Spectral characterization of volcanic earthquakes at Nevado del Ruiz Volcano using spectral band selection/extraction techniques

M. Orozco-Alzate[1,2], M. Skurichina[2] and R.P.W. Duin[2]

UNIVERSIDAD NACIONAL DE COLOMBIA

TU Delft
Delft University of Technology

[1]Department of Informatics and Computing, Universidad Nacional de Colombia Sede Manizales

[2]Information and Communication Theory Group, Delft University of Technology, The Netherlands

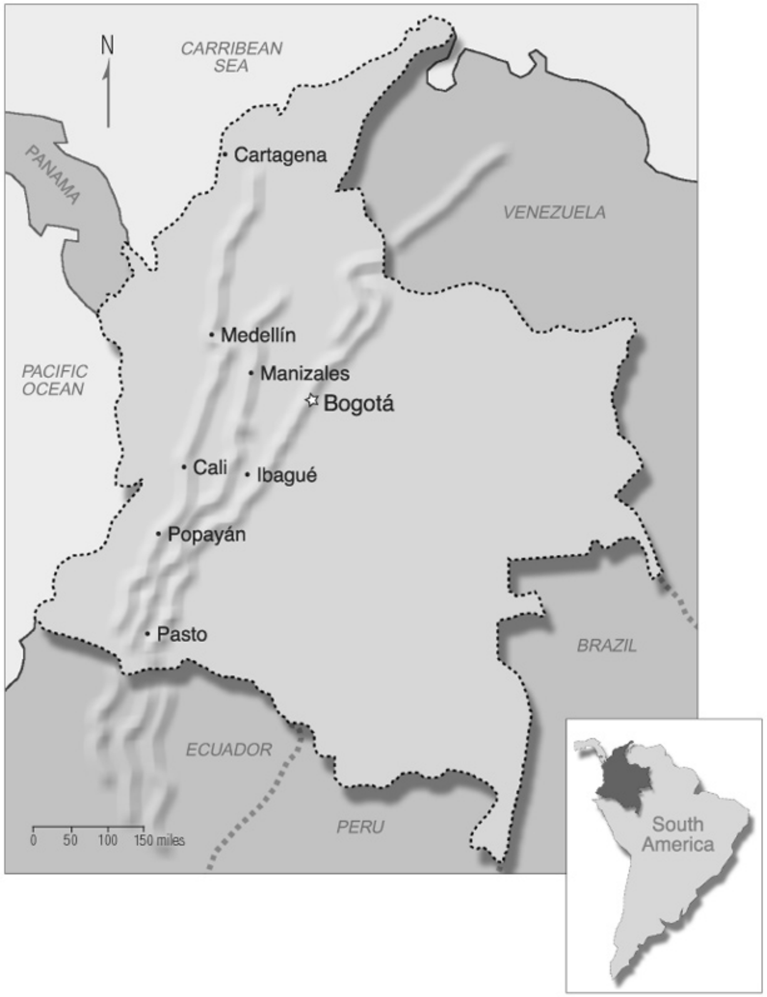
13th Iberoamerican Congress on Pattern Recognition
September 9–12, 2008
Havana, Cuba

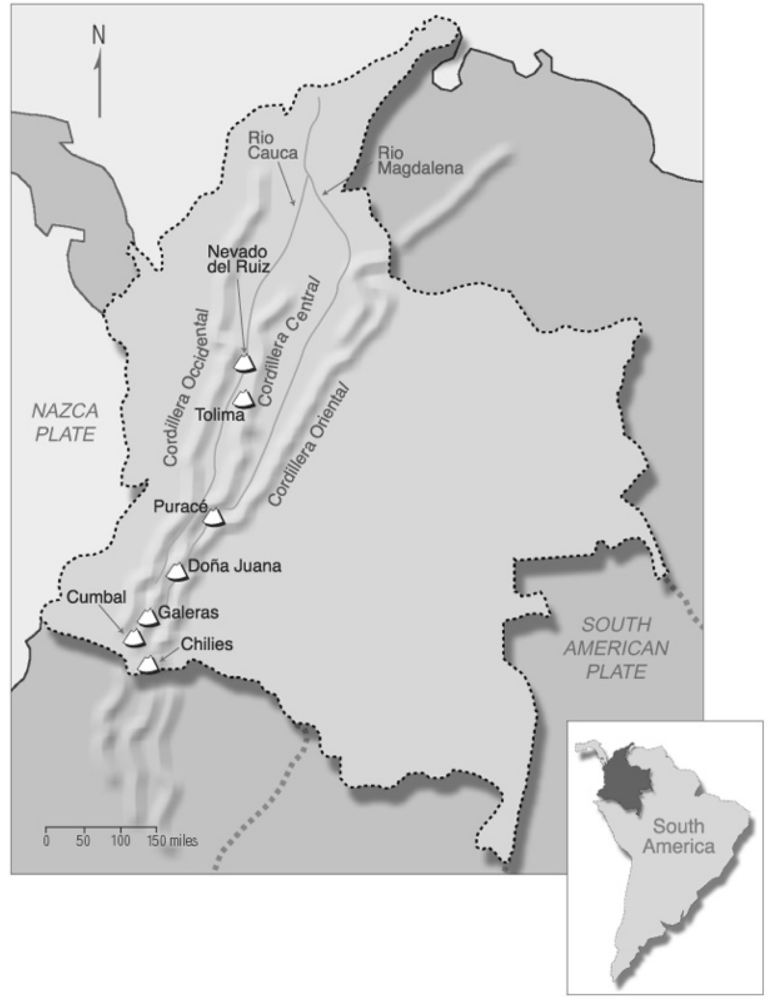
Outline

- Nevado del Ruiz volcano
- Seismic volcanic events
- Spectral representation and classification
- Data set
- Spectral representations
- Spectral band selection methods
- Experimental results and discussion
- Conclusion

Nevado del Ruiz volcano

Sept 12, 2008



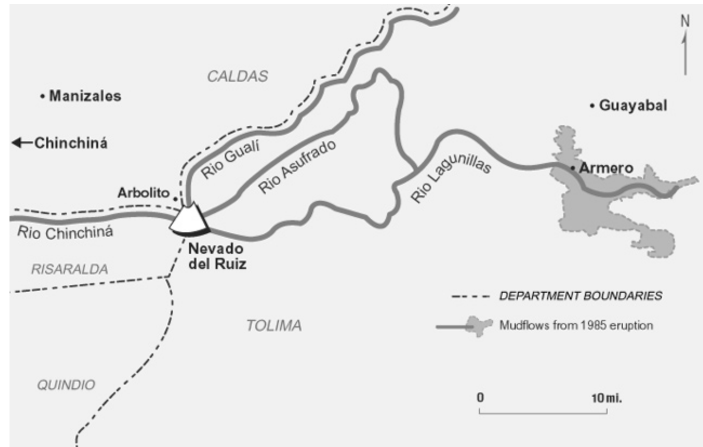


Nevado del Ruiz volcano

Elevation: 5.321 m
Eruptions: 1595, 1845, **1985**, 1991
Glacier: volume of about 1200 1500 million cubic meters. Less now due to Global Warming

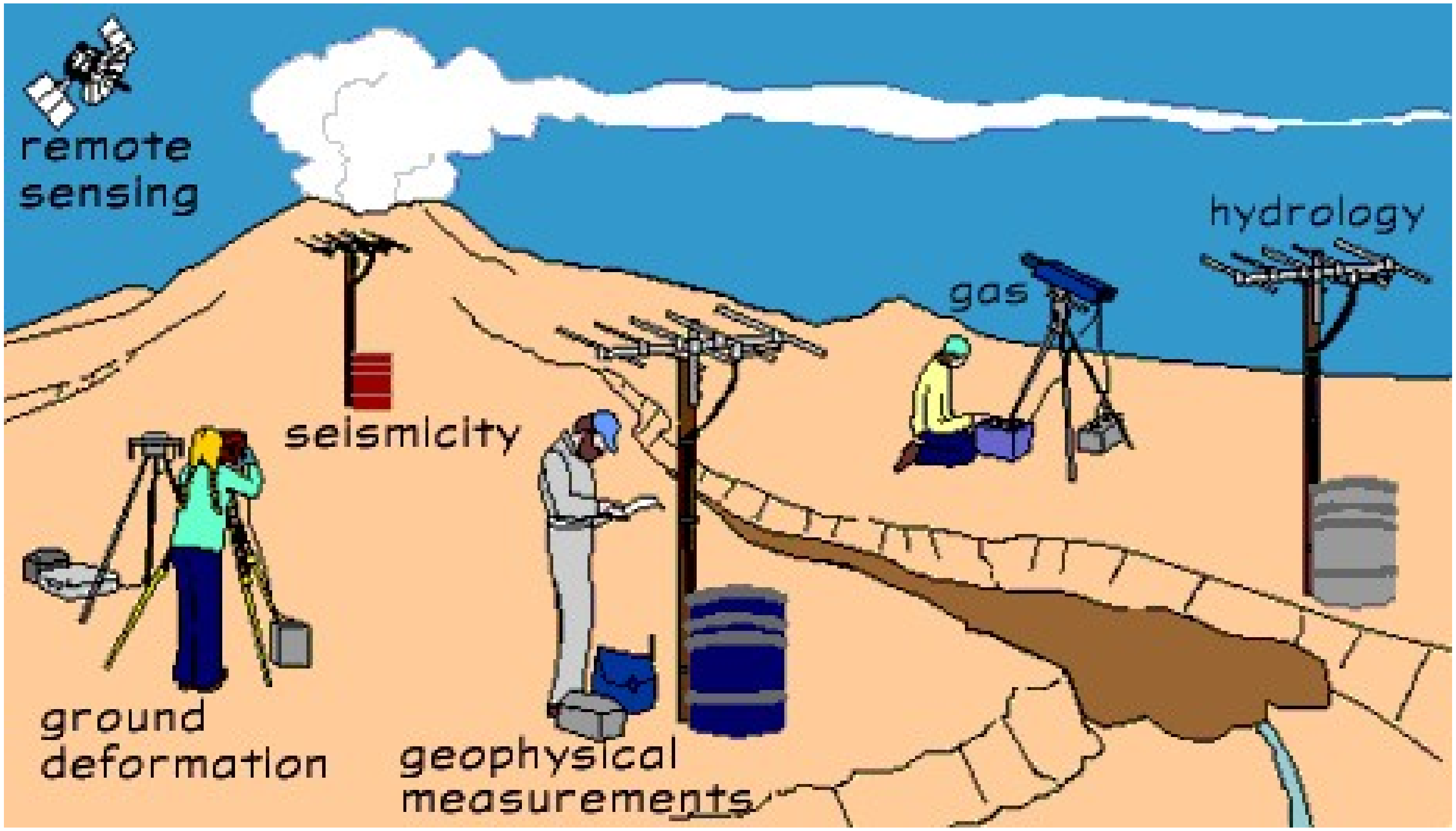
Deadly eruption: 1985

Although only moderately explosive, 1985 eruption produced mudflows which reached the town of Armero and killed more than 23,000 people.




85 12 18

Volcano-monitoring techniques



© B. Myers, http://volcanoes.usgs.gov/

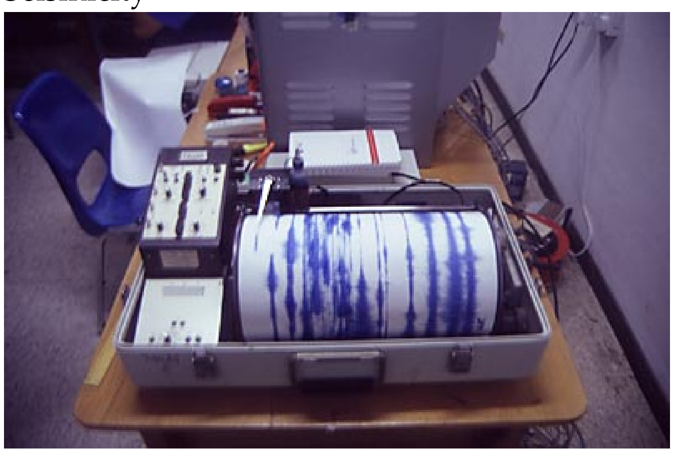
Monitoring Nevado del Ruiz volcano

Due to the tremendous disaster of 1985, the government created the volcanological observatories; before that, Colombian volcanoes were not daily and sufficiently monitored.

Ground deformation

Radon gas

Seismicity



Sulfur emissions

Chemical samples

John Makario Londoño-Bonilla
Observatorio Vulcanológico y Sismológico de Manizales (OVSM)

Pattern Recognizers in the field

Left: Robert (Bob) P.W. Duin
Delft University of Technology. The Netherlands

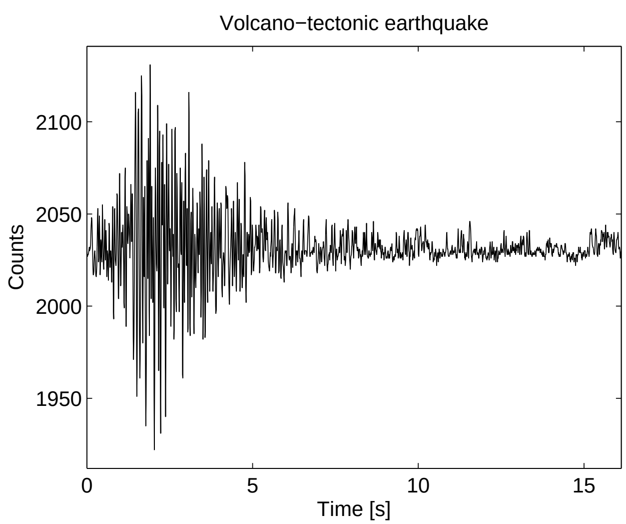
Right: Mauricio Orozco-Alzate
Universidad Nacional de Colombia Sede Manizales

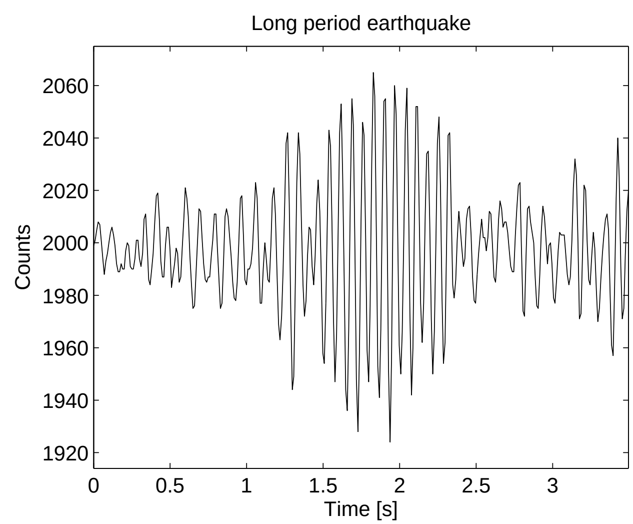
Seismic volcanic events

They correspond to manifestations of two types of physical processes: transport of fluids and fracture of solid rock.

Long-period (LP) events : essential for eruption forecasting

Volcano-tectonic (VT) : observed as a first sign of renewed volcanic activity as well as during the active period itself.





Analysis is essential to achieve an understanding of eruptive processes in andesitic volcanoes.

Spectral representation and classification

General problem: 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

Differences in spectral content allow a discrimination of different types of volcanic earthquakes

$$\Downarrow$$

- Spectral-based classification is a natural approach to face the problem.
- Some physical insight into magma properties may be derived from a spectral study.

Spectral representation and classification

The always increasing capacity of sensory systems and data storage provides high quality and densely sampled spectral measurements.

Particular problem: 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.

Fact: Spectral information is largely redundant by nature, i.e. its intrinsic dimensionality is low.

Solution: Reducing data redundancy and finding an understandable and reduced number of discriminative regions by using spectral band selection/extraction techniques.

Data

Signals were selected from data collected by the OVSM monitoring network deployed on the Nevado del Ruiz volcano

Composition: 483 VT events and 580 LP events.

Monitoring network: composed by 15 19 stations (seismometers) deployed in the volcanic complex.

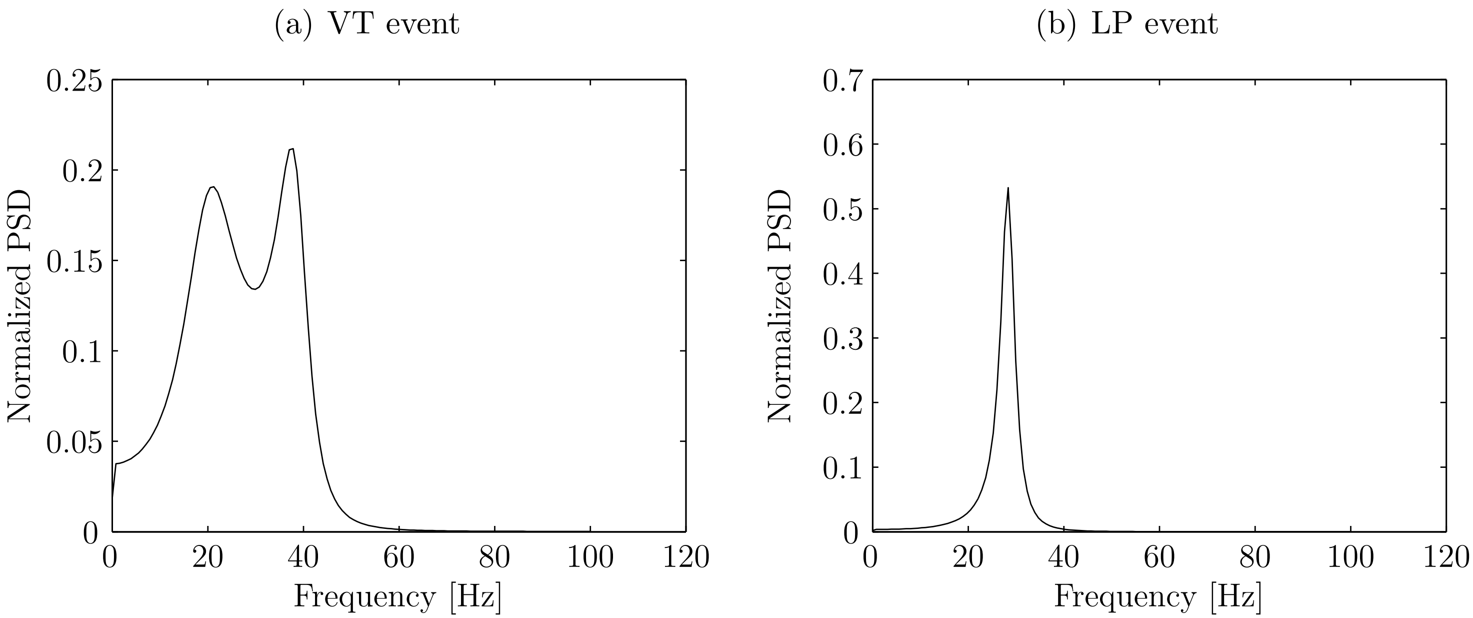
Selected station: *Olleta crater* station. It is considered a reference for the volcanic-related events. Besides, by considering signals recorded at the same location, we reduce the influence of the so-called path effect.

Location: 4.08 km from the active crater.
Digitization: 100.16 Hz sampling rate, 12 bits ADC.

Spectral representations

Fast Fourier Transform (FFT): N-point data-based spectral estimation.

Power Spectral Density (PSD): model-based spectral estimation. Particularly, a Yule-Walker AR method was used.



Spectral band selection methods

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

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

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.

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.

Spectral band selection methods

The following spectral band extraction/selection techniques are considered in this study:

1. *GLDB-TD*: Top-down multiresolution feature extraction algorithm
2. *GLDB-BU*: Bottom-up generalized local discriminant bases algorithm
3. *SP*: Sequential partitioning
4. *SPE*: Sequential partitioning and elimination of uninformative spectral bands
5. *SS*: Sequential selection
6. *SSN*: Sequential selection of non-overlapping discriminative spectral regions
7. *FP*: Floating partition

See proceedings for a detailed description of each method.

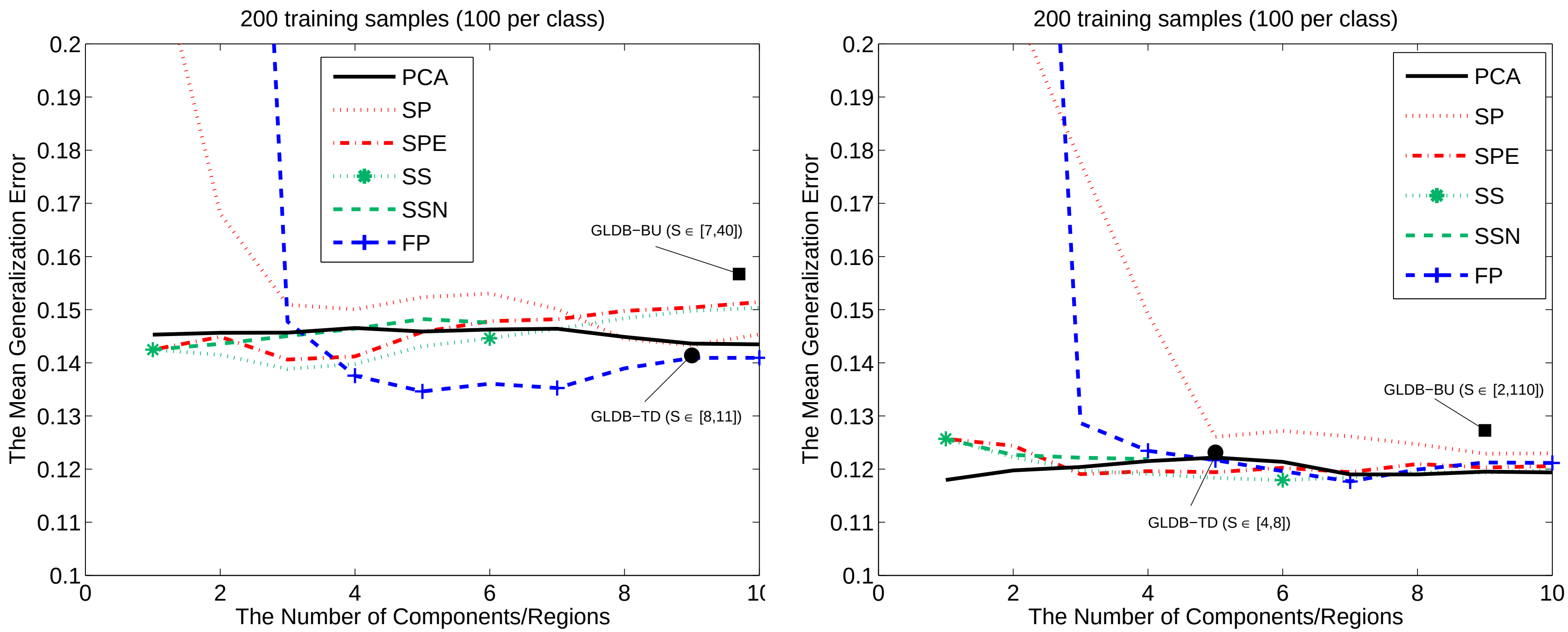
Experimental results and discussion

Experimental setup

- Training data sets with 100 objects per class, randomly chosen from the total set, for the two spectral representations: FFT and PSD.
- Remaining data used for testing.
- Prior class probabilities set to be equal as the data are unbalanced and the real prior class probabilities are unknown.
- Regularized Linear Classifier (RLC) for evaluating the performance of feature selection/extraction methods ($\lambda = 10^{-8}$)
- Experiments repeated 20 times on independent training sets.

Experimental results and discussion

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:

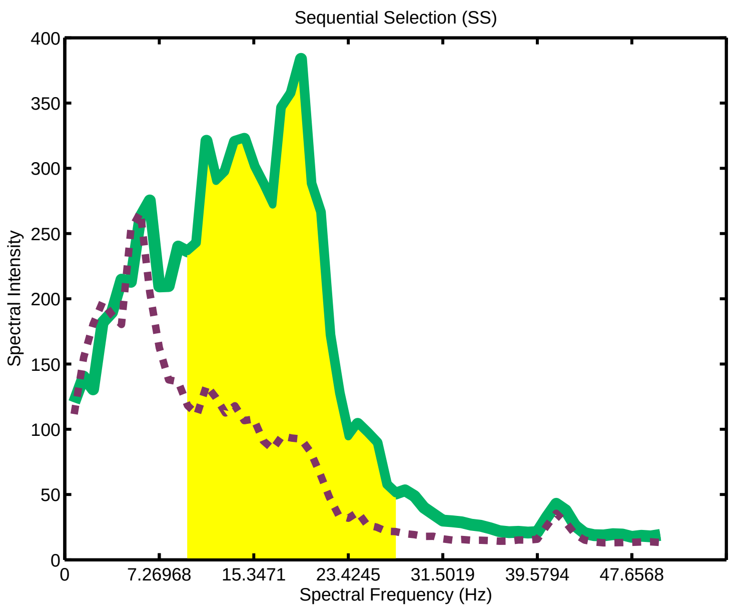


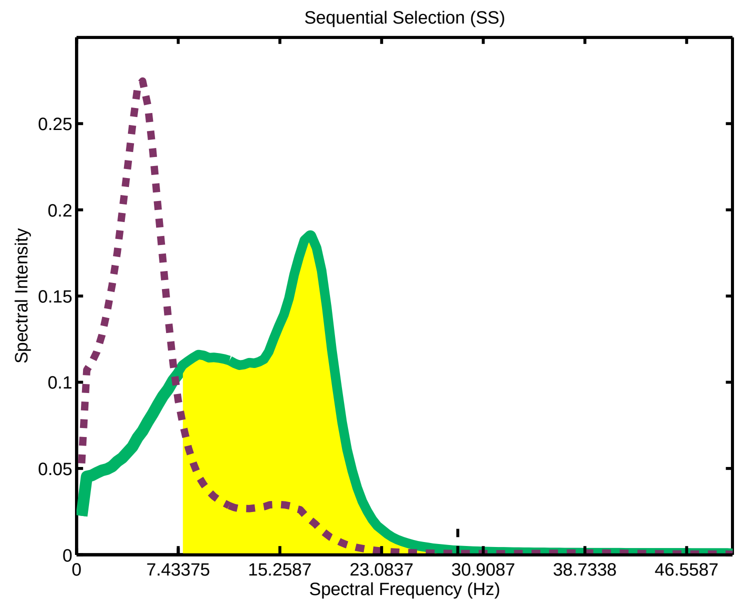
- 20 trials, maximum of 10 extracted spectral bands, $\sigma \cong 0.01$
- GLDB-TD and GLDB-BU terminate automatically according to a data-driven criterion $\rightarrow$ a single point is given in each plot

Experimental results and discussion

Selection of a single spectral region

- PCA (a single component), SS and SPE strategies are the best performing ones; in particular, PCA is the best one for the *Ruiz-PSD* dataset.
- There is no difference between SS, SSN and SPE, as they converge to the same solution.



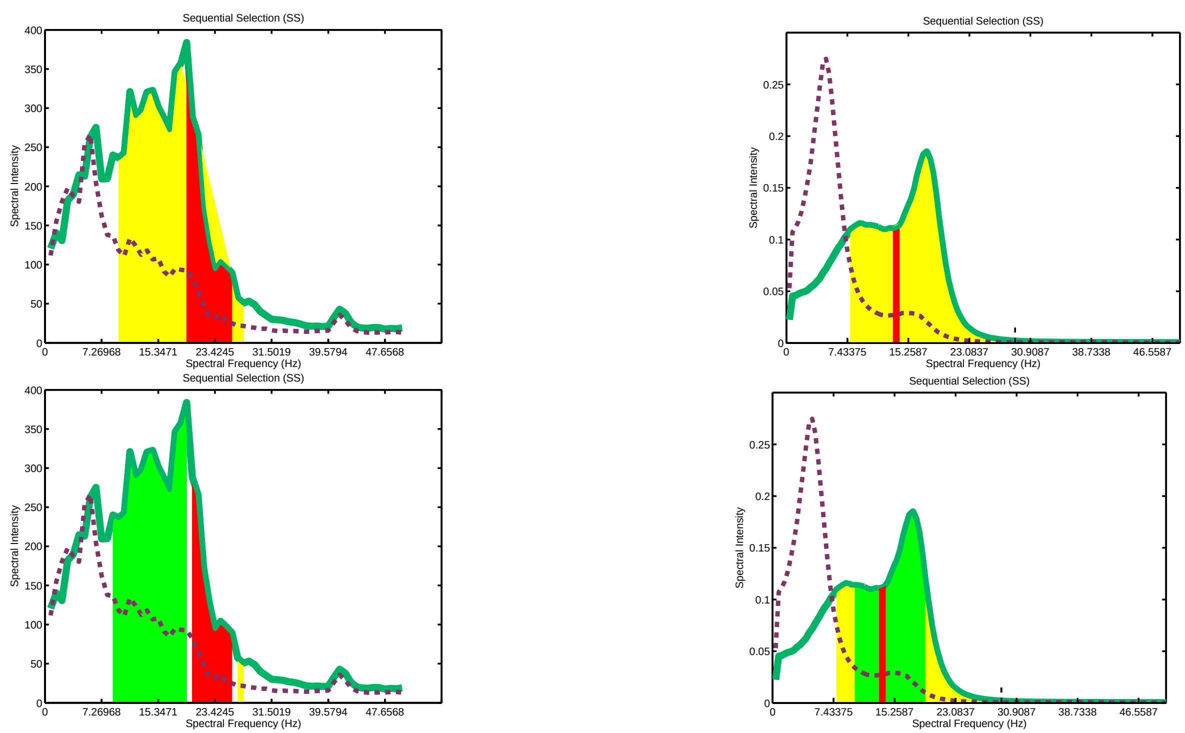


Consistent results: selected region spectral region roughly from 8 Hz up to 27.5 Hz, where a significant part of energy associated to the VT events is placed.

Experimental results and discussion

The SS strategy for 2–3 regions

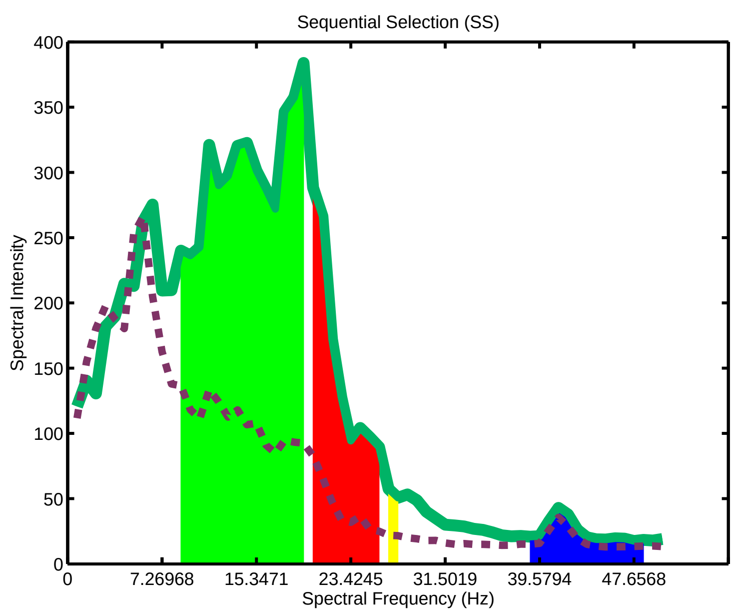
- One of the best performing spectral band selection methods. It benefits the most for small number of spectral regions
- Its performance gradually degrades because each new selected spectral region depends on the sequence of the regions selected before

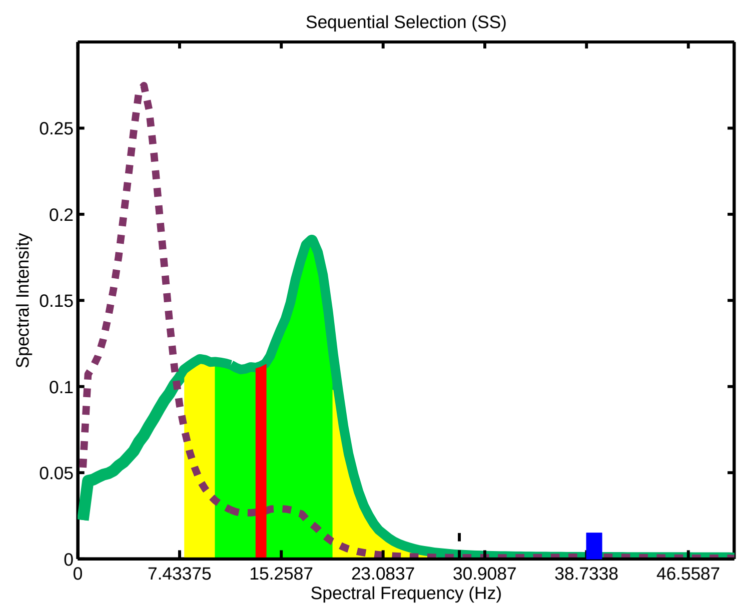


Experimental results and discussion

The SS strategy for 4 regions

- Interesting case!
- The first three bands 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.

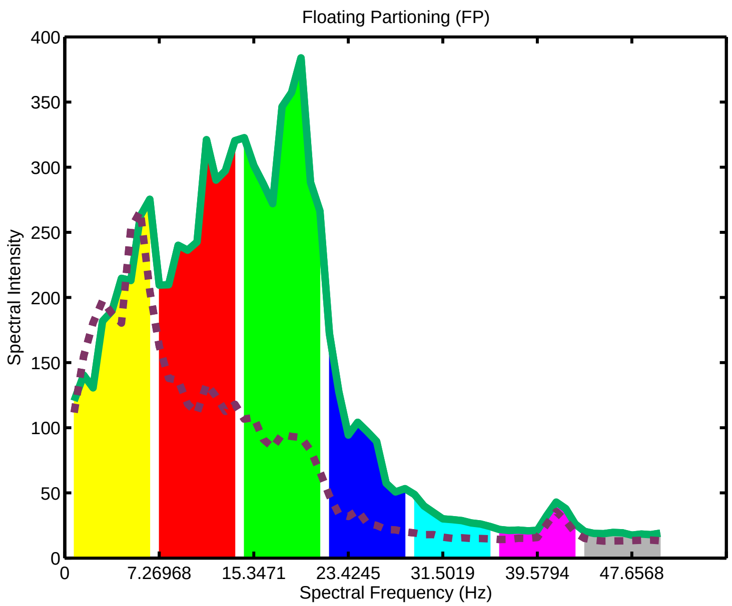


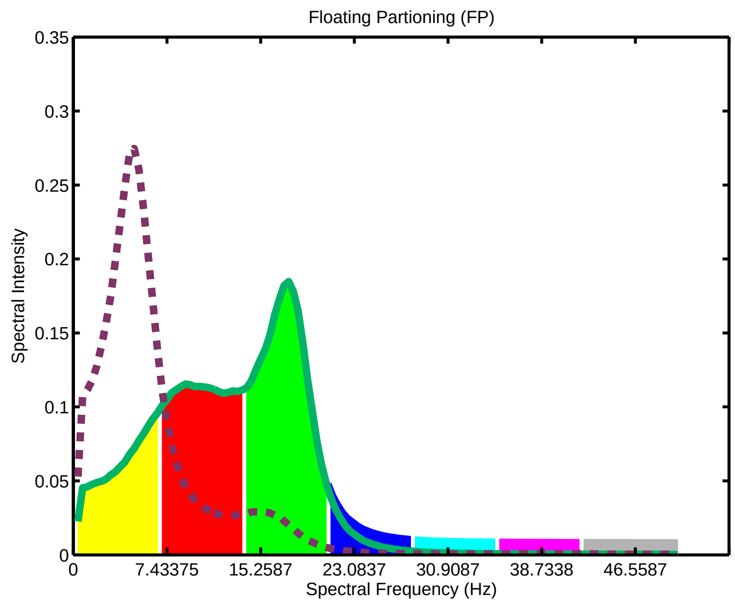


Experimental results and discussion

The FP strategy

- It depends very much on the initial partition
- It outperforms other techniques when spectra are split into five or more spectral regions
- Result is very close to the initial uniform partition.
- It is not very useful for interpretation and dimensionality reduction





Conclusion

- 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.

- Further discussion, based on the opinion of experts (geologists, volcanologists), might give some insight about the physical phenomena associated to VT and LP events. For instance, magma viscosity, rock hardness/composition among others properties.