Supplemental Materials: Temporal models for the occurrence of Etna eruptions

7.1 Data

In the Data availability section we provide the link to two csv files containing the onset time of the events considered in this paper:

1. File OnsetTimes_flank.csv is a 1-column csv file containing the year (Current Era, CE) of the known flank eruptions since 1760 CE, taken from Proietti and Branca (2024).

2. File OnsetTimes_paroxysm.csv is a 6-column csv file containing the year (CE), the month, the day, and, when known, the hour, the minutes and the seconds, of the known summit paroxysms from Andronico et al. (2021), Calvari et al. (2018), Calvari and Nunnari (2022) and Mereu et al. (2023).

7.2 k-means analysis of the summit paroxysms catalog

In order to identify an adequate number of clusters k that optimally partitions a dataset is a challenging task (Garcia-Aristizabal et al., 2017; Sandri et al., 2021). Here, we use the common Silhouette Method (Rousseeuw, 1987; Xu and Wunsch, 2005; Al- Zoub and al Rawi, 2008; Kaufman and Rousseeuw, 2008), based on the mean Silhouette coefficient. The method tends to prefer k values that partition the data in clusters where the average distance between each element and the elements belonging to other clusters is much larger than the average distance between elements belonging to the same cluster.

In figure S1 we show the mean silhouette coefficient as a function of the number of clusters k. The figure shows that the partition of the 259 events into k=12 clusters is the best one.

In figure S2b we show the silhouette coefficient specifically for each element divided into the 12 clusters (clusters are given as groups of horizontal blue bars; each blue bar marks the silhouette coefficient value for that event in this partition). We see that for clusters labelled 4 and 7 (respectively the second and fourth last in the temporal order) the partitioning is somehow poor: for example, in cluster labelled 4 (marked by yellow stars in panel a) the last event (i.e., the last yellow star in panel a, occurring just before the onset of the last cluster starting in the end of 2020) has even a negative silhouette value. However, for the remaining clusters the partitioning is quite good, as all the elements have positive silhouette coefficients and quite close to the average value (marked by a red bar).
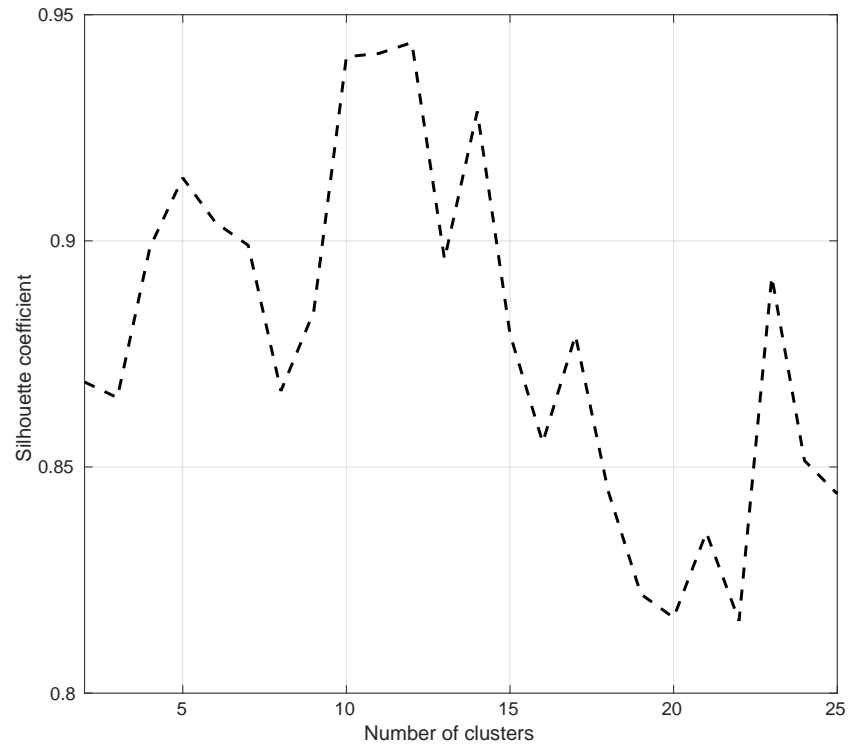
Figure S1: Mean silhouette coefficient (y-axis) as a function of the number k of clusters. Each point shows the mean silhouette coefficient among the 259 events when the best partition into k clusters is performed.
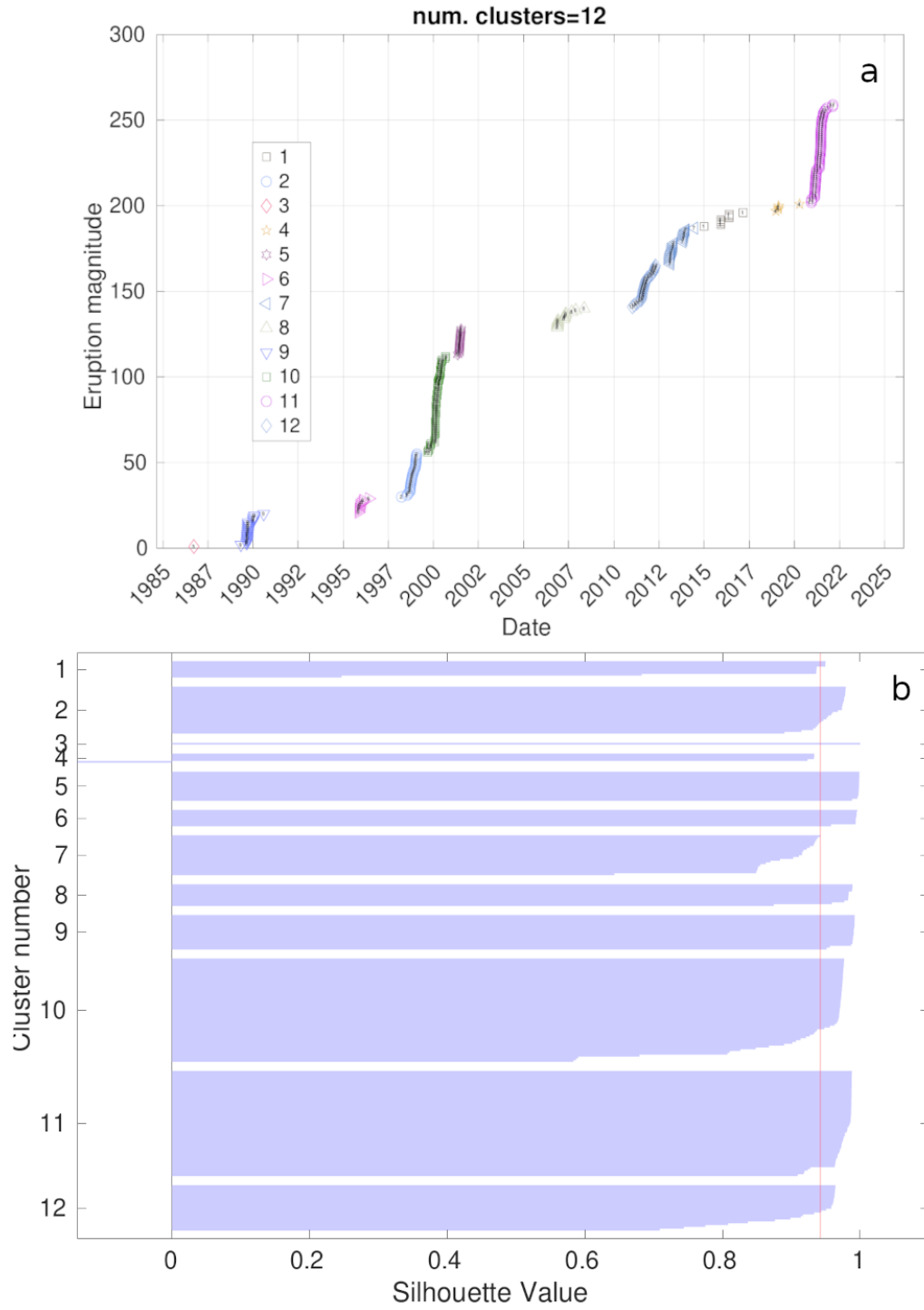
Figure S2: Panel a: result of the cluster analysis for $k=12$ (which is the best partition according to figure S1) in terms of ECDF of paroxysmal events (same as in Figure 3 in main text), but showing the different symbols and colors associated to the clusters' numbering used in panel b (for example, cluster labelled "4" is shown by yellow stars and is the second last in the temporal order). Panel b: blue horizontal bars show the silhouette coefficient for each of the 259 events when partitioned into $k=12$ clusters. The clusters are show as groups of horizontal blue bars. The red line shows the mean silhouette coefficient for this partitioning.