IN21D-0732

AGU 100 ADVANCING EARTH AND SPACE SCIENCE | FALL MEETING Washington, D.C. | 10-14 Dec 2018

ISTITUTO NAZIONALE DI GEOFISICA E VULCANOLOGIA

# Can we rely on machine learning to reveal short term features of volcanic activity on Mt. Etna?


**Flavio Cannavo'**[1] (flavio.cannavo@ingv.it), **Andrea Cannata**[1,2] (andrea.cannata@unipg.it)

(1) INGV National Institute of Geophysics and Volcanology, Osservatorio Etneo, Catania, Italy

(2) University of Perugia, Dipartimento di Fisica e Geologia, Perugia, Italy



## Abstract

*Volcanic eruptions are usually not easily predictable and this poses a significant hazard not only for exposures to the local population but due to possible presence of tephra also for airline traffic.*

*The significant investments of the last years in new monitoring techniques and networks have improved our capabilities to sense volcano health, but the path to automatically recognize signs of potentially hazardous unrest is still long.*

*On the other hand, machine learning is currently living a period of tumultuous growth and it is possible to find its applications practically in all the contexts where there is an overflow of data to be interpreted. Our aim is to exploit the capability of some established algorithms in machine learning to test their reliability in early detecting anomalous signals from the monitoring network on Mt. Etna (Italy). In particular, we evaluate the effectiveness of using random forest and support vector machine approaches to learn from the measured signals the complex dynamics of the Etnean volcanic environment without any a-priori information on the data relationships. Such models are then tested against real eruptive cases to assess their performance.*


## Introduction

Hunting for precursors/key-features is always the ultimate goal of any early warning system.

In the last decades, with the development of volcano monitoring networks, huge amount of data of different geophysical, geochemical and volcanological types have been collected and stored in large databases.

Having such big data sets with many examples of volcanic activity allows us to study volcano monitoring from a machine learning perspective. Thus, exploiting opportunities offered by the abundance of volcano monitoring time-series data we can try to address the following questions:

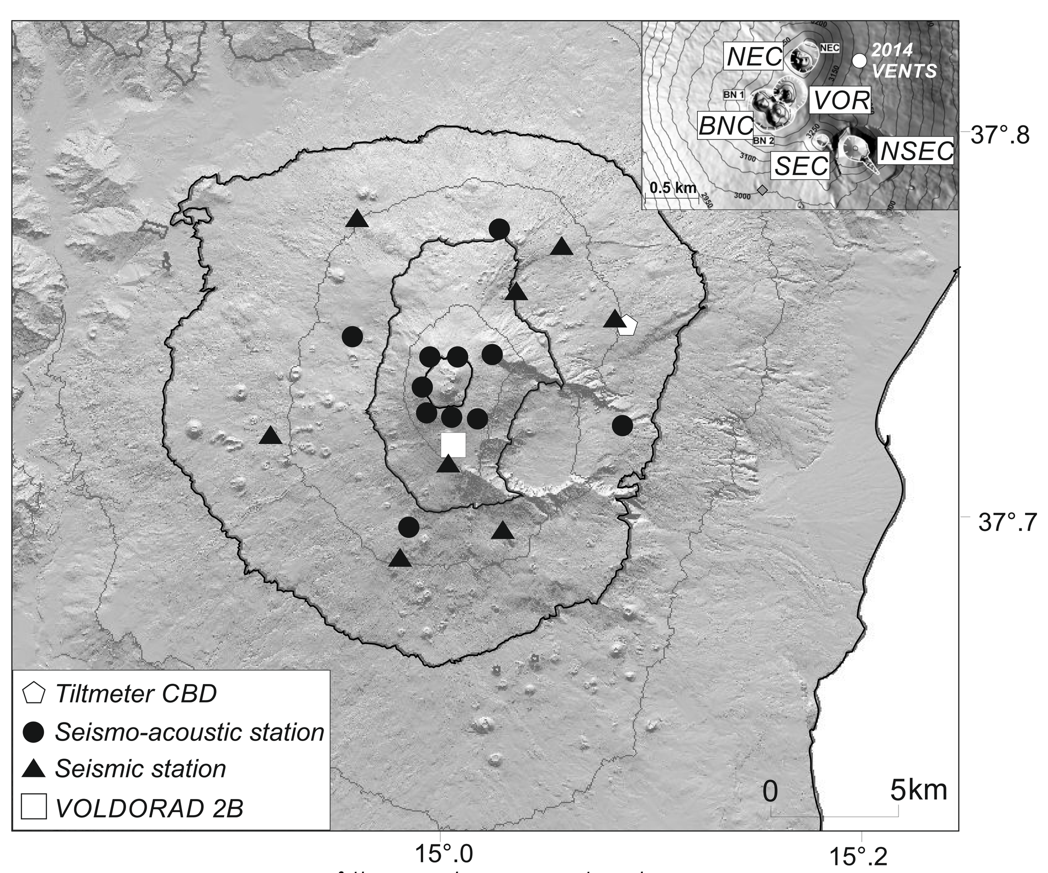
- Are the monitored parameters sufficient to discriminate the volcano state?
- Is it possible to infer/distinguish the volcano state only from the multivariate patterns of measurements?
- Are all the kind of measurements in the pattern equally useful for state assessment?
- How accurate would be an automatic classification system of state inference be based only on pattern recognition of data?

Here we present preliminary results of the data analysis we performed on a set of data and activity covering the period 2011-2017 at Mount Etna (Italy). In the considered period, we had 52 events of lava fountaining and long periods of Strombolian activity.

## Data

We used the following data routinely recorded and calculated at Istituto Nazionale di Geofisica e Vulcanologia (INGV), Osservatorio Etneo for real-time volcano survaillance and monitoring purposes:

- 1) number of shallow (<10 km) earthquakes in the volcanic area




*map of the used sensor networks*

- 2) derivative of tilt data recorded at CBD tiltmeter station installed at 10 m depth and equipped with a high resolution (< 0.005 μrad) tiltmeter (Ferro et al., 2011)
- 3) summit degassed $SO_2$ flux measured using an automated UV scanner array (Salerno et al., 2009)
- 4) total radar echo backscattered by particles in the atmosphere crossing the beam of the 1.274 GHz Doppler radar, called VOLDORAD 2B and installed in the upper southern flank of the volcano (2600 m asl) (Donnadieu et al. 2016)
- 5) seismic tremor location depth obtained by a grid search method, based on a spatial seismic amplitude distribution and assuming the propagation in a homogeneous medium, within 30-min-long time windows.
- 6) thermal index from infrared camera named SARATERM (Andò and Pecora, 2006)
- 7) rate of spatially-clustered infrasound events (Cannata et al., 2011)
- 8) RMS amplitude of seismic tremor in the band 0.5 – 5.5 Hz on the vertical component of the signal within 10-min-long time windows, averaged on 19 stations.
- Data were resampled and synchronized at a rate of 10 minutes. Not all the data cover the considered period so the analysis takes into account only the times when all the data used for the classification are available.

## Methods

We calculated the separability of the features measured during the paroxysms from the same features measured during the rest of the time.

We adopeted the **Silhouette distance** in an Euclidean space to analyse the separability between the two classes.

The Silhouette value is a measure of how similar an object is to its own cluster (cohesion) compared to other clusters (separation). The silhouette ranges from −1 to +1, where a high value indicates that the object is well matched to its own class and poorly matched to neighboring classes. If most feature patterns have a high value, then the classification is appropriate. If many feature patterns have a low or negative value, then the classification is not appropriate and may be an indication of either possible new classes or redundant unsuseful ones.

We considered the mean value of Silhouette distance for all the patterns for a feature configuration. The average distance over all points of a class is a measure of how tightly grouped all the points in the class are. Thus the average distance over all data of the entire dataset is a measure of how appropriately the data have been classified and thus how good the feature configuration is able to discriminate the class (i.e. the paroxysm).

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}}$$

a(i) is the average distance between the i-th data and all other data within the same class

b(i) is the smallest average distance of i-th data to all points in any other class of which i-th point is not a member.

To classify the volcanic activity we used two different algorithms well-established in machine learning field:

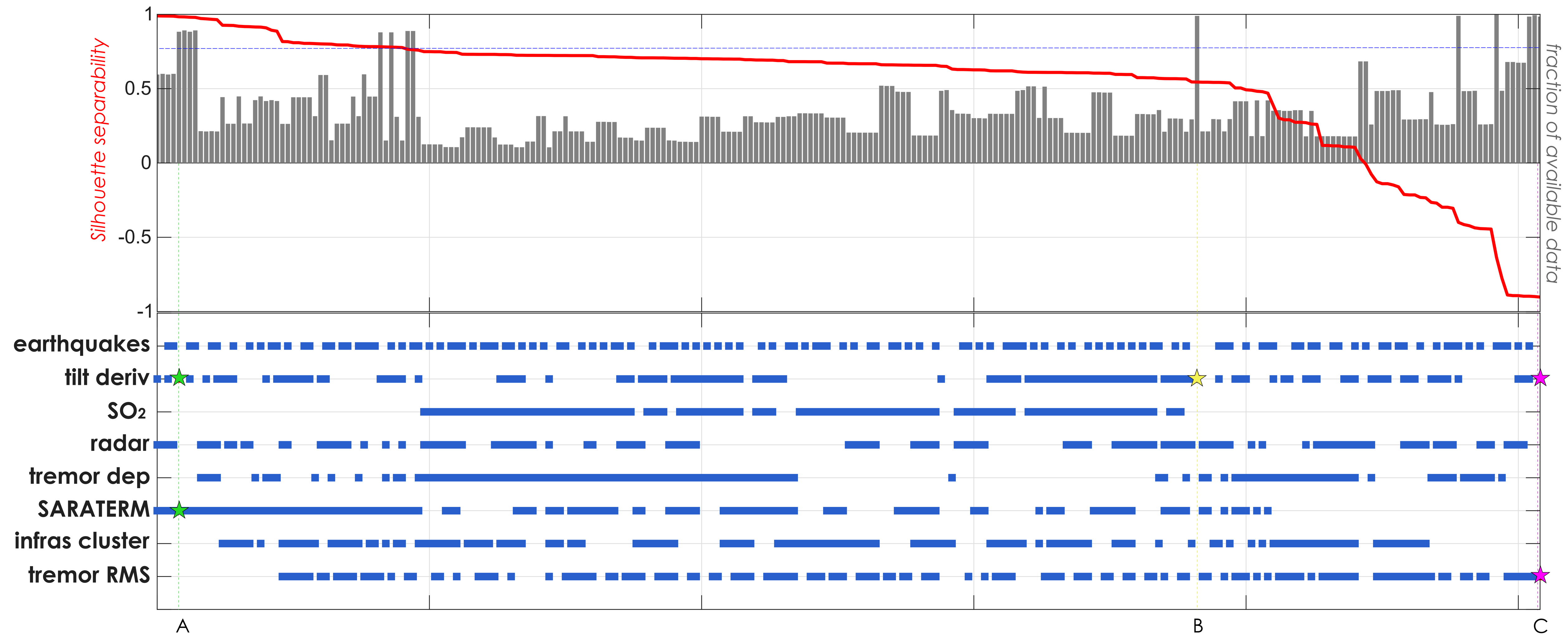
**Random forests** (Breiman, 2001)are an ensemble learning method for classification that operates by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes of the individual trees. This avoids the decision trees' habit of decision tree is a tree where each node represents a feature, each branch represents a rule and each leaf represents a class. The goal is to create a model that predicts the value of a target variable by learning simple decision rules inferred from the data features.

The maximum number of rules defines the complexity of each tree. In our case, we set to 500 the maximum number of splits (rules) in a single decision tree of the forest.

**Support Vector Machine (SVM)** (Cortes & Vapnik, 1995) are supervised discriminative classifiers formally defined by separating hyperplanes. In other words, given labeled training data, the algorithm outputs an optimal hyperplane which categorizes new examples.

## Results

We calculated the separability in terms of Silhouette (line red in the figure) for all the possible combinations among the features. The bars in gray represent the fraction of data available for the considered set of features (shown with blue squares).




The highest separabilities are trivially reached with features involving direct observations of the phenomenon by SARATERM thermal camera.

To evaluate the potential separability of the classes, we adopted the two classification algorithms on 3 crucial feature sets:

**A**: the feature set with the highest silhouette distance with a number of sample greater than 75% of the total

**B**: the feature set with the highest silhouette distance with a number of sample greater than 75% of the total that does not include direct visual observations as SARATERM and radar data

**C**: the feature set with the lowest silhouette distance with a number of sample greater than 75% of the total

We calculated the confusion matrixes for the 3 sets of features classified by the 2 different machine learning algorithms.

The matrixes show the number of samples (time series were sampled at 10-minute rate) classified or misclassified by the algorithms.

It is shown the percentage of samples with respect to the total number of real samples in the considered class. The total accuracy of the confusion matrix is also reported.

The perfomances follow the indication given by the separability measures, i.e. the higher separability the better classification performances.

It is worth noting that although the separability can assume high values (as for case A) it represents only an average over all the samples thus the classification can still suffer of unsatisfactory results depending on the classifier,

| Algorithm | Output Class | A – Target Class: No Paroxysm | A – Target Class: Paroxysm | B – Target Class: No Paroxysm | B – Target Class: Paroxysm | C – Target Class: No Paroxysm | C – Target Class: Paroxysm |
|---|---|---|---|---|---|---|---|
| Random Forest | | Accuracy: 99.94% | | Accuracy: 99.92% | | Accuracy: 99.86% | |
| Random Forest | NoParoxysm | 100.0% 630726 | 30.3% 371 | 100.0% 707077 | 42.4% 554 | 100.0% 703542 | 74.0% 967 |
| Random Forest | Paroxysm | 0.0% 6 | 69.7% 855 | 0.0% 9 | 57.6% 753 | 0.0% 0 | 26.0% 340 |
| SVM | | Accuracy: 99.86% | | Accuracy: 99.86% | | Accuracy: 99.85% | |
| SVM | NoParoxysm | 100.0% 630678 | 70.3% 862 | 100.0% 707079 | 75.3% 984 | 100.0% 703539 | 81.7% 1068 |
| SVM | Paroxysm | 0.0% 54 | 29.7% 364 | 0.0% 7 | 24.7% 323 | 0.0% 3 | 18.3% 239 |
| | | tilt derivative SARATERM | | tilt derivative | | tilt derivative tremor RMS | |

## Considerations

- The separability index is an useful metric to select the features more useful to discriminate automatically the volcanic activity.
- At the moment, the visual inspection from the camera remains the main information source to classify paroxysm activities.
- Even if the classes are apparently well-separated in average, the classification performance can depend on the classification algorithm.