Spectral Characterization of Volcanic Earthquakes at Nevado del Ruiz Volcano Using Spectral Band Selection/Extraction Techniques

Mauricio Orozco-Alzate[1], Marina Skurichina[2], and Robert P.W. Duin[2]

[1] Departamento de Informática y Computación, Universidad Nacional de Colombia
Sede Manizales, km 7 vía al aeropuerto, Manizales (Caldas), Colombia,
`morozcoa@unal.edu.co`
[2] Information and Communication Theory Group, Delft University of Technology,
P.O. Box 5031, 2600GA Delft, The Netherlands
{`m.skurichina,r.p.w.duin`}`@tudelft.nl`

Abstract. Spectral content of seismic signals contains essential information for discriminating different classes of volcanic earthquakes. Such an information is largely redundant; therefore, a reduce number of spectral regions may provide almost the same description of the original events. By reducing the number of bands considered, the amount of data to be processed is significantly decreased and the interpretability of the characterization results is enhanced as well. We consider several spectral band selection methods in a two-class classification problem of volcanic earthquakes recorded at Nevado del Ruiz Volcano. Selection approaches have been compared to each other in terms of classification accuracy as well as by looking at the resulting spectral divisions. Detailed discussions about the technical considerations of the selection approaches as well as regarding their possible physical interpretations have been conducted. Results show that the sequential selection approach is the most flexible and powerful for classifying and characterizing volcanic earthquakes.

Keywords: Signal processing and analysis, statistical pattern recognition, seismic-volcanic events, spectral analysis.

1 Introduction

Nevado del Ruiz is an ice-capped volcano placed in the Colombian Andes. Its historical eruptions, including the deadly one in November 1985, produced mudflows which travelled through the river valleys until reaching the low-lying plains. When the mudflows from 1985 eruption came into the Lagunillas river valley, the city of Armero was destroyed and about 23.000 people were killed. Due to the tremendous disaster, the government created the volcanological observatories; before that, Colombian volcanoes were not daily and sufficiently monitored. Nevado del Ruiz volcano is currently studied by the Volcanological and Seismological Observatory at Manizales (VSOM), which for several years has registered a large amount of data related to seismic activity.

Seismic volcanic events correspond to manifestations of two types of physical processes: transport of fluids and fracture of solid rock. The former produces long-period (LP) events, that are essential for eruption forecasting; the latter originates volcano-tectonic (VT) earthquakes, usually observed as a first sign of renewed volcanic activity as well as during the active period itself. Analyzing those seismic signals is essential to achieve an understanding of eruptive processes in andesitic volcanoes [1]. In order to use the forecasting potential of LP events fully, one must first be able to distinguish their signatures from those of VT earthquakes, a task made difficult by the extreme heterogeneity of volcanic media [2]. Differences in spectral content allow a discrimination of different types of volcanic earthquakes [3]; in consequence, a spectral-based classification is a natural approach to face the problem. In addition, magma properties are fundamental to explain the volcanic eruption style as well as the generation and propagation of seismic waves [4]; the reverse also holds, and some physical insight into these properties may be derived from a spectral study.

The always increasing capacity of sensory systems and data storage provides high quality and densely sampled spectral measurements but, at the same time, such an improvement in data acquisition challenges the performance of analysis algorithms, e.g. classifiers of volcanic events, that have to deal with a huge amount of data to be processed. However, spectral information is largely redundant by nature, i.e. its intrinsic dimensionality is low. In order to reduce data redundancy and also to find an understandable and reduced number of discriminative regions, the spectral band selection/extraction techniques have a good potential and should be preferred over the standard feature extraction techniques like Principal Component Analysis (PCA), which does not take into account the connectivity between neighbouring spectral bins.

In this paper, we restrict ourselves to the problem of selecting discriminative spectral bands for classifying seismic-volcanic signals. Since which spectral band selection technique is preferred seems to be defined by the problem [5], we study the performance of 7 different band selection strategies, applied for spectral classification of seismic events at Nevado del Ruiz volcano. We compare these results with those obtained without any dimensionality reduction, i.e. providing the entire spectrum to the classifier, as well as when PCA is used to extract the useful discriminative information from the spectral data. We also consider the spectral regions extracted by the best performing spectral band selection technique and discuss their possible interpretations.

2   Data

The seismic signals were selected from data collected by the VSOM monitoring network deployed on the Nevado del Ruiz volcano [6]. The dataset is composed of 483 VT events and 580 LP events. Stations of the monitoring network are located strategically; for instance near to the glaciers and craters. Signals from one single station (Olleta crater station) were selected for the experiments because this station is a reference for the volcanic-related events. Besides, by considering

signals recorded at the same location, we reduce the influence of the so-called path effect. Olleta crater station is located at a distance of 4.08 km from the active crater. Signals were digitized at 100.16 Hz sampling frequency by using a 12 bits analogue-to-digital converter.

Spectra were computed by using two different approaches: (i) N-point Fast Fourier Transform (FFT) and, (ii) parametric estimation of the power spectral density (PSD). The first one is a data-based spectral estimation; the second one is a model-based spectral estimation, in particular we used the Yule-Walker AR method. In both cases, the direct current bias was removed before computing the spectra. According to the spectral estimation method, datasets are referred to as Ruiz-FFT and Ruiz-PSD respectively.

3   Spectral Band Selection Methods

Each particular spectral band selection technique finds the discriminative spectral regions by using a particular criterion and merges them according to a different function. Such a criterion or a merging function is crucial for the performance of the band selection strategy. For the sake of comparison under equal conditions, we use the same criterion and the same merging function for all the studied spectral band selection strategies.

In order to evaluate a discriminative capacity of the extracted spectral regions, we use the Mahalanobis Distance (MD) between data classes as a discriminant measure (criterion):

$$MD = (\mu_A - \mu_B)' \left(p\Sigma_A + (1 - p)\,\Sigma_B \right)^{-1} (\mu_A - \mu_B),\tag{1}$$

where μ_A, μ_B and Σ_A, Σ_B are the means and the covariance matrices of data classes A and B, respectively; p is the prior probability of the data class A. The larger MD, the larger discriminative capacity between data classes. We use the mean function, i.e. taking the average of spectral intensities in the region, to reduce the dimensionality of each considered spectral region to a single value representation. The following spectral band extraction/selection techniques are considered in this study.

Approach 1. *GLDB-TD.* It consists in partitioning the original p-dimensional spectra into smaller subspaces by using a top-down recursive algorithm [7]. First, the best place to split spectra into two parts is found by computing a discriminant measure between data classes. The criterion value obtained on the parent space is compared with the criterion values calculated on the children subspaces. If the child subspace has a higher discrimination than the parent space, then it is partitioned further. The partitioning is stopped when no child subspace shows an improvement in its discrimination capacity compared to the parent space. The GLDB-TD algorithm is fast, but the final set of spectral regions found is suboptimal, because the optimization is performed only in one-dimensional way: a discrimination capacity is evaluated for each spectral region separately but not for a total set of selected spectral regions.

Approach 2. *GLDB-BU.* It merges p original bands in larger subspaces by using a bottom-up recursive algorithm [7]. First, the best pair to merge among all possible pairs of neighbouring single bands is found by computing a discriminant measure between data classes. The criterion value obtained on the best merged band is compared with the criterion values calculated on its component subspaces. If the merged space has a higher discrimination than the component subspaces, the merge is accepted and we move to the next level. Elsewise, the merge is denied and we consider the second best pair to merge on this level. If no merge is found that gives the better discrimination than component subspaces, then the merging procedure stops. This strategy has the same limitation as GLDB-TD: the optimization is performed only for one spectral group.

Approach 3. *Sequential Partitioning (SP)* [8]. It also performs multidimensional optimization for the spectral region selection. First, spectra are partitioned into two parts by finding the best split (with the optimal criterion value over all possible partitions) in the space of two features extracted from the two spectral regions. When the first split location is anchored, we look for the second optimal split in such a way that the criterion value in a three-dimensional space (on three features extracted from the three spectral regions) is the largest over all possible locations for the second split. We fix the second split location and repeat the procedure until the desired number of spectral regions S is found. In this approach, all spectral bands are used in the partitioning of spectra. However, some of them may be uninformative - introducing only noise. It is good to remove them, as they may deteriorate the classification when they are included in the extracted spectral regions. One way to do this is described below.

Approach 4. *Sequential Partitioning and Elimination of uninformative spectral bands (SPE)* [8]. After a desired number of spectral regions S is found by the previous approach, we shrink the spectral regions removing uninformative bands. We proceed sequentially from the most left region to the most right one. In order to shrink the spectral region, we consider all possible subregions of the reduced size in the region and find the subregion with the largest criterion value in S-dimensional space (where one feature represents a shrunk subregion of the spectral region under consideration and the rest $S - 1$ features are extracted from the other $S - 1$ spectral regions which definitions are fixed for a moment). After shrinking the first spectral region, we anchor its new definition and move to the next spectral region in order to exclude uninformative spectral bands. This method does not guarantee the optimal shrinking for all regions in general, because it is highly dependent on the proceeding order of spectral regions.

Approach 5. *Sequential Selection (SS)* [8]. The discriminative spectral regions are selected sequentially one by one. At each step s ($s = \overline{1, S}$) we consider all possible region definitions (of arbitrary size) in spectra. For each definition we calculate the discriminant measure in s dimensions: one feature represents a current potential pretender for the most discriminative spectral region and the other $s - 1$ features are extracted from the previously selected

spectral regions. The region (a potential pretender) with the largest criterion value in s-dimensional space is picked as the most discriminative spectral region (in combination with the $s - 1$ previously found optimal regions). In this approach the overlapping and non-overlapping spectral regions may be selected. Some spectral bands might be not selected at all to participate in spectral regions.

Approach 6. *Sequential Selection of Non-overlapping discriminative spectral regions (SSN)* [8]. This approach is identical to SS but the overlapping spectral regions are not allowed to be selected.

Approach 7. *Floating Partition (FP)* [9]. First, spectra are uniformly partitioned to a predefined number of spectral regions S. At each step, we allow the borders between spectral regions to float one spectral bin aside from the current position. Among 3^S possible mutations we select the partition that provides the highest discrimination according to the selected criterion. We repeat the procedure until no improvement in discrimination capacity can be found. This method performs multivariate optimization by simultaneously adjusting all spectral regions. However, it is still a suboptimal procedure because the drifting step for region borders is limited to one spectral bin. The efficacy of this method can be improved by enlarging the drifting step d. But this leads to computational burdens because one has to rank $(2d + 1)^S$ cases at each step of the procedure.

4 Experimental Results and Discussion

For our experiments, training data sets with 100 objects per class are considered for the two spectral representations studied: FFT and PSD spectral estimations. Each time the training objects are chosen randomly from the total set. The remaining data are used for testing. The prior class probabilities are set to be equal as the data are unbalanced and the real prior class probabilities are unknown. To evaluate the performance of different feature selection/extraction methods for the datasets described above, we have chosen the Regularized Linear Classifier (RLC) [10] which constructs a linear discriminant function assuming normal class distributions and using a joint class covariance matrix for both data classes. The value of the regularization parameter used is equal to 10^{-8}. All experiments are repeated 20 times on independent training sets.

Obtaining a reduced number of meaningful spectral regions is particularly useful for interpretation as well as for the sake of visualization. Fig. 1 shows the mean generalization errors over 20 trials, considering a maximum of 10 extracted spectral bands, for the *Ruiz-FFT* and the *Ruiz-PSD* datasets respectively. The standard deviation of the reported results (the mean per two data classes) is about 0.01 for each considered case. Since the GLDB-TD and GLDB-BU algorithms terminate automatically according to a data-driven criterion, only a single point is given in each plot.

In the case of selecting a single spectral region, PCA (a single component for this method), SS and SPE strategies are the best performing ones; in particular, PCA is the best one for the *Ruiz-PSD* dataset. In this case there is no
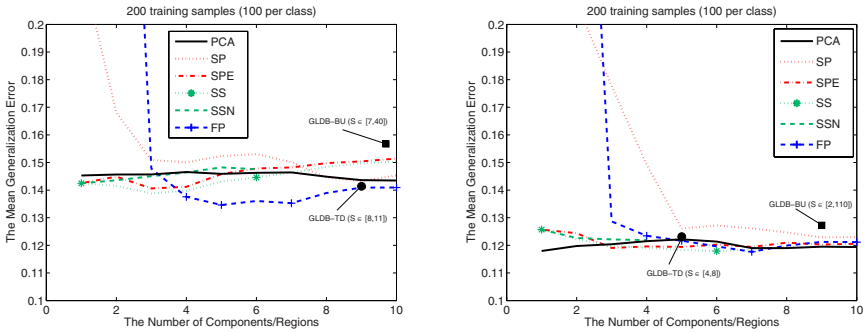
Fig. 1. Average generalization error of RLC using different methods to select discriminative spectral regions for the *Ruiz-FFT* (left plot) and the *Ruiz-PSD* (right plot) datasets

difference between SS, SSN and SPE, as they converge to the same solution. Such selected single region is shown in Fig. 2, for the FFT and PSD spectral estimations respectively. The results in both cases are approximately consistent and correspond basically to the spectral region where a significant part of energy associated to the VT events is placed, that is, roughly from 8 Hz up to 27.5 Hz. The SS method is the most flexible and powerful method from all the considered ones. It sequentially finds the most discriminative spectral bands (in combination with the previously found ones). By this, it is not surprising that it is one of the best performing spectral band selection methods. It benefits the most for small number of spectral regions. When the number of spectral regions increases, its performance degrades because each new selected spectral region depends on the sequence of the regions selected before. This does not guarantee the optimal set of regions in general. A better alternative would be to select spectral regions simultaneously, i.e. performing a full search; however, this way is computationally very expensive.

Due to space constraints, plots showing the regions selected by the SS method, for 2-4 spectral regions, had been left out of the paper. Nonetheless, it is available at the first author website (see `http://orozco.co.nr/`). The case of four selected regions is particularly interesting because a detailed insight, describing the spectral properties of the studied phenomenon, can be derived for both datasets. Even the bands are split in different ways, the three first ones indicate that most of the discriminative information is contained between 7.5 Hz and 25 Hz approximately. A narrow band around 40 Hz, associated to the peak observed in both classes, is also selected.

The FP strategy depends very much on the initial partition; nonetheless, it outperforms the other techniques when spectra are splitted into five or more spectral regions. As shown in Fig. 3, the result is very close to the initial uniform partition. Due to a small floating step, the FP method converges to its local minima. By using other initial split, the method may converge in a completely
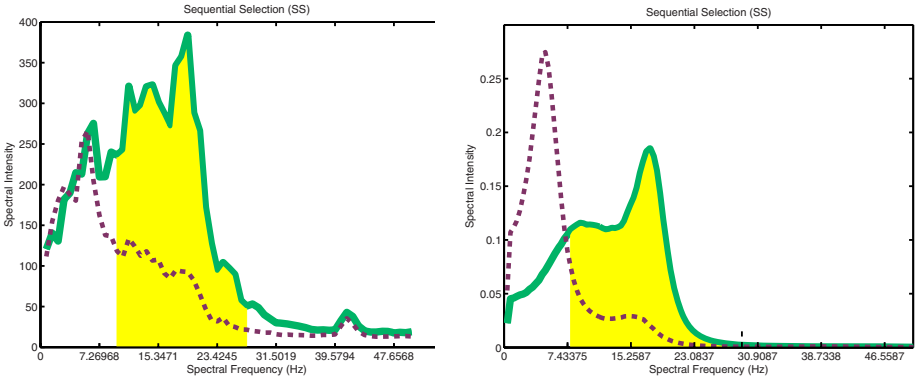
Fig. 2. A single spectral band selected by using the SS strategy for the *Ruiz-FFT* dataset (left plot) and the *Ruiz-PSD* dataset (right plot)
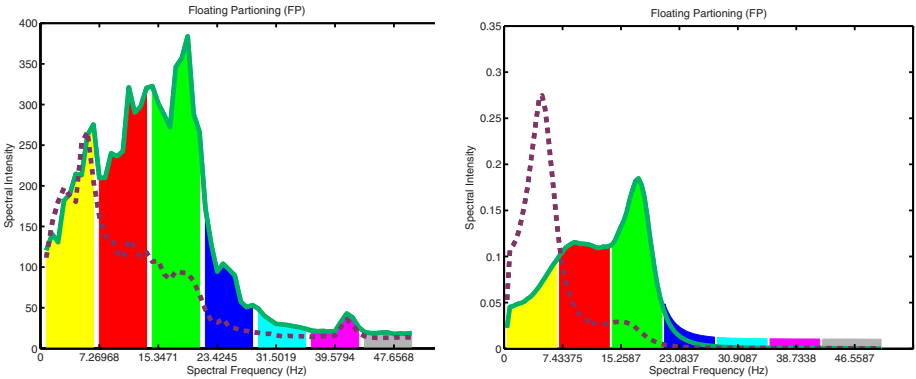

Fig. 3. Seven spectral bands selected by the FP strategy for the *Ruiz-FFT* dataset (left plot) and the *Ruiz-PSD* dataset (right plot)

different partition, in consequence it is not very useful for interpretation and dimensionality reduction purposes although it may converge by accident to a good partition. Differences between the results for both datasets show the influence of the spectral estimation method. Indeed, even though the global interpretation does not change significatively, narrow bands and average spectra are clearly different.

5   Conclusion

In this paper, we have studied a series of spectral band selection methods for characterizing the spectra of two different volcanic seismic signals –VT and LP events– recorded at Nevado del Ruiz Volcano. Such spectral band selection techniques allow us to find a small number of discriminative spectral regions as well

as to discard spectral peaks, apparently significant, which might be attributed to path effects and near surface resonance. Differences between selected bands and average spectra shapes suggest that the influence of the spectral estimation methods should be analyzed carefully. The two types of volcanic events share some dominant and discriminative regions, e.g. the narrow band near 40 Hz, suggesting a common source of volcanic process. Nonetheless, a further discussion, based on the opinion of experts such as geologists and volcanologists familiarized with the structure of the volcanic edifice, might give some insight about the physical phenomena associated to VT and LP events. For instance, magma viscosity, rock hardness/composition among others properties.

References

1. Zamora-Camacho, A., Espíndola, J.M., Reyes-Dávila, G.: The 1997–1998 activity of volcán de Colima, Western Mexico: Some aspects of the associated seismic activity. Pure Appl. Geophys. 164, 39–52 (2007)
2. Chouet, B.A.: Longperiod volcano seismicity: its source and use in eruption forecasting. Nature 380, 309–316 (1996)
3. Zobin, V.: Introduction to Volcanic Seismology. Elsevier, Amsterdam (2003)
4. Jousset, P., Neuberg, J., Jolly, A.: Modelling low-frequency volcanic earthquakes in a viscoelastic medium with topography. Geophys. J. Int. 159, 776–802 (2004)
5. Skurichina, M., Verzakov, S., Paclík, P., Duin, R.P.W.: Effectiveness of spectral band selection/extraction techniques for spectral data. In: Yeung, D.-Y., Kwok, J.T., Fred, A., Roli, F., de Ridder, D. (eds.) SSPR 2006 and SPR 2006. LNCS, vol. 4109, pp. 541–550. Springer, Heidelberg (2006)
6. Orozco-Alzate, M., García-Ocampo, M.E., Duin, R.P.W., Castellanos-Domínguez, C.G.: Dissimilaritybased classification of seismic volcanic signals at Nevado del Ruiz volcano. ESRJ 10, 57–65 (2006)
7. Kumar, S., Ghosh, J., Crawford, M.M.: Bestbases feature extraction algorithms for classification of hyperspectral data. IEEE Trans. Geosci. Remote Sensing 39, 1368–1379 (2001)
8. Skurichina, M., Paclík, P., Duin, R.P.W., de Veld, D., Sterenborg, H.J.C.M., Witjes, M.J.H., Roodenburg, J.L.N.: Selection/extraction of spectral regions for autofluorescence spectra measured in the oral cavity. In: Fred, A., Caelli, T.M., Duin, R.P.W., Campilho, A.C., de Ridder, D. (eds.) SSPR&SPR 2004. LNCS, vol. 3138, pp. 1096–1104. Springer, Heidelberg (2004)
9. Meloni, S.: Finding discriminative bands in autofluorescence spectra for automatic cancer diagnosis. Master's thesis, Cagliari University, Sardinia, Italy (2004)
10. Friedman, J.H.: Regularized discriminant analysis. J. Amer. Statist. Assoc. 84, 165–175 (1989)