Bulletin of Volcanology (2019) 81: 40
https://doi.org/10.1007/s00445-019-1299-4

RESEARCH ARTICLE

# K-CM application for supervised pattern recognition at Mt. Etna: an innovative tool to forecast flank eruptive activity

Alfonso Brancato[1] · Paolo Massimo Buscema[2,3] · Giulia Massini[2] · Stefano Gresta[4] · Giuseppe Salerno[1] · Francesca Della Torre[2]

Received: 27 August 2018 / Accepted: 18 May 2019 / Published online: 19 June 2019


**Abstract**
We investigated the relationship between the temporal monitoring series routinely recorded at Mt. Etna and the flank eruptions that occurred between January 2001 and April 2005 by the K-contractive map (K-CM) method pattern classifier with supervised learning. The reference dataset includes 28 variables and 1580 records collected over 52 months for a total of 301 eruptive days. A two-step analysis was performed. In the first step analysis, we used the 28 parameters of each day to recognize anomalies heralding a flank eruption. K-CM estimated a sensitivity higher than 95% and a specificity close to 100%. In the second step analysis, we considered each record comprising the 28 variables for 6 days as an input (for a total of 180 inputs) and the outcomes of the seventh day as an output to predict eruption or rest. In this case, K-CM showed sensitivity and specificity close to 98% and 100%, respectively. Results highlight the reliability of the K-CM method to build up a prediction algorithm able to alert the volcano experts a day before the occurrence of a potential flank eruption. The robustness of the two analyses was investigated by the behavior of the receiver operating characteristic curve. The relative area under the curve showed values close to 1, thus providing a valid measure of the performance of the classifier. Finally, a complete overview of the performance levels of the method used was explored analyzing the retrieved Molchan error diagram, in both cases, trajectories very close to the theoretical minimum.

**Keywords** Mt. Etna volcano · Flank eruption · Monitoring data complex system · Neural networks · Supervised pattern recognition

## Introduction

Machine Learning (ML) is a subset of the vast world called artificial intelligence, which is based on a natural computation (NC) that opposes classical computation. While classical computation entrusts the drafting of the rules of the studied phenomenon to a group of experts, which evaluates how much hypothesis are able to describe it (top down approach), NC lets the rules emerge spontaneously from data. This approach is made up of models that generate further models by adapting to

Editorial responsibility: S. Vergniolle

✉ Alfonso Brancato
alfonso.brancato@ingv.it

Paolo Massimo Buscema
m.buscema@semeion.it

Giulia Massini
g.massini@semeion.it

Stefano Gresta
gresta@unict.it

Giuseppe Salerno
giuseppe.salerno@ingv.it

Francesca Della Torre
f.dellatorre@semeion.it

[1] Istituto Nazionale di Geofisica e Vulcanologia-Osservatorio Etneo, Piazza Roma, 2 - 95123 Catania, Italy

[2] Semeion Research Center of Sciences of Communication, Via Sersale, 117 - 00128 Rome, Italy

[3] Department of Mathematical and Statistical Sciences, University of Colorado, Denver, CO, USA

[4] Dipartimento di Scienze Biologiche, Geologiche e Ambientali, Sezione di Scienze della Terra, Università di Catania, Corso Italia, 57 - 95129 Catania, Italy

the data examined (bottom-up approach; Ballard 1999; Arbib 1995; Buscema and Tastle 2013). Recent years have seen an exponential increase in data availability in several fields. This has led the need to determine specific models for each of the field considered which take into account the complexity of the information in recorded data (McKinsey Global Institute 2011). In particular, the power of these models has offered the opportunity to analyze and make explicit "many to many" relationships, i.e., to perform a multivariate analysis of complex phenomena. In this context, artificial adaptive algorithms, capable of handling large and complex set of data, have proven to be the ideal tool. Among these algorithms, deep learning has shown to be of considerable effectiveness (Bengio 2009). Artificial neural networks (ANNs) are one of the most significant tools in multivariate analysis thanks to their ability to deal with the non-linear behavior of a complex system (Buscema et al. 2014). In this paper, we explore the potential and emerging use of the K-Contractive Map (K-CM) classification method (Buscema et al. 2014) in volcanic processes.

A volcano can be considered as a complex system (Langer et al. 2003; Brancato et al. 2016), whose structure and components may change dynamically due to local and/or regional geodynamical interactions, while retaining its cohesion in space-time. Thus, through its changeable behavior, volcanoes generate new data over time, covering from a long- to short-temporal scale (e.g., Newhall and Hoblitt 2002; Martin et al. 2004; Brancato et al. 2011). We can thus define volcanoes as natural adaptive systems, which means that volcanic processes will interact spatially and temporally to produce outputs (Smits 2015). Due to its complexity, this type of system has a highly non-linear behavior and understanding of its dynamics may be achieved by inverting and cross-correlating recorded multiparametric time series (Brancato et al. 2016).

Processes of unrest occurring at volcanoes and preceding eruptions maybe considered as precursors that could be used for forecasting (Marzocchi et al. 2004; Martin et al. 2004; Brancato et al. 2011, 2012; Selva et al. 2012; Tonini et al. 2016). Information on processes occurring in the volcanic dynamics may be used to detect eventual changes in the volcano's state (Martin et al. 2004; Brancato et al. 2011, 2012; Tonini et al. 2016), though sub-surface processes are observably these directly (Mader 2006).

Mt. Etna is located along the Ionian coast of eastern Sicily (Fig. 1) and has undergone both effusive and frequent explosive basaltic eruptions over the last 500 ky ago (Branca et al. 2008, 2011). This persistent and dynamic activity has focused attention on the associated volcanic risk posing a challenge for the Italian Civil Protection, stakeholders, and other decision makers. Questions asked include "When and where will eruptions occur?" and "How long will eruptions last?", and there is a pressing need to reduce uncertainty in forecasting given the densely inhabited area around the volcano, where a million people live in vulnerable areas (e.g., Bonaccorso et al. 2004; Crisci et al. 2010; Barreca et al. 2012). Though summit eruptions (e.g., short-lived lava fountains) are the most frequent phenomena at Mt. Etna (e.g., Andronico and Corsaro 2011; Bonaccorso and Calvari 2013), major crises have been usually associated with lava flows on the volcano flanks, which may persist for weeks to years (Barberi et al. 1992; Calvari and Pinkerton 1998; Andronico and Lodato 2005; Aloisi et al. 2009; Harris et al. 2011).

Mt. Etna is one of the best-monitored volcanoes in the world (Bonaccorso et al. 2004; Mattia et al. 2015), and the vast quantity of multidisciplinary recorded available offers a good opportunity to thoroughly investigate hidden information potentially associated with the dynamics of a volcanic system (e.g., Patanè et al. 2013; Spampinato et al. 2015; Corsaro et al. 2017). The objective of this work is to pinpoint days during which flank eruptions at Mt. Etna might occur by looking at changes in the state of the volcano. Our aim is to define two states: "in a flank eruption" (i.e., *erupting*) and "not in a flank eruption" (i.e., *quiescent*, namely rest). The assumption is, if the volcano is not in a flank eruption today, it is very likely it will not be tomorrow, and conversely, if a flank eruption has started, it is very likely it will continue tomorrow. Furthermore, we aim to determine if any changes that occur before and during an eruption may be used to forecast the onset and the end of a flank eruption, respectively.

The K-CM method is an application of the most general Contractive Map (CM) technique (Buscema et al. 2018), being specifically designed to solve supervised pattern recognition (i.e., classification) problems by using the k-Nearest Neighbor (kNN) criterion (Hastie et al. 2009). Classification is defined as the process of assigning objects to a class. A number of features are used to mark out these objects, and the resulting set constitutes a vector, namely the pattern (also known as record). According to the contractive map technique, because of its equations, K-CM determines a contraction of the input during the learning phase of the network. This process, as better explained later in the text, determines what is called *learning* that is the ability of the network to generalize what seen in the known database, to unknown cases. K-CM has already been successfully applied in classification tasks (Buscema et al. 2014; Grossi et al. 2017). The philosophy behind the K-CM classification method and technical details is reported in Appendix 1.

K-CM classification is one of the current methodologies in multivariate analysis that can find a mathematical model to recognize the membership of samples to their proper class. Here, we present the results of an application of the K-CM pattern classifier with supervised learning at Mt. Etna. A two-step analysis was performed on data recorded between January 2001 and April 2005 focusing on the relationships between geophysical (seismic, ground deformation, gravimetric), geochemical ($SO_2$ and $CO_2$ fluxes), and volcanological

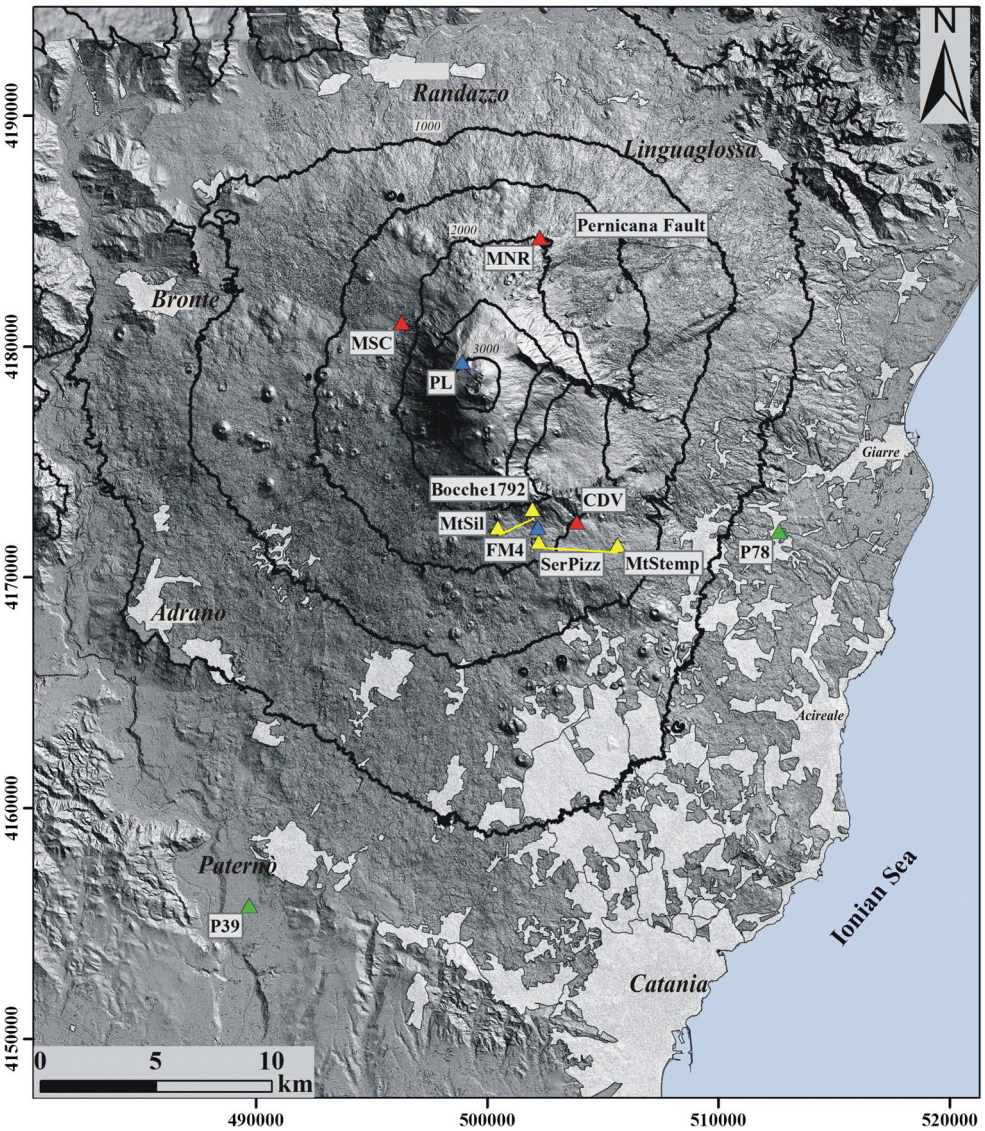


**Fig. 1** Sketch map of Mt. Etna volcano. Green, yellow, blue, and red triangles indicate the geochemical, ground deformation, gravity, and tilt sensor locations, respectively (as listed in Table 1). Yellow solid lines indicate the ground deformation profiles used in this study. The contour lines at 500 m intervals and the urbanized areas are shown. Main towns are also labelled (modified from Brancato et al. 2016)

activity (mainly ash emission) records and the status (i.e., eruptive or not) of the volcano. It means that, in our modelling, each record, i.e., each day characterized by geophysical, geochemical, and volcanological parameters recorded in the same day, must belong to one of the two classes: eruptive day or non-eruptive day. The K-CM classifier is trained to predict changes in volcanic activity (eruption or rest) forecasting being made 1 day in advance. Analysis was performed using the same monitoring dataset used in Bayesian modelling for eruption forecasting by Brancato et al. (2011), with the objective to identify any discrepancies between the two methods and/or new outcomes offered by the application of the K-CM approach. The use of a sophisticated neural network, such as K-CM, allows the creation of a fully data-driven model without the need to introduce any a priori theoretical knowledge regarding the problem or a priori beliefs. It is also a completely different strategy involving complex adaptive systems rather than being based on the concept of event tree as in the Bayesian Event Tree model previously used. Finally, we emphasize that the K-CM analysis is not focused on the likely feedback of the pattern recognition to inform on the underlying volcanic processes, meaning that scope of our study is not to model how Mt. Etna works but to deliver a forecasting tool.

## Material and methods

### Monitoring dataset

Intensive monitoring at Mt. Etna, both routinely and during dedicated campaigns (for the latter, a linear interpolation was used to estimate the missing data), has generated time series data of seismicity (earthquakes), ground deformation, gas emission, microgravity, and petrology as listed in Table 1.

**Table 1** List and relative units of the 28 variables of the dataset collected at Mt. Etna

| Variable | Units | Abbreviation[a] | $S_i$ | $C_i$ | $D_i$ |
|---|---|---|---|---|---|
| M. Silvestri-Bocche 1792 line | μstrain day$^{-1}$ | SerPiz_MtStemp | 0.07 | 1 | 1 |
| Serra Pizzuta-M. Stempato line | μstrain day$^{-1}$ | MtSil_Bocche1792 | 0.07 | 1 | 1 |
| Deformation Pernicana Fault | cm day$^{-1}$ | vel_def_Pern | 0.07 | 1 | 1 |
| Gravity (FM4 station) | μgal day$^{-1}$ | grav_FM4 | 0.07 | 1 | 1 |
| Clinometric variation (MSC station) | μrad day$^{-1}$ | var_MSC | 0.07 | 1 | 1 |
| Variation of the Clinometric mean (CDV, MNR, MSC stations) | μrad day$^{-1}$ | tilt_var_3stat | 0.07 | 1 | 1 |
| Clinometric variation (CDV station) | μrad day$^{-1}$ | var_CDV | 0.07 | 1 | 1 |
| $CO_2$ (P78 station) | g m$^{-2}$ day$^{-1}$ | $CO_2$_P78 | 0.13 | 1 | 1 |
| Areal dilatation | μstrain day$^{-1}$ | ar_dil | 0.14 | 1 | 1 |
| Number of VT earthquakes ($M = 2+$; W sector) | Event day$^{-1}$ | eqs_W_Sect | 0.20 | 1 | 1 |
| Number of VT earthquakes ($D < 5$ km) | Event day$^{-1}$ | eqs_D < 5 km | 0.20 | 1 | 1 |
| Number of VT earthquakes ($M = 1+$) | Event day$^{-1}$ | eqs_volc | 0.20 | 1 | 1 |
| Number of earthquakes ($D \geq 200$ km; $M = 5+$; Tyrrhenian slab) | Event day$^{-1}$ | eqs_slab_Tyrr | 0.20 | 1 | 1 |
| Number of VT earthquakes ($M = 3+$; Pernicana Fault) | Event day$^{-1}$ | eqs_Pern | 0.20 | 1 | 1 |
| Number of VT earthquakes ($D \geq 20$ km; $M = 3+$; NW sector) | Event day$^{-1}$ | eqs_NW_Sect | 0.20 | 1 | 1 |
| Ash emission | No units | ash | 0.20 | 3 | 2 |
| Clinometric variation (MNR station) | μrad day$^{-1}$ | var_MNR | 0.25 | 1 | 1 |
| $SO_2$ emission | t day$^{-1}$ | $SO_2$ | 0.25 | 1 | 1 |
| W flank dilatation | Strain day$^{-1}$ | dil_W | 0.73 | 4 | 6 |
| Clinometric variation (> 0.033 μrad day$^{-1}$; CDV station) | No units | tilt_var_0.033_CDV | 0.73 | 4 | 6 |
| Gravity (PL station) | μgal day$^{-1}$ | grav_PL | 0.73 | 4 | 6 |
| Clinometric measurement (MNR station) | μrad day$^{-1}$ | MNR | 0.94 | 3 | 4 |
| Clinometric mean (CDV, MNR, MSC stations) | μrad day$^{-1}$ | stat_mean | 0.94 | 4 | 7 |
| $CO_2$ (P39 station) | g m$^{-2}$ day$^{-1}$ | $CO_2$_P39 | 1.00 | 4 | 7 |
| Sideromelane | No units | sideromelane | 1.04 | 3 | 5 |
| Clinometric measurement (MSC station) | μrad day$^{-1}$ | MSC | 1.23 | 4 | 8 |
| Number of seismic swarms (> 30 earthquakes day$^{-1}$) | No units | eqs_swarm | 2.78 | 2 | 5 |
| Clinometric measurement (CDV station) | μrad day$^{-1}$ | CDV | 4.17 | 4 | 15 |

Columns 4–6 report the estimates of the *Association Strength* $S_i$, the *Node Centrality* $C_i$, and the *Node Degree* $D_i$, respectively. The list is sorted based on the $S_i$ column

[a] Code given to the variable name for software purposes

We are aware that interpolation is not appropriate for a prediction tool, since it cannot be used to project into the future; thus, we only used interpolation for short rest periods at Mt. Etna. All 28 selected monitoring parameters ($N$ variables; Table 1) represent a multidisciplinary dataset intended to provide a detailed picture of the quiescent background state of Mt. Etna. The reader has to focus that each of the 28 variables represents an operational variable for overall pattern recognition rather than providing individual independent field observations, as clearly indicated by the clinometric parameters. CDV data, recorded at CDV station (Fig. 1), perform both independently (Clinometric measurement (CDV station) or Clinometric variation (> 0.033 μrad day$^{-1}$; CDV station) or Clinometric variation (CDV station); CDV, tilt_var_0.033_CDV and var_CDV, respectively; Table 1) and together with other tilt data (Variation of the Clinometric mean (CDV=Casa Del Vescovo, MNR= Monte Nero, MSC= Monte Scavo stations) or Clinometric mean (CDV, MNR, MSC stations); tilt_var_3stat and stat_mean, respectively; Table 1) to derive different variables to describe different volcanic scenarios. The merging is useful as the tilt network reflects the final intrusive phase by causing changes at most stations with an amplitude related to source-station distance (i.e., a high variation for a closer station and a smaller one for a distal station; Ferro et al. 2011). In detail, these phases are usually preceded by large variations (up to over 100 μrad; Gambino et al. 2014).

The time series spans January 2001–April 2005 (a total of 1580 records), during which Mt. Etna underwent the July–August 2001, the October 2002–January 2003, and the September 2004–March 2005 flank eruptions, with a total of 301 eruptive days. This reference period was chosen as it includes most of the flank events, which have occurred at Mt. Etna over the last few decades. A further flank eruption that occurred between May 2008 and

July 2009 was not considered since there was a switch to continuous acquisition starting from late 2005. This led to a complete change in the dataset.

We point out that the analysis in Brancato et al. (2016), though used the same dataset, provided different results by using different techniques (i.e., genetic algorithms). Brancato et al. (2016) showed a pattern recognition rather than a predictive outcome, testing whether certain machine learning techniques performed on a series of signals (i.e., monitoring data) could estimate the typical distribution of eruptive activity onset. The authors implemented a selection of the most predictive monitoring series, reducing the initial dataset (i.e., 28 variables) usually run for Bayesian probability estimates to 11 variables (Brancato et al. 2011).

The dataset here is a revised form of that used to forecast flank eruptions at Mt. Etna with a Bayesian Event Tree approach, after an expert elicitation (Brancato et al. 2011). The current dataset includes only the temporal series (recorded and/or collected values) of any monitored parameter (Table 1), neglecting the inertia and anomalies information needed for the former application (see Brancato et al. 2011). Brancato et al. (2011) refer to inertia as the timescale during which the parameter evolves, according to the belief of an expert, towards a more anomalous value, which indicates a change in volcanic state (e.g., from a magmatic unrest to an eruption). For the sake of clarity, different timescales of days to weeks/months/years were reported (Brancato et al. 2011). The revision process exploits the capacity of the machine learning approaches to objectively pinpoint hidden links between the monitoring data and the state of Mt. Etna (quiescent or eruptive). The approach allowed also to avoid any modelling of the underlying volcanic processes and the associated dependence on the subjective view of the modeller, primarily in selecting the thresholds and inertias of the involved monitoring parameters. Specifically, the role of machine learning techniques is based on detecting anomalies in the dataset without any specific setting for the anomaly.

Seismic monitoring has convincingly demonstrated that flank eruptions at Mt. Etna are often preceded by earthquake unrest (e.g., Brancato and Gresta 2003; Patanè et al. 2003; Feuillet et al. 2006), on timescales of a few days to weeks. The different seismic parameters are spatially relevant. In other words, different locations imply different links to the various volcanic stages (from unrest phase to eruption) of Mt. Etna. Earthquakes occurring in the Tyrrhenian slab (Number of earthquakes ($D \geq 200$ km; $M = 5+$; Tyrrhenian slab); eqs_slab_Tyrr; Table 1) and along the Pernicana Fault (Number of VT earthquakes ($M = 3+$; Pernicana Fault); eqs_Pern; Table 1, as well as deep seismicity in the NW sector (Number of VT earthquakes ($D \geq 20$ km; $M = 3+$; NW sector); eqs_W_Sect; Table 1) of Mt. Etna, reveal early stages of deep magma movements (unrest phase). Conversely, shallower seismicity (depth less than 5 km; Number of VT earthquakes ($D < 5$ km); eqs_D < 5 km; Table 1) depicts a strong magma involvement (Brancato and Gresta 2003; Patanè et al. 2003). Such seismicity is mainly clustered (Number of seismic swarms (> 30 earthquakes day$^{-1}$); eqs_swarm; Table 1), being likely related to a fracturing field on the flanks of the edifice (Brancato and Gresta 2003). In particular, eqs_swarm features a variable with a time and frequency definition (30 events/day) of earthquakes clustered anywhere on the Mt. Etna edifice, usually indicating the location of the vent zone (flank eruption occurrence).

Ground deformation sampling campaigns on monthly timescales, which become daily when magma reaches the vent zone, are highly useful indicators of existing volcanic activity (Bonaccorso et al. 2004, 2006; Aloisi et al. 2006; Bonforte et al. 2007). We differentiate the relative monitoring by considering GPS, electronic distance measurements (hereafter, EDM), and tilt variables. The anomalies of the GPS variables (W flank dilatation and Deformation Pernicana Fault; dil_W and vel_def_Pern, respectively; Table 1) are usually associated with a ductile behavior induced by the plumbing system, while anomalies of the EDM variables (M. Silvestri-Bocche 1792 baseline, Serra Pizzuta-M. Stempato baseline and areal dilatation; SerPiz_MtStemp, MtSil_Bocche1792, and ar_dil; Table 1) give first-order information about the inflation/deflation state of Mt. Etna. Specifically, EDM monitors the horizontal component of ground deformation. From the end of the 1970s, three separate networks operated at Mt. Etna: on the northeastern, western, and southern flanks. In particular, the M. Silvestri-Bocche 1792 and Serra Pizzuta-M. Stempato baselines (Table 1 and Fig. 1) provide insights into the spreading of the SE and the middle part of Rift S, respectively, when an extension (due to a magmatic intrusion?) occurs on the well-known NNW-SSE structures. Conversely, tilt data, due to the timescales of a few hours to days, are indicators of the rapid evolution of the internal magma movements immediately preceding the onset of flank activity (Ferro et al. 2011; Gambino et al. 2014). The different sites cover the sectors of Mt. Etna (Fig. 1) that underwent most of the recent flank activity. In particular, tilt_var_0.033_CDV (Table 1) variable gives information about the early stage of an unrest phase, whereas CDV and var_CDV (Table 1) variables (similar to Clinometric measurement (MNR station) or Clinometric variation (MNR station) and Clinometric measurement (MSC station) or Clinometric variation (MNR station) (MNR, var_MNR, MSC, and var_MSC, respectively; Table 1) are linked to the final magma intrusion during the onset of flank activity at Mt. Etna. On the other hand, tilt_var_3stat and stat_mean (respectively; Table 1) may describe the stages linked to the internal movement of the magma towards the surface of the volcano.

Analysis of gas emissions, both observed remotely from the volcanic plume as bulk contribution from all summit

craters ($SO_2$ emission, $SO_2$; Table 1) and directly from the soil ($CO_2$ (P39 station) and CO2 (P78 station); CO2_P39 and CO2_P78; Table 1 and Fig. 1), provides insights into the magmatic feeding system of Mt. Etna from the deep to the shallow part of the volcano edifice, with timescales from weeks to months (e.g., Badalamenti et al. 2004; Caltabiano et al. 2004; Salerno et al. 2018). An increase in bulk $SO_2$ emission has been shown to relate to the ascent of a new batch of gas-rich magma from a shallow zone of 3–4 km beneath summit craters (e.g., Sutton et al. 2001; Caltabiano et al. 2004; Salerno et al. 2018). Conversely, anomalies in $CO_2$ flux suggest emplacement of undegassed magma at ~ 12 km (e.g., Notsu et al. 2006; Spilliaert et al. 2006; Giammanco et al. 2012). In our study, temporal anomalies at P39 site (Fig. 1) indicate exsolution from depths of ca. 12 km, while anomalies in $CO_2$ flux at P78 site indicate shallower magma movements from depths of ca. 4 km (Giammanco et al. 2012).

Anomalies in microgravity measurements (Gravity (FM4 station) and Gravity (PL=Punta Lucia station) variables; grav_FM4 and grav_PL, respectively; Table 1) have also been observed a few months before the eruption onset (Carbone et al. 2003; Carbone and Greco 2007). Both series are recorded at PL and FM4 stations (Fig. 1), which represent the sites where the maximum variation was observed (Carbone et al. 2003). The PL sensor belongs to the N-S profile, which is aimed at observing magma rising from any particular direction, while the FM4 sensor belongs to the E-W profile, designed to observe the dynamics of the southern flank of the Mt. Etna edifice.

Figure 2 shows the (normalized) monitoring time series used for the current analysis, which evidence variations in both magnitude and temporal scale in connection with the observed eruptive activity at Mt. Etna. In particular, all the selected series show significant variations before the July–August 2001 and October 2002–January 2003 flank eruptions. No particular trend (either increasing or decreasing) is observed, with the sole exception of CDV variable values (gray solid line; Fig. 2) which shows a continuous decreasing trend after the end of the July–August 2001 eruption. Finally, none of the selected data show anomalous values before the September 2004–March 2005 flank event but instead show significant variations a few days before the end of the relative eruptive activity (Fig. 2).

Other monitoring data commonly include petrologic observations (Ash emission and Sideromelane; ash and sideromelane, respectively; Table 1) of the volcanic rocks. Analysis of the physical and chemical properties of emitted volcanic rocks can provide important constraints for modelling. In particular, the presence of sideromelane in volcanic ash is indicative of a fresh magma injection, hence a possible

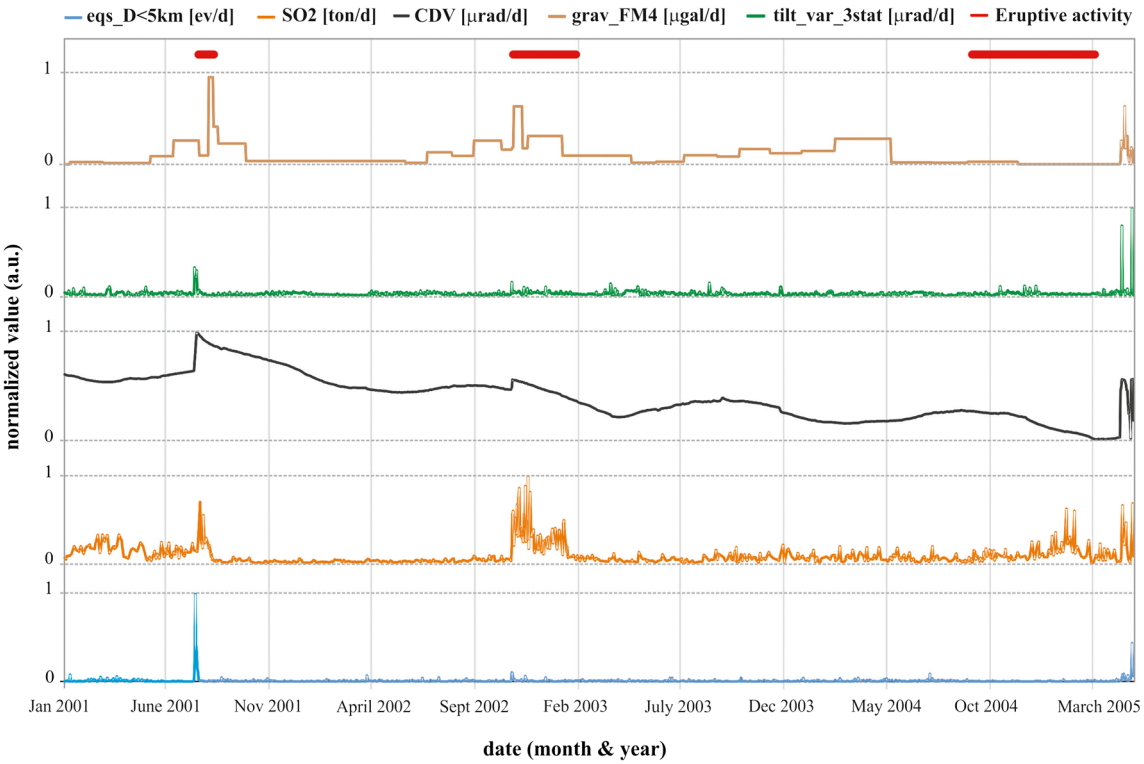


**Fig. 2** Selected time evolution of the monitoring parameters employed in the present study. For each series, data are normalized to the relative maximum value (for graphical display). Also shown (top of the panel) the eruptive activity at Mt. Etna (in terms of lava flow) which occurred during the January 2001–April 2005 period. It is evident how the September 2004–March 2005 flank episode was not heralded by any anomalous monitoring recording

precursor of imminent flank activity at Mt. Etna (Andronico et al. 2009; Corsaro et al. 2017).

## Auto-CM algorithm application

Time series, and their relationship with the eruptive activity of Mt. Etna, was explored performing a two-step analysis. The first analysis is a classical problem of pattern recognition. For any day (i.e., any record), we considered the 28 values of the variables of the day before as input for the K-CM algorithm. To evaluate the correct learning of a network, it is necessary to verify its ability to predict cases never seen before. For this reason, it is usual to divide the sample into two subsets, called training set and testing set. For both subsamples, we know the class of belonging of each record (in this case, eruptive or non-eruptive) called *target value*. The learning is carried out on the training set showing to the neural network both the record and the relative class of belonging. Then, we try to predict the target value of the testing set in a test called *blind*. The network, on the basis of what has been learned about the training set, will assign a class to each of the testing records. At this point, it will be possible to verify the percentage of correct forecasts and the percentage of errors. Both training and testing sets include 1580 patterns each. Leave One Out (LOO) testing protocol (see Appendix 1) was used to validate relative outcomes. This type of analysis allowed to assess whether the system is able to predict the state of the volcano the next day, based on the data recorded the previous day, and provide volcanologists with essential information and alerts. In the second step analysis, we applied a moving window with a length of 6 days to compute the probability of each target (eruption or rest) for the seventh day. Thus, for each day, we codified the $28 \times 6 = 168$ values of the variables and the relative $2 \times 6 = 12$ outcomes (both targets for each day) for a total of 180 inputs. After that, we codified the outcome of the seventh day as output. We then divided the sample of 1574 (= 1580–6) days into two subsamples (787 records each) following a chronological scheme. This type of experimentation focuses upon the possibility to assess whether the system is able to predict the state of the volcano on the seventh day, based on the data recorded during the previous week. The information in this case becomes even more valuable as the time for possible intervention is longer. For this step analysis, since each record is linked to the previous one by a time relation, it is not possible to perform a random training-testing set division, but it is necessary to consider the time. For this reason, the neural network was trained on the first 784 records and tested on the last 784.

The training set spans the period from 7 January 2001 to 4 March 2003 and the testing set from 5 March 2003 to 29 April 2005. The training set includes anomalies in the monitoring, better interpreted as precursors and linked to the early stages of the July–August 2001 and October 2002–January 2003 flank eruptions at Mt. Etna. The testing set covers a period during which no anomalies are present in the relative monitoring dataset despite the September 2004–March 2005 flank eruption occurrence. The dataset analyzed within the first experiment had a total of 44,240 values (i.e., $N = 28$ variables multiplied by 1580 records) to be processed. When submitted to the auto-contractive map (Auto-CM) algorithm for the learning session, the dataset was structured so that the variables were the hyperpoints and the records were the hyperpoint coordinates. This is because we are more interested in understanding how the variables are related to each other in terms of non-linear correlation rather than records.

After 24,625 epochs (i.e., a classic measure of ANNs learning phase: the number of times the ANN examines the whole dataset), the K-CM, with a contraction parameter $C = \sqrt{N} = \sqrt{28} = 5.292$, is fully trained (RMSE = 0.1322). The $C$ value is the only parameter in the Auto-CM neural network to be set before learning. It tunes the transmission of the input into the hidden layer by compressing it with respect to a factor that depends on $C$ itself and on the values of the connections. Therefore, the value of the parameter $C$ controls the actual extent of the contraction, thereby explaining its interpretation as the contraction parameter (Buscema et al. 2018) from which the name of the network derives. Experience shows that the choice of $C = \sqrt{N}$ is often the optimal one (Buscema et al. 2018).

Once the learning was finished, the $W$ weight matrix was determined. As explained in more detail in Appendix 1, the matrix $N \times N$ of the weights contains, for each entry, the value $s_{ij}$ of similarity between variable $i$ and variable $j$. Similarity denotes a measure of association between two values or how much two variables tend to be present together (Fig. 3). Similarity and distance are two closely related concepts because the more two variables are similar, the less distant they will be if thought as hyperpoints of a metric hyperspace. Conversely, the less similar they are, the more distant they will be. It is therefore possible to calculate the Maximum Regular Graph (*MRG*) graph (see Appendix 1) to better understand the hidden relations between the variables and to perform fuzzy profiling of each record for prediction operations. As for the graph, all the entries (weights) are scaled with respect to the theoretic maximum value (i.e., the contraction parameter $C$) into the range [0, 1] and all the similarity values are translated in terms of distances. This technical step is necessary because the algorithms used are based on the minimization of distances instead of the maximization of similarities even though they are conceptually the same feature. The classic method to obtain such a transformation is to consider the complement to one of the similarity values (i.e., $d_{ij} = 1 - s_{ij}$, so that high $s_{ij}$ values determine low $d_{ij}$ values and vice versa). The $d$ value is therefore a distance index providing a value in the range [0, 1] of how far two variables are from each other. We point out that we are considering a distance index and not

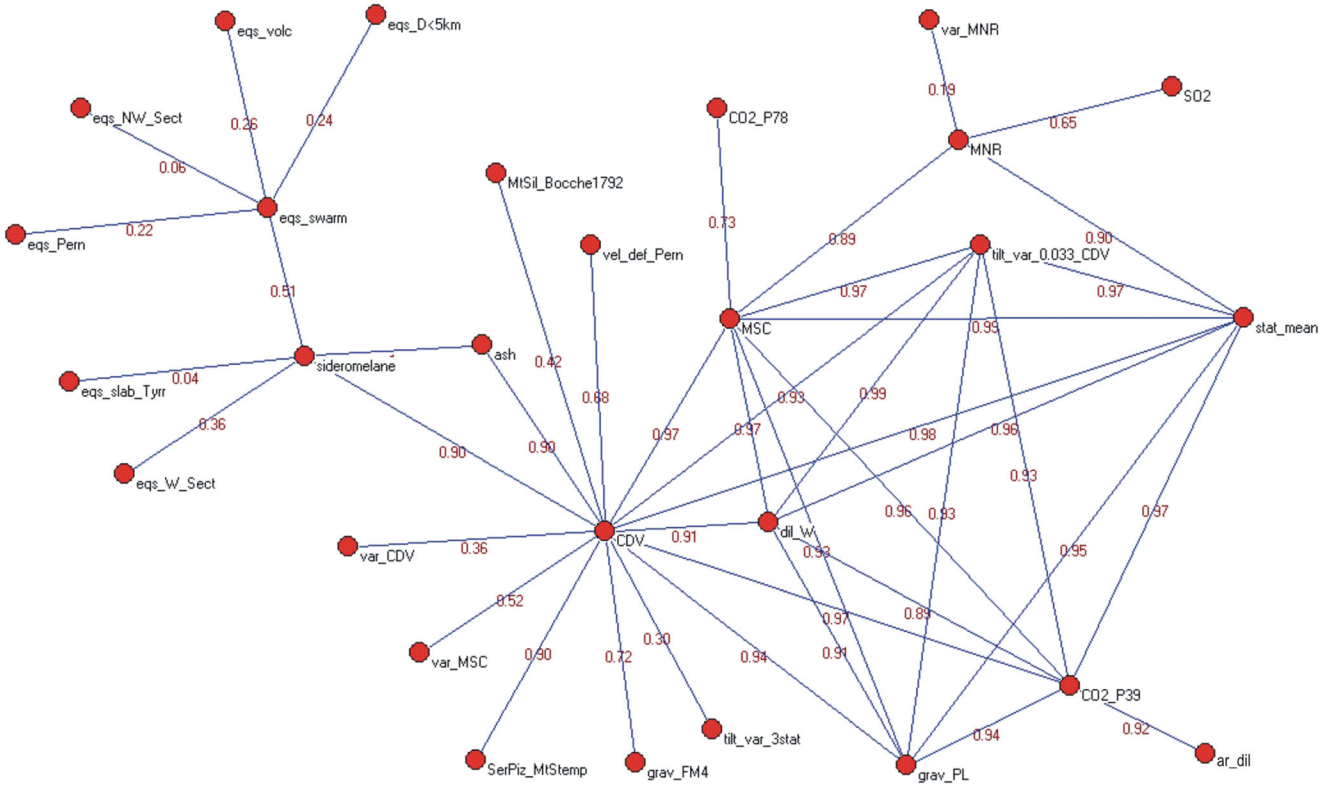


**Fig. 3** The semantic map of the 28 variables (for a relative legend, see Table 1) of the dataset that synthesizes their strongest similarity relationships providing important information about the relevance of the variables and how they relate to each other. It is possible to highlight a group made up of all the variables of seismological type, connected with the sideromelane node and weakly integrated with the rest of the variables. Moreover, given its location and number of connections, the CDV node seems to play a key role for the whole system

distance in the strict metric sense because some of the necessary properties cannot be verified. Then, the highest values among the variables of the weight matrix entries are considered in the *MRG* undirected graph construction. It should be noted that in the construction of the Minimum Spanning Tree (*MST*) (see Appendix 1), and therefore of the *MRG*, the aim is to minimize distances. Minimizing distances implies maximizing similarities, so we can assume that nearby or very connected nodes have much in common while distant and not directly connected ones are less well related. Figure 4 shows how non-linear correlation is estimated, with values even higher than 0.90, mainly associated to *CDV* variability. As a matter of fact, some of the values of the connections have rather weak values, despite the highest values being selected (Fig. 3; e.g., eqs_slab_Tyrr − sideromelane = 0.04). Conversely, ground deformation and geochemical variables show values generally close to 0.70, with very few exceptions (Fig. 3).

To understand the meaning of the *MRG* graph, we have used three known graph indices (Table 1 reports all the relative values):

1. The *Association Strength* ($S_i$): This index measures the strength by which each node attracts its neighbors.
2. The *Node Centrality* ($C_i$) measures how close each node is to the center of the graph. The higher the centrality of a node, the more it is a fundamental variable of the dataset represented by the graph.
3. The *Node Degree* ($D_i$) measures how far each node is a mediator (i.e., how many links for each node) among the other nodes.

The *Association Strength* $S_i$ of each of the displayed nodes is estimated by:

$$S_i = \frac{l_i}{\sum_k^{N_i} l_k} \tag{1}$$

where $S_i = 1$ is the baseline (at $i$th node), $l_i$ represents the number of links of the $i$th node, $N_i$ represents the number of nodes to which the $i$th node is directly linked, and $l_k$ stands for the number of links of the neighbors of the $i$th node. Therefore, the index is proportional to the number of its links and to the number of links of the nodes directly connected to it. If $S_i < 1$, a not very significant node is present due to the weak relative force of attraction for close connections, while a stronger node will

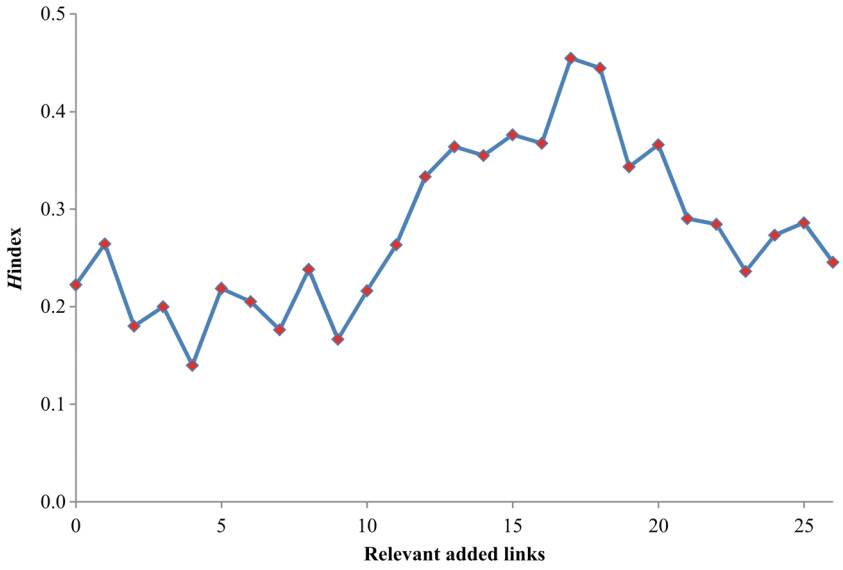


**Fig. 4** The *H* function of the *MRG* graph. The *x*-axis denotes the progressive arc number *k* that is added to the *MST*, and the *y*-axis represents the value of the function *H* (*k*), calculated on this new graph. The value $max_k$ $H(k) = H(17) = 0.45$ implies that the *MRG* will have 17 more connections than the *MST*. The maximum of the function corresponds to the graph of maximum complexity, that is, the one that best describes all the features of the dataset

have $S_i > 1$. It is therefore clear that the *CVD* variable is of considerable importance as characterized by $S_{CVD} = 4.17$ (Table 1). Furthermore, also eqs_swarm, MSC, and sideromelane variables also feature $Si > 1$ (Table 1).

Interpretation of the structure of the graph is issue in understanding complex datasets. One of the ways it simplifies visualization is investigating the relative *H* function (Buscema et al. 2016). This function is a new index to measure topological information (i.e., structural complexity, size measurements, and norms extractable from a graph without any reference to its relative geometry). In other words, the *H* function measures a graph's complexity. This function is calculated from a basic value related to the *MST* graph. Then, one by one, all discarded links are added and the function is recalculated. The *H* function takes on a maximum value when the number of links added determines the maximum complexity.

Figure 5 shows the value of the *H* function as a link is added. The value *H*(0) corresponds to the original *MST*, whereas *H*(*k*) corresponds to the *MST* to which *k* links of those discarded during its construction have been added. In our case, the *H* function reaches its peak when the 17th of the connections skipped during the *MST* generation is added back in. So, the *MRG* needs 17 extra connections to be added to the *MST* and, consequently, the *H* function value is almost 51.00% higher than its value at the original *MST* ($H(0) = 0.22$ vs. *MRG* $H(17) = 0.45$). The *MRG* depicts the most relevant associations among the variables and includes the *MST* representation, being the former characterized by links already reported in the *MST* and by the skipped 17 values. In particular, notice how the *MRG* separates the variables dataset into clear sub-trees whose boundaries are marked by *CDV* and *stat_mean*, respectively (Fig. 3).

## Results

The ground deformation parameter CDV (Table 1) produces a click (Buscema et al. 2016), a completely regular graph where each node is directly linked to any other node, including MSC, tilt_var_0.033_CDV, stat_mean, dil_W (i.e., ground deformation parameters; Table 1), CO2_P39 (i.e., geochemical parameter; Table 1), and grav_PL (i.e., gravimetric parameter; Table 1), among which the highest number of relationships is featured (six, shown in Fig. 3), as well the highest association values (higher than 0.95, on average). The minimum (0.89) and maximum (0.99) association values of the complete regular graph are referred to as dil_W versus $CO_2$_P39 and stat_mean versus MSC and dil_W versus tilt_var_0.033_CDV variables, respectively (Fig. 3). The variables connected through the complete graph show an *Association Strength* $S_i > 0.73$ (Table 1).

The CDV variable also marks a connection giving the association value of 0.90 with ash and sideromelane variables, both exclusively with a volcanic fingerprint (Fig. 3). Furthermore, ash versus sideromelane provides the highest correlation of 1.00 (Fig. 3), indicating that was produced by volatile-rich fresh magma and occurrence of explosive eruption (as demonstrated by the presence of ash). The sideromelane variable does not show a significant strength with any other variable (i.e., variables associated to different volcanic features are connected among themselves, thus displaying a kind of non-linear shared quality; Fig. 3).

Other links with weaker clustering strength (i.e., the *Association Strength* $S_i$) are marked throughout the *MRG* (Fig. 3). This means that the variables belonging to the main click of the graph represent the outline of the dataset, but the link between ash and sideromelane remains the main structural connection among data. As for the first step of pattern

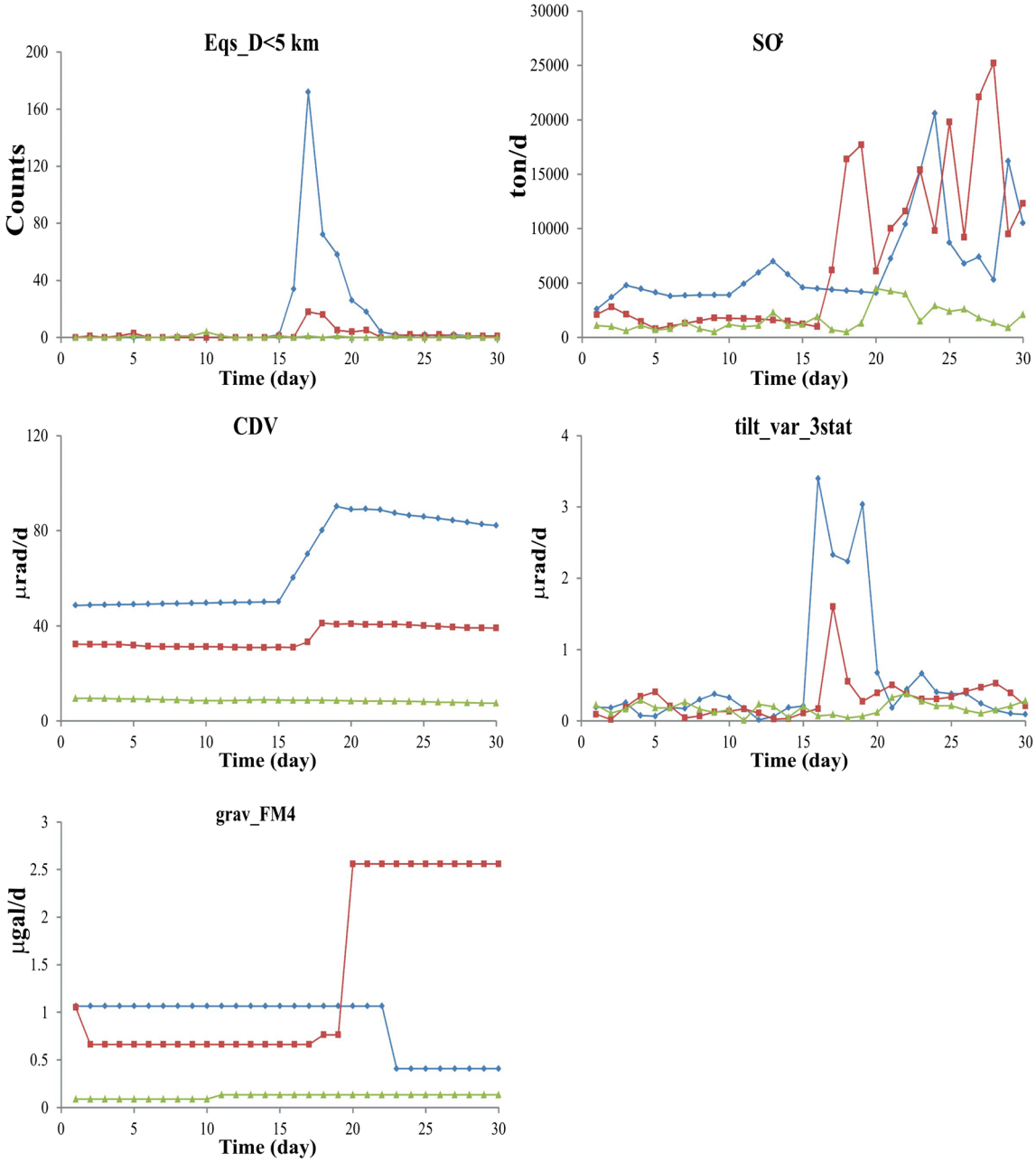


**Fig. 5** Selection of the monitoring values collected during different periods (see Table 1 for relative parameters). Each period spans 30 days of Mt. Etna volcanic activity. In particular, each time series starts 20 days before the onset of the July–August 2001 (blue line), October 2002–January 2003 (red line), and September 2004–March 2005 (green line) flank eruptions. In each panel, the *x*-axis represents time, with value $x = 20$ indicating the onset day of each eruption, while the *y*-axis indicates the value recorded in the considered variables

recognition task, each record must be characterized by a class (namely, a target), which the network must learn to recognize. The target assigned to each of the 1580 records specifies the state of the volcano on a certain day (i.e., a flank eruption at Mt. Etna would or would not occur on each day). The results obtained by the K-CM classifier application were validated using a LOO testing protocol (Appendix 1). Table 2 summarizes the relative calculations, mainly in terms of sensitivity (how many times the algorithm has correctly classified the *erupting* target) and specificity (how many times the algorithm has correctly classified the *quiescent* target). Following Brancato et al. (2016),

$$\text{Sensitivity} = \text{TP}/(\text{TP} + \text{FN}) \quad (2)$$

$$\text{Specificity} = \text{TN}/(\text{TN} + \text{FP}) \quad (3)$$

where TP, TN, FP, and FN stand for true positive, true negative, false positive, and false negative, respectively.

**Table 2** K-CM pattern recognition results, following the LOO validation protocol

| | Eruptive days (day) | Quiescent days (day) | Total days (row; day) | Error (day)[a] | Sensitivity[b] | Specificity[c] | A. mean[d] | W. mean[e] |
|---|---|---|---|---|---|---|---|---|
| Eruptive days | 294 | 7 | 301 | 7 | 97.67% | | 98.49% | 98.99% |
| Quiescent days | 9 | 1270 | 1279 | 9 | | 99.30% | | |
| Total days (column; day) | 303 | 1277 | 1580 | 16 | | | | |

[a] Average number of wrong classification

[b] Classified eruptive days

[c] Classified quiescent days

[d] Arithmetic mean (i.e. accuracy of the classification)

[e] Weighted mean

When considering *erupting*, a sensitivity of 97.67% (294 days out of 301 eruptive days in the investigated period) is estimated (Table 2). When considering the *quiescent* state, a specificity of 99.30% (1270 days out of 1279 non-eruptive days) is estimated (Table 2). The accuracy of classification is equal to 98.49%, and only 16 wrong classifications (seven and nine, respectively; Table 2) are observed. This means that, given the values recorded today, the network can predict with an accuracy of 98.49% if tomorrow there will be an eruption tomorrow or not. Having a high sensitivity value is essential in cases where the two classes are not perfectly balanced. In fact, this parameter takes account of the errors in relation to the number of elements in that class.

For the second step of this analysis, we estimated probabilities for each target (eruption or rest) for the seventh day, after an application of a mobile window of 6 days. Table 3 reports the results in terms of sensitivity and specificity for the last 787 days of the dataset (see above) and Table 4 compares the relative targets with the classified days. We should emphasize that this sample mostly focuses on the 183 eruptive days of the September 2004–March 2005 flank eruption at Mt. Etna. The results relating to the July–August 2001 and October 2002–January 2003 eruptions have to be used as an added value that completes the analysis in a more global perspective.

When looking at the probabilities of the next day, the most obvious result is the classification of the September 2004–March 2005 period, as highlighted in Table 4. Reported values show how the sample of each day is classified according to one of the two labels *erupting* or *quiescent*. The ANN classifies each day without knowing the real class to which it belongs. Most of the 183 eruptive days are classified as *erupting*, with the minimum of probability 14.00% occurring after the end of the eruption (Table 4). The onset of the eruption is classified as 10 September 2004 with a probability of 57.00%. Brancato et al. (2011) showed that the prediction of the September 2004–March 2005 flank activity failed because no anomalies were detected in any monitored parameters. The pattern recognition classified the restricted period as *erupting* with a relative sensitivity of 97.81% (179 out of 183 eruptive days; Table 3) and as *quiescent* with a specificity of 99.50% (601 records out 604 rest days). This high accuracy of classification (i.e., 98.66%) is reflected by only seven misclassifications (four and three, for *erupting* and *quiescent*, respectively).

These results are in line with those obtained after the first step (Tables 3 and 4). It is noteworthy that during the training phase of the Auto-CM algorithm, the monitoring dataset referring to the analyzed period can be regarded as anomalous,

**Table 3** K-CM pattern recognition results of the last 787 records of the dataset, after the application of a mobile window of 6 days (see text for more details)

| | Eruptive days (day) | Quiescent days (day) | Total days (row; day) | Error (day)[a] | Sensitivity[b] | Specificity[c] | A. mean[d] | W. mean[e] |
|---|---|---|---|---|---|---|---|---|
| Eruptive days | 179 | 4 | 183 | 4 | 97.81% | | 98.66% | 99.11% |
| Quiescent days | 3 | 601 | 604 | 3 | | 99.50% | | |
| Total days (column; day) | 182 | 605 | 787 | 7 | | | | |

[a] Average number of wrong classification

[b] Classified eruptive days

[c] Classified quiescent days

[d] Arithmetic mean (i.e., accuracy of the classification)

[e] Weighted mean

**Table 4** Selected comparison of eruptive activity occurring at Mt. Etna in the January 2001–April 2005 period among real eruption and real rest targeted (second and third, seventh and eighth columns) and classified (probability; fourth and fifth, nineth and tenth columns) days

| Date | Classes | | K-CM Output | | Date | Classes | | K-CM output | |
|---|---|---|---|---|---|---|---|---|---|
| | target_1 | target_2 | target_1 | target_2 | | target_1 | target_2 | target_1 | target_2 |
| 1 January 2001 | 0 | 1 | 0 | 1 | 24 November 2004 | 1 | 0 | 1 | 0 |
| ... | ... | ... | ... | ... | 25 November 2004 | 1 | 0 | 1 | 0 |
| 12 July 2001 | 0 | 1 | 0 | 1 | ... | ... | ... | ... | ... |
| 13 July 2001 | 0 | 1 | 0 | 1 | 21 January 2005 | 1 | 0 | 1 | 0 |
| 14 July 2001 | 0 | 1 | 0 | 1 | 22 January 2005 | 1 | 0 | 0.86 | 0.14 |
| 15 July 2001 | 0 | 1 | 0 | 1 | 23 January 2005 | 1 | 0 | 1 | 0 |
| 16 July 2001 | 0 | 1 | 1 | 0 | ... | ... | ... | ... | ... |
| ***17 July 2001*** | ***1*** | ***0*** | ***0*** | ***1*** | 29 January 2005 | 1 | 0 | 0.86 | 0.14 |
| 18 July 2001 | 1 | 0 | 1 | 0 | 30 January 2005 | 1 | 0 | 0.86 | 0.14 |
| 19 July 2001 | 1 | 0 | 1 | 0 | 31 January 2005 | 1 | 0 | 1 | 0 |
| 8 August 2001 | 1 | 0 | 1 | 0 | 6 February 2005 | 1 | 0 | 1 | 0 |
| ***9 August 2001*** | ***1*** | ***0*** | ***0*** | ***1*** | 7 February 2005 | 1 | 0 | 0.86 | 0.14 |
| 10 August 2001 | 0 | 1 | 1 | 0 | 8 February 2005 | 1 | 0 | 1 | 0 |
| 11 August 2001 | 0 | 1 | 0 | 1 | ... | ... | ... | ... | ... |
| ... | ... | ... | ... | ... | 7 March 2005 | 1 | 0 | 1 | 0 |
| 25 October 2002 | 0 | 1 | 0 | 1 | ***8 March 2005*** | ***0*** | ***1*** | ***0.86*** | ***0.14*** |
| ***26 October 2002*** | ***0*** | ***1*** | ***0*** | ***1*** | 9 March 2005 | 0 | 1 | 0.86 | 0.14 |
| 27 October 2002 | 1 | 0 | 0 | 1 | 10 March 2005 | 0 | 1 | 0.57 | 0.43 |
| 28 October 2002 | 1 | 0 | 1 | 0 | 11 March 2005 | 0 | 1 | 0 | 1 |
| 29 October 2002 | 1 | 0 | 0 | 1 | 12 March 2005 | 0 | 1 | 0 | 1 |
| 30 October 2002 | 1 | 0 | 1 | 0 | ... | ... | ... | ... | ... |
| 31 October 2002 | 1 | 0 | 1 | 0 | 14 April 2005 | 0 | 1 | 0 | 1 |
| ... | ... | ... | ... | ... | 15 April 2005 | 0 | 1 | 0.29 | 0.71 |
| 26 January 2003 | 1 | 0 | 0 | 1 | 16 April 2005 | 0 | 1 | 0.43 | 0.57 |
| 27 January 2003 | 1 | 0 | 1 | 0 | 17 April 2005 | 0 | 1 | 0.43 | 0.57 |
| ***28 January 2003*** | ***1*** | ***0*** | ***0*** | ***1*** | 18 April 2005 | 0 | 1 | 0.29 | 0.71 |
| 29 January 2003 | 0 | 1 | 0 | 1 | 19 April 2005 | 0 | 1 | 0.14 | 0.86 |
| 30 January 2003 | 0 | 1 | 0 | 1 | 20 April 2005 | 0 | 1 | 0 | 1 |
| ... | ... | ... | ... | ... | 21 April 2005 | 0 | 1 | 0 | 1 |
| 4 September 2004 | 0 | 1 | 0 | 1 | ... | ... | ... | ... | ... |
| 5 September 2004 | 0 | 1 | 0 | 1 | ... | ... | ... | ... | ... |
| ***6 September 2004*** | ***1*** | ***0*** | ***0*** | ***1*** | ... | ... | ... | ... | ... |
| 7 September 2004 | 1 | 0 | 0 | 1 | ... | ... | ... | ... | ... |
| 8 September 2004 | 1 | 0 | 0 | 1 | ... | ... | ... | ... | ... |
| 9 September 2004 | 1 | 0 | 0 | 1 | ... | ... | ... | ... | ... |
| 10 September 2004 | 1 | 0 | 0.57 | 0.43 | ... | ... | ... | ... | ... |
| 11 September 2004 | 1 | 0 | 1 | 0 | ... | ... | ... | ... | ... |
| 12 September 2004 | 1 | 0 | 1 | 0 | ... | ... | ... | ... | ... |
| 22 November 2004 | 1 | 0 | 1 | 0 | ... | ... | ... | ... | ... |
| 23 November 2004 | 1 | 0 | 0.86 | 0.14 | ... | ... | ... | ... | ... |

The onset and the end of the flank activities (see text for details) during the investigated time period are marked in bold and italics, respectively

with comparable values in line with the preceding flank eruption occurrences (namely, July–August 2001 and October 2002–January 2003). The onset day of the September 2004–March 2005 flank eruption (namely, 6 September 2004) and the following 3 days were classified as *quiescent* (Table 4). Only on 10 September 2004 did the K-CM algorithm classify it as *erupting* (Table 4). At the end of April (the eruption ended on 8 March 2005), misclassifications occur, classifying the

state of the volcano as *erupting* (Table 4). Flank activity at Mt. Etna has been observed to be usually preceded with changes in the volcano feeder system (Bonaccorso et al. 2004; Brancato et al. 2011). Figure 5 compares three periods, each of 30 days (i.e., monitoring values), relative to the flank activity here analyzed.

It is clear how monitoring parameters reflect the magma ascent a few days before the flank eruptions of the 2001–2003 period and, conversely, how no monitoring anomalies (i.e., precursors) were recorded before the 2004–2005 volcanic activity. In the first case, in fact, there are many sudden changes in the values recorded before the eruption, while in the second case, the observed variables maintained a constant value before, during, and after the eruption, making the event more difficult to predict.

It is noteworthy that K-CM classifies most of the 2004–2005 flank eruption as *erupting* (i.e., undergoing eruption) (Table 4), even though the classified onset is 5 days after the real onset (with a probability of 57.00%; Table 4) and the classified end 3 days after the real end (Table 4). This is the only relevant error in the K-CM prediction out of the 1580-day period. In this specific case, the network, trained on 6 days to predict the seventh, needed more time to provide the correct output. Since the sample used in training had only two eruptive flank episodes, the number of cases in which a change of state at Mt. Etna is present is rather small. It therefore seems reasonable to believe that with a larger monitoring dataset, the network could refine its ability to more effectively predict when the state changes. Furthermore, misclassifications occur at the end of April 2005, (Table 4). These are better interpreted as false positives (i.e., aborted eruptions), due to anomalous seismicity (S. Alparone, personal communication, 2010) and gas emissions (Brancato et al. 2011). These relate to the closing phases of a flank eruption at Mt. Etna. Relative confusion is generated because of a diffuse absence of anomalous values in the monitoring dataset. In some cases, this means the classified days have more uncertainty, the worst case (for decision makers) being when the ANN anticipates the end.

As further support for our study, we analyzed the behavior of the Receiver Operating Characteristic (ROC) curve (Hanley and McNeil 1982; Fig. 6) and Molchan diagram (Zhuang 2010; Schmid et al. 2012; de Arcangelis et al. 2016; Fig. 7). By considering that mistakes are likely in classifying cases, there will be some target patterns correctly classified as targets (i.e., TP) and some misclassified as non-targets (i.e., FN). Similarly, there will be non-target patterns correctly classified as non-targets (i.e., TN) and some misclassified as targets (i.e., FP).

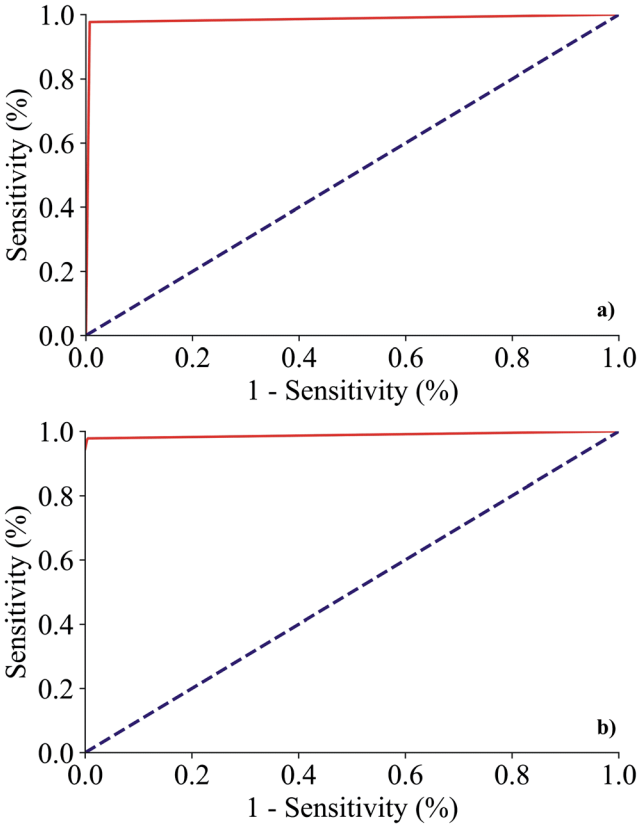


**Fig. 6** ROC curve (red solid line) for the first experiment (**a**) and second experiment (**b**), respectively. The dashed line means an AUC equals to 0.5 (i.e., TP and FP proportions are equal). Note that even for high specificity values, i.e., around zero on the *x*-axes, the sensitivity is very high. The value of the AUC is in fact close to 1

The solid red line in Fig. 6 corresponds to the ROC curve, relative to the trade-off between the probability of correctly classifying the erupting class (true positive rate, also known as sensitivity) and the false positive rate (1 – specificity) concerning, respectively, the first and second trials carried out. The Area Under the Curve (AUC), if compared with the area that lies under the gray dashed line (i.e., 0.5; no discrimination exists, which indicates a worthless test where the proportions of TP and FP are equal), provides a valid measure of diagnostic performance of the used classifier. In both cases, the area under the ROC plot, computed by means of the trapezoid rule (Tallarida and Murray 1987), has an extremely high value. In the best case, corresponding to the second test (Fig. 6b), it is equal to 0.9945, meaning that 99.45% of the classified erupting days are correctly rated for a given probability of false alarm. Following the Swets classification (Swets 1988), this value confirms that the diagnostic used is highly accurate.

The Molchan diagram features the error. It is drawn by considering values depending on the quantities TP, TN, FP, and FN, previously defined. The graph represents the

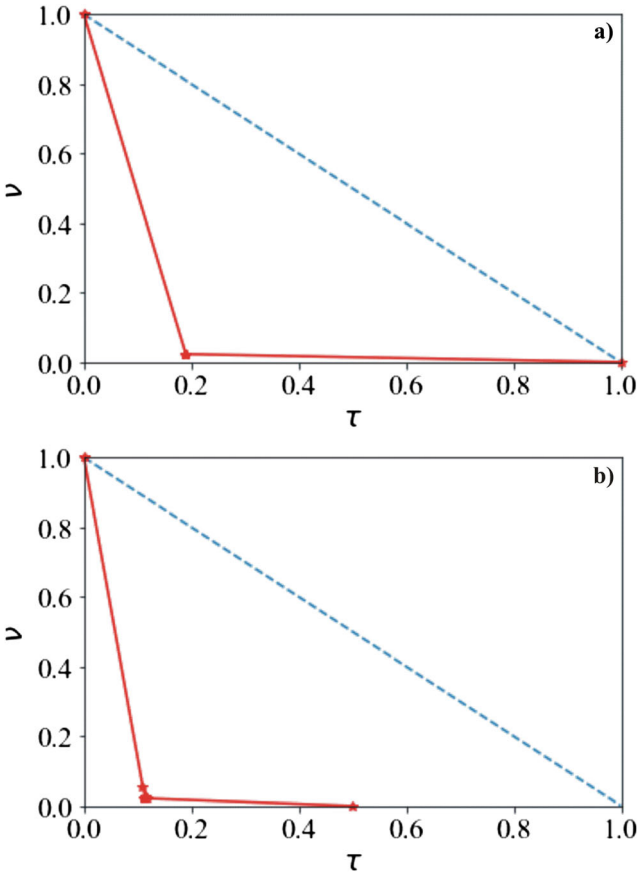


**Fig. 7** Molchan diagrams for the first (**a**) and the second experiment (**b**), respectively. The dashed line means a random guess (i.e., the curve of model equivalence; de Arcangelis et al. 2016). The red solid line shows the confidence level

pairs ($\tau$, $\nu$) obtained by varying a threshold value $\lambda$. The quantities $\tau$ and $\nu$ are defined as follows:

$$\nu = \mathrm{FN}/(\mathrm{TP} + \mathrm{FN}) \quad (4)$$

$$\tau = (\mathrm{TP} + \mathrm{FP})/(\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}) \quad (5)$$

where $\nu$ considers the amount of missed event and $\tau$ corresponds to the alarm time fraction (time covered by the alarms over the duration of the whole considered period). In our experimentation, the threshold value $\lambda$ was varied over the range [0, 1] with steps of 0.1 (Fig. 7). In the Molchan diagram, the better the forecast, the closer you get to the theoretical point (0.0). In our case, in both experiments, we obtained a Molchan curve that shows excellent network reliability. When the threshold changes, the points all fall into an area very close to the theoretical minimum. It should be noted that the output values of the K-CM network in the first experiment ($N = 28$) are binary (i.e., 0 or 1), while, in the case of the second experiment, the output values are continuous (i.e., they vary in the range [0, 1]). For this reason, both the ROC curve and the Molchan diagram relevant to the first test do not show particular variations when the threshold changes, maintaining constant and very good performances (Figs. 6 and 7). In the Molchan diagram, in particular, all the points are superimposed on a single pair of values $(\overline{\tau}, \overline{\upsilon})$, with the exception of the borderline cases (0, 1) and (1, 0).

## Discussion

Analyses of pattern classification (both supervised and unsupervised) have previously been performed at Mt. Etna by Langer et al. (2009) and Corsaro et al. (2013). In particular, the artificial neural networks were applied to volcanic tremor data recorded between July and August 2001 (Supporting Vector Machine (SVM) and Multi-Layer perceptron (MLP); Langer et al. 2009) and to geochemical data collected from 1995 to 2005 (Self-Organizing Map (SOM) and fuzzy clustering; Corsaro et al. 2013). The analyzed periods cover both summit eruptive activity (lava fountains) and flank eruptions. Even though both datasets represent a much reduced sample of that used here, the authors conclude that classification methodology is a suitable tool for understanding the non-linear relationships between volcanic data and the internal dynamics of a volcano. In particular, Langer et al. (2009), by applying the LOO testing scheme as used in this study, yielded high performances of 94.80% and 81.90% (for SVM and MLP, respectively), though with a slightly lower score relative to our value of 97.67% (Table 2). In addition, Corsaro et al. (2013) proposed that classifications performed by SOM and fuzzy clustering are extremely useful for an accurate interpretation of magma composition, its origin, and behavior during future eruptions at Mt. Etna, thus allowing a direct comparison between old and new erupted products.

This study applies and presents the theory behind an emerging supervised method for volcanological pattern recognition, named the K-CM, as applied to detecting the change from "in a flank eruption" (i.e., *erupting*) to the "not in a flank eruption" state (i.e., *quiescent*) at Mt. Etna. A two-step analysis was performed. First, the applied methodology exploits the variable connection weights provided by the Auto-CM neural network strategy to obtain the $z$-transforms (see Appendix 1) on which the kNN classifier is applied for class membership evaluation. The Auto-CM system reshapes the distances between variables or records in any dataset, considering their global vectorial similarities, and consequently drawing out the specific *warped space* (i.e., a non-Euclidean space) in which such variables or records lie, thereby providing a proper theoretical representation of their actual variability. We have shown how a known filter like the *MST* can be used to cluster a distance matrix, generated by means of Auto-CM from a dataset. In particular, how the *MST* can be regarded

as the minimal representation that takes into account the basic level of information below which the structure of the independence among variables loses coherence. We have used the *H* index, a function that evaluates the topological complexity of any kind of undirected graph, its mathematical consistency, and the potential for its applications (Fig. 4 and Table 1). Finally, by using the *H* function, we have represented our results on the *MRG* semantic graph. From an *MST*, generated from any metric, the *MRG* reshapes the links between nodes in order to maximize the fundamental and the most regular structures implicated in any dataset.

After analyzing the *MRG* graph, very robust relationships were found among variables (Fig. 3). In detail, the role of the *CDV* parameter (which monitors ground deformation) is predominant, around which a click is created featuring the highest number of connections with different parameters of the volcanic framework (Fig. 3). Further, an immediate link with *ash* and *sideromelane* variables generates one of the highest non-linear correlations (i.e., 0.90; Fig. 3).

Indeed, the most outstanding novelty concerns the September 2004–March 2005 flank eruption (Table 4). Brancato et al. (2011) failed to predict this activity because of the total absence of anomalies (due to fuzzy parameters) in the monitoring dataset. In contrast, the approach used in this study performs a high accuracy evaluation, with only a very few exceptions (only seven wrong classifications out of a total of 787 days; Table 3).

Brancato et al. (2011) performed a Bayesian Event Tree (BET) approach on the same time-interval analyzed here. A larger dataset, including also the volcanic tremor parameter (not used here because unavailable for the entire period), allowed the probabilistic volcanic hazard to be quantified, in terms of occurrence of flank eruptions. The results showed the major role of monitoring in forecasting short-time flank activity at Mt. Etna. It is worth noting that probability estimates of an impending flank eruption were affected by anomalous values in $CO_2$ emission at P39 site. Indeed, the presence of relative anomalies generated a sequence of long-lasting false alarms. Thus, Brancato et al. (2011) suggested that $CO_2$ emission at P39 site was incorrectly elicited by experts. Conversely, the K-CM approach points to the fundamental role of the parameter in the dataset, based on the *Association Strength* $S_i$ value of 1 and the highest $C_i$ value of 4 (Table 1), thus avoiding the relative confusion about the predicting aim of the parameter after the BET approach. In other words, even though the expert elicitation set the $CO_2$ emission at P39 site as one of the potential parameter to be used for predicting flank eruptions at Mt. Etna, the BET approach rose to the question about its reliability. The present approach highlights that a parameter, as discarded by previous approaches, might be a significant loss in terms of hidden links as better associated to the internal volcanic processes.

## Conclusion

This study highlights the performance of the modified K-CM classifier, applied for the first time to a volcanic context, using Mt. Etna as case study. We presented a new method to interpret the results of data represented on the *MRG* graph and to forecast flank eruptive activity. Improvements are mainly due to a many-to-many variable relationship. Given the complex nature of the phenomenon, in fact, we cannot be satisfied with the simple relationship between two or three variables, but we must investigate how a multiplicity of variables (many), interacting with others (to many), determines the phenomenon. Only by this approach, it will be possible to obtain a reliable predictive model. A lone parameter may not be anomalous (based on the mean variance), but the entire dataset of parameters, when defining the same non-linear function, can contribute incrementally to reveal deep global effects (the so-called meta-stable variables, whose behavior is strongly affected by the values of the other variables). Indeed, by comparing results by the K-CM from those obtained by the Bayesian approach (Brancato et al. 2011), the predominance of the algorithms of automatic learning becomes clear. This is due to the ability of the machine learning techniques to identify objectively the subtle links between the monitoring data and the state of Mt. Etna (quiescent or eruptive), thus avoiding the subjectivity of external operators, primarily in selecting thresholds and inertias for the monitoring parameters involved.

K-CM opens the possibility for bottom-up algorithms to provide a symbolic explanation of their paths and behavior. The symbolic level of the K-CM system is not, however, a set of naïve "if…then" rules traditionally applied to all training data. Understanding this symbolic level is relatively simple: the increase of a non-linear functional, from interpolation of the training set, results in the exponential growth of explicit rules suitable for deriving a description. A more interesting symbolic level is that which allows the explanation of the "fuzzy mental map" through which the learning algorithm is represented based on training data and, simultaneously, the subjective similarities on which the same algorithm can operate in new cases. In other words, a bottom-up algorithm can work in a symbolic way when it is able to self-generate a mental representation (weighted graph) of what it has learned (from the training set) and dynamically place it in a chart of new experiences (the test set), by inducing it to reorganize the initial map.

In other comparative studies (Buscema et al. 2014), K-CM showed the best valid classification performance for most of the considered datasets and, on average, out-performed the other classification methods. In practice, K-CM can improve the classification results especially in those cases where non-

linear relationships among variables are relevant. We are aware that this new approach, although very promising, leaves many questions unanswered and needs further scrutiny to investigate its properties and potential shortcomings. In forthcoming research, we plan to develop it from both the theoretical and the empirical angles, by better exploring and characterizing the structural features of the Auto-CM and the mathematical properties of the *H* function and the *MRG*. On the empirical side, we plan to apply this methodology to state-of-the art problems that are relevant in the literature of specific disciplines. It is, of course, of particular interest to evaluate how our approach performs against traditional methods in the analysis of networks, be they of a physical or social nature.

All the connections among variables (in particular, the click generated by *CDV*; Fig. 3) indicate that monitoring is fundamental for quantifying short-term volcanic hazard. In the same way, it is important to bear in mind that the pattern recognition approach improves the sensitivity, specificity, and accuracy estimated in previous applications of the ANN methodology (Brancato et al. 2016), thus confirming the high non-linear behavior of an open conduit volcanic complex system, such as Mt. Etna.

In light of this, this study represents one of the few attempts in processing long-time records of monitoring data collected at an active volcano by artificial neural network methodology. Results may offer a tool when decision makers face challenges in volcanic emergencies in terms of planning and/or mitigation of potential effects in areas prone to volcanic risk, mostly due to a robust near real-time estimate. Furthermore, results achieved here show how a supervised K-CM approach performs a reliable data mining when applied to a high-quality monitoring dataset. This might be a limiting factor in a poorly monitored volcano.

**Acknowledgments** The authors are deeply grateful to Alessandro Bonforte, G. Di Grazia, F. Greco, and S. Gambino of the INGV-OE for having provided data, beliefs, and theoretical models. The authors are also grateful to the Technical Staff of the INGV-OE for maintaining the monitoring networks. Semeion Research Center of Sciences of Communication is also thanked for providing the specialist ANN software for all the elaborations. Sylvie Vergniolle is sincerely acknowledged for the continuous encouragement and the constructive criticism. Andrew Harris is acknowledged as well for suggestions that deeply improved the final version of the paper. The authors wish to thank two anonymous reviewers for the careful revision of earlier versions of the manuscript. AB benefited from funding provided by GEOMONITOR Project. Semeion Research Center of Sciences of Communication funded PMB, GM, and FDT for this research. The authors kindly acknowledge Stephen Conway for the initial language revision of the manuscript.

## Appendix 1

### K-CM method overview

K-CM is a supervised version of Auto-CM neural networks (Buscema et al. 2014, 2018). K-CM has a training phase in a completely unsupervised way and then a blind testing validation, using the kNN algorithm (Hastie et al. 2009) to estimate to which of *N* classes each tested record belongs. Furthermore, by using the parameters derived from the Auto-CM learning, it is possible to obtain a graph that reflects the non-linear correlation values assumed by the neural network. The whole process is summarized in Fig. 8.

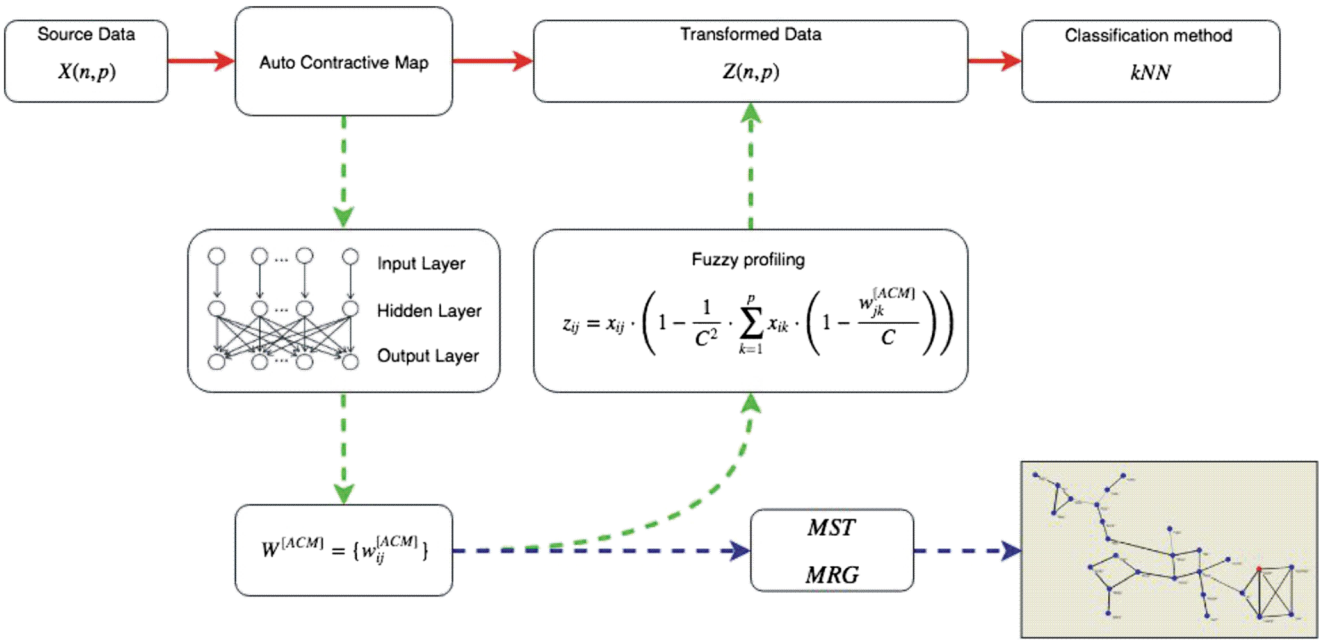


**Fig. 8** Flow chart of all the analyses carried out with K-CM

The entire dataset is initially used as input for an unsupervised neural network (i.e., Auto-CM), which synthesizes the non-linear correlation relationships between the variables in the $W$ connection matrix of its weights, as identified in the data. This matrix represents the Auto-CM knowledge about the whole dataset and can be used in two different ways: to rewrite the dataset through the function called $z$-transform or as a matrix of the weights to be linked to the graph whose nodes are the variables. The $z$-transformation allows the fuzzy profile of each record to be determined, taking into account the information about the relationships between the variables which emerged during the learning phase. The rewritten dataset is therefore more informative than the previous one. The kNN classifier is used on the new dataset to perform the pattern recognition task. The dataset is randomly subdivided into two subsets A and B. The subset A is trained and the subset B is blind tested, but the algorithm also works in reverse, with the subset B trained and the subset A subsequently blind tested. The average of these two tests represents the accuracy of the algorithm.

## The semantic map: data graphical visualization

The K-CM classifier is able to support its classification task projecting each tested record as a node of a weighted graph, where the more similar records are closer to each other. In other words, the resulting graph is a map of similarities among the predicted records and this map is very useful in understanding which records represent a general prototype of the whole dataset and which ones are outliers. On the other hand, Auto-CM is an ANN architecture that forces the non-linear correlation among variables into an embedding space where the relative associations are accurately reflected as closeness. Because of the Auto-CM architecture, the weight matrix has a dimension ($N \times N$), where $N$ is the number of variables. It is then possible to consider the graph $G$ ($n$, $E$), where $n$ is the set composed of all the $N$ variables (nodes of the graph) and $E$ is the set containing all the arcs (i.e., the couple ($i$, $j$) if the variables $i$ and $j$ are linked in the graph). All nodes are connected to each other and the weight of the connection between nodes $i$ and $j$ depends on the value of the $w_{ij}$ weight of the $W$ matrix ($w_{ij}$ values express similarities; therefore, it is necessary to make a transformation to translate them into terms of distance).

Since it is not easy to read a graph where all nodes are connected, filters are used to display only the most relevant connections, in particular the *MST* and the *MRG*. These two graphs work as an effective filter on the dataset correlation, drastically reducing the number of connections shown, and both of them have already shown promising applications in many fields (Buscema et al. 2014, 2016, 2018). While *MST* is a known algorithm for its application to real problems (an undirected graph without circuits or loops; see Mantegna 1999), *MRG* is a new one (see Buscema et al. 2014). The *MST* connects all the $N$ variables to each other, selecting the highest $N - 1$ connections that avoid the creation of cycles. *MRG* adds to *MST* those circuits representing the most important fingerprints of the whole analyzed dataset until the maximum of a complexity function is reached (the $H$ function, see below for details). In particular, the *MRG* reshapes the links between nodes in order to maximize the fundamental and most regular structures, thus highlighting the features of a dataset (Buscema et al. 2018).

Working with classification methods, it is common to deal with the overfitting phenomenon. This arises when the number of degrees of freedom of the classification technique (i.e., its complexity) is large enough to not be constrained in a stable way by the available data. To this end, cross-validation schemes represent a basic tool for the diagnosis of this overfitting. In a validation protocol, part of the patterns with a priori known class membership is set aside and not used during the training phase. These patterns represent the test set, which is used after the training to evaluate the performance of the classifier. If the degrees of freedom are well balanced with respect to the size of the dataset, then the classifier performance for patterns used during training and test phases will be comparable.

The K-CM classification method, used here, is modified according to a new validation protocol, namely the LOO estimator. This represents a particular case of the K-fold cross-validation estimator, when the number of patterns into which the original dataset is subdivided equals the number of cases of the same dataset. The model is then trained with all the patterns of the dataset but one (i.e., that is, the one on which the error is estimated), and the procedure runs for all the patterns. The average of the estimated errors for the entire pattern will provide the average error of the model.

## References

Aloisi M, Bonaccorso A, Gambino S (2006) Imaging composite dike propagation (Etna 2002, case). J Geophys Res 111. https://doi.org/10.1029/2005JB003908

Aloisi M, Bonaccorso A, Cannavò F, Gambino S, Mattia M, Puglisi G, Boschi E (2009) A new dyke intrusion style for the Mount Etna May 2008 eruption modelled through continuous tilt and GPS data. Terra Nova 21:316–321. https://doi.org/10.1111/j.1365-3121.2009.00889.x

Andronico D, Corsaro RA (2011) Lava fountains during the episodic eruption of South–East Crater (Mt. Etna), 2000: insights into magma-gas dynamics within the shallow volcano plumbing system. Bull Volcanol 73:1165–1178. https://doi.org/10.1007/s00445-011-0467-y

Andronico D, Lodato L (2005) Effusive activity at Mount Etna volcano (Italy) during the 20th century: a contribution to volcanic hazard assessment. Nat Hazards 36:407–443. https://doi.org/10.1007/s11069-005-1938-2

Andronico D, Scollo S, Cristaldi A, Ferrari F (2009) Monitoring ash emission episodes at Mt. Etna: the 16 November 2006 case study. J Volcanol Geotherm Res 180:123–134

Arbib MA (1995) The handbook of brain theory and neural networks, a Bradford book. MIT Press, Cambridge

Badalamenti B, Bruno N, Caltabiano F, Di Gangi S, Giammanco G, Salerno G (2004) Continuous soil CO2 and discrete plume SO2 measurements at Mt. Etna (Italy) during 1997-2000: a contribution to volcano monitoring. Bull Volcanol 66:80–89

Ballard DH (1999) An introduction to natural computation. MIT Press, Cambridge

Barberi F, Carapezza ML, Valenza L, Villan L (1992) L'eruzione 1991-1992 dell'Etna e gli interventi per fermare o ritardare l'avanzata della lava. Giardini, Pisa

Barreca G, Bonforte A, Neri M (2012) A pilot GIS database of active faults of Mt. Etna (Sicily): a tool for integrated hazard evaluation. J Volcanol Geotherm Res 251:170–186. https://doi.org/10.1016/j.jvolgeores.2012.08.013

Bengio Y (2009) Learning deep architectures for AI. Mach Learn 2(1):1–127

Bonaccorso A, Calvari S (2013) Major effusive eruptions and recent lava fountains: balance between expected and erupted magma volumes at Etna volcano. Geophys Res Lett 40:6069–6073. https://doi.org/10.1002/2013GL058291

Bonaccorso A, Calvari S, Coltelli M, Del Negro C, Falsaperla S (2004) Mt. Etna: volcano laboratory. American Geophysical Union, Washington DC

Bonaccorso A, Bonforte A, Guglielmino F, Palano M, Puglisi G (2006) Composite ground deformation pattern forerunning the 2004–2005 Mount Etna eruption. J Geophys Res. https://doi.org/10.1029/2005JB004206

Bonforte A, Gambino S, Guglielmino F, Obrizzo F, Palano M, Puglisi G (2007) Ground deformation modeling of flank dynamics prior to the 2002 eruption of Mt. Etna. Bull Volcanol 69:757–768. https://doi.org/10.1007/s00445-006-0106-1

Branca S, Coltelli M, De Beni E, Wijbrans J (2008) Geological evolution of Mount Etna volcano (Italy) from earliest products until the first central volcanism (between 500 and 100 ka ago) inferred from geochronological and stratigraphic data. Int J Earth Sci (Geol Rundsch) 97:135–152

Branca S, Coltelli M, Groppelli G (2011) Geological evolution of a complex basaltic stratovolcano: Mount Etna. Ital J Geosci 130(3):306–317

Brancato A, Gresta S (2003) High precision relocation of microearthquakes at Mt. Etna (1991-1993 eruption onset): a tool for better understanding the volcano seismicity. J Volcanol Geotherm Res 124:219–239

Brancato A, Gresta S, Alparone S, Andronico D, Bonforte A, Caltabiano T, Cocina O, Corsaro RA, Cristofolini R, Di Grazia G, Distefano G, Ferlito C, Gambino S, Giammanco S, Greco F, Marzocchi W, Napoli R, Sandri L, Selva J, Tusa G, Viccaro M (2011) BET_EF application at Mount Etna: a retrospective analysis (years 2001-2005). Ann Geophys 54:642–661

Brancato A, Gresta S, Sandri L, Selva J, Marzocchi W, Alparone S, Andronico D, Bonforte A, Caltabiano T, Cocina O, Corsaro RA, Cristofolini R, Di Grazia G, Distefano G, Ferlito C, Gambino S, Giammanco S, Greco F, Napoli R, Tusa G, Viccaro M (2012) Quantifying probabilities of eruption at a well-monitored active volcano: an application at Mount Etna (Sicily, Italy). Boll Geofis Teor Appl. https://doi.org/10.4430/bgta0040

Brancato A, Buscema PM, Massini G, Gresta S (2016) Pattern recognition for flank eruption forecasting: an application at Mount Etna volcano (Sicily, Italy). Open J Geol 06:583–597. https://doi.org/10.4236/ojg.2016.67046

Buscema M, Tastle WJ (2013) Data mining applications using artificial adaptive systems. Springer, Berlin

Buscema M, Consonni V, Ballabio D, Mauri A, Massini G, Breda M, Todeschini R (2014) K-CM: a new artificial neural network. Application to supervised pattern recognition. Chemom Intell Lab Syst 138:110–119

Buscema M, Asadi-Zeydabadi M, Lodwick W, Breda M (2016) The H0 function, a new index for detecting structural/topological complexity information in undirected graphs. Phys A 447:355–378. https://doi.org/10.1016/j.physa.2015.12.055

Buscema PM, Massini G, Breda M, Lodwick WA, Newman F, Asadi-Zeydabadi M (2018) Artificial adaptive systems using auto contractive maps: theory, applications and extensions. Springer, Berlin

Caltabiano T, Burton M, Giammanco S, Allard P, Bruno N, Murè F, Romano R (2004) Volcanic gas emission from the summit craters and flanks of Mt. Etna, 1987–2000. In: Calvari S, Bonaccorso A, Coltelli M, Del Negro C, Falsaperla S (eds) Mt. Etna volcano laboratory. Geophysics monography series, vol 143, pp 111–128

Calvari S, Pinkerton H (1998) Formation of lava tubes and extensive flow field during the 1991–1993 eruption of Mount Etna. J Geophys Res 103:27291–27301. https://doi.org/10.1029/97JB03388

Carbone D, Greco F (2007) Review of microgravity observations at Mt. Etna: a powerful tool to monitor and study active volcanoes. Pure Appl 164:769–790. https://doi.org/10.1007/s00024-007-0194-7

Carbone D, Budetta G, Greco F (2003) Bulk processes prior to the 2001 Mount Etna eruption, highlighted through microgravity studies. J Geophys Res 108. https://doi.org/10.1029/2003JB002542

Corsaro RA, Falsaperla S, Langer H (2013) Geochemical pattern classification of recent volcanic products from Mt. Etna, Italy, based on Kohonen maps and fuzzy clustering. Int J Earth Sci. https://doi.org/10.1007/s00531-012-0851-7

Corsaro RA, Andronico D, Behncke B, Branca S, Caltabiano T, Ciancitto F, Cristaldi A, De Beni E, La Spina A, Lodato L, Miraglia L, Neri M, Salerno G, Scollo S, Spata G (2017) Monitoring the December 2015 summit eruptions of Mt. Etna (Italy): implications on eruptive dynamics. J Volcanol Geotherm Res 341:53–69. https://doi.org/10.1016/j.jvolgeores.2017.04.018

Crisci GM, Avolio MV, Behncke B, D'Ambrosio D, Di Gregorio S, Lupiano V, Neri M, Rongo R, Spataro W (2010) Predicting the impact of lava flows at Mount Etna, Italy. J Geophys Res 115. https://doi.org/10.1029/2009JB006431

de Arcangelis L, Godano C, Grasso JR, Lippiello E (2016) Statistical physics approach to earthquake occurrence and forecasting. Phys Rep 628:1–91

Ferro A, Gambino S, Panepinto S, Falzone G, Laudani G, Ducarme B (2011) High precision tilt observation at Mt. Etna volcano, Italy. Acta Geophysica 59:618–632. https://doi.org/10.2478/s11600-011-0003-7

Feuillet N, Cocco M, Musumeci C, Nostro C (2006) Stress interaction between seismic and volcanic activity at Mt Etna. Geophys J Int 164:697–718. https://doi.org/10.1111/j.1365-246X.2005.02824.x

Gambino S, Falzone G, Ferro A, Laudani G (2014) Volcanic processes detected by tiltmeters: a review of experience on Sicilian volcanoes. J Volcanol Geotherm Res 271:43–54. https://doi.org/10.1016/j.jvolgeores.2013.11.007

Giammanco S, Neri M, Salerno GG, Caltabiano T, Burton MR, Longo V (2012) Evidence for a recent change in the shallow plumbing system of Mt. Etna (Italy): gas geochemistry and structural data during 2001–2005. J Volcanol Geotherm Res 251:90–97. https://doi.org/10.1016/j.jvolgeores.2012.06.001

Grossi E, Olivieri C, Buscema M (2017) Diagnosis of autism through EEG processed by advanced computational algorithms: a pilot study. Comput Methods Prog Biomed 142:73–79

Hanley JA, McNeil BJ (1982) The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 143(1):29–36

Harris A, Steffke A, Calvari S, Spampinato L (2011) Thirty years of satellite-derived lava discharge rates at Etna: implications for steady

volumetric output. J Geophys Res 116. https://doi.org/10.1029/2011JB008237

Hastie T, Tibshirani R, Friedman JH (2009) The elements of statistical learning: data mining, inference, and prediction. Springer, Berlin

Langer H, Falsaperla S, Thompson G (2003) Application of artificial neural networks for the classification of the seismic transients at Soufrie`re Hills volcano, Montserrat. Geophys Res Lett 30. https://doi.org/10.1029/2003GL018082

Langer H, Falsaperla S, Masotti M, Campanini R, Spampinato S, Messina A (2009) Synopsis of supervised and unsupervised pattern classification techniques applied to volcanic tremor data at Mt. Etna, Italy. Geophys J Int 178:1132–1144. https://doi.org/10.1111/j.1365-246X.2009.04179.x

Mader HM (2006) Volcanic processes as a source of statistical data. In: Mader HM, Coles SG, Connor CB, Connor LJ (eds) Statistics in volcanology, Special Publications of IAVCEI. Geological Society, London

Mantegna RN (1999) Hierarchical structure in financial markets. Eur Phys J B 11:193–197

Martin AJ, Umeda K, Connor CB, Weller JN, Zhao DP, Takahashi M (2004) Modeling long-term volcanic hazards through Bayesian inference: an example from the Tohoku volcanic arc, Japan. J Geophys Res 109. https://doi.org/10.1029/2004JB00320

Marzocchi W, Sandri L, Gasparini P, Newhall C, Boschi E (2004) Quantifying probabilities of volcanic events: the example of volcanic hazard at Mount Vesuvius. J Geophys Res 109. https://doi.org/10.1029/2004JB003155

Mattia M, Bruno V, Caltabiano T, Cannata A, Cannavò F, D'Alessandro W, Di Grazia G, Federico C, Giammanco S, La Spina A, Liuzzo A, Longo M, Monaco C, Patanè D, Salerno G (2015) A comprehensive interpretative model of slow slip events on Mt. Etna's eastern flank. Geochem Geophys Geosyst 16:635–658. https://doi.org/10.1002/2014GC005585

McKinsey Global Institute (2011) Big data: the next frontier for innovation, competition, and productivity. https://www.mckinsey.com/~/media/McKinsey/dotcom/Insightsandpubs/MGI/Research/TechnologyandInnovation/BigData/MGI_big_data_full_report.ashx

Newhall CG, Hoblitt RP (2002) Constructing event trees for volcanic crises. Bull Volcanol 64:3–20

Notsu K, Mori T, Do Vale SC, Kagi H, Ito T (2006) Monitoring quiescent volcanoes by diffuse CO2 degassing: case study of Mt. Fuji, Japan. Pure Appl Geophys 163:825–835. https://doi.org/10.1007/s00024-006-0051-0

Patanè D, Privitera E, Gresta S, Akinci A, Alparone S, Barberi G, Chiaraluce L, Cocina O, D'Amico S, De Gori P, Di Grazia G, Falsaperla S, Ferrari F, Gambino S, Giampiccolo E, Langer H, Maiolino V, Moretti M, Mostaccio A, Musumeci C, Piccinini D, Reitano D, Scarfì L, Spampinato S, Ursino A, Zuccarello L (2003) Seismological constraints for the dike emplacement of July-August 2001 lateral eruption at Mt. Etna volcano, Italy. Ann Geophys 46:599–608

Patanè D, Aiuppa A, Aloisi M, Behncke B, Cannata A, Coltelli M, Di Grazia G, Gambino S, Gurrieri S, Mattia M, Salerno G (2013) Insights into magma and fluid transfer at Mount Etna by a multiparametric approach: a model of the events leading to the 2011 eruptive cycle. J Geophys Res Solid Earth 118:3519–3539. https://doi.org/10.1002/jgrb.50248

Salerno GG, Burton M, Di Grazia G, Caltabiano T, Oppenheimer C (2018) Coupling between magmatic degassing and volcanic tremor in basaltic volcanism. Front Earth Sci 6. https://doi.org/10.3389/feart.2018

Schmid A, Grasso JR, Clarke D, Ferrazzini V, Bachèlery P, Staudacher T (2012) Eruption forerunners from multiparameter monitoring and application for eruptions time predictability (Piton de la Fournaise). J Geophys Res 117. https://doi.org/10.1029/2012JB009167

Selva J, Orsi G, Di Vito MA, Marzocchi W, Sandri L (2012) Probability hazard map for future vent opening at the Campi Flegrei caldera, Italy. Bull Volcanol. https://doi.org/10.1007/s00445-011-0528-2

Smits G (2015) Volcanic hazards as components of complex systems: the case of Japan. Asia Pac J 13(33):6

Spampinato L, Sciotto M, Cannata A, Cannavo F, La Spina A, Palano M, Salerno GG, Privitera E, Caltabiano T (2015) Multiparametric study of the February–April 2013 paroxysmal phase of Mt. Etna New South-East crater. Geochem Geophys Geosyst. https://doi.org/10.1002/2015GC005795

Spilliaert N, Allard P, Mètrich N, Sobolev AV (2006) Melt inclusion record of the conditions of ascent, degassing, and extrusion of volatile-rich alkali basalt during the powerful 2002 flank eruption of Mount Etna (Italy). J Geophys Res 111. https://doi.org/10.1029/5742005JB003934

Sutton AJ, Elias T, Gerlach TM, Stokes JB (2001) Implications for eruptive processes as indicated by sulfur dioxide emission from Kīlauea volcano, Hawaii, 1979–1997. J Volcanol Geotherm Res 108:283–302

Swets JA (1988) Measuring the accuracy of diagnostic systems. Science 240:1285–1293

Tallarida RJ, Murray RB (1987) Area under a curve: trapezoidal and Simpson's rules. In: Tallarida RJ, Murray RB (eds) Manual of pharmacologic calculations. Springer, New York, pp 77–81

Tonini R, Sandri L, Rouwet D, Caudron C, Marzocchi W, Suparjan (2016) A new Bayesian Event Tree tool to track and quantify volcanic unrest and its application to Kawah Ijen volcano. Geochem Geophys Geosyst. https://doi.org/10.1002/2016GC006327

Zhuang J (2010) Gambling scores for earthquake predictions and forecasts. Geophys J Int 181(1):382–390