Clustering of eruptive events from high precision strain signals recorded during the 2020-2022 lava fountains at Etna volcano (Italy)

Luigi Carleo[1], Gilda Currenti[1], Alessandro Bonaccorso[1]

[1]Istituto Nazionale di Geofisica e Vulcanologia, Sezione di Catania – Osservatorio Etneo, Catania, 95125, Italy

5 *Correspondence to*: Gilda Currenti (gilda.currenti@ingv.it)

Abstract. Explosive eruption events have been clustered by machine learning techniques applied on strain signal recorded by high-precision borehole strainmeters. We focus on the extraordinary intense and frequent eruptive activity at Etna in the period December 2020 – February 2022 when more than 60 lava fountains occurred. We apply the k-means algorithm on the associated strain variations which are representative of the eruptive dynamics. A novel procedure was developed to ensure a
10 high-quality clustering process and obtain robust results. The analysis identified four distinct groups of strain variations which characterize the events in terms of amplitude, and duration and time derivative of the signal. The temporal distribution of the clusters provides useful insights into the evolution of the volcano activity and reveals transitions in the eruptive style.

1 Introduction

In the last decade, lava fountains represented a typical eruptive style at the Etna volcano (i.e., Calvari et al., 2018; Andronico
15 et al., 2021). These eruptive events are powerful jets that can expel hot mixtures of solid particles and volcanic gases to heights ranging from tens to several hundred meters. The ash dispersal and fall-out deposits of the solid particles, known as tephra, cause critical hazards to civil infrastructures and to aviation, frequently provoking the temporary closure of southern Italy airports. The characterization of such eruptive events is thus fundamental for both monitoring the volcano activity and evaluating the associated hazards.

20 At Etna, lava fountain events produce short-term and small deformations of the ground (Bonaccorso and Calvari, 2017; Bonaccorso et al., 2013b; Bonaccorso et al., 2016; Bonaccorso et al., 2021) that are well-captured by the Sacks-Evertson dilatometer (Sacks et al., 1971), a widely employed geophysical instrument to study ground deformation processes associated with volcano unrest (i.e. Amoruso et al., 2015; Bonaccorso et al., 2012; Bonaccorso et al., 2020; Linde and Sacks, 1995; Linde et al., 2016; Linde et al., 1993; Voight et al., 2006). This dilatometer is particularly appropriate to monitor lava
25 fountains since it measures the volumetric strain with a very wide frequency range (10^{-7} to >20 Hz) and with the highest resolution (10^{-10} to 10^{-11}) achievable among geophysical instruments (i.e. NASEM, 2017; Roeloffs and Linde, 2007). Other geodetic techniques such as GPS and InSAR are unable to detect the deformations associated with lava fountains because of their lower accuracy (GPS > 0.5 cm) or lower frequency sampling (InSAR periodic passages). These technical characteristics

make the strain measurements fundamental for monitoring lava fountains, especially when images from surveillance cameras

30   do not allow the event detection because of poor visibility (Carleo et al. 2022b; Calvari and Nunnari, 2022).

From December 2020 to February 2022, Etna underwent an intense eruptive activity with more than 60 lava fountain events from the South East Crater (Calvari and Nunnari, 2022). A variability in terms of duration, degree of explosiveness and portion of effusive flows, has been observed (Calvari and Nunnari, 2022; Calvari et al., 2022), implying a different degree of the hazard associated with these eruptive events. Indeed, the onset and the dynamic of the lava fountain is usually a gradual

35   growing process, starting from weak Strombolian activity, continuing with transitional explosive activity, and eventually leading to sustained eruptive columns. The intensity and the duration of these three main phases are not always the same and characterize the temporal evolution of the episodes. A preliminary inspection on the strain signal recorded during the lava fountains reveals a similar pattern for all the events and a strict relationship with their temporal evolution (Bonaccorso et al., 2021; Calvari et al., 2021) allowing tracing the waxing and waning of each episode and marking the onset and the end of the

40   eruptions. On average, but not systematically, some differences arise in terms of amplitude and duration of the strain signal. For example, the lava fountains occurring in February – April 2021 were characterized by strain changes with high amplitudes (hundreds of nstrain) and temporal evolutions ranging from tens of minutes to 8-9 hours. Conversely, the strain changes accompanying the eruptions in May – June 2021 were lower in amplitude (tens of nstrain) and developed in intervals from 1 hour to less than 4 hours.

45   In the recent past, attempts to classify the lava fountains at Etna have been made manually by experts by comparing different geophysical and volcanological data (i.e. Andronico et al., 2021; Calvari et al., 2022). However, manual classification is time consuming and prone to subjective biases. With the aim of avoiding a classification biased by experts' belief, we investigate whether an objective cluster analysis on the instrumental dataset could help in discovering group of events with similar characteristics. Clustering analyses of monitoring signals have already been performed in volcanology (Cirillo et al. 2022;

50   Corradino et al., 2021; Langer et al., 2009; Nunnari, 2021; Romano et al., 2022; Unglert et al., 2016).

Here, we made use of clustering techniques applied on the strain variations recorded concurrently with the eruptive episodes from December 2020 to February 2022 in order to derive the key features that characterize the eruptive process and distinguish the events. In particular, we applied the k-means method, a widely employed unsupervised machine learning algorithm to solve clustering problems in several domains (Lloyd, 1982; MacQueen, 1967). One of the drawbacks of such

55   algorithm is that the optimal number of the clusters and also the optimal set of key features which lead to a high-quality clustering are not known *a priori*. We developed a procedure to appropriately identify the features and the number of clusters with high cohesion and separation. Moreover, since the clustering solution could depend on the initial position of the barycenter of the clusters (centroids) chosen to start the algorithm (Fränti and Sieranoja, 2019), we also investigated the influence of the initial position of the centroids on the k-means performance by comparing different initialization techniques.

60   Lastly, we discuss the implications that this result entails in the assessment of volcanic activity and the associated eruptive style.

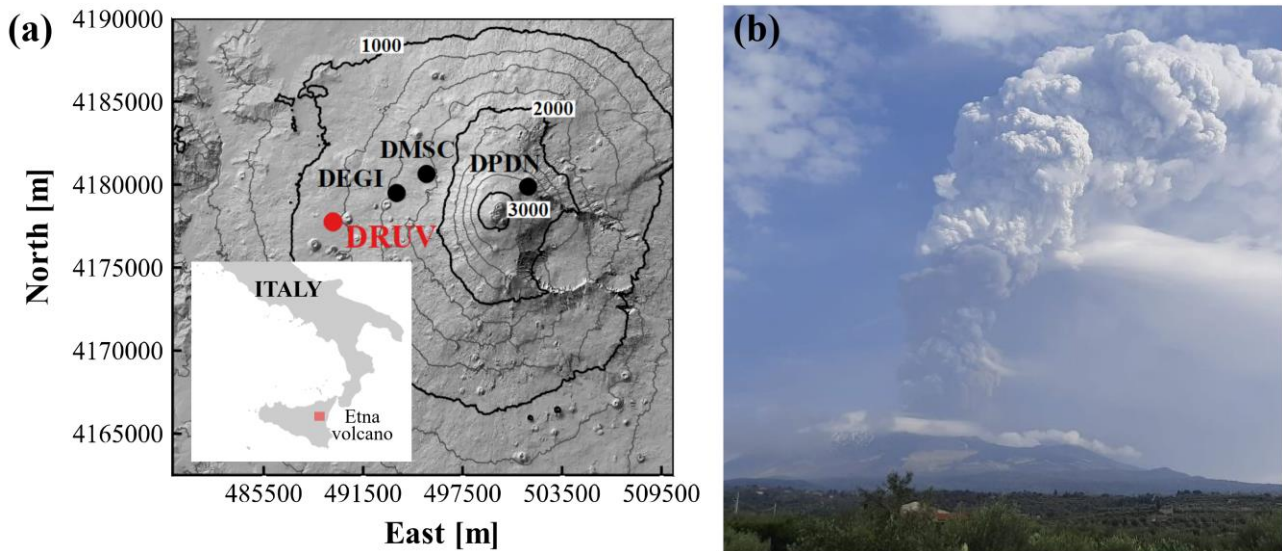



Figure 1: (a) Location of the borehole strainmeter stations installed at the Etna volcano. The coordinates system is WGS 84 UTM 33. (b) Lava fountain occurring on 23 October 2021 (photo from INGV internal report n° 43/2021).

2 Strain changes during the Etna lava fountains in 2020-2022

The December 2020 - February 2022 Etna eruptive activity was extraordinarily intense. It started with four lava fountains from 13 December 2020 to 16 January 2021 (period P_i). Successively, a first lava fountain sequence of 17 events took place up to 1 April 2021, with an average frequency of 0.39 events/day (sequence S_1). After 49 days of repose, Etna volcano reawakened and a second lava fountain sequence of 34 episodes occurred till 10 August 2021, with a frequency of about 0.42 events/day (S_2). Then, the eruptive activity diminished with 5 events occurring from 10 August 2021 to the end of February 2022 (P_f).

The borehole strainmeter network, operating at Etna since 2011 (Bonaccorso et al., 2016; Fig. 1), was fundamental in investigating the dynamics of the eruptions (Bonaccorso et al., 2021) and monitoring the eruptive events in near real-time for volcanic surveillance (Carleo et al., 2022b). In this study, we focus on the measurements recorded by the DRUV station, which is located quite far from the summit craters, at ~11 km, and installed in a massive rock layer (at ~180 m depth) guaranteeing high-efficiency in transferring deformation from the rock to the sensor. The strainmeter was calibrated with three different techniques (Bonaccorso et al., 2013a; Bonaccorso et al., 2016; Currenti et al., 2017) that confirmed its high sensitivity (~10^{-10}). The DRUV strain signal was filtered from the disturbing effects of both the Earth tides and the barometric pressure to highlight small strain variations related to the volcano activity (Currenti and Bonaccorso, 2019). We used the procedure proposed by Carleo et al. (2022a) to highlight tiny volcano-related strain changes up to 10^{-10} for time scales less than 1 day. Furthermore, we removed the long-term drift component from the strain signal due to the effect of both the curing of the cement and the relaxation of the drilled hole (Canitano et al., 2021).

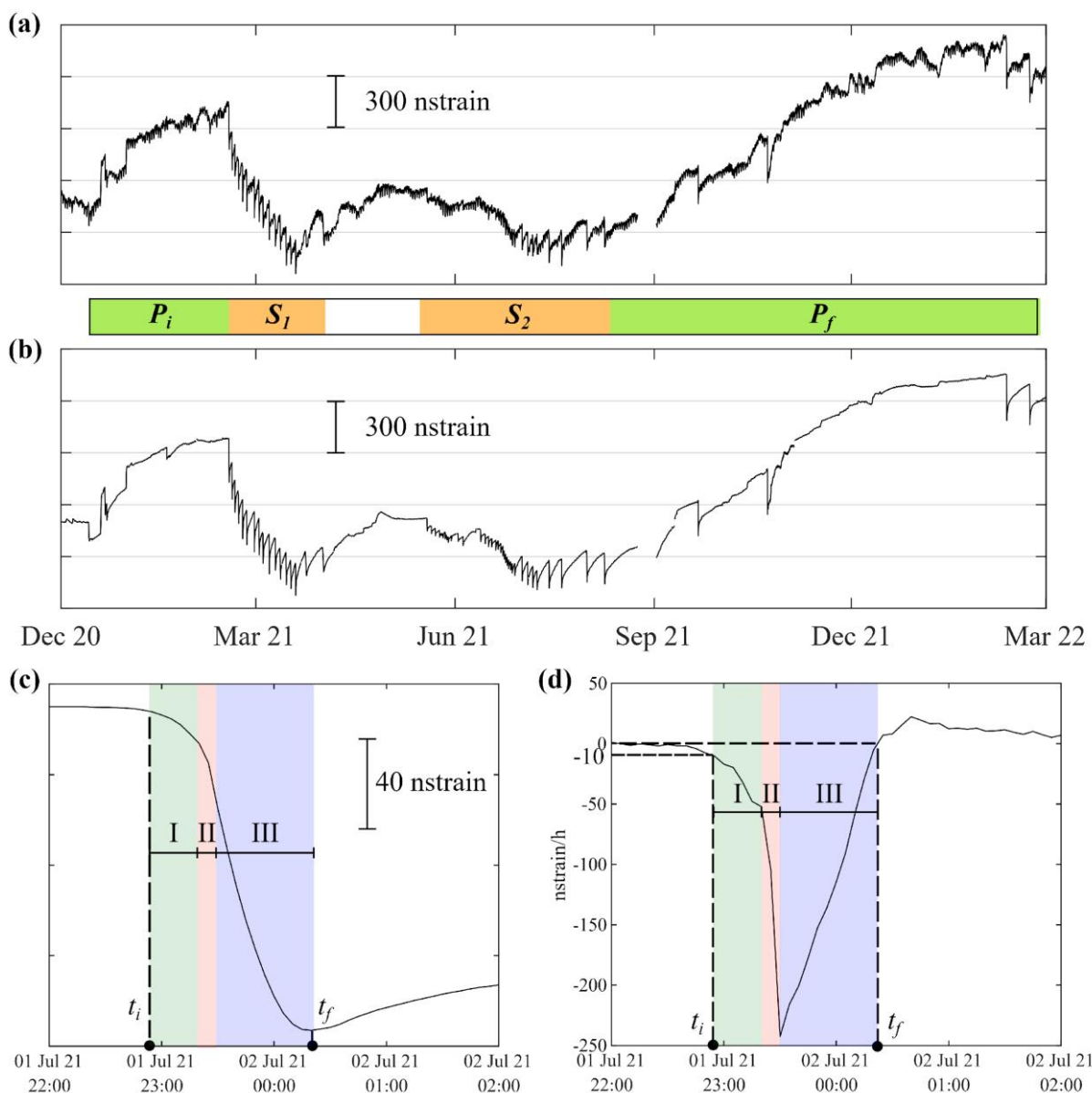



Figure 2: The detrended recorded (a) and filtered (b) DRUV signals in the study period. Negative step-like strain variations,

85 **clearly detected in the filtered signal, occurred concurrently with the lava fountain episodes. The lava fountains on 29 August and**
23 October 2021 were not recorded due to technical problems. The eruptive activity is divided into 4 periods: P_i (4 events), S_1 (17
events), S_2 (34 events) and P_f (5 events). The filtered strain signal (c) and the related strain rate signal (d) during the lava fountain
on 01 July – 02 July 2021 are shown as an example. t_i and t_f indicate the onset and the end of the strain variations. At the beginning
of the lava fountain (I), expansion rate gradually increases. In the central part (II), strain rate rapidly reaches its absolute
90 **maximum value. The expansion gradually diminishes and the strain rate reaches the pre-event level during the final part of the**
eruptive event (III).

In Fig. 2a and 2b, both the recorded and the filtered DRUV signals are presented. Each negative step-like strain variation, clearly visible in the filtered signal, happened concurrently with a lava fountain and correspond to rock expansion around the sensor. The recorded dilatation is the response to the decompression of the magmatic source feeding the lava fountain (Currenti and Bonaccorso, 2019). The near-real time detection algorithm developed by Carleo et al. (2022b) allowed the automatic detection of all the 58 strain variations related to the eruptive episodes in the analyzed period. In Fig. 2c and 2d, the filtered strain signal and its time derivative (strain rate signal) during the lava fountain on 1 – 2 July 2021 are shown as an example of the recorded strain variations. Typically, the strain and the strain rate signals show a sigmoid and a V shape, respectively, which can be described by separating the signals in three parts: in the initial part (I) the strain rate gradually decreases with time and the strain signal shows an elbow with a downward concavity; in the central part (II) the strain rate reaches the absolute maximum value; in the final part (III) the strain rate inverts its trend reaching the pre-event level. We define the onset of the variation t_i when the strain rate reaches the value of -10 nstrain/h, while the end of the variation t_f was set when the sign of the strain rate becomes positive.

We characterized each lava fountain event by extracting the main features from both the strain and strain rate signals. In particular, we focused on parts I and II of the signals. Feature extraction transforms each eruptive event into a strain data point in the feature domain which forms a dataset that is going to be clustered.

3 Clustering the lava fountain events through the related strain variations

We used the k-means algorithm to characterize the 2020-2022 lava fountain events using the associated strain changes. We varied the inputs of the algorithm, namely the k number of clusters, the employed set of features and the initial centroid positions. We analyzed the quality of the different clustering solutions by employing two methods, both described in Appendix A: the "Elbow method", based on the Sum-of-Squared Errors (SSE), and the Silhouette value (Sil; Kaufman and Rousseeuw, 2009). An iterative procedure (Appendix B) was developed to find the optimal number of clusters, k_{opt}, and the optimal subset of features, C_{opt}, from a set of 15 potential features (Table B1) whose values were previously normalized to warrant a balanced weight in the clustering process (Langer et al., 2020). We compared two different initialization techniques to evaluate the effects of the initial centroid position on the k-means performance: the random centroid (RC) position (MacQueen, 1967) and the method proposed by Yelda et al. (2010) (YC). We repeated the k-means clustering n times by varying the RC and the YC positions randomly and took the solution with the lowest SSE value. We tested the influence of different values of n for the quality of the clustering.

The iterative procedure converged in two steps and provided an optimal subset of features composed by the amplitude A and the duration D of the strain variation and the average strain rate in the time window ranging from t_i to the time when the strain rate reaches the 75% of the minimum strain rate, SS_{0-75}. The results are presented in Fig. B1. In the k – SSE plot (Fig. B1a), the elbow of most of the curves is found at $k = 4$ which can be selected as the optimal number of clusters, k_{opt}.

Table 3: Coordinates of the cluster centroids and mode values of the optimal features for each cluster.

Clusters	Amplitude A [nstrain]		Duration D [hour]		Strain rate SS_{0-75} [nstrain/h]	
	Centroid	Mode	Centroid	Mode	Centroid	Mode
Cluster 1	50.0	51.5	1.15	0.98	32.5	20.4
Cluster 2	177.5	181.8	1.92	1.78	59.1	40.5
Cluster 3	232.7	225.3	0.83	0.98	187.6	181.4
Cluster 4	116.8	116.7	9.25	8.94	12.3	20.4

125

The *Sil* values are all positives indicating a good clustering for all the strain data points. Moreover, the average Silhouette values for the four clusters are very high and equal to 0.90, 0.71, 0.82 and 0.99, denoting both a high cohesion in the same cluster and a high separation among the clusters (Text B1). The results of the analysis on the influence of the initial centroid position and on the number of repetitions of the k-means algorithm are summarized in Fig. B2. We reported the k – SSE and

130   the k – *Sil* plots related to both the RC and YC positions with value of n equal to 10, 10^4 and 10^6. The k – SSE plots for the different values of n (Fig. B2a,c,e) showed notably overlapped curves, indicating that the analyzed initial centroid positions do not affect the shape of the curves and, hence, the choice of k_{opt} can be easily identified at $k = 4$. In Fig. B2b,d,f the number of cluster k is plotted against the *Sil* value for n equal to 10, 10^4 and 10^6, respectively. The figures highlight that a high number of repetitions is necessary to make the clustering independent from the analyzed initialization techniques. Indeed, the

135   k – *Sil* curves overlap with n values higher than 10^4 and up to $k = 4$ which corresponds to k_{opt}. Therefore, the outputs of the procedure, k_{opt} and C_{opt}, can be considered reliable. The mode values of the frequency distributions of the C_{opt} features are presented together with their centroids locations in Table 1. In Fig. 3a, the strain changes are presented in the C_{opt} feature domain. A very good clustering can also be observed visually confirming the reliability of the procedure in providing high-quality results. The frequency distribution of the values of the C_{opt} features, A, D and SS_{0-75}, are presented in Fig. 3b-d,

140   respectively. In Fig. 3e, all the strain changes are plotted by aligning them with their initial time t_i for a further visual comparison. The analysis of the distributions of the cluster features allows us to identify the main characteristics of the events. Cluster 1 gathers lava fountain episodes of low strain amplitude and duration and characterized by small initial strain rate changes. All the features of Cluster 1 are located in the lower range of variations. Cluster 2 groups events whose features cover more the intermediate part of their ranges. Cluster 3 gathers events characterized by high deformations evolving in a

145   very short time window, less than 1 hour. Furthermore, the mode value of the SS_{0-75} feature for Cluster 3 (181.4 nstrain/h; Table 1) is 4.5 to 9 times higher than for the others. Cluster 4 groups the episodes with the highest mode value for the duration feature D, which is 5 to 11 times higher with respect to the ones related to the other clusters. The mode of the SS_{0-75} feature in the Cluster 4 shows the lowest value among the clusters.

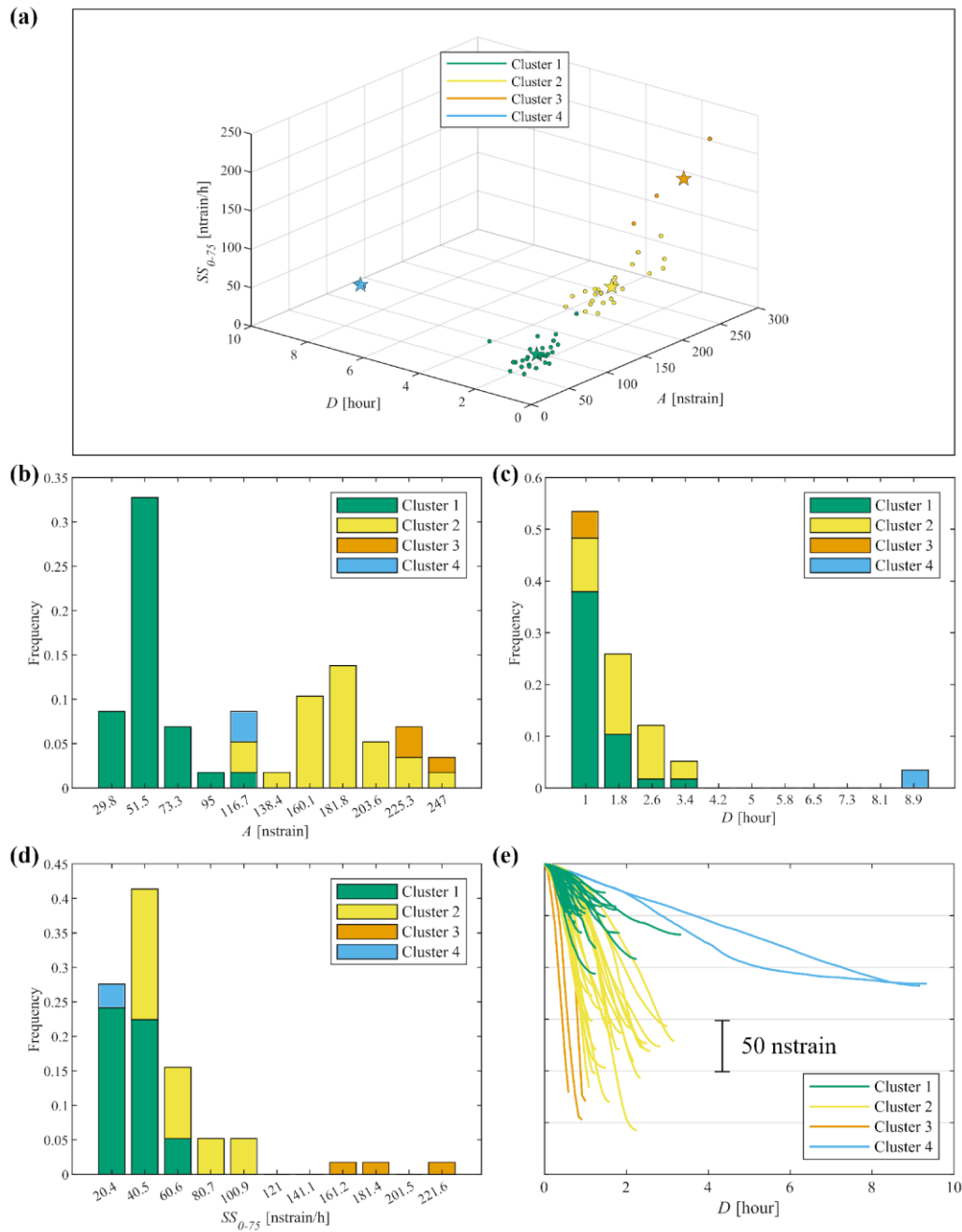
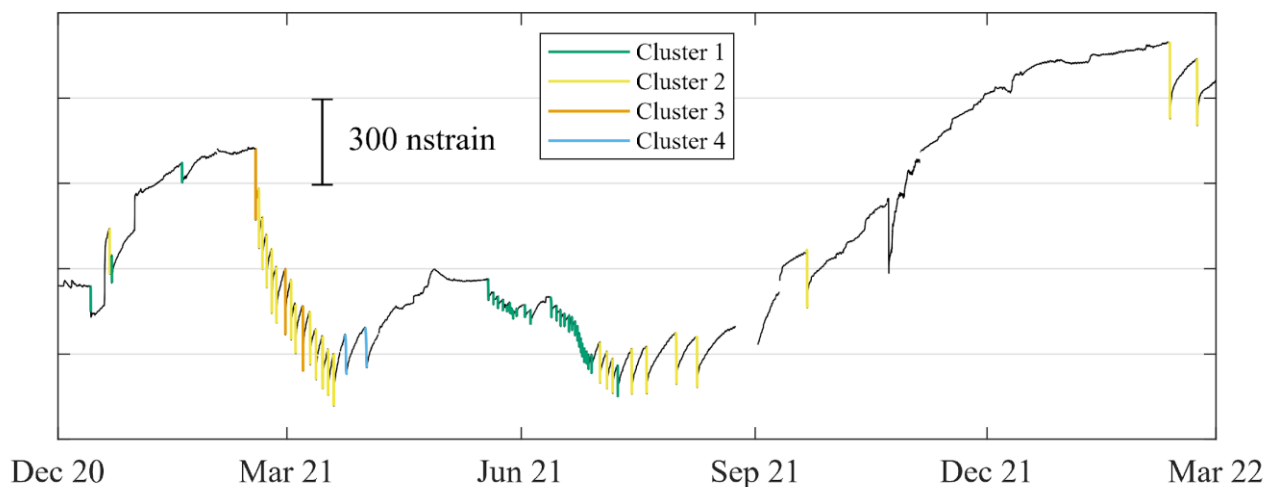



Figure 3: (a) All the clustered strain data points plotted as circles in the domain of the optimal features: amplitude A, duration D and strain rate SS_{0-75}. Stars represent the cluster centroids. Frequency distribution of A **(b)**, D **(c)** and SS_{0-75} **(d)** for the different clusters. Mode values of the distributions are reported in Table 1. In **(e)**, all the clustered strain variations are aligned with the initial time t_i.

155   **Figure 4: Temporal distribution of the clusters in the analyzed time period plotted together with the filtered DRUV signal.**

4 Discussions and Conclusion

We studied the eruptive activity occurring at Etna in the period December 2020 - February 2022, where a variability in the characteristics of the lava fountain events has been observed (Andronico et al., 2021; Calvari and Nunnari, 2022). A signature of this variability is found in the diversity of the ground deformation response imprinted in the strain signal

160   (Calvari et al., 2022). Using the extraordinary 2020-2022 strain dataset, we investigated the use of an automated clustering analysis to provide an objective quantitative measure of similarities and differences of the lava fountain events.

The clustering analysis allows us to methodically identify three key features (A, D and SS_{0-75}) that grouped the events in four distinct and coherent clusters. In particular, all the three features are required to distinguish Cluster 1 and 2 from the other clusters, while the SS_{0-75} and the duration features sharply identify Cluster 3 and Cluster 4, respectively.

165   The clustered events do not occur randomly but are grouped over time as shown in Fig. 4, denoting a transition in the eruptive dynamic. It appears that the clusters have an intimate relationship with the volcanic eruption style. In the period P_i+S_1 (December 2020 – March 2021), Andronico et al. (2021) manually identified three eruption styles classified as transitional activity (TA), sustained lava fountain (LF) and large-scale lava fountain (LSLF). Comparing the strain clustering and the eruption style classification reported in Andronico et al. (2021), we observed an interesting correspondence. The first

170   events recorded in P_i and classified as TA are all grouped in Cluster 1, except the 21 December lava fountain which falls into Cluster 2. Then, in the S_1 lava fountain sequence, the eruptive style turned into LF with episodic LSLF events. As well, the clustering highlights a transition, grouping the S_1 events in Cluster 2 and 3. The three events, that belong to Cluster 3, are all classified as LSLF in Andronico et al., (2021) and occurred closely in time on 16 February, 28 February and 7 March 2021. At the end of S_1 on 23 March and 31 March, two lava fountains belonging to Cluster 4 occurred. Although classified as LF

175 by Andronico et al. (2021), their eruptive behavior, in terms of long duration, large effusive volume emitted (Calvari and Nunnari, 2022) and very low strain-rate, is quite different from other lava fountain events. The clustering analysis on the strain signal is able to distinguish them and highlights the duration as their distinctive feature (Fig. 3a). Indeed, also the analysis of the thermal camera images (Calvari and Nunnari, 2022; in Fig. 13) evidenced their very long duration, well above the average value observed for the others. After a period of repose, the new sequence S_2 restarted in May 2021 with events

180 belonging to Cluster 1 that slowly over time turned into events belonging to Cluster 2. This transition is in agreement with a variation in the parameters estimated from the thermal camera images (Calvari and Nunnari, 2022).

Summarizing, the automatic clustering on the strain signal recognizes and identifies four classes of lava fountains confirming the three classes manually identified and adding a further class of longer events in Cluster 4. Moreover, the time occurrence of the clustered events is closely linked to the eruptive style evolution observed by the experts. Changes in the eruptive style

185 are regulated by many interrelated magmatic properties and processes (Cassidy et al., 2018). The exsolved and dissolved gas content, overpressure at depth, magma composition and conduit geometry of the plumbing system, all control the speed at which magma ascends, decompresses and outgasses, and hence, determine eruptive style and evolution. Our findings demonstrate that the clustering analysis on the strain signal may contribute in discovering persistence and transition in the eruptive style and thus may give indirect insights into the evolution of the magmatic plumbing system. A joint analysis,

190 together with other geophysical, geochemical, volcanological and petrophysical data, may help in confirming the persistence and transition of the magmatic system conditions and in identifying the most likely magmatic properties and/or processes that regulate the lava fountain behavior at Etna volcano.

Appendix A

The quality of the clustering solutions was evaluated using both the "Elbow method" and the Silhouette value (Kaufman and

195 Rousseeuw, 2009). The "Elbow method" is one of the most widely used techniques to find the optimal number of the clusters. The key quantity is the *sum-of-squared errors* (SSE), defined as the sum of the squared Euclidean distances between the data points and the centroids to which the data points belong. The k-means algorithm clusters data so that the SSE are as little as possible. Therefore, the less the SSE, the better the clustering. Normally, the higher the number of the clusters *k*, the lower the SSE. In a *k* - SSE plot, the elbow in the curve represents the optimal number of clusters for the

200 analyzed dataset and corresponds to the most effective clustering solution in terms of *k* and SSE.

The Silhouette value, *Sil*, for a single data point is defined as:

$$Sil = \frac{b-a}{max\{b,a\}} \,, \tag{A1}$$

where *b* is the average distance between the datum and the data of another cluster minimized over the clusters, and *a* is the

205 average distance between the datum and the data within the cluster to which the datum belongs. The *Sil* value is a measure of

how much a data point is cohesive within its own cluster (distance *a*) and, at the same time, separated from the other clusters (distance *b*). It ranges from -1 to 1, where -1 corresponds to a completely wrong clustering while 1 to an optimal clustering.

Appendix B

We developed an iterative procedure to extract an optimal subset of features, C_{opt}, from a set of 15 potential features $X = \{X_1,$
210 $X_2, \ldots, X_j\}$, where *j* represents the *j*-th feature of *X*, (Table B1), and to determine the optimal number of clusters, k_{opt}. Both the "Elbow method" and the Silhouette value *Sil* are used to evaluate the quality of the clustering solutions. The procedure is organized in the following steps:

1) create the most basic subset of features C_{start} composed by the amplitude *A* and the duration *D* of the strain variation;

2) if it is the first iteration of the procedure, the starting subset of feature related to the *i*-th iteration, C_i, is C_{start} otherwise
215 $C_{i-1,j\_max,}$ defined at point 7);

3) consider a new set of features $X_{left} = X - C_i$. Create all the possible subsets of features, $C_{i,j}$, composed by C_i plus one feature from X_{left};

4) cluster the dataset using C_i and all the $C_{i,j}$;

5) find the optimal number of clusters for the *i*-th iteration of the procedure, $k_{opt,i}$, by comparing all the $k - SSE$ curves;

220 6) at $k_{opt,i}$, calculate the Silhouette values averaged over the clusters related to C_i and all the $C_{i,j}$, $Sil_{a,Ci}$ and $Sil_{a,Ci,j}$ respectively;

7) if $\max\{Sil_{a,Ci,j}\} > Sil_{a,Ci}$ then define a new subset of features C_{i,j_max} composed by C_i plus the feature that provides $\max\{Sil_{a,Ci,j}\}$; repeat from point 2) to point 7) updating C_i with C_{i,j_max}. If $\max\{Sil_{a,Ci,j}\} \leq S_{a,Ci}$, stop the procedure and take C_i as the optimal set of features, C_{opt}, and $k_{opt,i}$ as the optimal number of clusters, k_{opt}.

225 Before proceeding with the analysis, since the features are in different units and ranges, we normalized them in the range [0 1] to ensure a balanced weight in the clustering process (Langer et al., 2020). The initial centroid position was chosen randomly. To have more robust results, for each clustering solution, we performed $n = 10^4$ repetitions of the k-means, setting different random initial seeds and keeping fixed the other inputs. Then, we chose the solution with the lowest SSE value. The robustness of the choice of *n* and of the RC initialization in providing reliable results was then tested.

230 The influence of the initialization was also investigated by comparing two seeding techniques: the random centroid (RC) position (MacQueen, 1967) and the method proposed by Yelda et al. (2010) (YC). The former method is the most popular and consists of locating the centroids randomly in the range of variation of the features, namely, in our case, between 0 and 1. The latter method involves first sorting data points in accordance with their distance from the origin and, then, partitioning them in *k* clusters with equal number of sorted points. Yelda et al. (2010) proposed to locate the initial centroid position in

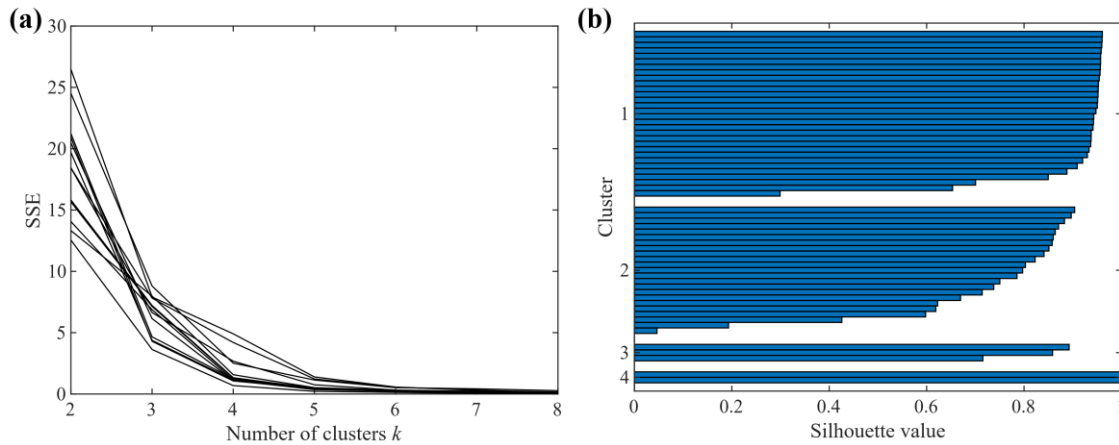
235 the barycenter of each cluster. We introduced more randomness by locating the centroids randomly in each cluster. The tests were performed using the optimal set of features C_{opt}. To investigate the importance of performing repetitions, we repeated k-means by considering *n* in the range [10 10^6].

240 **Figure B1: (a) k – SSE curves related to all the clustering solutions analyzed in the last step of the procedure. (b) Silhouette values for the clustering solution obtained with the optimal set of features C_{opt} and the optimal number of clusters $k_{opt} = 4$.**

Table B1: Features considered in the cluster analysis

Symbol	Description
A	Amplitude of the strain change from t_i to t_f
D	Duration of the strain change from t_i to t_f
Sr_{min}	Minimum strain rate from t_i to t_f
SA_{0-75}	Amplitude from t_i to the instant when 75% of Sr_{min} is reached
SA_{0-100}	Amplitude from t_i to the instant when 100% of Sr_{min} is reached
SD_{0-75}	Length of the time window from t_i to the instant when 75% of Sr_{min} is reached
SD_{0-100}	Length of the time window from t_i to the instant when 100% of Sr_{min} is reached
SS_{0-75}	Average strain rate from t_i to the instant when 75% of Sr_{min} is reached
SS_{0-100}	Average strain rate from t_i to the instant when 100% of Sr_{min} is reached
AS_{0-50}	Average strain rate from t_i to the instant when 50% of A is reached
AS_{0-75}	Average strain rate from t_i to the instant when 75% of A is reached
$ASr_{min,0-50}$	Minimum strain rate from t_i to the instant when 50% of A is reached
$ASr_{min,0-75}$	Minimum strain rate from t_i to the instant when 75% of A is reached
AD_{0-50}	Length of the time window from t_i to the instant when 50% of A is reached
AD_{0-75}	Length of the time window from t_i to the instant when 75% of A is reached

245

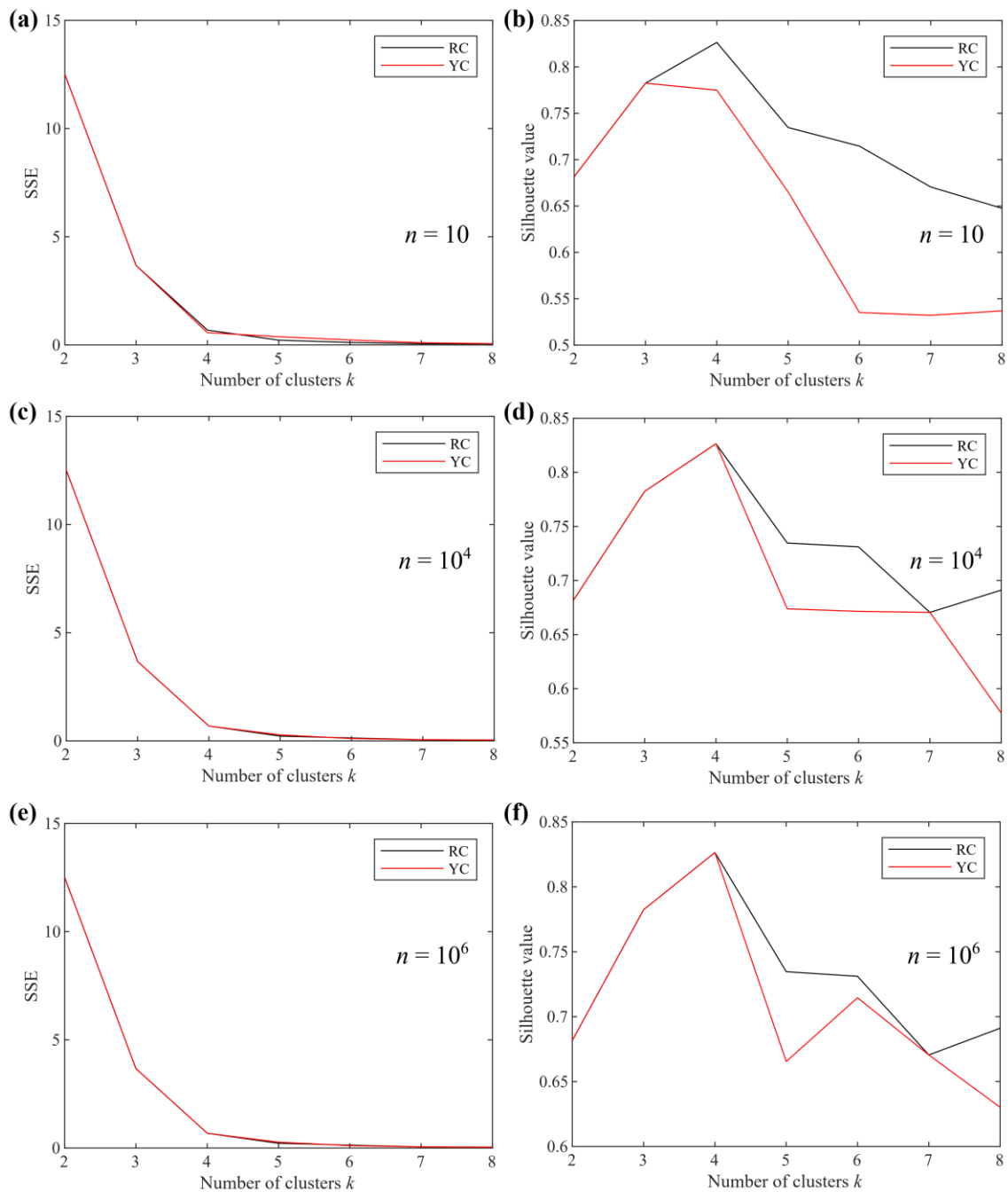



Figure B2: k – SSE and k – Sil plots obtained using the optimal set of features C_{opt}. The SSE and Silhouette values are presented considering the random initial centroid (RC) positions and centroids located as proposed by Yelda et al. (2010) (YC). The results

250 **obtained performing a number of repetitions n equal to 10 (a and b), 10^4 (c and d) and 10^6 (e and f) are shown as an example.**

Code and data availability

MATLAB scripts and data are available upon request to the corresponding Author

Author contribution

GC and LC conceived and conceptualized the study. LC developed the code and performed the analyses. AB managed and
255 administrated the funding acquisition for conducting the research. All Authors contributed to the writing of the manuscript
and the discussion of the results.

Competing interests

The authors declare that they have no conflict of interests.

Acknowledgements

260 This research benefited from funding provided by the 2019–2021 Agreement between INGV and Italian Presidenza del
Consiglio dei Ministri, Dipartimento della Protezione Civile (DPC), All. B2 - WP2 - Task 9 and also from the EC H2020-
FET OPEN project "SiC nano for picoGeo", grant agreement No. 863220.

References

Amoruso, A., Crescentini, L., Scarpa, R., Bilham, R., Linde, A. T., and Sacks, S. I.: Abrupt magma chamber contraction and
265 microseismicity at Campi Flegrei, Italy: Cause and effect determined from strainmeters and tiltmeters, J. Geophys. Res. Sol.-
Ea., 120(8), 5467–5478, https://doi.org/10.1002/2015JB012085, 2015.

Andronico, D., Cannata, A., Di Grazia, G., and Ferrari, F.: The 1986–2021 paroxysmal episodes at the summit craters of Mt.
Etna: Insights into volcano dynamics and hazard, Earth-Sci. Rev., 220, 103686,
https://doi.org/10.1016/j.earscirev.2021.103686, 2021.

270 Bonaccorso, A., and Calvari, S.: A new approach to investigate an eruptive paroxysmal sequence using camera and
strainmeter networks: Lessons from the 3–5 December 2015 activity at Etna volcano, Earth Planet. Sc. Lett., 475, 231–241,
https://doi.org/10.1016/j.epsl.2017.07.020, 2017.

Bonaccorso, A., Calvari, S., Linde, A. T., Sacks, S. I., and Boschi, E.: Dynamics of the shallow plumbing system
investigated from borehole strainmeters and cameras during the 15 March, 2007 Vulcanian paroxysm at Stromboli volcano,
275 Earth Planet. Sc. Lett., 357–358(2012), 249–256, https://doi.org/10.1016/j.epsl.2012.09.009, 2012.

Bonaccorso, A., Calvari, S., Currenti, G., Del Negro, C., Ganci, G., Linde, A. T., Napoli, R., Sacks, S. I., and Sicali, A.: From source to surface: dynamics of Etna's lava fountains investigated by continuous strain, magnetic, ground and satellite thermal data, B. Volcanol., 75(2), 1–12, https://doi.org/10.1007/s00445-013-0690-9, 2013a.

Bonaccorso, A., Currenti, G., Linde, A. T., and Sacks, S. I.: New data from borehole strainmeters to infer lava fountain
280   sources (Etna 2011–2012), Geophys. Res. Lett., 40(14), 3579–3584, https://doi.org/10.1002/grl.50692, 2013b.

Bonaccorso, A., Linde, A. T., Currenti, G., Sacks, S. I., and Sicali, A.: The borehole dilatometer network of Mount Etna: A powerful tool to detect and infer volcano dynamics, J. Geophys. Res. Sol.- Ea., 121(6), 4655–4669, https://doi.org/10.1002/2016JB012914, 2016.

Bonaccorso, A., Currenti, G., Linde, A. T., Sacks, S. I., and Sicali, A.: Advances in Understanding Intrusive, Explosive and
285   Effusive Processes as Revealed by the Borehole Dilatometer Network on Mt. Etna Volcano, Front. Earth Sci., 7, 357, https://doi.org/10.3389/feart.2019.00357, 2020.

Bonaccorso, A., Carleo, L., Currenti, G., and Sicali, A.: Magma Migration at Shallower Levels and Lava Fountains Sequence as Revealed by Borehole Dilatometers on Etna Volcano, Front. Earth Sci., 9, 740505, https://doi.org/10.3389/feart.2021.740505, 2021.

290   Calvari, S., and Nunnari, G.: Comparison between Automated and Manual Detection of Lava Fountains from Fixed Monitoring Thermal Cameras at Etna Volcano, Italy. Remote Sens. - Basel., 14(10), 2392, https://doi.org/10.3390/rs14102392, 2022.

Calvari, S., Cannavò, F., Bonaccorso, A., Spampinato, L. and Pellegrino, A. G.: Paroxysmal Explosions, Lava Fountains and Ash Plumes at Etna Volcano: Eruptive Processes and Hazard Implications, Front. Earth Sci., 6:107.
295   https://doi.org/10.3389/feart.2018.00107, 2018.

Calvari, S., Bonaccorso, A., and Ganci, G.: Anatomy of a paroxysmal lava fountain at Etna volcano: The case of the 12 March 2021, episode, Remote Sens. - Basel., 13(15), 3052, https://doi.org/10.3390/rs13153052, 2021.

Calvari, S., Biale, E., Bonaccorso, A., Cannata, A., Carleo, L., Currenti, G., Di Grazia, G., Ganci, G., Iozzia, A., Pecora, E., Prestifilippo, M., Sciotto, M., and Scollo, S.: Explosive Paroxysmal Events at Etna Volcano of Different Magnitude and
300   Intensity Explored through a Multidisciplinary Monitoring System, Remote Sens. – Basel., 14(16), 4006, https://doi.org/10.3390/rs14164006, 2022.

Canitano, A., Mouyen, M., Hsu, Y.-J., Linde, A. T., Sacks, S. I., and Lee, H.-M.: Fifteen Years of Continuous High-Resolution Borehole Strainmeter Measurements in Eastern Taiwan: An Overview and Perspectives, Agid. Geo. In., 2(3), 172–195, https://doi.org/10.3390/geohazards2030010, 2021.

305   Carleo, L., Bonaccorso, A., Currenti, G., and Sicali, A.: Near real-time filtering of high precision borehole strainmeter signals for volcano surveillance, Ann. Geophys. – Italy, 65(4), https://doi.org/10.4401/ag-8787, 2022a.

Carleo, L., Currenti, G., Bonaccorso, A., and Sicali, A.: Real-time signal processing of high-precision borehole strainmeters at Mt. Etna for volcanic surveillance and eruptive events detection, Nuovo Ciment. C, 45(6), https://doi.org/10.1393/ncc/i2022-22185-5, 2022b.

310   Cassidy, M., Manga, M., Cashman, K., and Bachmann, O.: Controls on explosive-effusive volcanic eruption styles, Nat. Commun., 9 (2839), https://doi.org/10.1038/s41467-018-05293-3, 2018.

Cirillo, F., Avvisati, G., Belviso, P., Marotta, E., Peluso, R., and Pescione, R. A.: Clustering of Handheld Thermal Camera Images in Volcanic Areas and Temperature Statistics, Remote Sens. – Basel., 14(15), 3789, https://doi.org/10.3390/rs14153789, 2022.

315   Corradino, C., Amato, E., Torrisi, F., Calvari, S., and Del Negro, C.: Classifying major explosions and paroxysms at stromboli volcano (Italy) from space, Remote Sens. – Basel. , 13(20), 4080, https://doi.org/10.3390/rs13204080, 2021.

Currenti, G., and Bonaccorso, A.: Cyclic magma recharge pulses detected by high-precision strainmeter data: the case of 2017 inter-eruptive activity at Etna volcano, Sci. Rep. – U.K., 9(1), 1–7, https://doi.org/10.1038/s41598-019-44066-w, 2019.

Fränti, P., and Sieranoja, S.: How much can k-means be improved by using better initialization and repeats? Pattern Recogn.,
320   93, 95–112, https://doi.org/10.1016/j.patcog.2019.04.014, 2019.

Currenti, G., Zuccarello, L., Bonaccorso, A., and Sicali, A.: Borehole Volumetric Strainmeter Calibration From a Nearby Seismic Broadband Array at Etna Volcano, J. Geophys. Res. Sol.- Ea., 112(10), 7729–7738, https://doi.org/10.1002/2017JB014663, 2017.

Langer, H., Falsaperla, S., Masotti, M., Campanini, R., Spampinato, S., and Messina A.: Synopsis of supervised and
325   unsupervised pattern classification techniques applied to volcanic tremor data at Mt Etna, Italy. Geophys. J. Int., 178(2), 1132–1144, https://doi.org/10.1111/j.1365-246X.2009.04179.x, 2009.

Langer, H., Falsaperla, S., and Hammer, C.: Advantages and Pitfalls of Pattern Recognition - Selected Cases in Geophysics, Elsevier, https://doi.org/10.1016/B978-0-12-811842-9.00003-0, 2020.

Linde, A. T., and Sacks, S. I.: Continuous monitoring of volcanoes with borehole strainmeters, In J. M. Rhodes and J. P.
330   Lockwood (Eds.), Mauna Loa Revealed: Structure, Composition, Hystory, and Hazards, American Geophysical Union, 92, 171–185, Washington D. C. https://doi.org/10.1029/GM092p0171, 1995.

Linde, A. T., Agustsson, K., Sacks, S. I., and Ragnar, S.: Mechanism of the 1991 eruption of Hekla from continuous borehole strain monitoring, Nature, 365, 737–740, https://doi.org/10.1038/365737a0, 1993.

Linde, A. T., Kamigaichi, O., Churei, M., Kanjo, K., and Sacks, S. I.: Magma chamber recharging and tectonic influence on
335   reservoirs: The 1986 eruption of Izu-Oshima, J. Volcanol. Geoth. Res., 311(2016), 72–78, https://doi.org/10.1016/j.jvolgeores.2016.01.001, 2016.

Lloyd, S. P.: Least Squares Quantization in PCM. IEEE Transactions on Information Theory, 28(2), 129–137, https://doi.org/10.1109/TIT.1982.1056489, 1982.

Kaufman, L., and Rousseeuw, P. J.: Finding groups in data: an introduction to cluster analysis, John Wiley and Sons, 2009.

340   MacQueen, J.: Some methods for classification and analysis of multivariate observation, 5th Berkeley Symposium on Mathematical Statistics and Probability, 281–297, 1967.

NASEM - National Academies of Sciences, Engineering, and Medicine: Volcanic Eruptions and Their Repose, Unrest, Precursors, and Timing, Washington, DC: The National Academies Press, https://doi.org/10.17226/24650, 2017.

Nunnari, G.: Clustering activity at Mt Etna based on volcanic tremor: A case study, Earth Sci. Inform., 14(3), 1121–1143,
345   https://doi.org/10.1007/s12145-021-00606-5, 2021.

Roeloffs, E. A., and Linde, A. T.: Borehole observations of continuous strain and fluid pressure, In Volcano deformation,
305–322, https://doi.org/10.1007/978-3-540-49302-0_9, 2007.

Romano, P., Di Lieto, B., Scarpetta, S., Apicella, I., Linde, A. T., and Scarpa, R.: Dynamic strain anomalies detection at
Stromboli before 2019 vulcanian explosions using machine learning, Front. Earth Sci., 10, 862086,
350   https://doi.org/10.3389/feart.2022.862086, 2022.

Sacks, S. I., Suyehiro, S., Evertson, D. W., and Yamagishi, Y.: Sacks-Evertson Strainmeter, Its Installation in Japan and
Some Preliminary Results Concerning Strain Steps, Pap. Meteorol. Geophys., 22(3–4), 195–208, 1971.

Unglert, K., Radić, V., and Jellinek, A. M.: Principal component analysis vs. self-organizing maps combined with
hierarchical clustering for pattern recognition in volcano seismic spectra, J. Volcanol. Geoth. Res., 320, 58–74.
355   https://doi.org/10.1016/j.jvolgeores.2016.04.014, 2016.

Voight, B., Linde, A. T., Sacks, S. I., Mattioli, G. S., Sparks, R. S. J., Elsworth, D., Hidayat, D., Malin, P. E., Shalev, E.,
Widiwijayanti, C., Young, S. R., Bass, V., Clarke, A., Dunkley, P., Johnston, W., McWhorter, N., Neuberg, J., and Williams,
P.: Unprecedented pressure increase in deep magma reservoir triggered by lava-dome collapse, Geophys. Res. Lett., 33(3),
L03312. https://doi.org/10.1029/2005GL024870, 2006.