Categories of Infrasound Signals at Mount Etna Inferred from Unsupervised Learning Techniques

Felix Eckel[1], Horst Langer[2], and Mariangela Sciotto[3]

[1]Christian-Albrechts-Universität zu Kiel - Institut für Geowissenschaften
[2]Istituto Nazionale di Geofisicia e Vulcanologia - Osservatorio Etneo
[3]Istituto Nazionale di Geofisica e Vulcanologia - Osservatorio Etneo

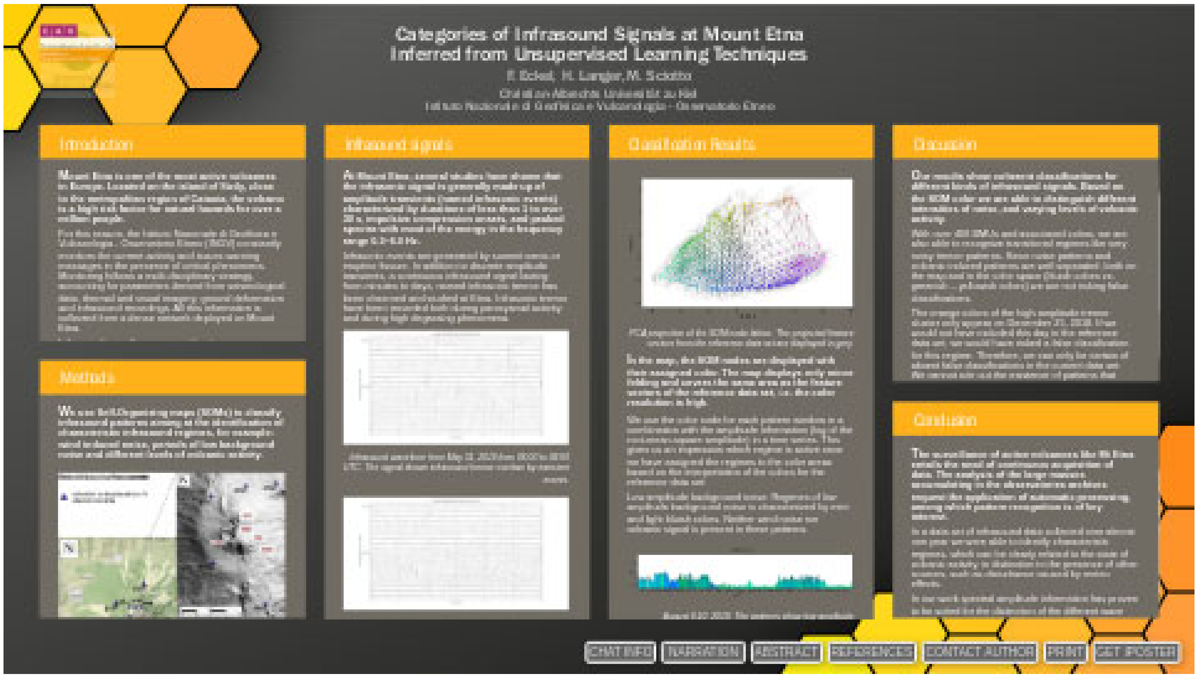
November 24, 2022

Abstract

The frequent activity taking place at Mount Etna may pose risks to tourists and specialists visiting the summit craters as well as to the nearby population and thus requires constant monitoring. Infrasound recordings play an important role in volcanic observation because shallow explosive activity, shallow tremor processes or other volcanic phenomena coupled with the atmosphere are easier to identify with sound waves travelling through the air than with seismic waves, which are more effective for the characterization of buried sources but are strongly affected by scattering within the volcano edifice. Similar to seismic waves, infrasound signals often interfere with noise sources, in case of Mount Etna mostly wind induced noise. The manual distinction of noisy data from real volcanic signals is often not easy, and becomes unrealistic when a large amount of data has to be processed. Currently five summit craters at Mt Etna are active, showing intermittent and fluctuating levels of activity. This leads to a wide variety of infrasound signal patterns coming along with changing noise levels. In order to distinguish between noise induced signals and different kinds of volcano induced signals – such as infrasound events and tremor – in the waveform data we apply unsupervised pattern recognition techniques. We show that by extracting features from the amplitude spectrum, a simple clustering analysis with K-Means yields a set of different activity regimes. A more indepth analysis of the patterns is possible by means of Self-Organizing maps (SOMs) which also allow for the identification of transitional activity regimes and provide the option to color-code the results for an intuitive interpretation. We create a reference data set from multiple months of infrasound waveforms to include as many activity regimes as possible. We apply the results to a test set of data showing how new data can be treated fast and efficiently.

Posted on 24 Nov 2022 — CC-BY-NC-ND 4 — https://doi.org/10.1002/essoar.10506280.1 — This a preprint and has not been peer reviewed. Data may be preliminary.

Categories of Infrasound Signals at Mount Etna Inferred from Unsupervised Learning Techniques



F. Eckel, H. Langer, M. Sciotto

Christian-Albrechts Universität zu Kiel
Istituto Nazionale di Geofisica e Vulcanologia - Osservatorio Etneo

PRESENTED AT:

AGU FALL MEETING
Online Everywhere | 1–17 December 2020

INTRODUCTION

Mount Etna is one of the most active volcanoes in Europe. Located on the island of Sicily, close to the metropolitan region of Catania, the volcano is a high risk factor for natural hazards for over a million people.

For this reason, the Istituto Nazionale di Geofisica e Vulcanologia - Osservatorio Etneo (INGV) constantly monitors the current activity and issues warning messages in the presence of critical phenomena. Monitoring follows a multi-disciplinary strategy, accounting for parameters derived from seismological data, thermal and visual imagery, ground deformation and infrasound recordings. All this information is collected from a dense network deployed on Mount Etna.

Infrasound recordings are sensitive to sources located close to the surface, in the context of volcanic activity degassing processes, lava fountains or Strombolian explosions. They are complementary to seismological data which are effectively applied monitoring sources buried at some depth. They also provide precious information on ongoing volcanic activity when surveillance based on video or thermal cameras is hindered for unfavorable meteo conditions or during night time.

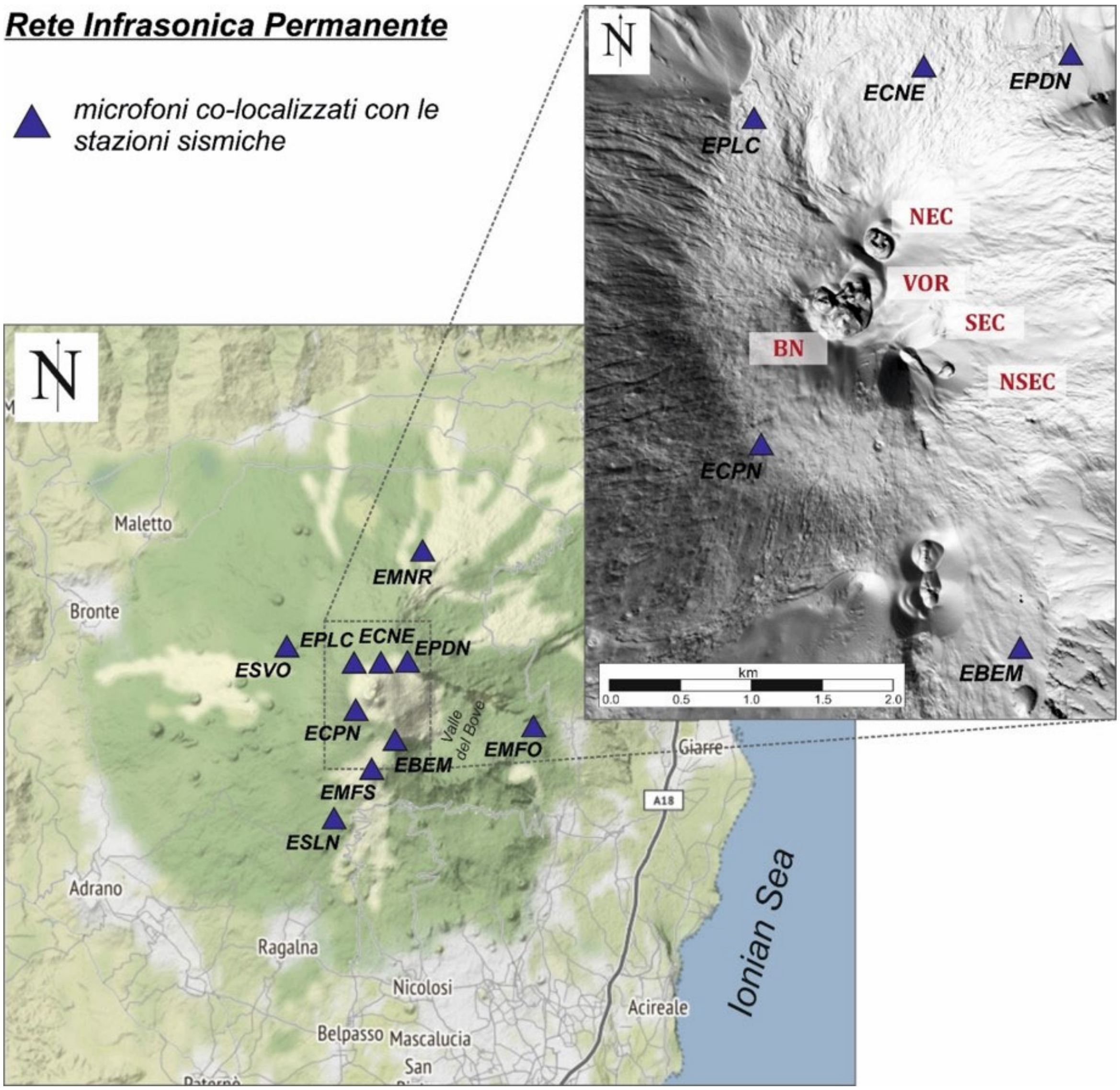
However, the interpretation of infrasound signals is difficult for the untrained eye, as we have to account for various sources, being them linked to volcanic activity with its varying styles or disturbance mainly due to meteorological phenomena – wind gusts, rainfall or thunder.

The problem of handling large and complex data sets can be tackled applying machine learning, namely pattern recognition techniques. Her we focus on so-called "Unsupervised Learning", where we identify groups of patterns being similar to each other. The degree of similarity is based on a metric measuring the distance among the features of the patterns.

Popular methods of Unsupervised Learning are Cluster analysis techniques, which are applied in data sets with strong heterogeneities. Self-Organizing Maps (SOM) or Kohonen Maps have been successfully applied to data sets where also transitional regimes occur.

METHODS

We use Self-Organizing maps (SOMs) to classify infrasound patterns aiming at the identification of characteristic infrasound regimes, for example: wind induced noise, periods of low background noise and different levels of volcanic activity.



Infrasound network on Mt Etna. We use station ESLN in this work.

The direct use of the time series in pattern recognition is often unfeasible. Therefore we obtain a set of feature for pattern windows of 250 s from the spectral content of the waveform.

After some tests we adopted the continuous wavelet transform using a Mexican hat wavelet and 10 logarithmically spaced wavelet scaling factors, which function essentially as frequency filters.

As a result we obtain a time sequence of coefficients related to the scaled wavelets. We use logarithmically spaced scaling factors to have both low and high frequency equally represented in the features. Finally, the sequence of coefficients are averaged over windows with a length of 250 s windows. Each window forms a pattern and is characterized by 10 values stored in the feature vector. The logarithm of these values are taken to account for the large dynamic range of amplitudes. To ensure that each feature dimension has equal weight each component of the feature vector is individually normalized with respect to the variance.

As with any pattern recognition technique, SOMs need a training based on a reference data set. We have selected the reference data set in a way that it should represent all relevant scenarios at best.

SOMs apply a data reduction by introducing so called nodes that function as prototype patterns for a set of input feature vectors. The node closest to any input vector is called best matching unit (BMU). During the training process the nodes are adjusted to minimize the sum of distances between vectors and BMUs making the BMU essentially the center of a micro clusters.

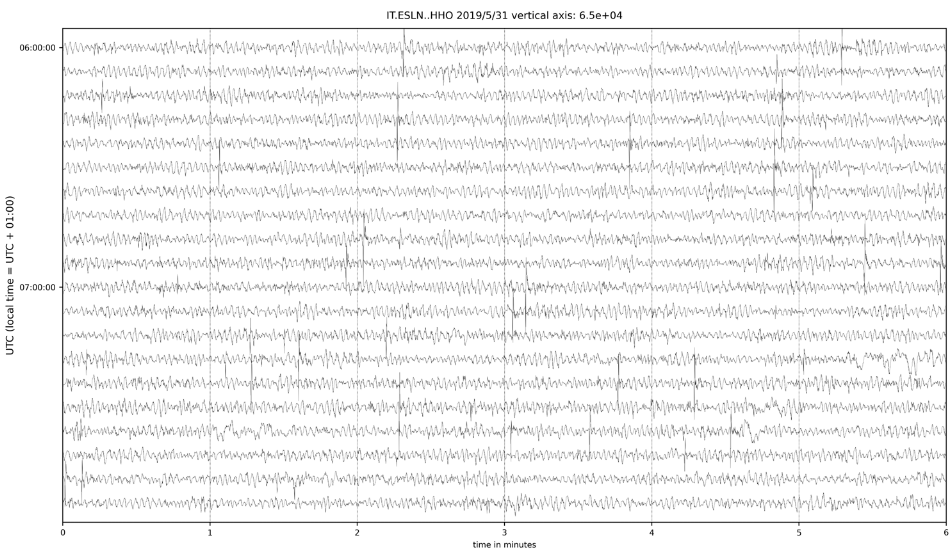
The nodes are then projected into a two dimensional representation space. Ideally, the topological information is conserved in this process, meaning that two nodes that are close in the feature space are also close in the representation space. Additionally, each node is assigned a color based on their position in the representation space. This is done solemnly for visualization purposes.

We analyze data recorded at the station ESLN in the time span from December 28, 2018 and February 29, 2020 (320 days). We focus on the station ESLN due to the data availability and the relatively low levels of wind noise affecting this station compared to stations around the summit. For the reference data set we selected a set of samples covering ca. 24 days.

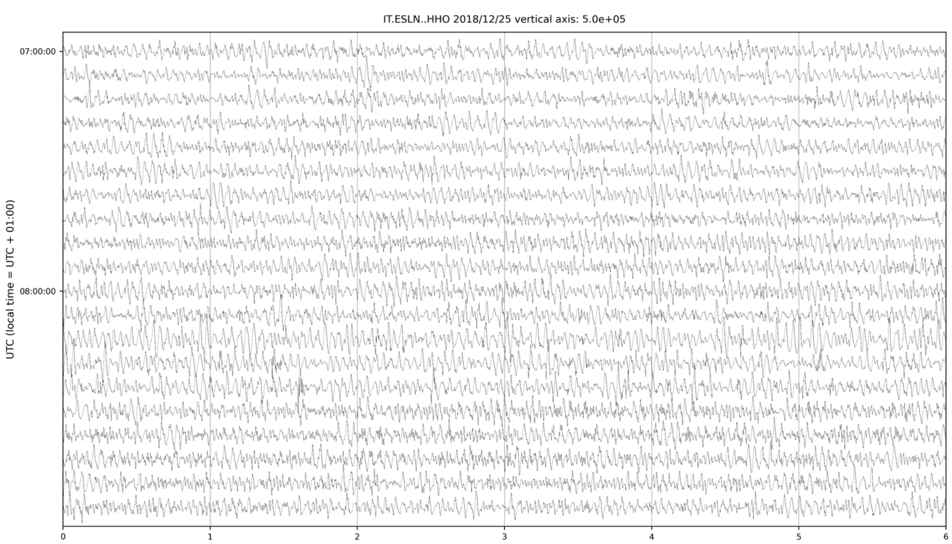
INFRASOUND SIGNALS

At Mount Etna, several studies have shown that the infrasonic signal is generally made up of amplitude transients (named infrasonic events) characterized by durations of less than 1 to over 30 s, impulsive compression onsets, and peaked spectra with most of the energy in the frequency range 0.3–6.0 Hz.

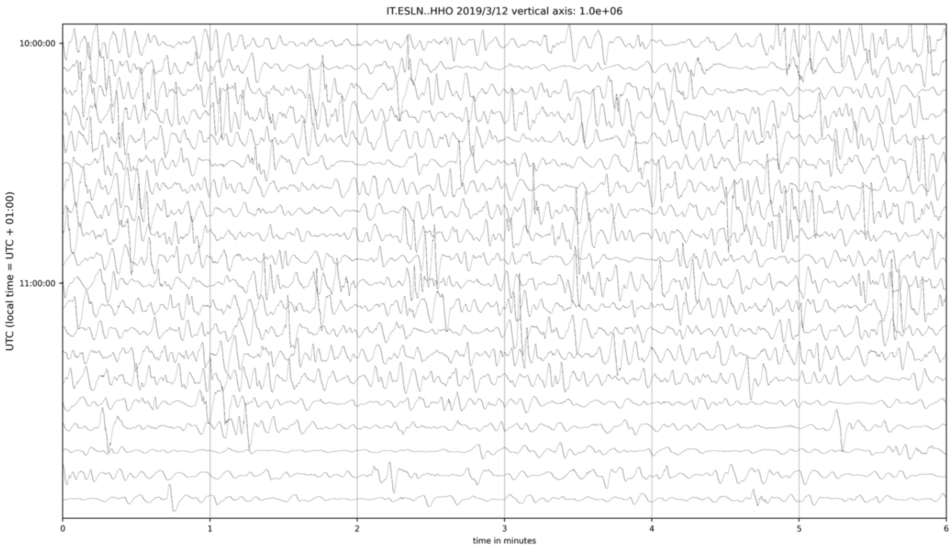
Infrasonic events are generated by summit vents or eruptive fissure. In addition to discrete amplitude transients, a continuous infrasound signal lasting from minutes to days, named infrasonic tremor has been observed and studied at Etna. Infrasonic tremor have been recorded both during paroxysmal activity and during high degassing phenomena.



Infrasound waveform from May 31, 2019 from 06:00 to 08:00 UTC. The signal shows infrasound tremor overlain by transient events.

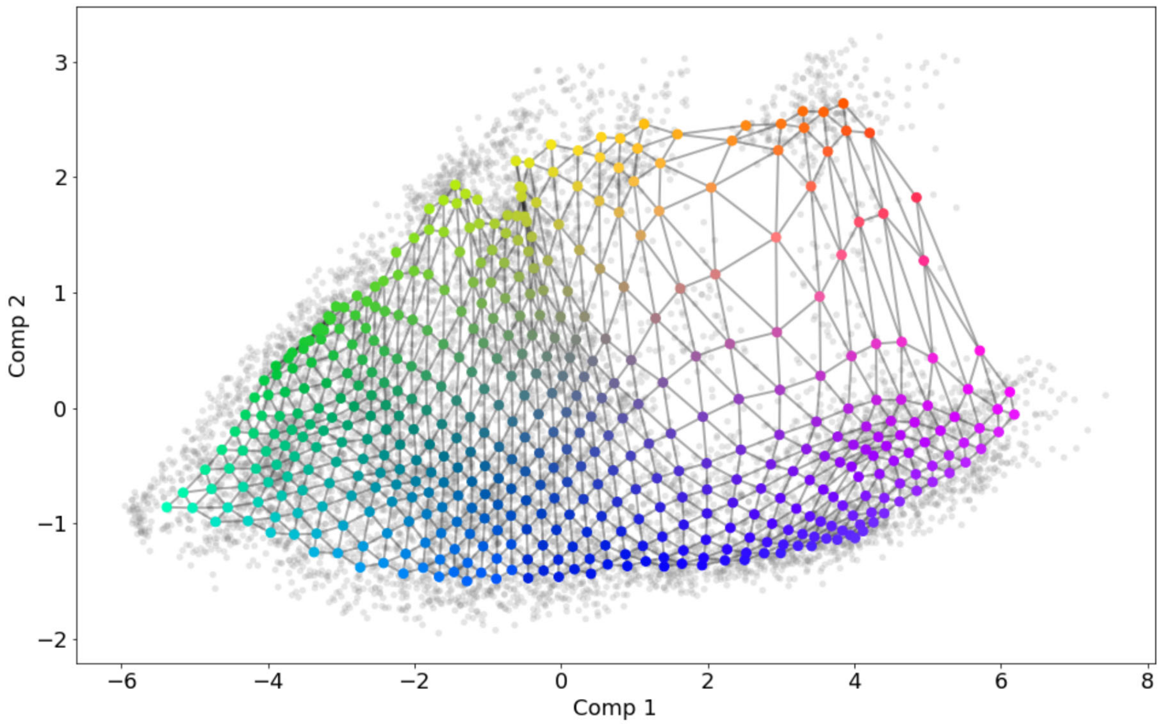


Infrasound waveform from December 25, 2018 from 07:00 to 09:00 UTC. The signal shows a high amplitude infrasound tremor related to the paroxysm on December 18, 2018.



Infrasound waveform from March 12, 2019 from 10:00 to 12:00 UTC. The signal shows energetic wind noise.

CLASSIFICATION RESULTS

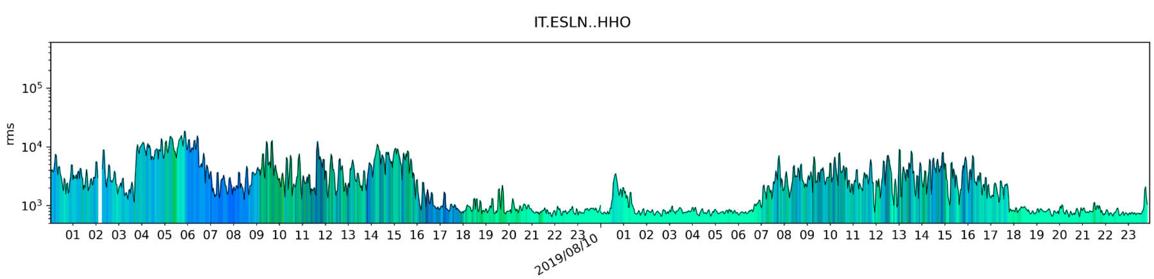


PCA projection of the SOM node lattice. The projected feature vectors from the reference data set are displayed in grey.

In the map, the SOM nodes are displayed with their assigned color. The map displays only minor folding and covers the same area as the feature vectors of the reference data set, i.e. the color resolution is high.

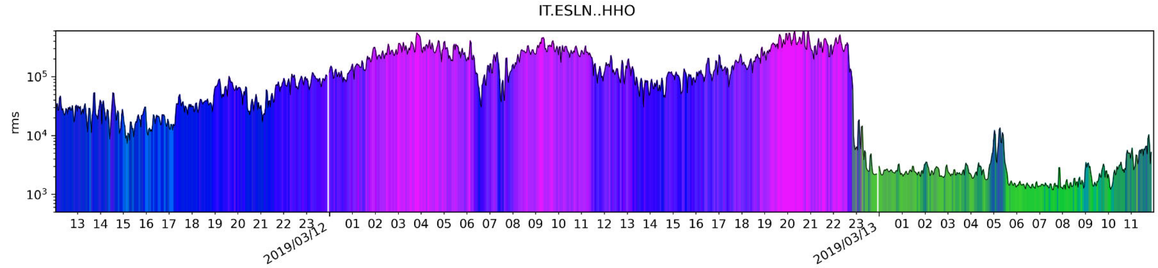
We use the color code for each pattern window in a combination with the amplitude information (log of the root-mean-square amplitude) in a time series. This gives us an impression which regime is active once we have assigned the regimes to the color areas based on the interpretation of the colors for the reference data set:

Low amplitude background noise: Regimes of low amplitude background noise is characterized by mint and light bluish colors. Neither wind noise nor volcanic signal is present in these patterns.



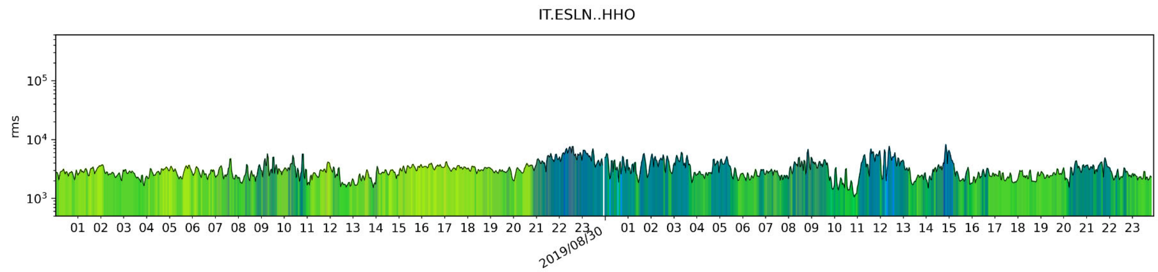
August 9-10, 2019. The patterns show low amplitude background noise.

Wind noise: During regimes with the presence of windy conditions our patterns show bluish and purple colors. The color corresponds to an increasing noise amplitude going from light blue to dark blue and finally into magenta.



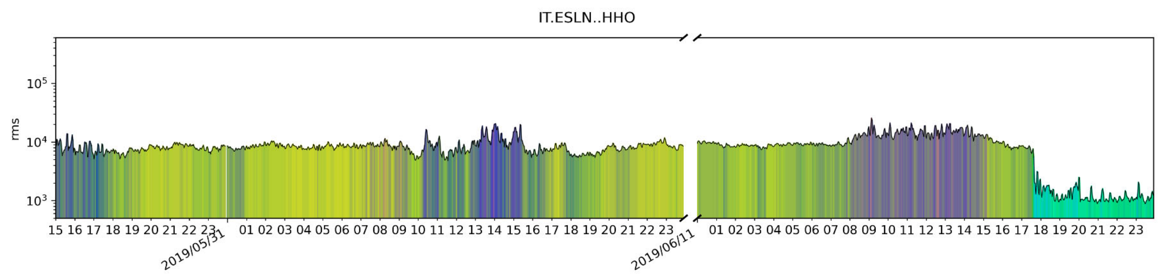
March 11-13, 2019. Blue and magenta colors indicate wind noise. (Compare with waveform in 3rd figure in section "Infrasound signals")

Low volcanic activity: Low amplitude volcanic tremor (for example generated by degassing processes) is displayed in darker green tones. We can say that these colors represent "low volcanic activity" in general.



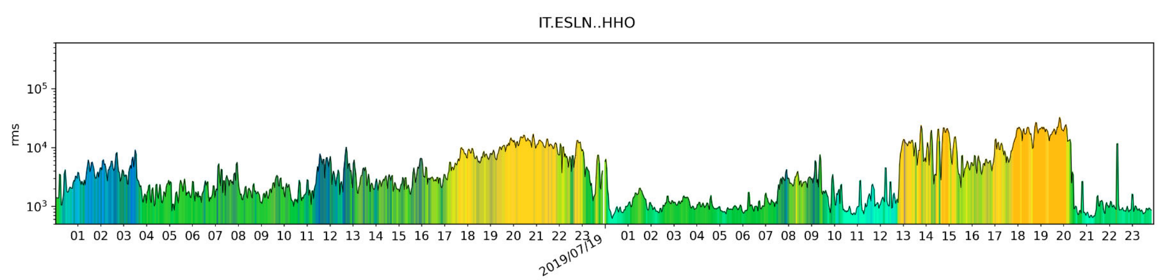
August 29-30, 2019. Dark and medium green patterns indicate infrasound tremor.

Medium volcanic activity: We identify the combination of infrasound tremors and infrasound events (caused by Strombolian explosions) as medium volcanic activity and recognize them by light greenish colors.



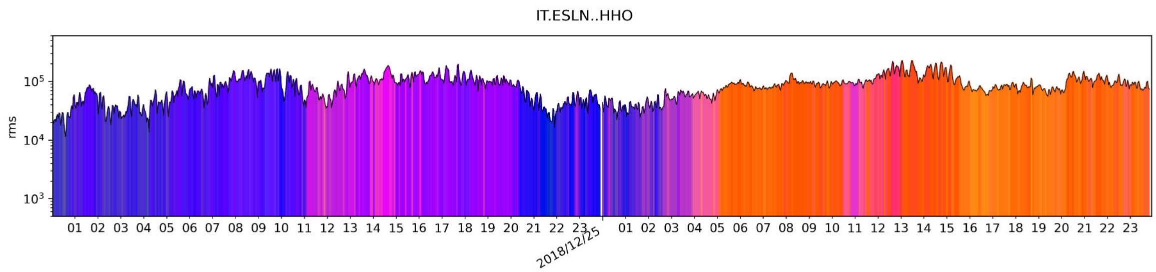
May 30 - June 1 & June 11, 2019. Light green indicate infrasound tremor and infrasound events. (Compare with waveform in 1st figure, section "Infrasound signals")

High volcanic activity: Rapid infrasound events with dominating periods of few seconds represent the occurrence of high volcanic activity. They are displayed in yellow.



July 18-19, 2019. Yellow patterns show increased volcanic acitivity in form of very frequent explosive events.

High amplitude infrasonic tremor: After a paroxysm on December 24n December, 2018, high amplitude, low frequency tremor was observed on December 25, 2018. The patterns observed during this day are the only samples for this regime. They are displayed by orange colors.



December 24-25, 2018. High amplitude infrasound tremor is displayed in orange. (Compare with waveform in 2nd figure, section "Infrasound signals")

DISCUSSION

Our results show coherent classifications for different kinds of infrasound signals. Based on the SOM color we are able to distinguish different intensities of noise, and varying levels of volcanic activity.

With over 468 BMUs and associated colors, we are also able to recognize transitional regimes like very noisy tremor patterns. Since noise patterns and volcanic induced patterns are well separated, both on the map and in the color space (bluish colors vs. greenish – yellowish colors) we are not risking false classifications.

The orange colors of the high amplitude tremor cluster only appear on December 25, 2018. If we would not have included this day in the reference data set, we would have risked a false classification for this regime. Therefore, we can only be certain of absent false classifications in the current data set. We cannot rule out the existence of patterns that have no representation in the current reference data set.

Throughout our testing of different parameters used in the feature generation and map creation, the shape of the map and the color coding showed only minor changes. This indicates a very stable result and confirms the suitability of spectral derived features.

CONCLUSION

The surveillance of active volcanoes like Mt Etna entails the need of continuous acquisition of data. The analysis of the large masses accumulating in the observatories archives request the application of automatic processing, among which pattern recognition is of key interest.

In a data set of infrasound data collected over almost one year we were able to identify characteristic regimes, which can be clearly related to the state of volcanic activity, in distinction to the presence of other sources, such as disturbance caused by meteo effects.

In our work spectral amplitude information has proven to be suited for the distinction of the different wave form patterns. In the Self Organizing Map (SOM) we are able to assign distinct areas which correspond to a regime. The application of SOM has proven to be particularly effective for visualization purposes as the characteristics of the patterns are represent just by a color code. We are thus able to follow the development of the pattern characteristics even over long times just by plotting a sequence of colored symbols – a task difficult to tackle with the crude waveforms or some spectrograms.

In a future development we envisage to carry out a sound validation process, still enlarging the data space and additional scenarios of volcanic activity. Finally we aim at the development of a near real-time processing scheme which can support the staff involved in monitoring Mt Etna's activity.

ABSTRACT

The frequent activity taking place at Mount Etna may pose risks to tourists and specialists visiting the summit craters as well as to the nearby population and thus requires constant monitoring. Infrasound recordings play an important role in volcanic observation because shallow explosive activity, shallow tremor processes or other volcanic phenomena coupled with the atmosphere are easier to identify with sound waves travelling through the air than with seismic waves, which are more effective for the characterization of buried sources but are strongly affected by scattering within the volcano edifice. Similar to seismic waves, infrasound signals often interfere with noise sources, in case of Mount Etna mostly wind induced noise. The manual distinction of noisy data from real volcanic signals is often not easy, and becomes unrealistic when a large amount of data has to be processed. Currently five summit craters at Mt Etna are active, showing intermittent and fluctuating levels of activity. This leads to a wide variety of infrasound signal patterns coming along with changing noise levels. In order to distinguish between noise induced signals and different kinds of volcano induced signals – such as infrasound events and tremor – in the waveform data we apply unsupervised pattern recognition techniques. We show that by extracting features from the amplitude spectrum, a simple clustering analysis with K-Means yields a set of different activity regimes. A more in-depth analysis of the patterns is possible by means of Self-Organizing maps (SOMs) which also allow for the identification of transitional activity regimes and provide the option to color-code the results for an intuitive interpretation. We create a reference data set from multiple months of infrasound waveforms to include as many activity regimes as possible. We apply the results to a test set of data showing how new data can be treated fast and efficiently.

REFERENCES

Cannavò F., Sciotto M., Cannata A., Di Grazia G., 2019, *An Integrated Geophysical Approach to Track Magma Intrusion: The 2018 Christmas Eve Eruption at Mount Etna*, Geophysical Research Letters, 46, 8009–8017 (tel:8009–8017). doi: 10.1029/2019GL083120 (https://doi.org/10.1029/2019GL083120)

Sciotto M., Cannata A., Gresta S., Privitera E., Spina L., 2013. *Seismic and infrasound signals at Mt. Etna: Modelling of North-East Crater conduit and its relation with the feeding system of the 2008-2009 eruption*, J. Volcanol. Geotherm. Res., doi: 10.1016/j.jvolgeores.2012.12.024. (https://doi.org/10.1016/j.jvolgeores.2012.12.024.)

Cannata A., P. Montalto, E. Privitera, G. Russo, and S. Gresta, 2009 *Tracking eruptive phenomena by infrasound: May 13, 2008 eruption at Mt. Etna*, GEOPHYSICAL RESEARCH LETTERS, VOL. 36, L05304, doi: 10.1029/2008GL036738 (https://doi:10.1029/2008GL036738)

Langer H., Falsaperla S., Hammer C., 2020, *Advantages and Pitfalls of Pattern Recognition*, Vol. 3 Computational Geophysics, Elsevier, doi: 10.1016/C2016-0-00254-1 (https://doi.org/10.1016/C2016-0-00254-1)

Messina A., Langer H., 2011, *Pattern recognition on volcanic tremor data on Mt. Etna (Italy) with KKAnalysis - A software program for unsupervised classification*, Computers & Geosciences, doi: 10.1016/j.cageo.2011.03.015 (https://doi.org/10.1016/j.cageo.2011.03.015)