# Classifying Three-way Seismic Volcanic Data by Dissimilarity Representation


Authors: Diana Porro Muñoz
Robert P.W. Duin
Mauricio Orozco Alzate
Isneri Talavera Bustamante
John Makario Londoño Bonilla

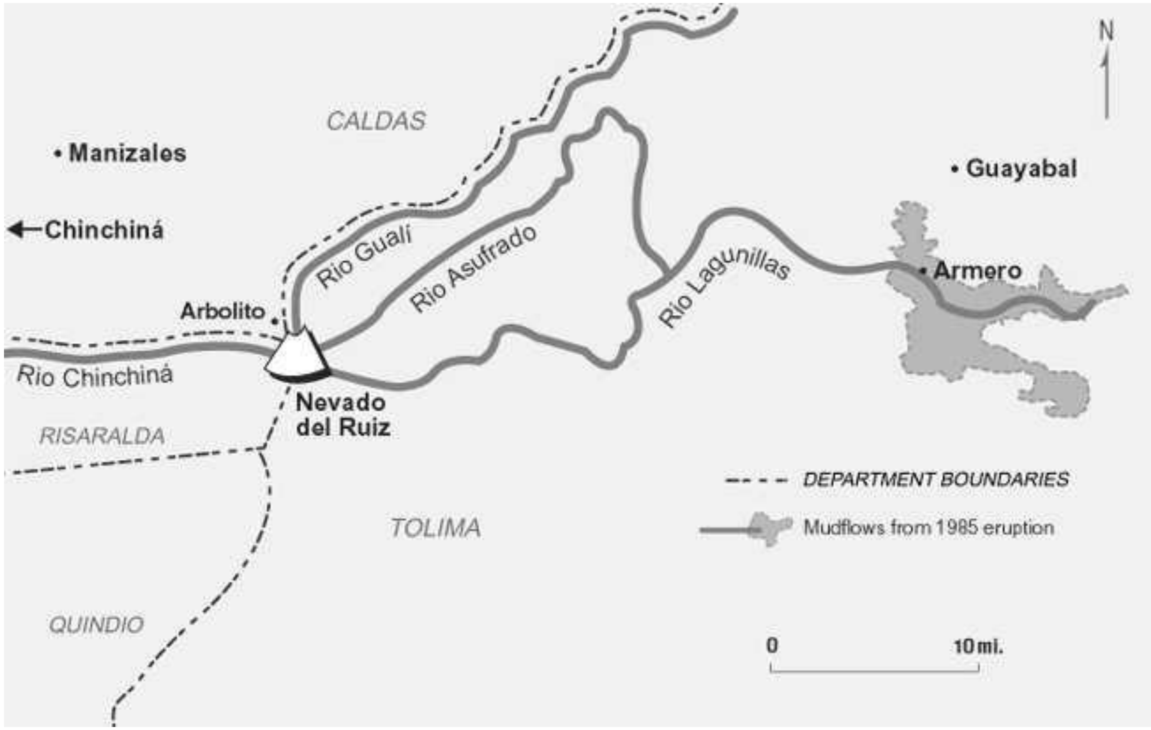


ICPR 2010, Turkey

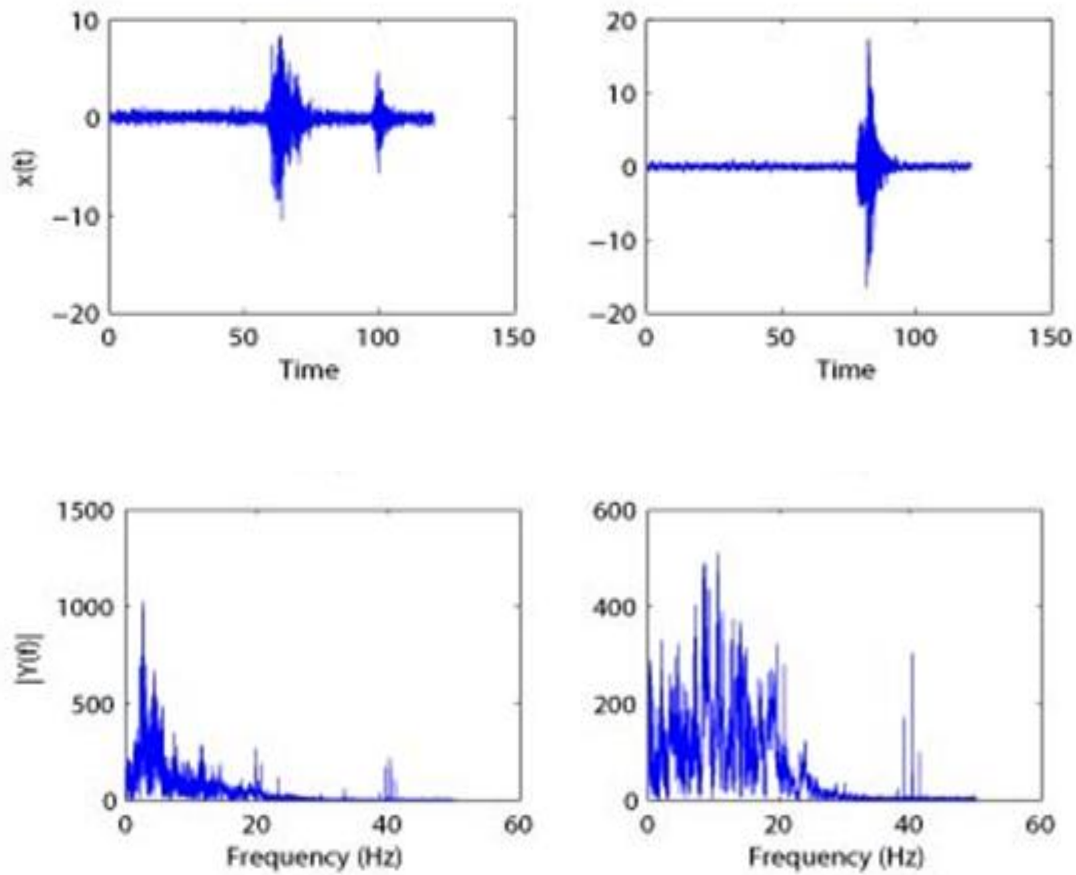
**August 2010**

# Classification of seismic volcanic data

The automatic classification of seismic volcanic signals is an essential task nowadays, with the goal of discovering the interaction between volcanic earthquakes and volcanic processes.



# Traditional Representation of seismic volcanic data

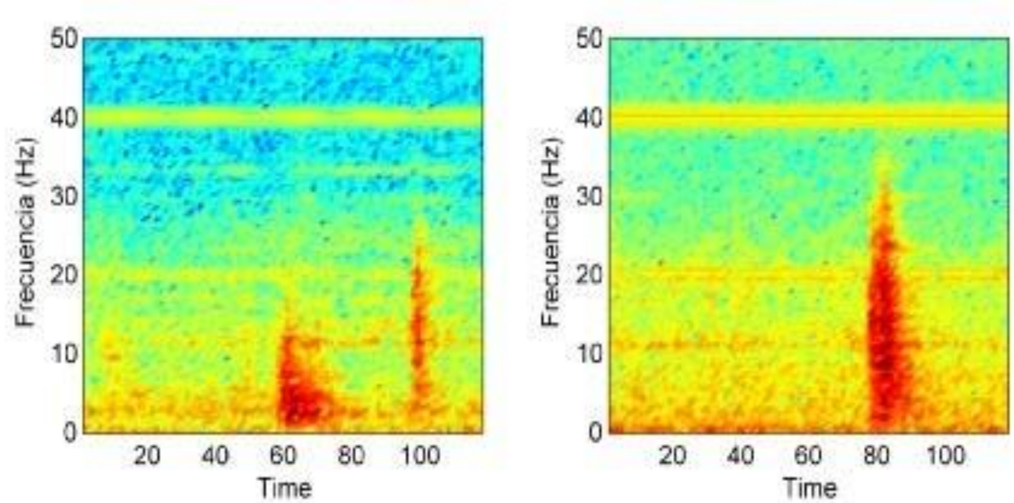



Does not take frequency changes along time into account

Does not take into consideration the time information

**Time or frequency representations alone may not be optimal for seismic signal analysis, since spectral energy changes in time**

# 2D Representation

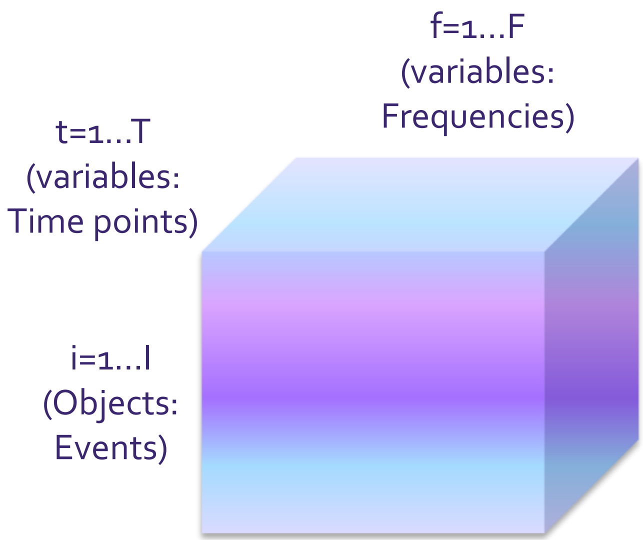



- Spectrograms: Time-frequency representation , showing frequency changes in time.

- So far, the spectrograms have just been averaged to obtain the spectral representation . The 2D object representations has not intensively been exploited as such in automatic classification systems.

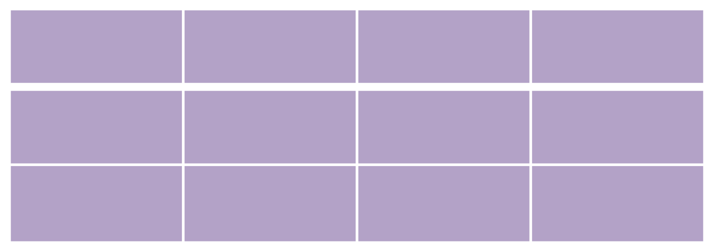
# Multi-way data analysis

- Arrays with more than two-dimensions.
- The most common is the three-way array where objects are in the first direction and the different types of measured variables in the two other.




- So, an intuitive way to represent this time-frequency relationship for all the signals would be in a three-way array

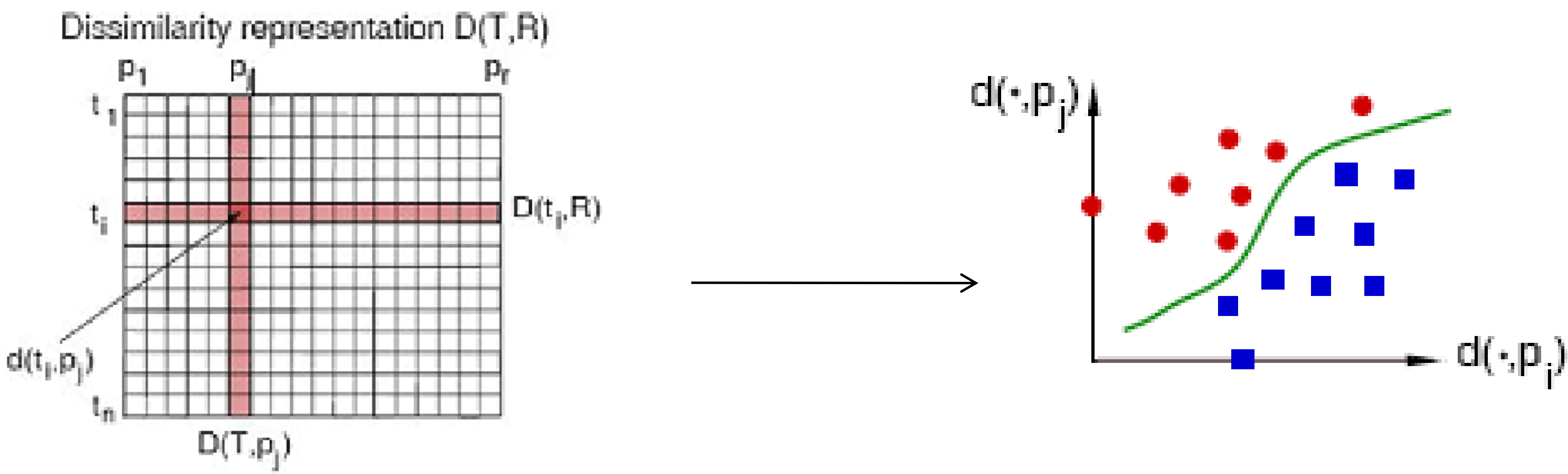
$$Y \in \mathbb{R}^{I\times F\times T}$$



$$y_i \in \mathbb{R}^{F\times T}$$

# Dissimilarity Representation (DR)

It proposes to train classifiers in the space of the proximities between objects, instead of the traditional feature.




Now, the objects are represented by a dissimilarity matrix $D\ T,R$ .

Objects are represented in this space by the column vectors of the dissimilarity matrix.

# Some advantages of DR for our problem

- ✓ High dimensionality spaces
- ✓ Any knowledge or information about the spectral background can be included into the dissimilarity measure.
- ✓ Any traditional classifier can work on the Dissimilarity Space!!!

+

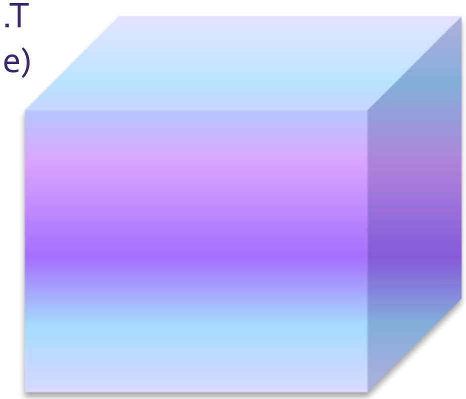
DR can be generated from many initial representations of the objects eg. numerical vectors, graphs, as long as a “suitable” measure is used.

# DR to classify three-way data?

f=1…F
(Frequency)

t=1…T
(Time)

i=1…I
(Events)



$$Y \in \mathbb{R}^{I \times F \times T}$$

$$y_i \in \mathbb{R}^{F \times T}$$

?

h=1…r
Prototypes (Representative objects)

i=1…I
(Events)

$$\begin{bmatrix} d_{11} & d_{12} & d_{13} & \cdots & d_{1r} \\ d_{21} & d_{22} & d_{23} & \cdots & d_{2r} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ d_{l1} & d_{l2} & d_{l3} & \cdots & d_{lr} \end{bmatrix}$$

# DR for multi-way data

For a t-dimensional array $Y \in \mathbb{R}^{I_1 \times I_2 \times \ldots, \times I_t}$, the theory of the DR holds.

Originally, each object is represented by a (t-1)-dimensional array, and all the objects together compose the t-dimensional array.

Consequently, to obtain the dissimilarity space, it is just done a mapping $\phi(\cdot, R) : \mathbb{R}^{I_1 \times I_2 \times I_{t-1}} \rightarrow \mathbb{R}^r$, such that for each object it is obtained

$$\phi(y_i, R) = [d(y_i, p_1), d(y_i, p_2), \ldots, d(y_i, p_r)]$$

# 2D Measure for three-way spectral data

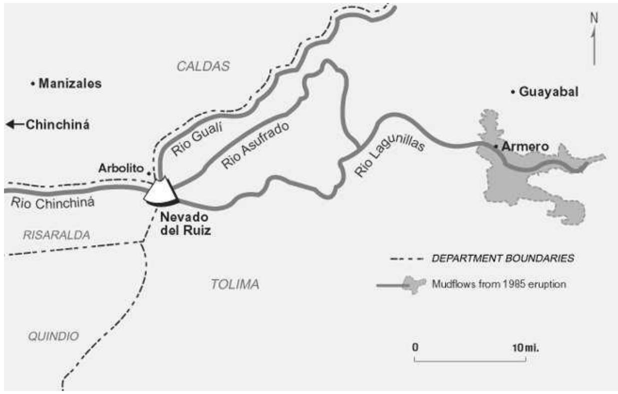
$$D^1_{a,b} = \left( \sum_{t=1}^{T} \left( \sum_{f=1}^{F} \left( \overset{\sigma}{y}_{a,f,t} - \overset{\sigma}{y}_{b,f,t} \right)^2 \right)^{p/2} \right)^{1/p}, \quad y^{\sigma}_{i,f,\cdot} = \frac{d}{d_f} G \; f,\sigma \; * y_{i,f,\cdot}$$

$$D^2_{a,b} = \left( \sum_{f=1}^{F} \left( \sum_{t=1}^{T} \left( \overset{\sigma}{y}_{a,f,t} - \overset{\sigma}{y}_{b,f,t} \right)^2 \right)^{p/2} \right)^{1/p}, \quad y^{\sigma}_{i,\cdot,t} = \frac{d}{d_t} G \; t,\sigma \; * y_{i,\cdot,t}$$

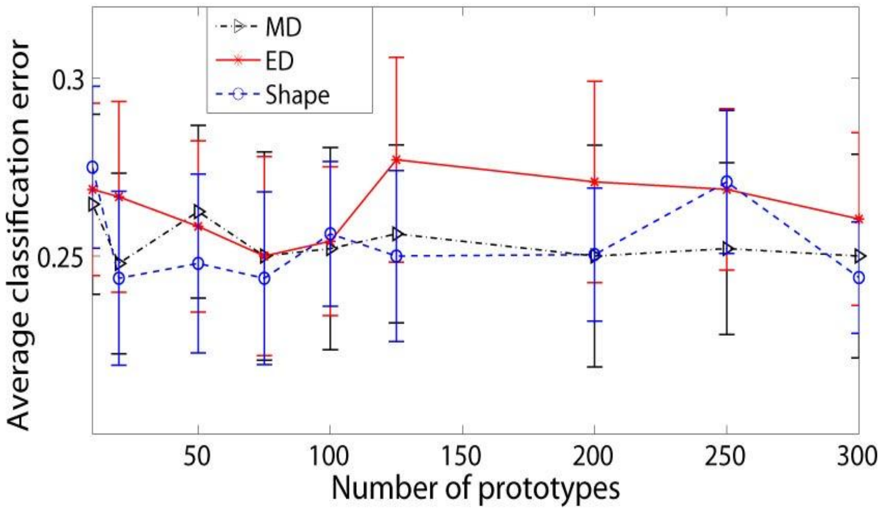
$$D = \frac{1}{w_1} D^1 + \frac{1}{w_2} D^2$$

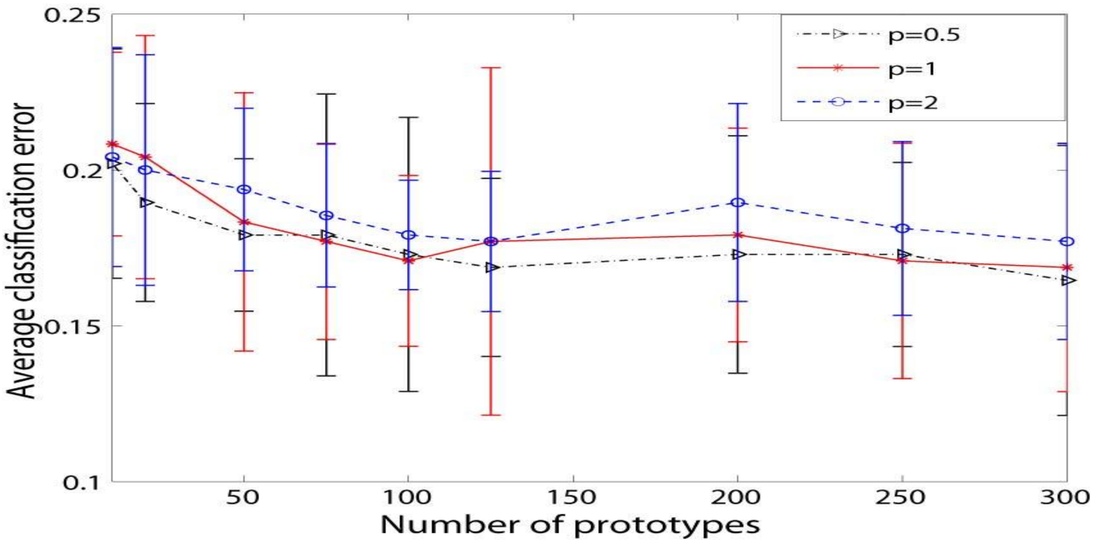
# Experiments with signals from Nevado del Ruiz Volcano

- Volcano: Nevado del Ruiz , Colombia
- Classes: Long Period, 235 events
  Volcano Tectonic earthquakes, 235 events
- Event: Time signals of 12032 points (120 sec.)
- 1D (spectral):12032-point Fast Fourier Transform (FFT).
- 2D (spectrogram): 256 short time Fourier transform
  Windows size: 256 points with 50\% of overlap.
- Data set: 470 x 129 x 93
- 2D (scalogram) : Morlet wavelet was used .
  Scale values: Major frequency components in the signals.
- Data set: 470 x 72 x 12032
- Fisher Linear classifier on DR.
- 10 times 10-fold cross-validation process.
- Prototypes were randomly chosen.



# Results . 1D (left) and 2D (right)



# Conclusions

CENATAV

- It was applied the Dissimilarity Representation as a tool for classifying three-way seismic volcanic data.
- The relationship between the different dimensions is analyzed i.e.change of frequency content in time.
- Information about the data that is missing in the actual representation e.g. shape and connectivity, can be taken into account in the proposed 2D dissimilarity measure.
- The experiments results ratify our hypothesis that the time-frequency relation is more discriminative than just the spectral information. Besides, the proposed 2D measure is capable of capturing this information.

# Conclusions

- However, a study of the influence of different overlaps and more precise techniques to obtain the time-frequency representation, could improve the three-way analysis results.

Thank you!!!!!