AGU Fall Meeting
San Francisco, USA
11 - 15 December 2006
Paper No: S13B-0228
Abs Ref. No: 2046

University of Bologna
Italy

Istituto Nazionale di
Geofisica e Vulcanologia
*Sezione di Catania*

# SUPERVISED AND UNSUPERVISED AUTOMATIC CLASSIFICATION METHODS APPLIED TO VOLCANIC TREMOR DATA AT MT ETNA, ITALY

M. Masotti, S. Falsaperla, H. Langer, S. Spampinato, R. Campanini

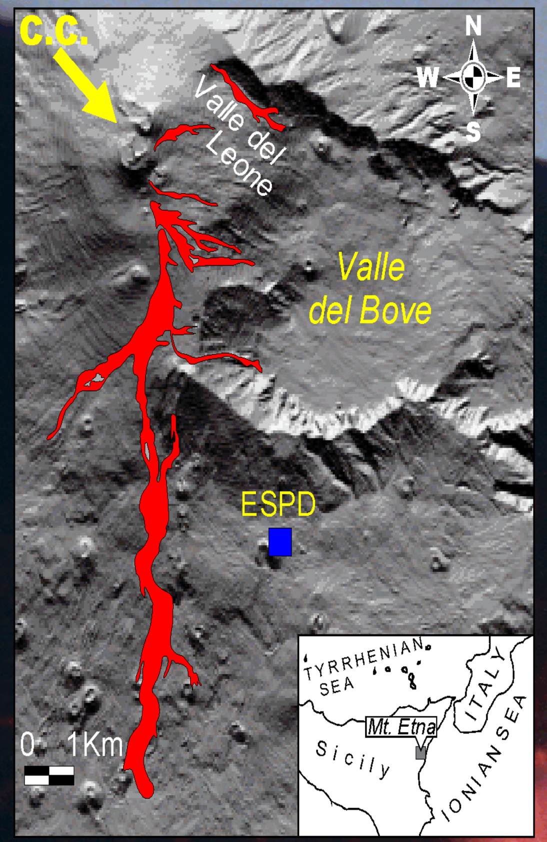
Medical Imaging Group, Department of Physics, University of Bologna, Italy
Istituto Nazionale di Geofisica e Vulcanologia (INGV) - Sezione di Catania, Italy

matteo.masotti@bo.infn.it
falsaperla@ct.ingv.it
langer@ct.ingv.it
spampinato@ct.ingv.it
renato.campanini@bo.infn.it

## Overview

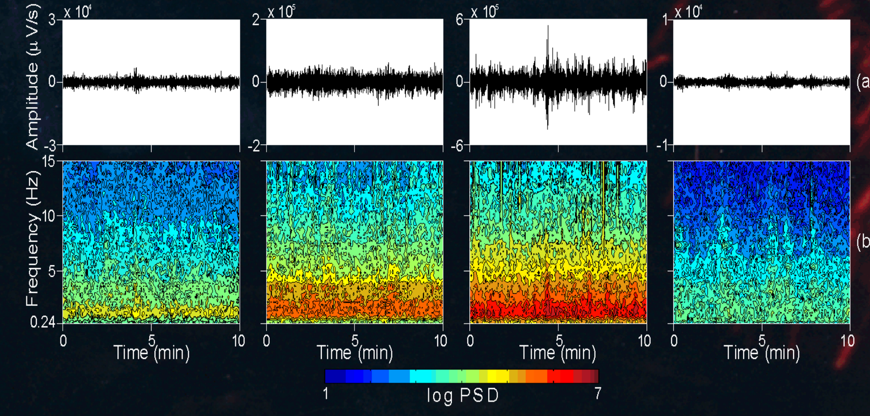
We applied automatic pattern recognition techniques to classify volcanic tremor data recorded during different states of activity at Etna volcano, Italy. The choice of automatic classification methods is determined by the necessity to solve rather complex discrimination problems using as little a-priori information as possible. Following Masotti et al. (2006), we trained a Support Vector Machine (SVM) classifier and compared its performance with those achieved by methods using supervised learning, such as Artificial Neural Networks (ANN), and unsupervised learning, such as Cluster Analysis (CA). The performance of SVM and ANN was also tested with and without the application of Genetic Algorithms (GA).

Our seismic data analysis ranged from 1 July to 15 August, 2001, and included 16 days before the onset and 7 days after the end of a flank eruption (F1). The persistent background radiation (called volcanic tremor) was recorded at a seismic station ESPD (F1) deployed 6 km southeast of the summit craters, over a time span encompassing episodes of lava fountains and a 23 day-long effusive activity.




F1 - Eruptive field at Etna in 2001. C.C. stands for Central Craters. The blue square marks the location of the seismic station ESPD.

Training and test of the classifiers were carried out using 425 spectrogram-based feature vectors. The latter were calculated from 142 time series with duration of 10 min from Z and NS components and 141 from EW. In our applications, each spectrogram of volcanic tremor represented a pattern belonging to a given class. We established four target classes, i.e., pre-eruptive, lava fountains, eruptive, and post-eruptive. Examples of the patterns are depicted in F2.



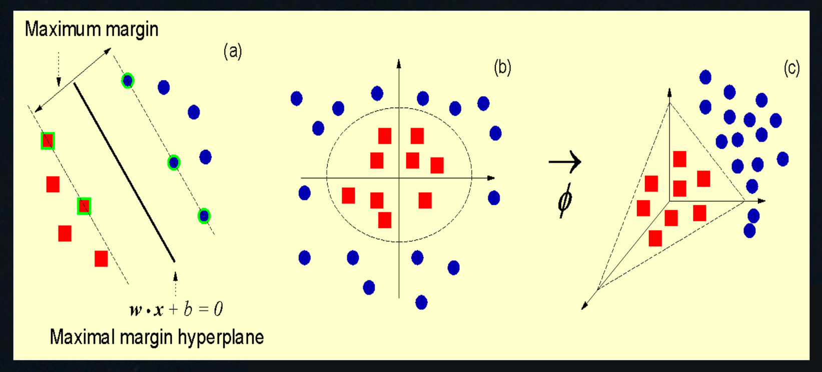

F2 - From left to right, examples of pre-eruptive, lava fountain, eruptive, and post-eruptive patterns: (a) time series, (b) spectrograms. The examples are taken from the Z component. PSD stands for power spectral density.

## Classification with supervised learning using SVM

SVM is a supervised classification method where nonlinearly separable classification problems are converted into linearly separable ones using a suitable transformation of the patterns (Vapnik, 1998). For a two-class classification problem, the decision function determined during the SVM training is the so-called maximal margin hyperplane, namely the hyperplane which causes the largest separation between itself and the border of the two classes under consideration (F3a). This border is defined by a few patterns, the so-called support vectors (F3a). As the hyperplane calculated by SVM is the farthest from the classes in the training set, it is also robust in presence of previously unseen patterns. Throughout the training, SVM computes the maximal margin hyperplane as

$$f(\boldsymbol{x}) = sgn(\boldsymbol{w}\cdot\boldsymbol{x}+b) = sgn\left(\sum_{i=1}^{l} y_i\alpha_i(\boldsymbol{x}\cdot\boldsymbol{x}_i)+b\right)$$

where the vector of weights $w$ is calculated in terms of the scalars $\alpha_i$ and $b$ by solving a quadratic programming problem (Vapnik, 1998).



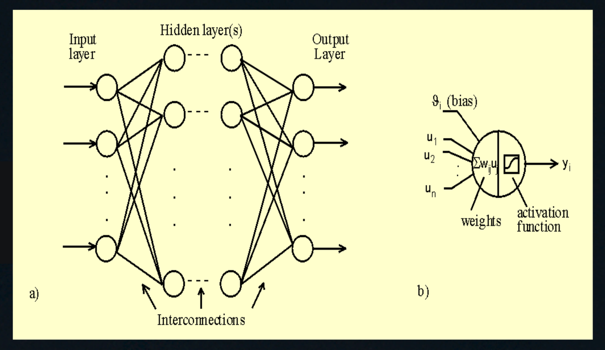

F3 - (a) Maximal margin hyperplane found by SVM; the green bordered patterns on the two margins are called support vectors, and are the only ones contributing to the determination of the hyperplane. (b and c) Transformation of a non-linear classification problem into a linear one applying the kernel function $\phi$ (from Masotti et al., 2006).

## Classification with supervised learning using ANN

A fundamental feature of ANN is that it is possible to solve classification problems of arbitrary complexity. The procedures followed for the estimation of the discrimination function do not require any a-priori knowledge about its mathematical structure. Formally the classification is carried out by applying a mapping function of the input vector **X** (which represents our signals) to an output vector **Y** (here the class membership values). The mapping function is given by

$$\hat{y}_k(\mathbf{X}) = \sum_{j=1}^{NH} c_j\sigma(\mathbf{w}_j^T\cdot\mathbf{X}+t_j)+c_0$$

where $\hat{y}_k$ is the k-th element of **Y** estimated by the network, **X** is the input vector, $\mathbf{w}_j$ are the vectors of the weights between input and hidden layer, $c_j$ are the weights between hidden and output layer, $t_j$ are biases, $\sigma(\cdot)$ is the sigmoid activation function $\sigma(z)=1/(1+e^{-z})$, $c_0$ is a constant. F4 depicts a simple architecture of ANN, the so-called Multi-layer Perceptron (Rumelhart et al., 1986).



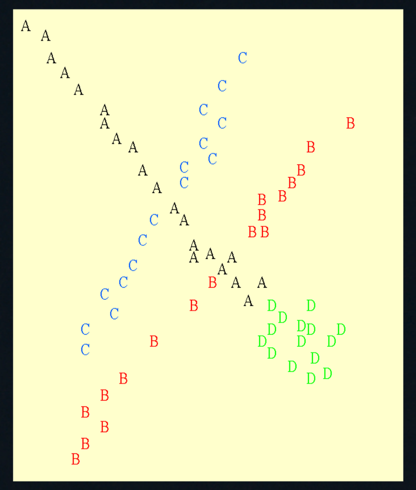

F4 - (a) Topology of an ANN, and (b) scheme of the single neuron (from Langer et al., 2006).

## Classification with unsupervised learning using CA

CA is an unsupervised classification tool, which requires only limited a-priori information about the structure of the data set. Its task is to partition the data into groups with maximum internal homogeneity. Accordingly, in our application each cluster has its typical pattern which is given by the rms amplitudes in three frequency bands between 0.24 and 15 Hz, and bandwidth of 5 Hz. As a measure of homogeneity we use the distances:

$$d^2_{ik} = (x_{ik}-x_k)^T G_k (x_{ik}-x_k)$$

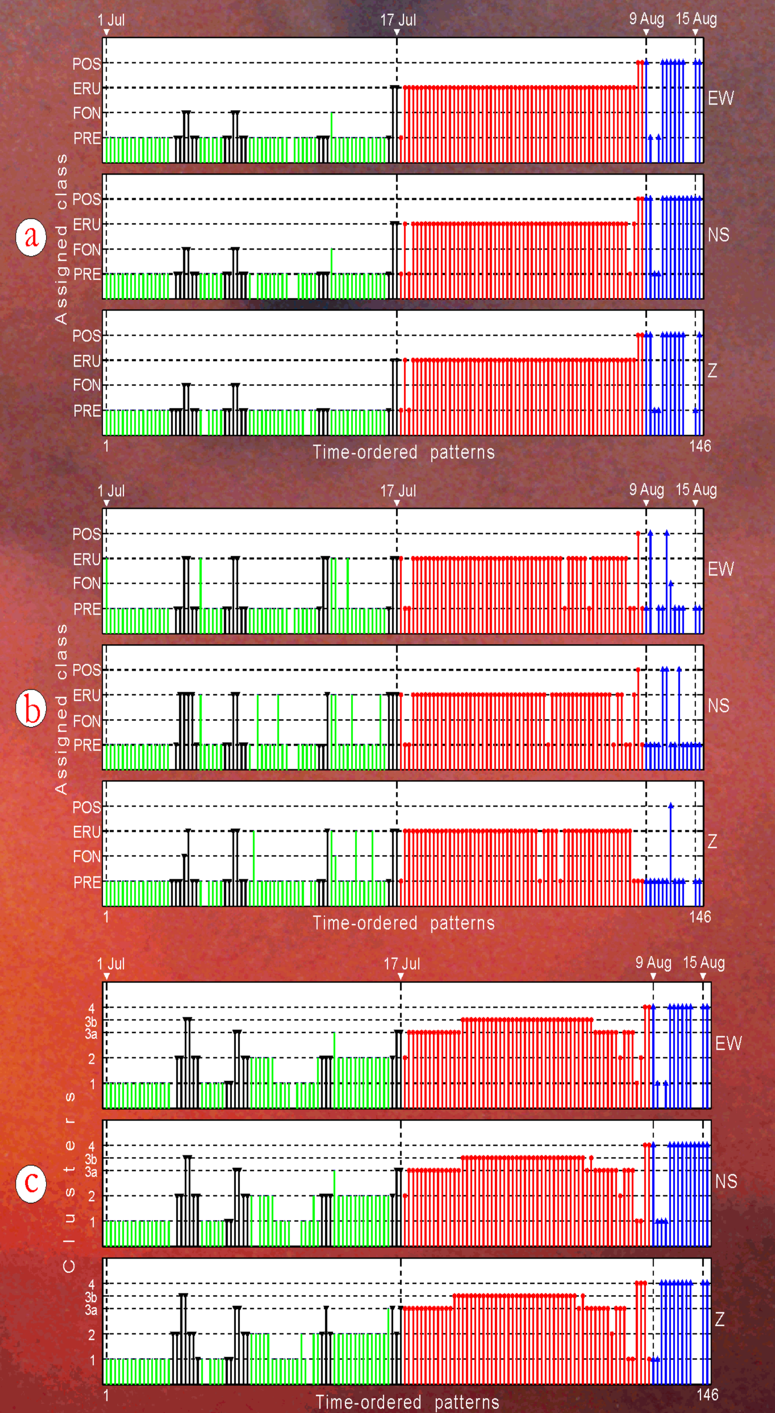
where T is the transpose and $G_k = (\det S_k)^{(1/n)} S_k^{-1}$, $S_k$ is the dispersion matrix of rank n = 3. An example of CA of a 2D demonstration data set taken from Späth (1983) is given in F5. The characters in the figure have no physical importance, as they just label elements belonging to a certain group.



F5 - CA of a 2D demonstration data set (from Späth, 1983).

## Results

Overall, our data set was composed of 425 patterns. For the supervised learning, the original data set was divided into 153 PRE, 55 FON, 180 ERU, and 37 POS. Following a leave-one-out strategy (Efron and Tibshirani, 1993), SVM and ANN were trained with the whole data set of patterns, except the one used for test. Training and test were then repeated a number of times equal to the number of patterns considered, by changing the test pattern in a round-robin manner. The results of the performance of each classifier are provided in the form of figures and tables. F6 depicts how each single pattern was classified by: SVM (F6a), ANN (F6b), and CA (F6c). The results in F6a -b were obtained after an optimization of the parameters using GA. The best performance of CA was achieved using a partition in five clusters; F6c depicts the possible matches between this partition and the a-priori classification. The tables provide the scores for each classifier summed over the three components: Table 1 (SVM), Table 2 (ANN), Table 3 (CA). Tables 1 and 2 are the confusion matrix of the leave-one-out strategy; rows and columns are to be read as a priori and assigned class membership, respectively. The number of elements found by CA for each of the five clusters 1, 2, 3a, 3b, and 4 is reported in Table 3. The five clusters are indicated as they might correspond to the classes PRE, FON, ERU, and POS. Note that the clusters 3a-b might take in the whole class ERU.




F6 - Classification results for SVM (a) and ANN (b) of the leave-one-out strategy, with scaling at the maximum value of the original feature-vectors; (c) results of CA using a partition in five clusters. The colors identify the a priori classification: PRE (green), FON (black), ERU (red), POS (blue). Patterns are time-ordered; gaps correspond to missing data. The arrows on top are time markers.

### Table 1

Confusion matrix for the SVM classifier.
Correct classifications in white in the diagonal elements.

| | PRE | FON | ERU | POST |
|---|---|---|---|---|
| PRE (153) | 151 | 2 | 0 | 0 |
| FON (55) | 37 | 12 | 6 | 0 |
| ERU (180) | 6 | 0 | 168 | 6 |
| POS (37) | 7 | 0 | 0 | 30 |

### Table 2

Confusion matrix for the ANN classifier.
Correct classifications in white in the diagonal elements.

| | PRE | FON | ERU | POST |
|---|---|---|---|---|
| PRE (153) | 136 | 1 | 16 | 0 |
| FON (55) | 30 | 1 | 24 | 0 |
| ERU (180) | 20 | 0 | 158 | 2 |
| POS (37) | 30 | 1 | 0 | 6 |

### Table 3

Number of elements belonging to each cluster found by CA.
Matches between the five clusters and the a-priori classification are indicated in white in the diagonal elements.

| | PRE | FON | ERU | POS |
|---|---|---|---|---|
| | *1* | *2* | *3a+3b* | *4* |
| *1* | 90 | 7 | 6 | 7 |
| *2* | 60 | 29 | 6 | 0 |
| *3a* | 3 | 13 | 67 | 0 |
| *3b* | 0 | 6 | 94 | 0 |
| *4* | 0 | 0 | 7 | 30 |

## Conclusions

- The results obtained with the application of GA for the optimization of the parameters used by SVM and ANN slightly improved (of a few percentage) those obtained with a trial-and-error procedure, with the drawback of much longer computation time. Nevertheless, for each method, the results with and without the optimization using GA do not change the overall performance.
- In the comparison between methods with supervised learning, SVM achieved the best performance over ANN, with both scaled and non-scaled data. Matches were on the order of 94.12% (SVM) against 70.8% (ANN).
- The scaling at the maximum value of the data forming the spectrogram-based feature-vectors does not change the results of the SVM performance obtained without scaling.
- Applying CA, we found a separation in clusters pretty close to the classes used in supervised learning. Adopting a partition in five clusters (F11), beginning and end of the eruptive stage were assigned to a cluster which corresponds to misclassifications given by SVM and ANN. This observation would confirm the existence of states of transition, which were postulated by Masotti et al. (2006).

*References*

Efron, B., and R. J. Tibshirani, (1993). *An introduction to the bootstrap*, Chapman & Hall, 1993.
Langer, H., S. Falsaperla, T. Powell, and G. Thompson, (2006). *Automatic classification and a-posteriori analysis of seismic event identification at Soufrière Hills volcano, Montserrat*. J. Volcanol. Geoth. Res., 153, 1-10, doi:10.1016/j.jvolgeores.2005.08.012.
Masotti, M., S. Falsaperla, H. Langer, S. Spampinato, and R. Campanini, (2006). *Application of Support Vector Machine to the classification of volcanic tremor at Etna, Italy*. Geophys. Res. Lett., 33, L20304, doi:10.1029/2006GL027441.
Rumelhart, D.E., G.E. Hinton, R. J. Williams, (1986). *Learning internal representation by error propagation*. In: Rumelhart, D.E., Mc Clelland, J.L (Eds), Parallel Distributed Processing: Explorations in the Microstructures of Cognitions. MIT Press, Cambridge, Mass, pp. 318-362.
Späth, H., (1983). *Cluster-Formation und Analyse, Theorie, FORTRAN-Programme*, Beispiele, Oldenbourg, München, 236 pp.
Vapnik, V., (1998). *Statistical Learning Theory*, Wiley, N.Y.

*The background photo is courtesy of Alfio Amantia (INGV-Sez. CT).*
*This work was financially supported by the projects INGV-DPC V4/02 and V4/03.*