

Feature Extraction Using Circular Statistics Applied to Volcano Monitoring

César San-Martin^{1,4}, Carlos Melgarejo^{1,2}, Claudio Gallegos^{1,2}, Gustavo Soto³,
Millaray Curilem⁴, and Gustavo Fuentealba⁵

¹ Information Processing Laboratory, Department of Electrical Engineering,
Universidad de La Frontera. Casilla 54-D Temuco, Chile
csmarti@ufro.cl

² Observatorio Volcanológico de los Andes del Sur Dinamarca 691, Temuco, Chile
cmelgarejo@sernageomin.cl, cgallegos@sernageomin.cl

³ Center for Mathematical Model, Universidad de Chile Casilla 412-3, Santiago, Chile
gsoto@dim.uchile.cl

⁴ Department of Electrical Engineering, Universidad de La Frontera. Casilla 54-D
Temuco, Chile
millaray@ufro.cl

⁵ Department of Physics, Universidad de La Frontera. Casilla 54-D Temuco, Chile
gustavo@ufro.cl

Abstract. In this work, the applicability of the circular statistics to feature extraction on seismic signals is presented. The seismic signals are captured from Llaima Volcano, located in Southern Andes Volcanic Zone at $38^{\circ}40'S$ $71^{\circ}40'W$. Typically, the seismic signals can be divided in long-period, tremor, and volcano-tectonic earthquakes. The seismic signals are time-segmented using a rectangular window of 1 minute of duration. In each segment, the instantaneous phase is calculated using the Hilbert Transform, and then, one feature is obtained. Thus, the principal hypothesis of this work is that the instantaneous phase can be assumed as a circular random variable in $[0, 2\pi)$ interval. A second feature is obtained using the wavelet transform due to the fact that seismic signals present high energy located in low frequency. Then, in the range 1.55 and 3.11 Hz the wavelet coefficients were obtained and their mean energy is calculated as the second feature. Real seismic data represented using this two features are classified using a linear discriminant with a 92.5% of correct recognition rate.

Keywords: seismic classifications, feature extraction, circular statistic, wavelet transform.

1 Introduction

An adequate study of the activity of an active volcano requires the use of indirect methods for evaluating information related to the dynamics of magma [1]. The scientific literature has shown that volcanic activity can generate a wide range of seismic signals [2]. The study of the waveforms of these signals differentiate the

various types of source mechanisms of earthquakes [3]. Until the late twentieth century, most studies were restricted to the spectral analysis and its variation over time. Today, new processing techniques are being incorporated into the analysis, in an attempt to automate the identification of the most important patterns of seismic signals. The main structure used to classify them considers a preprocessing stage and a classification stage.

The first one deparates the signals and performs its spectrum, amplitude (energy) and waveform feature extraction [4]. [5] proposed a linear predictive coding to extract spectral features and a parameterizations of the signal to extract information about the waveform. In [6] autocorrelation functions obtained by the FFT represent the spectral content and the short-term average to long-term average ratio (LTA/STA) relationship is used to distinguish between peaks and signals of long duration, often with similar spectral content. Some studies incorporate many frequency, amplitude and waveform features and use genetic algorithms to search a representative feature subset that improves the classifier performance [7,4]. Other methods are the wavelet transform [8,9], cross correlation methods [10] and hidden Markov models [11].

As seismic signals are inherently nonlinear and nonstationary [12] techniques such as artificial neural networks (ANN) [13] are being incorporated to perform classification. In many works, ANNs have outperformed traditional methods of analysis. ANN were used to predict the evolution of the seismicity of the Vesuvius, Campi Flegrei, Etna and Hawaii volcanoes [14] or to discriminate between local earthquakes and other Mt Vesuvius volcanic signals [5] and also for classifying transient signals representative of the Montserrat Volcano [6]. Some studies show that support vector machines (SVM) [15] have performed better than ANN [16]. [17] used spectral characteristics of four types of tremor of the Etna volcano and conducted a comparative study in which the SVM outperformed the ANN. [18] obtained the same result to classify three different seismic events of the Stromboli volcano.

It should be noted in the previous studies that the methods of analysis are difficult to generalize and every work is specialized in a certain type of patterns related to specific volcanoes. In this paper, we consider the Llaima Volcano, located in Southern Andes Volcanic Zone at $38^{\circ}40'S$ $71^{\circ}40'W$, at Chile. In particular, the applicability of circular statistics as feature extraction is studied. This kind of signal processing technique has received little attention in the literature maybe due to the representability of the original signal by the instantaneous phase, estimated by the use of Hilbert transform. Here, a feature vectors based on circular summary statistics taken from the instantaneous phase of the seismic signals, and the mean energy of the wavelet coefficients are used to distinguish among different seismic event. In this work, we address discrimination among long period events (LP) and volcano tectonic (VT) earthquakes from three components of Llaima volcano.

This paper has been organized as follow. In section 2 the Llaima volcano description is introduced. In section 3, mathematical concepts and their application



Fig. 1. Llama volcano located in the South of Chile

to seismic signal are presented. In section 4, feature extraction proposed method applied to real data is shown. Finally, the conclusion and future work are presented in section 5.

2 Volcano Monitoring

Llama volcano is located in Southern Andes Volcanic Zone at $38^{\circ}40'S$ $71^{\circ}40'W$. It is located 82 km northeast of Temuco and 663 km southeast of Santiago de Chile. Llama volcano is one of the largest and more active volcanoes in Southamerica. Llama volcano is a stratovolcano, a steep sloped cone composed of alternating layer of solidified ash, hardened lava and rocks ejected by previous eruptions. It has frequent eruptions but fortunately moderate explosive eruptions with occasional lava flows. One of the most recent eruption occurred in January 1, 2008, producing a column of smoke skyward, released a plume of sulfur dioxide and forcing the evacuation of dozens of tourists from volcano's base. On July 1, 2008, another eruption forced the evacuation of hundreds of people from nearby villages. An eruption occurred on April 5, 2009, with pyroclastic flows, ash and lava seen on the slopes.

The goal of volcano monitoring is allowing to know what is happening in the volcano. This implies the study of the past activity and keep a close watch on any current eruptions. The data are obtained from stations near the volcano that register seismic information, deformation and other events in real time. The data

are collected and sent to the Southern Andes Volcano Observatory (OVDAS) located in Temuco, where the specialists analyze the information. OVDAS depends on the Servicio Nacional de Geología y Minería (SERNAGEOMIN), the public statement in charge to carry out the actions based on the results of volcano information analysis.

2.1 Seismic Signals

Volcanic activity generates seismic events whose name and nature are common to all of them. However each volcano has its own seismic activity, that is, the characteristics of the events are particular. In this paper, we address discrimination among long period events (LP) and volcano tectonic (VT) earthquakes from three components of Llaima volcano. The data used have been collected using a three-component wide-band seismic sensor located in south-western of the Llaima volcano (Lave station), around 7.4 km from the vent. The seismometric used is the Oyogeospace SEIS-Monitor. The Lave station operates in continuous registration mode with 16 bit at 50 sample per second. All the events have been classified by an expert of OVDAS.

VT event is an earthquake originated in the center of a volcano. The VT event is associated to the fracture of solid of the volcano or the conduits of fluid ascent. The temporal signal present high amplitude with exponential decay with a lower time duration than LP event. In frequency domain, the VT signals presents a component higher than 8 Hz initially, but the most important range corresponds to the 6 - 8 Hz. In Llaima volcano they are very important because they are present in volcanic eruptions principally. Due to this, the amount of VT events is lower than LP, and in our database we have only 85 samples of VT events. The principal disadvantage in VT detection is the low possibility of being identified due to the presence of other events or noise.

LP events are very important to be detected because the type of LP allows identifying particular families of seismic signals which allow to know more precisely what is happening inside the volcano structure. The principal characteristic of an LP event is the high energy located in low frequency, typically below 3 Hz. From this characteristic intuitively it is possible to obtain this feature using a bandpass filter in order to estimate the energy in this frequency range.

In order to identify LP and VT events we built a pattern with two features based on circular statistic and wavelet transform applied to the seismic signal. In the next section we show in detail the procedure to obtain each feature.

3 Mathematical Background

In this section we briefly review the statistics moments of circularly distributed signals, wavelet transform and their applications to real seismic data.

3.1 The Analytic Signal

In order to obtain the analytic signal from sampled real data, the Hilbert transform [19] is used. The analytic signal represent all real-valuated signals

as complex signals, which turn out to have especially attractive properties for signal processing. Let $x(t)$ be a signal in the time domain t . The analytic signal $x_a(t)$ is given by:

$$x_a(t) = x(t) + j\mathcal{H}\{x\}(t), \quad (1)$$

where $\mathcal{H}\{x\}(t)$ is the Hilbert transform of $x(t)$ given by [20]:

$$\mathcal{H}\{x\}(t) = \text{p.v.} \int_{-\infty}^{\infty} \frac{x(\tau)}{t - \tau} d\tau, \quad (2)$$

where \mathcal{H} is the Hilbert transform operator. Because of the pole at $t = \tau$, the Cauchy principal value p.v. of the integral is used. The analytic signal of (1) can be expressed in the form:

$$x_a(t) = \psi(t) \exp(j\theta(t)), \quad (3)$$

where $\psi(t) = |x_a(t)|$ is the amplitude envelope and $\theta(t) = \arg(x_a(t))$, the instantaneous phase of the signal $x(t)$, with $\theta(t) \in [0, 2\pi)$. Thus, from this analytic signal, it is possible to define in a unique way the concepts of instantaneous amplitude and instantaneous phase.

3.2 Circular Statistics

The key assumption of this work is representing the seismic signals by features extracted from their random circular variables.

Circular statistics is similar to linear statistics [21], but in this case, sample trigonometric moments of a random circular variable are estimated from a well-known probability density function. Let $\Theta = \{\theta_n\}$ be a set of instantaneous phase values, where $n = 1, \dots, N$, then the p^{th} -order sample trigonometric moment is given by:

$$\mu_p = \frac{1}{N} \sum_{n=1}^N \exp(jp\theta_n). \quad (4)$$

As with linear statistics, measures of spread and symmetry, i.e., variance and skewness, can be defined in terms of sample trigonometric moments. Then, the sample circular variance of the data set Θ is defined as:

$$\sigma^2 = 1 - |\mu_1|, \quad (5)$$

The sample circular skewness is defined as:

$$\gamma = \frac{|\mu_2| \sin(\arg(\mu_2) - 2 \arg(\mu_1))}{(\sigma^2)^{\frac{3}{2}}}, \quad (6)$$

where $\arg(\mu_p)$ denotes the angle of the (complex valued) p^{th} -order sample trigonometric moment ($p = 1, 2$), and σ^2 is as defined in (5). Circular kurtosis, a measure

of peakedness in the circular density, can also be defined with sample trigonometric moments as follows:

$$\kappa = \frac{|\mu_2| \cos(\arg(\mu_2) - 2 \arg(\mu_1)) - |\mu_1|^4}{(\sigma^2)^2}. \tag{7}$$

A comprehensive description of trigonometric moments and other statistics for circular data can be found in [22].

3.3 Wavelet Transform

Briefly, the discrete WT can reconstruct any signal $x(t)$ by the formula:

$$x(t) = \sum_m \sum_n \langle x, \psi_{m,n} \rangle \psi_{m,n}(t), \tag{8}$$

where $\psi_{m,n}(t) = a^{-m/2} \psi(a^{-m}t - nb)$ is the wavelet function and a and b are the scale and shift parameters respectively. The wavelet coefficients are given by $\langle x, \psi_{m,n} \rangle$ and the multi resolution decomposition is obtained when $a = 2$ and $b = 1$. In this case we have an orthogonal decomposition. Two sequences exist that satisfies:

$$h_n = \langle \phi_{0,0}, \phi_{-1,n} \rangle, \phi(t) = 2^{1/2} \sum_n h_n \phi(2t - n), \tag{9}$$

$$g_n = \langle \psi_{0,0}, \phi_{-1,n} \rangle, \phi(t) = 2^{1/2} \sum_n g_n \phi(2t - n) \tag{10}$$

where h_n and g_n are the low and high pass filters with 0.25 cutoff frequency when 1 is the sampling rate. Then, the approximation and detail wavelet coefficients for the level k with sampling rate of 2^{-k} are respectively:

$$A_k^{[x(t)]}(t) = \sum_n 2^{-k} \langle x(t), \psi(2^{-k}t - n) \rangle \psi(2^{-k}t - n), \tag{11}$$

$$D_k^{[x(t)]}(t) = \sum_n 2^{-k} \langle x(t), \psi(2^{-k}t - n) \rangle \phi(2^{-k}t - n), \tag{12}$$

that can be calculated using (9) and (10). In our case, the wavelet filter used corresponds to the Daubechies with 10 taps, the sampling rate is 50 samples per second, and the detail for $k = 4$ corresponds to the frequency range of interest, i.e., between 1.55 and 3.11 Hz. Calculating the power of this coefficient, the mean energy in this band is obtained in the form:

$$E = \int (D_4^{[x(t)]}(t))^2 dt. \tag{13}$$

By the use of multi resolution analysis [23] is possible to obtain a band pass component and then, obtain the energy over the wavelet coefficient in the corresponding level.

4 Results

Real seismic data was captured from Llaima volcano between May 2009 and June 2009 and used to test the performance of the proposed method. A database with 2509 samples has been built, containing 404 samples of LP events, 85 of VT events in one component and 404 no-event signals. A sample corresponds to data interval of 1 minute of registered data non-continuous between intervals captured from the vertical, north-south, and east-west components denoted as Z, N-S, and E-W, respectively. The data has been labeled by expert from OVDAS. In this database, only LP and VT events are presented.

For each segment a pattern \mathbf{x} is built using equation (4) and (13) forming the pattern $\mathbf{x} = (u_1, E)$. Then, the database contains the 404 values $\mathbf{x}_1, \dots, \mathbf{x}_{404}$ for LP, 404 for no-events and 85 for VT events. In this case, only the Z component is considered. In Fig. 2 we illustrate the scatter plot for all data set for LP, VT and no-event patterns. Clearly, the VT events present a separation from LP events and no-events signals considering only the first circular moment. As a preliminary approach to the classification problem and based on the distribution of the classes shown in Fig. 2, a linear classifier was designed. Its performance reached 100% for VT identification and more than 90% for LP identification. More sophisticated classifiers will be evaluated during the research ongoing.

The discrimination between LP and no-events is given only by wavelet energy. But it is possible to include the first circular moment in order to improve the performance of the classifier. Table 1 presents the results using only wavelet energy and the incorporation of the first circular moment in three components.

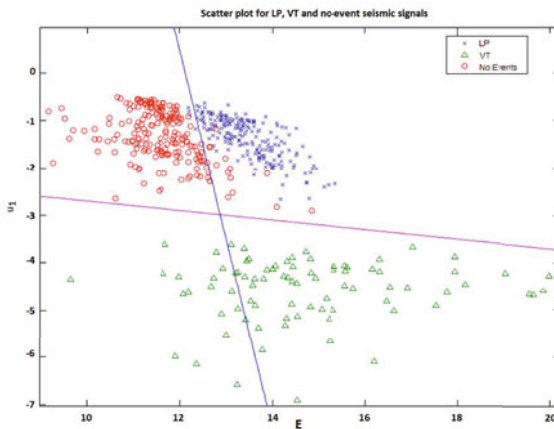


Fig. 2. Scatter plot for LP, VT and no-events patterns using wavelet energy and circular statistic

Table 1. The correct classification rate for LP and no-events identification

Component	% CC with $\mathbf{x} = E$	% CC with $\mathbf{x} = (u_1, E)$
Z	87.95	94.05
N-S	94.14	93.56
E-W	78.79	90.01

5 Conclusions and Future Work

In this paper we applied the circular statistic as a feature extraction technique to the seismic signal of the Llaima volcano. This implies that the instantaneous phase of seismic signal can be considered as circular random variable in the $[0, 2\pi)$ range. The first circular moment allows discriminating clearly the VT events from others with a high performance. For the case of LP events, a combination between energy located between 1.55 and 3.11 Hz is improved when the first circular moment is added. In fact, the correct identification increase from 86.96% to 92.54%. Future work include more extensive tests with a larger database, the evaluation of this feature extraction process to on line operation and to evaluate other classifier using neural networks among others.

Acknowledgments

The authors gratefully acknowledge to OVDAS Sernageomin for supplied the seismic data of Llaima Volcano, Temuco, CHILE.

References

1. Shick, R.: Volcanic Tremor- Source Mechanism And Correlation With Eruptive Activity. *Natural Hazard*, 125–144 (1988)
2. Minakami, T.: Seismology of Volcanoes in Japan. *Physical Volcanology. Developments In Solid Earth Geophysics 6* (1982)
3. Chouet, B.: Volcano Seismology As An Approach To Eruption Forecasting. Submitted To Nature, Draft Of (1993)
4. Orlic, N., Loncaric, S.: Earthquake-explosion discrimination using genetic algorithm-based boosting approach. *Computers and Geosciences* 36(2), 179–185 (2010)
5. Scarpetta, S., Giudicepietro, F., Ezin, E.C., Petrosino, S., Del Pezzo, E., Martín, M., Marinaro, M.: Automatic Classification of Seismic Signals at Mt Vesuvius Volcano, Italy, using Neural Networks. *Bulletin of the Seismological Society of America* 95(1), 185–196 (2005)
6. Langer, H., Falsaperla, S., Powell, T., Thompson, G.: Automatic classification and a-posteriori analysis of seismic event identification at Soufrière Hills volcano, Montserrat. *Journal of Volcanology and Geothermal Research* 153, 1–10 (2006)
7. Curilem, G.M.S., Vergara, J., Fuentealba, G., Acuña, G., Chacón, M.: Classification of Seismic Signals at Villarrica Volcano (Chile) using Neural Networks and Genetic Algorithms. *Journal of Volcanology and Geothermal Research* 180(1), 1–8 (2009)

8. Gendron, P., Nandram, B.: An empirical Bayes estimator of seismic events using wavelet packet bases. *Journal of Agricultural, Biological, and Environmental Statistics* 6(3), 379–402 (2001)
9. Erlebacher, G., Yuen, D.A.: A wavelet toolkit for visualization and analysis of large data sets. *Earthquake Research, Pure Appl. Geophys.* 161, 2215–2229 (2004)
10. Lesage, P., Glangeaud, F., Mars, J.: Applications of autoregressive models and time-frequency analysis to the study of volcanic tremor and long-period events. *Journal of Volcanology and Geothermal Research* 114, 391–417 (2002)
11. Ibanez, J.M., Benitez, C., Gutierrez, L.A., Cortés, G., García-Yeguas, A., Alguacil, G.: The classification of seismo-volcanic signals using Hidden Markov Models as applied to the Stromboli and Etna volcanoes. *Journal of Volcanology and Geothermal Research* 187(3-4), 218–226 (2009)
12. Dowla, F.U.: Neural networks in seismic discrimination. In: Husebye, E.S., Dainty, A.M. (eds.) *NATO ASI (Advanced Science Institutes). Series E*, vol. 303, pp. 777–789. Kluwer, Dordrecht (1995)
13. Bishop, C.: *Neural Networks for Pattern Recognition*. Oxford University Press, Inc., New York (1995)
14. Luongo, G., Marandola, C., Mazzarella, A.: Neural forecasting of seismicity and ground displacements in different volcanic areas. *Journal of Volcanology and Geothermal Research* 130, 133–146 (2004)
15. Vapnik, V.: *The Nature of Statistical Learning Theory*. Springer, New York (1995)
16. Cherkassky, V., Krasnopolsky, V., Solomatine, D.P., Valdes, J.: Computational intelligence in earth sciences and environmental applications: Issues and challenges. *Neural Networks* 19(2), 113–121 (2006)
17. Langer, H., Falsaperla, S., Masotti, M., Campanini, R., Spampinato, S., Messina, A.: Synopsis of supervised and unsupervised pattern classification techniques applied to volcanic tremor data at Mt Etna. *Italy Geophysical Journal International* 178(2), 1132–1144 (2009)
18. Giacco, F., Esposito, A.M., Scarpetta, S., Giudicepietro, F., Marinaro, M.: Support Vector Machines and MLP for automatic classification of seismic signals at Stromboli volcano. In: *Proceeding of the 2009 conference on Neural Nets WIRN 2009: Proceedings of the 19th Italian Workshop on Neural Nets. Frontiers in Artificial Intelligence and Applications*, vol. 204, pp. 116–123 (2009)
19. Nunes, J.-C., Naït-Ali, A.: Hilbert transform-based ECG modeling. *Biomedical Engineering* 39, 133–137 (2005)
20. Cohen, L.: *Time-frequency analysis*. Prentice-Hall, Englewood Cliffs (1995)
21. Fisher, N.I.: *Statistical Analysis of Circular Data*. Cambridge University Press, Cambridge (1995)
22. Jammalamadaka, S.R., SenGupta, A.: *Topics in circular statistics*. World Scientific Publishing, Singapore (2001)
23. Mallat, S.: *A wavelet tour of signal processing*, 2nd edn. Academic Press, London (1999)