Discrimination of seismic waves produced by volcanoes using self-organizing maps

A Salazar¹, E Salazar², O Cadena¹, and R Hernández²,³

¹ Departamento de Física, Universidad de Nariño, Pasto, Colombia
² Laboratorio de Investigación Tecnológica en Reconocimiento de Patrones (LITRP), Universidad Católica del Maule, Talca, Chile
³ Centro de Investigación de Estudios Avanzados del Maule, Universidad Católica del Maule, Talca, Chile

E-mail: edwin.salazar@alu.ucm.cl, rhernandez@ucm.cl

Abstract. The analysis and classification of seismic records of volcanoes allow us to determine the alert state in which it is. A timely study of these signs can contribute to decision-making to safeguard the integrity of people in the face of a natural disaster. The present work applies a methodology that combines the analysis of linear prediction coefficients and artificial neural networks to classify earthquakes. Two types of earthquakes that come from Galeras Volcano, Colombia, are studied: volcano-tectonic and long-period. The classification is made using the clustering technique based on unsupervised learning. The signals are transformed using the linear prediction filter coefficients technique, which has the function of reducing the size of the vector that contains the original data. MATLAB software is used to generate a self-organizing network that handles clustering. The results show that the best alternative in unsupervised learning is to use the linear prediction coefficients of order 5, 6, and 7 to represent a seismic signal. For lower orders, the necessary information is not captured and for higher orders, noise information is shown.

1. Introduction
The Galeras Volcano is located in the department of Nariño, approximately 9 km west of San Juan de Pasto city, Colombia. According to historical records, it is the volcano with the highest activity in Colombia [1]. The volcano needs permanent surveillance, given that in its surroundings, there are eight municipalities and more than seven townships in imminent danger. Hence it has daily documentation that is ongoing of its activity [1, 2]. Different types of earthquakes have been recorded, among which are: volcano-tectonic (VT) earthquakes, that are associated with rock fracture, and long-period (LP) earthquakes, which are generally produced by the transitory movement of the fluid [3–5].

The Observatorio Vulcanológico y Sismológico de Pasto (OVSP), Colombia, processes and classifies the seismic information, aiming to account for the volcanic activity for the taking of the different security to safeguarding people who are living near the volcano. The actions performed by OVSP require a significant amount of time for the constant supervision of one by one of the seismic records, being the time a vital factor when issuing an alert. Therefore, the study of different mechanisms that minimize the time in the analysis and classification of earthquakes contributes to the work done by OVPS [6]. There are different investigations associated with
the classification of earthquakes produced by volcanoes. In [7], for the first time, it proposes the combination of linear prediction filter coefficients (LPC) with artificial neural networks using supervised learning to solve the classification problem in volcanic seismology. In [6, 8] and [9], from its approaches is based part of the current research; the authors classify earthquakes using supervised learning, and different mechanisms for the extraction of characteristics are studied, considering the complete signal and signal fragments.

The present paper aims to propose a methodology to classify earthquakes as VT and LP through unsupervised learning. The proposed approach has many advantages compared to supervised learning; among the benefits are: the evolution of the network allows adapting to situations provided, no prior training is required, and eliminates the need for manual classification of training cases [10]. Computational analysis techniques are used, such as linear prediction analysis because it reduces the dimension of the input vectors to the classifier system [11] and the use of self-organized maps that uses neural networks that learn the topology and distribution of the data by grouping the different classes into clusters. The technique is based on unsupervised learning that automatically identifies groupings or clusters of elements according to one measure of similarity between these clusters [12].

2. Methodology
The proposed approach follow the methodology outlined below:

(i) The traces of earthquakes to be evaluated are selected.
(ii) Computational routines are carried out for the treatment and the pre-processing of the information.
(iii) An alternative representation of the signals is used to reduce the dimensionality of the input vectors employing LPC; later, they are passed into the classifier system.
(iv) The results obtained are analyzed in the self-organized maps, which are automatically generated by using the MATLAB nnstart GUI in the training phase.
(v) The maps (i.e. clusters) that are activated depending on the input information are labeled.
(vi) The effectiveness of the system for classifying new information is evaluated.

2.1. Selection of seismic records
The degree of similarity through their spectral and enveloping representation are the relevant aspects for selecting the samples in the training phase. Figure 1 and Figure 2 represent an LP and a VP earthquake, respectively, recorded by the OVSP seismographs with their corresponding spectrum: the Figure 1(a) corresponds to the original LP signal, and the Figure 1(b) depicts the spectrum and the envelope of the LP signal; the Figure 2(a) corresponds to the original VT signal, and the Figure 2(b) depicts the spectrum and the envelope of the VT signal. In the figures, it can be seen that the signals in the frequency domain show the differences between the two earthquakes. The selected samples were taken from 1046 traces of LP earthquakes and 713 traces for VT earthquakes; these signals come from the monitoring carried out by the OVSP of the Galeras volcano, Colombia. For the training phase, 500 samples of VT earthquakes and 500 samples of LP earthquakes were used; the remaining traces were used to validate the results.

2.2. Information pre-processing
The following activities are for the pre-processing information phase, designing and implementing computer routines:

- Removal of offset: in this stage, the signal moves to the origin of a coordinate system.
- Normalization of the units: the signal to this process oscillates over an interval between $[-1, 1]$ or $[0, 1]$. 
Filtering signals: signals with frequencies (in this case lower than 1 Hz) are eliminated by means of a digital filter, since in this frequency range, the signals do not have information of our interest.

![Figure 1](image1.png)  ![Figure 2](image2.png)

**Figure 1.** Representation of an LP-type earthquake: (a) corresponds to the original signal, and (b) depicts the spectrum and the envelope of the signal.

**Figure 2.** Representation of a VT-type earthquake: (a) corresponds to the original signal, and (b) shows the spectrum and the envelope of the signal.

2.3. Extraction of characteristic properties

The study is carried out based on the spectral bands where the most significant energy is concentrated since the essential information of this analysis is in these ranges. LPC represents the signals since it leads to a reduction in the vector dimension that contains the data and therefore benefits in processing the information by the computer, optimizing time.

2.4. Generation of training matrices

The training matrices to generate the self-organized maps are created from vectors that contain the spectral information of the earthquakes. In addition to the training matrices, it is necessary to generate test matrices; these can be created from the training matrices or from new information of which there is the certainty of the group to which it belongs. Then, the former matrices are applied in the validation phase of the selected networks.

The Figure 3 and Figure 4 show an example of a self-organized map trained with matrix 10 from the Table 1. Figure 3 shows the self-organizing map in which each hexagon in grey represents a neuron. Here a weight plane is indicated for each element of the input vector. In this case, a vector of 6 inputs corresponding to the order of the LPC was used. The figures represent the weights that connect each input to each of the neurons; the darker colors represent larger weights. If the connection patterns of two inputs are similar, it can be assumed that the inputs are highly correlated. Figure 4 shows the locations of the neurons in the topology and indicates how much training data is associated with each of the neurons. The topology is a 10 by 10 grid, so it has 100 neurons.
Figure 3. Connection patterns between the inputs and the neurons of the network.

Figure 4. Activation of neurons in the training process.

Table 1. Training matrices for the neural network using different configurations: depending on the number of examples (# Ej) and the LPC order.

| Matrix | LPC | # Ej VT | # Ej LP | Matrix | LPC | # Ej VT | # Ej LP |
|--------|-----|---------|---------|--------|-----|---------|---------|
| 1      | 4   | 100     | 100     | 18     | 7   | 300     | 300     |
| 2      | 5   | 100     | 100     | 19     | 8   | 300     | 300     |
| 3      | 6   | 100     | 100     | 20     | 9   | 300     | 300     |
| 4      | 7   | 100     | 100     | 21     | 10  | 300     | 300     |
| 5      | 8   | 100     | 100     | 22     | 4   | 400     | 400     |
| 6      | 9   | 100     | 100     | 23     | 5   | 400     | 400     |
| 7      | 10  | 100     | 100     | 24     | 6   | 400     | 400     |
| 8      | 4   | 200     | 200     | 25     | 7   | 400     | 400     |
| 9      | 5   | 200     | 200     | 26     | 8   | 400     | 400     |
| 10     | 6   | 200     | 200     | 27     | 9   | 400     | 400     |
| 11     | 7   | 200     | 200     | 28     | 10  | 400     | 400     |
| 12     | 8   | 200     | 200     | 29     | 4   | 500     | 500     |
| 13     | 9   | 200     | 200     | 30     | 5   | 500     | 500     |
| 14     | 10  | 200     | 200     | 31     | 6   | 500     | 500     |
| 15     | 4   | 300     | 300     | 32     | 7   | 500     | 500     |
| 16     | 5   | 300     | 300     | 33     | 8   | 500     | 500     |
| 17     | 6   | 300     | 300     | 34     | 9   | 500     | 500     |

2.5. Results validation
In order to validate the results of the proposed approach, VT and LP-type earthquake traces are included to evaluate each neural network; the above examples are not included in the training phase.

3. Results
The experiments considered for the classification of volcanic earthquakes were divided into three parts: the representation of the signals using their LPCs, the generation of the training matrices, and the validation of the results. In the following sections, we present the obtained results in each of them.
3.1. **Representation of signals using their linear prediction filter coefficients**

For the optimal representation of a signal using their LPCs, it is required to establish a proper order in the linear prediction. For this analysis, 30 signal traces corresponding to LP-type earthquakes and 30 traces from VT-type earthquakes were chosen, where the linear prediction order varies between 2 and 20. The results are shown in Figure 5 and Figure 6.

![Figure 5](image.png)

**Figure 5.** Variance as a function of the LPC order for 30 LP-type earthquake traces.

A range in the order of the LPC where the variances are simultaneously small is required. In Figure 5, the most appropriate range is around the order of LPC is 10, since the difference in variances is small. On the other hand, in Figure 6, the prediction of the upper range to 10 corresponds to considerable variances. The proper prediction order is within the defined interval between 4 and 10, where it is observed that both curves have a stable behavior and whose successive variance differences are relatively small.

![Figure 6](image.png)

**Figure 6.** Variance as a function of the LPC order for 30 VT-type earthquake traces.

3.2. **Generation of training matrices**

For the training phase, 34 matrices were created; the number of training examples for each matrix ranges between 100 and 500, where LPC order associated with each trace is in the range between 4 and 10, as shown in Table 1. All the traces that make up the training matrices were subjected to a digital signal processor. In this process, the offset was eliminated, the units were normalized, frequencies less than 1 Hz were extracted using the Butterworth filter of MATLAB and the characteristic properties were obtained. For the training of all the previous matrices, a self-organizing map of dimension 10x10 with 100 neurons was used.

3.3. **Results validation**

In results validation, 100 traces of VT-type earthquakes and 100 traces of LP-type earthquakes were included with 200 samples to evaluate the effectiveness of each neural network. The previous
signals were not included in the training phase described in Table 1. The effectiveness results of the different configurations are shown in Figure 7, where the training one intervenes and the of the LPC order.

In Figure 7, different plots correspond to different number of training traces for both LP and VT-type earthquakes: Figure 7(a) 100 traces, Figure 7(b) 200 traces, Figure 7(c) 300 traces, Figure 7(d) 400 traces, and Figure 7(e) 500 traces. For 100 earthquakes traces (Figure 7(a)), the best result is obtained for an LPC of order 9, which reaches a percentage of 88.5%, and the overall percentage average is 83.7%. With 200 traces (Figure 7(b)), an effectiveness up to 95.5% is achieved for an LPC of order 7. On the other hand, in Figure 7(c), an unstable behavior is observed with different orders of LPC. It should be noticed that for an order 6, the system reaches its best performance, but in order 7, the effectiveness drops drastically. For 400 traces (Figure 7(d)), the results show a significant variation in the order 4 to 5 of the LPC, where the effectiveness acquires its maximum and minimum values respectively; for higher orders, the effectiveness declines slightly. Finally, with 500 traces, (Figure 7(e)), the best result is obtained for an LPC of order 5, which reaches a percentage of 96.5%.

![Figure 7](image-url)

**Figure 7.** Results of effectiveness as a function of the LPC order by varying the number of training traces for both LP and VT-type earthquakes: (a) 100 traces, (b) 200 traces, (c) 300 traces, (d) 400 traces, and (e) 500 traces.

In unsupervised learning, the use of self-organized maps is one of the most applied techniques due to its versatility and reliability. During the development of this work, good results were obtained for the classification of LP and VT-type earthquakes. However, the clustering technique showed to be very susceptible to the quality of the training data, since one of the drawbacks of the method is that the system must locate the center of clusters. Hence, taking into account that the presented study was focused exclusively on spectral analysis, it is possible to improve the results by transforming the signals to other domains aiming to include more variety in the characteristics so that one set is differentiated from another.
4. Conclusion
The effectiveness in classifying traces of LP and VT-type earthquakes events of the proposed work varies from 77% to 97% in the best of cases. The variation of these results is a direct consequence of the quality of the training examples, the LPC order, and the total number of units included in each training matrix. The results of the effectiveness as a function of the order of the LPC show that the best alternative to represent a signal in terms of its LPC corresponds to orders 5, 6, and 7. The efficiency of the method in these orders of linear prediction lies in the fact that it was achieved to collect enough information to allow correct discrimination.

References
[1] Williams S 1997 Emplacement and petrological evolution of the andesitic dome of Galeras volcano, 1990-1992 Journal of Volcanology and Geothermal Research 77(1-4) 57-69
[2] Medina L, Arcos D, Battaglia M 2017 Twenty years (1990-2010) of geodetic monitoring of Galeras volcano (Colombia) from continuous tilt measurements Journal of Volcanology and Geothermal Research 344(15) 232-245
[3] Hidayati S, Ishihara K, Iguchi M 2007 Volcano-tectonic earthquakes during the stage of magma accumulation at the Aira caldera southern Kyushu Japan Bulletin of the Volcanological society of Japan 52(6) 289-309
[4] Kumagai H, Chouet B 1999 The complex frequencies of long-period seismic events as probes of fluid composition beneath volcanoes Geophysical Journal International 138(2) 7-12
[5] Brill K, Waite 2019 Characteristics of repeating long-period seismic events at fuego volcano Journal of Geophysical Research: Solid Earth 124(8) 8644-8659
[6] Curilem G, Vergara J, Fuentellaba G, Aucúa G, Chacón M 2009 Classification of seismic signals at Villarrica volcano (Chile) using neural networks and genetic algorithms Journal of Volcanology and Geothermal Research 180(1) 1-8
[7] Del Pezzo E, Esposito A, Giudicepietro F, Marinaro M, Martini M, Scarpetta S 2003 Discrimination of earthquakes and underwater explosions using neural networks Bulletin of the Seismological Society of America 93(1) 215-223
[8] Cueva L, Paillacho R, Valeria, Michelle V 2017 Towards an automatic detection system of signals at cotopaxi volcano DYNA 84(200) 176-184
[9] Cadena O 2011 Detección y Clasificación Automática de Registros Sísmicos en el Observatorio Vulcanológico y Sismológico de Pasto Utilizando Redes Neuronales Artificiales (Colombia: Universidad Nacional de Colombia)
[10] Olivas E, Guerrero J, Martinez M, Magdalena J, Serrano L 2009 Handbook of Research on Machine Learning Applications and Trends: Algorithms, Methods, and Techniques: Algorithms, Methods, and Techniques (Hershey: IGI Global) p 852
[11] da Silva, Wunsch D 2017 An information-theoretic-cluster visualization for self-organizing maps IEEE Transactions on Neural Networks and Learning Systems 29(6) 2595-2613
[12] Gorzalczany M, Rudzinski F 2017 Generalized self-organizing maps for automatic determination of the number of clusters and their multiprototypes in cluster analysis IEEE Transactions on Neural Networks and Learning Systems 29(7) 2833-2845