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:

$$p_c = \frac{exp(o_c)}{\sum\limits_{k} exp(o_k)} \tag{2}$$

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:

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

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.

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 waveform 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

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

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 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:

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**

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 ≈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.

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 $BERG$

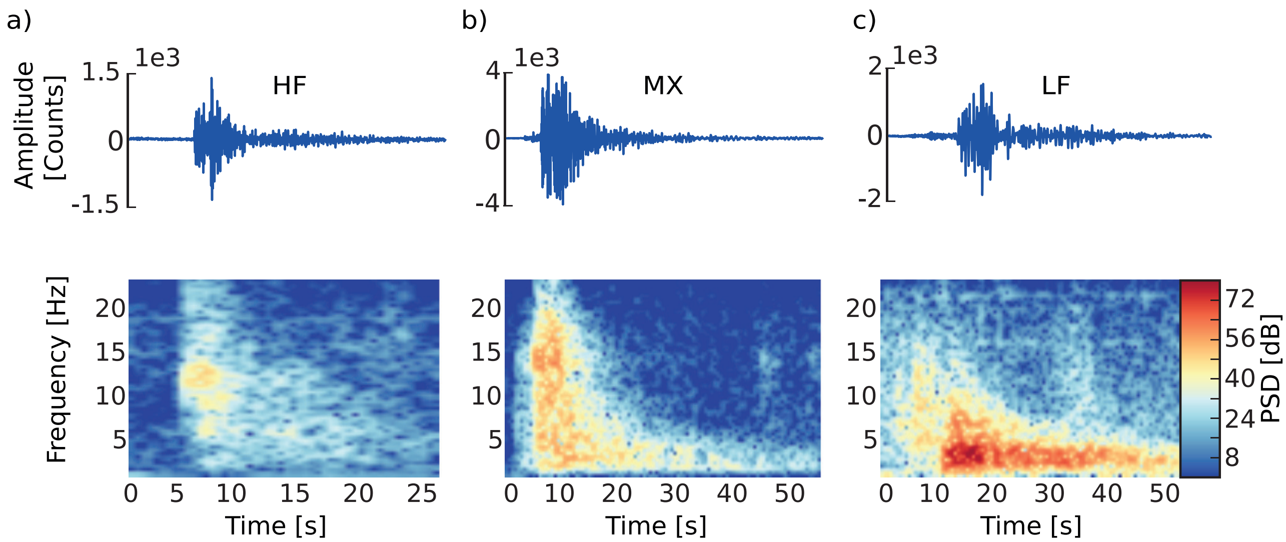

Figure 1. Waveform and spectrograms for three of the most representative volcano-seismic signals [1]. The frequency content of the events shifts from higher (a) to lower (c) frequencies. In the Mixed Frequency (MX) case (b), notice frequencies are mixed between lower and higher part of the frequency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

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).

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:

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

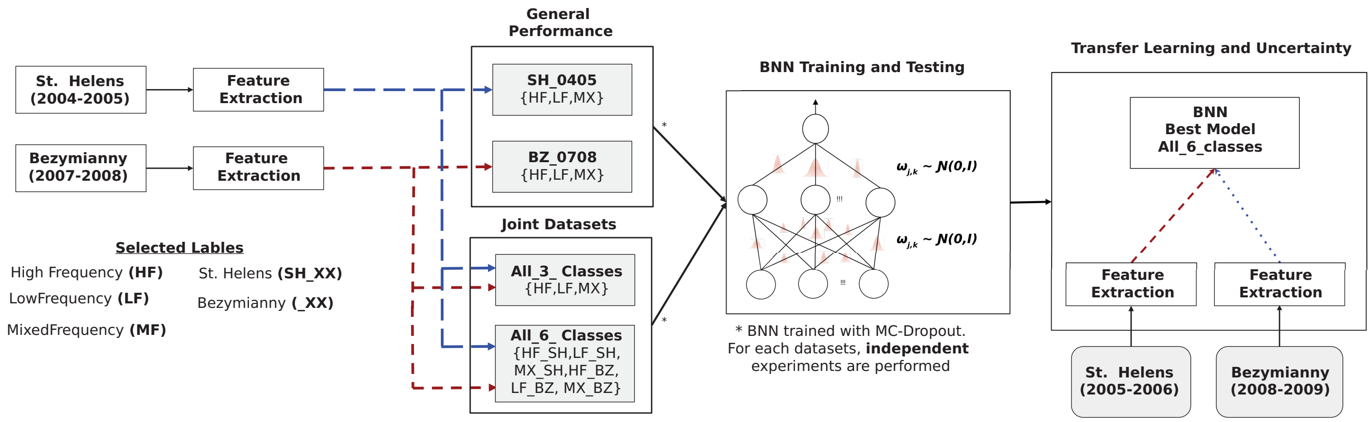

Figure 2. Experimental framework implemented in this work. For each volcano, frequency features are computed from pre-processed raw signals. Four datasets are created: SH (St.Helens), BZ (Bezymianny), similar labels (both_3_classes) and separately (both_6_classes). Best models from joint datasets are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

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
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

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.

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. 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 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

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

True \ Pred.	HF_SH *(a)*	LF_SH	MX_SH	HF_BZ	LF_BZ	MX_BZ	PR *(b)*	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

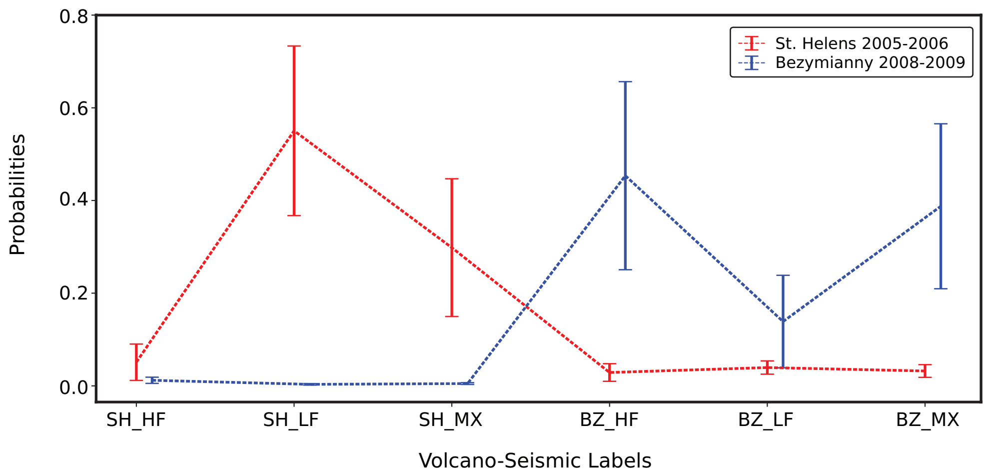

Figure 3. Per-class overall mean and variance within predictions for (a) St.Helens 2005-2006 new eruptive period, blue line (b) Bezymianny 2008-2009 post-eruptive period, red line. The axis order is the same as the assigned categorical labels in our training data. By following the lines, notice that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

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 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

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

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 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 de-

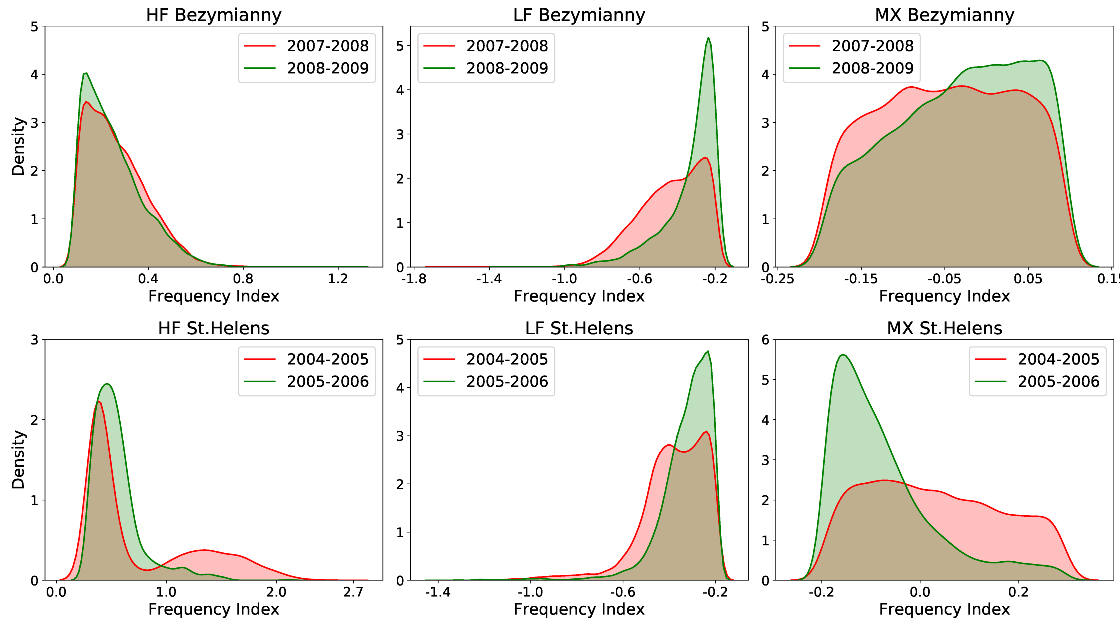

Figure 4. Frequency distribution variations for the two selected eruptive periods on Bezymianny (upper part) and St. Helens (lower part) volcanoes. Notice that, after the main eruptions, the frequency content in all selected events changes, shifting the frequency index distribution.

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

	(a)						***(b)***					
True \ Pred.	**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**

creasing 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.

[50] 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.

[51] K.P. Diederik and J. Ba. "Adam: A Method for Stochastic Optimization". In: *CoRR* abs/1412.6980 (2014). arXiv: 1412.6980. URL: http://arxiv.org/abs/1412.6980.