Application of Artificial Neural Networks for the classification of the seismic transients at Soufrière Hills volcano, Montserrat

H. Langer,[1] S. Falsaperla,[1] and G. Thompson[2]

Received 1 July 2003; revised 18 September 2003; accepted 6 October 2003; published 5 November 2003.

[1] Seismic activity at Soufrière Hills volcano is characterized by a variety of transients, such as tectonic earthquakes, long-period events, hybrid events, and rockfalls. The huge quantity of seismic data daily recorded on the volcano makes the application of automatic processing highly recommendable. We propose a method of supervised classification of the transients based on Artificial Neural Networks (ANN), which may be useful for processing the large data sets piled up in the past. Particularly, data sets recorded before the climactic eruptions from 1995 to 2002 may allow us to reconstruct the distribution of the different classes of seismic transients in time. We believe that this analysis may give useful insights into impending eruptive scenarios. The good performance of the ANN with 70% of transients correctly classified in a test set of 156 data, along with the opportunity to revise the misfits, make ANN a powerful tool for data processing. *INDEX TERMS:* 7280 Seismology: Volcano seismology (8419); 8419 Volcanology: Eruption monitoring (7280); 8499 Volcanology: General or miscellaneous. **Citation:** Langer, H., S. Falsaperla, and G. Thompson, 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, 2003.

1. Introduction

[2] Modern seismic monitoring of volcanoes is based on continuous data acquisition. Seismic signals on active volcanoes are indeed characterized by a wide variety of transients as well as a more or less persistent form of energy radiation, called volcanic tremor. Continuity and completeness of the data are essential conditions to relate seismic events to volcanic activity, and constrain models for eruptive scenarios. This brings along the problem of huge data masses to process and archive, on the order of 100 MB per day per seismic station. In classical earthquake seismology the problem of large data sets is partly solved by the compilation of data bases containing physically meaningful parameters, such as earthquake magnitude, seismic moment, moment tensor, stress drop, and so forth. Conversely, as the physical characteristics of the seismic sources acting in volcanoes remain controversial, data reduction and parameter extraction in volcano seismology have been hitherto based on signal characteristics.

[3] Moving on from these considerations, we test an automatic process for the classification of seismic transients recorded at Soufrière Hills volcano on the island of Montserrat, West Indies (Figure 1). The volcano unrest at Montserrat in 1995 after 400 years of relative quiescence has yielded climactic eruptions, which have repetitively led the local authorities to the evacuation of the southern half of the island. The dramatic events related to the eruptive activity from 1995 to 1999 [*Druitt and Kokelaar*, 2002] and the lava dome growth since, make Montserrat one of the most dangerous but also best monitored volcanic areas worldwide [*MVO Team*, 1997].

[4] The Soufrière Hills form an andesitic volcanic complex where seismic energy radiation is characterized by numerous transient signals. The various seismic transients recorded in the eruptive periods aforementioned offer a key to interpret the behavior of the volcanic system and assess seismic scenarios leading to impending eruptive episodes. These signals have been grouped according to their waveform, frequency content, and duration [*Miller et al.*, 1998]. Here we focus our attention on the classification of transients which in the Montserrat Volcano Observatory (MVO) terminology are referred to as: (i) volcano-tectonic events (hereafter VT), i.e., earthquakes occurring within the volcano edifice; (ii) regional events (REG), local earthquakes with origin outside the volcano edifice; (iii) long-period events (LPE) [e.g., *Baptie et al.*, 2002]; (iv) hybrid events (HYB), i.e., seismic signals with signatures in-between volcano-tectonic events and long-period events [*White et al.*, 1998], and (v) rockfalls (ROC) [*Luckett et al.*, 2002] (Figure 2). Additionally, we consider a sixth class of events (LPE + ROC) formed by a LPE shortly preceding a rockfall (Figure 2f). We tackle our task in the sense of a classification with supervisor, exploiting the expertise of human operators of the MVO. In so doing, we chose Artificial Neural Networks (ANN) as they have proven to be a versatile and straightforward tool for supervised classification [e.g., *Rumelhart et al.*, 1986]. Such a supervised classification allows the ANN to learn from a number of examples with a given class membership—defined by an expert—and to generalize their main characteristics.

2. Data Set

[5] The data set we analyze covers a time span which ranges from 2000 to 2002. The seismic signals were recorded at 9 digital stations belonging to the permanent seismic network run by MVO (Figure 1). The stations were equipped with Guralp CMG-40T 3C broadband sensors, having a bandwidth between 0.03 and 30 Hz, and dynamic range of 144 dB. Additionally, Integra LA-100, vertical component, 1 Hz seismometers were also used. The sam-

[1]Istituto Nazionale di Geofisica e Vulcanologia, Sezione di Catania, P.zza Roma 2, Catania, Italy.
[2]Montserrat Volcano Observatory, Montserrat, British West Indies.

0094-8276/03/2003GL018082$05.00

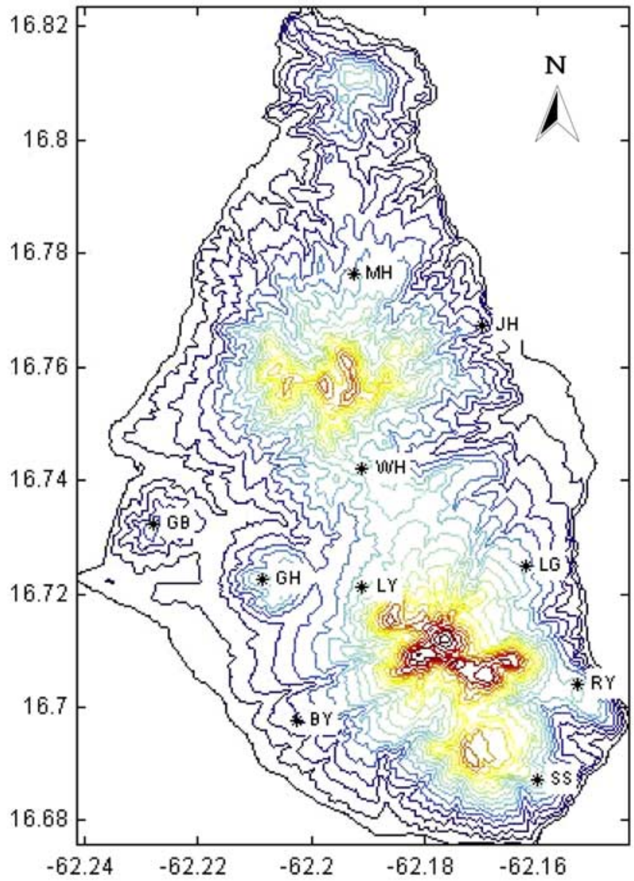

Figure 1. Sketch map of Montserrat. Stars mark location of the stations belonging to the permanent seismic network.

pling rate was 75 Hz. Signals from each station were telemetered to MVO, where they were recorded on a PC-based digital acquisition system.

[6] For our application we use 336 seismic records, considering the vertical component only. These records were originally divided into the six classes aforementioned by MVO's seismologists. The time series have durations from ca. one minute to several minutes and are representative of the large majority of the seismic transients usually recorded. It is worth noting that we chose to use the time series recorded at the different stations for each single transient as separate examples, i.e, the classification is carried out considering each single channel. This is certainly different from what is done by a human operator, who usually assigns different weights to the stations. The use of the single transients as separate examples permits us to explore the performance of the ANN at each station. The analysis of this performance provides information regarding whether reference stations are needed for the future applications of the ANN, and in that case which may be the most suitable stations.

3. Application of the ANN

[7] The architecture of ANN we use for our application is the so-called Multilayer Perceptron [*Rumelhart et al.*, 1986]. This ANN has the advantage to define discrimination functions of arbitrary complexity in principle, and the procedures used for the estimation of the discrimination function do not require any a-priori knowledge about its mathematical structure.

[8] For the set-up of the ANN, we carry out the three following steps: (i) choice of an ANN topology, (ii) identification of representative examples of the transients whose class membership is known, (iii) definition of a suitable form of data representation. We chose a simple topology consisting of three layers of nodes: an input layer, where the input data vectors are stored, a hidden layer, where a nonlinear weighting function is applied and which is necessary to guarantee the generality of the discrimination function, and the output layer representing the resulting vector of the ANN application. Consequently, we consider topologies in the form **U**-X-6, where **U** and X are the number of neurons in the input and hidden layers, respectively, and 6 are the classes of transients to identify. For each single transient, the output vectors contain the class memberships of the signals calculated with the ANN, which are compared to the target values defined by the human operator. We define our target output by assigning "1" to the class which the signal belongs to, and "0" to the other classes.

[9] Formally the classification is carried out by applying a mapping function to the input vector **U** (which represents our signals) to an output vector **Y** (here the class membership values). The mapping function is given by

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

where $\hat{Y}_k$ is the k-th element of **Y** estimated by the network, **U** 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. Previous studies

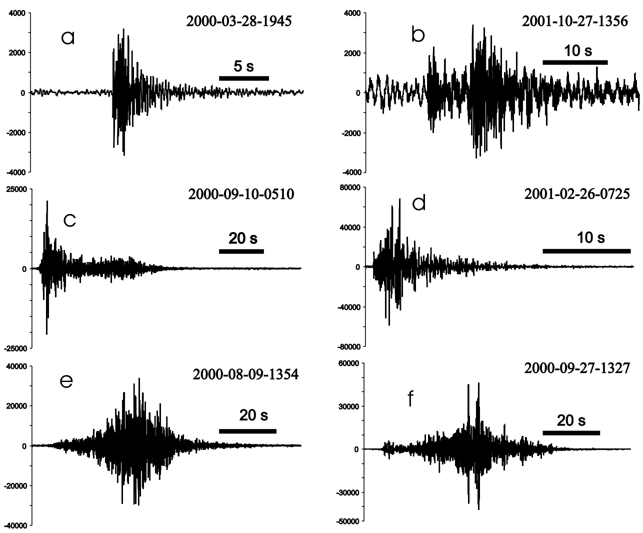

Figure 2. Typical waveforms of the six classes of transients: (a) VT; (b) REG; (c) LPE; (d) HYB; (e) ROC; and (f) LPE + ROC. Year, month, day, hour and minute at the onset of each transient are indicated at the upper right of each trace.

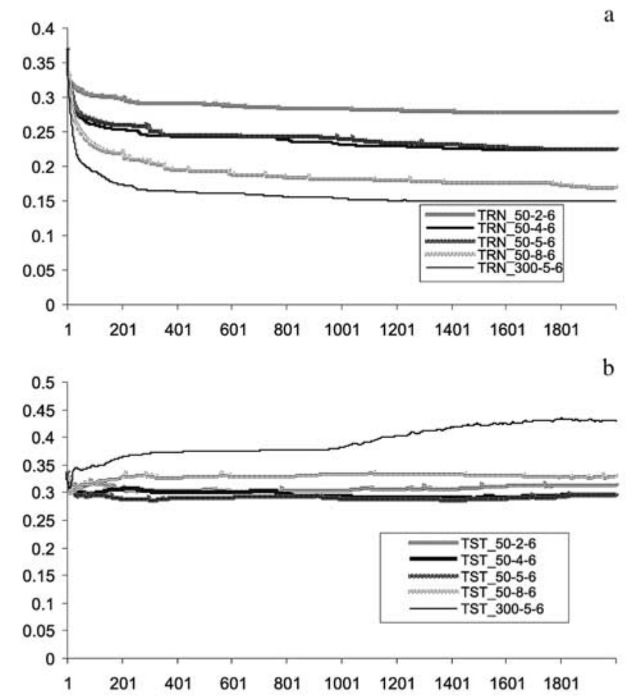

Figure 3. Error curves for training (a); and test (b) of the ANN according to various topologies. Training error curves and test error curves are labeled with TRN_**U**-X-6 and TST_**U**-X-6, respectively.

[e.g., *Falsaperla et al.*, 1996; *Langer and Falsaperla*, 2003] showed that the ANN performance improves when the length of the input data vectors can be limited and phase alignment problems are avoided. Additionally, it is convenient to use an information code which provides a constant length of all the input data vectors, regardless of the original signal duration. We use a combination of the autocorrelation function, statistical parameters (such as the sums of the amplitude signal A), and amplitude ratio of the bandpass filtered (between 0.2 and 0.6 Hz) and unfiltered traces, as input data vectors. We found that the filtered traces were not affected by noise due to sea-microseisms, whose spectral maxima are expected between 0.09 and 0.18 Hz in the ocean [*Kenneth*, 2001]. The autocorrelation functions (acf) represent the spectral content of the signal. The acf were obtained in the frequency domain using a window length of 16384 points for the FFT, and taking the first n points. Being zero-phase functions, they always have their maximum at the beginning of the trace. This avoids the aforementioned problems of phase alignment. The sums of the amplitude over A, A^2, A^3, A^4 resemble statistical moments. They should help to distinguish brief, peaked transients (like VT) from signals with long duration (like ROC), but having the same frequency content. Finally, the amplitude relations may help to separate classes showing a combination of signals, such as LPE + ROC. We divide our data set of 336 records in two groups, by randomly selecting a training set with 180 records, which was used for the estimation of the ANN coefficients, and a test set with 156 records. The random selection makes reasonable the assumption that training and test data sets belong to the same parent population, as requested by theory.

[10] The only parameter controlling the complexity of the ANN and its discrimination function is the number of nodes in the hidden layer, for the length of the input and output layers is defined by the data we use and the number of classes to identify, respectively. The size of the hidden layer is usually determined by an iterative trial and error procedure. Based on the training data set, we iteratively derive the coefficients of the ANN according to a scheme known as back-error propagation [*Werbos*, 1974]. The latter minimizes the error defined by the root-mean square differences between target and calculated output. During the iterative procedure, we simultaneously test the performance of the ANN with the test data set, containing transients which have not being used in the training set. The analysis of the error curves obtained with training and test data sets is a key for the choice of the final ANN's topology. A large number of nodes in the hidden layer usually yields low errors for the training data set and high errors for the test data set (Figure 3). This effect is known as overfitting. In the following, we discuss how the performance of the ANN clearly shows the effects of overfitting whether we use too many nodes in the hidden layer (Table 1).

4. Results and A-Posteriori Inspection of the Misfits

[11] Based on inspection of the output vectors of the training data set, we obtain the best performance of the

Table 1a. Topology 50-5-6

VT	REG	LPE	HYB	ROC	LPE + ROC
Training set					
17	0	0	0	0	1
5	**39**	1	0	3	0
0	0	**12**	0	0	5
0	0	4	**36**	0	0
0	0	2	1	**39**	10
0	0	0	0	0	**5**
Test set					
15	2	0	0	1	1
6	**36**	1	5	3	1
0	0	**6**	2	3	2
1	1	3	**16**	2	1
0	0	1	2	**32**	10
0	0	0	1	2	**2**

Table 1b. Topology 50-8-6

VT	REG	LPE	HYB	ROC	LPE + ROC
Training set					
21	0	0	0	0	0
1	**39**	0	0	0	0
0	0	**12**	0	0	2
0	0	3	**35**	0	0
0	0	3	2	**39**	2
0	0	1	0	3	**17**!
Test set					
11	12	0	0	0	0
11	**26**	0	3	1	0
0	1	**5**	1	3	3
0	0	3	**15**	4	1
0	0	2	4	**25**	4
0	0	1	2	9	**9**

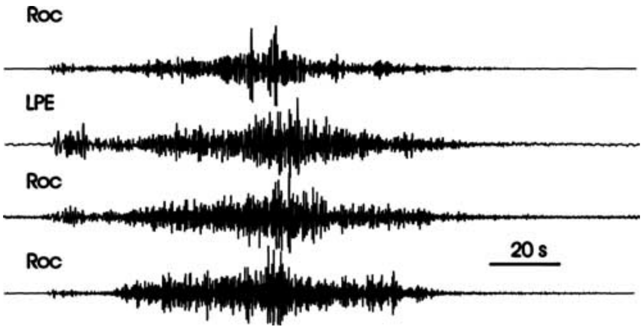

Figure 4. Examples of misfits for the class LPE+ROC. The wrong classification of each transient is indicated at the upper left of each trace.

ANN with the topology 300-5-6, (295 + 5 neurons between acf and the other selected parameters, 5 neurons for the hidden layer, 6 neurons for output). However, the results of the test set are the worst among all tested configurations. The topology which yields the best performance for the test data set is the 50-5-6 (45 + 5 neurons between acf and other parameters, 5 neurons for the hidden layer, 6 neurons for output). We report the results of the training and test data sets for this topology in Table 1a. The class membership assigned by the ANN was defined by taking the class where the highest score was encountered. The generic a_{ij} entry in Table 1 represents the number of transients classified as belonging to the ith class ($i \in$ HYB, LPE, LPE + ROC, REG, ROC, VT), and recognized as belonging to the jth class ($j \in$ HYB, LPE, LPE + ROC, REG, ROC, VT) by the ANN. Hence, the elements of the principal diagonal represent the number of transients correctly classified by the ANN in each class. Using the topology 50-5-6 the expected classification is matched for about 70% of the test data set, with 107 out of 156 transients correctly classified (Table 1a). A slightly larger ANN (50-8-6) provides excellent results during the training phase, whereas the performance of the test is, on the whole, slightly worse with 91 out of 156 transients correctly classified (Table 1b). However, for specific classes such as LPE + ROC, the success of the topology 50-8-6 for the test set is higher than that of the topology 50-5-6.

[12] A success rate of 70% is good enough to be considered encouraging. Nevertheless, we believe that the 30% failure may be dramatically reduced with the help of the performance analysis in the misfits. For example, the lower errors obtained with the topology 50-8-6 with respect to the 50-5-6 for the class LPE + ROC during the training are a hint that the 50-5-6 is not sufficient for resolving this class. The comparison of the results for the two topologies (Table 1), leads us to conclude that we should avoid problems of overfitting by augmenting the number of neurons in the hidden layer along with the overall number of data. As a side effect, an augmented training and test data set enhances the statistical significance of the results obtained with the ANN. Systematic misfits may be attributed to a not well constrained distinction between certain classes. This is the case of the VT which is often confused with REG (Table 1). This misfit, which is related to the choice of the examples for the learning and training of the ANN, is strongly affected by the station location and its distance from the source. A solution for the correct separation of transients such as VT and REG might be the use of some reference station/s with distinct S-P arrival times for each class. A similar solution should solve problems of erroneous classification of HYB, LPE, and ROC, which might be confused for propagation and/or site effects affecting the records. The example of LPE + ROC depicted in Figure 4 highlights how the seismic signature may undergo remarkable changes from station to station.

[13] For the limited data set analyzed here, the stations have more or less the same overall performance, besides MBLG (Figure 1), which has significantly lower misfits than the average and may be a good candidate for becoming a key station.

5. Conclusions

[14] Automatic classification using ANN has the advantage of being objective, in the sense that it is reproducible. It does not depend on the criteria adopted by human operators which may differ from person to person. The a-posteriori analysis helps to verify the significance of the a-priori classification and the criteria used. We surmise that the good performance of the ANN for the classification of seismic transients recorded at Soufrière Hills—with 70% of the transients correctly classified in a test data set of 156—may further improve using larger data sets both for the training and test of the ANN to avoid problems of overfitting.

[15] Our analysis of the misfits suggests the use of reference stations to separate classes that the ANN cannot resolve efficiently. According to our findings, MBLG may be a good reference station. The inspection of the misfits inevitably yields a continuous feedback between the performance of the ANN and what we learn from it as well as from other independent analyses. For example, hypocentral location and observations on the rockfalls may focus the ANN user on stations which can give a better resolution of the seismic source/s. Given the black box nature of the ANN, it is recommendable to check its performance from time to time. Encountering unexpected difficulties in the classification could mean that new kinds of transients have appeared and a new training of the ANN with enlarged data sets is necessary. In this light, the classification of seismic transients should not be seen as a static approach, but it achieves dynamic aspects. Our findings encourage this application to the large data set of approximately 184,000 transients recorded on MVO's analog seismic network between July 1995 and March 2001 and associated with the eruption of the Soufrière Hills. An analysis of this data set would take far too long to do manually, and is an ideal application for a neural network.

[16] **Acknowledgments.** We are grateful to two anonymous reviewers for their useful comments and suggestions. This paper was financially supported by the EC project MULTIMO (Contract No. EVG1-CT-2000-00021).

References

Baptie, B., R. Luckett, and J. Neuberg, Observations of low-frequency earthquakes and volcanic tremor at Soufrière Hills volcano, Montserrat, in *The Eruption of Soufrière Hills Volcano, Montserrat from 1995 to 1999*, edited by T. H. Druitt and B. P. Kokelaar, Geological Society, London, Memoirs, *21*, 611–620, 2002.

Druitt, T. H., and B. P. Kokelaar (Eds.), *The Eruption of Soufrière Hills Volcano, Montserrat from 1995 to 1999*, Geological Society, London, Memoirs, *21*, 2002.

Falsaperla, S., S. Graziani, G. Nunnari, and S. Spampinato, Automatic classification of volcanic earthquakes by using multi-layered neural networks, *Natural Hazards*, *13*, 205–228, 1996.

Kenneth, B. L. N., The seismic wavefield—Volume I: Introduction and Theoretical Development, *Cambridge Univ. Press*, USA, 370 pp., 2001.

Langer, H., and S. Falsaperla, Seismic monitoring at Stromboli volcano (Italy): A case study for data reduction and parameter extraction, *J. Volcanol. Geoth. Res.*, doi:10.1016/S0377-0273(03)00257-9, 2003.

Luckett, R., B. Baptie, and J. Neuberg, The relationship between degassing and rockfall signals at Soufrière Hills volcano, Montserrat, in *The Eruption of Soufrière Hills Volcano, Montserrat from 1995 to 1999*, edited by T. H. Druitt and B. P. Kokelaar, Geological Society, London, Memoirs, *21*, 595–602, 2002.

Miller, A. D., R. C. Stewart, R. A. White, W. Ambeh, L. Lynch, L. Luckett, B. Baptie, W. Aspinall, and R. Latchman, Seismicity associated with dome growth, *Geophys. Res. Lett.*, *25*, 3401–3404, 1998.

MVO Team, The ongoing eruption in Montserrat, *Science*, *276*, 371–372, 1997.

Rumelhart, D. E., G. E. Hinton, and R. J. Williams, Learning internal representation by error propagation, in *Parallel Distributed Processing: Explorations in the Microstructures of Cognitions*, edited by D. E. Rumelhart and J. L. Mc Clelland, MIT Press, Cambridge, Mass, 318–362, 1986.

Werbos, P., *Beyond regression: new tools for prediction and analysis in the behavioral sciences*, Ph.D thesis, Harvard, Cambridge, MA, 1974.

White, R. A., A. D. Miller, L. Lloyd, and J. Power, Observations of hybrid seismic events at Soufrière Hills volcano, Montserrat: July 1995 to September 1996, *Geophys. Res. Lett.*, *25*, 3657–3660, 1998.

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

G. Thompson, Montserrat Volcano Observatory, Montserrat, British West Indies.