Application of Support Vector Machine to the classification of volcanic tremor at Etna, Italy

M. Masotti,[1] S. Falsaperla,[2] H. Langer,[2] S. Spampinato,[2] and R. Campanini[1]

[1] We applied an automatic pattern recognition technique, known as Support Vector Machine (SVM), to classify volcanic tremor data recorded during different states of activity at Etna volcano, Italy. The seismic signal was recorded at a station deployed 6 km southeast of the summit craters from 1 July to 15 August, 2001, a time span encompassing episodes of lava fountains and a 23 day-long effusive activity. Trained by a supervised learning algorithm, the classifier learned to recognize patterns belonging to four classes, i.e., pre-eruptive, lava fountains, eruptive, and post-eruptive. Training and test of the classifier were carried out using 425 spectrogram-based feature vectors. Following cross-validation with a random subsampling strategy, SVM correctly classified $94.7 \pm 2.4\%$ of the data. The performance was confirmed by a leave-one-out strategy, with 401 matches out of 425 patterns. Misclassifications highlighted intrinsic fuzziness of class memberships of the signals, particularly during transitional phases. **Citation:** 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.

1. Introduction

[2] Continuous seismic monitoring has become a key tool for the surveillance of active volcanoes. On basaltic volcanoes like Etna, the interpretation of the persistent background radiation (called volcanic tremor) is of particular importance as its characteristics disclose insights into magma dynamics. Yet, the continuous acquisition of signals comes with the problem of accumulating a large mass of data difficult to handle on-line as well as off-line. Consequently, the automatic processing of data is the goal of any analysis encompassing the unrelenting flow of signals. To this purpose, we considered the application of an automatic classification of volcanic tremor following a supervised classification scheme. In this scheme, a data set which is used to determine the controlling parameters of the classifier (i.e., the training set) is prepared. Then, parameters are tuned through an iterative process during which the classifier is said to learn the classification problem. Eventually, the classifier is applied to test data, i.e., patterns which were not used during the learning phase, but are supposed to belong to the same parent population of the training set. A well-known approach for such an automatic supervised classification is Artificial Neural Networks (ANN) [e.g., *Langer et al.*, 2003; *Scarpetta et al.*, 2005]. They generate arbitrarily complex mapping functions, but turn out to be sensitive to overfitting, giving excellent performance on the training set, yet being unstable on the test set.

[3] Here, we discuss the application of Support Vector Machine (SVM hereafter), an automatic classifier largely adopted by the pattern classification community and recently applied by two of us (R. Campanini and M. Masotti) in medical applications [*Bazzani et al.*, 2001; *Campanini et al.*, 2004; *Angelini et al.*, 2006]. SVM is a supervised classification method where nonlinearly separable classification problems are converted into linearly separable using a suitable transformation of the patterns. Apart from the low complexity of the resulting classification curve, an important benefit of the SVM approach over automatic classifiers based on non-linear discrimination functions is that this classification curve is the farthest from the border of the classes to separate. As a result, SVM tends to be less prone to problems of overfitting than other methods do. We outline basic characteristics of SVM in section 3, and address the interested reader to textbooks like *Duda et al.* [2000] and *Hastie et al.* [2002].

2. Tremor Data Analysis

[4] By 17 July, 2001 a volcano unrest began at Etna after five days of intense tectonic seismicity heralding the opening of the eruptive fractures. Episodes of lava fountains – with duration ranging from hours to about one day – shortly preceded and accompanied the onset of the lava effusion as well. Lava flows poured out in Valle del Leone, Valle del Bove, and middle-upper southern flank of the volcano (Figure 1). The effusive activity stopped on 9 August after the emission of $\sim 25 \times 10^6$ m^3 of lava and $5-10 \times 10^6$ m^3 of pyroclastics [*Behncke and Neri*, 2003]. Our seismic data analysis covered the time-span from 1 July to 15 August, 2001, and included 16 days before the onset and 7 days after the end of the flank eruption. The data were recorded at the digital station ESPD, deployed 6 km southeast from the summit craters (Figure 1). ESPD belonged to the permanent seismic network run by Istituto Nazionale di Geofisica e Vulcanologia. The station was a Lennartz PCM 5800, equipped with a Lennartz LE-3D broadband (20s), three-component seismometer. The signal was sampled at a frequency of 125 Hz, and transmitted by digital telemetry to Catania, where it was stored on a PC-based acquisition system. We chose this station for: (i) its continuity of acquisition and good signal-to-noise ratio throughout the time span investigated, and (ii) the broadband characteristics of the recordings unavailable for the other stations.

[1]Medical Imaging Group, Department of Physics, University of Bologna, Bologna, Italy.
[2]Istituto Nazionale di Geofisica e Vulcanologia, Sezione di Catania, Catania, Italy.
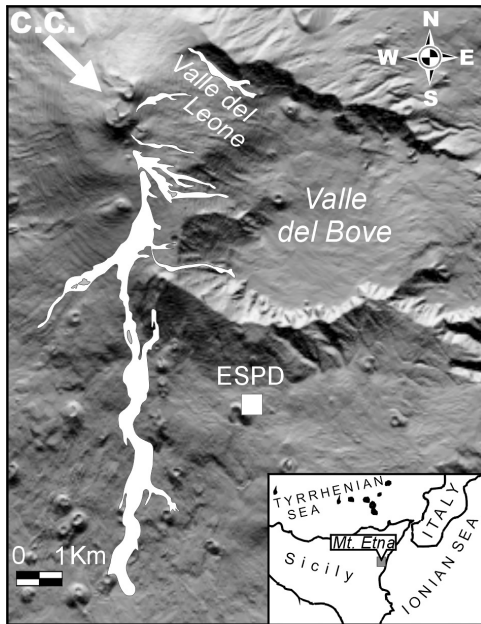
Figure 1. Eruptive field at Etna in 2001. C.C. stands for Central Craters. The square marks the location of the seismic station ESPD.

Over the 46 days investigated, we extracted 142 time series with duration of 10 min from Z and NS components and 141 from EW, achieving a total of 425 time windows. As the data targets for our application had to be representative of the range of signal characteristics in each class, the duration was reduced up to 2 min in a few cases (i.e., during the seismic swarm between 12 and 17 July) to exclude earthquakes from volcanic tremor. Then, the seismic records were divided into four distinct classes: pre–

eruptive (PRE) between 1 and 16 July, lava fountains (FON) both in the pre-eruptive (4, 5, 7, 12, 13, and 16 July) and eruptive stages (17 July), eruptive (ERU) between 17 July and 8 August, and post-eruptive (POS) between 9 and 15 August. Based on this separation, the class ERU encompassed the time series recorded throughout the whole flank eruption with the only exception of those related to the episodes of lava fountains.

[5] Previous investigations found that spectrogram analysis is particularly informative to discriminate different styles of volcanic activity [*Falsaperla et al.*, 2005]. Accordingly, we calculated the Fast Fourier Transform and obtained spectrograms from successive time windows of 1024 points, with overlap of 50%. Each spectrogram had a range of frequencies between 0.24 and 15 Hz, with resolution of approximately 0.24 Hz. By considering each spectrogram as a distinct pattern associated with an a priori defined class, our resulting data set was composed of 153 PRE, 55 FON, 180 ERU, and 37 POS. Figures 2a and 2b depict a typical seismic record and spectrogram for each class. The frequency range of the signal was usually between 0.5 and 3 Hz throughout the whole time span analyzed. Spectrograms relative to the pre-eruptive stage had already warm colors in the frequency range between 0.5 and 2 Hz. In comparison, the episodes of lava fountains had higher energy of the signal. However, the highest values of the energy radiation characterized the eruptive stage, especially in the bands between 1 and 2 Hz (Figure 2b). Finally, the post-eruptive stage marked a condition of relatively low energy radiation (cooler colors prevailing).

3. The SVM Classifier

[6] SVM is a powerful supervised classification technique developed by V. Vapnik in the late 1990s [*Vapnik*,
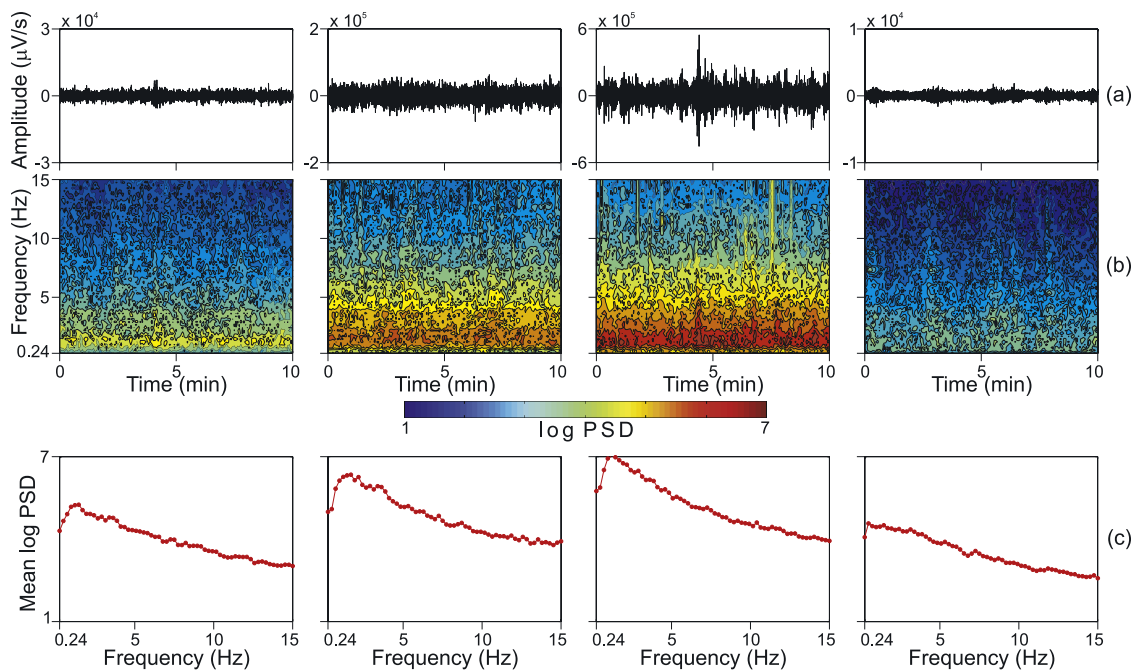

Figure 2. From left to right, examples of pre-eruptive, lava fountain, eruptive, and post-eruptive patterns: (a) time series, (b) spectrograms, and (c) corresponding 62-dimensional feature vectors. The examples are taken from the Z component. PSD stands for power spectral density.
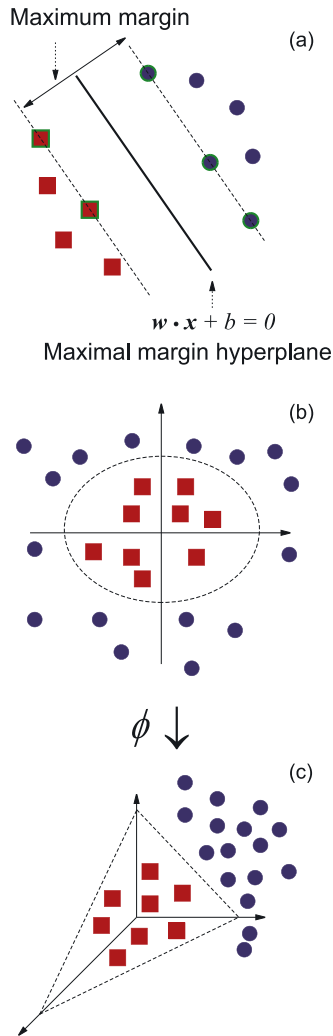
Figure 3. (a) Maximal margin hyperplane found by SVM; the green bordered patterns on the two margins are called support vectors, and are the only ones contributing to the determination of the hyperplane. (b and c) Transformation of a non-linear classification problem into a linear one applying the kernel function ϕ.

1998], and ever since extensively adopted and successfully used within the pattern recognition community. In our SVM application, each spectrogram of volcanic tremor represented a pattern belonging to a given class; it had frequency-time dimensions of 62×145 points, except for those calculated over time series of 2 min whose frequency-time dimensions were of 62×28 points. To work with a homogeneous number of features, we averaged the rows of each spectrogram, ending up with a vector x_i of 62 features (Figure 2c). To accomplish its classification goal, SVM requires a set of l labeled patterns $(x_i, y_i) \in \mathbf{R}^N \times \mathbf{Z}$, $i = 1, \ldots, l$, where x_i is the N-dimensional feature vector associated with the i-th pattern (in this work, a $62-$dimensional spectrogram$-$based feature vector), and the integer label y_i assigned to its class membership (in this work, the volcanic state associated with the i-th pattern, namely PRE, FON, ERU, or POS). $\mathbf{R}$ and $\mathbf{Z}$ are the realm of real and integer numbers, respectively. The training of the

automatic classifier implies the determination of a decision function $f\colon \mathbf{R}^N \to \mathbf{Z}$, such that the l labeled patterns of the training set are all correctly classified or, at least, the error rate (empirical risk) over this set is minimized. The SVM performance is evaluated on a diverse test set (i.e., a set of patterns not used during the training) by comparing the a priori class membership y of each new pattern analyzed with the class membership $f(x)$ assigned.

[7] For a two-class classification problem, i.e., $y \in \{1; -1\}$, the decision function $f\colon \mathbf{R}^N \to \{1; -1\}$ 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 (Figure 3a). This border is defined by a few patterns, the so-called support vectors (Figure 3a). 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, achieving better generalization capabilities. Throughout the training, SVM computes the maximal margin hyperplane as

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

where the vector of weights $\mathbf{w}$ is calculated in terms of the scalars α_i and b by solving a quadratic programming problem [*Vapnik*, 1998]. After the training is completed, the classification of a new pattern x is achieved according to the integer value (i.e., ± 1) resulting from $f(x)$ in equation 1. The α_i coefficients are non-zero only for the small fraction of training patterns (the so-called support vectors) which contribute to the determination of the maximal margin hyperplane. Consequently the number of dot products $\mathbf{x} \cdot \mathbf{x}_i$ which must be actually computed in equation (1) is sensibly smaller than l, and thus assigning a label to x is quite fast.

[8] When patterns are not linearly separable in the feature space, a non-linear transformation $\phi(x)$ is used to map feature vectors into a higher dimensional feature space where they are linearly separable [*Vapnik*, 1998]. With this approach, classification problems – which appear quite complex in the original feature space – can be tackled by using simple decision functions, namely hyperplanes (Figures 3b and 3c). To implement this mapping, the dot products $\mathbf{x} \cdot \mathbf{x}_i$ of equation (1) are substituted by a non-linear function $K(\mathbf{x}, \mathbf{x}_i) \equiv \phi(\mathbf{x}) \cdot \phi(\mathbf{x}_i)$ named kernel [*Vapnik*, 1998]. Admissible and typical kernels are the linear $K(\mathbf{x}, \mathbf{x}_i) = \mathbf{x} \cdot \mathbf{x}_i$, the polynomial $K(\mathbf{x}, \mathbf{x}_i) = (\gamma \, \mathbf{x} \cdot \mathbf{x}_i + r)^d$, the exponential $K(\mathbf{x}, \mathbf{x}_i) = \exp(-\gamma \|\mathbf{x} - \mathbf{x}_i\|^2)$, etc., where γ, r, and d are kernel parameters selected by the user.

[9] The two$-$class approach described above can be easily extended to any k-class classification problem by adopting methods such as the one-against-all or the one-against-one, which basically construct a $k-$class SVM classifier by combining several two-class SVM classifiers [*Weston and Watkins*, 1999]. In this work, with $k = 4$, we used the one$-$against$-$one approach. This method constructs $k(k - 1)/2$ SVM classifiers where each one is trained on patterns from two classes only. A test pattern x is then associated with the class to which it is more often associated by the different SVM classifiers.
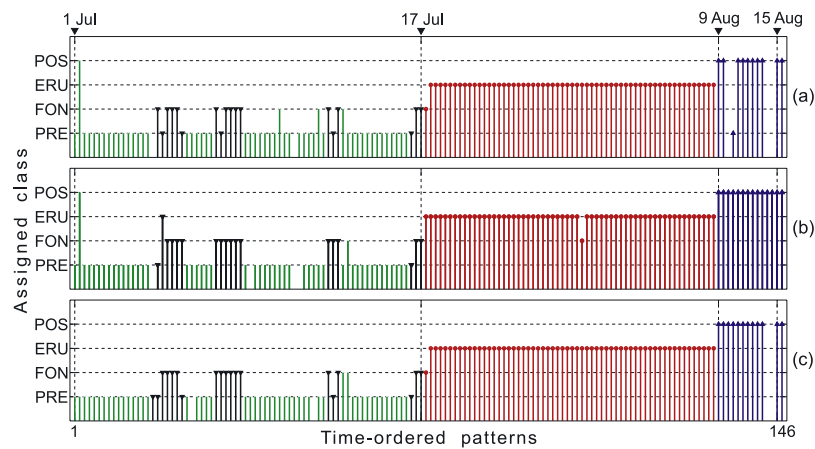
Figure 4. Classification results of the leave-one-out strategy for the (a) EW, (b) NS, and (c) Z component. The colors identify the a priori classification: PRE (green), FON (black), ERU (red), POS (blue). Patterns are time-ordered; gaps correspond to missing data. The arrows on top are time markers.

[10] Although SVM conveys the same basic concepts of other methods with supervised learning, the SVM classifiers have numerous advantages over the latter. Being the solution of a quadratic programming problem, the decision function f found by SVM is unique, hence no local minima solutions occur. By using low-complexity and largest margin decision functions (i.e., using maximal margin hyperplanes), it can be demonstrated that the solution found by SVM is the one with the best trade-off between accuracy on training patterns and generalization capabilities [*Vapnik*, 1998]. Furthermore, the results achieved using SVM are repeatable, as there is no need for random initialization of weights.

4. Results

[11] We evaluated the SVM performances using cross–validation with a random sub-sampling strategy [*Efron and Tibshirani*, 1993]. Accordingly, a training set (ca. 80% of the entire data set) and a test set (ca. 20%) were randomly selected 100 times. This partition corresponded to fivefold cross-validation, which is recommended as a good compromise between bias and variance of the prediction error [*Hastie et al.*, 2002]. The rationale behind the use of this evaluation strategy was that, with respect to a traditional training-validation-test scheme, larger portions of the dataset could be used for training; furthermore, classification performances were estimated as average error rate over the 100 test repetitions, preventing problems arising from spurious splits of the data set. For each repetition, SVM was trained with 123 PRE, 44 FON, 144 ERU, and 30 POS. Then, it was tested on 30 PRE, 11 FON, 36 ERU, and 7 POS. After an initial trial-and-error experimentation, we opted for a polynomial SVM kernel with degree 3 ($d = 3$, $\gamma = 10$, $r = 0$). Classification results showed that, on the entire test set, the class membership assigned by SVM matched the actual class membership $94.7 \pm 2.4\%$ of the times. In particular, $94.2 \pm 5.2\%$ of PRE, $99.6 \pm 0.4\%$ of ERU, and $100.0 \pm 0.0\%$ of POS were correctly recognized. Conversely, FON were correctly recognized for about $76.4 \pm 13.7\%$, whilst for $20.2 \pm 12.6\%$ of the times were misclassified as PRE. Similar classification performances

were achieved on average when the data of the three components EW, NS, and Z were taken into account separately. By considering the EW component only, for example, we obtained a score of $92.0 \pm 5.1\%$ of correct classification. For the NS and Z components, the corresponding scores were $95.6 \pm 4.1\%$ and $93.7 \pm 4.2\%$, respectively.

[12] For further assessing the classification accuracy, we applied a leave-one-out strategy [*Efron and Tibshirani*, 1993]. In this case, SVM was 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 (425), by changing the test pattern in a round-robin manner. The classification results obtained with this procedure matched the actual class membership for 401 out of 425 cases, i.e., 94.4%. Figure 4 depicts how each single pattern was classified by SVM, whilst Table 1 provides the scores for each class, summed over the three components.

5. Discussion

[13] Misclassifications were mostly concentrated near class transitions, particularly between PRE and FON. In Figure 4, we observed that misclassifications were generally not isolated, but rather marked the transition from one volcanic state to the other. This is the case, for example, of the misclassifications associated with the transition between PRE and the first or third FON, as well as those located at the transition between PRE and ERU on 17 July, 2001. A possible reason for that may be the intrinsic

Table 1. Confusion Matrix of the Leave-One-Out Strategy Summed Over the EW, NS, and Z Components[a]

	PRE	FON	ERU	POS
PRE	**145**	4	0	2
FON	12	**41**	2	0
ERU	0	3	**157**	0
POS	1	0	0	**36**

[a]Rows and columns read as a priori and assigned class membership, respectively. Correct classifications in bold in the diagonal elements.

fuzziness (i.e., the ambiguity of the event class membership) particularly in the transition from one stage to the other. This yield a non-null intra-class variability likely responsible for the misclassifications afore-mentioned. In particular, by looking at the plots of the different feature vectors (such as those in Figure 2c), we noted that the intra-class variability of PRE and FON was quite high, with a high number of PRE very similar to FON and vice versa. Conversely, the intra-class variability of ERU and POS was much lower, with just a few ERU similar to FON and very few POS similar to PRE.

6. Conclusions

[14] By achieving less than 6% of classification error on volcanic tremor data recorded at Etna in 2001, SVM performances are quite interesting. Useful applications can be envisaged for on-line processing of data. Off-line classifications of large, past data sets are affordable as well, and may take into account additional classes identified by time-history reports. Besides, the identification of different states of the volcano using just simple spectrograms makes this approach a useful tool for monitoring also other volcanoes where the state of the system can be associated with typical volcanic tremor patterns. Finally, SVM results can be used as a starting point for in-depth investigations on the critical definition of state transitions.

References

Angelini, E., et al. (2006), Testing the performances of different image representations for mass classification in digital mammograms, *Int. J. Mod. Phys. C*, 17(1), 113–131, doi:10.1142/S0129183106009199.

Bazzani, A., et al. (2001), An SVM classifier to separate false signals from microcalcifications in digital mammograms, *Phys. Med. Biol.*, 46(6), 1651–1663, doi:10.1088/0031-9155/46/6/305.

Behncke, B., and M. Neri (2003), The July–August 2001 eruption of Mt. Etna (Sicily), *Bull. Volcanol.*, 65, 461–476, doi:10.1007/s00445-003-0274-1.

Campanini, R., et al. (2004), A novel featureless approach to mass detection in digital mammograms based on support vector machines, *Phys. Med. Biol.*, 49(6), 961–976, doi:10.1088/0031-9155/49/6/007.

Duda, R. O., P. E. Hart, and D. G. Stork (2000), *Pattern Classification*, 654 pp., John Wiley, Hoboken, N. J.

Efron, B., and R. J. Tibshirani (1993), *An Introduction to the Bootstrap*, CRC Press, Boca Raton, Fla.

Falsaperla, S., et al. (2005), Volcanic tremor at Mt. Etna, Italy, preceding and accompanying the eruption of July–August, 2001, *Pure Appl. Geophys.*, 162, 2111–2132, doi:10.1007/s00024-005-2710-y.

Hastie, T., R. Tibshirani, and J. Friedman (2002), *The Elements of Statistical Learning*, 533 pp., Springer, New York.

Langer, H., S. Falsaperla, and G. Thompson (2003), Application of Artificial Neural Networks for the classification of the seismic transients at Soufrière Hills volcano, Montserrat, *Geophys. Res. Lett.*, 30(21), 2090, doi:10.1029/2003GL018082.

Scarpetta, S., et al. (2005), Automatic classification of seismic signals at Mt. Vesuvius volcano, Italy, using neural networks, *Bull. Seismol. Soc. Am.*, 95(1), 185–196, doi:10.1785/0120030075.

Vapnik, V. (1998), *Statistical Learning Theory*, John Wiley, Hoboken, N. J.

Weston, J., and C. Watkins (1999), Multi-class support vector machines, in *Proceedings of ESANN99*, edited by M. Verleysen. pp. 219–224, D. Facto Press, Brussels.

————————————

R. Campanini and M. Masotti, Medical Imaging Group, Department of Physics, University of Bologna, Viale Berti-Pichat 6/2, I-40127, Bologna, Italy.

S. Falsaperla, H. Langer, and S. Spampinato, Istituto Nazionale di Geofisica e Vulcanologia, Sezione di Catania, P.zza Roma 2, I-95123 Catania, Italy. (falsaperla@ct.ingv.it)