Bayesian Monitoring of Seismo Volcanic Dynamics

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

***Abstract*—Methods for volcano monitoring - based on the analysis of geophysical data - often rely on deterministic approaches without considering the complex and dynamical 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 with the specifc purpose to detect and classify individual earthquake transients from continuous seismic data recorded in volcanic environments. Here, we test this new data analysis framework by analysing seismic data associated with eruptions at Bezymianny volcano (Russia) during 2007. Our study demonstrates efficient signal detection and classification accuracy, and effective detection of changes in the volcanic system in the hours preceding eruptive activity. Our results can be extended to other volcanoes and earthquake-prone areas, and demonstrate new applications 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 successful implementation of such data analysis frameworks for pattern recognition in real or quasi-real-time, [1]. Volcano seismology remains the most popular tool for volcano monitoring. Volcanic activity is known to generate a variety seismic signals, which represent evidence of multiple complex processes acting within volcanic systems. Changes in the style of seismicity, its rates of occurrence and magnitude, are frequently recognized as precursors to eruptions [2]. Recent technological advances, lower costs of equipment and improved open access policies for access to 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 exploiting the feature learning capabilities 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 performances in well-studied, selected periods of seismic unrest (*snapshots*): they are selected time-interval associated with well-identified volcanic activity, used to *fine-tune* monitoring algorithms [12], [13], [14], [15]. These workflows were, however, implemented without consideration for the assessment of uncertainty, which can affect their performances in adverse conditions [16]. These algorithms can experience significant performance drops due to limitations in the use of training seismic databases and procedures that do not account for the non-stationary evolution of volcanic unrest and its seismic fingerprint. These issues can be partly mitigated through constant updating of the training datasets, an arduous task that involves continual manual inspection of large amount of seismic data and re-training of the neural networks. A probabilistic approach to Deep Learning can helps to partly tackle 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 of the previous issues it still requires expert-validated, manually segmented training datasets.

In the field of seismo-volcanic monitoring, the concept of uncertainty can be associated with the result of a measurement that characterizes the dispersion of the values that could 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. As opposed to this, 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 concepts to the field of 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 will vary according to whether there is an increase in the homogeneity of the data or, on the contrary, the data cease to resemble known conditions. It is well known that this evolution is associated with changes in the characteristics of seismo-volcanic signals, for example, increase of the energy, the existence of volcanic tremor, shift

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 Istituto Nazionale di Geofisica e Vulcanologia (INGV), Sezzione di Pisa, Pisa, Italy.

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

J. M. Ibáñez is with 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), 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.

of the frequency of the signals, among many others. In this case, an increase in seismic noise, for example, would make a signal stop looking like another of the same class because a random noise level has been added to it. In this case, it is evident that the aleatory uncertainty will have to grow, but as we have seen, the growth of epistemic uncertainty will be more evident and more useful to define a change in the volcanic system. However, the nominal value of the aleatory uncertainty and the difference with respect to the previous period can tell the volcanologist about the significance of the produced change; that is, if the variations in the epistemic is due to new data heterogeneities (mild increments of the SNR noise) or to more significant internal variations (volcanic inflation, change in seismic impedance, ...). For this reason, the randomness in the monitoring implies a change of data and the environment, always conditioned to the available training data. If an analyst can expand the volcanic knowledge by taking more refined measurements, it may be sensible to consider variables that exhibit dependence on those measures and can explain stochasticity in the data.

In this study, we introduce a new Bayesian Deep Learning (BDL) method for detection, segmentation and classification of seismo-volcanic data streams, with uncertainty quantification. We will introduce a new hybrid architecture that combines segmentation and temporal sequence classification for the simultaneous identification and separation of seismic signal from background noise. We will test and confirm 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 will discuss the potential future use of Bayesian uncertainty in volcano monitoring.

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 is presented. 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, learning the seismic background tremor is motivated by its presence as a natural 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, learning the levels 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 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 spectra 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. The Bayesian approach permits to quantify the total uncertainty of the model in both, classification and segmentation tasks, to 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 context, where *background* could have very diverse meanings. For example, vocal or music [23], speech denoising [24] and very recent work in 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 requires 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 doubled from the previous one: from the first to the last layer in *C*, 16, 32 and 64 filters are used respectively, 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 the 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

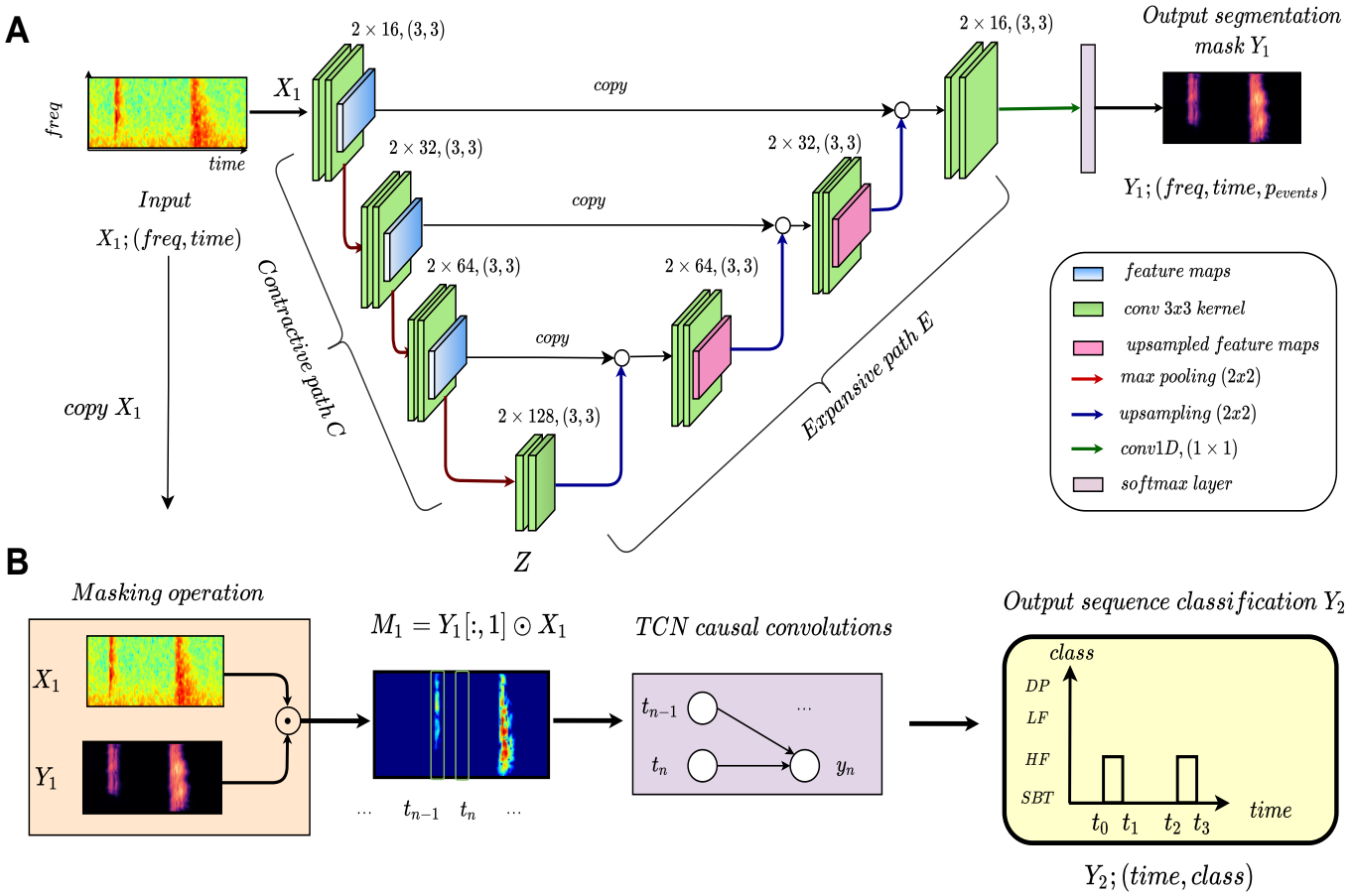


Fig. 1. Diagram of the proposed hybrid architecture to perform seismic event detection, A, and continuous sequence classification B, of seismo-volcanic events. The input is given by the two-dimensional spectrogram, X_1, size $(freq, time)$. The outputs are given by the segmentation mask Y_1 of size $(freq, time, p_{event})$ and Y_2 the labeled sequence $(time, class)$. The X_1 matrix is embedded into a latent representation Z, in which the network learn 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 TCN probes this representation to perform seismo volcanic sequence recognition, producing Y_2, size $(t, n_{classes})$, the per-frame classified data stream.

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. Further, 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, respectively. 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 indexes 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 modelling 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, 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 from previous frames, $m_{t0}, m_{t1}, ..., m_t$, but not from future inputs $m_{t+1}, ...m_t$. 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:

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

with $f(i)$ the i^{nth} filter in layer i, k the filter size and d 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 our seismic dataset D, as the set of pairs data samples (x, y), with x the processed input data stream and y the target labels, this probabilistic approximation allows the computation of the posterior distribution of the network weights, $p(w|D)$ as:

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

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

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

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 research work by [30] has 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:

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

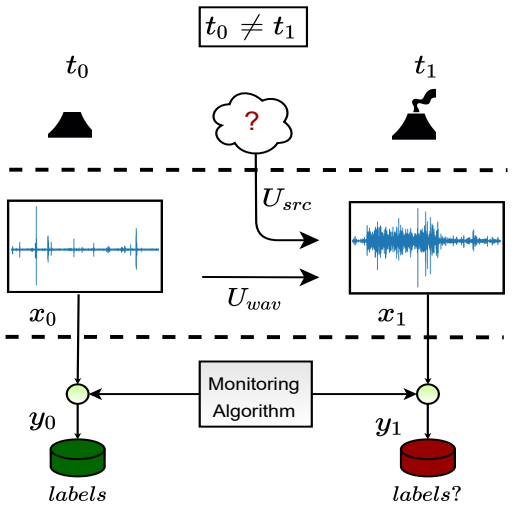


Fig. 2. The standard seismo-volcanic monitoring architectures are based on static supervised learning. In practice, when an unforeseen *change* in a volcano alter the seismograms, this translated into a potential data shift that can compromises the performance of the monitoring algorithm. Our framework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

Note that the second term of the KL divergence acts a regularizer over 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 l^{th} layer $(\omega = \{W_l\}_{l=1}^{L})$, and θ as the variational approximate in which the optimisation has to be performed:

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

with V_l the set of variational parameters, dimension $K_i \times K_{i-1}$, $p_{l,i}$ the distribution of Bernouilli parameters. This mathematical formulation yields a VI foundation to approximate 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 parameterised weights at T stochastic forward passes:

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

with $(W_l)_t \sim q(W_l)$. 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].

III. Seismo-Volcanic Uncertainties

The 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 drifts, arises if the data samples (seismograms) from an initial data support distribution are altered due 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 departed 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 are composed by an unknown number of latent, heterogeneous variables that contribute to the overall alteration on the seismic observable [2], [21]. From statistics, 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 covariance and mean vector. This approach, derived by [32], decomposes the model uncertainty in aleatoric and epistemic, offering a probabilistic proxy to uncertainty estimates without additional network parameterization. Hence, given the realized 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 step, $\tilde{p}_t = p(\hat{\omega}_t)$; the aleatoric and epistemic uncertainty can be computed as:

$$U_t = \frac{1}{T}(\sum_{t=1}^{T} diag\,(\tilde{p_t}) - \tilde{p_t}^2) + \frac{1}{T}(\sum_{t=1}^{T}(\tilde{p_t} - \overline{p})^2) \quad (8)$$

with $\overline{p}$ the averaged p_t over all the T stochastic forward passes from the dropout variational distribution. The sum both terms composes U_t, 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, that is, 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:

$$U_t = U_{src} + U_{wav} \quad (9)$$

with U_{wav} the uncertainties associated with the seismic wavefield, and U_{src} the uncertainties linked to the monitoring process. This equation 9 intersects with previous equation 8 of statistical uncertainties 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 described framework, the aleatory uncertainty estimates a single value that reflects the aggregated contribution of all the overall irreducible sources of uncertainties, 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 be leading to different seismograms, and thus, assume that the volcano has changed.

IV. Bezymianny Dataset

In order to study the potential changes in the uncertainty according to the eruptive and non-eruptive stages of the volcano and to observe possible similar behaviour in the seismic wavefield, we have selected the 2007 eruptive sequence at Bezymianny volcano, with three well-identified eruptions: 25th September, 14th to 16th October and 5th November. Figure 3 summarises the whole eruption chronology of this work. This eruptive crisis represents a non-stationary environment with seismic re-activations, eruptions, structural failures and generic unrest in a short period. The seismic network at Bezymianny volcano is composed of numerous seismic stations, and from them, we selected the BELO station according 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 the waveforms, as it is placed 1.7 kilometers afar from the eruptive centre and can be considered a very near field seismic station with negligible attenuation [34].

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

A. Bezymianny seismic events categorization

Based on preliminary geophysical knowledge of this volcano, the whole dataset has been curated with precise annotations and temporal onsets at the waveform level. Active

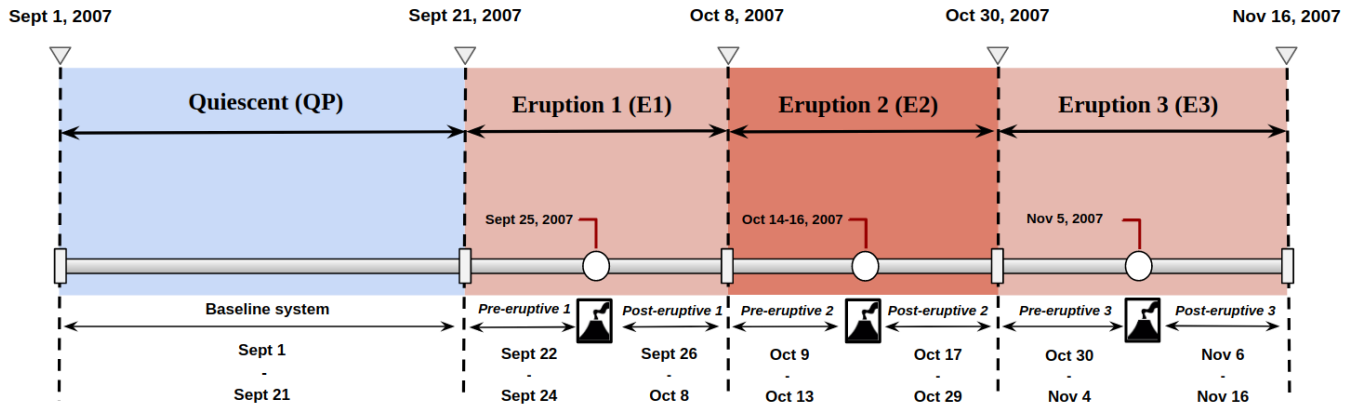


Fig. 3. 2007 eruptive chronology at Bezymianny volcano. We subdivided the eruptive sequence into four minor groups: Quiescent Period (QP), Eruption 1 (E1), Eruption 2 (E2) and Eruption 3 (E3). The QP period covers a time-span of usual background seismic data. The unrest periods E1, E2 and E3 are subdivided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

volcanic systems generate a full extent of seismo-volcanic signals. They receive different labels according to their generative volcanic sources, 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 an 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:

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

with A_{low} and A_{high} the mean amplitude of high and low spectral bands, respectively. We adopt the labelling criteria by [19], with A_{low} and A_{high} defined as $[1-5]\ Hz$ and $[6-13.5]\ Hz$. Each extracted event is labelled according to the FI value ($FR = 0.5$) and duration into 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 mechanism (i.e., brittle fracture). In contrast, lower values correspond to seismic events with narrow lower frequency band (i.e., soft ruptures [37]). The full eruptive dataset is curated in two main steps. First, a semi-supervised segmentation and categorisation at the stated FI frequency bands are 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 to manually confirm or modify the annotated labels, along with the onset times of any detected events [39]. The expert-reviewed data catalogue from September to November 2007 is compiled into a sequence-level dataset.

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

V. EXPERIMENTAL METHODOLOGY

This section describes the two main experiments to investigate how the proposed hybrid framework performs in a non-stationary seismo-volcanic monitoring application. The first set of experiments aim to obtain a broad understanding of the performance of the method. By establishing a baseline for each of the chronological periods as independent *data-snapshots*, we simulate the generic situation in which an algorithm, fine-tuned for initial monitoring conditions, is evaluated after major changes had happened in 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 systems, 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. 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 is done separated and isolated from the others baselines, with no data leakage

from past or future periods (see figure 3) during the training stage. Thus, each baseline system is trained and tested on each continuous *data-snapshots*, independently, with 75-25% data cross-validation split, and report the test set results. Then, we blind-test with the whole post-eruptive period after the next eruption in time (see figure 3). Hence, for the b-QP, b-E1 and b-E2, we blind-test with the post-1, post-2 and post-3 periods. For the b-E3 system, the blind test set corresponds to the entire QP period, where the monitoring conditions differ after three eruptions [19].

In the second experiment, we aim to determine if our framework, starting from the quiescent period, can detect when a data drift is happening and how severe affect the performance of monitoring systems, for the whole eruptive chronology. To this end, we select the best model from the QP period, (25% of the whole Bezymianny dataset) and blind-test with the rest of the eruptive periods, on each temporal sub-partitions of the dataset. This testing strategy is motivated to provide the model with a highly granular resolution of the data partitions to refine the time interval in which data change can be detected. We introduce the *monitoring uncertainty maps* as visual means to provide monitoring interpretability about the connections between waveforms and uncertainty. As a final step, regardless of how the dataset is organized, we evaluate the temporal evolution of the uncertainty for the whole dataset with a broadly used monitoring tool (energy), to link the monitored process with the estimated uncertainty.

A. Feature extractions

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 five-minute window (30000 samples), with overlapping of two and a half minutes. From these windows, we compute the input magnitude spectrograms X_1, using windows of 2.5 seconds with 1.5 seconds overlapping. As a final result, each X_1 is characterized by a matrix of dimensions $(freq, time)$, in our case, $(256, 256)$. We create 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 frequencies masks, composes the input Y_1. The per-class target sequence classification Y_2 of size, $(time, event)$ is also given during training time, with $event$ belonging to any of the classes described in section IV. In this work, we consider SBT as a single class, part of the mask.

B. Optimization procedure.

The training is perform as a multitask approach, in which a double loss function $\mathcal{L}_t$ is used: $\mathcal{L}_{seg}$ for segmentation and $\mathcal{L}_{class}$ for classification, with a softmax with a cross-entropy respectively. In the case of $\mathcal{L}_{seg}$, the softmax layer is applied over the frequency map of the target Y_1, whereas $\mathcal{L}_{class}$ is applied frame-wise. We selected *Adam* [40] optimiser with initial learning rate of 0.01, *ReLU* activation function and mini-batch size of 64. The dropout probability is set to $p = 0.25$. We use early-stopping with a patience interval of 5 epochs over 300 training epochs to prevent over-fitting. A random search over the most promising hyper-parameters derived from a Bayesian optimization procedure towards best classification and segmentation performance [41]. The Bayesian Inference procedure has been implemented as described in Section II, with T=20 *MC-dropout* sampling steps. We followed this optimization procedure for all the system trained on each period, independently. Our entire experimental methodology is simulated in 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, $F1$ stands as 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:

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

with $\cap$ the intersection between the pixels in the target and prediction mask, and $\cup$ 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

This subsection evaluates the monitoring capabilities of the proposed model to perform continuous seismic event recognition. Table 2 depicts the attained metrics for each selected eruptive periods and test sets. 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 the test sets and 94% for the blind-test partitions. Such IoU generalization capabilities are due to 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 decreases a 4% for the blind test period of the b-QP and 1% for the b-E1 systems.

TABLE II
OVERALL SEGMENTATION IOU AND CLASSIFICATION METRICS (PR, RC, F1, ACC) FOR THE BASELINE SYSTEMS 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

In terms of classification metrics, the performance gap in the b-QP system is less evident, with only 3% less in F1, PR, and RC for the blind-test partition.

Similarly, the b-E2 system experience a performance drop of 1%, 2% and 3% for PR, RC and F1, respectively. However, all the classification metrics drop 10% for the b-E1 system when blind-tested on the post-eruptive 2, which may be a mild indicator that a change has occurred in the volcano. Note that PR, RC and F1 metrics are calculated from the enhanced M_1 spectrogram representation, and hence, a change in the frequency spectra of the post-eruptive events, despite the good segmentation metrics (above 97% in 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 shifts the frequency spectra of events, compromising the performance of the model. The attained results in the blind-test for b-E2 evidence that predictions become less accurate after significant eruptions, pointing towards potential data drift.

B. Visualizing predictions

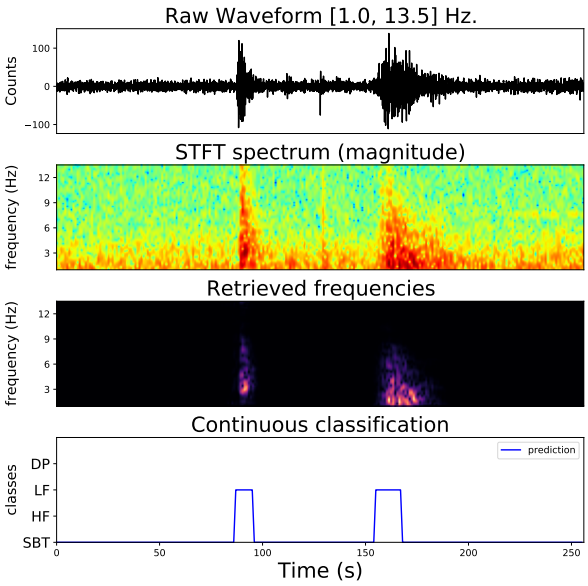
Previous results have highlighted that our hybrid framework can segment and classify seismic events from continuous data streams. From top to bottom, figure 4 displays the raw waveform, the STFT magnitude spectrum, the retrieved frequencies of events (M_1) and continuous predictions are displayed for 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 retrieved from the main trace. For example, in Fig. 4 (a) we can notice the clear low-frequency band of the LF earthquakes; ranging from 1.0 to the 3.5 Hz for the second LF earthquake. In Fig. 4 (b) we can observe the high frequency of spectra of the DP and the HF classes. The detection of target frequencies components from a broad spectra yields a denoised feature representation for the continuous TCN to perform enhanced seismic event segmentation and classification. Our network generalizes for data streams containing seismic events with copious background tremor noise as well. The classification prowess 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

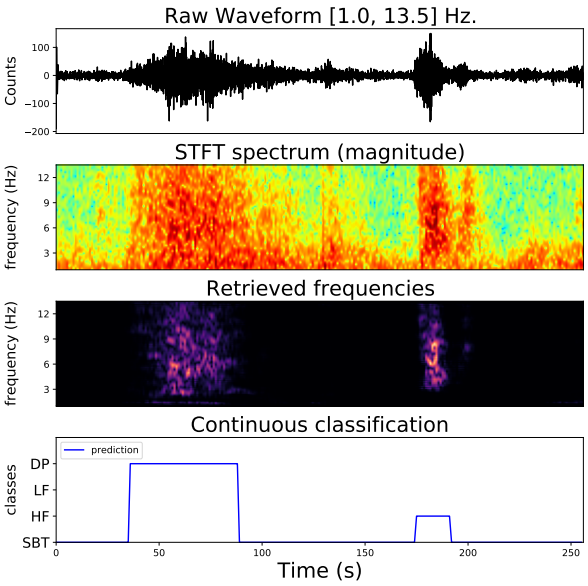
Table 3 presents the performance metrics, along with the epistemic and aleatory uncertainty; the situation in which an observatory would not have time to analyze the sheer volumes of eruptive/post-eruptive data, given the substantial annotation effort required to provide new data to retraining any monitoring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

Overall, the segmentation metric IoU remains above 96% thorough all the test-set periods. The classification performance attains high PR and $F1$ score, with RC being at an adequate level, above 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 explosion and is recovering towards stability [19]. Nonetheless, the system experiences a generalization gap in segmentation and classification metrics for all the eruptions E1, E2 and E3, up to 31% in the F1 classification metric for the E2 eruption. Hence, the 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.

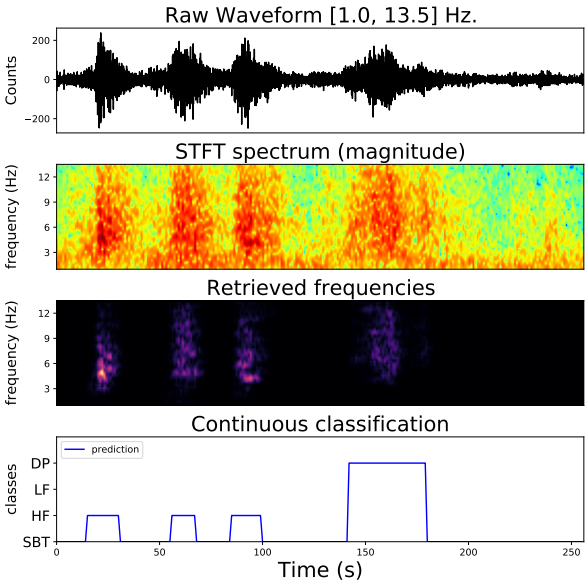
The uncertainties U_{ep} and U_{al} in both, segmentation and classification modules, act as a coupled mechanism. Higher uncertainty values correspond to lower monitoring performances. The variation of U_t with the eruptive chronology can thus be explained by comparing the variations in frequency and noise levels for each time-lapse. Note that even if the FR, the initial frequency distribution, remains approximately the same, the uncertainty is amenable to the RMS levels of seismic background tremor. For example, the 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



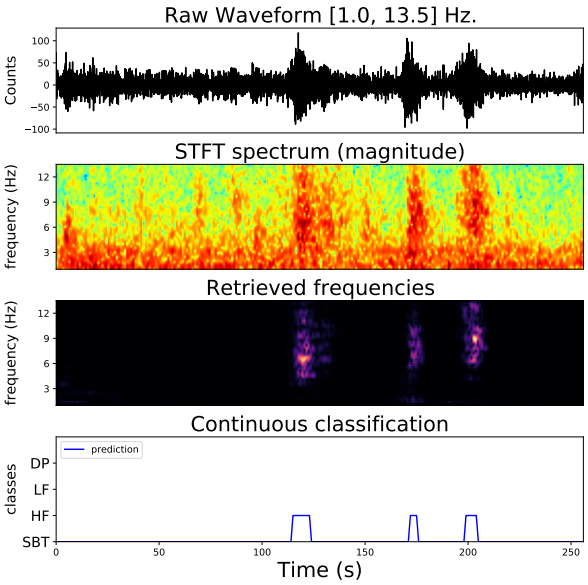
(a)



(b)



(c)



(d)

Fig. 4. Normalised waveform, input magnitude spectrum, retrieved frequencies and continuous recognition of the signal. We can notice that the network can segment the events onset, and produce a segmentation mask highlighting the active frequencies in the trace. The seismic background noise is learnt efficiently, and the TCN can track the events despite copious background noise. Besides, the network boundary segmentation for larger like DP is tight with event onsets and frequencies onsets.

TABLE III
CHRONOLOGICAL PERFORMANCE, FREQUENCY RATIOS, NOISE RMS AND OVERALL UNCERTAINTY QUANTIFICATION FOR BEZYMIANNY 2007 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

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 the eruptions, but is significantly greater than its counterpart in the classification module. The larger values in segmentation U_{ep} are due to the learning of the intra-frequency range of the volcanic tremor and seismic signals, on a very narrow spectrum which can overlap. The evolving wavefield yields seismograms and noise levels dependent on geophysical parameters, in which frequency, energy, or duration could have changed due to volcanic sources 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, position, gas content, surface temperature or differences in the open or closed conduit will imply a change in the waveform and the spectral content, with substantial differences between signals 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 difference of the power level in the tremor and noise conditions across eruptive periods is equivalent, by analogy, to the distortion of images 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

In the previous experiment, we have gathered numerical evidence that volcanic changes imply changes in the data space of the signals. We take one step further to decompose equation 9 and associate waveforms to seismo-volcanic uncertainties. The visual interpretation of the uncertainty can help *post hoc* seismic analysis and 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 on the low detected energies, which is expected as it is a very narrow spectral band carrying the information about the continuous volcanic tremor. For this volcano, U_{wav} and U_{src} presents as a coupled, synchronous mechanism with the input waveform: U_{wav} displays notable uncertainty levels at the abrupt transition from background to a seismic event. The uncertainty U_{src} is linked to the potentially co-existent sources in the seismic wavefield that the model has learnt to detect. Figure 5 (a) illustrates the seismic uncertainty maps at two LF earthquake boundaries. At the onset of the LF earthquakes, the epistemic uncertainty is emitted in the SBT and LF classes; and switches immediately to LF class up to event offset. On 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 U_{wav} and U_{src} uncertainty maps.

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): the aleatoric uncertainty tracks the high spectra 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 exhibit a similar behavior on frequency components, high or low, that are concurrent with the event.

These *monitoring uncertainty maps* serve as an interpretable, multi-source activation map that highlights potential co-existent events at any given time. The maps reflect the intrinsic non-linearity about seismic energy release, providing refined visual information about the behavior of U_{wav} and U_{src} in the seismo-volcanic wavefield. The *seismic monitoring uncertainty maps* complements the segmentation mask M_t and the classification outcome, enriching the details on the aspects of seismo-volcanic monitoring.

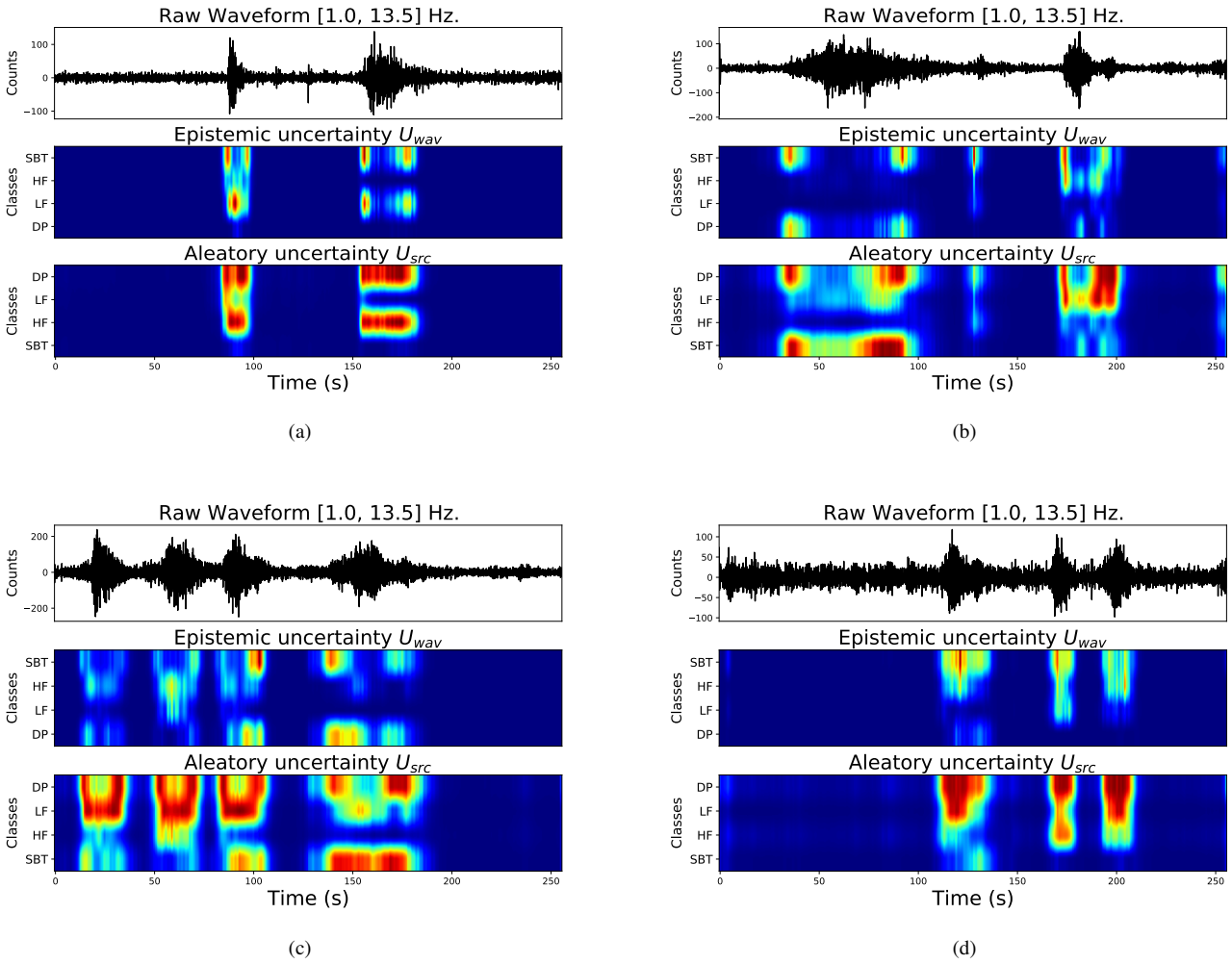


Fig. 5. Raw waveform (up), epistemic (middle) and aleatoric (bottom) uncertainty seismic maps. The association uncertainty-energy, for each event, is clearly visible. All uncertainties are associated to both, seismic energy release and signal onset.

E. Monitoring seismic wavefield

The previous sections have demonstrated how uncertainty can detect drift, for each period, independently. In this section, investigate if the estimated uncertainty is correlated to other geophysical parameters, and it evolves through time. Figure 6 depicts the normalised temporal variation of U_{ep} and U_{al}, along with the RMSE (energy). The x-axis of figure 6 represents time-units (TU); time scales over which the uncertainty has been estimated. On all three main eruptions, the estimated short-term evolution of the uncertainty exhibit similar temporal behaviour with the RMSE seismic energy. The sustained energy exchange from the volcano with the medium depends on the waveform propagation and signal frequency, thus being 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 the seismograms. These waveform distortions translate into an alteration of the monitored variable and thus into higher uncertainties 6. These graphs indicate *when* the data drift is happening. The accurate identification of such drifts is an essential input to improve monitoring adaptability to indicate 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 was preceded by minor increases in seismicity, moving towards a sustained, but smaller eruption E1. Note that this sustained seismicity is reflected in the gradual data drift. The principal, classic eruption E2, which is the highest peak for U_t, corresponds to the maximum release of seismic energy during the 2007 eruptive cycle, with high rates of seismic events recording and a massive, longer explosion. Further, previous research has reported steady increments in the background tremor energy [34]. The seismic activity leading to eruption E3 was shallow, which can be noticed in the almost constant temporal evolution of the uncertainty. Eruption E3 was a relatively small energy release that the seismological observatory could not visually verify due to poor atmospheric conditions, but only via thermal camera anomaly detection over the dome and RMSE measures [19] [35]. Therefore, the continuous trend in the uncertainty abruptly rises simultaneously as the RMSE, thus indicating an explosion in the volcano that changed the monitored conditions

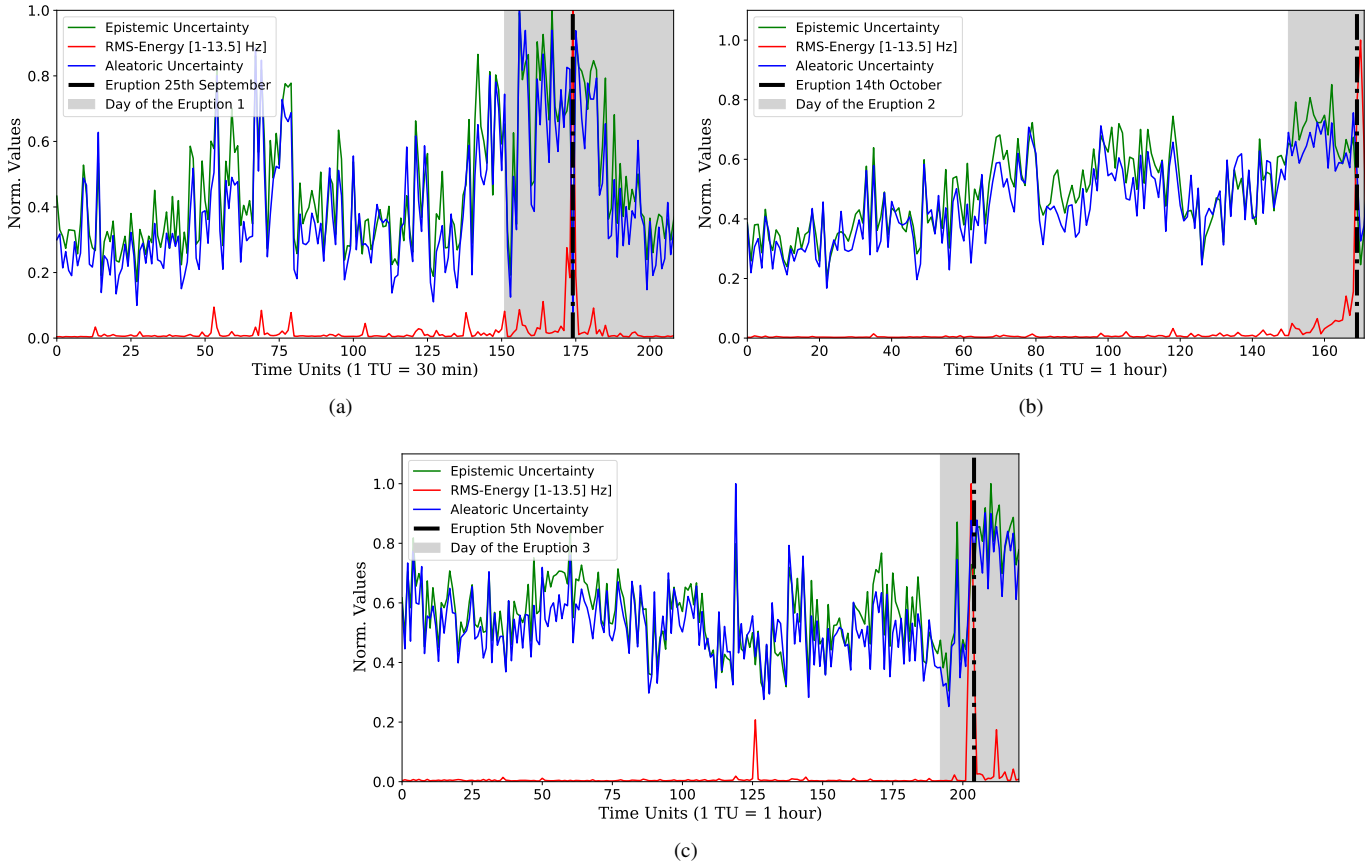


Fig. 6. Hourly normalised plots of the overall epistemic (green), aleatoric (blue) and RMSE energy (red), for the pre-eruptive and eruptions E1 (a), E2 (b) and E3 (c). There is a direct correspondence across energy and uncertainties, aligned as a coupled mechanism, as a rise of both uncertainties is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

abruptly, thus confirming past seismological bulletins.

VII. Conclusion

We have 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, 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. We group our findings in three main points. First, the designed convolutional network can learn and detect the full intra-frequency dynamic range of seismo-volcanic events whilst performing seismic event recognition.

Second, the designed framework is a probabilistic surrogate to estimate 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 change. This formulation has allowed us to introduce *monitoring uncertainty maps* as a supportive tool about 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 the monitoring outcome.

Last, we have shown that principled uncertainty estimates are useful for categorizing changes in the seismic signals. The estimated short-term evolution of the uncertainty connects with the RSEM, a direct waveform parameter independent from the estimated uncertainty.

Acknowledgment

The authors are grateful to Dr Alejandro Díaz-Moreno for continuous support and advice on Bezymianny data, along with all the people who contributed with 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 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 (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.