

# Detection and Classification of Continuous Volcano-Seismic Signals with Recurrent Neural Networks

Manuel Titos\*, Angel Bueno\*, Luz García\*, Carmen Benítez\*, Jesús Ibañez†.

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

†Instituto Andaluz Geofísica, University of Granada, Spain

**Abstract**—This paper introduces recurrent neural networks (RNN), long short term memory (LSTM) and gated recurrent unit (GRU), to detect and classify continuous sequences of volcano-seismic events at Deception Island Volcano, Antarctica. A representative dataset containing volcano-tectonic earthquakes, long period events, volcanic tremors and hybrid events was used to train these models. Experimental results show that RNN, LSTM and GRU can exploit temporal and frequency information from continuous seismic data, attaining close to 90%, 94% and 92% events correctly detected and classified. A second experiment is presented in this work. The architectures described above, trained with data from campaigns of seismic records obtained in 1995-2002, have been tested with data from the recent seismic survey performed at Deception Island Volcano in 2016-2017 by the Spanish Antarctic scientific campaign. Despite the variations in the geophysical properties of the seismic events within the volcano across eruptive periods, results provide good generalization accuracy. This result expands the possibilities of RNNs for real-time monitoring of volcanic activity, even if seismic sources change over time.

## I. INTRODUCTION

Seismic monitoring of potentially active volcanoes is the most popular remote sensing technique to forecast eruptions. Seismic anomalies are often associated to volcanic eruptions, as they reflect energy exchanges between volcanoes and the environment [1]: for example, gases might be accelerated within the volcano edifice, leading to heavy explosions and intense earthquakes. These energy exchanges have their source in stress and relaxation processes, pressure changes or fluid movements [2], [3]. As a result, seismometers can record a wide range of volcano-seismic signals that reflect the nature and underlying physics of the source process. By analyzing these seismic events, we can classify them into classes and sometimes identify the active sources of emission, and thus, improve our knowledge about the state of the volcano. Machine learning and signal processing techniques provide an appropriate framework to analyze such data.

Machine Learning algorithms have emerged as a reliable approach to classify volcano-seismic data. Support Vector Machines with Gaussian Kernels were proposed by [4] as computational models to discriminate volcanic tremors, landslides and explosions at Stromboli Volcano. Artificial Neural Networks (ANNs) for binary classification of noise and volcano-tectonic earthquakes were introduced by [5]. More recently, [6] used ANN to discriminate hybrid events at Stromboli volcano, whilst [7] proposed deep neural networks based on stacking of unsupervised pre-trained simpler models to classify isolate events from Volcán de Fuego de Colima

(Mexico). Variations of Stromboli tremor features were highlighted using a combined ANN-dynamical approach by [8]. Similar approaches employ one-hidden-layer neural networks to discriminate explosions from noise at Stromboli Volcano like [9], [10], or to train a Self Organized Maps (SOM) for identification and interpretation of events attributes and correlations [11], to analyze Stromboli VLP events [12], volcanic tremor at Etna [13], at Raoul Island [14], at Ruapehu [15] and at Tongariro [16].

Research work by [17] introduced Hidden Markov Models (HMMs) to classify seismic events at Merapi. [18] highlighted their potential role in volcano alert level decision-support. [19] applied continuous HMMs to study the continuous seismic signal at Tenerife. [20] used HMMs to analyze Etna flank eruptions while in [21], HMMs are used as statistical models to classify temporal sequences of seismic data at *Deception Island* volcano. Using feature vectors based on log-scale cepstral coefficients, a HMM was trained to discriminate between five different types of seismic events at *Deception Island* volcano. Similar approaches by [22], [23] applied HMMs to discriminate seismic events at *Saint Cristobal* and *Popocatepetl* volcanoes. Research by [24] studied the robustness of HMMs to discriminate earthquakes and explosions from noise background, and combined both datasets to test the portability of HMMs among volcanoes. The time dependence is best exploited by the hidden semi-Markov models (HSMM) proposed by [25]. [26] proposed for Galeras a HMM based on a hybrid generative–discriminative classification paradigm. Also interesting is the learning-while-recording approach by [27] that can identify events as soon as they first appear. [28] at Nevado del Ruiz exploited the temporal contribution of features.

Sequence modeling requires efficient models that are able to capture the temporal evolution of seismic data. Detecting volcano-seismic events from seismic data is a sequential problem which involves complex and highly dimensional dynamic signals. Recurrent Neural Networks (RNNs) have arisen as neural networks with temporal modeling capabilities, being able to translate input sequences into output sequences [29], [30]. This mapping is performed by computing non-linearity functions through time. However, for very long-duration sequences, RNNs had problems capturing long-term dependencies since, computing the error derivatives, the gradient might vanish or explode, acting as a constrain [31] [30]. To solve this problem, two recurrent alternatives known as Long Short Term Memory (LSTM) and Gated-Recurrent Units (GRU)

were proposed by [29] and [30]. Using internal "memory cells"; both models can better capture long-term dependencies.

RNN have become the state-of-the-art in many scientific disciplines, such as speech recognition [32], image generation [33], natural language processing [34] and machine translation [35], [36]. In remote sensing, RNNs have been used in a wide range of tasks: research works by [37], [38] applied RNNs to predict the time and location of *moderate-to-large* earthquakes. Moreover, RNNs have been successfully implemented as rain-fall forecast models, [39], [40]; cyclone track prediction [41]; river flow forecasting [42]; ocean wave forecasting [43] and to classify satellite image data [44].

In this research, we present a new volcano-seismic recognition framework based on RNNs for five different types of seismic events recorded at *Deception Island* volcano. To develop this system, we needed to take into account two aspects:

- Environmental noise, faulty sensors and the fact that soil conditions corrupt recorded signals by introducing unwanted information.
- Performance of the system is highly dependent on how accurately parameters of the model can be estimated using the available training data.

In order to address these considerations, log-filter banks (LFB) features are extracted, in a similar process to [21]. Once features are extracted, we aim to address the capabilities of vanilla-RNNs, LSTMs and GRUs to map temporal volcano-seismic sequences into human-readable labels, and determine how usefully LSTM and GRU model long-term signals, such as volcanic tremors.

The rest of the paper is organized as follows: section II provides a theoretical framework of RNN architectures, and how they can be used for temporal modelling of volcano-seismic events. Section III describes from the geophysical point of view, the seismic signals recorded at *Deception Island* volcano. Section IV describes the experimental setup. Section V presents the results and discussions. Section VI concludes the study.

## II. RECURRENT NEURAL NETWORKS

RNNs are feed forward neural networks that process sequential data. A RNN specializes to map a given sequence  $X = x_1, x_2, \dots, x_n$ , into an output sequence  $Y = y_1, y_2, \dots, y_n$ , by computing layers of non-linearities through time-steps (see Figure 1). Furthermore, each time-step can be considered as an additional layer in a deep feed-forward neural network, with weights shared across time [29].

RNNs parameters  $\theta$  are defined by three weights matrices:  $U$  the input to hidden connections,  $W$  the hidden-to-hidden recurrent connections between layers, and  $V$  from hidden layer to the output layer. Then,  $h^{(t)}$  (hidden states at time  $t$ ), can be computed following Equation 1:

$$h^{(t)} = \sigma(x^{(t)} * U + h^{(t-1)} * W + b) \quad (1)$$

Being  $h^{(t-1)}$  the hidden state of the network computed in the previous time-step  $t - 1$ ;  $x^{(t)}$  the input vector at current time-step  $t$ ;  $W, U$  the recurrent and input weight matrices

above-mentioned;  $\sigma$  the non-linearity function, and  $b$  the biases of the network. As a result,  $y^{(t)}$  (the output of the time-step  $t$ ) can be obtained using Equation 2:

$$y^{(t)} = \text{softmax}(V * h^{(t)}) \quad (2)$$

Where  $V$  is weight matrix from hidden to the output layer, and  $\text{softmax}()$  is the softmax function used over the outputs to compute the normalized per-class output probabilities. Notice that matrix  $W$  keeps the weights shared across all time-steps, and RNN takes into considerations previous information to predict  $y^{(t)}$ . Weights sharing acts as a "memory", forcing hidden units to be sensitive to input variations *through time*.

In a traditional feed-forward neural network, training stage is performed using stochastic gradient descent (SGD) [29]. However, in a RNN the setting of the gradient depends not only on the current time-step, but also the previous ones. This procedure is known as *backpropagation through time* (BPTT) [30]. Therefore, the error derivatives are propagated through time, and matrices  $U, W$  and  $V$  are updated.

In this sense, the computation of the error derivatives faces mathematical constraints: the gradients with respect to the parameters in the early layers become extremely small (vanish) or extremely big (explode), effectively preventing the weight from changing its value, acting as a constrain for the RNN to learn long-range dependencies. If the gradient vanishes, its value will decrease exponentially after few time-steps, and long-range dependencies will not be learnt. This problem, known in the literature as *exploding and vanishing gradient*, is widely studied by [31] and [30].

To alleviate this problem, several techniques such as clipping and regularization gradients were suggested by [31]. In practice, for large temporal sequences, it is more effective to use specific models to tackle the vanishing gradient problem. These models, known as Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU), employ internal activation gates to control information flow through time, efficiently learning long-range dependencies.

### A. Long Short-Term Memory

Long Short-Term Memory or LSTM networks are a variant of RNN in which the hidden states  $h^{(t)}$ , has been replaced by a memory cell (see Figure 1(b)). LSTM networks do have the ability to remove or add information inside the memory cell state, carefully regulated by a gated mechanism which allows the information to flow through time, and explicitly model long-term dependencies. The memory content  $\tilde{c}_t$ , jointly with the input  $i_t$ , forget  $f_t$  and output  $o_t$  gates are defined as:

$$i_t = \sigma(x_t * U^i + h_{t-1} * W^i) \quad (3)$$

$$f_t = \sigma(x_t * U^f + h_{t-1} * W^f) \quad (4)$$

$$o_t = \sigma(x_t * U^o + h_{t-1} * W^o) \quad (5)$$

$$\tilde{c}_t = \tanh(x_t * U^g + h_{t-1} * W^g) \quad (6)$$

Being  $\sigma$  the sigmoid function used to compute non-linear complex functional mappings between the inputs and response variable. The matrices  $U^i, W^i, U^f, W^f, U^o, W^o, U^g, W^g$  are

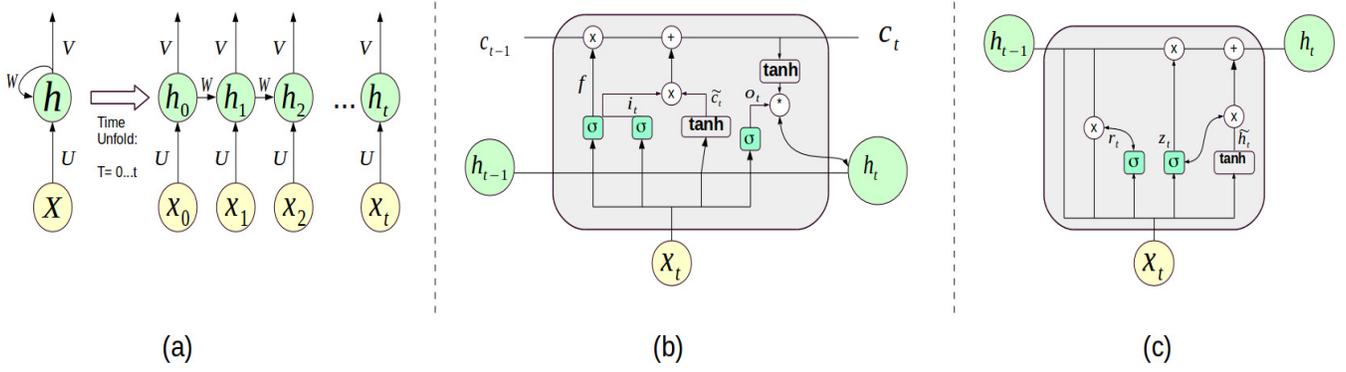


Figure 1. (a) Classical vanilla-RNN time unfolding operation, defined by the matrices  $U$ ,  $V$  and the recurrent connections  $W$ . Hidden states  $h$  are computed over time, and predictions  $Y = y_1, y_2, \dots, y_n$  are made at each time step  $t$ . (b) LSTM cell with input  $i$ , output  $o$ , forget gates  $f$  and memory candidates  $c$  are depicted. (c) GRU cell with input, reset gates  $r$  and update hidden states gate  $\tilde{h}_t$ . Notice that GRU is a simplified version of the LSTM, with less gates and internal connections.

initialized at the beginning of the learning process and they are used to compute the values of each gate. By computing a memory candidate  $\tilde{c}$ , the aim of these gates is to determine which quantity of both, old and new memory, should be ignored in order to compute the current hidden state  $h_t$ . The memory cell content  $c_t$ , at time  $t$ , can be updated as:

$$c_t = c_{t-1} * f + \tilde{c} * i \quad (7)$$

Being  $c_{t-1}$  the memory content at the previous time-step. Given the associated candidate state of memory cell  $\tilde{c}_t$  and the output gate  $o_t$ , the hidden state  $h_t$  is computed by:

$$h_t = \tanh(c_t) * o \quad (8)$$

Output states are computed using a softmax layer following the same procedure used for the vanilla-RNN (Equation 2), where  $h_t$  (Equation 8) is the hidden state computed using memory cells. Training stage is done via BPTT with a defined loss-function. For categorical classification task, cross-entropy cost function is used [45].

### B. Gated Recurrent Unit

After empirical evaluation of how internal gates affect performance, Gated Recurrent Units (GRU) were proposed as a faster and computational lightweight version of the LSTM by [36]. GRU reduces the number of gates by keeping the essential gates to model long-term sequences, deprecating the rest. Figure 1.c depicts the GRU scheme. GRU has a reset  $r_t$  and update gates  $z_t$ , given by:

$$r_t = \sigma(x_t * U^r + h_{t-1} * W^r) \quad (9)$$

$$z_t = \sigma(x_t * U^z + h_{t-1} * W^z) \quad (10)$$

Where matrices  $U^r, W^r, U^z, W^z$  are initialized at the beginning of the learning process. The candidate activation  $\tilde{h}$  is computed as in Equation 1. Note that GRU model is simpler than standard LSTM networks. As its name suggest, the reset gate controls the amount of information to flush from the memory. The update gate controls how much information needs to be stored at current time step for future computations.

The current hidden state is computed using the update gate, the candidate activation  $\tilde{h}$ :

$$h_t = (1 - z) * h + z * \tilde{h}_t \quad (11)$$

Similar to LSTM models, output states are computed using a softmax layer (Equation 2), being  $h_t$  (Equation 11) the hidden state of the network. Training stage is also done via BPTT with a defined loss-function. According to empirical evaluations by [30] and [46], both architectures LSTMs and GRUs yield similar performance, being the fine-tuning and hyper-parameters optimization the most time-consuming parts in both architectures. It also important to emphasize that GRUs may train a bit faster than LSTMs, as they have fewer parameters and less gates.

## III. VOLCANO-SEISMIC DATA AT DECEPTION ISLAND

Volcanoes are the surface manifestation of dynamic and complex processes occurring in the Earth's interior coupling physical and chemical processes. Due to the complexity of these processes, a large variety of different seismic signals can be recorded in these environments [47] [48], but no uniform global classification scheme has been done. In despite of this variety, there is a remarkable observation: many volcanoes show comparable seismic characteristics that can be associated to different seismo-volcanic sources [49]. Therefore seismic signals are often classified into event families that could help to evaluate potential seismic sources and their relationship to the present volcanic process [22]. In our case, at *Deception Island*, the main volcano-seismic events can be grouped as:

- 1) *Long period events (LP)* (Figure 2(a)): The source model is related to fluid dynamics within the volcano edifice: from cracks in which resonances occur when liquids are ascending towards the surface, to existence of pressure transients within the fluid-gas mixture, also causing resonance phenomena [2] [50]. They are located at the shallow part of the volcano, and their frequency content is restricted to a narrow band between 0.5 and 5 Hz.

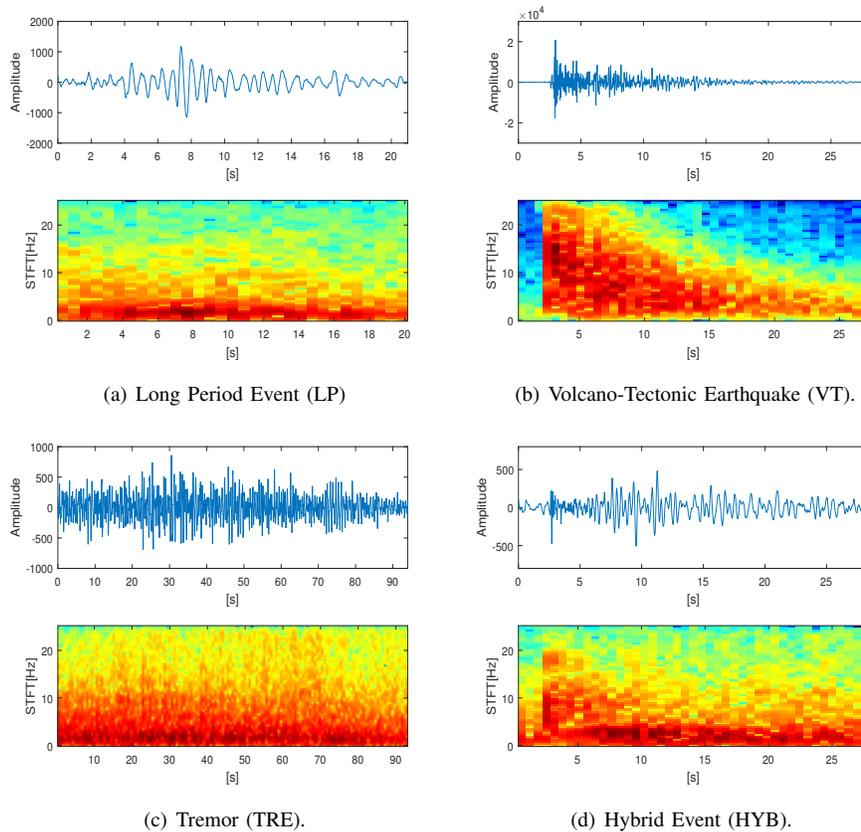


Figure 2. Spectrogram of volcano-seismic signals recorded at *Deception Island* volcano, during three seismic surveys: 1994-1995, 1995-1996 and 2001-2002.

- 2) *Volcano tectonic earthquakes (VT)* (Figure 2(b)): VT are very impulsive signals originated within a range of depths. They are produced by seismic stress; when a solid fracture takes place, it produces a seismic wave in which P (Primae) and S (Secundae) waves can be identified. The spectral content of this signal is very wide, reaching up to 30 Hz.
- 3) *Volcanic tremors (TRE)* (Figure 2(c)): Volcanic tremor is a sustained sing always present in active volcanoes [51]. Their spectral content is below 5 Hz. Their duration is highly variable, lasting from few minutes to months. Some theories consider they are caused by magma movements, whereas other suggest gas fluctuations. Since their source mechanisms are still unknown, the importance and timing between the first appearance of tremor and possible eruptive activity is still a matter of discussion [52]. In many cases to identify and to distinguish it from background noise is a quite complex task, requiring advanced signal analysis. For the case of *Deception Island* volcano, we used the well-know result described by [53], [49]). In this sense, there is an indubitable distinction between TRE and Noise in this database.
- 4) *Hybrid events (HYB)* (Figure 2(d)): These signals are characterized by an initial high-frequency phase, with short duration, and followed by a second signal identical to that of the long period event. Their origin can be explained by the increments of pressure that leads to

earthquakes. This pressure-induced fracture is filled with volcanic fluids, producing LP events. Hybrid events are related to imminent pre-eruptive episodes.

- 5) *Silence (SIL)*: These type of signals, mainly of low amplitude, are recorded when the internal seismic sources within the volcano does not emit any seismic information.

The seismic waves contain information not only on the volcanic dynamics but also about the inner complex structure of the volcanic edifice affecting the seismic wave-field and its interpretation [54]. In most volcanoes, a pronounced and rough topography introduces new complex effects, such as interference, severe attenuation effects, or changes in the path followed by the direct seismic waves [55]. As final result, even at same volcano, the same original seismic signal is recorded with different shape and wave-field characteristics according to the site of the placed seismometer. In addition, at the same seismic station, similar seismic-sources generate different signal patterns according to the way in which the source radiates energy [56]. All of these effects can be mostly classified according to path effects (attenuation) and source effects (energy and radiation pattern).

#### A. Attenuation Effects.

The main phenomenon conditioning the spectral content and shape of a seismogram is the seismic attenuation [57]. Seismic attenuation is the contribution of both, the energy lost by in-elasticity (intrinsic attenuation) and the energy

lost by dissipation (scattering attenuation). The effect of the attenuation is a visible loss of energy, being more effective at higher frequencies and directly dependent of the distance receiver-source. In volcanic environments, where the complexity and heterogeneity of the medium is more pronounced, this effect strongly modifies the seismic waveform, producing several phenomena such as: arrivals of scattered seismic waves in the last part of the seismograms [58]; reduction of high frequencies energy contribution in the seismograms [59]; changes of the magnitude and laws of scale [60]; or distortions in the spectra of the LP events [61], among others. These effects increases the difficulty to discriminate the type of event.

### B. Source Effects.

The source of volcano-seismic signals is associated to the interaction of geophysical systems. Source effects can be related to the interaction of water and hot rocks, among many others factors. At *Deception* island, evidences of aquifers and hot materials placed near the surface have been widely confirmed using seismic tomographies in velocity and attenuation [62]. Interactions between water and hot rocks generates a sudden change of phase at depth, with its associated pressure step and radiation of high frequency seismic waves. In addition, the presence of several and complex fault systems in the area [63] induces low-frequency seismic waves swarms as the result of fluid auto oscillations filling the crack. When the interaction between water and hot-rocks is simple, simple oscillations are recorded. However, in case of multiple interactions, the continuous change of phase and resonance of the faults generates an overlapping of signals.

### C. Deception Island dataset

*Deception Island* is located at  $62^{\circ} 59' S$  and  $60^{\circ} 41' W$  in the South Shetland Islands. It is considered as one of the most active volcanoes in the Antarctic Peninsula, with more than six eruptions in the last 160 years. The dataset was collected during three seismic Antarctic surveys in 1994–1995, 1995–1996, and 2001–2002. In [49] a full description of sensor and acquisition systems can be obtained. Data labelling was performed by a committee of geophysicists, based on their professional knowledge and experience of *Deception Island* volcano. All volcano-seismic events have been selected as the most representative of each class. As a result, a total of 512 continuous data streams were obtained, containing a total of 2193 events, with the following per-class distribution: 75 VT, 765 LP, 77 TRE, 54 HYB and 1222 SIL. Recorded data streams are *continuous*, that is, each recorder seismic signal is composed by different volcano-seismic signals. Figure 3 shows the time (by number of frames) distribution histograms for each volcano-seismic class recorded described. Notice that the histograms of LP class in Figure 3(a) and SIL class in Figure 3(e) contain many short frames, with a duration less than 5 seconds (three overlapped frames).

## IV. EXPERIMENT DETAILS

### A. Data processing and feature extraction.

The overwhelming success of recurrent neural networks and the replacement of hand-crafted features with features learned directly from data might be a good approach to classify volcano-seismic events. However, volcano-seismic events are related to complex geophysical processes, which yield to specific particularities from a signal processing perspective (see subsection III). Temporal duration of events within the same seismic campaign can be very different, with long-range temporal dependencies that are hard to model. During an eruption, events can be recorded simultaneously (e.g., explosions can be associated to lava flows and/or rock falls), making the classification problem harder. In the context of volcanic-seismology, data parameterization is still advisable in order to enhance the learning process. Experiments are performed with raw data, linear prediction coefficients (LPC) and log-filter banks (LFB).

Raw data has been windowed with 4 seconds Hamming windows and 3.5 seconds overlapping. In the case of LPC features approach, once the signal had been windowed, 5 LPC coefficients for each window are computed, yielding to a feature vector with 5 components. LFB data parameterization is automated in a *pipeline*: the input is the volcano-seismic raw waveform, and the output is the parametric representation with less redundant information. LFB carry rich information and are computationally simple, being this an important advantage when deployed in volcano observatories. These features have been suggested by [21] and [24] as a robust approach to parameterize volcano-seismic signals. Figure 4 depicts the feature extraction *pipeline*, which can be summarized as follows:

- 1) Once the signal has been windowed a 512-points FFT is computed for each frame. The magnitude of the spectrum is used as a set of 16 triangular filter bank uniformly distributed, on a logarithmic frequency scale with 50% overlapping between adjacent filters. The purpose of this filter bank is to give the average of the energy of the signal at a given frequency band.
- 2) Optionally, we can compute the logarithm of the output filter-bank energies and apply the discrete cosine transform (DCT) in order to decorrelate the features. In the case of ANN, this step is not necessary since DCT is a linear operation, being counterproductive with the non-linear nature of the activation functions of the hidden units.

As a result, each window provides a feature vectors of 16 components. During training stage, and for a specific temporal sequence in the dataset, successive 16-*dimensional* frames are fed to the recurrent architectures. Given this set of extracted features, we will train three recurrent architectures: vanilla RNNs, LSTM and GRU networks. Our main goal in this paper is to show the robustness of RNNs architectures as temporal classifiers of five different types of seismic events and to understand if internal gates in LSTM and GRU can capture specific temporal patterns within the seismic data.

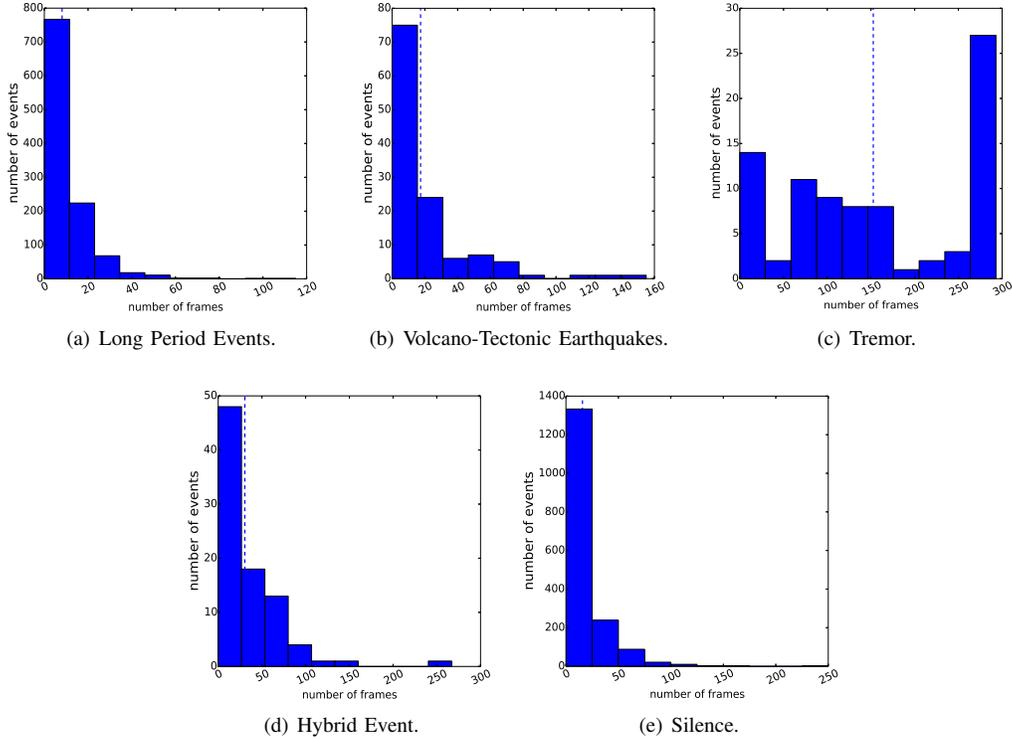


Figure 3. Histograms summarizing the length distribution by frame for the volcano-seismic signals recorded during 1994–1995, 1995–1996, and 2001–2002 seismic campaign at *Deception Island* volcano. Dashed lines are centered on the mean duration of each event. All signals have been windowed with a 4 seconds Hamming window, with 3.5 seconds overlapping, as explained in subsection IV-A.

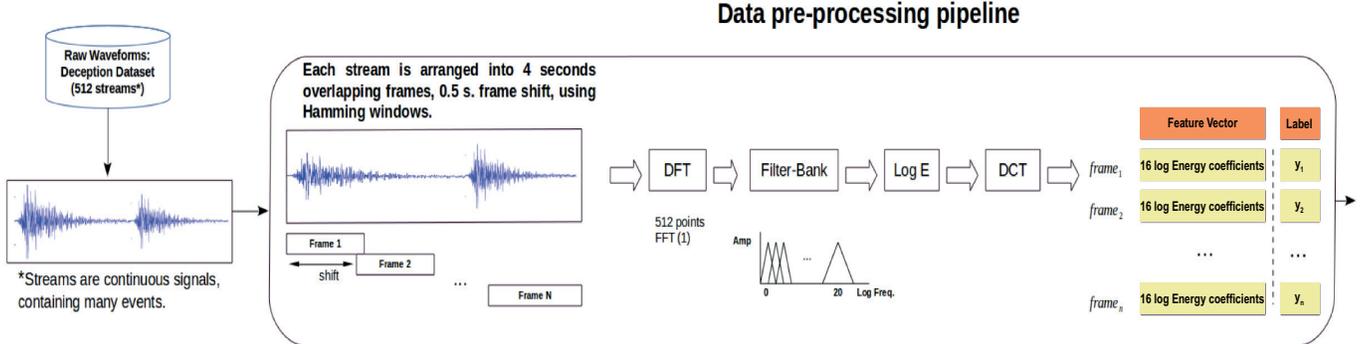


Figure 4. Data pre-processing pipeline. Each signal is windowed every 4 seconds, with 3.5 seconds overlapping, using a Hamming window. A 512-FFT spectrogram is computed for each frame, and 16 log-filter banks features are used. At the end of this pipeline, all volcano-seismic signals are encoded into frames, each frame containing a feature vector of 16 coefficients, being these the inputs to our recurrent neural networks. Thus, any given seismic signal is represented by a matrix with dimension (number of frames x 16).

## B. Experimental Setup

All recurrent neural network were implemented using *Theano* [64], a deep learning framework written in Python. Data parameterization and feature extraction have been implemented in Python.

Experiments to determine the temporal modelling capabilities of recurrent architectures are performed using data from *Deception Island* volcano (section III). Before training stage, data is pre-processed as described in section IV-A. Once processed, the dataset is split into training (80%), and test (20%) sets. A random shuffle of the data is done to balance both sets. In order to test model capabilities to generalize

on unseen data, leave-one-out cross-validation with four test partitions is used.

Vanilla RNNs, GRU and LSTM are optimized following a similar approach as described in [65], evaluating several models between 10 and 300 hidden units. A *softmax* probability layer with five target outputs, corresponding to each class of our dataset is added on top. Learning rates have been tested within the range [0.1, 0.001], with a momentum of 0.9 and learning rate decay of 0.001. L2 regularization and early-stopping criterion with a patience interval of 10 epochs were used to prevent over-fitting. These models have been trained with a batch size of 10 training instances. Given the elevated

number of computations required by recurrent architectures, we train these models on two *Graphics Processor Unit* (GPU): NVIDIA K40c and NVIDIA GEFORCE GTX 1080.

### C. Defining the metric for evaluation.

The reported metric is given in terms of *accuracy*, defined as:

$$Acc(\%) = \frac{C - I}{T} * (100) \quad (12)$$

where  $C$  is the number of correct predictions,  $I$  the number of insertion errors and  $T$  the total number of events present in the test. With this setup, we determine the performance of recurrent architectures, and provide an in-depth study of how these models generalize to recent seismic periods at *Deception Island* volcano. After a certain length of time, the dynamics of the volcano can change, causing certain characteristics of the events differ from previous ones even being the same type of event. Therefore, training the models with volcano-seismic data from 1994 to 1996 campaigns, and evaluate those architectures with data from 2017, permits us to test if recurrent architectures generalize across different seismic campaigns.

### D. Adding geophysical knowledge to RNN predictions

At each time step, recurrent architectures assign a label for each incoming frame. Whilst from a machine learning perspective this procedure is correct, it may not consider the nature of volcano-seismic signals (see section III). Based on geophysical knowledge of *Deception Island* volcano, a set of rules can be incorporated to improve the interpretability of all models: the average duration of seismic events allows to check that predictions are coherent with the expected lengths of events. In Figure 3, histograms and average duration for each volcano-seismic event are depicted (see section III). These values have been used to establish the following rules:

- *Predictions of events with very short duration in-between two well recognized events*: in this case, spurious incoming frames have triggered the recognition of an event without the expected duration. We correct the wrong frame by assigning the highest class probability between the two well recognize events.
- *Many consecutive predictions with different labels and short duration, in-between well recognized events*: seismometers record overlapped events, which translate into signals with very heterogeneous frequency contents. Thus, RNNs architectures output a prediction for each incoming frame, switching between overlapped events. To solve this, we introduced the class *unknown event*: a special event considered as an insertion, that can be later analyzed by geophysics experts.

## V. RESULTS AND DISCUSSION

This section reports the classification results obtained by the recurrent architectures described in subsection II. Besides the attained classification results, an in-depth review of how these

Table I  
PER-FRAME ACCURACY (ACC %) PERFORMANCE OBTAINED BY VANILLA-RNN, GRU AND LSTM ARCHITECTURES WITH RAW DATA, LPC AND LFB.

	vanilla-RNN	GRU	LSTM
Raw	73.83	77.10	77.64
LPC	77.15	80.25	79.76
LFB	79.83	84.07	83.56

Table II  
ARCHITECTURE DIAGRAM SHOWING THE BEST TOPOLOGIES OBTAINED BY VANILLA-RNN, GRU AND LSTM ARCHITECTURES WITH RAW DATA, LPC AND LOG-FILTER BANKS FEATURES.

	vanilla-RNN	GRU	LSTM
Raw	170	110	240
LPC	250	220	290
LFB	60	20	130

models work internally, and how they generalize to recent seismic data is also presented. In addition, we perform an experiment with recent seismic data. After a careful parameter optimization process, results are presented for the best obtained models in terms of accuracy. Best models have been found with 20 hidden units for the GRU, 60 hidden units for the vanilla-RNN and 130 hidden units for the LSTM architecture using LFB features (Table II).

### A. General performance of the system

Table I shows per-frame recognition results for parameterized and non-parameterized seismic data. Despite promising results have been attained using raw data, notice that LFBs features yield to higher results. Based on these results, we decided to develop this work using LFB features, but we also include the LPC approach results.

Table III shows  $Acc$  and  $Cor$  with LPC and LFB features, for the best obtained models. The term  $Acc$  is computed as equation 12, and  $Cor$  refers to the number of correct events with respect to the total. Average performances of all systems are also included for the four test sets.  $Acc$  results show that recurrent architectures attain similar performance, with 74.15 % for the vanilla RNN, 79.28 % for the LSTM and 81.76 % for the GRU. The training times obtained by these models are summarized in Table IV.

Some insertion errors might be produced by the capability of the model to detect information not seen by human operators. Deletion errors evidence that models are missing events from the input data stream. Thus, if we consider only the events correctly predicted by the models ( $Cor$ ), vanilla RNN achieves a classification accuracy of 89.54 %, whilst LSTM and GRU attain 93.81% and 91.60 % respectively. An in-depth analysis of these results must be done to provide a better understanding of how RNN architectures detect seismic data.

### B. Detailed analysis of results

Notice that the number of inserted events is much greater than the number of deleted events. Nature of the seismic data might play an essential role in the elevated number of those insertions: the RNNs detect overlapped events, which

Table III  
PERFORMANCE OBTAINED BY RNNs ARCHITECTURES. RESULTS WITH THE BEST CONFIGURATIONS ARE REPORTED.

		LFB		LPC	
		Acc (%)	Cor (%)	Acc (%)	Cor (%)
RNN-Vanilla	test <sub>1</sub>	82.01	95.23	84.12	92.06
	test <sub>2</sub>	75.58	88.95	72.09	84.88
	test <sub>3</sub>	69.10	88.20	70.78	87.07
	test <sub>4</sub>	69.94	85.79	76.50	85.79
	avg(%)	74.15	89.54	75.87	87.45
GRU-RNN	test <sub>1</sub>	85.18	93.65	79.36	94.17
	test <sub>2</sub>	77.32	87.20	71.51	86.62
	test <sub>3</sub>	83.14	92.13	71.91	87.64
	test <sub>4</sub>	81.42	93.44	75.95	89.07
	avg(%)	81.76	91.60	74.68	89.37
LSTM-RNN	test <sub>1</sub>	85.18	97.88	83.06	91.53
	test <sub>2</sub>	77.32	90.69	74.41	84.30
	test <sub>3</sub>	79.77	92.69	80.33	88.76
	test <sub>4</sub>	74.86	93.98	68.85	87.97
	avg(%)	79.28	93.81	76.66	88.14

Table IV  
TRAINING TIMES OBTAINED BY THE BEST CONFIGURATIONS USING VANILLA-RNN, GRU AND LSTM ARCHITECTURES WITH LFB AND LPC FEATURES. VALUES ARE EXPRESSED IN SECONDS.

	vanilla-RNN	GRU	LSTM
LPC	2600.21	8086.50	27777.81
LFB	437.21	577.62	6810.25

increases the quantity of frames labelled as *unknown events*, and therefore, the insertions. Another hypothesis suggests that RNNs are detecting events that were not originally recognized by expert geophysicists. We have performed a detailed analysis of some signals in which the models have the highest number of insertions. Figure 5 depicts an example of them:

- Figure 5(a): the labels associated with this part of signal in the dataset are SIL-TRE. However, vanilla-RNN, LSTM and GRU recognize them as SIL-VT-TRE. After a posterior supervision by a geophysical expert, we can consider as correct the output obtained by the classifiers. At the beginning of the tremor, a short and overlapped VT can be recorded by the seismometer, but the signal has been labelled as a TRE since often, the source of these types of tremors are preceded by a small earthquake. In our classification system based on the labels originally provided, the event VT had been erroneously considered an insertion, decreasing the performance of the architectures.
- Figure 5(b): the labels associated with this chunk of signal are SIL-VT-SIL-LP. The second event labelled as silence (SIL) has a duration shorter than the range of duration considered in the grammar applied, being able to be considered as part of the coda of the VT. Therefore recurrent architectures ignore it and return the classification sequence as SIL-VT-LP. After applying the grammar, they return SIL-HYB, introducing a deletion error.
- Figure 5(c) and 5(d): these spectrograms correspond

to two different signals, SIL-TRE and SIL-HYB-TRE, which all architectures recognize as SIL-TRE, introducing several deletions on the classification accuracy. These deletions are easy to explain: by looking at both spectrograms, the presence of HYB events can not be easily distinguished, and waveform is needed to classify them. In this case, human factor has an essential role: some labels in this type of signals include the HYB events, whereas other not, which depends of the geophysical subjectivity of the human operator.

### C. Generalization capabilities of RNNs for recent campaigns.

One of the most important challenges in automatic volcano-recognition systems is to build robust computational models that can easily re-adapt themselves to highly dynamical internal seismic sources inside volcanoes. Thus, even if volcano properties do change over time, the system trained with data from past campaigns should be able to provide an efficient monitoring.

A 3.5 hours seismic record from the 2016-2017 Spanish Antarctic campaign (January 2017) at *Deception Island* has been used as testing data. All signals have been filtered between 1 and 50 Hz, in order to work in the same frequency range as the dataset described in subsection III-C. This dataset is selected from a raw seismic records, with unknown events, and no prior human supervision. This might influence the predictions, as this new dataset contains events that may differ from the carefully chosen prototype events used to train the system-. Best vanilla-RNN, LSTM and GRU are trained on 1995-1996 and 2001-2002 campaigns, have been tested with data from this recent seismic survey.

In order to compare the results with other architectures, several experiments using HMM -Hidden Markov Models have been tested. In doing so, we have used the Hidden Markov Model Toolkit (HTK) [66]. Two were the approaches followed:

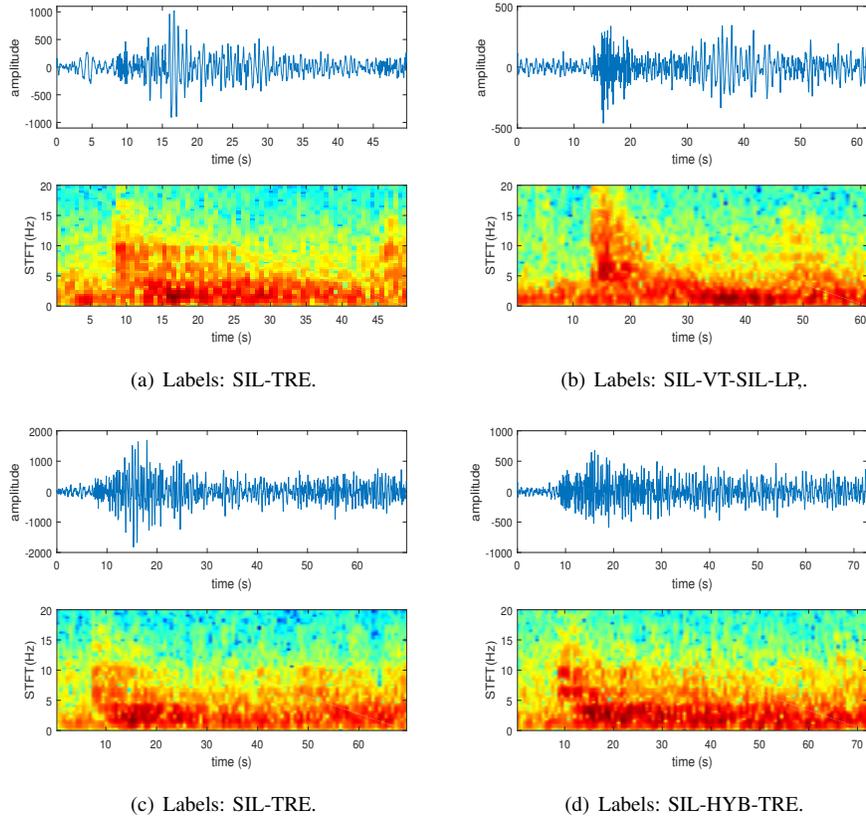


Figure 5. Spectrogram of seismic signals in the *Deception Island* volcano dataset. These spectrograms were selected from the test set to explain the high number of insertions by the recurrent architectures.

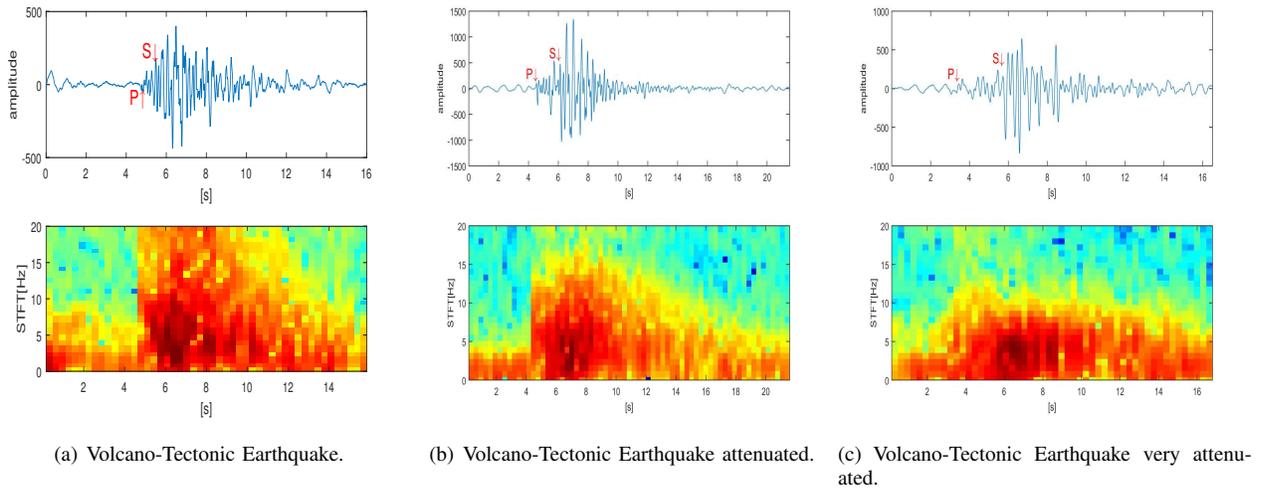


Figure 6. Attenuation effects with data from the 2016-2017 Spanish Antarctic campaign. Seismograms and spectrograms which summarize how seismometer location affects recorded shape and wave-field characteristics of volcano-seismic signals.

Table V  
CLASSIFICATION PERFORMANCE OBTAINED BY RNNs WITH TEST DATA FROM 2016-2017 SPANISH ANTARCTIC CAMPAIGN.

	LFB	
	Acc (%)	Cor (%)
LSTM-RNN	75.22	80.42
GRU-RNN	54.12	74.31
RNN-Vanilla	54.12	77.37
HMM	42.55	60.79

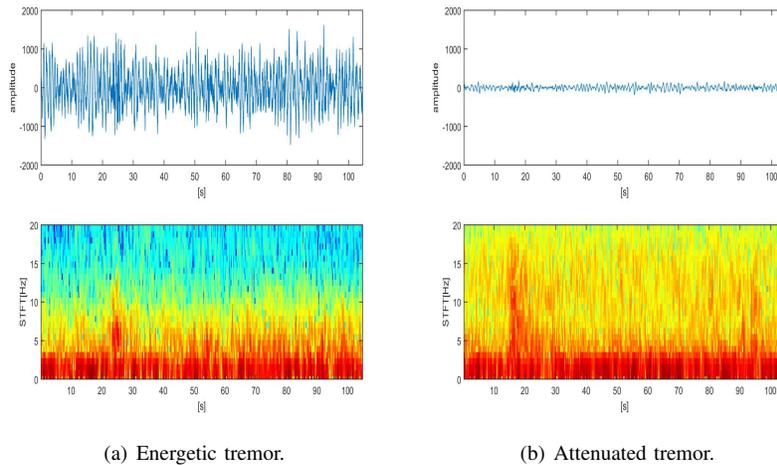


Figure 7. Attenuation effects. Seismograms and spectrograms summarizing how two different volcanic tremors, with similar source mechanism in their frequency pattern as shown in their spectrogram, but with evident differences in the energy level due to attenuation effects.

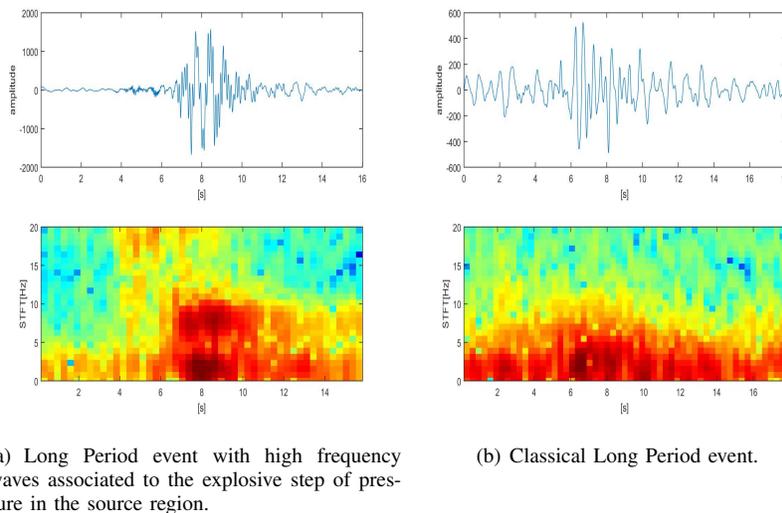


Figure 8. Source effects. Seismogram and spectrogram summarizing how the source of the seismic signal influences the recorded waveform and the spectral content.

- Finding a single or general purpose competitive configuration. To do that, we evaluated models with 5,9 and 11 hidden states. In this sense, to model the emission probabilities for a feature vector in any state, we evaluated between 1 and 16 multivariate Gaussian probability density (pdfs) functions with diagonal covariance matrices.
- Taking into account the wide variability of the volcano-seismic signals, we also considered the implementation of a HMM for each kind of event, that is, knowing that

each event has different temporal features, we proposed to use different model configurations based on the average duration of events. For relatively short events as LP and VT (less than three minutes), we proposed models with 5 states-model. For intermediate events, with duration between 3 and 7 minutes (SIL and HYB), we used 10 states, and finally, for large events like TRE, we chose 15 states. As above mentioned, to model the emission probabilities in any state, we evaluated between 1 and 16

multivariate Gaussian probability density functions (pdfs) with diagonal co-variance matrices.

The best result was obtained using the second approach and 12 multivariate Gaussian pdfs by state. Table V summarizes the test results. Whilst recognition results are good, there is a noticeable drop in performance by the GRU and vanilla-RNN architectures. After human expertise reviewing, this degradation could be explained from a geophysical perspective.

As mentioned in sections III-A and III-B, attenuation and source effects condition the recognition accuracy of all recurrent models. In this sense, Figure 6 depicts three VT events, recorded by the same seismic station, but with different hypocenter (source to receiver) distance (whose estimation has been done using the velocity model derived by [67], [68],[69], [70]). Notice how the attenuation effect introduces a bias in the recognition pattern of VT class:

- Figure 6(a) plots a VT with S-P time of around 1 second, i.e., a received-source distance lower than 3-4 km. As it can be visible in its spectrogram, the spectral content reach more than 20 Hz, but higher frequency content can be visible. RNNs architectures consider this event VT class with a probability higher than 66%.
- In Figure 6(b) the VT event was recorded with S-P time of around 2 seconds (distance close to 8 km). The spectral shape is different and the high frequency content does not reach beyond 15 Hz. All RNNs assign a lower class probability of VT (52%).
- Finally, in Figure 6(c), the S-P time is close to 3 seconds (at least 12 km of distance). The distance is not too large for the size of *Deception Island* volcanic environment, and many VT events can be recorded at further distances. However, in the spectrogram, the attenuation of the high frequency contents is evident, with the main presence of frequencies lower than 6 Hz. This yields VT signals with clear similarities with some of the trained LP events. In this case, RNNs architectures assign lower but similar (40%) per-class probabilities for VT and LP classes, leading to a confusion in the recognition.

Another important aspect related to attenuation effects can be observed by the peak to peak amplitude degradation of the recorded signal. Figure 7 depicts two examples of volcanic tremor (TRE) recorded by the same seismic station, at same scale for visualization purposes. Whilst the frequency pattern shown in their spectrogram suggest they have been generated by the same source mechanism, the noticeable differences in their energy level indicates strong attenuation effects. Figure 7(a) shows a volcanic tremor (TRE) of high amplitude that all RNNs architectures identified as a tremor. Instead, Figure 7(b) shows another volcanic tremor, which has very low amplitude, forcing all RNNs to recognize it as silence (SIL). However, expert geophysicist indicated as potential tremor. In this case, both (automatic system and operator) are correct: from the spectral point of view, this signal resembles a TRE, but can be considered as SIL since the amplitude of the signal is almost ten times lower than the expected amplitude level for a volcanic tremor. Results from Table V reflect this type

of inconsistency between human operator and RNN-based recognition systems, and it accounts for 5% of the total error (mostly deleted events). Thus, if we consider as correct the output of the RNNs, recognition accuracy would improve noticeably. These examples of how seismic attenuation could affect to the spectral content of VT and TRE events can be extensible to the case of HYB events, where in the nearest stations, a signal can be identified as HYB and in a more distant one as LP. For future some conditions must be included when a whole seismic network will be analyzed providing highest weight to the nearest stations of the seismic source.

The source effect addressed in section III-B is a direct influence in the accuracy drop by the vanilla-RNN and GRU in Table V. At the volcanic environment of *Deception Island*, LP events generated near the station have a first arrival of high frequency waves that resemble a hybrid event (HYB). Figure 8(b) shows a recorded LP event with a short distance to the generation source. It is visible the first arrival of this high frequency waves associated to the explosive step of pressure in the source region. This signal is identified as a package of high frequency signals (up to 20 Hz) in the spectrogram. No visible exponential decay in frequency is observed. After that, a well identify low frequency arrival is observed and directly associated to *classical* LP event. In Figure 8(a) we plot the same LP event recorded at a more distant station. The attenuation effect has decreased the energy of the high frequency waves (and in general of the total energy, see the lower vertical scale in the seismogram) showing a *classical* LP event.

In this scenario, it is plausible the potential confusion between near source LP and HYB events: Vanilla RNN and GRU, with less internal gates, recognize incoming frames being able to detect rapid seismic changes. Therefore, a potential VT event with short duration is inserted, decreasing the accuracy of vanilla-RNN and GRU. In the case of the LSTM, such short packages are ignored and no internal updates within their states cell memory is done. However, vanilla-RNN and GRU predictions should not be associated to a mistake, revealing the necessity to create a large dataset with near field LP events included.

## VI. CONCLUSION

This work focuses on how RNNs can be applied as statistical models to exploit temporal information of volcano-seismic signals, and explores their generalization capabilities over different seismic periods. Three recurrent architectures are studied: vanilla-RNN, LSTM and GRU. Experiments have been performed with seismic data from *Deception Island* volcano, with seismic records from 1994-1996 and 2001-2002, and further tested with data from a recent seismic survey in 2017 by the Spanish Antarctic scientific campaign. Using LFB features, attained results have shown that vanilla-RNN, LSTM and GRU classify volcano-seismic events with good accuracy, and memory cells (LSTM and GRU) enhance the detection of long-term signals. However, volcano-seismic data nature, specially path and source effects influences predictions. We incorporated a set of rules, based on geophysical knowledge,

which take into consideration the particularities of volcano-seismic data. These results demonstrate RNNs capability to generalize for recent recorded seismic data, presenting them as robust monitoring tools to enhance current early warning systems in real time.

#### VII. ACKNOWLEDGEMENTS

We would like to thank Prof. Hugo Larochelle and all the members of SMART Laboratory at University of Sherbrooke for their continuous support and advice when developing deep learning models in Theano. We thank *Instituto Andaluz de Geofísica* for providing us with *Deception* dataset and invaluable geophysical insight. This research was funded by TEC2015-68752-R (MINECO/FEDER).

#### REFERENCES

- [1] J.M. Ibáñez, 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: *Journ. Vol. Geoth. Res.* 187.3 (2009), pp. 218–226.
- [2] B. Chouet. “Volcano Seismology”. In: *Pure and Applied Geophysics* 160.3-4 (2003), pp. 739–788.
- [3] J. Wassermann. “IASPEI New manual of seismological observatory practice”. In: ed. by P. Bormann. Vol. 1. GeoForschungsZentrum Potsdam, 2002. Chap. Chapter 13: Volcano seismology, 42 pp.
- [4] M. Masotti, S. Falsaperla, H. Langer, S. Spampinato, and R. Campanini. “Application of Support Vector Machines to the classification of volcanic tremor at Etna, Italy”. In: *Geophysical research letters* 33.20 (2006).
- [5] S. Scarpetta, F. Giudicepietro, EC. Ezin, S. Petrosino, E. Del Pezzo, M. Martini, and M. Marinaro. “Automatic classification of seismic signals at Mt. Vesuvius volcano, Italy, using neural networks”. In: *Bull. Seism. Soc. Am.* 95.1 (2005), pp. 185–196.
- [6] A. Esposito, F. Giudicepietro, S. Scarpetta, and S. Khilnani. “A neural approach for hybrid events discrimination at Stromboli volcano”. In: *Multidisciplinary Approaches to Neural Computing*. Springer, 2018, pp. 11–21.
- [7] 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.
- [8] R. Carniel. “Neural networks and dynamical system techniques for volcanic tremor analysis”. In: *Annals of Geophysics* 39.2 (1996).
- [9] E. Del Pezzo, A. Esposito, F. Giudicepietro, M. Marinaro, M. Martini, and S. Scarpetta. “Discrimination of earthquakes and underwater explosions using neural networks”. In: *Bull. Seism. Soc. Am.* 93.1 (2003), pp. 215–223.
- [10] S. Diersen, E. Lee, D. Spears, P. Chen, and L. Wang. “Classification of seismic windows using Artificial Neural Networks”. In: *Procedia Computer Science* 00 (2001), pp. 1–10.
- [11] M. Kuroda, A. Vidal, A. Maria, and A. De Carvalho. “Interpretation of seismic multiattributes using a neural network”. In: *Journal of Applied Geophysics* 85 (2012), pp. 15–24.
- [12] AM. Esposito, F. Giudicepietro, L. D’Auria, S. Scarpetta, MG. Martini, M. Coltelli, and M. Marinaro. “Un-supervised neural analysis of very-long-period events at Stromboli volcano using the self-organizing maps”. In: *Bull. Seism. Soc. Am.* 98.5 (2008), pp. 2449–2459.
- [13] H. Langer, S. Falsaperla, M. Masotti, R. Campanini, S. Spampinato, and A. Messina. “Synopsis of supervised and unsupervised pattern classification techniques applied to volcanic tremor data at Mt Etna, Italy”. In: *Geophy. J. Int.* 178.2 (2009), pp. 1132–1144.
- [14] R. Carniel, L. Barbui, and AD. Jolly. “Detecting dynamical regimes by Self-Organizing Map (SOM) analysis: an example from the March 2006 phreatic eruption at Raoul Island, New Zealand Kermadec Arc.” In: *Bollettino di Geofisica Teorica ed Applicata* 54.1 (2013).
- [15] R. Carniel, AD. Jolly, and L. Barbui. “Analysis of phreatic events at Ruapehu volcano, New Zealand using a new SOM approach”. In: *Journ. Vol. Geoth. Res.* 254 (2013), pp. 69–79.
- [16] AD. Jolly, P. Jousset, JJ. Lyons, R. Carniel, N. Fournier, B. Fry, and C. Miller. “Seismo-acoustic evidence for an avalanche driven phreatic eruption through a beheaded hydrothermal system: an example from the 2012 Tongariro eruption”. In: *Journ. Vol. Geoth. Res.* 286 (2014), pp. 331–347.
- [17] M Ohrnberger. “Continuous automatic classification of seismic signals of volcanic origin at Mt”. In: *Merapi, Java, Indonesia: Doctoral thesis, University of Potsdam* (2001).
- [18] WP. Aspinall, R. Carniel, O. Jaquet, G. Woo, and T. Hincks. “Using hidden multi-state Markov models with multi-parameter volcanic data to provide empirical evidence for alert level decision-support”. In: *Journ. Vol. Geoth. Res.* 153.1-2 (2006), pp. 112–124.
- [19] M. Beyreuther, R. Carniel, and J. Wassermann. “Continuous hidden Markov models: application to automatic earthquake detection and classification at Las Cañadas caldera, Tenerife”. In: *Journ. Vol. Geoth. Res.* 176.4 (2008), pp. 513–518.
- [20] M. Bebbington. “Identifying volcanic regimes using hidden Markov models”. In: *Geophy. J. Int.* 171.2 (2007), pp. 921–942.
- [21] C. Benítez, J. Ramírez, J.C. Segura, A. Rubio, J.M. Ibáñez, J. Almendros, and A. García-Yeguas. “Continuous HMM-Based Volcano Monitoring at Deception Island, Antarctica”. In: *Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings. 2006 IEEE International Conference on*. Vol. 5. 2006, pp. V–V. DOI: 10.1109/ICASSP.2006.1661384.
- [22] L. Gutiérrez, J.M Ibáñez, G. Cortés, J. Ramírez, C. Benítez, V. Tenorio, and A. Isaac. “Volcano-seismic signal detection and classification processing using hidden Markov models. Application to San Cristóbal volcano, Nicaragua”. In: *Geoscience and Remote Sensing Sym-*

- posium, 2009 IEEE International, IGARSS 2009. Vol. 4. IEEE. 2009, pp. IV–522.
- [23] G. Cortés, R. Arámbula, L. Gutiérrez, C. Benítez, J. Ibáñez, P. Lesage, I. Alvarez, and L. Garcia. “Evaluating robustness of a HMM-based classification system of volcano-seismic events at Colima and Popocatepetl volcanoes”. In: *Geoscience and Remote Sensing Symposium, 2009 IEEE International, IGARSS 2009*. Vol. 2. IEEE. 2009, pp. II–1012.
- [24] J.M. Ibáñez, 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: *Journ. Vol. Geoth. Res.* 187.3 (2009), pp. 218–226.
- [25] M. Beyreuther and J. Wassermann. “Hidden semi-Markov model based earthquake classification system using weighted finite-state transducers”. In: *Nonlinear Processes in Geophysics* 18.1 (2011), p. 81.
- [26] M. Bicego, C. Acosta-Muñoz, and M. Orozco-Alzate. “Classification of seismic volcanic signals using hidden-Markov-model-based generative embeddings”. In: *IEEE Transactions on Geoscience and Remote Sensing* 51.6 (2013), pp. 3400–3409.
- [27] C. Hammer, M. Beyreuther, and M. Ohrnberger. “A seismic-event spotting system for volcano fast-response systems”. In: *Bull. Seism. Soc. Am.* 102.3 (2012), pp. 948–960.
- [28] D. Cárdenas-Peña, M. Orozco-Alzate, and G. Castellanos-Dominguez. “Selection of time-variant features for earthquake classification at the Nevadodel-Ruiz volcano”. In: *Computers & geosciences* 51 (2013), pp. 293–304.
- [29] Y. LeCun, Y. Bengio, and G. Hinton. “Deep learning”. In: *Nature* 521.7553 (2015), pp. 436–444.
- [30] J. Schmidhuber. “Deep learning in neural networks: An overview”. In: *Neural networks* 61 (2015), pp. 85–117.
- [31] R. Pascanu, T. Mikolov, and Y. Bengio. “On the difficulty of training recurrent neural networks”. In: *arXiv preprint arXiv:1211.5063* (2012).
- [32] A. Graves, A. Mohamed, and G. Hinton. “Speech recognition with deep recurrent neural networks”. In: *Acoustics, speech and signal processing (icassp), 2013 IEEE international conference on*. IEEE. 2013, pp. 6645–6649.
- [33] K. Gregor, I. Danihelka, A. Graves, D. Rezende, and D. Wierstra. “DRAW: A recurrent neural network for image generation”. In: *Proceedings of the 32 nd International Conference on Machine Learning, Lille, France, 2015. JMLR: WCP volume 37*. (2015).
- [34] T. Mikolov, S. Kombrink, L. Burget, J. Černocký, and S. Khudanpur. “Extensions of recurrent neural network language model”. In: *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on*. IEEE. 2011, pp. 5528–5531.
- [35] I. Sutskever, O. Vinyals, and QV. Le. “Sequence to sequence learning with neural networks”. In: *Advances in neural information processing systems*. 2014, pp. 3104–3112.
- [36] K. Cho, B. Van Merriënboer, C. Gulcehre, Caglar, D. Bahdanau, F. Bougares, H. Schwenk, Holger, and Y. Bengio. “Learning phrase representations using RNN encoder-decoder for statistical machine translation”. In: *arXiv preprint arXiv:1406.1078* (2014).
- [37] K. Kurach and K. Pawłowski. “Predicting dangerous seismic activity with Recurrent Neural Networks”. In: *Computer Science and Information Systems (FedCSIS), 2016 Federated Conference on*. IEEE. 2016, pp. 239–243.
- [38] A. Panakkat and H. Adeli. “Recurrent neural network for approximate earthquake time and location prediction using multiple seismicity indicators”. In: *Computer-Aided Civil and Infrastructure Engineering* 24.4 (2009), pp. 280–292.
- [39] M. Karamouz, S. Razavi, and S. Araghinejad. “Long-lead seasonal rainfall forecasting using time-delay recurrent neural networks: a case study”. In: *Hydrological Processes* 22.2 (2008), pp. 229–241.
- [40] J. Cao and J. Wang. “Global asymptotic stability of a general class of recurrent neural networks with time-varying delays”. In: *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications* 50.1 (2003), pp. 34–44.
- [41] R. Lee and J. Liu. “Tropical cyclone identification and tracking system using integrated neural oscillatory elastic graph matching and hybrid RBF network track mining techniques”. In: *IEEE Transactions on Neural Networks* 11.3 (2000), pp. 680–689.
- [42] D. Kumar, K. Raju, and T. Sathish. “River flow forecasting using recurrent neural networks”. In: *Water resources management* 18.2 (2004), pp. 143–161.
- [43] C. Balas, L. Coç, and L. Balas. “Predictions of missing wave data by recurrent neuronets”. In: *Journal of waterway, port, coastal, and ocean engineering* 130.5 (2004), pp. 256–265.
- [44] J. Simpson and T.J. McIntire. “A recurrent neural network classifier for improved retrievals of areal extent of snow cover”. In: *IEEE Transactions on Geoscience and Remote Sensing* 39.10 (2001), pp. 2135–2147.
- [45] G. Hinton, Li Deng, D. Yu, G-E. Dahl, A. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, and T. Sainath. “Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups”. In: *Signal Processing Magazine, IEEE* 29.6 (2012), pp. 82–97.
- [46] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio. “Empirical evaluation of gated recurrent neural networks on sequence modeling”. In: *arXiv preprint arXiv:1412.3555* (2014).
- [47] B. Chouet. “Dynamics of a fluid-driven crack in three dimensions by the finite difference method”. In: *J. Geophys. Res. Sol. Earth.* 91.B14 (1986), pp. 13967–13992.
- [48] J. Ibáñez, E. Carmona, J. Almendros, G. Saccorotti, E. Del Pezzo, M. Abril, and R. Ortiz. “The 1998–1999 seismic series at Deception Island volcano, Antarctica”. In: *Journ. Vol. Geoth. Res.* 128.1 (2003), pp. 65–88.

- [49] J. Ibañez, E. Del Pezzo, J. Almendros, M. La Rocca, G. Alguacil, R. Ortiz, and A. Garcia. “Seismovolcanic signals at Deception Island volcano, Antarctica: Wave field analysis and source modeling”. In: *J. Geophys. Res. Sol. Earth.* 105.B6 (2000), pp. 13905–13931.
- [50] J. W. Neuberg, H. Tuffen, L. Collier, D. Green, T. Powell, and D. Dingwell. “The trigger mechanism of low-frequency earthquakes on Montserrat”. In: *Journ. Vol. Geoth. Res.* 153.1-2 (2006), pp. 37–50.
- [51] D. Rouland, D. Legrand, M. Zhizhin, and S. Vergnolle. “Automatic detection and discrimination of volcanic tremors and tectonic earthquakes: An application to Ambrym volcano, Vanuatu”. In: *Journ. Vol. Geoth. Res.* 181.3-4 (2009), pp. 196–206.
- [52] D. Seild, R. Schick, and M. Riuscetti. “Volcanic Tremor at Etna, a model for hydraulic origin”. In: *Bulletin Volcanologique* 44.1 (1981), pp. 43–56.
- [53] J. Almendros, J. Ibañez, G. Alguacil, E. Del Pezzo, and R. Ortiz. “Array tracking of the volcanic tremor source at Deception Island, Antarctica”. In: *Geophysical Research Letters* 24.23 (1997), pp. 3069–3072.
- [54] G. Saccorotti, L. Zuccarello, E. Del Pezzo, J. Ibañez, and S. Gresta. “Quantitative analysis of the tremor wavefield at Etna Volcano, Italy”. In: *Journ. Vol. Geoth. Res.* 136.3 (2004), pp. 223–245.
- [55] A. Yeguas, A. Garcia, J. Almendros, R. Abella, and J. Ibañez. “Quantitative analysis of seismic wave propagation anomalies in azimuth and apparent slowness at Deception Island volcano (Antarctica) using seismic arrays”. In: *Geophy. J. Int.* 184.2 (2011), pp. 801–815.
- [56] B. Chouet. “Resonance of a fluid-driven crack: Radiation properties and implications for the source of long-period events and harmonic tremor”. In: *J. Geophys. Res. Sol. Earth.* 93.B5 (1988), pp. 4375–4400.
- [57] E. Del Pezzo, M. Simini, and J. Ibañez. “Separation of intrinsic and scattering Q for volcanic areas: a comparison between Etna and Campi Flegrei”. In: *Journ. Vol. Geoth. Res.* 70.3-4 (1996), pp. 213–219.
- [58] E. Del Pezzo, M. La Rocca, and J. Ibañez. “Observations of high-frequency scattered waves using dense arrays at Teide volcano”. In: *Bull. Seism. Soc. Am.* 87.6 (1997), pp. 1637–1647.
- [59] C. Martinez-Arevalo, F. Bianco, J. Ibañez, and E. Del Pezzo. “Shallow seismic attenuation and shear-wave splitting in the short period range of Deception Island volcano (Antarctica)”. In: *Journ. Vol. Geoth. Res.* 128.1 (2003), pp. 89–113.
- [60] J. Havskov, J. Peña, J. Ibañez, L. Ottemoller, and C. Martinez-Arevalo. “Magnitude scales for very local earthquakes. Application for Deception Island Volcano (Antarctica)”. In: *Journ. Vol. Geoth. Res.* 128.1 (2003), pp. 115–133.
- [61] P. Jousset, J. Neuberg, and A. Jolly. “Modelling low-frequency volcanic earthquakes in a viscoelastic medium with topography”. In: *Geophy. J. Int.* 159.2 (2004), pp. 776–802.
- [62] J. Prudencio, L. De Siena, J. Ibañez, E. Del Pezzo, A. Garcia-Yeguas, and A. Diaz-Moreno. “The 3D attenuation structure of deception Island (Antarctica)”. In: *Surveys in Geophysics* 36.3 (2015), pp. 371–390.
- [63] J. Almendros, E. Carmona, and J. Ibañez. “Precise determination of the relative wave propagation parameters of similar events using a small-aperture seismic array”. In: *J. Geophys. Res. Sol. Earth.* 109.B11 (2004).
- [64] Theano Development Team. “Theano: A Python framework for fast computation of mathematical expressions”. In: *arXiv e-prints* abs/1605.02688 (May 2016). URL: <http://arxiv.org/abs/1605.02688>.
- [65] 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*. Corvallis, Oregon, USA: ACM, 2007, pp. 473–480. ISBN: 978-1-59593-793-3. DOI: 10.1145/1273496.1273556. URL: <http://doi.acm.org/10.1145/1273496.1273556>.
- [66] S. Young and S. J. Young. *The HTK hidden Markov model toolkit: Design and philosophy*. University of Cambridge, Department of Engineering, 1993.
- [67] J. Ibañez, E. Del Pezzo, J. Almendros, M. La Rocca, G. Alguacil, R. Ortiz, and A. Garcia. “Seismovolcanic signals at Deception Island volcano, Antarctica: Wave field analysis and source modeling”. In: *Journal of Geophysical Research* 105.B6 (2000), pp. 13–905.
- [68] T. Ben-Zvi, W. Wilcock, A. Barclay, D. Zandomenighi, J. Ibañez, and J. Almendros. “The P-wave velocity structure of Deception Island, Antarctica, from two-dimensional seismic tomography”. In: *Journ. Vol. Geoth. Res.* 180.1 (2009), pp. 67–80.
- [69] D. Zandomenighi, A. Barclay, J. Almendros, J. Ibañez, W. Wilcock, and T. Ben-Zvi. “Crustal structure of Deception Island volcano from P wave seismic tomography: Tectonic and volcanic implications”. In: *J. Geophys. Res. Sol. Earth.* 114.B6 (2009).
- [70] E. Carmona, J. Almendros, I. Serrano, D. Stich, and J. Ibañez. “Results of seismic monitoring surveys of Deception Island volcano, Antarctica, from 1999–2011”. In: *Antarctic Science* 24.5 (2012), pp. 485–499.



**Manuel Titos** received the M.Sc. degree in Computer Engineering and the Master degree in Computer Networks Engineering from University of Granada, Spain, in 2012 and 2013, respectively. He currently is a PhD candidate at University of Granada, working on development of advanced signal processing algorithms for description and characterization of seismo-volcanic signals. His main research interests include deep learning techniques (machine learning) and computational intelligence in remote sensing signals, particularly

in the volcanic eruption early warning area. He also interested in data analysis and high-resolution image processing.



Angel Bueno received his MSc. in Telecommunications engineering with specialization in signal processing from University of Granada, Spain. He is working towards his Ph.D. degree in the intersection of deep learning and geophysics. His current main research interest is the optimization of deep learning algorithms, acoustic modeling for robust signal recognition and bayesian methods.



Luz García received the M.Sc. degree in telecommunication engineering from the Polytechnic University of Madrid, Spain, in 2000. She received the Ph. D. degree from the University of Granada, Spain, in 2008. After working as a Support Engineer for communication networks at Ericsson-Spain for 5 years, she joined a European research project at the University of Granada. She has belonged to the Department of Signal Theory, Telematics and Communications, University of Granada, Spain, since 2005, having worked

first as Assistant Professor and currently as an Associate Professor. Her research interests are signal processing, pattern recognition, and machine learning in the fields of speech and geophysics.



M. Carmen Benítez received the M.Sc. and Ph.D. degrees in physics from the University of Granada, Spain, in 1991 and 1998, respectively. From 1990 to 2004, she was a part of the Department of Electrónica y Tecnología de Computadores, Faculty of Sciences, University of Granada. Since 2004, she has been a part of the Department of Signal Theory, Telematics, and Communications, (ETSIT); RST as a Researcher and an Assistant Professor and then as an Associate Professor since 2003. Since 2015, she has been the Head

of the Department. She was a Visiting Researcher at the International Computer Science Institute, Berkeley, CA, USA, and with the USGS, Menlo Park, CA, USA. Her research interests include signal processing, geophysical signal processing, speech processing, machine learning, and pattern recognition.



Full Professor at the University of Granada in the area of Physics of the Earth. Specialist in Seismology and mainly in Volcanic Seismology, including High definition seismic tomography of volcanic areas using active and passive data, in velocity and attenuation. Seismo-volcanic sources, volcano tectonic earthquakes, long period events, volcanic tremor and explosive signals. Seismic antennas, inversion of volcano seismic signal, seismic pattern and seismic swarm analysis. Seismic signal analysis and processing, pattern

recognition, seismo-volcanic signals classification. Automatic and Real-time volcano-seismic signals classification and identification. Seismic attenuation, seismotectonic, site effects and seismic source. Tectonic activity of subducting regions. I have been responsible of more than 20 Research projects funded by the Spanish Government, European Union and other thirds counties focused in the above mentioned research lines. I have been the head of several seismic experiment carried out in active volcanoes such as Deception Island (Antarctica), Teide, Lanzarote and El Hierro (Canary Islands, Spain), Stromboli, Vesuvius, Campi Flegrei and Etna (Italy), Sao Miguel (Azores, Portugal), Copahue (Argentina), Colima (Mexico) and Fogo Volcano (Cape Verde) among other experiment in which I participated as researcher.