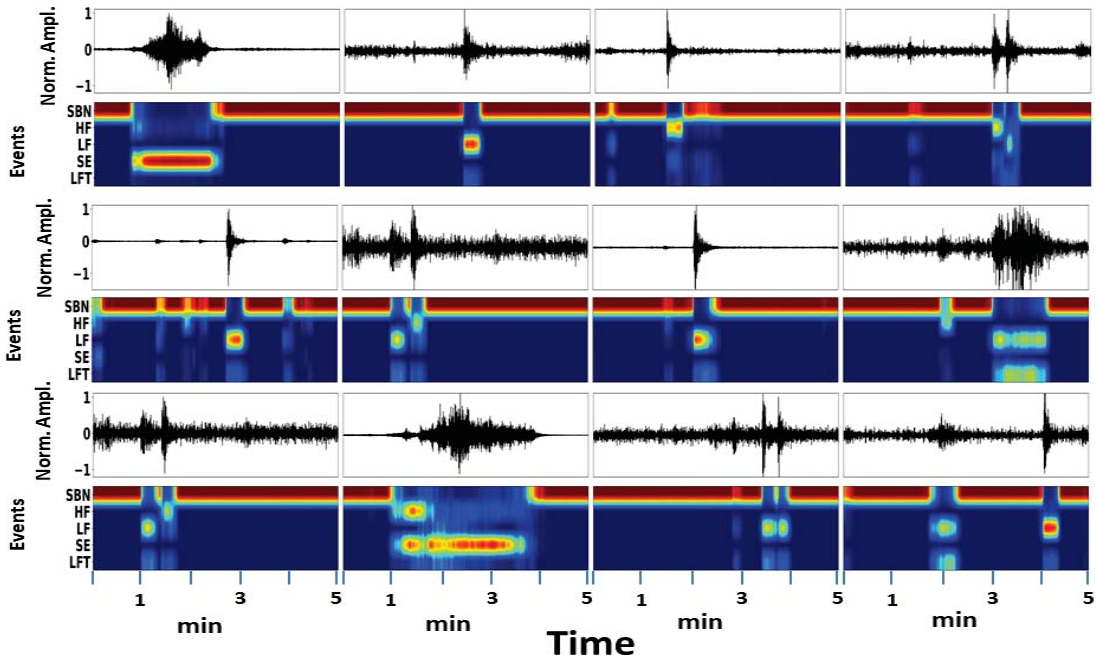


Fig. 3. **Pre-eruptive heat maps for Bezymianny volcano (Kamchatka, Russia) from 21 to 24 September 2007**. Normalized waveforms and per-class probabilistic heat-maps for the pre-eruptive sequence at Bezymianny volcano from 21 to 24 September 2007. The neural network systematically isolates fundamental seismo-volcanic events, even when they occur in rapid succession. The scattering transform can identify multiple active sources—including low frequency (LF) earthquakes—and uses the recurrent structure to unmask the coupled mechanisms from the background noise or tremor. The polyphonic multi-output provides invaluable geophysical information about potential sources and refines our understanding of volcanic unrest.

event information. The polyphonic approach is evident in the multiple frequency components isolated as sparse probabilistic maps, concentrating higher output probabilities to higher spectral content. For example, the waveforms of HF and LF have durations that are longer than the actual duration of the source because they are composed of direct waves incoming from the source (being the most energetic portion of the signal) and successive arrivals of waves coming from the resonance of the system, or scattered waves generated by heterogeneities of the medium, among other possible phenomena. Monitoring challenges at Bezymianny include the energetic background tremor and noise level. However, our results show that the effect of these on intra-class separability is negligible.

In Figure 4, we show seismic traces from the first eruption (25 September). Our approach succeeded in isolating and categorizing the presence/absence of volcano-seismic signals despite

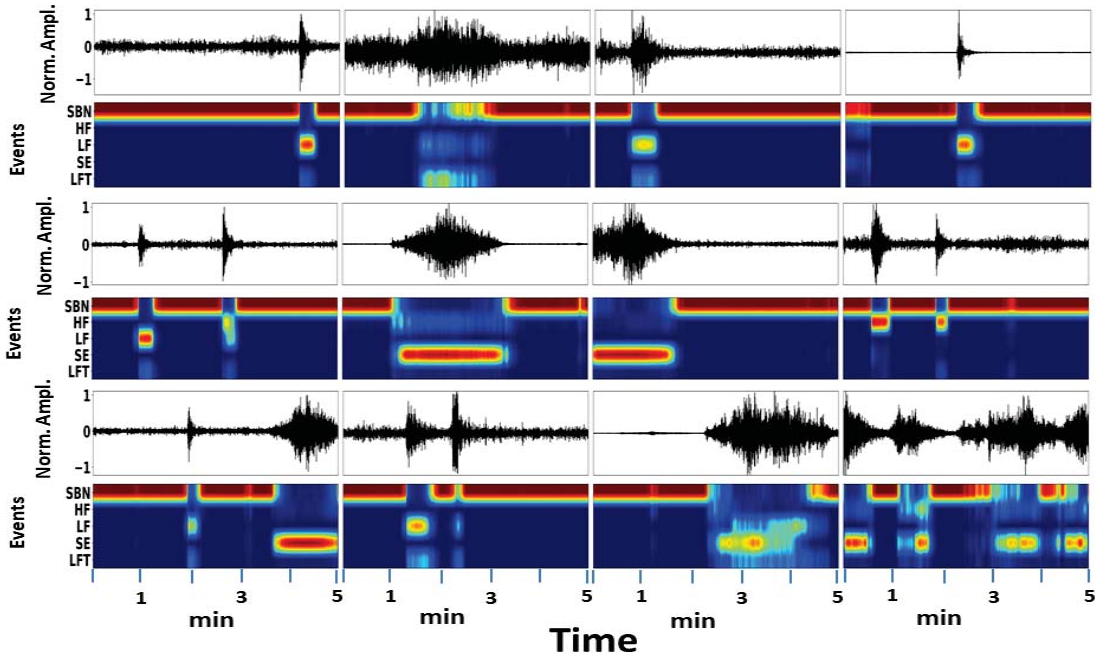


Fig. 4. **Heat maps for the 24–25 September 2007 eruptive period at Bezymianny volcano (Kamchatka, Russia).** Normalized waveforms and per-class probabilistic heat-maps for the eruptive sequence at Bezymianny volcano from 24 to 25 September 2007. The first two rows represent reported mild pre-eruptive seismicity (24 September). Earthquakes are classified as low frequency (LF) and high frequency (HF) events, with occasional surficial debris processes (SE) identified during dome inflation. The bottom row (25 September eruption) represents co-existent debris processes and low frequency tremor (LFT) during the main eruptive episode; the LFT component is recognized over the recorded exogenous signal, effectively demonstrating how a polyphonic approach enhances volcano monitoring by providing all potential sources active at any given time.

the per-class probabilistic heat maps showing many recorded concurrent processes. As an example, elevated confusion occurs within the lower left seismic trace, in which there exist many simultaneous processes with mixed frequencies, influencing intra-class separability and predictive performance. The recorded SE activity could be caused by increased deformation through dome growth paired with energetic rockfalls recorded prior to an eruption.

In Figure 5, we show a set of polyphonic heatmaps for data streams from 9 to 16 October (first two rows) and 30 October to 5 November (last two rows). Polyphony is evident in the heat maps for all of the earthquake transients; our neural network detected LF and HF events co-existent with LFT. Furthermore, for the third eruptive period (the last two rows), polyphony is also clearly visible, with both HF and energetic SE classes. As mentioned, in each eruptive period, a series of changes in the seismic wavefield are used for forecasting. Alongside every

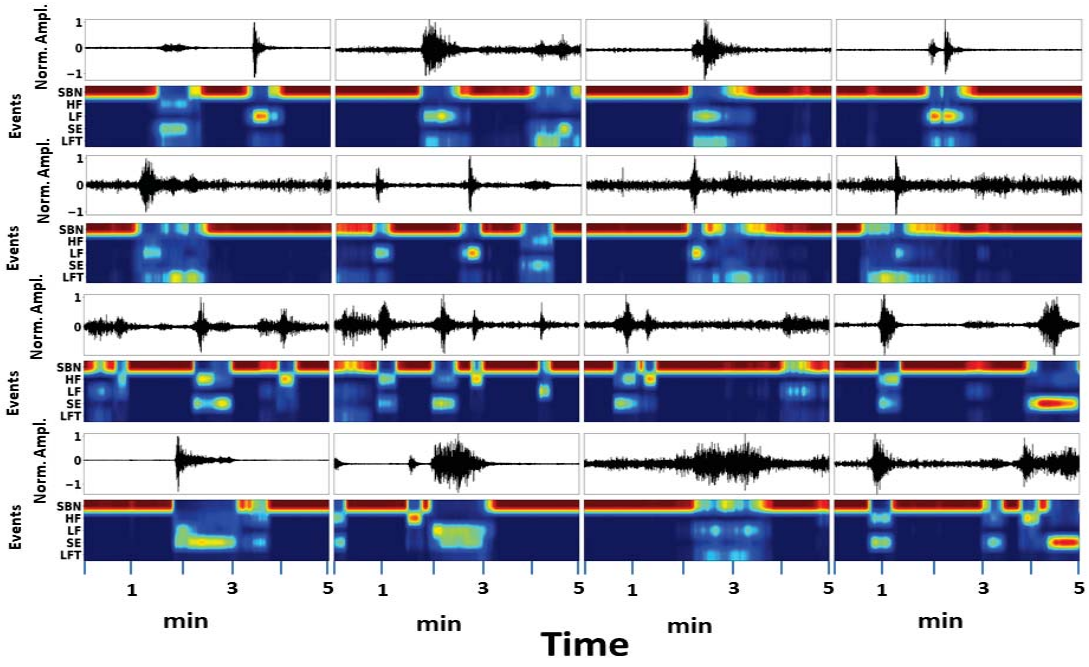


Fig. 5. **Heat maps for Bezymianny volcano (Kamchatka, Russia) from 9 October to 5 November.** Normalized waveforms and per-class probabilistic heat-maps for Bezymianny volcano from 9 October to 5 November 2007. Note that the waveforms present concurrent processes. Multisourced signals arise in the low-frequency components, as the frequencies of low frequency (LF) and high frequency (HF) events (i.e., earthquakes) overlap with the those of low frequency tremor (LFT). Similarly, surficial events (SE) are correctly segmented, as their arrivals and subsequent envelopes differ from those of earthquakes.

data stream that has been extracted and classified is the ever-present volcanic tremor. The tremor wavefield also changes as the volcano evolves, and in some cases can become highly energetic and relevant from an identification and classification point of view. For example, in the second eruptive period, such a background signal was manifested and identified as the LFT. These results show that our neural network succeeds in intra-class separability of earthquake transients and background signals, even under intense seismic fluctuations. Our neural network learns, without supervision, the energetic onset/offset times of events, being sensitive to subtle fluctuations in seismic transients from background noise and surficial events. Indeed, our neural network can operate as an automatic event detector with the maximum probabilities of temporally aligned and concentrated seismic signals. An interesting observation is that when our approach switches the highest probabilities from volcanic tremor to a different class, a low probability emission is generated before the potential arrival of the event in this different class, as it can be noticed from

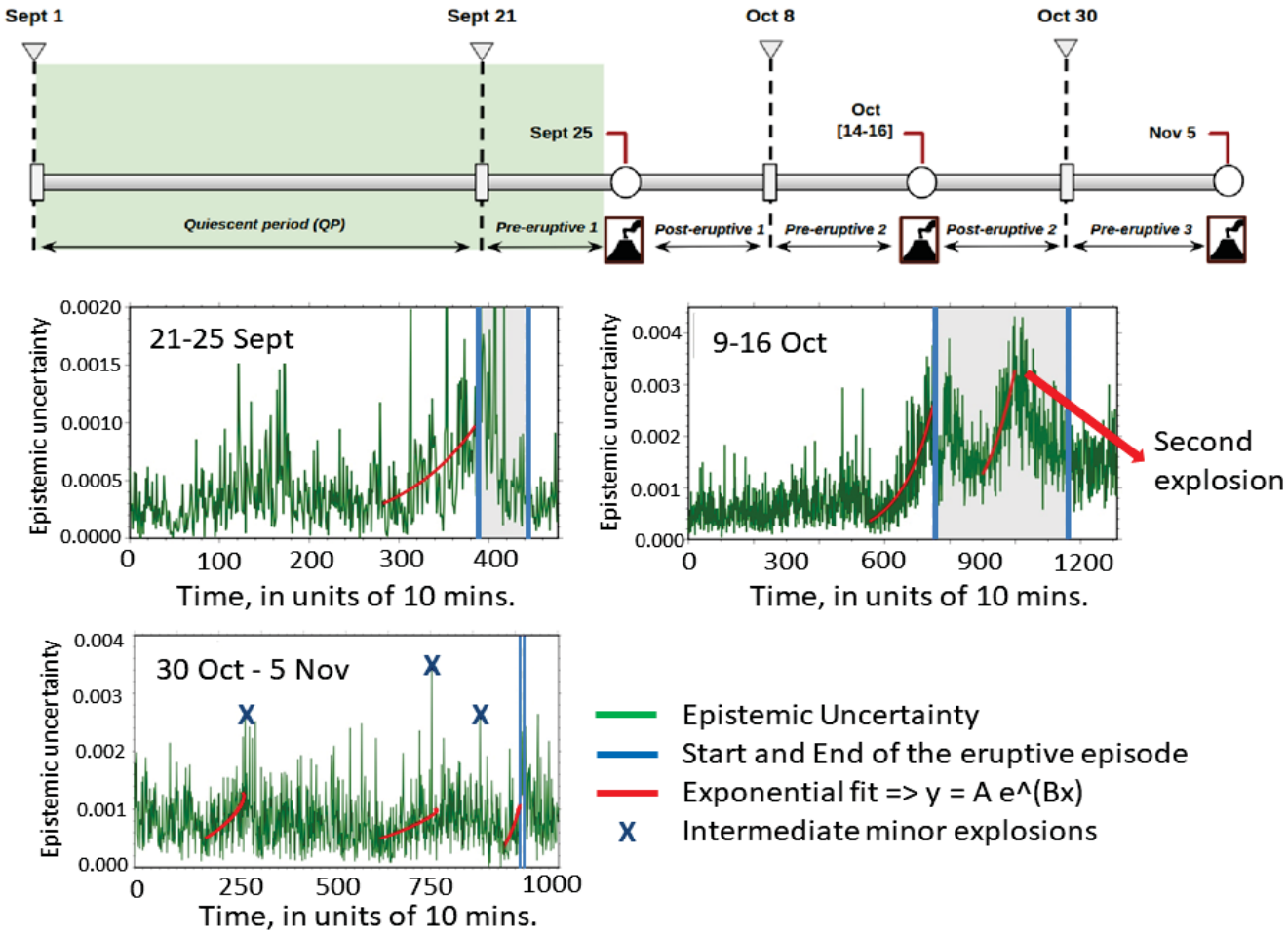


Fig. 6. **Temporal epistemic uncertainty variation and detected drifts in pre-eruptive and eruptive sequences at Bezymianny volcano.** The main reported eruptions are indicated in gray. Multiple explosions for the third eruptive episode are indicated by ×. Fits to power laws preceding eruptions and explosions are indicated in red. The Bayesian recurrent scattering network successfully detected change prior to the main eruptions, revealing that changes in the uncertainty are related to unrest.

figures 3, 4 and 5. The learned scattering transform provides a physics-informed identification of the distortions in the near seismic wave field. Such information serves as an a *a-priori* estimate of the phase arrival time, covering an extensive low-probability range in time until the energetic arrival. These probability switches are consistently present despite signal-to-noise ratio (SNR) and spectral differences between seismic signals and background tremor.

D. Drifts in Pre-eruptive Epistemic Uncertainty

Epistemic uncertainty is uncertainty in model predictions reflecting the degree of data knowledge of the model [57]; here, the model is the classification module of our designed recurrent

scattering network. The data stream is mapped to the tensor S_t, which is forwarded to the temporal classification module (corresponding to BiLSTM1 and BiLSTM2 in Figure 1). The tensor S_t contains, for any given frame time t, the mixture of all the scattering coefficients that best explain the raw observable signal. From a Bayesian perspective, S_t is used as an embedding in which compressed information can be used to approximate epistemic uncertainty. The greater the data availability, the better we can approximate the training data distribution, with lower epistemic uncertainty values. Regardless, the Bayesian approach is always conditioned to the training data distribution, in which probabilistic shifts are detectable via epistemic uncertainties [57]. For our learned scattering recurrent network, and for each day in the validation/test streaming data sets, we computed the predictive map from each of the 20 stochastic Monte-Carlo dropout samples, chronologically at 10 min time intervals. Our hypothesis was that a drift in epistemic uncertainty would be directly correlated with changes in the volcanic processes.

Figure 6 depicts the results for the pre-eruptive and eruptive sequences for each of the eruptive periods considered here; each point on the horizontal axis corresponds to a 10 min interval and eruptions are marked in dark blue. In the pre-eruptive sequences, the predictions are remarkably robust, reflecting the invariances encoded by the scattering transform module of the network. Drifts in epistemic uncertainty preceding eruptions signify detectable changes in volcanic processes. Remarkably, these drifts exhibit power-law behavior in time, directly indicative of metastable behavior. Prominent peaks in epistemic uncertainty correlate with high-energy events that are, clearly, not present in the training data sets. The general time fluctuations in the uncertainty mostly relate to the complexity associated with polyphony in the predictions. The uncertainty may not return to values encountered in a pre-eruptive period; this can be attributed to changes in material properties and structural conditions after eruption. In this case, our deep neural network might need to be retrained, for example, through transfer learning. The plots in Figure 7 are consistent with those of previous studies [62], [64], [65].

For the first sequence (21–25 September), only minor pre-eruptive seismicity was recorded, with no significant post-eruptive changes. Our epistemic uncertainly shows a sustained drift in the form of a power law (in red) for 208 time units (or 34.6 h) prior to the eruption; the uncertainty peak coincides with the main explosion. This first eruption ends at 442 time units (73.6 h), while the uncertainty returns values encountered in the pre-eruptive period. For the second sequence, the eruption of 14 October was preceded by increased seismicity and tremor, including relatively large magnitude earthquakes. The eruption was characterized as explosive and lasted until the

17 October. The uncertainty again shows a clear drift following a power law (in red) for 197 time units (32.83 h) prior to the eruption on 14 October. A second, short-lived but high-energy eruption is clearly visible in the uncertainty on 15 October. Preceding this second eruption, the drift in epistemic uncertainty is attributed to an increment of tremor amplitude. Finally, the third eruption at Bezymianny, that on 5 November, was reported as weak with minor fumarole activity. The activity could not be confirmed owing to weather conditions, but it was postulated that a sequence of explosive events occurred. We can confirm such explosive activity based on two spikes in the uncertainty at 700 and 810 time units along with other minor explosive activity that could have preceded the main explosion at 914 time units. Again, the individual explosions were preceded by a power-law drift in the uncertainty (in red) at 63 and 123 time units (10.5 and 20.5 h) for the first and second individual eruptions, respectively.

E. Power-law and Quasi-exponential Uncertainty Curves

In volcanology, eruption prediction methodologies include one based on modeling the rate of change of selected observables using a differential equation with exponent, α [66]. Exponential behavior occurs if $\alpha = 1$, and power-law behavior occurs otherwise. We carried out careful fitting [67] of α to the epistemic uncertainty obtained for the three eruptive periods (Figure 6; note that the exponents of the eruptions are not 1), and argue that α estimated from epistemic uncertainty is indicative of eruptive type.

A power law defines a polynomial relationship between two quantities, in which relative changes in one leads to a proportional change in the other, modeled by exponent α. In our volcanic setting, the quantities are time and uncertainty. Power-law models are often referred to as a "material failure forecast" and can be used for the prediction of volcanic eruptions [68]. Using this approach, it is more important to determine if there is an acceleration in the change than if there is a change in itself; therefore, the second derivative is critical. As such, an exponential of greater than 1 is often interpreted as a precursor element. Power laws are universal across volcanoes, permit scaling, and implicitly encoding acceleration. Their use is not new [8], [68]; however, by using uncertainty as an early warning element, our method can reduce the uncertainty whether or not there is an increase in the homogeneity of the data. If the uncertainty grows, it implies the appearance of new classes of events or that the events within a class have begun to differ from each other. Therefore, an increase in uncertainty, which initially reflects a lack of classifier quality, can be used as a good indicator for early warning of volcanic

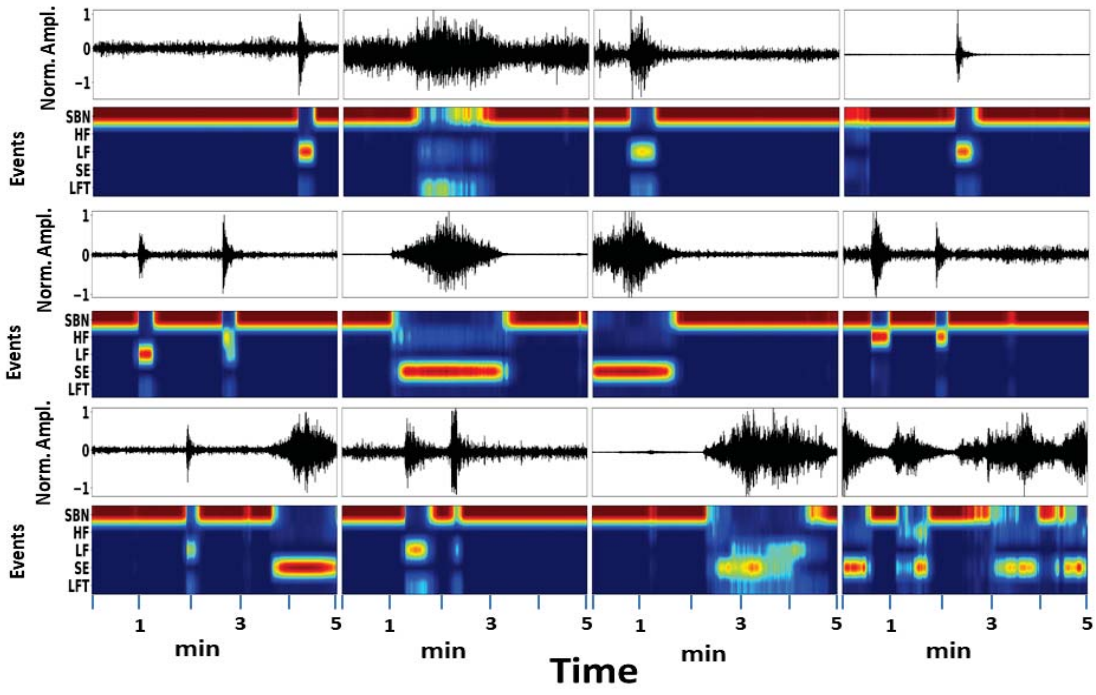


Fig. 7. **Generalization and exportability test of the implemented system using data from Mt. Saint Helens.** Normalized waveforms and per-class probabilistic heat-maps for Mt. St. Helens volcano from 0000 PDT to 23:59 PDT on 9 September 2005. Seismic waveforms are from station JUN. Seismic signals are characterized by regularly spaced events (or "drumbeats").

eruptions. It is important to note that this approach -measuring the variation of the uncertainty- can be exported between volcanoes. While the nominal values of the uncertainty will differ among volcanoes, observing the derivatives is more important than the absolute nominal value of the uncertainty.

Figure 6 shows changes in uncertainty before each of the three eruptive events. The nominal values of uncertainty differ, but similarities can be observed in the acceleration processes immediately before each eruption. Among the three eruptions, field observations indicated that the second was the most energetic, while the third was the least energetic. The shapes of change for each uncertainty curve differ and are associated with the energy of the eruption. Both the forecasting time and exponent of the second eruption exceeded those of the other two events. For the third eruption, we identified potential previous failures and a law much closer to 1. These observations are promising as they support the use of this approach for early warning.

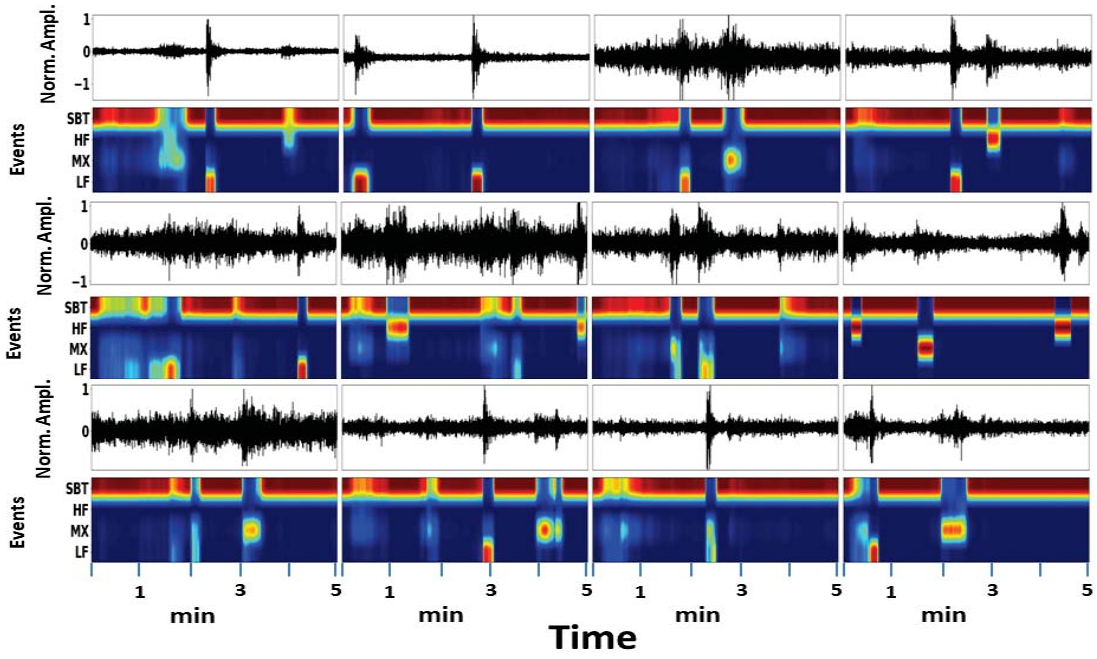


Fig. 8. **Generalization and exportability test of the implemented system using data from Mt. Etna.** Normalized waveforms and per-class probabilistic heat-maps for Mt. Etna from 14 to 24 July 2019. Seismic waveforms are from station ENCR. The polyphonic segmentation of the events permitted the identification of low frequency (LF) and MX frequency events, demonstrating the existence of coupled mechanisms, such as hybrid-type events.

F. Blind Test: Mount Saint Helens, 2005

Bezymianny and Mount Saint Helens (Mt. St. Helens) share a similar composition and eruptive style; both primarily andesitic and present a classic stratovolcano-type morphology. Both produce explosive eruptions of high viscosity magma associated with a high rate of volcano seismicity.

The seismic data from Mt. St. Helens were gathered during a vent-clearing phase recorded from 9 September (starting from 0000 PDT) to 5 October 2005 at the JUN station. This seismic sequence comprises regularly spaced earthquakes, of very similar magnitude and waveform also known as drumbeats, resulting from stick-slip motion of a conduit spine. During the vent-clearing episode, the event rate increased to up to 3 earthquakes per minute, towards a sustained eruption. Figure 7 shows the polyphonic results for the blind-test.

For explosions registered on 10 September 2005, although the arrival of the waveform is clearly visible and detectable by less sophisticated algorithms [69], it is within the polyphonic categorization where similarities across volcanoes are perceivable. Our pre-trained network can

detect events that share waveform properties with those at Bezymianny, and, despite the differences in geophysical processes, it successfully highlights the current mechanisms via probabilistic heatmaps. We note that with the pre-trained network we reliably detected the earthquakes in a polyphonic setting, assigning a high probability to LF, HF, and SE classes. These experimental results demonstrate the success of the proposed methodology for volcanoes that share similar geophysical features, and thus being easily exportable between volcanoes without re-training the whole architecture with a pre-existing dataset.

G. Transfer Learning, Switching Volcano Type: Mt. Etna

We also applied our Bezymianny-trained deep neural network to data from Mt. Etna. The internal morphology of Mt. Etna frequently changes owing to intra-crater volcanic activity [70], [71], [72], [73], which provides a challenge for our approach of detecting change with epistemic uncertainty. The significant changes in eruptive activity and correlation with other volcanic data are described in [74]. We used data gathered at the south-east of Mt. Etna (Bocca Nuova, BC) during the paroxysm recorded from 4 to 24 July 2019, at the ENCR station. The main eruption occurred at 23:09 UTC on 18 July. The seismic events and style of the eruption were fundamentally different from those at Bezymianny and Mt. St. Helens. The activity comprised explosive degassing preceding ash-rich explosions. Signals were characterized by HF and LF events, along with a new class of MX event, a hybrid frequency event that shares the spectra of LF and HF events. Hence, for this new volcanic scenario, we have the following classes: HF, LF, MX, and SBT.

H. Uncertainty blind-test exportability

As observed with Mt. St. Helens, our architecture and approach can be exportable and produce heatmaps to another volcano when no prior knowledge is available. Now we wonder if the study of the temporal evolution of uncertainty can be exportable without having prior knowledge of the new volcanic scenario. To do this, we performed a blind test for the selected Mt. Etna data using the pre-trained system in Bezymianny volcano. We first window the data stream as in the Bezymianny case but average the estimated uncertainty for 10 minutes. No training or additional fine-tuning of the system is further performed in this test.

Figure 9.B represents the temporal evolution of this exported uncertainty along time. The main observation is the uncertainty drops to zero when the main energetic moment of the eruption will

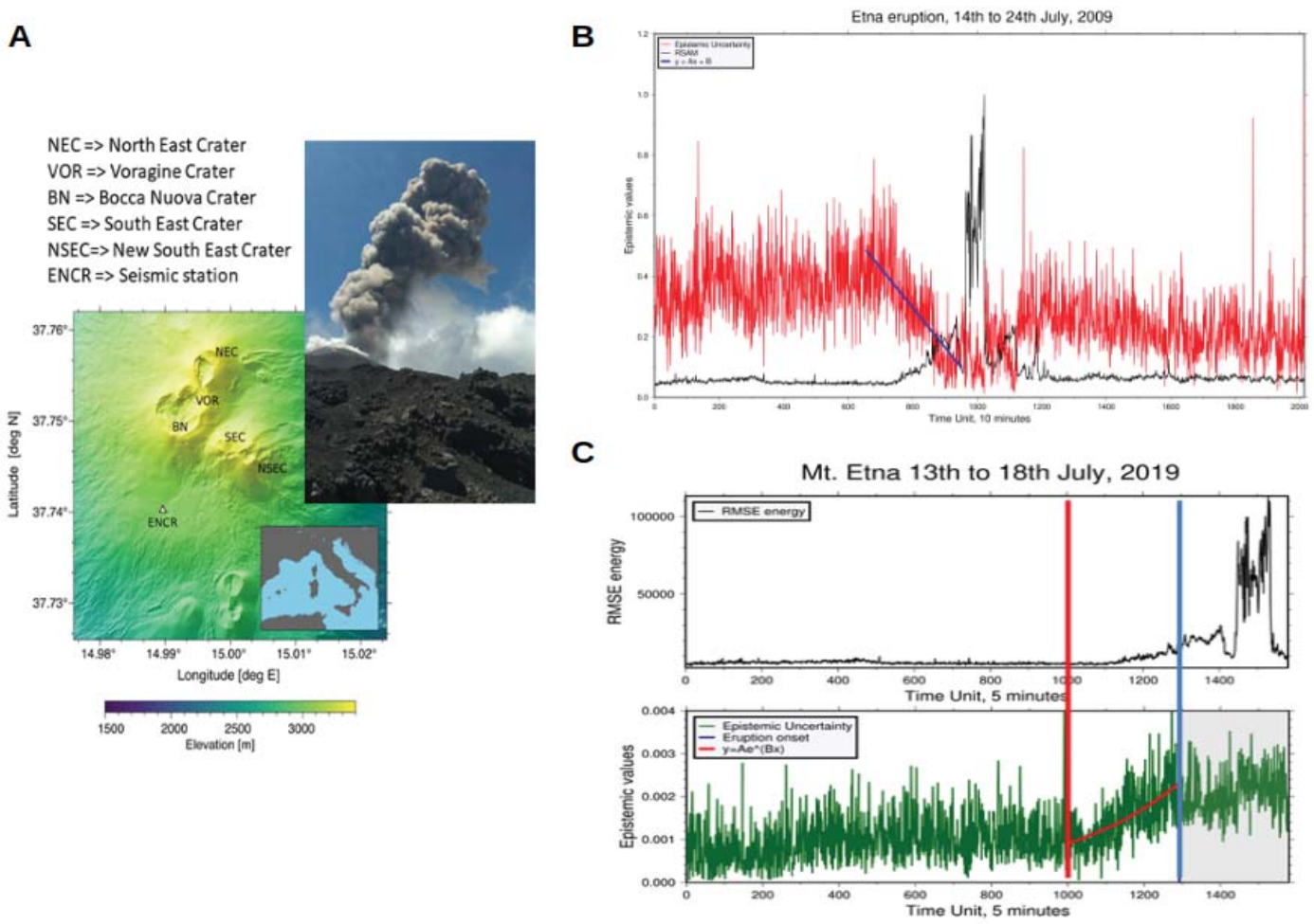


Fig. 9. **Switching volcano and eruption style with transfer learning:** (A) Image of the Mt. Etna eruptive column from on 18 July 2019, along with a map of seismic stations and the location of the ENCR station shown. (B) Temporal evolution of the uncertainty for an exportability test to Mt. Etna using the pre-trained system on Bezymianny. The blue line represents the linear fit when uncertainty drops to zero. Observing this line indicates in advance when the more explosive episode of the eruption will start. (C) RMSE energy and temporal uncertainty variation for the 2019 eruption after the system has been trained.

happen (observing the sudden increase of the energy), i.e. the network is effectively tracking the frequency bands of the seismic data stream through time, which is why the uncertainty remains relatively high at the beginning and drops to zero towards the eruption. The results of the blind test do not indicate when the eruptive process will begin but when the most energetic event will occur. If a simple linear fit is performed to the decay of the uncertainty in the blind test (blue line in figure 9.B), we observe that this line marks precisely the moment of greatest energy of

the eruptive process. This observation is fundamental because although it does not say when the eruption will occur, it does forecast when the paroxysmal moment of the eruption will occur. This observation cannot be generalized yet, but it is a significant result and opens a new venue to the forecasting processes of volcanic eruptions. From the perspective of physical models that could explain this behaviour, we do not yet have a conclusive answer. We believe that it is associated with the fact that at the paroxysmal moment of the eruption, the system begins to register again a high content of high-frequency signals, compared to the pre-eruptive moment where low-frequency events dominate. Therefore, in the blind text, as we approach the explosive paroxysm, high frequencies dominate again in the seismic data stream, similar to those already-known already known by the system. Hence, the exported experience with the uncertainty tends towards zero. We must insist that this test of exportability and the success as a predictor element (onset of the eruption and paroxysmal moment) is a promising result, and tests should continue in new volcanic scenarios to see it as a universal tool.

I. Transfer learning on Mount Etna

Applying our developed solution on different volcanoes can be done very simply by using our pre-trained model and only adapting the parameters of the last layers in the architecture. We adopted a transfer learning approach by replacing the whole temporal classification module with two new bi-directional LSTMs and a dense layer with the number of classes at this volcano (HF, LF, HYB, and SBT). Hence, the temporal classification module (dashed-square) in Figure 1 has been substituted by a new temporal classification module; but changing the number of hidden units to the same number of classes in this volcano. The rest of the seed network remained unmodified, that is, the recurrent scattering layers, the volumetric layer, the ConvLSTM 1_λ and ConvLSTM 2_λ; the time-skip connection and the feature fusion module. We followed the same training procedure as that used for Bezymianny and retrained the new temporal classification layers using the quiescent period of 10 days at Mt. Etna (4–14 July).

Figure 8 shows per-class heatmaps for the eruptive sequence recorded at Mt. Etna. Our system detected LFT and MX classes attributed to fluid interactions (i.e., tremor and hybrid events). Note that the MX class is well recognized and distinguished from the SBT class. The principled recognition of MX events is present for all of the depicted waveforms, confirming that the network has adapted to a new domain in which the broad spectra of seismo-volcanic events is different from that of the seed dataset. As Mt. Etna volcano has different volcanic dynamic

compared with Bezymianny or Mt. St. Helens volcanoes, there can be different geophysical interpretations for seismic signals with the same label. However, while the target dataset is significantly different from the original dataset from a machine learning perspective, the pre-trained model already has learned features relevant to the monitoring problem. These results show that the proposed network can reliably detect seismo-volcanic signals from different origins after transfer learning.

Figure 9.C depicts the temporal variation of the epistemic uncertainty for period before the main eruption. The root mean square (RMSE) energy is computed for the filtered trace between 0.5 and 12 Hz. First, it is noticeable that the uncertainty shows a gradual drift, following a power law (in red), at 1005 time units (or 23.91 h). Note that the energy does not exhibit such a clear drift from the usual background energy until 200 time units (7.6 h) prior to the main eruption. As discussed, it is not possible to compare nominal values of uncertainty from one volcanic system to another. However, we use uncertainty to define when our system can be considered sufficiently well trained and adapted to the seismic data domain. One possible indication is that the uncertainty value remains stable regardless of whether new events are used within the trainer system. If the uncertainty remains stable when increasing the number of events, the system is considered well trained. For Mt. Etna, the uncertainty at the main eruption is similar to that of the second eruption at Bezymianny. The fact that it remains constant prior to the eruption implies that the system is sufficiently well trained for use with the Mt. Etna dataset.

1) The evolution of uncertainty is a good indicator for volcanic early warning.
2) The laws of change for this uncertainty are exportable from one system to another.
3) The form and timing of the change are related to the apparent energy in each eruptive process.
4) Comparing figure 6, figure 9.B and figure 9.C, we can confirm that the temporal behavior of the uncertainty is opposite to when the system is trained and adjusted from one volcano to a new, different one.
5) In terms of eruption pre-warning time, uncertainty is ahead of other observables, rendering it potentially more effective for volcanic alert systems. For example, at Mt. Etna, the uncertainty started to increase 16.6 hours before the eruption, that is, 200 time units before the RMSE. The acceleration of uncertainty with such an hourly anticipation constitutes the pillars to create exportable and universal early-warning systems in poorly-monitored regions, and other volcanoes world-wide. Finally, we have shown that our technique was

able to predict many hours in advance when the most energetic explosive process of this eruption was expected.

VIII. Conclusion

We present a new neural network architecture rooted in a learnable scattering transform to perform temporal modeling and introduce a multi-modular recurrent architecture to implement polyphonic detection, segmentation and classification of seismic signals with the purpose of seismo-volcanic data exploration. Our architecture demonstrates that the flexibility introduced in learning layer-wise knots and filter-bank design, jointly with specialized recurrent dynamic convolutions, yields optimal, robust features or representations in a frame-wise fashion. The application of our architecture to data from three volcanoes, of different types and on different continents, shows that our approach generalizes well and properly adapts to different environments. The non-uniform data taxonomies in seismo-volcanic applications are collapsed into a generic but well-known categorization scheme to enable the computation of invariant, robust, and universal scatter-grams. Our neural network guarantees the rapid recognition of events, and is robust against sparse data taxonomies and the presence of background noise, which is an active topic of research in machine learning applications [75].

Nonetheless, the designed recurrent scattering networks allows for seamless integration with modern seismic data workflows, and includes an online streaming data approach that can provide direct warning system statements. This requires a straightforward adjustment of the architecture, namely, changing the bidirectionality to an unidirectional frame-wise sequential recurrent network. Hence, our system is a universal framework that bridges the gap between deep learning and online monitoring; moreover, despite the myriad of complex physical mechanisms involved in volcanic unrest, our framework allows for simultaneous seismic events to be fully detected and characterized.

In addition to studying the complexities in mechanisms that drive volcano dynamics, the timing of eruptions remains an active research topic. Established forecasting algorithms are designed to find mathematical relationships involving accelerated strain, energy, or frequency variations; they assume handcrafted features that do not consider the class-membership of events and discard hidden information that might offer insight in the source mechanisms that generate seismic signals and drive volcanism. Our deep recurrent scattering network departs from the traditional perspective and opens new directions for designing forecasting methods based on

the connections between epistemic uncertainty and volcano dynamics. The power-law drift in epistemic uncertainty associated with seismic data streams implies that volcanic processes preceding eruptions are detectable. With no prior assumptions about signal distribution, deep learning can identify such behavior without supervision or parametrization by data alone. The epistemic uncertainty generated by our deep neural network holds promise for forecasting eruptions, although challenges remain. As our approach can be modified to act on real-time streaming data, this goes in concert with the development of a novel early-warning strategy.

APPENDIX A
BAYESIAN NEURAL NETWORKS

Mathematical notation: First, we establish the mathematical notation that we will follow in the formulation of Bayesian modelling for neural networks. We define our dataset (assuming we are working with seismic data) as a set of N points, $D = (\mathbf{X}, \mathbf{Y})$ where $\mathbf{X} = \{x_1, x_2, \ldots, x_N\}$ the matrix whose rows correspond to a set of seismic events (in samples) extracted from a continuous seismic record, and $\mathbf{Y} = \{y_1, y_2, \ldots, y_N\}$ is the matrix whose rows contain the corresponding labels that are assigned to every seismic event and that are categorized over a set of C classes. We refer as L to the total number of layers in the neural network, being i the sub-index of any of its layers, $i = \{1, \cdots, L\}$. Therefore, in this appendix, we specify our neural network as a function parameterized by its weights, $\mathbf{y} = f^{\omega}(\mathbf{x})$, where ω represents all weights matrices associated with the hidden layers of the neural network, that is, $\omega = \{W_i\}_{i=1}^{L}$, $\mathbf{x}$ the input feature vector of the network (i.e. a row of matrix $\mathbf{X}$) and $\mathbf{y}$ the associated polyphonic vector labels (i.e. a row of matrix $\mathbf{Y}$). In this appendix, we refer as θ to the parameters that define the approximate distribution $q_\theta(\omega)$.

Bayesian modelling in neural networks: It is well known that Bayesian methods provide a measure of uncertainty for each input and output of a given model, based on all observed data. In most of the so-called *frequentist approaches*, commonly used in neural networks, the final optimization result is a set of best-fitting parameters. Unlike frequentist methods, the result of a Bayesian fit is a probability distribution of each parameter of the model, called the *posterior distribution*. For a given set of parameters in our neural network, and the dataset D defined, the posterior $p(\omega|\mathbf{X}, \mathbf{Y})$ is determined by using the Bayes Theorem:

$$p(\omega|\mathbf{X}, \mathbf{Y}) = \frac{p(\mathbf{Y}|\mathbf{X}, \omega)p(\omega)}{p(\mathbf{Y}|\mathbf{X})}, \tag{7}$$

with $p(\mathbf{Y}|\mathbf{X},\omega)$ is defined as the *model likelihood distribution*, that is, the knowledge of the model on the data distribution, and therefore, the assignment of probabilities for each $\mathbf{X}$ and $\mathbf{Y}$, given the parameters of the model. The term $p(\omega)$ is known as *prior* and constitutes the initial, known probability distribution of the parameters of the network. Hence, in a Bayesian neural network (BNN), the prior distribution is specified as a set of probability distributions located on their weights [76]. The denominator of (7) corresponds to the *model evidence* or *marginal likelihood*, a normalizing constant that can be obtained by marginalizing the likelihood over the parameters ω:

$$p(\mathbf{Y}|\mathbf{X}) = \int p(\mathbf{Y}|\mathbf{X},\omega)p(\omega)\,\mathrm{d}\omega. \tag{8}$$

Theoretically, the marginalization in (8) involves the average with respect to all possible parameters of the model ω, weighted by $p(\omega)$. For complex models, such as BNNs, an approximation is required [77]. Defining all the terms in the numerator and denominator of (7), a BNN can predict the outputs y^* for any new input x^* through the predictive function by integration over the parameters of the network ω:

$$p(y^*|x^*,\mathbf{X},\mathbf{Y}) = \mathbb{E}_{p(\omega|\mathbf{X},\mathbf{Y})} = \int p(y^*|x^*,\omega)p(\omega|\mathbf{X},\mathbf{Y})\,\mathrm{d}\omega, \tag{9}$$

where $p(y^*|x^*,\omega)$ is the data likelihood for this new point x^*. The prediction of new seismic events for multiple instances in y^* is known as *inference*. However, the exact inference in (9) is impossible given that the posterior is part of the integral. The computation of (9) with $p(\omega|\mathbf{X},\mathbf{Y})$ is equivalent to evaluate an infinite number of neural networks with all the possible parameter configurations. This is computationally intractable for neural networks of any size. For this reason, in Bayesian modelling, an *approximate inference* procedure is required. This type of inference entails an optimization conditioned to the training of the architecture, that is, the approximation of this integral. Variational inference methods are used to approximate $p(\omega|\mathbf{X},\mathbf{Y})$ and therefore the equation (9).

Variational inference in BNNs: Variational Inference (VI) focuses on obtaining an approximation to $p(\omega|\mathbf{X},\mathbf{Y})$ by using optimization procedures [78]. Formally, this optimization aims at determining a probability density, $q_\theta(\omega)$, that should be as close as possible to $p(\omega|\mathbf{X},\mathbf{Y})$. The measure of closeness is given by the Kullback-Leibler (KL) divergence between both distributions:

$$KL(q_\theta(\omega)\;||\;p(\omega|\mathbf{X},\mathbf{Y})) = \int q_\theta(\omega)\;\log\left\{\frac{q_\theta(\omega)}{p(\omega|\mathbf{X},\mathbf{Y})}\right\}d\omega. \tag{10}$$

with $q_\theta(\omega)$ known as the *variational distribution*. We minimize the (10) by optimizing the variational parameters θ of our variational distribution $q_\theta(\omega)$:

$$\hat{\theta} = \underset{\theta}{\operatorname{argmin}} \ \mathbb{E}_{q_\theta(\omega)}[log\ q_\theta(\omega) - log\ p(\omega|\mathbf{X}, \mathbf{Y})] \tag{11}$$

$$\hat{\theta} = \underset{\theta}{\operatorname{argmin}} \ KL(q_\theta(\omega)\ ||\ p(\omega|\mathbf{X}, \mathbf{Y})) \tag{12}$$

with $\hat{\theta}$ the parameters that results in the minimum KL divergence. Once we obtained our variational approximation $q_\theta(\omega)$, and the KL in (10) has been minimized, the predictive distribution is given as:

$$p(y^*|x^*, D) \approx \int p(y^*|x^*, \omega) q_{\hat{\theta}}(\omega)\, \mathrm{d}\omega =: q_{\hat{\theta}}(y^*|x^*). \tag{13}$$

However, we can verify that the evaluation of the KL divergence in (10) requires the computation of the posterior distribution for our network's parameters, which are precisely the distribution that we want to approximate. To circumvent this, we can minimize a function similar to (10) added to a constant term. This function is known as Evidence Lower Bound (ELBO). The mathematical relationship between the ELBO and KL divergence, $KL(q_\theta(\omega)\ ||\ p(\omega|\mathbf{X}, \mathbf{Y}))$ can be derived from (10) by invoking Bayes rule and taking logarithm, yielding [78]:

$$log\, p(\mathbf{Y}|\mathbf{X}) - \underbrace{\int q_\theta(\omega)\ log \left\{ \frac{p(\mathbf{Y}|\mathbf{X}, \omega)\, p(\omega)}{q_\theta(\omega)} \right\} d\omega}_{\text{ELBO or } \mathcal{L}_{ELBO}(\theta)} \tag{14}$$

Therefore it is observable from the above equations that the KL divergence is equal to the ELBO ($\mathcal{L}_{ELBO}(\theta)$) and a constant which is given by the marginal log-likelihood of our data. Since the KL divergence is a probabilistic distance and always positive, we can thus write:

$$log\, p(\mathbf{Y}|\mathbf{X}) \geq \mathcal{L}_{ELBO}(\theta) + KL(q_\theta(\omega)\ ||\ p(\omega|\mathbf{X}, \mathbf{Y})). \tag{15}$$

with $\mathcal{L}_{ELBO}(\theta)$ becoming the objective of our optimization problem. In addition, minimizing the divergence of KL is also equivalent to maximizing ELBO with respect to the variational parameters of the distribution $q_\theta(\omega)$. We can expand the term $\mathcal{L}_{ELBO}(\theta)$, and obtain its closed-form expression:

$$\mathcal{L}_{ELBO}(\theta) := -\int q_\theta(\omega)\ log\, p(\mathbf{Y}|\mathbf{X}, \omega) d\omega\ +\ KL(q_\theta(\omega)\ ||\ p(\omega)). \tag{16}$$

The optimization of the first integral term conditions the Bayesian model to better fit our data. The second KL term acts as a regularizer, keeping $q_\theta(\omega)$ from extreme deviations of $p(\omega)$. This

analytical representation can be used to rewrite (10), $KL(q_\theta(\omega) \,||\, p(\omega|\mathbf{X}, \mathbf{Y}))$ in approximative terms for the parameters of our neural network:

$$-\sum_{n=1}^{N} \int q_\theta(\omega) \, \log \, p(y_n|f^\omega(x_n))d\omega + KL(q_\theta(\omega) \,||\, p(\omega)) \tag{17}$$

where $f^\omega(x_n)$ is the output of the neural network for a given arbitrary input x_n, and the summatory term defined as the *expected log likelihood*. Once all the parameters for variational optimization in a BNN have been established, it is necessary to choose the prior and explicitly define the variational $q_\theta(\omega)$ distribution to optimize in (17). In a BNN, $q_\theta(\omega)$ is always conditioned to the distribution given by the matrices of its neural connections. The multiple non-linearities and the evaluation of the first integral in (17) with N events of a dataset entails a prohibitive, non-scalable computation. However, we can consider the Monte Carlo sampling estimators and their connections to regularization techniques in deep neural networks. Monte Carlo estimators permit an approximation of the expected log-likelihood for neural network models with multiple hidden layers and their derivatives with respect to the variational parameters θ. The so-called Monte-Carlo dropout (MC-dropout) is thus a variational estimation that connects dropout regularization and standard neural network optimization with the inference procedure in (17) [24].

Monte Carlo dropout: The dropout technique can be used in a BNN as a Bayesian approximation of the posterior distribution of the network parameters. Initially, the dropout is formulated by [79] as a stochastic regularization technique for deep neural networks, randomly deactivating the parameters of a neural network with a given probability, p_i. A Bernoulli distribution can model this probability p_i, selecting which of the hidden units remain active in the network. The key result for this reasoning is derived by [24] and [80]: the integral and KL terms in (17) can be linked to standard dropout training in deep neural networks. This permits scalable and robust inference for large datasets in very complex networks.

Formally, our neural network is composed of a set of weight matrices in all its layers, $\omega = \{W_i\}_{i=1}^{L}$. Each weight matrix has a dimension $K_i \times K_{i-1}$. We define the variational distribution $q_\theta(\omega)$ as the factorization over the weight matrices of all the hidden layers conditioned to the dropout technique:

$$W_i = M_i \cdot \mathrm{diag}([z_{i,j}])_{j=1}^{K_i}, \tag{18}$$

$$z_{i,j} \sim \text{Bernoulli}(p_l), \quad i = 1, \ldots, L, \;\; j = 1, \ldots, K_{l-1}, \tag{19}$$

where $z_{i,j}$ represents the dropout masks (matrices of zeros and ones drawn from the Bernoulli distribution) which disable the hidden element j on layer $i-1$. The term M_i is a *mean weight matrix*, whose set $\theta = \{M_i\}_{i=1}^{L}$ are the variational parameters. Finally, having defined the variational distribution, we can use Monte-Carlo estimation to approximate the integral of the expected log-likelihood in (17):

$$-\int q_\theta(\omega)\ \log\ p(\mathbf{Y}|\mathbf{X},\omega)d\omega = \frac{1}{N}\sum_{n=1}^{N} - log\, p(y_n|f^{\hat{\omega}}(x_n)) \tag{20}$$

where $\hat{\omega}$ is not a maximum posterior estimate, but multiple realisations of random variables from the Bernoulli distribution, $\hat{\omega} \sim q_\theta(\omega)$. This reasoning is identical to applying successive dropout masks to the network weights. Hence, the averaged sum of $log\, p(y_n|f^{\hat{\omega}}(x_n))$ represents, by definition, the cost function of a neural network.

In order to link the variational inference optimization $\mathcal{L}_{ELBO}(\theta)$ to the optimization objective of standard neural networks with dropout, $\mathcal{L}_{dropout}(\theta)$, it is necessary that the mathematical relation known as KL-condition in the regularizer term is fulfilled. The KL-condition links the derivatives of the optimization objective in (10) with standard loss functions in neural networks. In this Appendix, we do not cover the entire proof in detail and refer the reader to the original work by [24], pages 150-152, Appendix A. The KL-condition establishes that the regularizer KL term in (10) can be approximated as a standard dropout regularizer weighted by a normalization constant λ. Our objective of variational minimization is defined as:

$$\mathcal{L}_{dropout}(\theta) = \frac{1}{N}\sum_{n=1}^{N} - log\, p(y_n|f^{\hat{\omega}}(x_n)) + \lambda \sum_{l=1}^{L}(\|M_l\|_2^2 + \|b_l\|_2^2) \tag{21}$$

Therefore, approximate inference procedures result in an optimization goal identical to that of a neural network using the loss function $\mathcal{L}_{dropout}(\theta)$. This function is defined to optimize the parameters of the neural network and find the best $q_{\hat{\theta}}$ that minimizes the KL divergence, $KL(q_\theta(\omega)\ ||\ p(\omega|\mathbf{X},\mathbf{Y}))$ in equation (17). Finally, we can use the approximation learned by our network to evaluate the predictive function in (13), using Monte-Carlo sampling with T sampling steps:

$$q_{\hat{\theta}}(y^*|x^*) = \int p(y^*|f^{\hat{\omega}}(x^*))q_{\hat{\theta}}(\omega)d\omega \approx \frac{1}{T}\sum_{t=1}^{T} p(y^*|f^{\hat{\omega}_t}(x^*)) \tag{22}$$

or equivalently $\hat{\omega}_t \sim q_\theta(\omega)$. Therefore, at the time of inference, the dropout layers are applied to the M_i matrices, generating a Monte-Carlo sample from the posterior distribution (see equation 9). In practice, the average of these samples can be interpreted as the prediction of the network, although a single estimate is not obtained, as many as T sampling steps are performed. We can

use the probabilities obtained by MC-dropout to estimate the uncertainty in the application of seismo-volcanic recognition.

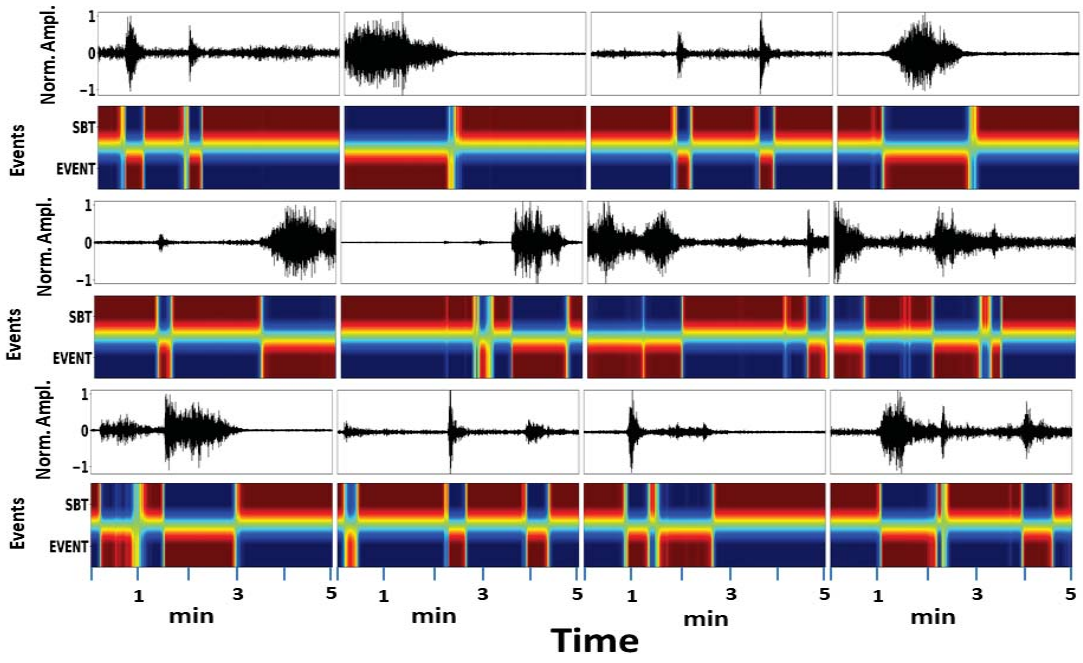


Fig. 10. **Pre-eruptive event detection heat maps**. Probabilistic heat maps for the first eruptive sequence of Bezymianny volcano. Data were obtain on 25 September 2007.

APPENDIX B
HYPERPARAMETERS OF THE NETWORK

Our neural network architecture contains two paths or branches. The first, named path A, analyzes via dynamic convolutions a sequence of local scattering coefficient variations, outputting a refined feature map. The second, path B, is a bypass skip connection with a learnable convolution to perform data reduction and per-frame time analysis to capture precise onsets of events. The training is done, for all the data sets, with the raw seismic wave forms.

The number of knots k, the number of octaves J and wavelets per octave Q are carefully chosen after some signal analysis of the input waveforms. We selected, for the *first scattering layer L1*, $k = 8$, $J = 6$ and $Q = 7$, and for the *second scattering layer L2*, $k = 2$, $J = 5$ and $Q = 5$. These parameters give enough bandwidth and frequency resolution to provide a scattering representation where the main frequencies of the events are discernible from the background tremor. After the

concatenation of both scattering feature vectors, a volumetric pooling of size $(1, 1, 100)$ is applied to remove redundant information. Both $ConvLSTM_\lambda$s contain 16 filters, with a kernel size of $(5, 5)$ and strides $(1, 1)$. The temporal skip connection is a single filter with a kernel size of $(1, 1)$ to perform dimensionality reduction. The successive BiLSTMs layers are composed of 92 hidden units in each direction (128 in each BiLSTM). A per-class *sigmoid* activation function is applied to compute the final detection matrix: 5 hidden units for Bezymianny and Mount Saint Helens. The scattering transform module produces a sparse vector of 30000 samples pooled by 100 to achieve a 1s time resolution. Layer normalization is applied after all successive recurrent layers. The dense, predictive layer contains a sigmoid activation function in which a binarization threshold of 0.5 over the per-class event activity probabilities is applied to produce the event detection matrix allowing multiple class memberships to be detected. We adopt the polyphonic training framework in [42]; with binary cross-entropy loss function. For the transfer learning setup in subsection VII-I, we use the same hyper-parameter configurations but changing the per-class *sigmoid* activation function to 4 hidden units corresponding to SBT, HF, MX and LF classes.

APPENDIX C
EVENT DETECTION

Seismo-volcanic event detection is a monophonic task. Our neural network is trained to detect and segment the presence or absence of events, where the classes become event and no event. We use the same hyper-parameter that those on the polyphonic class, except that we change the number of hidden units in the sigmoid class to two, event/no event. Table II contains the attained metrics for each learnable configuration on the test data set associated with the Bezymianny eruption of 25 September 2007. The results improve on those obtained using a polyphonic setting. Implicitly, the neural network becomes more sensitive to the onset of events as well as to the duration. It can thus be applied to seismology as an alternative to traditional algorithms [69], feature-based algorithms [81], or data-mining based pipelines [82]. Traditional STA/LTA methods require some prior knowledge about the signals to which they are applied, which is avoided by using a learning-based approach. Feature-based algorithms do improve the segmentation boundaries in the time of events, but require prior knowledge about the frequency range in which events are distinguished from noise. An algorithm based on fingerprint similarity, where fingerprints are generated through a network analogue from signals using data mining [82],

TABLE II

SEISMIC EVENT DETECTION PERFORMANCE ON THE ERUPTION ON SEPTEMBER 25TH

System	F1 (1s)	PR (1s)	RC (1s)
Vanilla	88.17	88.16	88.18
Filters	88.11	88.12	88.09
Knots + Filters	88.18	88.18	88.19

relies on the assumption that all events are sufficiently frequently represented in the data set. Our neural network for event detection can be pre-trained and is exportable so that it can be applied ubiquitously to many volcanic systems, regardless of eruptive style. Results on test data from Bezymianny are illustrated in Figure S1. The pre-eruptive event detection heat maps show that our neural network can systematically detect and segment seismo-volcanic events even in the presence of significant background noise.

ACKNOWLEDGMENT

The facilities of IRIS Data Services, and specifically the IRIS Data Management Center, were used for access to the waveforms, related metadata, and/or derived products used in this study. IRIS Data Services were funded through a Seismological Facilities for the Advancement of Geoscience (SAGE) Award of the National Science Foundation under Cooperative Support Agreement EAR-1851048.

The Mt. Etna data were collected within the framework of the project: Volcanic Emissions Analysis through Seismic and Infrasound Advanced monitoring (VOSSIA), supported by the Trans-National Access component of the EUROVOLC project (European Network of Observatories and Research Infrastructures for Volcanology, EU Horizon 2020 Research Infrastructure Project grant No 731070).

REFERENCES

[1] Mie Ichihara. “Seismic and infrasonic eruption tremors and their relation to magma discharge rate: A case study for sub-Plinian events in the 2011 eruption of Shinmoe-dake, Japan”. In: *Journal of Geophysical Research: Solid Earth* 121.10 (2016), pp. 7101–7118.

[2] Azusa Mori and Hiroyuki Kumagai. “Estimating plume heights of explosive eruptions using high-frequency seismic amplitudes”. In: *Geophysical Journal International* 219.2 (2019), pp. 1365–1376.

[3] Florent Brenguier, Nikolai M Shapiro, Michel Campillo, Valérie Ferrazzini, Zacharie Duputel, Olivier Coutant, and Alexandre Nercessian. “Towards forecasting volcanic eruptions using seismic noise”. In: *Nature Geoscience* 1.2 (2008), pp. 126–130.

[4] Clara Gómez-García, F Brenguier, P Boué, NM Shapiro, DV Droznin, SYa Droznina, SL Senyukov, and EI Gordeev. “Retrieving robust noise-based seismic velocity changes from sparse data sets: synthetic tests and application to Klyuchevskoy volcanic group (Kamchatka)”. In: *Geophysical Journal International* 214.2 (2018), pp. 1218–1236.

[5] Tomoya Takano, Florent Brenguier, Michel Campillo, Aline Peltier, and Takeshi Nishimura. “Noise-based passive ballistic wave seismic monitoring on an active volcano”. In: *Geophysical Journal International* 220.1 (2020), pp. 501–507.

[6] Bernard Chouet. “Volcano seismology”. In: *Pure and applied geophysics* 160.3 (2003), pp. 739–788.

[7] RSJ Sparks, J Biggs, and JW Neuberg. “Monitoring volcanoes”. In: *Science* 335.6074 (2012), pp. 1310–1311.

[8] S.R. McNutt, G. Thompson, J. Johnson, Silvio S. De Angelis, and D. Fee. “Seismic and infrasonic monitoring”. In: *The Encyclopedia of Volcanoes (Second Edition)*. Elsevier, 2015, pp. 1071–1099.

[9] Jesús M Ibáñez, Edoardo Del Pezzo, Javier Almendros, Mario La Rocca, Gerardo Alguacil, Ramón Ortiz, and Alicia García. “Seismovolcanic signals at Deception Island volcano, Antarctica: Wave field analysis and source modeling”. In: *Journal of Geophysical Research: Solid Earth* 105.B6 (2000), pp. 13905–13931.

[10] M Palo, JM Ibáñez, M Cisneros, M Bretón, E Del Pezzo, E Ocana, J Orozco-Rojas, and AM Posadas. “Analysis of the seismic wavefield properties of volcanic explosions at Volcan de Colima, Mexico: insights into the source mechanism”. In: *Geophysical Journal International* 177.3 (2009), pp. 1383–1398.

[11] Vyacheslav M Zobin. *Introduction to volcanic seismology*. Vol. 6. Elsevier, 2012.

[12] Kostas I Konstantinou. “Tornillos modeled as self-oscillations of fluid filling a cavity: Application to the 1992–1993 activity at Galeras volcano, Colombia”. In: *Physics of the Earth and Planetary Interiors* 238 (2015), pp. 23–33.

[13] Robin S Matoza, Alejandra Arciniega-Ceballos, Richard W Sanderson, Gerardo Mendo-Pérez, Alejandro Rosado-Fuentes, and Bernard A Chouet. “High-Broadband Seismoacoustic Signature of Vulcanian Explosions at Popocatépetl Volcano, Mexico”. In: *Geophysical Research Letters* 46.1 (2019), pp. 148–157.

[14] Gilberto Saccorotti and Ivan Lokmer. “A review of seismic methods for monitoring and understanding active volcanoes”. In: *Forecasting and Planning for Volcanic Hazards, Risks, and Disasters* (2020), pp. 25–73.

[15] G. Cortés, R. Carniel, M. Ángeles Mendoza, and P. Lesage. “Standardization of Noisy Volcanoseismic Waveforms as a Key Step toward Station-Independent, Robust Automatic Recognition”. In: *Seismological Research Letters* 90.2A (Jan. 2019), pp. 581–590. ISSN: 0895-0695. DOI: 10.1785/0220180334. eprint: https://pubs.geoscienceworld.org/srl/article-pdf/90/2A/581/4655551/srl-2018334.1.pdf. URL: https://doi.org/10.1785/0220180334.

[16] M. Malfante, M. Dalla Mura, J. Metaxian, J. I. Mars, O. Macedo, and A. Inza. “Machine Learning for Volcano-Seismic Signals: Challenges and Perspectives”. In: *IEEE Signal Processing Magazine* 35.2 (2018), pp. 20–30. ISSN: 1053-5888.

[17] M. Titos, A. Bueno, L. García, M. C. Benítez, and J. M. Ibáñez. “Detection and Classification of Continuous Volcano-Seismic Signals With Recurrent Neural Networks”. In: *IEEE Transactions on Geoscience and Remote Sensing* (2018), pp. 1–13. ISSN: 0196-2892. DOI: 10.1109/TGRS.2018.2870202.

[18] Muhammad Salman Khan, Millaray Curilem, Fernando Huenupan, Muhammad Farhan Khan, and Nestor Becerra Yoma. “A Signal Processing Perspective of Monitoring Active Volcanoes [Applications Corner]”. In: *IEEE Signal Processing Magazine* 36.6 (2019), pp. 125–163.

[19] Noel Perez, Pablo Venegas, Diego Benítez, Román Lara-Cueva, and Mario Ruiz. “A new volcanic seismic signal descriptor and its application to a data set from the cotopaxi volcano”. In: *IEEE Transactions on Geoscience and Remote Sensing* 58.9 (2020), pp. 6493–6503.

[20] John C Lahr, Bernard A Chouet, Christopher D Stephens, John A Power, and Robert A Page. “Earthquake classification, location, and error analysis in a volcanic environment: Implications for the magmatic system of the 1989–1990 eruptions at Redoubt Volcano, Alaska”. In: *Journal of Volcanology and Geothermal Research* 62.1-4 (1994), pp. 137–151.

[21] Benjamin A Heath, Emilie EE Hooft, and Douglas R Toomey. “Autocorrelation of the seismic wavefield at Newberry Volcano: Reflections from the magmatic and geothermal systems”. In: *Geophysical Research Letters* 45.5 (2018), pp. 2311–2318.

[22] Haruhisa Nakamichi, Masato Iguchi, Hetty Triastuty, Muhamad Hendrasto, and Iyan Mulyana. “Differences of precursory seismic energy release for the 2007 effusive dome-forming and 2014 Plinian eruptions at Kelud volcano, Indonesia”. In: *Journal of Volcanology and Geothermal Research* 382 (2019), pp. 68–80.

[23] Andrea Bevilacqua, Marcus Bursik, Abani Patra, E Bruce Pitman, and Ryan Till. “Bayesian construction of a long-term vent opening probability map in the Long Valley volcanic region (CA, USA)”. In: *Statistics in Volcanology* 3 (Apr. 2017). DOI: 10.5038/2163-338X.3.1.

[24] Yarin Gal and Zoubin Ghahramani. “Dropout as a Bayesian approximation: Representing model uncertainty in deep learning”. In: *international conference on machine learning*. 2016, pp. 1050–1059.

[25] S Mostafa Mousavi and Gregory C Beroza. “Bayesian-Deep-Learning Estimation of Earthquake Location From Single-Station Observations”. In: *IEEE Transactions on Geoscience and Remote Sensing* (2020).

[26] M. C. Benitez, J. Ramirez, C. Segura, J.M. Ibanez, J. Almendros, A. Garcia-Yeguas, and G. Cortes. “Continuous HMM-based seismic-event classification at Deception Island, Antarctica”. In: *IEEE Transactions on Geoscience and remote sensing* 45.1 (2006), pp. 138–146.

[27] Joan Bruna and Stéphane Mallat. “Invariant scattering convolution networks”. In: *IEEE transactions on pattern analysis and machine intelligence* 35.8 (2013), pp. 1872–1886.

[28] Y. Zhang, W. Chan, and N. Jaitly. “Very deep convolutional networks for end-to-end speech recognition”. In: *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 2017, pp. 4845–4849.

[29] S. Xingjian, Z. Chen, H. Wang, DT. Yeung, W-K Wong, and W c Woo. “Convolutional LSTM network: A machine learning approach for precipitation nowcasting”. In: *Advances in neural information processing systems*. 2015, pp. 802–810.

[30] Alex Graves and Jürgen Schmidhuber. “Framewise phoneme classification with bidirectional LSTM networks”. In: *Neural Networks, 2005. IJCNN’05. Proceedings. 2005 IEEE International Joint Conference on*. Vol. 4. IEEE. 2005, pp. 2047–2052.

[31] K.J. Bergen, T. Chen, and Z. Li. “Preface to the Focus Section on Machine Learning in Seismology”. In: *Seismological Research Letters* 90.2A (2019), pp. 477–480.

[32] K.J. Bergen, P.A. Johnson, M.V. de Hoop, and G. Beroza. “Machine learning for data-driven discovery in solid Earth geoscience”. In: *Science* 363.6433 (2019). ISSN: 0036-8075. DOI: 10.1126/science.aau0323. eprint: https://science.sciencemag.org/content/363/6433/eaau0323.full.pdf. URL: https://science.sciencemag.org/content/363/6433/eaau0323.

[33] D.C. Bolton, P. Shokouhia, Bertrand Rouet-Leduc, Claudia Hulbert, Jacques Rivière, Chris Marone, and Paul A Johnson. “Characterizing acoustic signals and searching for precursors during the laboratory seismic cycle using unsupervised machine learning”. In: *Seismological Research Letters* 90.3 (2019), pp. 1088–1098.

[34] Y. Wu, Y. Lin, Z. Zhou, D.C. Bolton, J.Liu, and P.A Johnson. “DeepDetect: A cascaded region-based densely connected network for seismic event detection”. In: *IEEE Transactions on Geoscience and Remote Sensing* 57.1 (2018), pp. 62–75.

[35] Guillermo Cortés, Luz García, Isaac Álvarez, Carmen Benítez, Ángel de la Torre, and Jesús Ibáñez. “Parallel System Architecture (PSA): An efficient approach for automatic recognition of volcano-seismic events”. In: *Journal of Volcanology*

and Geothermal Research 271 (2014), pp. 1 –10. ISSN: 0377-0273. DOI: https://doi.org/10.1016/j.jvolgeores.2013.07.004. URL: http://www.sciencedirect.com/science/article/pii/S0377027313002229.

[36] Christopher X Ren, Aline Peltier, Valerie Ferrazzini, Bertrand Rouet-Leduc, Paul Allan Johnson, and Florent Brenguier. "Machine learning reveals the seismic signature of eruptive behavior at piton de la fournaise volcano". In: *Geophysical Research Letters* 47.3 (2020), e2019GL085523.

[37] Grace F Manley, David M Pyle, Tamsin A Mather, Mel Rodgers, David A Clifton, Benjamin G Stokell, Glenn Thompson, John Makario Londoño, and Diana C Roman. "Understanding the timing of eruption end using a machine learning approach to classification of seismic time series". In: *Journal of Volcanology and Geothermal Research* (2020), p. 106917.

[38] T. Sainath, R.J. Weiss, K.W. Wilson, A. Narayanan, and M. Bacchiani. "Factored spatial and spectral multichannel raw waveform CLDNNs". In: *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE. 2016, pp. 5075–5079.

[39] Neil Zeghidour, Nicolas Usunier, Iasonas Kokkinos, Thomas Schaiz, Gabriel Synnaeve, and Emmanuel Dupoux. "Learning filterbanks from raw speech for phone recognition". In: *ICASSP*. IEEE. 2018, pp. 5509–5513.

[40] Mirco Ravanelli and Yoshua Bengio. "Speaker recognition from raw waveform with sincnet". In: *2018 IEEE Spoken Language Technology Workshop (SLT)*. IEEE. 2018, pp. 1021–1028.

[41] Florent Jaillet and Bruno Torrésani. "Time-frequency jigsaw puzzle: Adaptive multiwindow and multilayered Gabor expansions". In: *Int. J. of Wavelets, Multiresolution and Inf. Proc.* 5.02 (2007), pp. 293–315.

[42] E. Cakir, E. C. Ozan, and T. Virtanen. "Filterbank learning for deep neural network based polyphonic sound event detection". In: *2016 International Joint Conference on Neural Networks (IJCNN)* (2016), pp. 3399–3406.

[43] Haidar Khan and Bulent Yener. "Learning filter widths of spectral decompositions with wavelets". In: *Advances in Neural Inf. Proc. Sys. 31*. Ed. by S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett. 2018, pp. 4601–4612.

[44] Mirco Ravanelli and Yoshua Bengio. "Interpretable convolutional filters with sincnet". In: *arXiv preprint arXiv:1811.09725* (2018).

[45] Soo-Chang Pei and Shih-Gu Huang. "STFT with adaptive window width based on the chirp rate". In: *IEEE Transactions on Signal Processing* 60.8 (2012), pp. 4065–4080.

[46] Randall Balestriero, Herve Glotin, and Richard Baraniuk. "Interpretable and Learnable Super-Resolution Time-Frequency Representation". In: *Proceedings of Machine Learning Research*. Ed. by Joan Bruna, Jan Hesthaven, and Lenka Zdeborova. Vol. 107. Conference on Mathematical and Scientific Machine Learning. JMLR, 2021, pp. 1–25.

[47] W Ye, J Cheng, F Yang, and Y Xu. "Two-Stream Convolutional Network for Improving Activity Recognition Using Convolutional Long Short-Term Memory Networks". In: *IEEE Access* 7 (2019), pp. 67772–67780.

[48] A. Chattopadhyay, P. Hassanzadeh, and S. Pasha. "Predicting clustered weather patterns: A test case for applications of convolutional neural networks to spatio-temporal climate data". In: *Scientific Reports* 10.1 (2020), pp. 1–13.

[49] Z. Liang, Z. Guangming, M. Lin, S. Peiyi, S. Syed, and B. Mohammed. "Attention in Convolutional LSTM for Gesture Recognition". In: *NIPS* (2018).

[50] Annamaria Mesaros, Toni Heittola, and Tuomas Virtanen. "TUT Database for Acoustic Scene Classification and Sound Event Detection". In: *24th European Signal Processing Conference 2016 (EUSIPCO 2016)*. Budapest, Hungary, 2016.

[51] Joakim Andén and Stéphane Mallat. "Deep scattering spectrum". In: *IEEE Transactions on Signal Processing* 62.16 (2014), pp. 4114–4128.

[52] Stéphane Mallat. "Understanding deep convolutional networks". In: *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 374.2065 (2016), p. 20150203.

7 | Continuous active learning for seismo-volcanic monitoring

This chapter introduce the active learning framework with B-TCNs to significantly reduce annotation time, speed up algorithmic training, and boost monitoring adaptability to unforeseen situations; that is, when the change has been detected. This article is in press at IEEE Geoscience and Remote Sensing Letters, with the following journal metrics: (IF), JCR 2020: (3.833). Remote Sensing (Rank 10/30) (Q2). Geochemistry and geophysics (Rank 13/85) (Q1). Imaging Science and photographic technology (Rank 5/27) (Q2). We reproduce the draft in press by the journal, which can be cited as:

1. **A. Bueno**, M. Titos, C. Benitez and J.M. Ibánez. Continuous Active Learning for Seismo-Volcanic Monitoring. IEEE Geoscience and Remote Sensing Letters (in press).

Continuous Active Learning for Seismo-Volcanic Monitoring

Angel Bueno, Manuel Titos, Carmen Benítez, and Jesús M. Ibáñez

Abstract

The advancements in Deep Learning have boosted the field of volcano-seismology to unprecedented levels. Nevertheless, curated data catalogs still require substantial annotation efforts, often delayed in time due to the ever changing seismic data conditions. The selective segmentation of which earthquake transients has to be reviewed by an expert can significantly reduce annotation time, speed up algorithmic training, and boost monitoring adaptability to unforeseen situations. In this work, we propose a Bayesian temporal convolutional neural network (B-TCN) to perform continuous detection and classification while extracting the most uncertain events from the continuous data stream. Formulated as an active learning (AL) procedure, our B-TCN outputs an uncertainty map over time, highlighting the class memberships that need to be reviewed. We attain a significant improvement in monitoring metrics, with only a fraction of the initial dataset to achieve a performance of 83% for five seismo-volcanic events.

I. Introduction

Seismo-volcanic monitoring involves the individual identification of earthquake transients in the continuous data stream, [1]. The typical monitoring framework comprises an algorithm that classifies, supervised or unsupervised, seismic transients into a specific category. The advent of deep learning has made neural networks the preferred choice of the seismological community to perform such tasks in both isolated waveform classification, and continuous seismic recognition [2]. Well-established deep learning algorithms work well under the assumption that training and

A. Bueno and C.Benítez are with the Department of Signal Theory, Telematic and Communications, University of Granada, 18071, Spain. e-mail: angelbueno@ugr.es.

M. Titos is with the Icelandic Met Office, Division of Processing and Research, 150 Reykjavik, Iceland.

J. M. Ibáñez is with the Andalusian Institute of Geophysics and Department of Theoretical Physics and Cosmos, 18071, University of Granada, Spain

This work is supported by TEC2015-68752 (KNOWAVES) and PID2019-106260GB-I00 (FEMALE)

test data follow the same distribution. Nevertheless, modelling the radiated seismic wavefield harbor a set of complex factors, such as data challenges (non-stationary environment changing conditions) [3] and geophysical constraints (underlying physical factors that influence recordings) [4]. Besides, seismic waves are affected by data shifts due to seismic changes in the volcanic settings, hindering monitoring capabilities that sometimes require a complete re-training of the deployed system [5].

The procedural approach to recover monitoring systems from severe data shifts is often based on the manual selection of events followed by a *fine-tuning* of the monitoring algorithm. However, in a period of seismic changes, the severe alterations of the monitored variables can lead to situations where there are not enough data samples for most classes of interest. Thus, a practical problem is the optimization of a monitoring system with fewer annotation efforts. Recent approximations to bridge performance degradation and annotation efforts focuses on semi-supervised categorization [6] [7], transfer learning (TL) procedures [3] [8], novel feature descriptors [9], [10] or feature selection [11]. TL approaches have proven useful in reducing training time, but they require substantial annotation efforts to provide a meaningful dataset to perform transfer learning procedures. In this regard, removing redundant information by selecting the optimum set of significant new seismic signals could appreciably reduce annotation efforts and increase monitoring adaptability.

In this work, we propose an active learning (AL) procedure based on temporal convolutional neural networks, TCN, to systematically select the most informative seismic samples to maximize the performance, even if limited training data is available. AL for seismo-volcanic monitoring has not been extensively studied. The main hindrance in the applicability of AL from continuous streams is rooted in the reliable extraction and selection of relevant events within the continuous data stream. To solve this, we exploit the Bayesian dropout approximation, previously applied in seismic monitoring [5] [12], to generate an uncertainty representation to segment and select those events that need to be queried for re-training. We test the capabilities of our model on a dataset from Bezymianny volcano (Russia), recorded in 2007 by a near-field seismic station (BELO). An AL scenario is simulated by only using a small portion of available annotations. The amount of waveforms used for training the TCN is increased based on the decisions of the proposed AL approach as if they were generated online by human annotators.

This paper is organized as follows: Section 2 introduces seismic sequence modeling and the TCN model. Section 3 presents the active AL framework. Section 4 summarizes the datasets

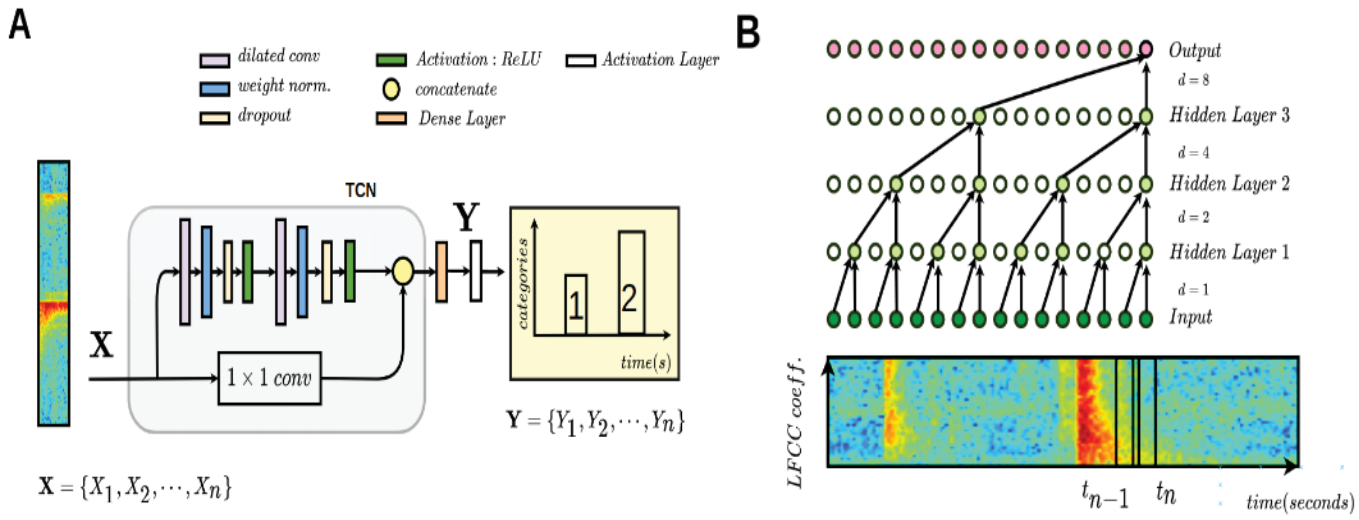


Fig. 1. **A**: Implemented TCN architecture consisting of two stacks of dilated convolutions with weight normalization and dropout layers. A skip connection of a 1×1 convolution is added for gradient stability. **B**: Dilated causal convolutions operator applied over the input feature vector of our network, at time t_n. Each time bin is convoluted by the successive hidden layers, at different dilation rates $[1, 2, 4, 8]$, with a pyramidal increase of the number of analyzed frames. The black lines correspond to the causal convolutions, in which it is noticeable that no future information leakage is allowed to the model.

studied. In section 5, we introduce the experimental methodology and the obtained results. Finally, section 6 closes the study.

II. Seismic Sequence Modeling

Seismic monitoring is cast as an automatic sequence recognition problem. Let it be our data domain $\boldsymbol{D} = (\boldsymbol{X}, \boldsymbol{Y})$, with $\boldsymbol{X}$ the dataset of n continuous seismic streams over time, with $\boldsymbol{X} = (X_1, X_2, ..., X_n)$ the input feature vectors from the seismic signal, and $\boldsymbol{Y} = (Y_1, Y_2, ..., Y_n)$ the corresponding sequence of labels. In monitoring tasks, an algorithm is trained to find the temporal relationships between sequences within (X_n, Y_n). TCNs offer a flexible, robust, and scalable approach for the automatic recognition of earthquake transients [5]. The temporal modeling prowess of TCNs, rooted in the formulation of neural convolutions, permits the analysis of longer input sequences in an online fashion. The mathematical formulation of TCNs builds upon the dilated convolution operator F, parametrized by a *dilation factor* d, which supports larger receptive fields to reach broader temporal ranges. Figure 1.A depicts the proposed TCN architecture with the stacks of convolutions operators and additional layers. The TCN can thus ingest seismic data sequentially, with integrated causal constraints and residual connections

to magnify seismo-volcanic long-term dependencies. The third component in this architecture, *residual connections*, acts as a rectified shortcut that keeps learning stable.

A. TCN convolutions

The convolutions in a TCN facilitate online sequence recognition; *causal* property bounds the prediction at time t to past frames, and the dilated convolution operator enables larger receptive fields for long-term memory. Figure 1.B portrays these operations over the input feature vectors of a seismo-volcanic stream. Hence, given the input sequence X_t, the dilated convolution operator F at time t is defined as:

$$F(t) = \sum_{i=0}^{k-1} f(i) X_{t-d \cdot i} \tag{1}$$

with $f(i)$ the i^{nth} filter in layer i, k the filter size and d the dilation factor. Note that each filter $f(i)$ operates over $t - d \cdot i$ frames of the seismic signal, with d controlling the number of frames that can be accessed by the i^{nth} filter for contextual information. Remark that if $d = 1$ in equation 1, $F(t)$ corresponds to standard convolutions. The pyramidal configuration of the dilation factor permits the TCN to cover broader input ranges, exponentially increasing the number of analyzed frames from the first to the predictive output layer.

B. Residual connections

Retaining long-term temporal information of seismic sequences is essential to learn the temporal correlation between distant events. Yet, gradient stability in longer sequences can be compromised as the numerical values of the gradient decrease over time, a problem known as vanishing gradient [13]. In this regard, TCNs incorporate *residual connections* to solve the vanishing gradient problem. Formally defined in [14], a residual connection aggregates the input X_n to the convoluted feature representation $\phi(X_n)$:

$$o = X_n + \phi(X_n) \tag{2}$$

However, if X_n and $\phi(X_n)$ do have different size, the addition cannot be done. To ensure same output o, the TCN includes a 1×1 convolution kernel on the residual connection, rectifying any variable length by squeezing the input X_n to a specific shape.

III. Active learning and monitoring

The applicability of AL in continuous seismo-volcanic monitoring presents several challenges. First, the monitoring algorithm needs to perform reliable and quasi-real-time event detection and segmentation, besides data scarcity conditions [3]. The definition of AL has to embrace a Bayesian perspective so that the uncertainty from temporal sequence modeling can be maximally informative of seismic changes and re-training requirements. AL encompasses the definition of a quintet, (M, Q, O, L, U), with M the classifier model trained on the initial labeled dataset L. The component Q or *acquisition function* selects (or query) the meaningful samples from the unlabeled data pool, U. For queried samples, the oracle O assigns the correct labels. This procedure is repeated until the exhaustion of U or to fulfill operational requirements *stopping-criterion*.

Hence, the classifier M and the querying function Q are two key components that need to be carefully defined in seismo-volcanic applications. We propose to incorporate the implemented TCN in an AL framework as a Bayesian model. The uncertainty of the model is then exploited to select the seismic events that need to be revisited. In the dynamic of the volcanic unrest, it is known that the uncertainty grows with time due to data drift [3] [5]. Hence, from an AL perspective, the initial support distribution is continuously shifted towards the newly acquired data. Hence, the model can *forget* the previous learning, and thus, estimate high uncertainty from initial, already revised conditions. To avoid alternate sampling across time-span, we incorporate a *cost-effective active learning* (CEAL) strategy: the model does not learn only those uncertain data samples, but very certain points are also concatenated as pseudo-labels to maintain a uniform sampling data distribution [15].

A. Bayesian TCN

We adopt the Bayesian seismic monitoring framework described in [5], which exploits the predictive uncertainty to detect change in a volcano. Formally, we define ω as the prior probability distributions, $\omega = \{\boldsymbol{W}_l\}_{l=1}^{L}$; with ω the weights matrices at each layer l. In a BNN, one aims to find an approximation to the posterior distribution of the network weights, $p(\omega|D)$ [16]. With this posterior, we can define the predictive distribution for a new pair of data stream samples, (X^*, Y^*), as:

$$p(Y^*|X^*, D) = \int p(Y^*|X^*, w)\; p(\omega|D)\; d\omega \tag{3}$$

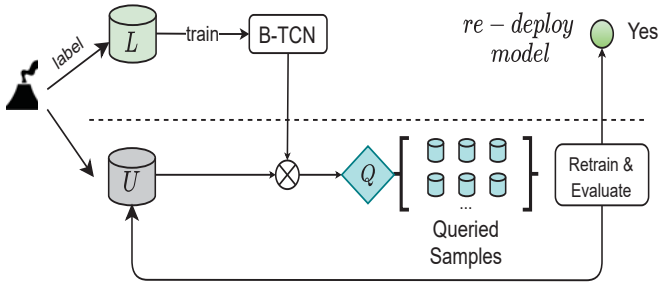


Fig. 2. **Proposed AL framework**: the initial monitored volcanic state, s_0, suddenly morph into a new monitoring condition s_1 with a geo-induced seismic data drift. The new unlabelled data, U is evaluated by the trained TCN *uncertainty-sampling* via the Q *acquisition function* by selecting a batch of maximally informative samples to retrain the model. This process, known as *acquisition step*, is repeated until fulfillment of the *stopping criteria*.

Yet, estimating $p(\omega|X,Y)$ is a very challenging task, in which few approximations exists. MC-dropout is equivalent to a variational inference approximation in which a probabilistic function $q_\theta(\omega)$, parameterized by θ is used to approximate the Kullback-Leibler (KL) divergence to the neural network posterior $p(\omega|X,Y)$ [16]. The uncertainty in the weights yields predictive uncertainty by marginalizing over the approximate posterior,

$$p(Y^*|X^*,D) \approx \int p(Y^*|X^*,\omega)\; q_\theta(\omega) d\omega \approx \frac{1}{K}\sum_{k=1}^{K} p(Y^*|X^*,\bar{\omega}_k) \tag{4}$$

with K *stochastic forward* sampling steps at test time. Hence, MC-dropout generate a range of predictions which can be used to select which earthquakes are necessary for further analysis and system re-training.

B. Acquisition Functions

The models in AL are based on a canonical set of Q functions designed to evaluate the performance of the classifier M. This is done by defining multiple venues of how the uncertainty is exploited as an acquisition function [16]:

1) **Random sampling:** Random sampling is equivalent to perform a random choices of a seismic waveform from the unlabeled set U at any given monitoring time.
2) **Predictive Entropy:** Rooted in information theory measures, it quantifies the information within the predictive distribution, sampling those events that maximize the entropy of the predictions.

$$\mathbb{H}[y|x, D] = -\sum p(y|x, D) \log p(y|x, D) \tag{5}$$

3) **Bayesian active learning by disagreement (BALD):** This functions samples seismo-volcanic events that are maximally informative with respect to predictive information and *expected* model posterior, with probabilities assigned high variance across multiple seismic classes.

$$\mathbb{I}[y, \omega|x, D] = \mathbb{H}[y|x, D] - \mathbb{E}_{p(\omega|D)}[\mathbb{H}[y|x, D] \tag{6}$$

Figure 2 summarizes the AL strategy implemented in this work. When unforeseen data shifts occurs in a volcano, a new unlabelled data pool streaming U from the new distribution is presented to the algorithm, M. The already presented query functions Q, batch a set of N samples forwarded to the oracle, O to assign a label. We keep the same notation as in descriptive AL frameworks. However, the oracle, O in our AL pipeline has been substituted by expert elicitation, where the best label has been selected and validated.

IV. VOLCANO-SEISMIC DATASET

Volcanic sources are highly variable, unpredictable and nearly chaotic, and thus, there is not a uniform criterion in the classification of seismic signal. To circumvent this problem, other classification taxonomies advocate for a rapid categorization scheme based on frequency properties instead of source-based event categorization, [1]. Nonetheless, the frequency categorization obliquely shares the type of fracture and the presence or not of fluid interaction. This study focuses on Bezymianny volcano, located on the Kamchatka Peninsula (Russia). We focus this study to the closest station to the eruptive center (BELO, Z component), on the period from 17th October to 29th October, 2007, that contains 288 hours with the following composition: 3698 high frequency (HF), 2383 low frequency (LF), 2979 debris processes (DP) and 333 low frequency tremor (LFT) [17]. In summary, these labels capture the common interactions recorded in volcano monitoring [1]:

1) **High frequency events (HF)**: Broad frequency range from 1 to up to 30 Hz. These events are the results of fracturing processes related to stress propagation in the volcanic edifice.
2) **Low frequency events (LF)**: Their energy is concentrated in the [1-5] Hz band. These events are associated with fluid-driven processes in the volcanic plumbing system. In

TABLE I

AL HYPER-PARAMETER SETUP

Data partition	**AL config.**	**Windowed waveforms**
X_{seed}	Labelled Set (L)	*31*
	Un-labelled Set (U)	*11969*
X_{test}	Test set	*10147*

Bezymianny volcano, the tremor has been renamed as low-frequency tremor (LFT), characterized as sustained signal (several minutes to hours) containing a swarm of LF events in a time-span of seconds.

3) **Debris process (DP)**: This category gathers seismic events with very high spectra and large duration, often associated with rockfalls, avalanches and other external processes that offer insights about volcanic unrest.

We refer the reader to [1] for a detailed geophysical explanation of seismic sources and [3] for a complete summary of seismic categories and potential associated mechanisms.

V. EXPERIMENTS AND DISCUSSION

A. Training details

The raw signal is preprocessed to generate an input feature representation for the TCN model (see Figure 1.A). Each day of the dataset is filtered in the frequency range of $[0.5 - 13.5]\ Hz$, and then divided into five minutes frames with 50% overlapping windows. We then partion the data into two chronological subsets: one for baseline and testing purposes, X_{seed}, with 12000 waveforms of five minutes each; and a held-out test, X_{test}, of 10147 waveforms, that corresponds to the last 28 hours of the available seismic period. The model accuracy on the held-out test set is recorded, and remains the same for all the baseline experiments and the active learning procedure.

We then compute, for each of the seismic windows in X_{seed} and X_{test}, the log-energy frequency-band features as the input of the TCN. Following [3], neural network training is performed on the full dataset X_{seed}, using Adam optimizer, with a 75%-25% data split over three k-fold splits. The learning rate linearly halves every five epochs, starting at 0.001; with early stopping added every ten epochs to avoid over-fitting. The dilation factor d is kept fixed

TABLE II

TCN BASELINE RECOGNITION PERFORMANCE

Conv. Filters	TCN structure	F1	PR	RC	ER
128	*Causal + Skip*	85.47	88.19	82.92	0.176
	Non-Causal + Skip	85.58	88.16	83.14	0.174
	Causal + No Skip	84.2	86.51	82.01	0.197
	Non-Causal + No Skip	83.91	85.55	82.33	0.201
64	***Causal + Skip***	**85.61**	**89.06**	**82.48**	**0.176**
	Non-Causal + Skip	**86.14**	**89.37**	**83.14**	**0.171**
	Causal + No Skip	85.38	88.55	82.53	0.178
	Non-Causal + No Skip	85.67	88.99	82.59	0.178
32	*Causal + Skip*	85.42	88.99	82.12	0.180
	Non-Causal + Skip	85.70	89.11	82.55	0.176
	Causal + No Skip	85.77	88.91	82.85	0.174
	Non-Causal + No Skip	85.79	88.94	82.86	0.174
64, 2 stacks	*Causal + Skip*	85.14	87.61	82.8	0.173
	Non-Causal + Skip	85.86	88.93	83.0	0.171

to $[1, 2, 4, 8]$. This setup is exported to the active learning experiments: the hyper-parameters for the active learning setup are given in Table I.

TABLE III

ACTIVE LEARNING RECOGNITION PERFORMANCE FOR EACH ACQUISITION STEPS

Acquisition Steps	Random				BALD				Entropy			
	F1	PR	RC	ER	F1	PR	RC	ER	F1	PR	RC	ER
1	50.28	48.32	52.42	0.88	50.28	48.32	52.42	0.88	50.28	48.32	52.42	0.88
2	72.13	73.90	70.45	0.38	72.24	73.87	70.68	0.40	77.44	77.33	77.56	0.27
3	76.92	76.93	76.90	0.29	78.62	79.58	77.68	0.26	78.08	79.03	77.16	0.25
4	79.86	83.33	76.67	0.24	80.39	83.32	77.66	0.22	79.43	84.51	74.93	0.22
5	80.2	83.83	76.87	0.23	81.34	83.62	79.18	0.21	81.66	84.05	79.41	0.20
6	81.04	82.47	79.65	0.25	83.03	85.30	80.87	0.21	82.37	84.73	80.15	0.17
7	82.24	84.90	79.75	0.22	82.00	84.15	79.95	0.17	82.69	84.62	80.85	0.21
8	81.70	80.31	83.13	0.22	83.74	83.81	83.68	0.17	83.45	83.34	83.57	0.18
Time per iteration	T_{Random}**: 544.85 s / it**				T_{BALD}**: 531.85 s / it**				$T_{Entropy}$**: 532.06 s / it**			

B. Classification results

The model performance is measured using the per-frame precision (PR), recall (RC), ($F1$) and error rate (ER) [4] [5]. The $F1$ metric represents a metric between the refinement of our classifier (PR) and the number of correctly detected events (RC) for any number of specific classes. The ER is calculated based on the sum of the total number of insertions I, deletions D, and substitutions S over the N data streams within X_{test}.

Table II shows the recognition results for different configurations of the TCN models. The terms *skip* and *no-skip* refers to the residual connection in the TCN; whereas *causal* and *no-causal* refers to the type of convolution employed (see Figure 1). The TCN model with causal and non-causal convolution bestows outstanding monitoring performance. When the dilation factors increase exponentially, the performance diminishes. This is due to the TCN model accessing a very broad temporal context in time, a very large feature vector sequence, implying that the TCN hinders the modeling of all the intra-event temporal dependencies. With non-causal convolutions, higher performance increases as future frames are incorporated into the convolutional operator. Such temporal aggregation implies a better learning of the energy variations dynamics of the LFCCs, thus yielding lower ER and higher PR, RC and $F1$ metrics. Finally, the advantages of skip connections are evident since, in addition to maintaining the robustness of the model, the aggregation of the input signal through a simple convolution improves monitoring performance. If we increase the number of TCN blocks to make the TCN deeper, although the system is indeed sufficiently fast and adaptable to the known initial data distribution, the TCN tends to decrease performance due to additional complexity added.

C. Active learning

In this subsection, we present the AL results. We start with 31 chronological waveforms from X_{seed}, keeping the rest as the unlabeled data pool U. The selected waveforms correspond to a reasonable time to collect geophysical evidence to request a re-training of the monitoring algorithm. Table I resumes the hyper-parameters for the dataset used in these experiments. The iterative procedure starts with initializing the TCN model with the initial L dataset and select, from the data pool U those events that have to be revised. The TCN samples 100 waveforms from U for each iteration, according to the acquisition functions (see section III.B). Hence, according to the CEAL strategy, the sampled data vector consists of 100 waveforms (50 minutes of the data stream), of which 50 are uncertain, and 50 are certain and thus, acting as *pseudo-labels*.

Table III presents the performance results for each acquisition step and the time required for each iteration. As the number of acquisition steps progresses, the performance metrics improve. Hence, the acquisition steps imply a very significant time gain, in the order of hours or even days of human-based data analysis. Notice that for 8 acquisition steps, that is, 4000 minutes of data curation in a seismological observatory, the BALD attains an F1 of 83.74%, and an ER of 0.17, whereas the *entropy* criteria achieve an F1 of 83.45%, and an ER of 0.18. The AL procedure yields a similar performance if compared to baselines in Table II trained with the entire dataset. Finally, it is observed that for each acquisition step, a steady percentage increase of the monitoring metrics above random sampling criteria demonstrates that the model is queried with the correct events, thus promotes adaptability when there are changing monitoring conditions

VI. Conclusion

This research presents an active learning framework based on temporal convolutions to perform seismo-volcanic event segmentation and classification. The dropout predictive uncertainty from the TCN is exploited to overcome the limitations of how the monitoring model must select uncertain events within the trace for re-training. The experimental results have shown that the TCN integrated with AL provides compelling performance with minimal dataset samples. The presented results illustrate that pragmatism in the design of deep learning monitoring algorithms overcome the ever-present data scarcity issues in automatic seismo-volcanic recognition.

References

[1] S.R. McNutt, G. Thompson, J. Johnson, S. De Angelis, and D. Fee. “Seismic and infrasonic monitoring”. In: (2015), pp. 1071–1099.

[2] Gilberto Saccorotti and Ivan Lokmer. “A review of seismic methods for monitoring and understanding active volcanoes”. In: *Forecasting and Planning for Volcanic Hazards, Risks, and Disasters* (2020), pp. 25–73.

[3] A. Bueno, C. Benítez, S. De Angelis, A. Díaz Moreno, and J. M. Ibáñez. “Volcano-Seismic Transfer Learning and Uncertainty Quantification With Bayesian Neural Networks”. In: *IEEE Transactions on Geoscience and Remote Sensing* 58.2 (2020), pp. 892–902.

[4] M. Titos, A. Bueno, L. García, M. C. Benítez, and J. Ibañez. “Detection and Classification of Continuous Volcano-Seismic Signals With Recurrent Neural Networks”. In: *IEEE Transactions on Geoscience and Remote Sensing* 57.4 (2019), pp. 1936–1948.

[5] A. Bueno, C. Benítez, L. Zuccarello, S. De Angeli, and J.M. Ibanez. “Bayesian Monitoring of Seismo-Volcanic Dynamics”. In: *IEEE Transactions on Geoscience and Remote Sensing* (2021), pp. 1–14. DOI: 10.1109/TGRS.2021.3076012.

[6] A. Bueno, A. Díaz-Moreno, S. De-Angelis, C. Benítez, and J. M. Ibáñez. “Recursive Entropy Method of Segmentation”. In: *Seism. Res. Lett.* 90.4 (2019), pp. 1670–1677.

[7] G. Cortés, L. García, I. Álvarez, C. Benítez, Á. de la Torre, and J.M. Ibáñez. "Parallel System Architecture (PSA): An efficient approach for automatic recognition of volcano-seismic events". In: *Journal of Volcanology and Geothermal Research* 271.12 (2014), pp. 1 –10. DOI: https://doi.org/10.1016/j.jvolgeores.2013.07.004.

[8] M. Titos, A. Bueno, L. García, C. Benítez, and J.C. Segura. "Classification of Isolated Volcano-Seismic Events Based on Inductive Transfer Learning". In: *IEEE Geoscience and Remote Sensing Letters* 17.5 (2020), pp. 869–873. DOI: 10.1109/LGRS.2019.2931063.

[9] N. Pérez, F. Granda, D. Benítez, F. Grijalva, and R. Lara-Cueva. "Toward Real-Time Volcano Seismic Events' Classification: A New Approach Using Mathematical Morphology and Similarity Criteria". In: *IEEE Transactions on Geoscience and Remote Sensing* (2021), pp. 1–13. DOI: 10.1109/TGRS.2020.3048107.

[10] N. Perez, P. Venegas, D. Benítez, R. Lara-Cueva, and M. Ruiz. "A new volcanic seismic signal descriptor and its application to a data set from the cotopaxi volcano". In: *IEEE Transactions on Geoscience and Remote Sensing* 58.9 (2020), pp. 6493–6503.

[11] M. Malfante, M. Dalla Mura, J. Metaxian, J. I. Mars, O. Macedo, and A. Inza. "Machine Learning for Volcano-Seismic Signals: Challenges and Perspectives". In: *IEEE Signal Processing Magazine* 35.2 (2018), pp. 20–30.

[12] S Mostafa Mousavi and Gregory C Beroza. "Bayesian-Deep-Learning Estimation of Earthquake Location From Single-Station Observations". In: *IEEE Transactions on Geoscience and Remote Sensing* (2020).

[13] Razvan Pascanu, Tomas Mikolov, and Yoshua Bengio. "On the difficulty of training recurrent neural networks". In: *International conference on machine learning*. 2013, pp. 1310–1318.

[14] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. "Deep residual learning for image recognition". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 770–778.

[15] K. Wang, D. Zhang, Y. Li, R. Zhang, and L. Lin. "Cost-Effective Active Learning for Deep Image Classification". In: *IEEE Transactions on Circuits and Systems for Video Technology* 27.12 (2017), pp. 2591–2600. DOI: 10.1109/TCSVT.2016.2589879.

[16] Y. Gal. "Uncertainty in deep learning". PhD thesis. PhD thesis, University of Cambridge, 2016.

[17] W. Thelen, M. West, and S. Senyukov. "Seismic characterization of the fall 2007 eruptive sequence at Bezymianny Volcano, Russia". In: *J. Vol. Geoth. Res.* 194.4 (2010), pp. 201 –213.

8 | Geophysical applications

This section of the thesis presents the supporting scientific software that we have implemented for data processing and labelling of the studied datasets, including the automatic detection, segmentation, and classification of volcanic signals. This chapter provides these two research works and their presented key concepts.

8.1 REMOS: Recursive Entropy Method of Segmentation

This section is devoted to the description of the entropy-based automatic detector, segmentation and classification of seismo-volcanic events across eruptions and volcanoes. The frequency-index and visualization techniques based on non-linear embeddings are also presented here. This article is published and available online at Seismological Research Letters (SRL), with the following journal metrics: (IF) 2019: (3.131). Geochemistry and Geophysical (Rank 29/85) (Q2). We reproduce the draft by the journal, which can be cited as:

1. **A. Bueno**, L. Zuccarello, A. Díaz-Moreno, S. De Angelis, J.Woolam, I. Alvarez, M. Titos, M.C. Benitez, J.M. Ibánez. PICOSS: Python Interface for the Classification of Seismic Signals. Computer & Geosciences. 2020 Sep 1; 142-104531.

Recursive Entropy Method of Segmentation for Seismic Signals

A. Bueno[1], **A. Díaz-Moreno**[2], **S. De Angelis**[2], **C. Benítez**[1], **J.M.Ibañez**[3]

[1]Department of Signal Theory, Telematic and Communications, University of Granada, Spain.

[2]Department of Earth, Ocean and Ecological Sciences, University of Liverpool, UK

[3]Instituto Andaluz de Geofísica, University of Granada, Spain

Key Points:

- Entropy-based segmentation
- Automatic Detection
- Data visualization

Corresponding author: Angel Bueno, angelbueno@ugr.es

Abstract

A wealth of data collected over the past three decades have demonstrated that volcanic unrest is often associated with elevated levels of seismicity. Volcano seismic networks commonly record intense swarms of earthquakes in the weeks to months before eruptions; peak rates of more than one event per minute are not uncommon. The ability to readily detect and classify these signals is crucial to effective monitoring operations, and hazard assessment. The sheer volume of information collected, however, poses a challenge to volcano observatories due to the unrealistically large number of staff required for manual inspection of these data. Here, we present REMOS (Recursive Entropy Method of Segmentation), a computationally efficient Python workflow to detect, extract, and classify volcanic earthquakes starting from raw, continuous, waveform data. Within REMOS, seismograms are first analyzed using the well-established Short-Term Average/Long-Term Average method to identify trigger times of candidate earthquakes. A new algorithm based on measurements of seismic energy and minimum entropy is then employed to investigate large amounts of earthquake triggers, and to discriminate and parse events into individual waveforms for further analyses. REMOS also includes a facility for classification of the extracted waveforms, based on simple frequency-domain metrics. Finally, the results can be visualized using t-Distributed Stochastic Neighbor Embedding (t-SNE), a technique for dimensionality reduction that is particularly well-suited to inspection of high-dimensional datasets. In this work we demonstrate the use of REMOS with seismic data recorded in 2007 during a period of unrest and eruption at Bezyminany volcano. Our results show that REMOS can efficiently detect, segment and classify earthquakes at scale, and at very low computational cost.

Introduction

Volcano and earthquake observatories routinely archive large amounts of seismic data that contain informative evidence on the likelihood of forthcoming eruptions. Arising from this plethora of observations there is a growing appreciation of how earthquake swarms reflect a variety of physical processes underneath volcanoes (e.g., *McNutt et al.* [2015]). Earthquake catalogs associated with intense periods of unrest and elevated seismicity at volcanoes, however, remain largely incomplete. The sheer volume of data recorded by volcano observatories poses a challenge due to the unrealistic amount of human resources that would be required for manual inspection in real-time. Similarly, efforts to retrospectively analyse decade-long time-series, and re-assess catalogs, are rarely undertaken. On the other hand, complete earthquake catalogs in near real-time would allow computation of parameters to form the input for, or allow validation of, advanced statistical and physical models of volcanic activity (*Bell and Kilburn* [2012]). Understanding the time history of seismicity at

volcanoes can, thus, provide valuable constraints for the assessment of the likelihood and timing of eruptions, and the prognosis of their related hazards. During volcanic crisis, the timeliness of response to unrest is critical, and large volumes of continuous data must be rapidly processed and interpreted (*Tepp, G.* [2018]).

Here, we introduce REMOS (Recursive Entropy Method of Segmentation), a complete workflow for detection, segmentation, classification, and visualization of seismic data. Starting from continuous data streams and the activation times of STA/LTA (Short-Term Average/Long-Term Average (*Withers et al.* [1998])), we identify a set of energy and entropy-based criteria to parse data into individual waveforms (segmentation); segmented data are then classified according to their frequency content, using well-established frequency domain metrics (*Buurman and West* [2010]). In order to assess the quality of the resulting catalog, and to provide direct knowledge of the high-dimensional structure of the dataset, we extend the capabilities of REMOS by including an exploratory data visualization tool based on t-Distributed Stochastic Neighbor Embedding (t-SNE) and frequency attributes (*Maaten and Hinton* [2008]). REMOS is developed in Python, and leverages its rich scripting syntax and scientific libraries, including Numpy, Scikit and Obspy (*Beyreuther et al.* [2010]). Here, the applicability of REMOS is tested using data collected at Bezymianny volcano (Kamchatka, Russia) (*Thelen et al.* [2010]). In particular, we analyze data recorded during a pre-eruption sequence between September-December 2007, which includes variable rates of seismic activity, and a comprehensive range of waveform types. This example demonstrates the use of REMOS, and provides a preliminary benchmark for its scalability when applied to comparatively large datasets.

Introduction

Volcano and earthquake observatories routinely archive large amounts of seismic data that contain informative evidence on the likelihood of forthcoming eruptions. Arising from this plethora of observations there is a growing appreciation of how earthquake swarms reflect a variety of physical processes underneath volcanoes (e.g., *McNutt et al.* [2015]). Earthquake catalogs associated with intense periods of unrest and elevated seismicity at volcanoes, however, remain largely incomplete. The sheer volume of data recorded by volcano observatories poses a challenge due to the unrealistic amount of human resources that would be required for manual inspection in real-time. Similarly, efforts to retrospectively analyse decade-long time-series, and re-assess catalogs, are rarely undertaken. On the other hand, complete earthquake catalogs in near real-time would allow computation of parameters to form the input for, or allow validation of, advanced statistical and physical models of volcanic activity (*Bell and Kilburn* [2012]). Understanding the time history of seismicity at

volcanoes can, thus, provide valuable constraints for the assessment of the likelihood and timing of eruptions, and the prognosis of their related hazards. During volcanic crisis, the timeliness of response to unrest is critical, and large volumes of continuous data must be rapidly processed and interpreted (*Tepp, G.* [2018]).

Here, we introduce REMOS (Recursive Entropy Method of Segmentation), a complete workflow for detection, segmentation, classification, and visualization of seismic data. Starting from continuous data streams and the activation times of STA/LTA (Short-Term Average/Long-Term Average (*Withers et al.* [1998])), we identify a set of energy and entropy-based criteria to parse data into individual waveforms (segmentation); segmented data are then classified according to their frequency content, using well-established frequency domain metrics (*Buurman and West* [2010]). In order to assess the quality of the resulting catalog, and to provide direct knowledge of the high-dimensional structure of the dataset, we extend the capabilities of REMOS by including an exploratory data visualization tool based on t-Distributed Stochastic Neighbor Embedding (t-SNE) and frequency attributes (*Maaten and Hinton* [2008]). REMOS is developed in Python, and leverages its rich scripting syntax and scientific libraries, including Numpy, Scikit and Obspy (*Beyreuther et al.* [2010]). Here, the applicability of REMOS is tested using data collected at Bezymianny volcano (Kamchatka, Russia) (*Thelen et al.* [2010]). In particular, we analyze data recorded during a pre-eruption sequence between September-December 2007, which includes variable rates of seismic activity, and a comprehensive range of waveform types. This example demonstrates the use of REMOS, and provides a preliminary benchmark for its scalability when applied to comparatively large datasets.

The REMOS algorithm

Segmentation of continuous seismic data streams relies on signal processing methods that exploit the internal structure of the data. From an information theory perspective, when earthquakes are generated by sources within a volcano the system is characterized by low-levels of chaos; in contrast, when volcano-seismic sources are not active seismometers record the continuous Earth's background noise, characterized by high entropy levels. Based on these principles, REMOS calculates the pseudo-energy of the seismic signal within specified time windows along the continuous waveforms in order to discriminate the boundaries between events (low entropy) and background noise (high entropy). The pseudo-energy of a seismogram, on the component being analysed, is calculated as

$$E = < s(n) \cdot s(n) > = \sum_{n=1}^{N} |s(n)|^2 \tag{1}$$

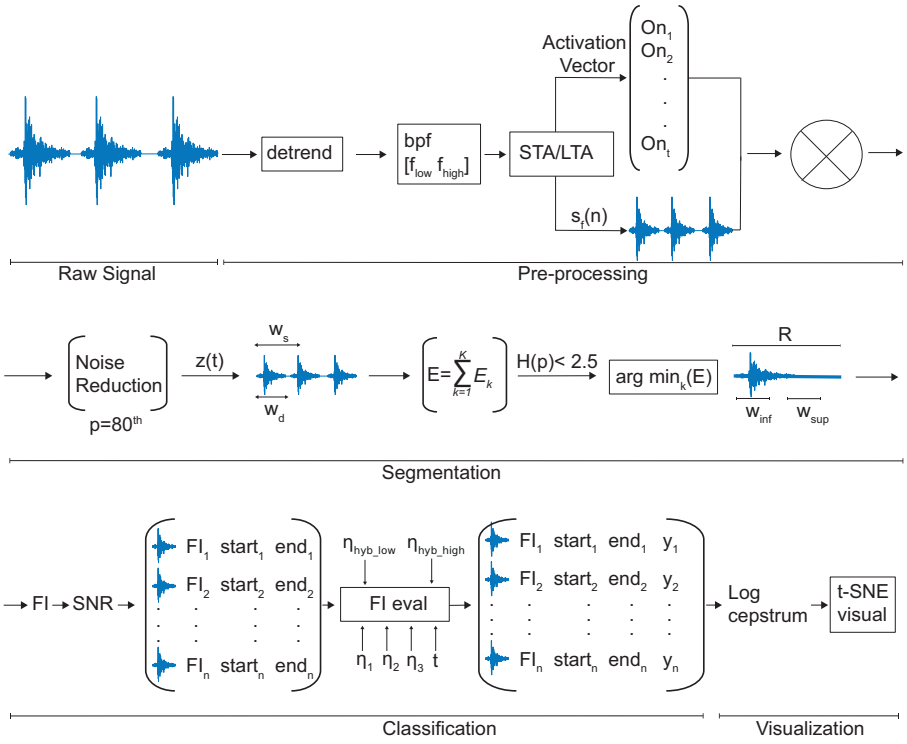


Figure 1. Recursive Entropy Method of Segmentation (REMOS) workflow. The algorithm is divided into four different stages: preprocessing, segmentation, classification, and visualization. From continuous raw data streams, initial processing aims to increase signal quality and produce an activation vector (on) from application of short-term average/long-term average (STA/LTA). Each element in on is considered to explore a segmentation region for potential earthquakes. Candidate events are segmented according to a minimum energy criterion within these regions. When all events are segmented, frequency-based classification and t-distributed stochastic neighbor embedding (t-SNE) visualization are performed. Bpf, bandpass filter; FI, frequency index.

where s is a vector of ground velocities, and N the number of samples in s. Note that E is proportional to seismic energy but it does not have physical dimensions of energy. In the next sections we present a description of the four main steps in REMOS: pre-processing, segmentation, classification, and visualization.

Pre-processing

Figure [1] illustrates the workflow of REMOS. Preliminary processing steps are applied to the continuous data before segmentation. Linear trends associated with effects such as very long-period instrument drift are removed. A bandpass filter in a user-selected frequency range $[f_{lo}\ f_{high}]$ $(Hz.)$

is applied to enhance the presence of earthquake signals; during our tests we observed that for most volcanoes the frequency band [1-15] Hz. is appropriate. The filtered data are then scanned using a recursive STA/LTA algorithm to obtain trigger times of potential events (*Withers et al.* [1998]). The parameters for the STA/LTA algorithm are selected by the user according to the specific dataset (*Trnkoczy* [1999]). For volcano-seismic signals recorded at relatively close distance from the source ($< 10km$), a short-term window of $0.5 - 2s$ and a long-term window of $8 - 15s$ are, frequently, appropriate; the STA/LTA threshold to declare a trigger is dependent on signal-to-noise ratio. This procedure yields a one-dimensional activation vector, $on = [on_1, on_2, on_3, ..., on_n]$, of earthquake trigger times.

Segmentation

In the second step of REMOS the vector of activation times, on, is used in combination with the filtered signal to investigate regions within the continuous data stream that contain earthquakes, and to parse these events into individual waveforms. Data segmentation in REMOS depends on two parameters: the maximum search window, W_s, and the minimum duration window, W_d. The maximum search window W_s represents the time window that REMOS explores to detect an event; W_d captures the minimum duration that REMOS uses to calculate signal energy. A noise reduction procedure is applied to the data stream (Fig. 1) to mitigate the influence of external noise sources (e.g: electronic spikes). Data are corrected according to:

$$z(n) = \begin{cases} 0 & s_f(n) <= T \\ \frac{s_f(n)}{T} & s_f(n) > T \end{cases} \tag{2}$$

where $s_f(n)$ is the filtered signal, and T a threshold value that corresponds to the n-th percentile of the data stream. This threshold is selected to be large enough to represent the background noise levels for the data stream. We, then, consider all activation times in on, and extract segments of data (exploration regions) with duration W_s starting at these times. Each of these regions is then windowed into k frames with duration W_d. The energy of the signal is computed according to (1) within all exploration regions, for each individual frame. The entropy, $H(p)$, is then calculated from the pseudo-probability normalized vector, p, as:

$$H(p) = \sum_{i=1}^{k} -p_i \, log(p_i) \tag{3}$$

where p_i is the normalized energy of each frame. If the entropy is low, below a pre-defined threshold, ϵ, then the onset of an event is declared. In this instance, the candidate event is extracted from the exploration region and parsed into an individual waveform according to the minimum of the energy distribution:

$$\arg\min_k E_k \tag{4}$$

The end of an event is declared when the energy reaches a minimum. If the exploration region contains multiple earthquakes, using the minimum energy as a factor to discriminate the end of an event ensures correct segmentation. To overcome long segmentations when taking the lowest energy, we compute the ratio R of the energies of the initial and final parts of the segmented signal. This ratio would be close to zero if no earthquake signal was present in the final section of the segmented data. In that case, the original candidate is segmented again. Otherwise, for ratios closer to one, the original candidate is selected, as it could be for instance an episode of long-duration tremor. Once the candidate is segmented, results are stored into a matrix. REMOS continues segmenting signals along the main trace until the activation vector *on* is exhausted. As a final step, candidates with low SNR can be erased from the final dataset. In REMOS, the SNR is computed as the ratio of the peak amplitude within a time window containing the surface wave signals to the root-mean-square of the noise trailing the signal arrival window.

Classification

The individual traces extracted from the continuous data are then classified based on a simple metric, the frequency index (FI) (*Buurman and West* [2010]). Automated data classification has the advantage of producing objective classifications at scale (*Titos et al.* [2018]), (*Ibáñez et al.* [2009]). The time required to analyze data is greatly reduced, and any mismatch in event classification due to the bias introduced by a human analyst are potentially eliminated. REMOS computes the FI as:

$$FI = log_{10}\left(\frac{A_{high}}{A_{low}}\right) \tag{5}$$

where A_{high} and A_{low} are spectral energies above and below a certain threshold value, respectively. We define a set of adjustable parameters $\eta = \{\eta_1, \eta_{hybrid\,low}, \eta_{hybrid\,high}, \eta_2, \eta_3\}$ as the FI thresholds that control how events are categorized according to their FI. Events are classified as Low-Frequency (LF) if the FI is below than a given threshold η_1. Similarly, if the FI its greater

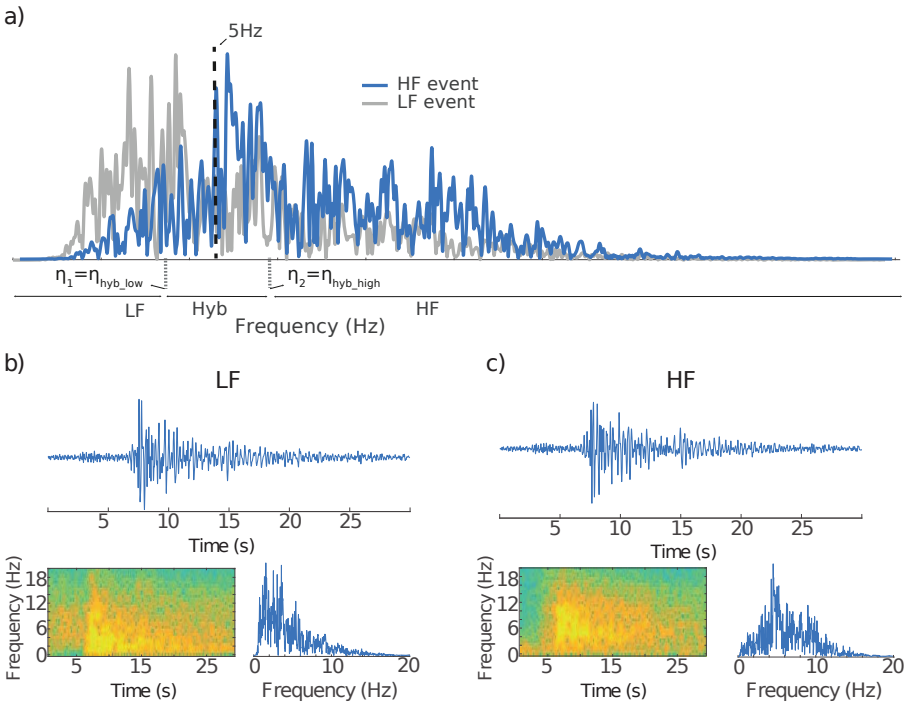


Figure 2. Comparison of High Frequency and Low Frequency spectrum and threshold selection for frequency index calculation. In order to include hybrids, $\eta_1 = \eta_{hybrid\ low}$ and $\eta_{hybrid\ high} = \eta_2$, but other configurations are possible. Figure 2.b and 2.c depict the seismograms, spectrograms and power spectrum for two segmented signals at BELO station, classified as low frequency (2.b) and high frequency (2.c) events.

than the given threshold η_2, earthquakes are classified as High-frequency (HF). The events with FI between η_1 and η_2 deliver energy at both, low and high frequencies, and could be categorized as hybrids. The parameters η_1 and η_2 can be set with the same values, thus acting as a single threshold for HF and LF events.

Tremors and Rockfalls are discriminated according to their frequency index and duration t. Additionally, η_3 is a threshold to control when an event is declared to be a rockfall or a tremor. The minimum duration t over which to consider an event either low frequency or tremor depends of the user and the volcanic-environment (*McNutt et al.* [2015]). Figure 2.a illustrates the frequency content of a HF and LF events at Bezymianny volcano. In the case of LF events, most of the frequency content is below 5 Hz, making the FI from equation 5 to be below zero or negative. Similarly, HF events result in FI to be greater than zero, or positive. Thus, LF's events have negative FI's, whereas HF will have positives FI's. The range of values between η_1 and η_2 could represent hybrid events.

REMOS supports the configuration of $\eta_{hybrid\,low}$ and $\eta_{hybrid\,high}$ parameters to quantify hybrid events and overcome the limitation of a single threshold for HF and LF events. Hybrid earthquakes could be labeled by selecting any intermediate values between η_1 and η_2. However, one can set the values of $\eta_{hybrid\,low}$ and $\eta_{hybrid\,high}$ to zero to not categorize any hybrid events, or set $\eta_{hybrid\,low}$ and $\eta_{hybrid\,high}$ to any value between η_1 and η_2 in order to categorize hybrids with FI between the threshold for low and high frequency events. The value of t could be set by the analysts based on geophysical knowledge of the specific study volcano. Typically, $30s$ should suffice for most volcanoes. Volcanic tremors are often observed as a precursor to and during eruptions; tremor that can last from minutes to days, or even longer, and delivers energy at frequencies, typically, between 1 and 10 Hz. In the case of harmonic tremor, multiple, equally spaced, spectral peaks can be observed. Classification of these signals is beyond the scope of REMOS, which focuses on the segmentation of seismic signals at a higher level; specific detectors exist to discriminate sub-categories within the general classification scheme (*D. C. Roman* [2017]).

Data exploration and visualization

Exploratory analysis holds potential to provide insights about topological hierarchies in high dimensional data. Here, for the purpose of visualization of REMOS outputs, we adopt t-SNE, a dimensionality reduction method that has proved very useful to discover subtle differences in highly complex data, and to obtain meaningful information about how data points are distributed in high dimensional spaces. The algorithm t-SNE aims to find a representation ($embedding - vector$) that can minimize the distance between multidimensional data (with hundred of features) and a lower representation. This distance can be inferred using the Kullback-Leibler (KL-divergence) as it measures the difference between two probability distributions. Therefore, the $embedding - vectors$ could be found as the minimization of the KL-divergence divergence between the joint probability distribution from data points in high dimensional space, P, and the joint probability distribution Q from the embeddings in the low dimensional space:

$$KL[P||Q] = \sum_i \sum_j p_{ij} \log \frac{p_{ij}}{q_{ij}} \tag{6}$$

with P and Q the probability distributions between the high- and low-dimensional space, p_{ij} and q_{ij} the pairwise similarities across data points in the high- and low-dimensional space. This minimization process helps to preserve topology and information from complex data by projecting representative embeddings into a low dimensional space that faithfully capture hidden similarities.

The reader is referred to *Maaten and Hinton* [2008] for further details on the t-SNE algorithm and visualization methodology. We use log-cepstrum frequency features (cepstral coefficient on a logarithmic scale of the given data *Beyreuther et al.* [2010]), as the input data for the t-SNE algorithm. The natural logarithms of the filter-bank energies are calculated, producing a 16-parameter feature vector. The Discrete Cosine Transform (DCT) is used to de-correlate the features and reduce the number of components from the feature to 13 coefficients. Therefore, log-cepstrum coefficients would enhance lower frequencies whilst providing highly refined information for the visualization.

Application of REMOS

In this section we demonstrate the use of REMOS with data collected at Bezymianny volcano (Russia) during 2007; this dataset contains a large number earthquakes, including high frequency (HF), low frequency (LF), rockfalls (R), and tremors (T). Unrest at Bezymianny during 2007 and the dataset used here have been previously investigated by *Thelen et al.* [2010], thus providing a reliable benchmark for the performance of REMOS and geophysical insight for the configuration of W_d, W_s and ϵ parameters. A Python script and notebook examples, along with seismic data from this example, are included as supplementary material. We perform our experiments on a 64-bit computer with a i7-8700k CPU (3.70GHz), 32GB RAM and Ubuntu 16.04. The unprocessed trace is filtered between $f_{low} = 1Hz$ and $f_{high} = 12Hz$. The recursive STA/LTA algorithm is used with STA of 3 seconds, LTA of 15 seconds, 2.0 as the on trigger, and 1.0 as off trigger thresholds, respectively. A vector of activation times of length 818 (the number of regions to explore) is generated. Individual events are then extracted using the segmentation method implemented in REMOS:

```
>>> W_d = 10
>>> W_s = 80
>>> windows = [W_d, W_s]
>>> remos_segmentation(original, processed, on, [W_d, W_s], epsilon=2.5,
plot=False, cut="original", delay_in=4.5 )
```

The REMOS short window (W_d) is set as 10 seconds, whereas the region of exploration W_s is 80 seconds. The parameter $delay_{in}$ can be configured to segment the signal before the detection time. In this work, and following *Thelen et al.* [2010], we define $delay_{in}$ as 4.5 seconds before the trigger time, given our STA/LTA is configured to detect very impulsive events. Being completely optional, the parameter $delay_{in}$ can be set to zero. If the parameter is included, the difference between the *on* time and $delay_{in}$ is included in the segmented waveform. The normalized $80th$ percentile from

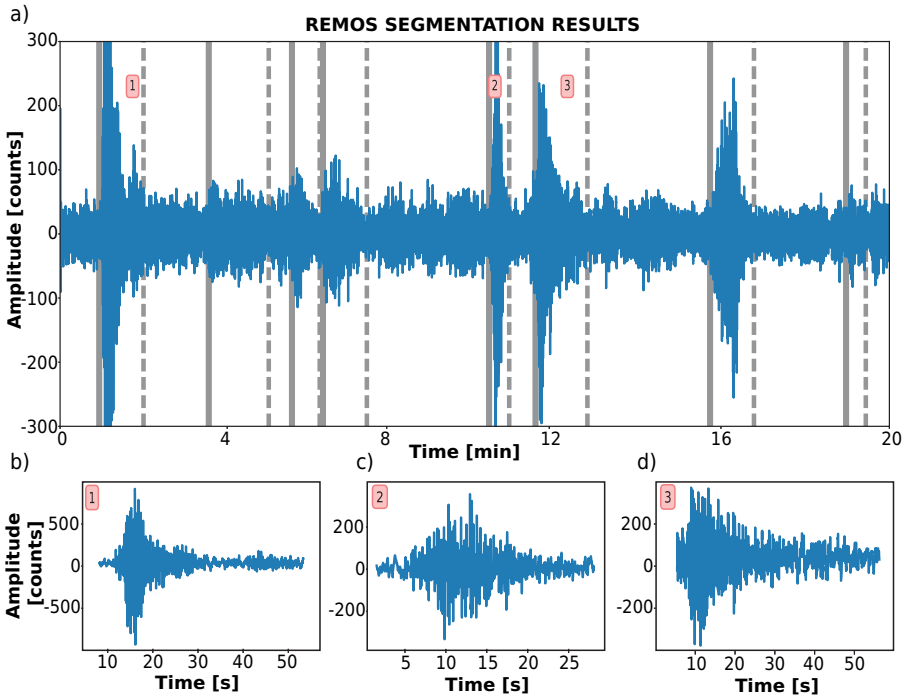


Figure 3. (a) Segmentation results with REMOS triggers on the continuous trace (17th October 2007) from BELO station before classification and visualization stage, filtered between *1* and *12 Hz*. Different examples of segmented waveforms by REMOS, with (b) a signal with a small energetic coda, (c) a short event and (d) exponential decay segmented waveform.

the seismic trace is used to reduce noise, and processed trace along with activation vector *on* are used for discriminative segmentation. Regions of exploration with duration W_s are segmented and windowed with 10 seconds duration (W_d). The entropy $H(p)$ is computed using the short-term energy as in equation 3. If $H(p)$ is below $\epsilon = 2.5$, the minimum of the energy is taken (see equation 1) within the W_s region and potential candidates are cut. The value of the ϵ threshold is selected to minimize the excessive influence of background noise, and can be adjusted by the analyst. Further, each candidate is windowed by half value of W_d, and energy ratios computed. If the parameter *cut* is set to *original*, the candidate is taken from the raw data, and a 0.5 *Hz* high-pass filter is applied to reduce background noise. Otherwise, the event is selected from the filtered trace. When all the events are segmented, signals with low SNR are erased from the final data matrix. Start and end times (in samples) and frequency index of each event are also stored.

Figure 3 shows the results for the segmentation of a 20-minute seismogram recorded on 17th October, 2007 at station BELO. All events are taken from the filtered data by setting the parameter

cut to $filtered$. The entropy criteria, used jointly with the minimum energy search, results in very effective data segmentation. Figure 3.b illustrates how REMOS preserve the coda decay of seismic signals. Figure 3.c, for instance, shows how REMOS can detect short-duration events.

The next step is FI classification. Events at Bezymianny volcano are characterized by a frequency range between 1 and 10 Hz. As the parameter *cut* was set to $original$, we use the high-pass filtered signal (0.5 Hz) to perform classification. Therefore, FI is calculated by considering a low-frequency band between 1 to 5 Hz, and high frequencies between 6 and 10 Hz. The FI cutoffs η_1 (for LF), η_2 for (HF) and η_3 (for R, T) are selected as $\eta_1 = -0.2$ and $\eta_2 = 0.2$, and $\eta_3 = 0.2$. These cutoffs values were empirically defined after manual data inspection. As no hybrid events were present in the study by *Thelen et al.* [2010], $\eta_{hybrid\,low}$ and $\eta_{hybrid\,high}$ were set to zero. Figure 2 shows the classification results for two segmented signals. Notice that the ratio of spectral energy above and below $5Hz$ suffices to discriminate events at BELO station.

We note that the FI thresholds must be fine-tuned and calibrated according to the specific dataset. We refer the reader to (*Buurman and West* [2010]) for further discussion on the choice of the correct FI parametrization. In simple words, when the frequency spectrum is dominated by energy below $5Hz$, the event is labeled as LF (figure 2.b). When the frequency content is mostly above $5Hz$, the events are labeled as HF (figure 2.c). The duration of the segmented waveform serves to discriminate rockfalls and tremor from other events. REMOS was able to detect 3259 HF, 3796 LF, 1442 R and 864 T. The classification matrix is stored in Numpy format ($.npy$). The average processing time for these steps, (segmentation and classification) was 7.04 seconds.

In the final step of the REMOS workflow we combine t-SNE and log-cepstrum features to assess the quality of the labeled dataset. The non-linear nature of t-SNE reduces the dimensionality of the log-cepstrum features whilst providing a good understanding of global and local data structures (*Maaten and Hinton* [2008]). Log-cepstrum features enhance frequencies below 20 Hz, especially at the lower end of the spectrum (*Ibáñez et al.* [2009]). A total of 13 log-cepstrum coefficients are computed using the *Obspy* signal processing package. A normalized t-SNE visualization map is shown in figure 4. At BELO station, t-SNE reveals the presence of four distinct clusters. Notice that few tremors (clear blue) fall within the low-frequency cluster (turquoise). Similarly, some rockfalls (blue) appear within the high frequency region (darkest blue). This map allows rapid visualization of what type of events are present in a database, and serves as an exploratory data analysis that could be used to refine posterior statistical analysis. Thus, given a specific classification criteria, REMOS produces an objective data-driven classification consistent with time and different observed features.

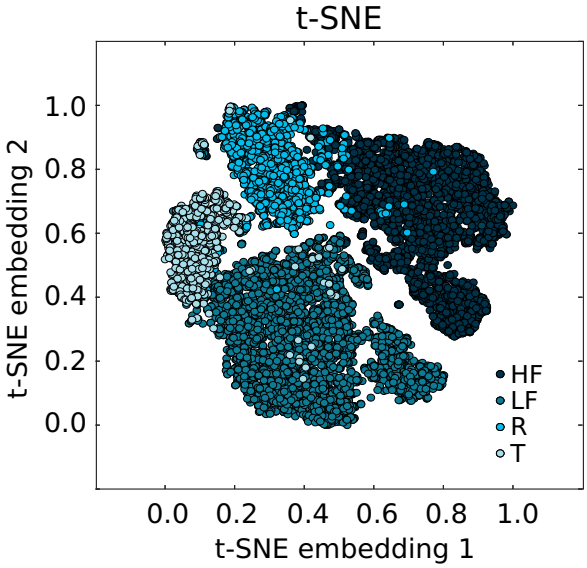


Figure 4. t-SNE visualization on BELO station

Conclusions

We have introduced REMOS, a Python module for the detection, segmentation, classification and visualization of continuous volcano-seismic data streams. REMOS makes extensive use of optimized scientific libraries, including Obspy, SciPy and NumPy. REMOS has a highly modular structure that can be easily adapted to different case studies. The results of our tests suggest that REMOS is capable of consistently segmenting and classifying large amounts of volcano-seismic data, with minimum supervision. REMOS provides a computationally cost-effective solution to enhance volcano monitoring systems, allowing rapid assessment of volcano-seismic unrest. Here, we have demonstrated the use of REMOS with data gathered during a period of unrest at Bezymianny volcano in 2007. The robustness and scalability of REMOS were tested against previous analyses of the same dataset. Our results show that REMOS can effectively detect, segment and classify volcano-seismic events, even when the signal-to-noise ratio is poor or event occurrence rates are high. The methodology described here is easily transferable to other domains, such as infrasound or hydroacoustic data. We suggest that real-time implementation of algorithms like REMOS is crucial to improving our ability to monitor unrest at active volcanoes.

Data and Resources

The data used in this study can be obtained from IRIS DMC. The facilities of IRIS Data Services, and specifically the IRIS Data Management Center, were used to access the waveforms, related metadata, and/or derived products used in this study. IRIS Data Services are funded through the Seismological Facilities for the Advancement of Geoscience and EarthScope (SAGE) Proposal of the National Science Foundation under Cooperative Agreement EAR-1261681.

The REMOS implementation, along with example data, extensive documentation and conda environment can be found at *https://github.com/srsudo/remos*. Scientific libraries, including NumPy, SciPy and ObsPy are accessible via their web-pages: https://www.numpy.org, https://www.scipy.org and https://www.obspy.org. Scikits.talkbox can be accessed at https://github.com/cournape/talkbox, and installed via Python package manager, *pip*. Last access on all websites on August 2018.

Acknowledgments

We would like to thank all the members at the Seismic Laboratory at University of Liverpool for continuous support and advice when developing REMOS algorithm. We would like to thank Weston Thelen and an anonymous reviewer for their detailed and constructive comments, which were very helpful in improving the original article. This research was funded by the "Spanish Ministerio de Economiay Competitividad" grant TEC2015-68752 (MINECO/FEDER) and by the UK Natural Environment Research Council (NERC) Grant NE/P00105X/1.

References

Bell, A. F., and C. R. J. Kilburn (2012), Precursors to dyke-fed eruptions at basaltic volcanoes: insights from patterns of volcano-tectonic seismicity at Kilauea volcano, Hawaii, *Bulletin of Volcanology*, *74*(2), 325–339, doi:10.1007/s00445-011-0519-3.

Beyreuther, M., R. Barsch, L. Krischer, T. Megies, Y. Behr, and J. Wassermann (2010), Obspy: A python toolbox for seismology, *Seismological Research Letters*, *81* (3), 530–533.

Buurman, H., and M. West (2010), Seismic precursors to volcanic explosions during the 2006 eruption of Augustine volcano, *1769*.

Roman, D. C. (2017), Automated detection and characterisation of harmonic tremor in continuous seismic data: Harmonic tremor detection, *44*.

Ibáñez, J.M., C. Benítez, L. Gutiérrez, G. Cortés, A. García-Yeguas, and G. Alguacil (2009), The classification of seismo-volcanic signals using hidden Markov models as applied to the stromboli

and etna volcanoes, *Journal of Volcanology and Geothermal Research*, *187*, 218–226.

Maaten, L. v. d., and G. Hinton (2008), Visualizing data using t-sne, *Journal of machine learning research*, *9* (Nov), 2579–2605.

McNutt, S. R., G. Thompson, J. Johnson, S. D. Angelis, and D. Fee (2015), Chapter 63 - seismic and infrasonic monitoring, in *The Encyclopedia of Volcanoes (Second Edition)*, edited by H. Sigurdsson, second edition ed., pp. 1071 – 1099, Academic Press, Amsterdam, doi: https://doi.org/10.1016/B978-0-12-385938-9.00063-8.

Thelen, W., M. West, and S. Senyukov (2010), Seismic characterization of the fall 2007 eruptive sequence at Bezymianny volcano, Russia, *Journal of Volcanology and Geothermal Research*, *194* (4), 201–213.

Tepp, G. (2018) A Repeating Event Sequence Alarm for Monitoring Volcanoes. *Seismological Research Letters*, *89* (5), 1863–1876. doi:https://doi.org/10.1785/0220170263.

Titos, M., A. Bueno, L. García, and C. Benítez, (2018), A deep neural network approach to automatic recognition systems for volcano-seismic events, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *11* (5), 1533–1544, doi:10.1109/JSTARS.2018.2803198.

Trnkoczy, A. (1999), Topic understanding and parameter setting of sta/lta trigger algorithm, *New Manual of Seismological Observatory Practice*, *2*.

Withers, M., R. Aster, C. Young, J. Beiriger, M. Harris, S. Moore, and J. Trujillo (1998), A comparison of select trigger algorithms for automated global seismic phase and event detection, *88*, 95–106.

8.2 PICOSS: DATA CURATION PLATFORM

This section present the data-curation open-source interface implemented as part of this thesis. This article is published and available online at Computer& Geosciences, with the following journal metrics: (IF), JCR 2019: (2.991). Computer science, interdisciplinary applications (Rank 49/109) (Q1). Geoscience Multidisciplinary (Rank 58/200) (Q2).

PICOSS: Python Interface for the Classification of Seismic Signals

A. Bueno[a], L. Zuccarello[a,b,c], A. Díaz-Moreno[d], J. Woollam[e,d], M. Titos[a], C. Benitez[a], I. Álvarez[a], J. Prudencio[b], S. De Angelis[d]

[a] *Department of Signal Theory, Telematic and Communications, University of Granada, Spain.,* [1]
[b] *Department of Theoretical Physics and Cosmos, University of Granada, Spain.*
[c] *Istituto Nazionale di Geofisica e Vulcanologia, Sezione di Pisa, Italy*
[d] *Department of Earth, Ocean and Ecological Sciences, University of Liverpool, UK*
[e] *Geophysical Institute, Karlsruhe Institute of Technology, Karlsruhe, Germany.*

Abstract

Over the last decade machine learning has become increasingly popular for the analysis and characterization of volcano-seismic data. One of the requirements for the application of machine learning methods to the problem of classifying seismic time series is the availability of a training dataset; that is a suite of reference signals, with known classification used for initial validation of the machine outcome. Here, we present PICOSS (Python Interface for the Classification of Seismic Signals), a modular data-curator platform for volcano-seismic data analysis, including detection, segmentation and classification. PICOSS has exportability and standardization at its core; users can select automatic or manual workflows to select and label seismic data from a comprehensive suite of tools, including deep neural networks. The modular implementation of PICOSS includes a portable and intuitive graphical user interface to facilitate essential data labelling tasks for large-scale

[1] **corresponding author: Angel Bueno angelbueno@ugr.es.**

volcano seismic studies.

Keywords: Volcanoes, Software, Classification, Segmentation, Detection

1. Introduction

Seismic networks are the backbone of volcano monitoring programs. Rapid technological advances over the past two decades have made the installation of geophysical networks with multiple sensors increasingly more affordable; thus, large amounts of data are routinely generated and archived at volcano and earthquake observatories. Earthquakes at volcanoes may occur at rates as high as hundreds of events per hour during periods of unrest (*Bell and Kilburn*, 2012); their waveforms must be extracted from the continuous seismic records and classified. Volcano-seismic signals are traditionally classified based on either waveform appearance (e.g., amplitudes, duration, and the number and type of seismic phases visible) or frequency domain features (e.g., the frequency band over which most energy is delivered). With different terminology but similar classification criteria, different authors (e.g: (*Minakami*, 1974), (*Lahr et al.*, 1994), (*Chouet, B.*, 2003), (*McNutt et al.*, 2015)) have proposed classification schemes for volcanic earthquakes. Today's most commonly adopted volcano-seismic classifications include high-frequency (HF), hybrids or mixed frequency (MF), low-frequency (LF) earthquakes and tremor (T), in addition to a number of other signals generated by surface processes (landslides, lahars, pyroclastic flows), and explosions (E). Over the years, different solutions for analysing continuous seismic data have been implemented (*Goldstein et al.*, 1998, *Havskov et al.*, 1999, *Lesage*, 2009, *Abdelwahed, M.F.,*, 2012, *Olivieri et al.*, 2012, *Álvarez et al.*, 2013, *Romero*

et al., 2016). Most recently, the Python language has gained popularity in the field of seismology and new toolboxes have been developed; the most popular example is Obspy (*Beyreuther et al.*, 2010), a flexible and modular environment to access waveform data in different formats, and to perform both basic and high-level data analysis tasks on multi-channel seismic data (*Zuccarello et al.*, 2016). In this manuscript, based on our experience with the development of APASVO (*Romero et al.*, 2016), a Python and Obspy-based interface designed for manual picking of P- and S-wave arrival times, we have built PICOSS (Python Interface for the Classification of Seismic Signals), a modular open source interface designed to support visualization, detection and characterization of volcano-seismic signals. In PICOSS, a single graphical user interface allow the user to span a wide range of tasks in seismic analysis, significantly simplifying the seismologist's workflow and decreasing the discovering of geophysical insights. PICOSS expands our first-generation interface by implementing additional signal processing and machine learning techniques. Additionally, the simplicity of its GUI offers opportunities for its use in higher education classroom settings.

2. PICOSS Description

PICOSS is a program conceived to detect and label volcano-seismic datasets, which offers a high-level of modularity as modern seismology toolboxes. All outputs are stored in easily accessible, platform-independent, standard format. This format includes segmentation times, along with the identified type of event, quality of the event, and seismological information used to build robust datasets. PICOSS can access seismic data from community wave-

form data servers (e.g., the Incorporated Research Institution for Seismology Data Management Center, IRIS DMC), and from off-line data structures in multiple formats. PICOSS is built to enable the following work-flow loop:

1. **Manual Inspection**: PICOSS provides a fully-functional GUI interface for manual inspection of seismic events, allowing the analyst to associate custom loaded labels and organise relevant seismic information to annotate the data. Supervised earthquake analysis capabilities include time-frequency analysis, and assignment of classification labels and picking times.
2. **Detection**: PICOSS supplies algorithms, such us STA/LTA (Short-Term Average/Long-Term Average; (*Allen*, 1982)) or adaptative multiband processing algorithm (AMPA; (*Álvarez et al.*, 2013)) to perform automatic detection and picking. Additional infrastructure and binding modules are also provided to ease the integration with new algorithms.
3. **Classification**: The program includes a classification module that categorizes seismic signals according to Frequency Index analysis (*Buurman and West*, 2010) or a multi-volcano pre-trained neural network (*Bueno et al.*, 2019b) specifically designed to refine datasets and decrease manual inspection time.

Our software can read labels and/or phase arrival times from existing catalogues, and associate them to waveform data. By default, PICOSS include the most common labels in volcanic-seismology. However, specific categorization schemes can be loaded via the *"Extra Info"* menu, which permits a broader range of applications (i.e, infrasound or hydro-acoustic data). Finally, the classification module permits grouping of the segmented signals

based on unsupervised learning techniques, with an intuitive visualization tool of data hierarchies. Specifically, active learning, i.e., the continuous and interactive querying of data following an acquisition function criteria, is incorporated as a procedure to decrease the amount of time required for data labelling and model adaptation (*Neal, R.M,*, 2012). A Command Line Interface (CLI) is provided with a set of auxiliary task for data conversion between formats and add scalability to the whole workflow.

2.1. PICOSS GUI

PICOSS can display single-channel continuous streams, currently up to 24h, along with their spectrogram and Fourier power spectrum. Figure 1 shows an example of continuous seismic data recorded on 19th October, Montserrat Volcano observatory (MVO), station *MBGA*, vertical component *Z* (*Luckett et al.*, 2007). The analyst can drag-select any part of the loaded stream, visualize the spectrogram and the spectrum, and annotate the data. Above the main trace panel (*main*), PICOSS includes a toolbar from which the user can access most of the program facilities, i.e: loading traces, signal processing routines, detection and picking routines or add extra (custom) information. The default seismic labels offered by PICOSS follows common classification practice and terminology in volcano-seismology, although custom labels can be loaded via the *"Extra Info"* menu. Additionally, a quality factor, Q, can be assigned to the signal as a qualitative measure of the analyst's confidence in the classification; Q ranges from 1 (very poor) to 5 (excellent). Segmented events can be compared with data recorded on other stations/components; the options *Visualize other components* and *Visualize other stations* open an auxiliary small GUI in which the user can explore

other components and/or stations.

PICOSS is written in Python, and built on Obspy (*Beyreuther et al.*, 2010), a popular Python-based seismic data analysis package. Libraries for segmentation, classification and detection are programmed in NumPy and Scipy. The CLI provides access to these specific routines, with a set of auxiliary tasks, including the conversion of segmented results from NumPy to MATLAB, earthquake detection, picking routines and automated classification of volcanic signatures.

3. Earthquake detection and segmentation

The earthquake detection functionality allow the configuration of earthquake detection and picking based on STA/LTA (*Allen* (1982)), AMPA (*Álvarez et al.* (2013)) and sophisticated detectors for low frequency events based on wavelet decomposition (*Kanwaldip et al.*, 1997). In particular, due to the necessity to discriminate volcano seismic events with respect to regional and teleseismic earthquakes (*Giudicepietro et al.*, 2017), we have configured the STA/LTA and AMPA parameters, following (*Aki and Richards*, 2002), (*Küperkoch et al.*, 2012) and (*Álvarez et al.*, 2013).

Figure 2 shows results from the application of the STA/LTA triggering algorithm to a seismic swarm recorded at Mount St. Helens in 2004. Detection results can be further extended with automated segmentation with the REMOS (Recursive Entropy Method of Segmentation) algorithm (*Bueno et al.*, 2019a). REMOS is an end-to-end approach which uses the detection time as the anchor point of an exploration window to determine the exact boundaries (beginning and end of an event) based on seismic entropy. By

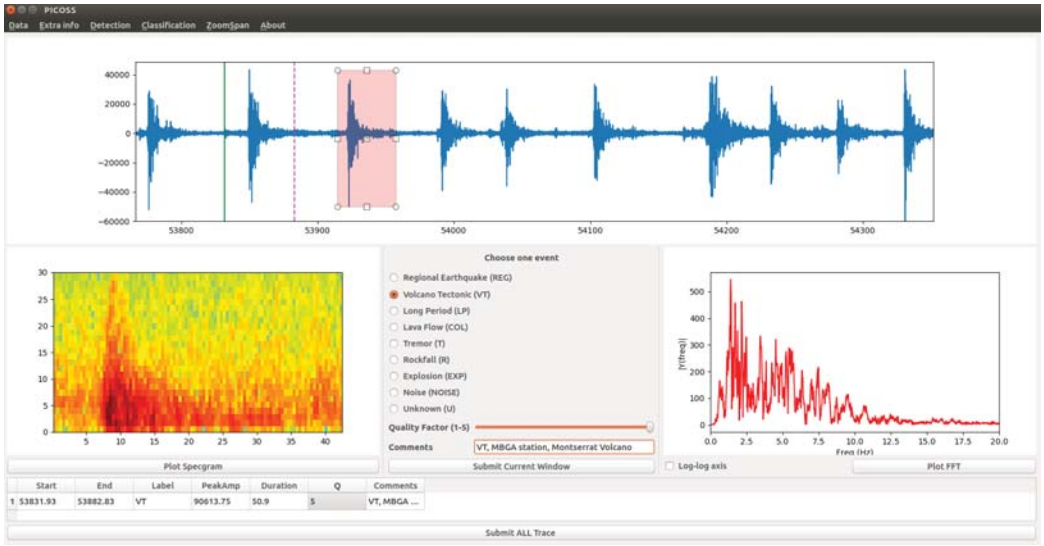


Figure 1: (a) Screenshot of PICOSS Main Interface, showing a sequence of manually segmented high-frequency events from Soufrière Hills Volcano, Montserrat, recorded on June 25, 1997, MBGA station.

taking measurements of the seismic energy, a minimum entropy criterion is used by REMOS to investigate large amounts of earthquake triggers and to discriminate and parse events into individual waveforms. REMOS configuration parameters shall be selected by the analyst to meet operational requirements specifically for each volcano.

4. Spectral Analysis

Whilst an analyst can manually segment and classify data from the main interface, PICOSS incorporates a semi-automated classification module based on spectral analysis. This classification is following the well-known categorization scheme proposed by (*McNutt et al.*, 2015). PICOSS automatic classification is performed using a pre-trained Bayesian Neural Network (BNN;

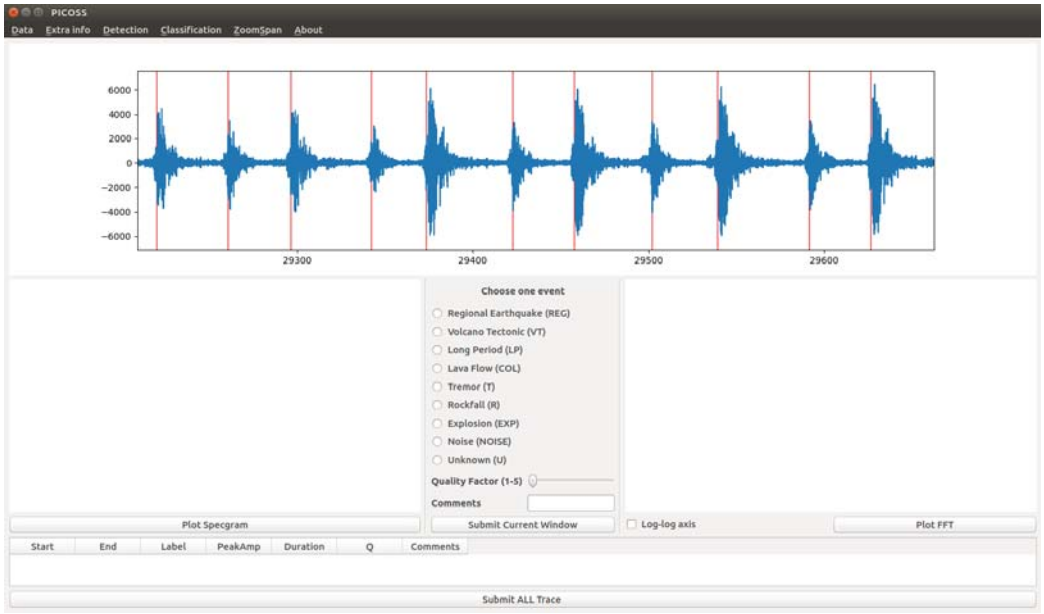


Figure 2: Example of application of STA/LTA, using the PICOSS interface, to continuous data recorded at Mount. St. Helens volcano on the 15th November, 2004, YEL station. PICOSS outputs a standard format which includes onset detection times. Alternative earthquake triggering methods (e.g, AMPA), along with automated data segmentation (i.e., extraction of earthquake waveforms with REMOS) are also included.

Bueno et al. (2019b)) and a frequency-index (FI) analysis (*Buurman and West*, 2010), in which the logarithm of the ratio of spectral energies over user-specified frequency bands is computed:

$$FI = log_{10}\left(\frac{A_{high}}{A_{low}}\right) \tag{1}$$

where A_{high} and A_{low} are the spectral energies above and over user-specified frequency bands. A set of threshold values $\eta = \{\eta_1, \eta_{hybrid\,low}, \eta_{hybrid\,high}, \eta_2, \eta_3\}$ are defined as the FI thresholds that control how events are categorized ac-

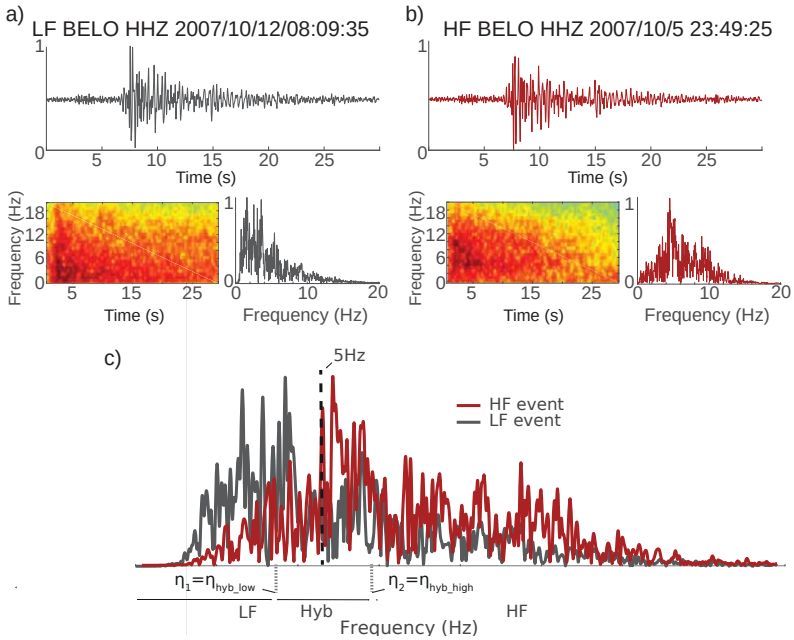


Figure 3: Figure 3.a and 3.b show seismograms, spectrograms and power spectrum for two segmented signals at BELO station. These earthquakes are classified as low-frequency (3.a) and high-frequency (3.b), respectively. Seismogram and power spectrum amplitudes are normalized to the unit. Figure 3.c shows high-frequency and low-frequency spectra along with the threshold selection for frequency index calculation (see main text). Mixed frequency events are confined within the interval $[\eta_1, \eta_2]$.

cording to their FI value. Events are classified as Low-Frequency (LF) if the FI is below the given threshold η_1. Similarly, if the FI its greater than threshold η_2, it is classified as High-Frequency (HF). The events with FI between η_1 and η_2 correspond to events with significant energy at both low and high frequencies and are categorized as hybrids. Parameters $\eta_{hybrid\,low}$ and $\eta_{hybrid\,high}$ can be configured to quantify hybrid events and overcome the limitation of a single threshold for HF and LF events (see Fig.3.c). The sufficient time t over which to consider low frequency or tremor depends of

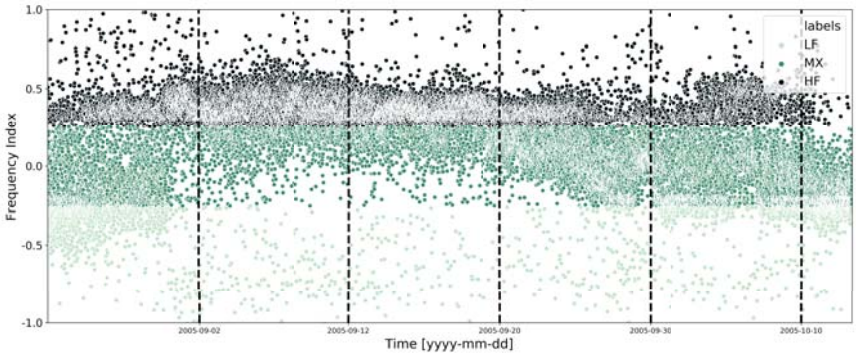


Figure 4: Temporal evolution of the frequency of detected events in the 2005 St.Helens eruption, station S02.

the analyst and the volcanic-environment (*McNutt et al.*, 2015). As tremor and rockfall require different frequency characterization as those of high/low frequency events, η_3 is a threshold to control whether an event is classified as a rockfall or tremor. For each event identified in the initial continuous trace, PICOSS extracts their waveform, computes the FI as in equation 1, assigns a label, and stores the results in *.npy* format.

The implemented module includes a deep neural network trained in a transfer learning setup on a multi-volcano dataset, with shared knowledge representations across similar seismic events (*Bueno et al.*, 2019b). This has the advantage of reducing data labelling time while allowing the model to generalize for new data samples rapidly, even under intense seismic activity. The trained network follows a Bayesian approach, in which the deterministic neural connections are substituted by probability distributions, thus bridging the gap between Bayesian modelling and deep learning architectures (*Neal, R.M,*, 2012). The extracted signals are characterized using a set of 13 cep-

stral coefficients on a logarithmic scale. This procedure is as follows: using a Hamming window (4.0 s), the spectrum of the seismic signal is computed, and a logspaced filter bank (16 triangular weighting function and 50%adjacency) is designed to compute the spectral frequencies. A cepstral analysis is performed to derive the 13 cepstral coefficients that best define each earthquake in the database. The network is trained following Adam optimization procedure, initial learning rate of 0.01, ReLU activation function, mini-batch size set of 32. The cross-entropy loss is used as a cost function for the classification of seismic events.

The computed uncertainty is exploited by PICOSS under an active learning setting: using a maximum entropy acquisition strategy, those segmented events with the highest classification uncertainty are selected (queried) to the user, which ultimately decides the samples that shall be used for re-training purposes. The included neural network is trained on very characteristic events from Bezymianny and St. Helens volcanoes, under the generalization scheme proposed by *McNutt et al.* (2015) (see Figure 3). Figure 4 depicts the temporal evolution of the FI for the volcano-seismic segmented events recorded at Mount St. Helens volcano, *S*02 station, using the pre-trained neural network. Only Q>2 are selected for this figure, as those below exhibit low SNR. Note the evolution of the frequency index over time; this type of analysis provides valuable, rapid, initial assessment of if and how seismic unrest evolves over time.

5. Output formats

The ultimate objective of PICOSS is to assist users to produce high-quality volcano seismic datasets. PICOSS stores earthquake segmentation and classification results by serializing the *segmentation table* (manual or automatic) in a high and efficient Python format, the *pickle* format. PICOSS saves the *start*, *end* times (in seconds), along with the assigned *label*, the *peak amplitude* ($PeakAmplitude$), the *duration*, quality Q values and additional comments for further post-processing. Additionally, PICOSS allows the user to further edit the *segmentation table*. Finally, *pickle* formats can be converted to NumPy arrays, MATLAB data formats (*.mat*), text CSV or numerical Pandas *dataframe* framework. The trigger times, along with the visualization, if required, are stored as NumPy arrays and can also be converted using the CLI utility functions included within *"convert data"* script. Each stream is saved according to its metadata within the seismic network, including the station, component, year, day and last segmentation time.

6. Conclusions

At present, a wealth of available seismic data acquired at volcanoes worldwide remain largely underutilized. In this paper, we present PICOSS, a Python Interface for the Classification of Seismic Signals. PICOSS is a modular open source software, with a graphical user interface designed for detection, segmentation and classification, focus on exportability and standardization of data formats. PICOSS includes functionalities that are programmed as independent modules that can be easily adapted for operational requirements at volcano observatories. The implemented modules are independent

from each other, and provide a collection of tools to analyse volcano-seismic data. The user can switch between automatic or manual modes by incorporating this suite of tools to efficiently compile complete catalogues of labelled volcano-seismic events. PICOSS could also be used for educational purposes, or refine other types of seismic data. All the modules of this interface are designed to reduce data labelling fatigue while increasing data classification efficiency.

7. Acknowledgments

PICOSS has benefited greatly from helpful feedback and advice from all the members at the Seismic Laboratory at the University of Liverpool and *Instituto Andaluz de Geofísica* (Granada). The authors are indebted to Prof. Jesús M. Ibáñez, without a doubt, the functionalities of PICOSS has been greatly expanded with his feedback. This interface is currently being used to produce the datasets used in the following projects: TEC2015-68752 and VOLCANOWAVES (Marie Sklodowska-Curie Grant Agreement no 798480). This research is funded by TEC2015-68752 (MINECO/FEDER), by NERC Grant NE/P00105X/1 and by European Union's Horizon 2020 Research and Innovation Programme Under the Marie Sklodowska-Curie Grant Agreement no 798480.

References

Aki, K., and Richards, P.G., (2002), Quantitative Seismology, Second Edition, University Science Books, Sausalito, California.

Abdelwahed, M.F., (2012), SGRAPH (SeismoGRAPHer): Seismic waveform analysis and integrated tools in seismology, *Computers and Geosciences*, *40*, pp. 153–165.

Allen, R., (1982), Automatic phase pickers: their present use and future prospects, *Bulletin of the Seismological Society of America*, *72* (6B), pp. 225–242.

Álvarez, I., García, L., Mota, S., Cortes, G., Benítez, C and De la Torre, A., (2013), An automatic P-phase picking algorithm based on adaptive multiband processing, *IEEE Geoscience and Remote Sensing Letters*, *10* (6), pp. 1488–1492.

Bell, A. F., and C. R. J. Kilburn (2012), Precursors to dyke-fed eruptions at basaltic volcanoes: insights from patterns of volcano-tectonic seismicity at Kilauea volcano, Hawaii, *Bulletin of Volcanology*, *74*(2), 325–339, doi: 10.1007/s00445-011-0519-3.

Beyreuther, M., Barsch, R., Krischer, L., Megies, T., Behr, Y., and Wassermann, J. (2010), Obspy: A python toolbox for seismology, *Seismological Research Letters*, *81* (3), pp. 530–533.

Bueno, A., Diaz-Moreno, A., De Angelis, S., Benitez, C., and Ibañez, J.M. (2019), REMOS: Recursive Entropy Method of Segmentation, *Seismological Research Letters*, (accepted), https://doi.org/10.1785/0220180317.

Bueno, A., Benítez, C., De Angelis, S., Díaz-Moreno, A., and Ibañez, J.M. (2019), Volcano-Seismic Transfer Learning and Uncertainty Quantification

With Bayesian Neural Networks. IEEE Transactions on Geoscience and Remote Sensing. (IEEE), pp. 892–902.

Buurman, H., and West, M. (2010), Seismic precursors to volcanic explosions during the 2006 eruption of Augustine volcano, *1769*.

Chouet, B., (2003), Volcano seismology. *Pure applied Geophysics*, *160*, pp. 739-778. doi:https://doi.org/10.1007/PL00012556.

Giudicepietro, F., Esposito A. M., Ricciolino P (2017). Fast Discrimination of Local Earthquakes Using a Neural Approach. *Seismological Research Letters*; 88 (4): 1089–1096. *doi: https://doi.org/10.1785/0220160222*

Goldstein, P., Dodge, D., Firpo, M., and Ruppert, S., (1998), What's new in SAC2000? Enhanced processing and database access, *Seismological Research Letters*, *69* (3), pp. 202–205.

Havskov, J. and Ottemoller, L., (1999), SEISAN earthquake analysis software, *Seismological Research Letters*, *70* (5), pp. 532–534.

Kanwaldip Singh, A., and Dowla, F., (1997). Wavelet transform methods for phase identification in three-component seismograms. Bulletin of the Seismological Society of America. *87* (6), pp. 1598–1612.

Küperkoch, L., Meier, T., Diehl, T., (2012). Automated Event and Phase Identification. New Manual of Seismological Observatory Practice 2 (NMSOP-2), Potsdam : Deutsches GeoForschungsZentrum GFZ, 1-52.

Lahr, J.C., Chouet, B., Stephens, C.D., Power, J.A., Page, R.A. (1994), Earthquake classification, location, and error analysis in a volcanic envi-

ronment: implications for the magmatic system of the 1989–1990 eruptions at redoubt volcano, Alaska. Journal of Volcanology and Geothermal Research, *62*, pp. 137–152.

Lesage, P. (2009), Interactive Matlab software for the analysis of seismic volcanic signals, *Computers and Geosciences*, *35* (10), pp. 2137–2144.

Luckett, R., Baptie, B., Ottemoller, L., and Thompson, G., (2007), Seismic Monitoring of the Soufriere Hills Volcano, Montserrat, *Seismological Research Letters*, *78* (2), pp. 192–200.

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 (Second Edition)* pp. 1071 – 1099, Academic Press, Amsterdam, doi:https://doi.org/10.1016/B978-0-12-385938-9.00063-8.

Minakami, T. (1974), Prediction of volcanic eruptions, *Developments in Solid Earth Geophysics*, *6*, pp. 313–333.

Neal, R.M., (2012), Bayesian learning for neural networks. *Springer Science & Business Media*, *118*.

Olivieri, M., and Clinton, J., (2012), An almost fair comparison between Earthworm and SeisComp3, *Seismological Research Letters*, *83* (4), pp. 720–727.

Roman, D. C. (2017), Automated detection and characterisation of harmonic tremor in continuous seismic data: Harmonic tremor detection, *44*

Romero, J.E., Titos, M., Bueno, A., Álvarez, I., García, L., De la Torre, A., Benítez, C., (2016), APASVO: A free software tool for automatic P-phase picking and event detection in seismic traces, *Computer and Geosciences*, *90*, pp. 213–220.

Zuccarello, L., Paratore, M., La Rocca, M., Ferrari,F., Messina, A., Branca, S., Contrafatto, D., Galluzzo, D., and Rapisarda, S., (2016), Shallow velocity model in the area of Pozzo Pitarrone, Mt. Etna, from single station, array methods and borehole data, *Annals of Geophysics*, DOI: 10.4401/ag-7086.

9 | Conclusion

This thesis has studied Bayesian deep learning as the core algorithmic component of seismo-volcanic monitoring algorithms. We developed unified machine-learning pipelines that blend ideas from seismology, signal processing, Bayesian theory, and neural networks to produce a broad range of monitoring models for sequential data. The conclusions extracted from this research work are presented in section 9.1. Section 9.2 highlights the contributions of this work, and section 9.3 explores open questions and future work.

9.1 Conclusions

While state-of-the-art techniques in other waveform data domains offer highly specialized deep learning routines for specific tasks, we identified a few steps behind productization. By drawing parallels to speech, or music recognition, we believe that the methods developed here fit very well in monitoring volcanoes or other seismological tasks, especially if we consider that no prior assumptions about the seismic data distribution or supervision are required in the detection of change. It is always a challenge to know precisely which of these techniques could be used as the core component of a monitoring data pipeline, but recent publications from seismological observatories are geared towards the methods considered in this thesis: CNNs, ConvLSTMs, TCNs, bi-LSTMs, and Bayesian approaches. Below, we provide the conclusions for each of the presented work, which we believe will strengthen the consideration of these methods as core components in modern observatories:

1. In **publication 1** [14], Chapter 4, we show that a BNN, trained on cepstral coefficients, can attain significant classification performance for three of the most representative volcano-seismic signals registered in two volcanoes (Mount Saint Helens and Bezymianny); high, low, and mixed frequencies earthquakes. When fine-tuning independent systems for each volcano, the proposed BNN attains an F1 score performance beyond 90% for each volcano, independently. Moreover, when the two datasets are combined into a larger one with the same labels, the BNN supersedes the previous F1 score. Then, when the datasets of both volcanoes are separated according to their volcanic origin, the proposed BNN trained over the sparse labels achieves high accuracy and can discern seismic events according to their recorded volcano. The seismic signals presumably linked to the same geophysical mechanism can be used with the same class to augment the datasets and increase the monitoring capacity, as data samples belong to a similar feature space. By drawing parallels with audio domains, similar seismic transients of different volcanoes can be treated as recordings of the same acoustic entity but under different conditions. On the other hand, these results show that the data taxonomy adopted in this work [42] is robust and offers intra-class separability with high accuracy for the models trained with this taxonomy.

   We then illustrate frequency content variations that are well-sensed by the uncertainty associated with the proposed BNN. The uncertainty derived from the proposed BNN has two main implications: it stands as a feature to be considered an unrest precursor and as a threshold level to determine when transfer learning algorithms should be used. Finally, we observe that the application of transfer learning allows exporting knowledge from one volcano or one stage to another and, simultaneously, can track how signals evolve, even for mixed seismo-volcanic datasets.

2. In **publication 2** [10], Chapter 5, we focus our problem on sequential data streams from three eruptions at Bezymianny volcano. Hence, we extend the previous approach and expand the discovered insights between non-stationary seismo-volcanic environments and uncertainty

quantification. We propose a hybrid end-to-end Bayesian convolutional neural network (B-CNN) to learn and detect the full intra-frequency dynamic range of seismo-volcanic events while performing seismic event recognition with a temporal modeling network. This architecture is based on a segmentation U-net type framework fused with a temporal convolution TCN (B-TCN). We tested on different eruptive periods, with very high frame-wise segmentation ($90\%in IoU metrics$)$and classification metrics$($80\%in one second frame - wise F1 - score$)

The designed framework yields a probabilistic approximation of the total uncertainty of the data streams, U_t, as the contribution of the uncertainty associated with seismic wavefield variations (reducible with more data) and the randomness of the monitored volcanic process. This formulation has allowed us to introduce *monitoring uncertainty maps* as an additional tool to illustrate how the estimated uncertainty behaves at the waveform level. The epistemic uncertainty acts as a complementary onset detector, whereas the aleatory uncertainty highlights the potential sources available in the seismic data stream.

Finally, the estimated short-term evolution of the uncertainty is consistent with the real-time seismic energy measurement, a direct waveform parameter independent from the estimated uncertainty. The temporal evolution of the uncertainties in pre- and post-eruptive periods is in good agreement with seismological bulletins for the studied eruption. Thus, changes in the seismic data streams due to volcano dynamics are detectable by the model as data drift, with the uncertainty of the model associated with volcanic unrest.

3. In **publication 3** [11], Chapter 6, we investigate the feature learning capabilities of deep learning methods and find if the uncertainty follows similar patterns as the previously discovered for the three eruptions at Bezymianny volcano. This work presents a novel end-to-end deep learning architecture embedded with a learnable scattering transform to implement multi-label detection, segmentation, and classification of seismic signals. The flexibility introduced in learning layer-

wise knots and filterbank design, jointly with specialized recurrent dynamic convolutions, yields optimal, robust feature representations in a frame-wise fashion. The application of this architecture to data from three volcanoes of different types and on different continents shows that this architecture generalizes well and appropriately adapts to different environments. Our neural network ensures prompt recognition of events and is robust against sparse data taxonomies and background noise, a very active research topic in acoustic machine learning applications.

Our deep recurrent scattering network departs from the traditional perspective and opens new directions for designing forecasting methods based on the connections between epistemic uncertainty and volcano dynamics. The power-law drift in epistemic uncertainty associated with seismic data streams implies that volcanic processes preceding eruptions are detectable. With no prior assumptions about signal distribution, deep learning can identify such behavior without supervision or parametrization by data alone. Further, for the Bezymianny volcano, the temporal evolution of uncertainty behaves similarly as the case studied in [14], already verified with seismological bulletins [63], [27]. When performing transfer learning on Mt. Etna volcano, the uncertainty at the main eruption is similar to that of the second eruption at Bezymianny. These results confirm that the Bayesian approach proposed in this thesis is robust, exportable, and holds promise for forecasting eruptions, although operational challenges remain. Nonetheless, this architecture can be modified to act on real-time streaming data; this goes in concert with developing a novel early-warning strategy.

4. In **publication 4** [12]; we investigate how the system can overcome new situations after a volcanic eruption after the data drift has decreased monitoring performance [12]. We propose an active learning framework based on TCNs. The total uncertainty associated with the model, from [12] is exploited to gain insight about how the model must select seismo-volcanic events within the continuous data stream, necessary for retraining. We noted that seismic data stream evolution influences the sampling of active learning procedures, shifting the dynamics of the sampling process. We incorporate cost-effective active learning to mitigate this influence, selecting cor-

rect, well-labeled examples as pseudo-labels. We note that for eight acquisition steps, equivalent to 4000 minutes of data curation in a seismological observatory, our active learning framework attains frame-wise F1 score recognition performance above 83.00% for all acquisition functions tested. Thus, we achieve similar performance to that of the baseline system trained with the entire dataset (288 hours) but with a fraction of the data. Hence, this approach answers the earlier claims on the necessity of high-quality data for fine-tuning monitoring algorithms.

5. In **publication 5** [13] and **publication 6** [15], albeit being research works devoted by their design to signal-processing-based geophysical applications, they comprise a significant, practical work of this thesis. First, the algorithm [13] had highlighted the complexity of the seismic signals to achieve correct segmentation and time boundaries, notably when no broad baseline seismo-volcanic data catalogs were available at the time of implementation. Dimensionality reduction and visualization techniques in [15] have demonstrated the necessity of supervisory control seismic data curation architectures comprising computers, networked accessible waveforms, and data labeling platforms for high-level seismic data processing.

9.2 CONTRIBUTIONS

In this thesis, we investigate the effectiveness of Bayesian deep learning methods for various monitoring tasks. These tasks can be listed as seismic waveform classification, temporal separation of seismic events from background noise, sequence recognition, seismic feature learning, and transfer learning. Further, we have undertaken Bayesian theory to analyze seismic data streams, aiming to detect change patterns before eruptions and characterize volcanic unrest. We thus have covered very early examples of how Bayesian deep learning methods can be applied in seismo-volcanic monitoring. The main contributions of this thesis are summarized as follows:

Design of three novel architectures to extend the current state-of-the-art monitoring in volcanoes. Exportability of the implemented architectures across different volcanoes and eruptive

styles.

BNNs, TCNs, hybrid architectures, and the proposed recurrent scattering neural networks provide a significant numerical advantage in monitoring performance, with the ability to capture the uncertainty within the learning model. All these methods have been blind-tested across different eruptive periods or volcanoes with different geological conditions. The results confirm the exportability of our frameworks.

The evolution of uncertainty is a good indicator for volcanic early warning.

The uncertainty estimation has led to the discovery and identification of the data drifts before, during, and after volcanic eruptions, regardless of the features used. In terms of eruption prewarning time, uncertainty is ahead of other monitoring observables, rendering it potentially more effective for volcanic alert systems. The laws of change for this uncertainty are exportable from one system to another. Independent of the framework or features used, the temporal evolution of the uncertainty with other monitoring parameters or visual evidence reported in seismological bulletins implies that uncertainty is an excellent early-warning descriptor.

Overall uncertainty as a descriptor to know when monitoring algorithms need to be retrained.

The monitoring algorithms that operate in volcanic settings need to incorporate mechanisms to detect, adapt and overcome evolving data changes over time; otherwise, the performance accuracy will degrade. As time passes, the algorithm needs to be updated by considering operational requirements or be replaced entirely to meet the new situation. We proposed the uncertainty as a *threshold* of the model from the properties of the classifier and the data. Our three implemented networks have demonstrated the capacity to detect seismic changes and adapt when required via transfer or active learning.

formulation of a Bayesian monitoring framework to associate and identify which uncertainties are due to seismogram variability or the monitored process.

The assessment of change and early warning outcomes is incomplete without knowing what is

driving the change. We had proposed a time-uncertainty representation, where we were able to identify onsets and sources, with notable phases of uncertainty related to the continuous volcanic tremor.

Identification of power-law behaviour. Estimated uncertainty holds the potential for volcano forecasting.
This work has characterized *when* and *how* the dynamics of the volcano are changing. However, challenges remain in the accurate forecasting of volcanic eruptions. Remarkably, the discovery of power-law drift in the uncertainty associated with seismic data streams implies that volcanic processes preceding eruptions are detectable, and thus, advanced procedures can be implemented to perform forecasting.

The duality of the uncertainty discovered by shifting volcano dynamics and active learning.
The estimated uncertainty in volcanoes does have a double side. In the active learning frameworks, the temporal evolution of uncertainty implies that the monitoring algorithms tend to sample those seismic waveforms in which the data drift occurs. Therefore, the initial support of the known data distribution, T0, is shifted towards the newly acquired data, T1, and *forgets* the previous learning in T0. The next phase reacquires signals from T0, which yields a recurrent problem of back and forth time sampling. Cost-effective active learning strategies maintain the uniform sampling of events to attain significant improvements over the trained baseline with the entire dataset.

9.3 FUTURE WORK

Based on the research work conducted herein and observations made in the process, this thesis raises more questions than answers. In the spirit of guiding future research work, we would like to highlight several perspectives on deep bayesian learning and seismo-volcanic monitoring, with excursions to other acoustic domains, which we believe are crucial to advance the monitoring

science further.

9.3.1 VOLCANO FORECASTING

The appearance of seismic swarms or the increment of background tremors are precursors that often precede eruptions at volcanoes worldwide [20]. This thesis has addressed algorithms to detect seismic change ending in eruptions from continuous data streams. However, we did not mention another critical problem: *how does the uncertainty behave for seismic sequences that do not end in an eruption*?. Many eruptions can occur without such precursors, and there exist certain volcanoes in which such behavior is not observable. Thus, a persistent challenge to build generalizable forecasting models to quantify the probability of an eruption is a lack of datasets containing failed eruptions.

A failed eruption can occur due to the intrusion of exogenous elements in the volcano or seasonal effects. Volcanoes are placed in geological areas containing water, subterranean aquifers, or melting ice. These elements increase tremor levels and accelerate seismic rates that do not end in an eruption [20]. Although these issues are commonly known, efforts to quantify from a forecasting perspective remain very limited. The developed Bayesian monitoring framework can help tackle this challenge by identifying exotic, non-volcanic signals via *uncertainty-maps*; and through the exploitation of the uncertainty to identify seasonality of the time series. Then, additional models can ingest this information and be trained to discern eruptions from seasonal effects or failed eruptions. An accurate prediction of volcanic eruptions would benefit from three themes of this thesis: recurrent scattering, end-to-end deep learning, and uncertainty.

9.3.2 EXPORTABILITY OF THE BAYESIAN METHOD

Uncertainty monitoring can be a universal indicator of change that is not unique nor specific to seismic monitoring. The developed Bayesian deep learning workflows can also be applied to acoustic surveillance in which the prompt identification of sounds outside the common acoustic background is an operational requirement. For example, condition-monitoring methods that

rely on data collections and sensors are difficult to implement due to the background noise in the industrial environment. Environmental sound identification with multi-source sounds or polyphonic music applications are acoustic data domains in which our deep learning methods can be further explored. We provided an example in Appendix A, but other refined architectures with the methodology developed in this thesis can enhance the monitoring outcome.

Another exciting application is speech recognition, as human biology can significantly impact speech production: accent, articulation, pronunciation, nasality, pitch, speed, or spoken language, to name a few examples. Besides, human speech is often influenced by emotions and external causes, including reverberation, distant sources, sound attenuation, or noise contamination. Drawing parallels to seismic monitoring, waveform alteration is a recurrent problem that our architectures are obliquely designed to cope with. Applying the developed models to these complex and challenging tasks can drive the research on many of these acoustic domains.

9.3.3 Refining seismic data domains

From a transfer learning perspective, the hidden states of a neural network can be used across different data domains to identify shared activation patterns in the *hidden* layers. These patterns of neural activities subsequently cause a high, secondary level of activation in the hidden units for waveforms that share common frequencies and waveform attributes. Thus, by providing means to extract distributed patterns, unsupervised clustering can find families of events across different domains, for example, to know what infrasonic and seismic domains have in common.

As demonstrated in this research, volcanoes yield a different data distribution due to data drift after eruptions. The infrasound domain can be treated similarly: an acoustic domain that shares properties with the seismic one. Hence, from a transfer learning lens, we raise the possibility of how exportable is a system trained in seismic signals to a data stream of infrasonic waveforms, and similar classification schemes such as those in volcanoes case correlate with infrasound. These ideas are also extensible to poorly monitored volcanoes, and the real classes remain largely unknown. This would undoubtedly help to create infrasonic data catalogs and

explore the similarities between volcanic sources.

9.3.4 Refinement of architectures

It has been highlighted in this research that the developed algorithms attain very high performance in a gamut of tasks and configurations. Furthermore, while the theoretical and empirical benefits of our models have been thoroughly explained and tested, we have built upon the knowledge of well-known machine learning guidelines in the seismo-volcanic monitoring field in terms of hyperparameter fine-tuning, feature extraction, or waveform input selection. In the future, it may be wise to spend research effort finding answers to questions such as: how can we adapt the long-term temporal modeling over more extensive sequences of seismic data streams? How can we change the segmentation configurations for the temporal modeling of seismic events? Can we apply very sparse data taxonomies? While the implicitly answer the first question in Appendix A, with the benchmarked data (in essence, a waveform at a higher sampling rate, and thus larger number of time samples), the author believes these research directions would be crucial for developing better and more robust monitoring algorithms in the future.

9.3.5 Data fusion techniques

Data fusion is defined as the aggregation of multiple data sources to produce more precise, scalable, and valuable information than that provided by any individual data source. Based on the processing stage in which they are fused, the level of refinement in which features are aggregated, data fusion techniques can be classified as low, intermediate, or high. As a result, data fusion techniques can increase the applicability of the Bayesian methods developed here, for example, by identifying signs of unrest using radar satellite images and seismic data streams or after processing embeddings from both camera-based and seismic surveillance methods.

A high-definition camera (with thermal capabilities) can survey the volcanic cone looking for ash clouds, degassing, fumaroles, or volcanic deformation that can precede an eruption. From the frameworks developed here, the degassing images do not correspond to the known data distribu-

tion by the model. Again, we are tackling an *out-of-distribution* problem in which the uncertainty can be exploited to trigger an alarm. If fused with seismic data streams, we can refine the uncertainty obtained and gain compelling visual evidence that if only the seismic stream is used for this task. Similarly, these camera-based systems are also exportable to radar satellites that can monitor the volcanoes and their surroundings to detect hot spots that correlate to new variations in the seismic stream. Remark that the deep scattering network can be applied to images as the scattering operator takes multiple convolutions. As a direct application, the proposed recurrent scattering can analyze the multiple sequences of video frames and yield a feature vector representation that can be fused in the *feature-fusion* module with the ones obtained by the seismic analysis. The fused vector can be parsed to obtain the required monitoring outcome.

A | Appendix: Benchmark results recurrent scattering network

We evaluated the performance of our neural network on two publicly available audio data sets: AudioMNIST for audio classification and URBAN-SED for polyphonic sound event detection [1]. Our SED analysis is focused on the most challenging task: identifying multiple time onsets and segmentation of sounds of interest. These benchmark tests demonstrate the generalizability and exportability of our approach to audio tasks beyond seismo-volcanic monitoring. [2] On both data sets, we trained on the available raw wave forms; that is, we did not apply any pre-processing. We refer as *vanilla* to the model in which no spline approximation is computed, *filter* to learnable filters and *filter-knots* to the learnable knots approximation

A.1 Audio classification: AudioMNIST

This benchmark data set consists of 30000 audio samples of spoken digits $(0-9)$ of 60 different speakers, with 50 repetitions for each speaker, recorded at a sampling frequency of 48kHz [3]. We

[1] Online datasets:
AudioMNIST: https://github.com/soerenab/AudioMNIST.
URBAN-SED: http://urbansed.weebly.com

[2] This research section has been done in collaboration with R. Balestriero, R. Baraniuk and M.V. de Hoop.

adopt the same configuration scheme as the baseline benchmark, [3]. Our network was trained with an *Adam optimizer* and a learning rate of 0.001. The batch size was 32 and training was carried out for 50 epochs with a validation split of 20%.

A.2 Polyphonic sound event detection: URBAN-SED

This benchmark data set comprises a total of 27.8 hours of audio organized into 10,000 soundscapes of $10s$. duration and sampling frequency of 44.1kHz. Each soundscape contains between 1˜9 foreground sound events from the UrbanSound8k data set, mixed with a background Brownian noise or *hum* as heard in urban environments. Training, validation and test sets are pre-sorted in this audio data set, comprising 6000, 2000 and 2000 soundscapes, respectively.

A hybrid C-RNN trained on 40 log-normalized mel-based filterbanks computed over 40ms with 50% overlapping has been the established baseline for the URBAN-SED dataset [57]. In another architecture proposed by [41], a family of adaptive pooling operators has been embedded in a dynamic CNN to promote smooth label interpolation so that the model can adapt to the characteristics of the sound sources. In our architecture, the time resolution of the wavelets to achieve a specific target accuracy can be adjusted via the pooling of the scattering transform [1], in a similar fashion as explained in [12].

A.3 Recurrent scattering network training

Our neural network architecture (Chapter 6, [12]), contains two paths or branches. The first, named path A, analyzes via dynamic convolutions a sequence of local scattering coefficient variations, outputting a refined feature map. The second, path B, is a bypass skip connection with a learnable convolution to perform data reduction and per-frame time analysis to capture precise onsets of events. Switching between audio classification or sound detection and seismo-volcanic

Table A.1: AudioMNIST digit classification benchmark

System	Acc (%)
AlexNet (Spectrogram)	95.82
AudioNet (Waveform)	92.52
Vanilla	**99.05**
Filters	**99.22**
Knots + Filters	**99.12**

monitoring only involves changing the numbers of filters in the scattering layers.

Urban-SED and AudioMNIST: For the *vanilla* configuration, the number of knots k, the number of octaves J and wavelets per octave Q are chosen as follows. *First scattering layer L1*: $k = 4$, $J = 5$ and $Q = 8$; *second scattering layer L2*: $k = 2$, $J = 2$ and $Q = 5$. After the concatenation of both scattering feature vectors, a pooling operation of size $(1, 1, 441)$ is applied to reduce redundant information and align the processed feature vectors with the time resolution. Standard dropout with a factor of 0.5 is applied at the conv-LSTMs and BiLSTMs layers of the network. Both $ConvLSTM_\lambda$s following path A contain 16 filters, a kernel size of $(5, 5)$ and strides $(1, 1)$. The time skip connection following path B contains a single filter with a kernel size of $(5, 5)$. A dense layer with a sigmoid activation function with the number of classes (here, 10) as hidden units is used as predictive layer.

The training is done, for all the data sets, with the raw audio waveforms. In URBAN-SED, we down-sample the data by a factor of 2 to reduce redundancy while keeping memory efficiency. We adjust the precision of the wavelets to achieve a specific time resolution via the pooling of the scattering transform. In URBAN-SED, the scattering transform module produces a sparse vector of 22050 samples pooled by a factor of 441 for a window overlap of 20ms. We selected the *Adam* optimizer with an initial learning rate of 0.01 and halved it every 10 epochs, and early stopping criterion of 100 epochs if no further improvement over the F-measure over one-second resolution was obtained.

Table A.2: URBAN-SED segmentation metrics benchmark

System	F1 (1s)	PR (1s)	RC (1s)
Max [41]	0.463	0.774	0.330
RAP [41]	0.526	0.650	0.442
CAP [41]	0.533	0.622	0.466
Auto Pooling [41]	0.504	0.738	0.382
C-RNN [57]	0.568	0.533	0.607
CNN [57]	0.569	0.691	0.483
Vanilla	0.571	0.544	0.601
Filters	0.568	0.539	0.601

A.4 AUDIOMNIST RESULTS

Table A.1 contains the classification results for the digit identification task, benchmarked with two different input feature vectors, spectrograms and raw wave forms, against two popular audio tagging networks, AlexNet and AudioNet. The classification results show a relative improvement of 3% over the spectrogram inputs, and 7% over the raw wave form representation. The performance of our network can surpass the specifically designed networks, with competitive performance in classification audio tasks.

A.5 URBAN-SED RESULTS

Table A.2 contains the comparative results; the RC, PR and $F1$ metrics shows that our neural network is able to pick up events whilst maintaining high accuracy and recognition rate. Note that our model model is able to detect events correctly since its RC is more significant in all cases. On the other hand, the adaptability of our architecture is made manifest. Even without adaptative pooling nor label smoothing, it attains higher recall and F1-metrics than very specialized architectures.

B | Appendix: Additional results

This appendix of the thesis presents additional results that were not included in the publications due space constraints. However, the author believes these experiments shall be included here to support previous experimental research.

B.1 Hybrid U-NET comparative table

There exist a plethora of algorithms to classify seismic events, very well-known and tested, with high classification accuracy [56]. The work presented in this thesis focuses mainly on the Bayesian approximation and uncertainty exploitation to gain scientific insight into impending unrest. In [10], we focus on learning the background tremor while performing change detection. This double task renders to be very challenging in static supervised learning scenarios with deterministic algorithms. Moreover, the uncertainty approach in Chapter 2 is formulated within the convolutional neural network frameworks, in which its applicability to seismic signatures, volcanic or not, has already been very well tested. This section of Appendix.B includes a comparative table of each of the components in [10], independently. Here, we refer as TCN as the temporal sequence classifier part B of our architecture in [10]. Hence, the number of classes are 4: HF, LF, DP and SBT. We incorporate the best LSTM proposed in [64] (128 hidden units), with the same hyperparameter configurations. Then, the encoder part of our hybrid u-net architecture, that is, the encoder with 16, 32 and 128 filters, is concatenated with a fully dense layer. To do this, we had to design the encoder part by down-sampling only in frequency (the y-axis component),

Table B.1: Comparative classification performance for different architectures.

System	test b-QP				test b-E2			
	PR	**RC**	**F1**	**Acc**	**PR**	**RC**	**F1**	**Acc**
Encoder+Dense	0.88	0.87	0.87	0.88	0.81	0.80	0.80	0.80
LSTM	0.92	0.91	0.91	0.92	0.88	0.87	0.87	0.87
TCN	0.93	0.94	0.93	0.93	0.84	0.83	0.83	0.83
Ours	**0.97**	**0.95**	**0.96**	**0.97**	**0.91**	**0.87**	**0.89**	**0.89**

thus changing the down-sampling operation from (2×2) to (2×1). The output is flattened and forwarded to a bottleneck hidden layer with 512 – 128 – 64 hidden units. A time-distributed layer is used to make frame-wise predictions for each time step.

Following [10], we report the b-QP and the b-E2 systems. The quiescent system, b-QP, is the less noisy period, whereas the b-E2 is the seismic period with the highest noise. Training and testing on the relatively mild QP seismic activity can highlight the classification capabilities of the implemented architecture. Similarly, the b-E2 can provide knowledge about how these systems surpass the tested approaches with noise. We follow the exact same training procedure as in [10], with 75% - 25% for the b-QP and the b-E2 periods.

As shown in Table B.1, the proposed model in [10] presents a very high PR and RC for the b-QP period; which is also the case for the noisy period of the b-E2. These results highlight that learning the mask provides the classifier with an enhanced, denoised representation in which the model can perform very accurate classification, despite the higher noise levels. Note that these results in the classification part of our architecture only show the metric for that task, but not the results from the segmentation module, related to event detection (basically a noise vs event), which from [10], remains high.

B.2 TCN FINE TUNING PROCEDURE

In the manuscript [11]; we have shown that a TCN trained with AL can achieve very similar accuracy to that of baselines levels, but with fewer annotated examples. Here, we show additional results for the fine-tuning procedure on the selected dataset. The best model are selected based on

the trade-off between complexity and error rate, that is, we favor low error rate for less complex neural structures.

Table B.2: TCN fine-tuning on the period from 17th October to 29th October, Bezymianny, 2007.

Hyperparameters	Type of convs	Dilations	F1	PR	RC	ER
128, 1 stack	*Causal + Skip*	*1, 2, 4, 8*	85.47	88.19	82.92	0.176
		1, 2, 4, 8, 16, 32, 64	85.98	89.59	82.64	0.175
	Non-Causal + Skip	*1, 2, 4, 8*	85.58	88.16	83.14	0.175
		1, 2, 4, 8, 16, 32, 64	85.44	88.31	82.74	0.180
	Causal + Non Skip	*1, 2, 4, 8*	84.20	86.51	82.01	0.197
		1, 2, 4, 8, 16, 32, 64	84.78	88.71	81.19	0.188
	Non-Causal + Non- Skip	*1, 2, 4, 8*	83.91	85.55	82.33	0.201
		1, 2, 4, 8, 16, 32, 64	85.51	88.51	82.70	0.181
64, 1 stack	*Causal + Skip*	***1, 2, 4, 8***	**85.61**	**89.06**	**82.48**	**0.176**
		1, 2, 4, 8, 16, 32, 64	85.50	89.16	82.13	0.179
	Non-Causal + Skip	***1, 2, 4, 8***	**86.14**	**89.37**	**83.14**	**0.171**
		1, 2, 4, 8, 16, 32, 64	85.94	90.07	82.17	0.180
	Causal + Non Skip	*1, 2, 4, 8*	85.38	88.55	82.53	0.178
		1, 2, 4, 8, 16, 32, 64	84.74	88.18	81.55	0.185
	Non-Causal + Non- Skip	*1, 2, 4, 8*	85.67	88.99	82.59	0.178
		1, 2, 4, 8, 16, 32, 64	85.56	89.09	82.30	0.180
32, 1 stack	*Causal + Skip*	*1, 2, 4, 8*	85.42	88.99	82.12	0.180
		1, 2, 4, 8, 16, 32, 64	85.25	88.68	82.08	0.180
	Non-Causal + Skip	*1, 2, 4, 8*	85.70	89.11	82.55	0.176
		1, 2, 4, 8, 16, 32, 64	85.62	89.29	82.24	0.178
	Causal + Non Skip	*1, 2, 4, 8*	85.77	88.91	82.85	0.174
		1, 2, 4, 8, 16, 32, 64	85.22	88.53	82.15	0.180
	Non-Causal + Non- Skip	*1, 2, 4, 8*	85.79	88.94	82.86	0.174
		1, 2, 4, 8, 16, 32, 64	85.23	88.62	82.09	0.180

BEST MODELS ON 2 STACKS (two TCN blocks)

Hyperparameters	Type of convs	Dilations	F1	PR	RC	ER
64, 2 stack	*Causal + Skip*	*1, 2, 4, 8*	85.14	87.61	82.8	0.173
	Non-Causal + Skip	*1, 2, 4, 8*	85.86	88.93	83.0	0.171

B.3 EVENT CLASSIFICATION - FINE TUNING

In the manuscript [14]; we have shown that a BNN trained on log-cepstral coefficients attains excellent performance in event identification on data gathered at Bezymianny and Mount Saint Helens. Here, in table B.3, we present additional results for the fine-tuning procedure on these datasets. We show the best models obtained after following the optimization procedure explained in Chapter 2, trained with log-cepstral input feature vectors [14].

Table B.3: Best models seismic event classification on isolated, mixed and sparse datasets

Volcano	Features	Best models	Test Loss	Test Acc
Bezymianny, 3 classes	*CC*	256-512-256-2000	0.15	0,939
		2000-64-128	**0.14**	**0.941**
		128-1024-512	0.21	0,927
		64-512-256-256	0.18	0.930
Mount St-Helens, 3 classes	*CC*	128-1024-2000-64	0.15	0.948
		1024-128-64	**0.10**	**0.957**
		512-128	0.11	0.955
Both volcanoes, 3 classes	CC	512-64	0.143	0.946
		1024-128-64	**0.142**	**0.946**
		1024-1024-256-64	0.164	0.936
Both volcanoes, 6 classes	CC	1024_64	0.22	0.93
		512-256-128	**0.21**	**0.92**
		1024-512-256-128	0.26	0.91

C | MODEL INTERPRETABILITY

In this section of the thesis, we provide additional results to provide interpretability to the recurrent scattering network. The application of deep learning in acoustic monitoring has long been considered as a black-boxes without interpretability. However, the operational monitoring requirements shall provide a compelling gateway to view and analyze their inner workings. This appendix demonstrates that our network captures the dynamic range of events through visualization of the internal layers. The results obtained in seismo-volcanic monitoring (and by extension, to audio) can be interpretable tracked through the internal connections of the network.

C.1 THE RSN VOLUMETRIC LAYER

The RSN volumetric layer learns the best arrangement possible for the coefficients from the scattering layer. From [12], the first scattering layer, known as $\lambda_{1,n}$ only produces, by definition, a single feature map. This set of scatter-grams is further concatenated and organized in a volumetric layer for the second and subsequent layers. This module arranges each $\lambda_{n,m}$ into λ_v, a dynamic structure that presents the variations of the scattering coefficients frequency for the events of interest, which implicitly also has information about the location in time. Hence, this structure is arranged in a volumetric fashion. Figure C.1 depicts the volumetric layer representation for a signal in the URBAN-SED dataset.

We can see that the set of all maps contains a dimensional shape equivalent to $(N, J \times Q, time, channels)$; where N is the number of seismic signals per batch, $J \times Q$ the scales components, $time$ given as the

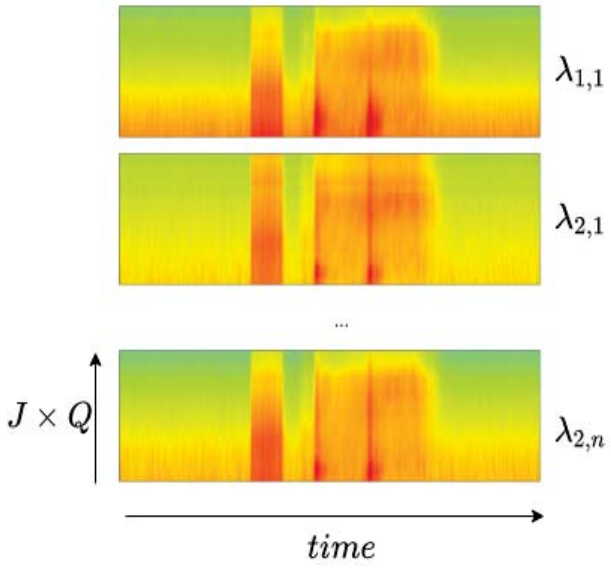
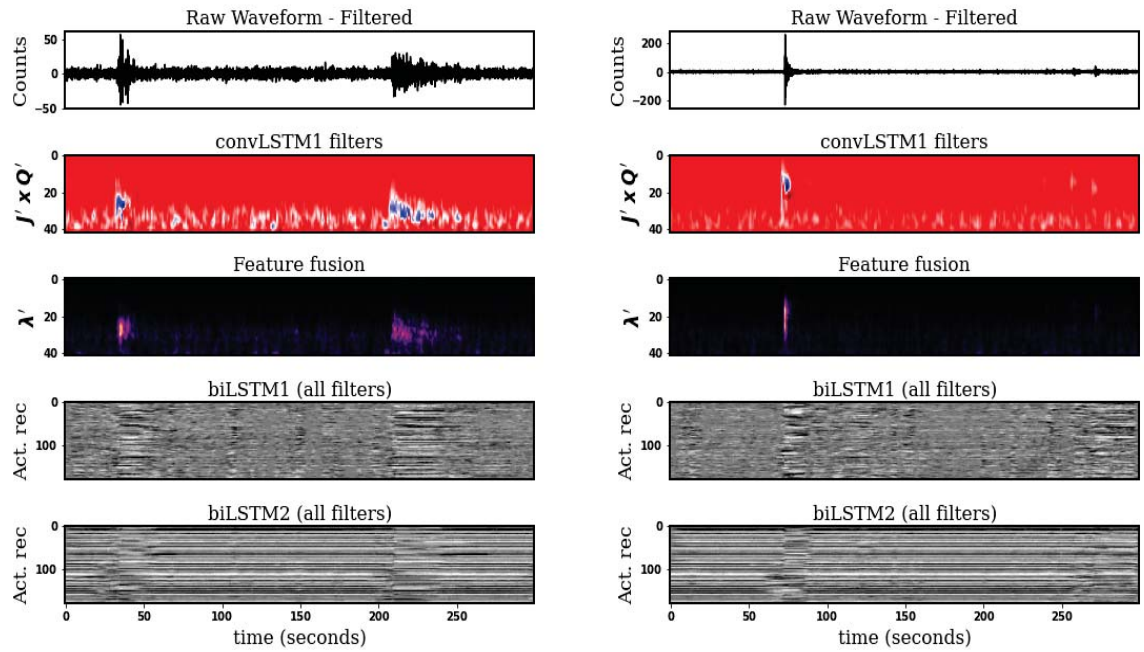


Figure C.1: RSN volumetric layer for the learned filters, and on a random signal sampled from the URBAN-SED dataset. The ground truth labels are jackhammer, dog bark, and drilling.

(number of time samples) and $channels$ each of the $\lambda_{n,m}$ independently. Recall that the volumetric layer can be seen as data augmentation, as each $\lambda_{n,m}$ codifies a set of invariant features within the raw signal.

C.2 RSN VISUALIZATION

Once we have shown that our RSN network can learn an input feature representation (see figure C.1) from audio data, we then visualize the middle components of the RSN network proposed in [12], but the application context of this thesis. In figure C.2.(a), we show a map of processed scattering feature vectors for convLSTM1, the output of the feature fusion module (before the classification), and the internal activations of the temporal classification module, for a random signal of the Bezymianny, 2007. The system that outputs these feature maps corresponds to the so-called *system eruption 1, with Knots+Filters* in [12], with the training data covering the period from September 1st to September 24th. Thus, we can verify that, for the learned scattering, the neural network can detect and track all the processed scattering feature vectors that defined these



(a) **(b)**

Figure C.2: RSN intermediate steps for a set of signals recorded at Bezymianny.

two events over time. On the other hand, it is observed that the refinement of our convLSTMs, together with the skip connections, produces an excellent feature representation to be exploited by the temporal classification module. Remarkably, the biLSTM1 in the temporal classification module can detect some small scattering contributions at a shallow scale, but these are indeed smoothed out and eliminated by the second biLSTM1 before the final classification module. We can see similar behavior with the signal in figure C.1.(b) the very energetic waveform shows different feature fusion maps, a consequence of the convLSTMs tracking much higher frequencies captured by learned scattering feature vectors.

D | Appendix: Tesis en español

Este apéndice en español se incluye como parte de la Memoria de tesis con la finalidad de cumplir con la normativa de elaboración proveniente de la Escuela de Posgrado. Este resumen debe incluye las siguientes secciones: introducción, objetivos, estructura de la memoria, metodología y conclusiones. Para la introducción, objetivos, estructura de la memoria, se ha empleado el texto del capítulo 1 (*Chapter 1*) de la tesis. Para la metodología, se incluye un resumen de los trabajos científicos. Por último, las conclusiones se han escrito en base al capítulo 9 ((*Chapter 9*).

D.1 Introducción

D.1.1 Motivación

Desde una perspectiva de aprendizaje automático, el dominio de datos sísmicos y volcánicos es escaso, dado que la disponibilidad de registros sísmicos de acceso abierto a la comunidad científica y que incluyan erupciones es limitado. Sin embargo, existe un creciente interés en obtener estos conjuntos de datos para mejorar los sistemas de monitoreo e incluir aplicaciones.

La complejidad volcánica implica años de estudio por parte de expertos para suministrar a la comunidad científica modelos teóricos que expliquen los observables las erupciones volcánicas. Los modelos teóricos producen taxonomías de datos escasas, no uniformes. La validez de un modelo teórico en un volcán no implica una generalización a otros entornos volcánicos. Por último, las barreras económicas condicionan los equipos electrónicos disponibles para el monitoreo: no todos los volcanes del mundo son monitoreados de la misma manera, llegando incluso

a sobrescribir registros almacenados en favor de otros períodos de interés sísmico. Todas estas deficiencias resultan en catálogos de datos volcánicos bastante modestos, limitados en el tiempo, y a menudo no utilizables desde una perspectiva de aprendizaje profundo. Como resultado, los desafíos de monitoreo asociados a la dinámica evolutiva de la serie sísmica no han recibido la atención suficiente. En una aplicación crítica para la seguridad humana, como es la vigilancia sísmica, el algoritmo de monitoreo puede generar predicciones en las que el modelo "piensa" que son correctas, a pesar de que la evidencia visual indica todo lo contrario. La detección de cambios y la interpretación de los resultados es uno de los pilares fundamentales sobre los que debemos construir los sistemas de monitoreo.

Esta tesis aborda estos retos y desarrolla una metodología que permite la detección de cambios en los volcanes mediante una aproximación bayesiano. Proponemos un conjunto de arquitecturas novedosas que incorporan el aprendizaje Bayesiano como una parte esencial del modelado de datos. Se proponen procedimientos de aprendizaje por transferencia (transfer learning) para facilitar el aprendizaje, la inferencia y la predicción a través de volcanes donde no existe ningún conocimiento previo disponible. También exploramos la capacidad de nuestros algoritmos para seleccionar mediante incertidumbre el conjunto óptimo de señales sísmicas, con la finalidad de reducir esfuerzos de anotación a la vez que se aumenta la adaptabilidad del sistema en condiciones cambiantes. Finalmente, la incertidumbre de la modelo condicionada a los datos de entrada es considerada como un potencial precursor y una herramienta de forecasting exportable a través de los volcanes.

D.1.2 OBJETIVOS

El objetivo principal de esta tesis es investigar, construir y ampliar las técnicas de aprendizaje automático recientemente propuestas en el contexto de sistemas de reconocimiento volcánico. Por otro lado, nuestro objetivo es investigar cómo las técnicas Bayesianas pueden integrarse en un marco de monitoreo volcánico que emplee la señal sísmica en bruto, aprendiendo aquellas características que mejor definen los datos de entrada. Aunque el principal campo de aplicación

de esta tesis es monitorización sísmica y volcánica, los algoritmos introducidos en esta investigación son exportables a otros dominios de datos acústicos. En el Apéndice A , presentamos dos ejemplos de aplicación para nuestra nueva arquitectura recurrente de scattering con datos de voz (audioMNIST) y el monitoreo acústico ambiental (UrbanSED). La amplia gama de tareas de monitorización abordadas en esta tesis, presentan una idea clara de la escalabilidad y robustez de las técnicas presentadas. Además, nuestro objetivo es proponer una detección de cambios basada en la incertidumbre del flujo de datos sismo-volcánicos, comprobando la exportabilidad de la incertidumbre en sistemas de alerta temprana. Por último , se exploran estrategias para mitigar los problemas de escasez de datos mediante técnicas de active learning y de transfer learning.

D.1.3 Estructura de la memoria

La estructura de esta tesis está organizada de la siguiente manera:

1. Los antecedentes teóricos sobre los métodos bayesianos de aprendizaje profundo y la estimación de incertidumbre, junto con la metodología propuesta, se presentan en el capítulo 2.

2. La disciplina científica del sismo-volcánica y, cómo abordamos la creación de los catálogos de datos, clase de eventos y entorno de monitoreo Bayesiano se presenta en el capítulo 3.

3. El capítulo 4 describe la red neuronal profunda bayesiana implementada para la clasificación de eventos pre-segmentados sísmicos. Los vínculos entre la incertidumbre y la dinámica volcánica son presentados en este capítulo. El aprendizaje de transferencia se utiliza para exportar el rendimiento del sistema a través de volcanes, y tras cambios producidos en una erupción.

4. En el capítulo 5, presentamos el trabajo de investigación con el modelo híbrido de monitorización bayesiana, creado para detectar y segmentar señales volcánicas. La conexión entre la incertidumbre y la forma de onda, junto a la variación temporal de la incertidumbre, son presentados en este capítulo.

5. En el capítulo 6, introducimos una arquitectura recurrente multimodular para realizar la detección, segmentación, y la clasificación de señales sísmicas mediante una aproximación multi-fuente. La arquitectura es novedosa puesto que aprende los mapas de características más relevantes para la captación de estos eventos. También proporcionamos aplicabilidad de la arquitectura y la incertidumbre estimada a otros volcanes.

6. En el capítulo 7, presentamos el enfoque de aprendizaje activo y la estrategias propuestas para realizar un reentrenamiento en el sistema, aumentando la adaptabilidad a las nuevas condiciones cuando el volcán ha cambiado.

7. En el capítulo 8, se incluyen las aplicaciones geofísicas desarrolladas como parte de esta tesis y que avalan los conjuntos de datos y la investigación sísmica de este trabajo de investigación.

8. En el capítulo 9 figuran las conclusiones y futuras líneas de investigación. aplicaciones de la vigilancia sísmica.

En los apéndices de este documento, incluimos resultados y figuras adicionales para completar la comprensibilidad de los resultados presentados.

D.2 METODOLOGÍA

En esta sección, se detalla la metodología investigada y propuesta a lo largo de esta tesis: las redes Bayesianas BNN , el muestreo basado en Monte Carlo dropout y la estimación de la incertidumbre.

La mayoría de los problemas de la estadística bayesiana consisten en estimar densidades de probabilidad que son difíciles de calcular. Este tipo de enfoque incluye un conjunto de algoritmos para determinar cantidades desconocidas. En el caso de las redes neuronales, las múltiples capas ocultas inducen no linealidad en sus representaciones aprendidas, lo que conlleva a que el conocimiento exacto de los parámetros de la red no sea directamente computable. Sin embargo, el

conocimiento de estos parámetros es de interés en el uso de redes neuronales, por ejemplo, para realizar predicciones y calcular incertidumbres. Por tanto, es necesario realizar una aproximación precisa, escalable y robusta para arquitecturas de redes neuronales profundas que trabajen con datos sísmicos en continua. Empleando inferencia variacional, una metodología de optimización para funciones complejas, se establecen las técnicas de regularización estocástica y de MC-dropout empleadas en el aprendizaje profundo de nuestra red. Construimos conexiones entre las redes profundas y la teoría de detección de cambio mediante la estimación de incertidumbre. Nuestra contribución teórica en este apartado consiste pues en la asociación matemática de cambios geofísicos unidos a la incertidumbre del modelo, condicionada a los datos disponibles. Posteriormente, presentamos el contexto sismológico de esta tesis, junto con las implicaciones geofísicas del marco de estimación de incertidumbre anteriormente propuesto. Finalmente, presentamos el concepto de deriva de datos y sus implicaciones en la ciencia del monitoreo de volcanes.

D.3 Transfer learning y cuantificación de incertidumbre

En este capítulo, proponemos la mejora de los sistemas de monitorización sísmicos a través del Deep Learning Bayesiano. Presentamos las redes neuronales bayesianas para realizar la identificación y clasificación de eventos sísmicos, además de estimaciones de incertidumbre en datos recopilados en dos volcanes activos. Demostramos que la arquitectura propuesta logra un rendimiento excelente (92,08%) al discriminar tanto el tipo de evento como su origen cuando los dos conjuntos de datos se fusionan y no se proporciona información adicional al modelo entrenado. Finalmente, demostramos que las representaciones de datos aprendidas por las BNN son transferibles a través de diferentes períodos eruptivos. También encontramos que la incertidumbre estimada está relacionada con cambios en las características de los eventos analizados, y proponemos que pudiera usarse para medir si los modelos aprendidos pueden exportarse a otros escenarios eruptivos. Ilustramos las variaciones de contenido de frecuencia durante los períodos pre y post eruptivo para

ambos volcanes, demostrando que la incertidumbre epistémica del modelo detecta el cambio. Por último. la incertidumbre epistémica derivada esta primera aproximación tiene dos implicaciones principales: no solo se presenta como una característica que debe considerarse como un detector de cambios, sino también como un nivel de umbral para determinar cuándo deben usarse los algoritmos de aprendizaje por transferencia.

D.4 Monitoreo Bayesiano en señales sísmicas

Los métodos para el monitoreo de volcanes suelen estar basados en enfoques deterministas que pueden no considerar la compleja dinámica de los sistemas volcánicos. Para detectar cambios sutiles no solamente en los eventos, sino en las secuencias sísmicas asociadas con datos volcánicos, extendemos nuestra aproximación anterior y presentamos un marco de inferencia para monitorizar a lo largo del tiempo de la actividad sísmica. Esta arquitectura ha sido diseñada y entrenada para detectar y clasificar múltiples señales sísmicas con registros sísmicos continuos. Comprobamos esta nueva arquitectura analizando datos sísmicos asociados con erupciones en el volcán Bezymianny (Rusia) durante 2007. Aprendemos el fondo del tremor y desenmascaramos la señal sobre el ruido. Para ello, ha sido esencial construir arquitecturas de tipo autoencoder que permitan reducir los datos de entrada, espectrogramas en este caso, a una representación latente que captura las propiedades y estructura jerárquica de los datos. A partir de esta representación, se emplea una arquitectura temporal convolucional que realiza el reconocimiento de la entrada. Nuestros resultados demuestran una detección eficiente de señales y precisión de clasificación. La incertidumbre total del sistema, U_t, definida como la suma de las variaciones del campo de ondas sísmicas (reducibles con más datos) y la aleatoriedad del proceso volcánico monitoreado, proporciona una detección efectiva de cambios en el sistema volcánico en las horas previas a una erupción. Esta formulación teórica U_t nos ha permitido introducir los mapas de incertidumbre, como una herramienta de apoyo para ilustrar la presencia o ausencia de fuentes simultáneas. La incertidumbre epistémica actúa como un detector señales complementario, mientras que la incertidumbre aleatoria representa la identificación de múltiples fuentes que mejora en gran medida

los resultados del monitoreo.

D.5 Redes de scattering y comportamiento metaestable

Presentamos una nueva arquitectura de red neuronal profunda extremo a extremo (E2E) diseñada para realizar un monitoreo sismo-volcánico enfocado en detectar cambios mediante el aprendizaje de un banco de filtros avanzado, a partir de la señal en bruto. A través de la evolución de la incertidumbre epistémica, invocando una estrategia de red bayesiana, detectamos el cambio y demostramos su importancia como indicador para el posible pronóstico de erupciones utilizando datos de los volcanes Bezymianny y Etna. Específicamente, proponemos modificar la transformada scattering, integrada en redes de scattering profundas, en una nueva red de aprendizaje recurrente e híbrida E2E que permita capturar las dependencias temporales de múltiples escalas de la forma de onda, en muestras. El aprendizaje de la transformada scattering y del banco de filtros está condicionada en cierto sentido por la física del campo sísmico, es decir, a través de la representación de características aprendidas a lo largo del tiempo. Al mismo tiempo, con una arquitectura LSTM convolucional profunda que ha sido cuidadosamente diseñada, aprendemos la dinámica temporal y las características intra-evento. Verificamos la efectividad del aprendizaje de transferencia cambiando entre volcanes, confirmando nuestra hipótesis de tratar a los volcanes como distribuciones de datos generadoras. Estas observaciones de incertidumbre concuerdan bien con informes científicos anteriores en boletines sismológicos. La erupción principal, clásica (E2), para el volcán Bezymianny, registrada en octubre y que contiene el pico más alto de la incertidumbre total, U_t, se corresponde a la máxima liberación de energía sísmica durante el ciclo eruptivo de 2007. Por lo tanto, la tendencia continua en la incertidumbre se eleva abrupta y simultáneamente junto con la energía, indicando una explosión en el volcán que cambió repentinamente las condiciones iniciales. La energía, siendo un parámetro independiente de la incertidumbre y asociado a la forma onda, demuestra que nuestro framework de estimación es universal.

D.6 Aprendizaje activo

Los avances en el aprendizaje profundo han impulsado el campo de la sismología volcánica a niveles sin precedentes. Sin embargo, los catálogos de datos seleccionados aún requieren esfuerzos sustanciales de etiquetado, a menudo retrasados en el tiempo debido a las condiciones cambiantes de los datos sísmicos, ya demostradas en esta tesis. En este trabajo, proponemos una red neuronal convolucional temporal con aproximación bayesiana para realizar tareas de detección y clasificación en continua mientras se extraen los eventos más inciertos de la traza sísmica. La segmentación selectiva de qué terremotos deben de ser revisados por un experto reduce significativamente el tiempo de anotación, acelera el entrenamiento de redes neuronales y aumentar la adaptabilidad del monitoreo a situaciones imprevistas. Formulado como un procedimiento de aprendizaje activo (AL), la arquitectura propuesta genera un mapa de incertidumbre a lo largo del tiempo, destacando las membresías de clase que deben revisarse. Logramos una mejora significativa en las métricas de monitoreo, con únicamente una fracción del conjunto de datos inicial para lograr un rendimiento del 83 % para cinco eventos sismo-volcánicos. Esto contrasta con el sistema base, que requiere datos por un total de 288 horas para alcanzar un rendimiento parecido, aunque nuestro sistema lo alcanza con datos de un total de 6.6 horas.

D.7 Otras aplicaciones geofísicas

Por último, presentamos el software científico de apoyo que hemos implementado para el procesamiento de datos y el etiquetado de los conjuntos de datos estudiados, incluida la detección, segmentación y clasificación automática de señales volcánicas. En primer lugar, hemos empleado un reconocedor de señales sísmicas para segmentar formas de onda volcánicas, y clasificar, de manera semi-supervisada, señales sísmicas a gran escala. Posteriormente, hemos empleado una plataforma de creación propia para verificar los resultados de segmentación, y clasificar de nuevo en caso de error.

D.8 Conclusiones

Las conclusiones que pueden extraerse de todo el trabajo desarrollado en esta tesis son las siguientes:

1. En la publicación 1, Capítulo 4 [14], hemos demostrado que una red BNN, entrenada con características cepstrales alcanza un rendimiento de clasificación significativo para los tres eventos de dinámica cercana de un volcán, HF, LF y MX, en dos volcanes (Monte Saint Helens y Bezymianny); terremotos de frecuencia alta, baja y mixta. El sistema ha demostrado que cuando se entrena para cada volcán de forma independiente, logra rendimientos (F1-score) más allá del 90% para cada volcán. Cuando los datos de ambos volcanes se combinan, la BNN alcanza una mejora sobre la métrica respectiva obtenida anteriormente. Por otro lado, si los conjuntos de datos de ambos volcanes se separan de acuerdo con su origen volcánico, la BNN propuesta entrenada sobre las etiquetas dispersas (en este caso, seis), logra una alta precisión y puede discernir eventos sísmicos en función del tipo de volcán que fueron registrados. Las señales sísmicas presuntamente vinculadas al mismo mecanismo geofísico se pueden usar con la misma clase para aumentar los conjuntos de datos y mejorar la capacidad de monitoreo, ya que las muestras de datos pertenecen a un espacio de características similar. Si comparamos con otros dominios de audio, las señales sísmicas de diferentes volcanes pueden tratarse como grabaciones de la misma entidad acústica (ej.: una canción) pero en diferentes condiciones de grabación. Por otro lado, estos resultados demuestran que la taxonomía de datos adoptada en este trabajo es robusta puesto que ofrece separabilidad intraclase con alta precisión para los modelos entrenados con esta taxonomía de datos.

   Por último, ilustramos las variaciones del contenido de frecuencia que son bien detectadas por la incertidumbre asociada al conjunto de datos y el modelo BNN propuesto. La incertidumbre derivada en esta red BNN tiene dos implicaciones: se erige como una característica para ser considerada un precursor de cambio a la vez que un nivel de umbral para determi-

nar cuándo deben usarse los algoritmos de aprendizaje por transferencia. Finalmente, observamos que la aplicación del aprendizaje por transferencia permite exportar conocimiento de un volcán o de una etapa eruptiva a otra.

2. En la publicación 2 [10], capítulo 5, enfocamos nuestro problema en reconocimiento sísmico en continua para datos secuenciales provenientes de tres erupciones en el volcán Bezymianny. Proponemos una red neuronal convolucional bayesiana e híbrida (B-CNN) para aprender el ruido de fondo, y que por tanto contiene el tremor, a la vez que se realiza el reconocimiento continuo de eventos sísmicos. Esta arquitectura se basa en un marco de segmentación tipo U-net fusionado con una TCN de convolución temporal. Probamos en diferentes periodos eruptivos, con métricas de segmentación y clasificación muy altas para todos ellos. El marco diseñado produce una aproximación probabilística de la incertidumbre total del sistema, U_t, como la contribución agregada de la incertidumbre asociada con las variaciones sísmicas del campo de ondas (reducible con más datos) y la aleatoriedad del proceso volcánico monitoreado. Esta formulación nos ha permitido introducir mapas de incertidumbre de monitoreo como una herramienta adicional para ilustrar cómo se comporta la incertidumbre del modelo a nivel de forma de onda. La incertidumbre epistémica actúa como un detector de la llegada de la señal, mientras que la incertidumbre aleatoria destaca las fuentes potenciales que pueden estar presente en el flujo de datos sísmicos.

   Finalmente, la evolución de la incertidumbre a corto plazo es consistente con la medición de energía sísmica en tiempo real, un parámetro de forma de onda que es independiente de la incertidumbre estimada. La evolución temporal de la incertidumbre en los períodos antes y después de una erupción concuerda con los boletines sismo-volcánicos del observatorio ruso, para la erupción estudiada. Por lo tanto, los cambios en las señales debido a la dinámica interna del volcán son detectables por el modelo y caracterizados como una deriva de datos, asociando incertidumbre con erupciones volcánicas.

3. En la publicación 3 [12], capítulo 6, investigamos si la incertidumbre sigue patrones sim-

ilares aquellos descubiertos previamente para las tres erupciones del volcán Bezymianny, con otros vectores de características. Este trabajo presenta una arquitectura novedosa de aprendizaje profundo de extremo a extremo incorporada con un filtro de bancos wavelet que puede ser aprendido mediante optimización para mejorar las capacidades de gestión del sistema en términos de detección, segmentación y clasificación de múltiples señales sísmicas. La flexibilidad introducida en el aprendizaje en el banco de filtros, junto con la red temporal recurrente cuidadosamente diseñada, generan representaciones de características óptimas y robustas. La aplicación de esta arquitectura a datos de tres volcanes de diferentes tipos y en diferentes continentes demuestra la generalización y adaptación a distintos entornos. Nuestra red neuronal garantiza el reconocimiento rápido de eventos y es robusta contra taxonomías de datos escasos y ruido de fondo, temas de investigación muy activos en aplicaciones de aprendizaje automático para reconocimiento acústico.

En este trabajo partimos de la perspectiva tradicional en el monitoreo sísmico, y abrimos una nueva perspectiva para diseñar métodos de pronóstico basados en las conexiones entre la incertidumbre y la dinámica de los volcanes. La incertidumbre epistémica sigue una la ley de aceleración, asociada con los flujos de datos sísmicos, lo que implica que los procesos volcánicos que preceden a las erupciones son detectables. Sin suposiciones previas sobre la distribución de señales, el aprendizaje profundo puede identificar tal comportamiento de manera no supervisada. Para el volcán Bezymianny, la evolución temporal de la incertidumbre se comporta de manera similar al caso estudiado en [14], ya verificado con boletines sismológicos [63], [27].

Al aplicar metodologías de transfer learning con datos del Etna, la evolución temporal de la incertidumbre epistémica para esta erupción es parecido a la segunda erupción en el volcán Bezymianny, con una aceleración de la incertidumbre antes de la erupción principal. Estos resultados confirman que el enfoque bayesiano propuesto en esta tesis es exportable y prom-

etedor para pronosticar erupciones, aunque persisten desafíos operativos. No obstante, esta arquitectura se puede modificar para actuar sobre datos de transmisión en tiempo real; esto va de la mano con el desarrollo de una nueva estrategia de alerta temprana.

4. En la publicación 4 [11]; investigamos cómo el sistema puede superar situaciones derivadas de un rendimiento bajo como consecuencia de un cambio anterior, es decir, el nuevo stream de datos tras una erupción volcánica. Para ello, proponemos un marco de aprendizaje activo (active learning) basado en redes convolucionales temporales. La incertidumbre total asociada con este modelo se emplea en la obtención de información sobre aquellos eventos que han de ser seleccionados, dentro de la traza sísmica, que contienen una incertidumbre elevada y pueden mejorar el rendimiento del sistema. Incorporamos una estrategia de aprendizaje efectivo para mitigar la influencia de la dinámica volcánica en la selección de eventos. Hemos comprobado que durante ocho pasos de adquisición (muestreo), equivalentes a 4000 minutos de análisis de datos en un observatorio sismológico, nuestro marco de aprendizaje activo alcanza un rendimiento de reconocimiento por encima del 83,00% (F1-score), con todas las funciones de adquisición probadas. Por lo tanto, logramos un rendimiento similar al del sistema inicial que, por el contrario, ha sido entrenado con el conjunto de datos completo (unas 288 horas). Esta investigación responde a las afirmaciones anteriores sobre la necesidad de datos de alta calidad para reentrenar los algoritmos de monitoreo tras cambios en los datos monitorizados.

5. En la publicación 5 [13], publicación 6 [15] y publicación 7 [16], si bien son trabajos de investigación dedicados a la creación de aplicaciones geofísicas, constituyen un trabajo eminentemente práctico y significativo de esta tesis. Primero, el algoritmo [13] había resaltado la complejidad de las señales sísmicas para lograr una segmentación adecuada, con límites de tiempo correctos Y EXACTOS, especialmente cuando no se disponía de catálogos amplios de datos sismo-volcánicos de referencia en el momento de la implementación. Las técnicas de visualización y reducción de dimensionalidad en [13] han demostrado la necesidad de plataformas de etiquetado de datos de alto nivel, y de forma automática. Incluso

en el reconocimiento sísmico infrasónico, un dominio de datos estrechamente relacionado con el puramente sísmico, el algoritmo [15] ha resaltado las complejidades asociadas con otras fuentes de monitoreo de volcanes, requiriendo aproximaciones muy especializadas para captar información significativa.

D.8.1 CONTRIBUCIONES

Las contribuciones de esta tesis pueden resumirse en las siguientes:

1. Diseño de tres arquitecturas novedosas para ampliar la monitorización de última generación actual en volcanes. Exportabilidad de las arquitecturas implementadas en diferentes volcanes y estilos eruptivos

2. La evolución de la incertidumbre caracterizada como un buen indicador de alerta temprana volcánica.

3. Incertidumbre general como descriptor para saber cuándo es necesario volver a entrenar los algoritmos de seguimiento.

4. Formulación de un marco de monitoreo bayesiano para asociar e identificar qué incertidumbres se deben a la variabilidad del sismograma o al proceso monitoreado.

5. Identificación del comportamiento de la incertidumbre del modelo y los datos sísmicos como una power-law. La incertidumbre estimada tiene el potencial para el pronóstico de volcanes, y caracteriza cuándo y cómo está cambiando el monitoreo volcánico.

6. La dualidad de la incertidumbre que ha sido descubierta al emplear metodologías de aprendizaje activo. La evolución temporal de la incertidumbre implica que los algoritmos de monitoreo tienden a muestrear aquellas formas de onda sísmica en las que ocurre la deriva de datos, lo que implica un muestreo recurrente e infinito sin alcanzar un buen rendimiento.

Bibliography

[1] R. Balestriero and R. G. Baraniuk. "A Spline Theory of Deep Networks". In: *Proc. Int. Conf. Mach. Learn.* Vol. 80. July 2018, pp. 374–383.

[2] C. J. Bean et al. "Long-period seismicity in the shallow volcanic edifice formed from slow-rupture earthquakes". In: *Nature Geoscience* 7 (2013).

[3] Sören Becker et al. "Interpreting and Explaining Deep Neural Networks for Classification of Audio Signals". In: *CoRR* abs/1807.03418 (2018).

[4] Boris Behncke et al. "The 2011–2012 summit activity of Mount Etna: Birth, growth and products of the new SE crater". In: *Journal of Volcanology and Geothermal Research* 270 (2014), pp. 10–21.

[5] K.J. Bergen et al. "Machine learning for data-driven discovery in solid Earth geoscience". In: *Science* 363.6433 (2019). ISSN: 0036-8075. DOI: 10.1126/science.aau0323.

[6] James Bergstra et al. "Algorithms for hyper-parameter optimization". In: *25th annual conference on neural information processing systems (NIPS 2011)*. Vol. 24. Neural Information Processing Systems Foundation. 2011.

[7] Andrea Bevilacqua et al. "Bayesian construction of a long-term vent opening probability map in the Long Valley volcanic region (CA, USA)". In: *Statistics in Volcanology* 3 (Apr. 2017). DOI: 10.5038/2163-338X.3.1.

[8] Moritz Beyreuther et al. “ObsPy: A Python toolbox for seismology”. In: *Seismological Research Letters* 81.3 (2010), pp. 530–533.

[9] David M Blei, Alp Kucukelbir, and Jon D McAuliffe. “Variational inference: A review for statisticians”. In: *Journal of the American statistical Association* 112.518 (2017), pp. 859–877.

[10] A. Bueno et al. “Bayesian Monitoring of Seismo-volcanic Dynamics”. In: *IEEE Transactions on Geoscience and Remote Sensing* (2021). DOI: 10.1109/TGRS.2021.3076012.

[11] A. Bueno et al. “Continuous active learning for seismo-volcanic monitoring”. In: *IEEE Geoscience and Remote Sensing Letters* (2021), pp. 1–22. DOI: 1.

[12] A. Bueno et al. “Recurrent Scattering Network detects metastable behavior in polyphonic seismo-volcanic signals for volcano eruption forecasting”. In: *IEEE Transactions on Geoscience and Remote Sensing* (2021), pp. 1–22. DOI: 1.

[13] A. Bueno et al. “Recursive Entropy Method of Segmentation”. In: *Seism. Res. Lett.* 90.4 (2019), pp. 1670–1677.

[14] A. Bueno et al. “Volcano-Seismic Transfer Learning and Uncertainty Quantification with Bayesian Neural Networks”. In: *IEEE Transactions on Geoscience and Remote Sensing* (2019), pp. 1–13. DOI: 10.1109/TGRS.2019.2941494.

[15] Angel Bueno et al. “PICOSS: Python Interface for the classification of seismic signals”. In: *Computers & Geosciences* 142 (2020), p. 104531.

[16] Angel Bueno et al. “VINEDA—Volcanic INfrasound Explosions Detector Algorithm”. In: *Frontiers in Earth Science* 7 (2019), p. 335. ISSN: 2296-6463. DOI: 10.3389/feart.2019.00335.

[17] Helena Buurman and Michael West. “Seismic Precursors to Volcanic Explosions During the 2006 Eruption of Augustine Volcano”. In: 1769 (Jan. 2010).

[18] Annalisa Cappello et al. “Changing eruptive styles at the south-east crater of Mount Etna: Implications for assessing lava flow hazards”. In: *Frontiers in Earth Science* 7 (2019), p. 213.

[19] Rich Caruana, Steve Lawrence, and Lee Giles. “Overfitting in neural nets: Backpropagation, conjugate gradient, and early stopping”. In: *Advances in neural information processing systems* (2001), pp. 402–408.

[20] Bernard Chouet. “Volcano seismology”. In: *Pure and applied geophysics* 160.3 (2003), pp. 739–788.

[21] G. Cortés et al. “Standardization of Noisy Volcanoseismic Waveforms as a Key Step toward Station-Independent, Robust Automatic Recognition”. In: *Seismological Research Letters* 90.2A (Jan. 2019), pp. 581–590. ISSN: 0895-0695. DOI: 10.1785/0220180334.

[22] Silvio De Angelis et al. “Uncertainty in Detection of Volcanic Activity Using Infrasound Arrays: Examples From Mt. Etna, Italy”. In: *Frontiers in Earth Science* 8 (2020), p. 169. ISSN: 2296-6463. DOI: 10.3389/feart.2020.00169.

[23] John Eichelberger, Warner Marzocchi, and Paolo Papale. “Identifying best practices in short-term eruption forecasting”. In: *EOS Transactions* 93 (Jan. 2012), pp. 5–5. DOI: 10.1029/2012EO010011.

[24] Yarin Gal. “Uncertainty in deep learning”. In: (2016).

[25] Yarin Gal and Zoubin Ghahramani. “Dropout as a Bayesian approximation: Representing model uncertainty in deep learning”. In: *international conference on machine learning*. 2016, pp. 1050–1059.

[26] Elisabetta Giampiccolo et al. “Dyke intrusion and stress-induced collapse of volcano flanks: The example of the 2018 event at Mt. Etna (Sicily, Italy)”. In: *Scientific reports* 10.1 (2020), pp. 1–8.

[27] Olga A Girina. “Chronology of Bezymianny volcano activity, 1956–2010”. In: *Journal of Volcanology and Geothermal Research* 263 (2013), pp. 22–41.

[28] Charles R. Harris et al. “Array programming with NumPy”. In: *Nature* 585.7825 (Sept. 2020), pp. 357–362. DOI: 10.1038/s41586-020-2649-2.

[29] Benjamin A Heath, Emilie EE Hooft, and Douglas R Toomey. “Autocorrelation of the seismic wavefield at Newberry Volcano: Reflections from the magmatic and geothermal systems”. In: *Geophysical Research Letters* 45.5 (2018), pp. 2311–2318.

[30] Geoffrey E Hinton and Drew Van Camp. “Keeping the neural networks simple by minimizing the description length of the weights”. In: *Proceedings of the sixth annual conference on Computational learning theory*. 1993, pp. 5–13.

[31] Jesús M Ibáñez et al. “Seismovolcanic signals at Deception Island volcano, Antarctica: Wave field analysis and source modeling”. In: *Journal of Geophysical Research: Solid Earth* 105.B6 (2000), pp. 13905–13931.

[32] R. M. Iverson et al. “Dynamics of seismogenic volcanic extrusion at Mount St Helens in 2004-05”. In: *Nature* 444.7118 (2006), pp. 439–443. ISSN: 1476-4687. DOI: 10.1038/nature05322.

[33] Muhammad Salman Khan et al. “A Signal Processing Perspective of Monitoring Active Volcanoes [Applications Corner]”. In: *IEEE Signal Processing Magazine* 36.6 (2019), pp. 125–163.

[34] Diederik Kingma and Jimmy Ba. “Adam: A method for stochastic optimization”. In: *arXiv preprint arXiv:1412.6980* (2014).

[35] John C Lahr et al. “Earthquake classification, location, and error analysis in a volcanic environment: Implications for the magmatic system of the 1989–1990 eruptions at Redoubt Volcano, Alaska”. In: *Journal of Volcanology and Geothermal Research* 62.1-4 (1994), pp. 137–151.

[36] Hao Li et al. “Visualizing the Loss Landscape of Neural Nets”. In: *CoRR* abs/1712.09913 (2017).

[37] Laurens van der Maaten and Geoffrey Hinton. “Visualizing data using t-SNE”. In: *Journal of machine learning research* 9.Nov (2008), pp. 2579–2605.

[38] David JC MacKay. “Bayesian methods for backpropagation networks”. In: *Models of neural networks III*. Springer, 1996, pp. 211–254.

[39] M. Malfante et al. “Machine Learning for Volcano-Seismic Signals: Challenges and Perspectives”. In: *IEEE Signal Processing Magazine* 35.2 (Mar. 2018), pp. 20–30. ISSN: 1053-5888.

[40] Martın Abadi et al. *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. Software available from tensorflow.org. 2015.

[41] Brian McFee, Justin Salamon, and Juan Pablo Bello. “Adaptive pooling operators for weakly labeled sound event detection”. In: *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 26.11 (2018), pp. 2180–2193.

[42] S.R. McNutt et al. “Seismic and infrasonic monitoring”. In: *The Encyclopedia of Volcanoes (Second Edition)*. Elsevier, 2015, pp. 1071–1099.

[43] Stephen R McNutt. “Volcanic seismology”. In: *Annu. Rev. Earth Planet. Sci.* 32 (2005), pp. 461–491.

[44] Stephen R. McNutt et al. “Chapter 63 - Seismic and Infrasonic Monitoring”. In: *The Encyclopedia of Volcanoes (Second Edition)*. Ed. by Haraldur Sigurdsson. Second Edition. Amsterdam: Academic Press, 2015, pp. 1071–1099. ISBN: 978-0-12-385938-9. DOI: 10.1016/B978-0-12-385938-9.00063-8.

[45] Rasmiranjan Mohakud and Rajashree Dash. “Survey on Hyperparameter Optimization Using Nature-Inspired Algorithm of Deep Convolution Neural Network”. In: *Intelligent and Cloud Computing*. Springer, 2021, pp. 737–744.

[46] SC. Moran et al. “Seismicity associated with renewed dome building at Mount St. Helens, 2004-2005”. In: *US Geological Survey professional paper* 1750 (Nov. 2008), pp. 27–60.

[47] Haruhisa Nakamichi et al. “Differences of precursory seismic energy release for the 2007 effusive dome-forming and 2014 Plinian eruptions at Kelud volcano, Indonesia”. In: *Journal of Volcanology and Geothermal Research* 382 (2019), pp. 68–80.

[48] Paolo Papale. “Rational volcanic hazard forecasts and the use of volcanic alert levels”. In: *Journal of Applied Volcanology* 6 (Dec. 2017). DOI: 10.1186/s13617-017-0064-7.

[49] Iseul Park et al. “Temporal variations of repeating low frequency volcanic earthquakes at Ngauruhoe Volcano, New Zealand”. In: *Journal of Volcanology and Geothermal Research* 373 (2019), pp. 108–119. ISSN: 0377-0273. DOI: https://doi.org/10.1016/j.jvolgeores.2019.01.024.

[50] Noel Perez et al. “A new volcanic seismic signal descriptor and its application to a data set from the cotopaxi volcano”. In: *IEEE Transactions on Geoscience and Remote Sensing* 58.9 (2020), pp. 6493–6503.

[51] Lutz Prechelt. “Early stopping-but when?” In: *Neural Networks: Tricks of the trade*. Springer, 1998, pp. 55–69.

[52] D Lo Presti et al. “Muographic monitoring of the volcano-tectonic evolution of Mount Etna”. In: *Scientific Reports* 10.1 (2020), pp. 1–11.

[53] Emmanuel G Ramos et al. “The low-frequency earthquake swarms at Mount Pinatubo, Philippines: implications for magma dynamics”. In: *Journal of Volcanology and Geothermal Research* 92.3 (1999), pp. 295–320. ISSN: 0377-0273. DOI: https://doi.org/10.1016/S0377-0273(99)00091-8.

[54] Mel Rodgers et al. “Stable and unstable phases of elevated seismic activity at the persistently restless Telica Volcano, Nicaragua”. In: *Journal of Volcanology and Geothermal Research* 290 (2015), pp. 63–74. ISSN: 0377-0273. DOI: https://doi.org/10.1016/j.jvolgeores.2014.11.012.

[55] O. Ronneberger, P. Fischer, and T. Brox. “U-net: Convolutional networks for biomedical image segmentation”. In: *International Conference on Medical image computing and computer-assisted intervention*. Springer. 2015, pp. 234–241.

[56] Gilberto Saccorotti and Ivan Lokmer. “A review of seismic methods for monitoring and understanding active volcanoes”. In: *Forecasting and Planning for Volcanic Hazards, Risks, and Disasters* (2020), pp. 25–73.

[57] Justin Salamon et al. “Scaper: A library for soundscape synthesis and augmentation”. In: *2017 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*. IEEE. 2017, pp. 344–348.

[58] Salvatore Scudero, Giorgio De Guidi, and Agust Gudmundsson. “Size distributions of fractures, dykes, and eruptions on Etna, Italy: Implications for magma-chamber volume and eruption potential”. In: *Scientific reports* 9.1 (2019), pp. 1–9.

[59] R. Sobradelo and M. Joan. “Short-term volcanic hazard assessment through Bayesian inference: retrospective application to the Pinatubo 1991 volcanic crisis”. In: *Journal of Volcanology and Geothermal Research* 290 (2015).

[60] R. Sobradelo and M. Joan. "Understanding causality and uncertainty in volcanic observations: An example of forecasting eruptive activity on Soufrière Hills Volcano, Montserrat". In: *Journal of Volcanology and Geothermal Research* 341 (2017).

[61] RSJ. Sparks and WP. Aspinall. "Volcanic activity: frontiers and challenges in forecasting, prediction and risk assessment". In: *The State of the Planet: Frontiers and Challenges in Geophysics* abs/1703.04977 (2004).

[62] Nitish Srivastava et al. "Dropout: a simple way to prevent neural networks from overfitting." In: *Journal of Machine Learning Research* 15.1 (2014), pp. 1929–1958.

[63] W. Thelen, M. West, and S. Senyukov. "Seismic characterization of the fall 2007 eruptive sequence at Bezymianny Volcano, Russia". In: *Journal of Volcanology and Geothermal Research* 194.4 (2010), pp. 201–213. ISSN: 0377-0273. DOI: https://doi.org/10.1016/j.jvolgeores.2010.05.010.

[64] M. Titos et al. "Detection and Classification of Continuous Volcano-Seismic Signals With Recurrent Neural Networks". In: *IEEE Transactions on Geoscience and Remote Sensing* (2018), pp. 1–13. ISSN: 0196-2892. DOI: 10.1109/TGRS.2018.2870202.

[65] Amadej Trnkoczy. "Topic Understanding and parameter setting of STA/LTA trigger algorithm". In: *New Manual of Seismological Observatory Practice* 2 (1999).

[66] M. E. West. "Recent eruptions at Bezymianny volcano—A seismological comparison". In: *Journal of Volcanology and Geothermal Research* 263 (2013). Magma System Response to Edifice Collapse, pp. 42–57. ISSN: 0377-0273. DOI: https://doi.org/10.1016/j.jvolgeores.2012.12.015.

[67] Michael E West. "Recent eruptions at Bezymianny volcano—A seismological comparison". In: *Journal of Volcanology and Geothermal Research* 263 (2013), pp. 42–57.

[68] Jason Weston et al. "Deep learning via semi-supervised embedding". In: *Neural Networks: Tricks of the Trade*. Springer, 2012, pp. 639–655.

[69] Le Zhang and P.N. Suganthan. “A survey of randomized algorithms for training neural networks”. In: *Information Sciences* 364-365 (2016), pp. 146–155. ISSN: 0020-0255. DOI: https://doi.org/10.1016/j.ins.2016.01.039.

[70] Vyacheslav M Zobin. *Introduction to volcanic seismology*. Vol. 6. Elsevier, 2012.

[71] Barret Zoph and Quoc V Le. “Neural architecture search with reinforcement learning”. In: *arXiv preprint arXiv:1611.01578* (2016).