



DISSIMILARITY-BASED CLASSIFICATION OF SEISMIC SIGNALS AT NEVADO DEL RUIZ VOLCANO

Mauricio Orozco^{1,3}, Marcelo E. García², Robert P.W. Duin³, and César G. Castellanos¹

¹ *Universidad Nacional de Colombia Sede Manizales, grupo de Control y Procesamiento Digital de Señales, Campus La Nubia, km 7 vía al Magdalena, Manizales, Colombia.*

² *INGEOMINAS - Observatorio Vulcanológico y Sismológico de Manizales, Avenida 12 de Octubre No. 15-47, Manizales, Colombia.*

³ *Information and Communication Theory Group, Mekelweg 4, 2628 CD Delft, Delft University of Technology, The Netherlands.*

Corresponding author: Mauricio Orozco, e-mail: morozcoa@unal.edu.co

ABSTRACT

Automatic classification of seismic signals has been typically carried out on feature-based representations. Recent research works have shown that constructing classifiers on dissimilarity representations is a more practical and, sometimes, a more accurate solution for some pattern recognition problems. In this paper, we consider Bayesian classifiers constructed on dissimilarity representations. We show that such classifiers are a feasible and reliable alternative for automatic classification of seismic signals. Our experiments were conducted on a dataset containing seismic signals recorded by two selected stations of the monitoring network at Nevado del Ruiz Volcano. Dissimilarity representations were constructed by calculating pairwise Euclidean distances and a non-Euclidean measure on the normalized spectra, which is based on the difference in area between spectral curves. Results show that even though Euclidean dissimilarities have advantageous properties, non-Euclidean measures can be beneficial for matching spectra of seismic signals.

Key words: Classification, dissimilarity, Nevado del Ruiz Volcano, seismic signals.

RESUMEN

La clasificación automática de señales sísmicas se ha llevado a cabo típicamente sobre representaciones de características. Trabajos de investigación recientes han mostrado que construir clasificadores sobre representaciones de disimilitud es una solución más práctica y, algunas veces, más precisa para ciertos problemas de reconocimiento de patrones.

En este artículo consideramos clasificadores Bayesianos construidos sobre representaciones de disimilitud. Mostramos que tales clasificadores son una alternativa viable y confiable para la clasificación automática de señales sísmicas. Nuestros experimentos fueron llevados a cabo sobre

Manuscript received July 5 2006.

Accepted for publication October 24 2006.

una base de datos que contiene señales sísmicas detectadas por dos estaciones seleccionadas de la red de monitoreo del Volcán Nevado del Ruiz. Las representaciones de disimilitud fueron construidas mediante el cálculo de distancias Euclidianas y de una medida no Euclidiana sobre los espectros normalizados, ésta última está basada en la diferencia de área entre curvas espectrales. Los resultados muestran que aunque las disimilitudes Euclidianas tienen propiedades ventajosas, las medidas no Euclidianas pueden resultar benéficas para comparar espectros de señales sísmicas.

Palabras claves: Clasificación, disimilitud, Volcán Nevado del Ruiz, señales sísmicas.

INTRODUCTION

Nevado del Ruiz Volcano (NRV) is capped by a large volume of snow and ice, forming a glacier which has a volume of about 1200~1500 million cubic meters. NRV has three craters: Arenas—the currently active vent—, and two parasite craters: Olleta and Piraña. Since the start of digital recording at the Volcanological and Seismological Observatory in Manizales (VSOM), a large and increasing amount of data has been recorded by the monitoring networks. Classification of seismic signals is a crucial issue in order to discover the interaction between volcanic earthquakes and volcanic processes. The database available from VSOM is suitable for applying automatic classification/learning techniques.

In this study, we consider three classes of seismic signals originating from NRV: Volcano-Tectonic (VT) earthquakes, Long-Period (LP) earthquakes and Icequakes (IC); of course, for every seismically monitored volcano, seismologists use their own classification with more detailed descriptions of every subtype of earthquakes (Zobin, 2003). VSOM staff currently classifies volcanic earthquakes by visual inspection; such a method imposes a great deal of workload for the seismic analysts. In consequence, an automatic classification tool dramatically reduces this arduous task and also makes the classification reliable and objective, removing errors associated to tedious evaluations and changing of personnel.

Among the applications of pattern recognition techniques to volcanic-seismic signals, two recent works are highlighted: Automatic classification of seismic signals at Mt. Vesuvius volcano, Italy (Scarpetta *et al.*, 2005) and automatic classification of seismic events at Soufrière Hills volcano, Montserrat (Langer *et al.*, 2006). Both of

them propose the application of Artificial Neural Networks (ANN) to classify seismic events. The former work uses a multilayer perceptron (MLP) to distinguish between VT events and transient signals due to other sources such as underwater explosions, quarry blasts, and thunders; spectral features and amplitude parameters are used for characterization. In the latter work, an ANN is used to classify five fundamental types of signals: VT events, regional (RE) events, LP events, hybrid (HB) events and Rockfalls (ROC); autocorrelation functions, high order statistical moments and amplitude ratios are introduced as features to the input nodes; a mismatch rate of 30% is reported, which was reduced up to 20% after a manual revision of the original a-priori classification. Typically, in the context of volcanic seismology, neural networks have been preferred rather than other classical statistical pattern recognition methods; they are still being used for discrimination of seismic signals, including modifications in the feature-based representation (e.g. the modified approach used in Benbrahim *et al.*, 2005). The popularity of neural networks models to solve pattern recognition problems has been primarily due to their seemingly low dependence on domain-specific knowledge and also because of the availability of efficient learning algorithms (Jain *et al.*, 2000).

Recently, a number of studies showed advantages of learning from dissimilarity representations as opposed to learning from feature-based representations (Duin *et al.*, 1998; Pękalska *et al.*, 2001; Pękalska and Duin, 2002; Paclík and Duin, 2003; Pękalska and Duin, 2005). A dissimilarity representation of objects, seismic events in our particular case, is based on pairwise comparisons and can be expressed as a $N \times N$ dissimilarity matrix $D(T, T)$, where each entry corresponds to dissimilarity between pairs of objects. Dissimilarity representations are more general

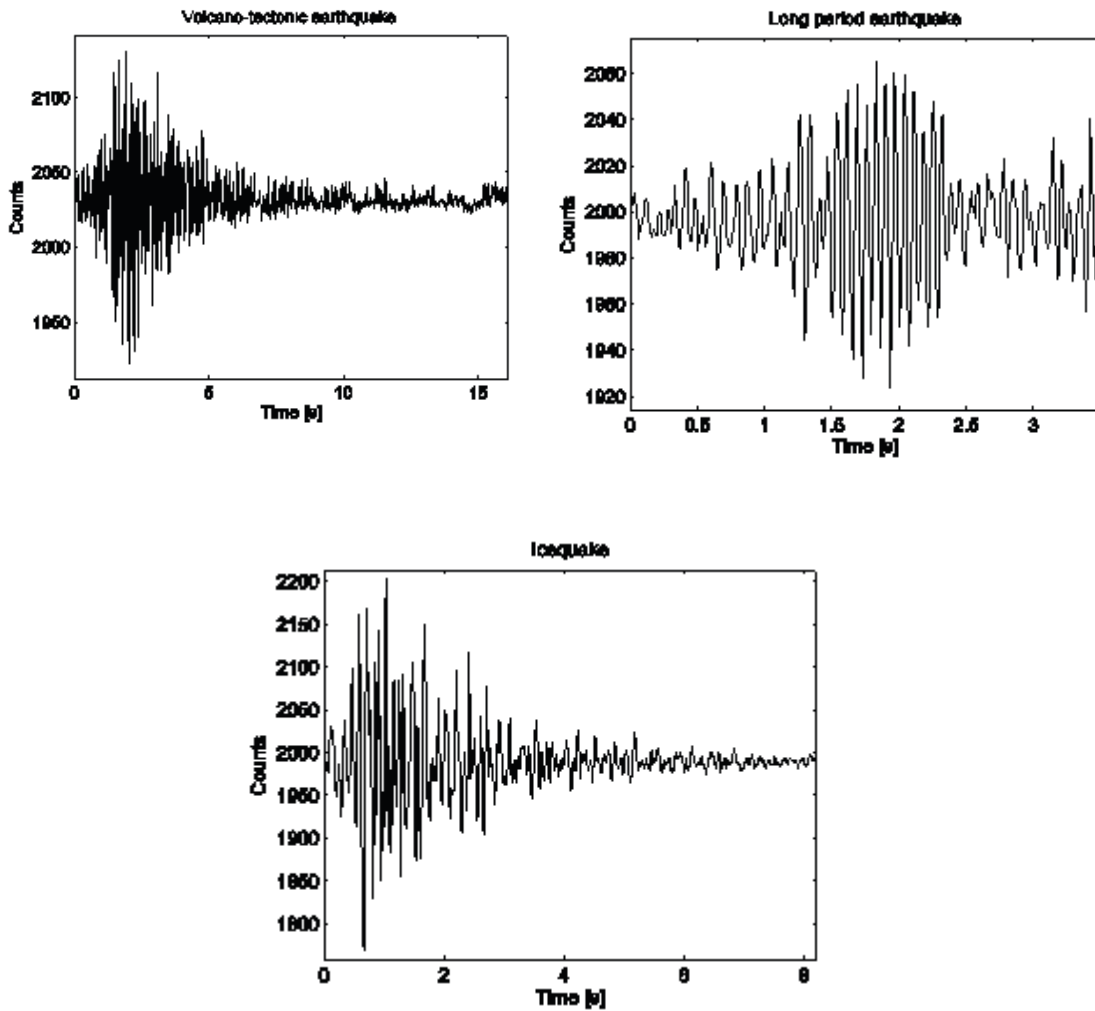


Figure 1. Typical waveforms of the three classes considered: VT, LP and IC events.

than feature-based representations; in fact, the notion of dissimilarity is more fundamental than that of a feature (Pełalska and Duin, 2005). For dissimilarities the geometry is contained in the definition, giving the possibility to include physical background knowledge; in contrast to feature-based representations which usually suppose a Euclidean geometry. This paper is devoted to explore dissimilarity representations to classify volcanic-seismic signals. In dealing with this particular problem, we advocate the dissimilarity-based classification of seismic signals as an advantageous and feasible alternative to the feature-based classification.

DATASET

The signals were selected from data collected by the monitoring network deployed by VSOM staff on NRV. The stations of the NRV monitoring network are located strategically; for instance near to glaciers and craters. Signals from two stations (Olleta crater station and Glacier station) have been selected for the experiments because, according to the experiences reported by VSOM staff, these two stations are a reference for the volcanic and ice-related events. The stations are located at distances of 4.08 km and 1.8 km from the active crater, respectively. Signals were digitized at 100.16 Hz sampling frequency by using 12 bits analogue-to-digital converter. A description of the NRV data set is provided in Table 1. Typical waveforms are shown in Figure

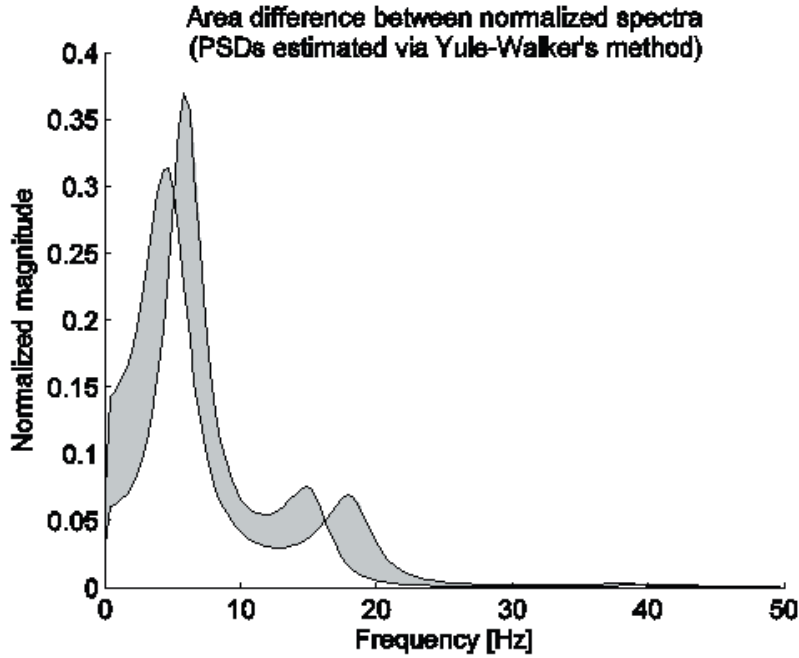


Figure 2. Dissimilarity measured as the difference between normalized spectra.

1.

TABLE 1. Composition of the data set (number of events per class). VT and LP events were detected by the Olleta crater station. IC events were detected by the Glacier station.

VT	LP	IC
483	580	782

VT and LP earthquakes are commonly detected at NRV. In our experiments, the VT and LP events are analyzed as a separate problem (the *Ruiz-LP,VT* two-class problem), and we also analyze all classes together (the multi-class *Ruiz-all* problem).

DISSIMILARITY REPRESENTATIONS AND CLASSIFIERS

Dissimilarity representations can be derived in many ways, e.g. from raw (sensor) measurements such as images, histograms or spectra or, from an initial representation by features, strings or graphs (Pełkalska *et al.*, 2006); nonetheless, the particular way in which dissimilarities are computed is crucial, and relies on the

additional knowledge that experts —volcanic seismologists— have about the problem.

The spectra of seismic signals are commonly used for classification and monitoring of volcanic activity. Since differences in spectral content allow a visual discrimination of different types of volcanic earthquakes (Zobin, 2003), we have calculated the spectrum for each signal by using two different approaches: (i) N-point Fast Fourier Transform (FFT) and, (ii) parametric estimation of the power spectral density (PSD). In such a way, we explore the difference between deriving dissimilarities from a data-based spectral estimation and from a model-based spectral estimation such as the Yule-Walker AR method. DC bias was removed before computing the spectra; in addition, when spectra are directly compared, they are required to be normalized and to have the same length. In consequence, considering the length of the shortest event and a length-resolution trade-off, we calculated 128-point spectra.

Two different dissimilarities measures have been computed between spectra: (i) pointwise Euclidean distance and (ii) area difference: the

area of non-overlapping parts (L^1 -norm) as shown in Figure 2.

The k Nearest Neighbor classifier (k -NN)

k -NN is considered a direct approach for dissimilarity-based classification. This rule classifies a new object by assigning it the class label most frequently represented among the k nearest prototypes (e.g., by finding the k neighbors with the minimum distances between the new object and all the prototypes). For $k = 1$, the rule is called 1-NN. Even though k -NN is asymptotically optimal in the Bayesian sense, it is sensitive to noise and erroneously labelled prototypes.

Linear and Quadratic normal density based classifiers

Previous studies (Pełalska *et al.*, 2001; Pełalska and Duin, 2002; Paclík and Duin, 2003) have shown that Bayesian (normal density-based) classifiers, particularly the linear (LDC) and quadratic (QDC) normal based classifiers, perform well in dissimilarity spaces. For a 2-class problem, the LDC based on the representation set R is given by

$$f(D(x,R)) = \left[D(x,R) - \frac{1}{2}(\mathbf{m}_{(1)} + \mathbf{m}_{(2)}) \right]^T \times C^{-1}(\mathbf{m}_{(1)} - \mathbf{m}_{(2)}) + \log \frac{P_{(1)}}{P_{(2)}} \quad (1)$$

and the QDC is derived as

$$f_{D(x,R)} = \sum_{i=1}^2 (-1)^i (D(x,R) - \mathbf{m}_{(i)})^T C_{(i)}^{-1} (D(x,R) - \mathbf{m}_{(i)}) + 2 \log \frac{P_{(1)}}{P_{(2)}} + \log \frac{|C_{(2)}|}{|C_{(1)}|} \quad (2)$$

where C is the sample covariance matrix, $C_{(1)}$ and $C_{(2)}$ are the estimated class covariance matrices, and $\mathbf{m}_{(1)}$ and $\mathbf{m}_{(2)}$ are the mean vectors, computed in the dissimilarity space $D(T,R)$, $P_{(1)}$ and $P_{(2)}$ are the class' a-priori probabilities. If C is singular, a regularized version must be used. In this study, the following regularization is used:

$$C_{reg}^1 = (1 - \lambda) C + \lambda \text{diag}(C) \quad (3)$$

We have fixed λ to be 0.01. Nonetheless, this regularization parameter should be optimized in order to obtain the best possible results for the normal density-based classifiers.

EXPERIMENTAL RESULTS

Experiments were conducted to compare the results of the k -NN rule and the LDC and QDC classifiers built on the dissimilarity representations described above. Experiments were performed 25 times for randomly chosen training and test sets. Since we are particularly interested in accuracy recognition rather than computational complexity and storage requirements, the entire training set T has been used as the representation set R . Nonetheless, R may be properly reduced by prototype selection procedures (Pełalska *et al.*, 2006). Training and testing sets were generated by selecting equal partitions for the classes.

The results of our experiments are shown in Figures 3 and 4. They present the generalization errors as a function of the number of randomly-chosen training objects. Figure 3 presents the results for four dissimilarity representations of the *Ruiz-VT,LP* two-class problem; similarly, the results for the *Ruiz-all* problem are shown in Figure 4. Standard deviations for averaged test errors decrease rapidly, varying around 0.02 after at least 10 training objects per class become available; for clarity reasons, standard deviations are not presented in Figures 3 and 4. Final errors and their standard deviations are summarized in Table 2.

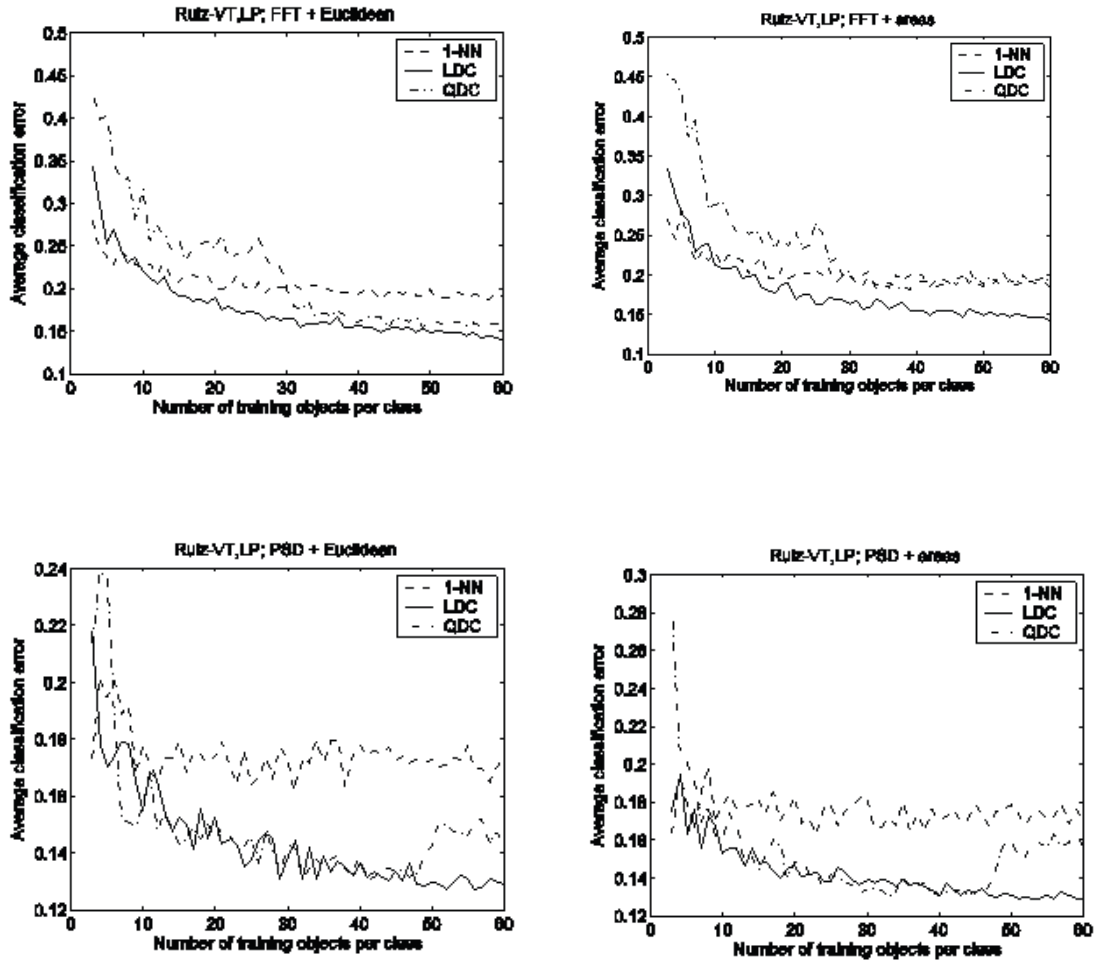


Figure 3. Ruiz-VT, LP data. Average classification error of the Bayesian classifiers and the 1-NN classifier in different dissimilarity spaces (FFT+Euclidean, FFT+areas, PSD+Euclidean and PSD+areas) as a function of the number of prototypes per class.

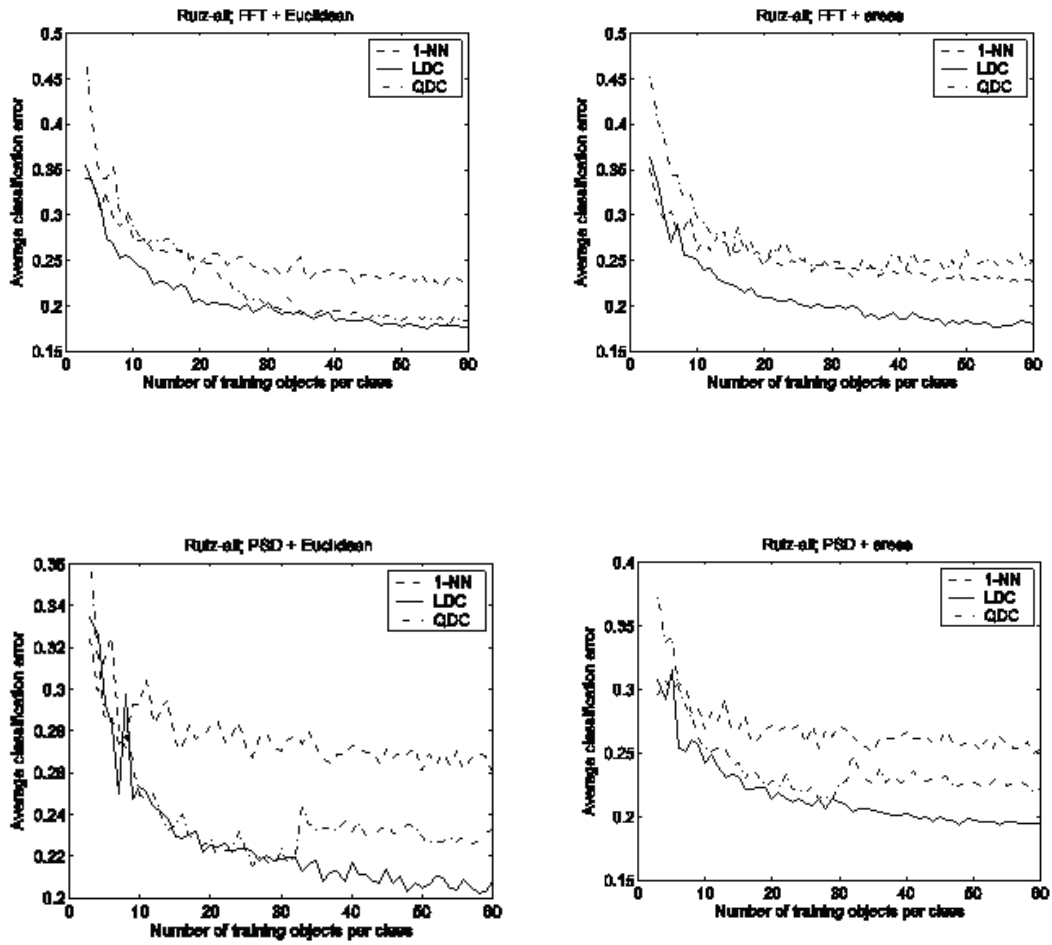


Figure 4. Ruiz-all data. Average classification errors of the Bayesian classifiers and the 1-NN classifier in different dissimilarity spaces (FFT+Euclidean, FFT+areas, PSD+Euclidean, and PSD+areas) as a function of the number of prototypes per class.

Table 2. Averaged final classification error (in %) with its standard deviation for the four representations and the three classifiers considered.

<i>Ruiz-VT,LP</i> two-class problem			
Representation	<i>k</i> -NN	LDC	QDC
FFT + Euclidean	19.85 (1.84)	14.79 (1.55)	16.36 (2.15)
FFT + Areas	19.90 (2.04)	14.94 (1.00)	19.35 (1.98)
PSD + Euclidean	17.99 (2.03)	13.35 (1.24)	15.45 (1.38)
PSD + Areas	17.48 (2.09)	13.40 (1.06)	16.23 (2.05)
<i>Ruiz-all</i> problem			
FFT + Euclidean	24.53 (1.66)	18.57 (1.13)	19.31 (1.33)
FFT + Areas	24.28 (1.52)	18.94 (1.02)	26.38 (2.73)
PSD + Euclidean	27.76 (2.08)	21.19 (0.80)	24.10 (1.47)
PSD + Areas	26.70 (1.67)	20.11 (1.25)	22.99 (1.85)

DISCUSSION AND CONCLUSIONS

We have explored and tested a dissimilarity-based strategy for classifying three different types of volcanic-seismic signals recorded by the monitoring network at NRV. Two classification problems were conducted: A two-class problem including VT earthquakes and LP seismic events, and a multi-class problem including ice-related seismic events. Four dissimilarity representations were derived by combining two different approaches for spectral estimation: N-point FFT and parametric PSD estimation, as well as two dissimilarity measures: Euclidean distance and area difference between spectral curves. These dissimilarity representations $D(T, T)$ allowed the usage of traditional statistical decision rules, particularly normal density based classifiers. The 1-NN rule was employed as a reference for performance comparison.

The two-class *Ruiz-VT, LP* problem seems the

easiest because it contains signals detected and identified by the same station (Olleta crater station); thus it is expected that sensor and noise conditions are the same influencing the subsequent steps for representation and classification. In addition, it is well known that, in general, multi-class problems are more difficult to deal with.

For the two-class problem, experiments based on parametric PSD estimation outperform those based on the FFT. This makes sense because event lengths are, in general, short and, consequently, a parametric spectral estimation yields a higher resolution; in addition, the autoregressive methods (AR) tend to adequately describe spectra of peaky data, which is precisely the spectral nature of many volcanic-seismic signals. In contrast, for the multi-class problem, the FFT yields better results but, in these particular cases, differences are not significant.

Our experiments confirm that Bayesian classifiers outperform the 1-NN classifier, when a sufficient number of prototypes are provided. The LDC constructed on the different dissimilarity representations, for both *Ruiz-VT,LP* and *Ruiz-all* problems, always outperforms the 1-NN rule. LDC accuracies for the *Ruiz-VT,LP* problem vary between 85% and 87% when the average classification error curve reaches a steady state. Similarly, classification accuracies for the *Ruiz-all* problem vary between 81% and 84%.

QDC shows accuracy loss when certain number of prototypes is provided. Therefore, a further study on a proper regularization for the QDC should be conducted in order to obtain an improvement of this classifier. LDC accuracies could be an intrinsic limit of our classification problem; however, a further study on other dissimilarity-based classifiers is needed as well as a re-analysis of the original a-priori classification, in order to find more suitable classifiers and to confirm the labels assigned by the experts.

ACKNOWLEDGEMENTS

This work was partially supported by the research project: “Técnicas de computación de alto rendimiento en la interpretación automatizada de imágenes médicas y bioseñales”. We thank the VSOM staff for providing the raw data set, Prof. Ricardo Henao for his valuable suggestions, and two anonymous reviewers for their useful comments.

REFERENCES

- Benbrahim, M., Daoudi, A., Benjelloun, K., and Ibenbrahim, A. (2005). Discrimination of Seismic Signals Using Artificial Neural Networks, *2nd World Enformatika Conference*, Istanbul, Turkey, February, 4-7.
- Duin, R. P. W., De Ridder, D., and Tax, D. M. J. (1998). Featureless pattern classification, *Kybernetika*. **34** (4). 399-404.
- Jain, A. K., Duin, R. P. W., and Mao, J. (2000). Statistical pattern recognition: a review, *IEEE Trans. Pattern Anal. Machine Intell.* **22** (1). 4-37.
- Langer, H., Falsaperla, S., Powell, T., and Thompson, G. (2006). Automatic classification and a-posteriori analysis of seismic event identification at Soufrière Hills volcano, Montserrat, *Journal of volcanology and geothermal research*. **153**. 1-10.
- Paclík, P. and Duin, R. P. W. (2003). Dissimilarity-based classification of spectra: computational issues, *Real Time Imaging*. **9**. 237-244.
- Pękalska, E., Paclík, P., and Duin, R. P. W. (2001). A generalized Kernel approach to dissimilarity based classification, *J. Mach. Learn. Res.* **5**. 175-211.
- Pękalska, E. and Duin, R.P.W. (2002). Dissimilarity representations allow for building good classifiers, *Pattern Recognition Lett.* **23**. 943-956.
- Pękalska, E. and Duin, R. P. W. (2005). *The Dissimilarity Representation for Pattern Recognition: Foundations and Applications*. World Scientific, Singapore.
- Pękalska, E., Duin, R. P. W., and Paclík, P. (2006). Prototype selection for dissimilarity-based classifiers. *Pattern Recognition*. **39**. 189-208.
- Scarpetta, S., Giudicepietro, F., Ezin, E. C., Petrosino, S., Del Pezzo, E., Martini, M., and Marinaro, M. (2005). Automatic classification of seismic signals at Mt. Vesuvius volcano, Italy, using neural networks. *Bulletin of the Seismological Society of America*. **95** (1). 185-196.
- Zobin, V. (2003). *Introduction to volcanic seismology*. Elsevier, The Netherlands.

