



# TREMO<sub>r</sub>EC: A software utility for automatic classification of volcanic tremor

**M. Masotti and R. Campanini**

*Medical Imaging Group, Department of Physics, University of Bologna, Viale Berti-Pichat 6/2, I-40127, Bologna, Italy*

**L. Mazzacurati**

*Department of Computer Science, University of Bologna, Mura Anteo-Zamboni 7, I-40127, Bologna, Italy*

**S. Falsaperla, H. Langer, and S. Spampinato**

*Istituto Nazionale di Geofisica e Vulcanologia, Sezione di Catania, Piazza Roma 2, I-95123, Catania, Italy  
(falsaperla@ct.ingv.it)*

[1] We describe a stand-alone software utility named TREMO<sub>r</sub>EC, which carries out training and test of a Support Vector Machine (SVM) classifier. TREMO<sub>r</sub>EC is developed in Visual C++ and runs under Microsoft Windows operating systems. Ease of use and short time processing, along with the excellent performance of the SVM classifier, make this tool ideal for volcano monitoring. The development of TREMO<sub>r</sub>EC is motivated by the successful application of the SVM classifier to volcanic tremor data recorded at Mount Etna in 2001. In that application, spectrograms of volcanic tremor were divided according to their recording date into four classes associated with different states of activity, i.e., pre-eruptive, lava fountain, eruptive, or post-eruptive. During the training, SVM learned the a priori classification. The classifier's performance was then evaluated on test sets not considered for training. The classification results matched the actual class membership with an error of less than 6%.

**Components:** 3306 words, 5 figures, 1 animation.

**Keywords:** Volcanic tremor; Etna; Support Vector Machine; classification.

**Index Terms:** 8419 Volcanology: Volcano monitoring (7280); 0555 Computational Geophysics: Neural networks, fuzzy logic, machine learning; 7280 Seismology: Volcano seismology (8419).

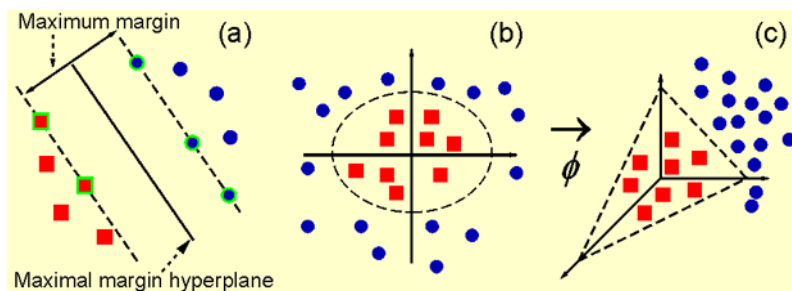
**Received** 15 October 2007; **Revised** 13 December 2007; **Accepted** 23 January 2008; **Published** 3 April 2008.

Masotti, M., R. Campanini, L. Mazzacurati, S. Falsaperla, H. Langer, and S. Spampinato (2008), TREMO<sub>r</sub>EC: A software utility for automatic classification of volcanic tremor, *Geochem. Geophys. Geosyst.*, 9, Q04007, doi:10.1029/2007GC001860.

## 1. Introduction

[2] Volcanic tremor is the background seismic radiation continuously recorded on basaltic volcanoes with persistent activity [e.g., McNutt, 2000]. Its analysis and interpretation have proven to

provide reliable information in case of volcano unrest and/or modifications of eruptive activity [e.g., Falsaperla et al., 2005]. Hence, an automatic system which can recognize different regimes of volcanic activity from tremor data might be useful for observatories as well as for purposes of Civil



**Figure 1.** Example of a two-class classification problem solved by SVM. (a) The maximal margin hyperplane is found to separate the two classes; the green bordered patterns on the two margins are the so-called support vectors. (b and c) Application of the kernel function  $\phi$  to transform a nonlinear classification problem into a linear one (from Masotti *et al.* [2006]).

Defense. To verify whether an automatic classifier could be implemented with this aim, a Support Vector Machine (SVM), a classifier based on supervised learning, was trained using volcanic tremor data recorded at Mount Etna in 2001 [Masotti *et al.*, 2006]. The application was carried out considering a time span of 46 days, during which the volcanic system changed from a relatively quiet state to episodes of lava fountains, lava emissions, and eventually returned to quietness. Assuming spectrograms of volcanic tremor as representative of these different states, the data were divided into four classes: pre-eruptive (PRE), lava fountain (FON), eruptive (ERU), and post-eruptive (POS). The results achieved showed that SVM recognized correctly the class membership for 401 out of 425 cases, i.e., 94.4% [Masotti *et al.*, 2006].

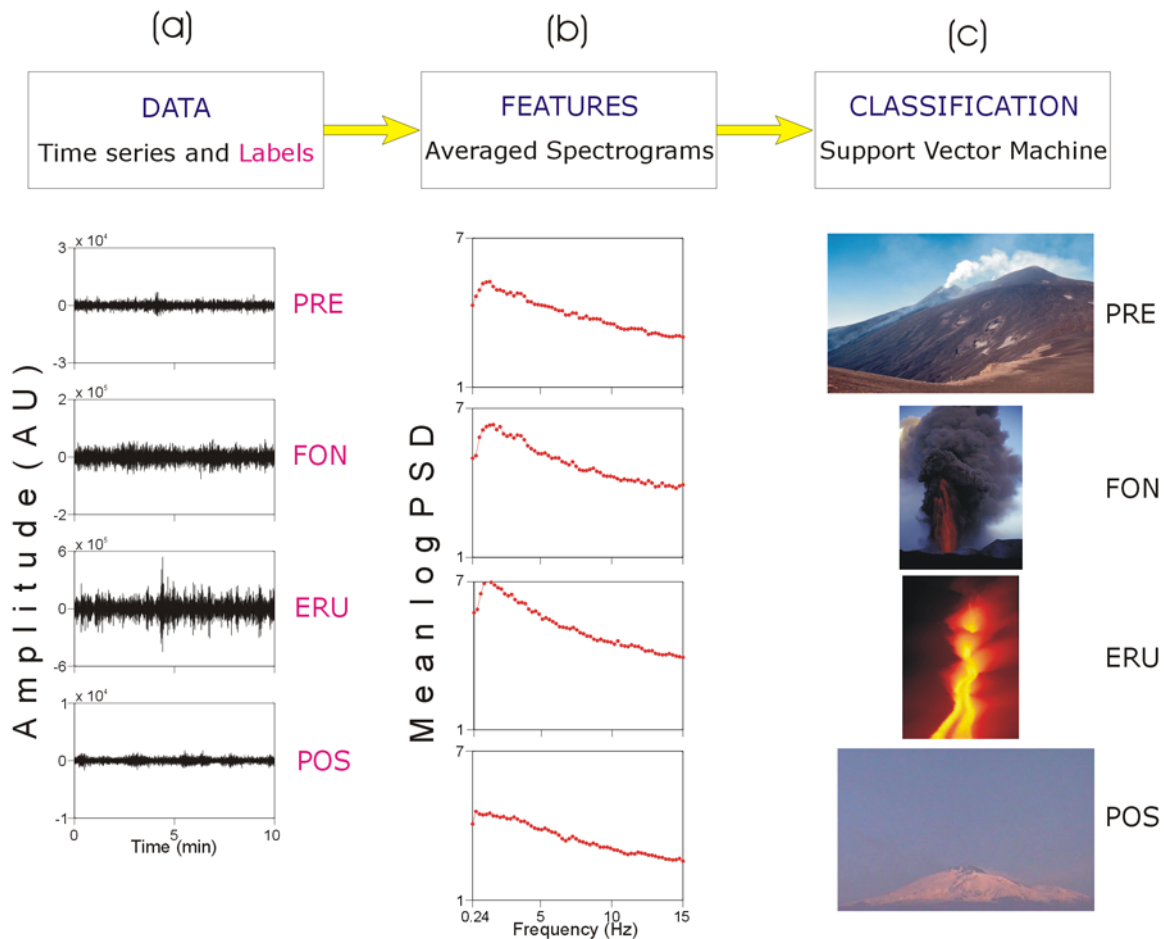
[3] Moving on from these promising results, and to exploit the excellent performance of the SVM classifier, we developed an automatic tool for the classification of volcanic tremor data. Particularly, the use of such an automatic processing accomplishes the goal of easily handling of large data sets. Besides, the acquisition of information on volcanic activity through the classification of volcanic tremor data is helpful for remote control, especially when weather conditions hinder the volcano and/or field surveys are not recommended. Here, we describe the implementation of this tool as a stand-alone software utility named TREMO<sub>r</sub>EC. To this purpose, first, we introduce the SVM classifier to the reader, then we describe how our software works, and finally we provide a sample session, which reproduces the classification results achieved by Masotti *et al.* [2006]. TREMO<sub>r</sub>EC can

be freely downloaded from <http://www.earthref.org> and <http://www.earth-prints.org>.

## 2. Support Vector Machine Classifier

[4] SVM is a supervised classification algorithm with which nonlinearly separable classification problems are converted into linearly separable ones using a suitable transformation of patterns [Vapnik, 1998]. As for the majority of supervised classification algorithms, the use of the SVM classifier encompasses two phases: (1) training on patterns with known a priori labels (i.e., selected classes) and (2) test on patterns not used for training and to which a label is assigned by the automatic classifier.

[5] For a two-class classification problem, SVM training consists in finding the so-called maximal margin hyperplane between the two classes in the  $n_D$ -dimensional feature space (Figure 1a). The maximal margin hyperplane is the linear decision function for which there is (1) the largest possible number of patterns with the same a priori label on each side and (2) maximum distance between the hyperplane and the patterns of each class. A regularization parameter  $C$  allows the trade-off between the amount of patterns erroneously classified and the maximization of the distance. As the hyperplane calculated by SVM is the farthest from the class boundaries in the training set, it is particularly robust in presence of previously unseen test patterns. As a result, the algorithm has excellent generalization capabilities. In case the patterns are not linearly separable in the feature space, a nonlinear transformation is used to map them into a higher dimensional feature space where



**Figure 2.** Main building blocks of TREMO<sub>r</sub>EC: (a) loading of the time series of the signals and a priori class memberships; (b) calculation of spectrograms and averaged spectrograms; and (c) test of the SVM classifier. In our sample session, the data recorded on Mount Etna in 2001 referred to four distinct classes: pre-eruptive (PRE), lava fountains (FON), eruptive (ERU), and post-eruptive (POS).

they are linearly separable (Figures 1b and 1c). This transformation is carried out by using the kernel functions  $K(\mathbf{x}, \mathbf{y})$  [Vapnik, 1998]. Indicating with  $\mathbf{x}$  and  $\mathbf{y}$  two generic patterns in the feature space, the dot product  $K(\mathbf{x}, \mathbf{y})$  in the mapped space can be defined as  $K(\mathbf{x}, \mathbf{y}) = \mathbf{x} \cdot \mathbf{y}$  (linear),  $K(\mathbf{x}, \mathbf{y}) = (\gamma \mathbf{x} \cdot \mathbf{y} + r)^d$  (polynomial),  $K(\mathbf{x}, \mathbf{y}) = \exp(-\gamma \|\mathbf{x} - \mathbf{y}\|^2)$  (exponential), where  $\gamma$ ,  $r$ , and  $d$  are parameters selected by the user.

[6] During the test phase, the SVM classifier assigns a label to patterns not used for training: this is done according to the position of each pattern in the feature space with respect to the maximal margin hyperplane. The performance of the classifier is hence evaluated by comparing the a priori label with that assigned by the classifier. Various strategies are reported in the literature for

test, such as the  $n$ -fold cross-validation or the leave-one-out [Efron and Tibshirani, 1993] described in the following.

### 3. TREMO<sub>r</sub>EC

[7] TREMO<sub>r</sub>EC is a software utility developed in Visual C++, which runs under Microsoft Windows 2000 or higher. It offers some preprocessing options, which allow the user to run it starting from either the raw time series, the spectrograms, or directly from the averaged spectrograms. The description in the following encompasses the whole functionality of the package, assuming time series as basic data input.

[8] The complete scheme of the basic architecture of the package is shown in Figure 2 and is detailed

in Figure S1 in the auxiliary material<sup>1</sup>. First, the time series of a signal belonging to a given class and its a priori class membership are selected (Figure 2a). Second, each time series is transformed into a spectrogram by means of the Short Time Fourier Transform (STFT) in its classical formulation [e.g., *Oppenheim and Schaffer*, 1989]. After the average value of each row of the spectrogram is calculated, the training of the SVM classifier starts using all the averaged spectrograms representative of the different classes (Figure 2b). Finally, the performance of the SVM classifier is evaluated on averaged spectrograms not previously used for training, then classes are separated (Figure 2c).

[9] The setting of TREMO<sub>r</sub>EC is handled within the three sections of the “Configuration Panel”: “Data Box,” “STFT Box,” and “SVM Box” (Figure 3). The Data Box concerns the setup of the input data. According to the basic information from which one can start (time series, spectrograms, average spectrograms), the proper value for the parameter “Data Type” has to be chosen. Any sequence of data points recorded at successive and uniform time intervals can be provided as time series. The label representative of the class membership of each time series can be any tag, e.g., integer, character, string of characters. Then, the section Data Box (Figure 3) allows (1) browsing the mass storage disks in search of the folders where the time series are saved, (2) loading automatically all the time series in the selected folders as distinct patterns, and (3) typing two or more arbitrary a priori labels (e.g., “Class1”, “Class2”, . . . , “ClassN”), which will be assigned to the time series according to the folder from which they are loaded.

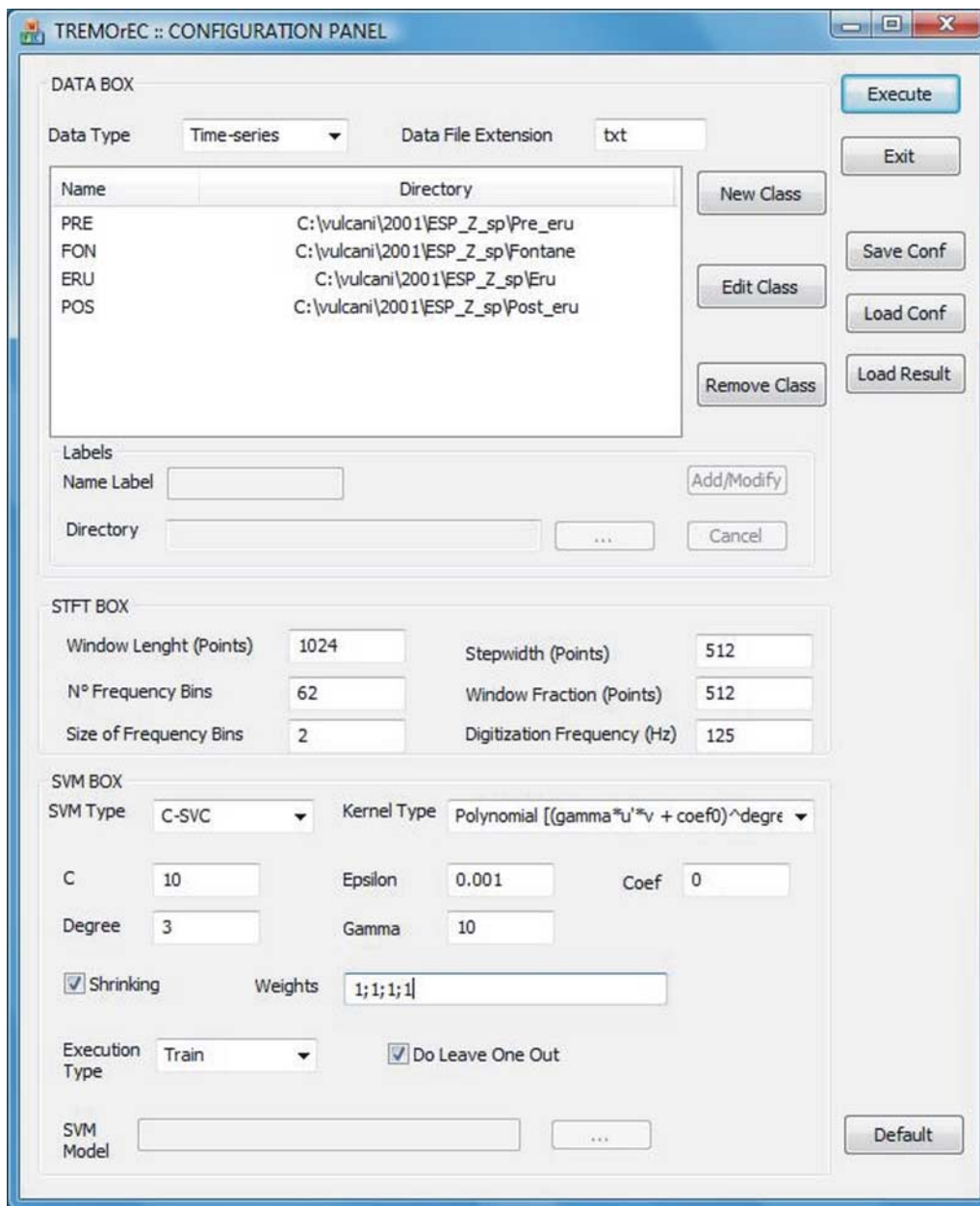
[10] Spectrograms are then calculated as consecutive spectra using STFT, with logarithmic scale of the amplitude to avoid underrating the weight of small spectral amplitudes. The STFT Box of the Configuration Panel (Figure 3) allows the setup of several parameters, such as “Window Length” (integer, power of 2) for the length in points of the time window used to ravel each time series, “Stepwidth” (integer) for the number of points to overlap in two adjacent windowed segments, “Window Fraction” (integer) for the number of sampling points used to calculate the Fast Fourier Transform of each windowed segment, “Digitiza-

tion Frequency” (real) for the sampling frequency in Hertz of the time series, “Number of Frequency Bins” (integer) for the number of frequency intervals composing the spectrogram, and “Size of Frequency Bins” (integer) for the width of the single frequency interval. For each time series loaded, the STFT is calculated and spectrograms with frequency-time dimensions of  $n_F \times n_T$  points are produced, where  $n_F$  is the Number of Frequency Bins chosen by the user and  $n_T$  is the number of time steps over which spectrograms are calculated. Hence, the rows of each spectrogram are averaged to obtain vectors of  $n_F$  points, namely, the classification features on which discrimination will be based. In doing so, the size of the input vector (i.e., the average spectral values) is equal for all time series, as requested by SVM [*Masotti et al.*, 2006]. In this light, the use of average spectrograms in TREMO<sub>r</sub>EC is advantageous, as their size is independent of the length of the original time series, which can be of arbitrary duration. Both spectrograms and averaged spectrograms are saved as plain ASCII files in the same folder of the time series.

[11] To apply the SVM classifier, TREMO<sub>r</sub>EC refers to LIBSVM, a free library written in C language [*Fan et al.*, 2005; C. C. Chang and C. J. Lin, LIBSVM: A library for support vector machines, 2001 (available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>)]. The SVM Box in the Configuration Panel recalls this library, providing easy and clear handling of the SVM classifier through the setup of a few parameters (Figure 3). “Kernel Type” allows choosing a specific kernel function among all the possible ones (e.g., “Linear,” “Polynomial,” “Exponential”). Once the kernel function is chosen, other parameters, such as the regularization parameter  $C$  (“C”), the polynomial bias  $r$  (“Coef”), the polynomial degree  $d$  (“Degree”), the multiplicative constant  $\gamma$  (“Gamma”), and the different weights assigned to each class (“Weights”) can be also easily tuned. Other numerical optimization parameters [*Fan et al.*, 2005; Chang and Lin, 2001], such as the tolerance of the termination criterion (“Epsilon”) and whether to use the shrinking heuristics (“Shrinking”), can be set as well; the latter is an optimization technique particularly useful for large data sets, as it speeds up the training process of SVM by ignoring patterns not specifically representative of the classification problem considered.

[12] In “Execution Type” there are two main working modalities, namely, training (“Train”)

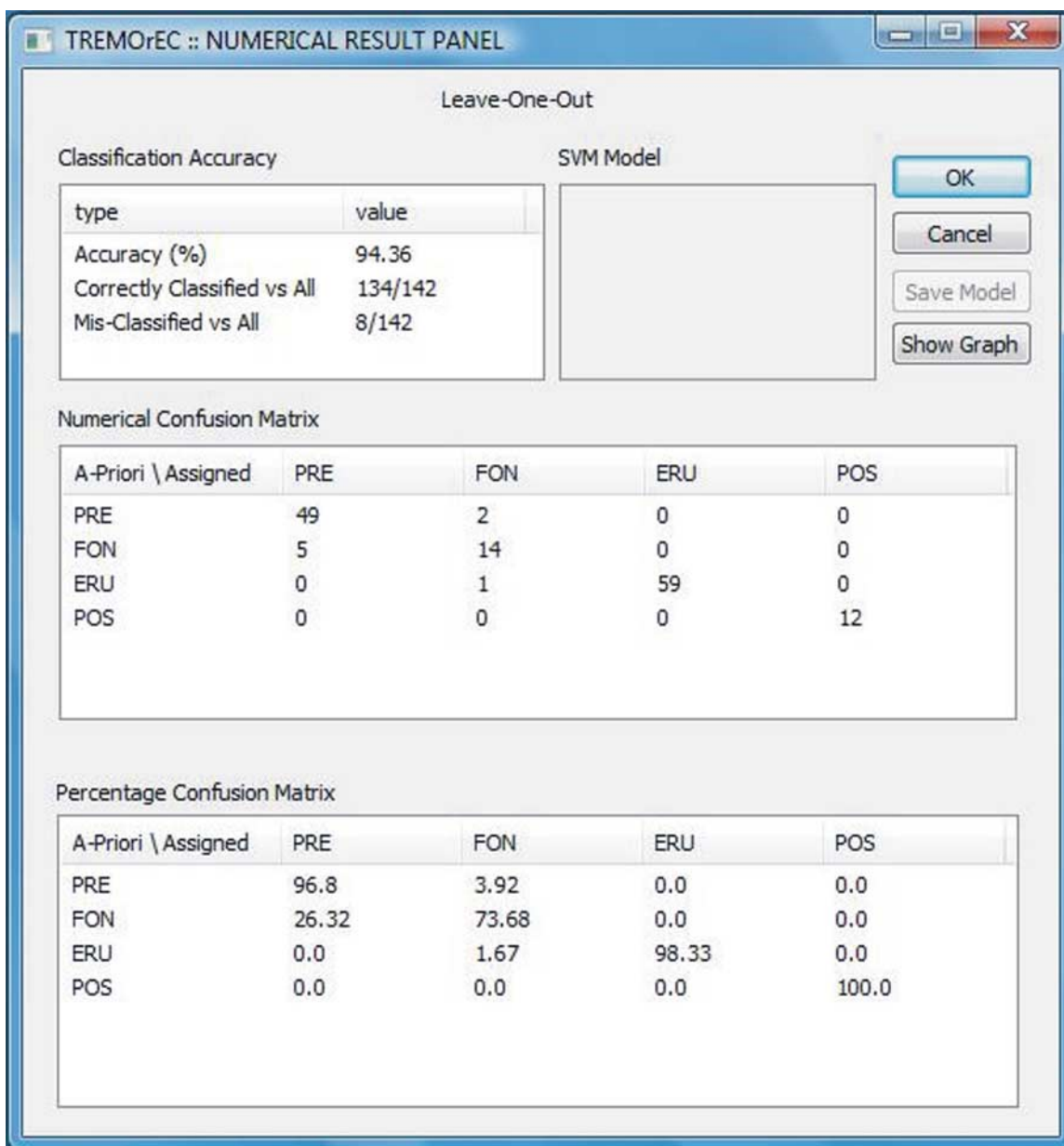
<sup>1</sup>Auxiliary materials are available in the HTML. doi:10.1029/2007GC001860.



**Figure 3.** Configuration Panel. (top) Data Box to load time series, spectrograms, or averaged spectrograms, assigning them an a priori class. (middle) STFT Box to set STFT parameters. (bottom) SVM Box to set SVM parameters and start execution specifics.

and test (“Test”) (Figure 3). If Train is chosen, then the SVM training is carried out on the data specified in the Data Box, after their processing with STFT. The resulting trained SVM classifier (“SVM Model” in Figure 3) is saved as a plain ASCII file. In case the parameter Execution Type is set up as Test, a previously trained SVM classifier can be loaded, then tested on the data specified in the Data Box after their processing with STFT. To

evaluate the performance of the classifier, a leave-one-out strategy [Efron and Tibshirani, 1993] can also be chosen setting Execution Type as Train and selecting the “Do Leave One Out” option (Figure 3). In this case, the SVM classifier is trained with the whole amount of averaged spectrograms except the one used for test, changing the test averaged spectrogram in a round-robin manner; training and test are then repeated a number of



**Figure 4.** Numerical Results Panel. (top) Accuracy, number of patterns correctly classified, and number of patterns misclassified. (middle and bottom) Numerical and percentage confusion matrices.

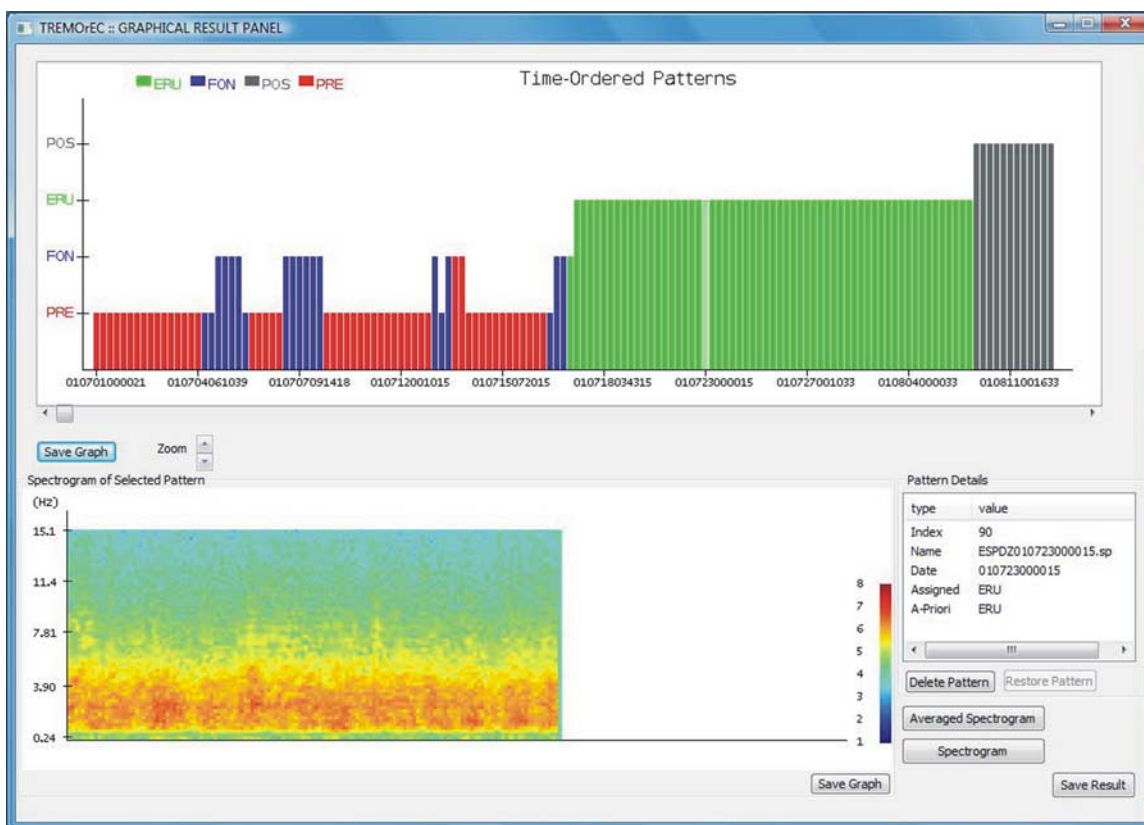
times equal to the number of averaged spectrograms considered.

#### 4. A Sample Session

[13] A tutorial video showing a sample session is provided as Movie S1<sup>2</sup>. We demonstrate the use of our software referring to the results recently achieved by *Masotti et al.* [2006], who classified

tremor data recorded at Mount Etna in 2001 using an SVM classifier. The volcanic tremor data were recorded at the three-component broadband station ESPD deployed 6 km southeast of the summit craters (see Figure 1 in *Masotti et al.* [2006]). The data covered the time span from 1 July to 15 August, 2001, and encompassed episodes of lava fountains as well as lava emissions. There were four distinct a priori labels depending upon the recording date: pre-eruptive (PRE) between 1 and 16 July, lava fountain (FON) in the pre-eruptive (4, 5, 7, 12, 13, and 16 July) and eruptive

<sup>2</sup>Movie 1 is available in the HTML.



**Figure 5.** Graphical Results Panel. (top) Classification of each pattern. (bottom) Spectrogram (or averaged spectrogram) of the selected pattern. In the top panel colors mark the a priori classification, while the assigned class is reported on the  $y$  axis.

stages (17 July), eruptive (ERU) between 17 July and 8 August, and post-eruptive (POS) between 9 and 15 August. Spectrograms with 62 frequency bins were calculated using a window of 1024 points, an overlap of 512 points, and sampling frequency of 125 Hz. Accordingly, averaged spectrograms were vectors of 62 points. Following *Masotti et al.* [2006], who chose SVM parameters after an initial trial-and-error experimentation, the classification was carried out by means of a polynomial SVM kernel with degree  $d$  set equal to 3, multiplicative constant  $\gamma$  10, bias  $r$  0, regularization parameter  $C$  10, and Weights set equal to 1 for all the classes; as far as the numerical optimization parameters, Epsilon was set equal to 0.001 and Shrinking was activated. Finally, the performance was evaluated through a leave-one-out strategy, applying the proprietary functions of the Matlab suite (The MathWorks Inc., MATLAB R2007a, 2007; available at <http://www.mathworks.com>).

[14] To check functionality and performance of TREMOREC, we consider specifically the data of the vertical component of ESPD processed by

*Masotti et al.* [2006] and start selecting time series as Data Type in the Data Box. The data shown in Figure 2 provides examples of the original time series, the corresponding label and average spectrogram for each pattern. Over the 46 days investigated, there are 142 patterns divided into 51 PRE, 19 FON, 60 ERU, 12 POS. We set STFT and SVM options using the STFT and SVM Box. The “Numerical Results Panel” (Figure 4) and the “Graphical Results Panel” (Figure 5) return the leave-one-out classification results. In particular, the former gives “Accuracy” (%) of the data correctly classified (94.36), number of “Correctly Classified Versus All” (134/142), and number of “Misclassified Versus All” (8/142) (Figure 4). The Numerical Results Panel also provides two confusion matrices in which rows and columns are read respectively as a priori and assigned classes (Figure 4). In the Graphical Results Panel the sequence of time-ordered patterns is displayed on top, with the classification of each pattern plotted as a stem graph (Figure 5). The color identifies the a priori class (red for PRE, blue for FON, green for ERU, and black for POS), whereas the assigned

class is given on the  $y$  axis. Clicking each of these stems, the corresponding spectrogram or averaged spectrogram is also displayed. This is useful to have a visual feedback about the characteristics of each pattern.

[15] The results achieved with TREMO<sub>r</sub>EC are identical to those claimed by *Masotti et al.* [2006], namely, 134 out of 142 patterns (96.07%) correctly classified, with 49 out of 51 (96.08%) of PRE, 14 out of 19 (73.68%) of FON, 59 out of 60 (98.33%) of ERU, and 12 out of 12 (100.00%) of POS.

## 5. Conclusions

[16] TREMO<sub>r</sub>EC offers ease of use and short time processing. It can directly handle ready-to-use spectrograms of volcanic tremor as well as compute them from the original time series. It carries out the training of the SVM classifier with different parameters and kernels, and allows evaluating the performance of the classifier either on an independent test set or using a leave-one-out procedure. Classification results are presented by means of a graphical user interface in the form of overall classification accuracy, confusion matrices, and with the help of graphs showing the a priori and assigned class for each considered pattern. The identification of different states of volcanic activity

using just simple spectrograms makes this tool extremely useful for monitoring volcanoes where the state of the system can be associated with typical volcanic tremor patterns.

## Acknowledgments

[17] We thank Olivier Jaquet and Manfred Joswig for their useful suggestions. This work was financially supported by Istituto Nazionale di Geofisica e Vulcanologia and Dipartimento per la Protezione Civile (projects V4/02 and V4/03).

## References

- Efron, B., and R. J. Tibshirani (1993), *An Introduction to the Bootstrap*, CRC Press, Boca Raton, Fla.
- Falsaperla, S., S. Alparone, S. D'Amico, G. Di Grazia, F. Ferrari, H. Langer, T. Sgroi, and S. Spampinato (2005), Volcanic tremor at Mt. Etna, Italy, preceding and accompanying the eruption of July–August, 2001, *Pure Appl. Geophys.*, *162*(11), 2111–2132.
- Fan, R. E., P. H. Chen, and C. J. Lin (2005), Working set selection using second order information for training SVM, *J. Mach. Learning Res.*, *6*, 1889–1918.
- 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.
- McNutt, S. R. (2000), Volcanic seismicity, in *Encyclopedia of Volcanoes*, edited by H. Sigurdsson et al., pp. 1095–1119, Academic, New York.
- Oppenheim, A. V., and R. W. Schaffer (1989), *Discrete-Time Signal Processing*, Prentice-Hall, Englewood Cliffs, N. J.
- Vapnik, V. (1998), *Statistical Learning Theory*, John Wiley, New York.