



OPEN

# On catching the preparatory phase of damaging earthquakes: an example from central Italy

Matteo Picozzi<sup>1✉</sup>, Antonio G. Iaccarino<sup>1</sup>, Daniele Spallarossa<sup>2</sup> & Dino Bindi<sup>3</sup>

How, when and where large earthquakes are generated remain fundamental unsolved scientific questions. Intercepting when a fault system starts deviating from its steady behavior by monitoring the spatio-temporal evolution and dynamic source properties of micro-to-small earthquakes can have high potential as tool for identifying the preparatory phase of large earthquakes. We analyze the seismic activity that preceded the Mw 6.3 earthquake that hit L'Aquila on 6 April 2009 in central Italy, and we show that the seismic catalog information can be transformed into features allowing us to track in a statistical framework the spatio-temporal evolution of seismicity. Features associated to foreshocks show different patterns from the background seismicity that occurred in the previous years. We show that features ensemble allows to clearly capture the activation phase of the main event. Nonetheless, foreshocks share similar clustering properties of previous seismic sequences not culminating in large earthquakes, and thus generating questions on their use as potential precursor for earthquake sequences prone to evolve into catastrophic sequences.

Among natural phenomena, earthquakes are one of the most impressive. Large earthquakes have strong impact on society, not only due to the loss of human lives, but also because the domino effects on our globalized society causes billions of euros of damage. Retrospective and theoretical studies on megathrust earthquakes have shown that patterns in seismicity and crustal deformation precede large earthquakes<sup>1–8</sup>.

The complex multi-scale generation process of large earthquakes might have different dominant features depending on the tectonic environment, such as foreshocks or slow slip events and creep phenomena<sup>9,10</sup>. Laboratory<sup>11,12</sup> and simulation<sup>13</sup> results suggest the occurrence of foreshocks being primarily related to structural and stress heterogeneities over rupture surfaces (i.e., where stress heterogeneity is higher, the foreshocks activity is the more prominent). Upscaling laboratory results to crustal faults, we can figure out that on faults with few stress heterogeneities, the preparatory process is driven by seismic and aseismic slip around a nucleation zone (i.e., aseismic model). On the contrary, on faults with diffuse stress and surface heterogeneities, we can find foreshocks triggering each other in a cascading process (i.e., cascade model). When foreshocks occur at a large and stressed asperity, the process can culminate in a large earthquake.

More in general, the large earthquake generation process is proposed being related to a progressive localization of shear deformation around a rupture zone that progressively evolves into a final rapid loading (i.e., generating the small magnitude earthquakes called foreshocks) of a crustal volume localized nearby the hypocenter of the major dynamic rupture<sup>14</sup>. Similar patterns of damage evolution have also been observed by studying acoustic emissions during triaxial tests on rock samples<sup>11</sup>, suggesting that the process generating earthquakes may be universal.

However, the non-systematic foreshocks appearance and the lack of systematic precursory patterns in seismicity and ground deformation in tectonic context different from megathrust areas are demonstrating us that the background physical processes generating large earthquakes are not fully understood yet<sup>14</sup>. Therefore, the prediction of large magnitude earthquakes remains an unresolved fundamental scientific question that needs to be investigated.

Thanks to the strategies implemented by the seismological community in establishing dense seismic networks to monitor regions known to be prone to large earthquakes, standardizing formats for data transmission and archiving, and creating open data repositories for sharing real-time and archived data streams, nowadays it is greatly improved our possibility to study the multi-scale (spatial and temporal) generation process of large earthquakes, and new scientific avenues have been opened. A meta-analysis of foreshocks data<sup>15</sup> has shown that a preparatory phase is potentially identified when seismic catalogs are complete for at least three magnitude

<sup>1</sup>University of Naples Federico II, Naples, Italy. <sup>2</sup>DISTAV, University of Genoa, Genoa, Italy. <sup>3</sup>Helmholtz Centre Potsdam, GFZ German Research Centre for Geosciences, Potsdam, Germany. ✉email: matteo.picozzi@unina.it

units less than the mainshock unit. It must keep in mind, however, that ‘foreshock’ is merely a label assigned to earthquake retrospectively.

For unveiling if preparatory processes are ongoing, instead of trying to establish whether an event is a foreshock or not, Picozzi and Iaccarino<sup>16</sup> proposed to study the spatial and temporal collective behavior of small magnitude earthquakes. Applying this concept to microseismicity from The Geysers geothermal field in California, where the seismicity was characterized by features carrying information about crustal stress conditions, it was shown as a recurrent neural network could identify the preparatory phase of moderate earthquakes<sup>16</sup>, which raises hope for identifying the preparatory phase of future large earthquakes.

Although some success on intercepting the preparatory phase of earthquakes starts to be achieved, recent analyses on two large Italian earthquakes (i.e., the Mw 6.3 L'Aquila 2009 and the Mw 6.1 Amatrice 2016), which occurred at few tens of kilometers of distance and in the same extensional environment in the Apennines in central Italy, have highlighted important differences in their initiation<sup>17,18</sup>, with a clear preparatory phase identified for the former and only a long-lasting quiescence without a clear activation phase for the second. The latter examples highlight the enigmatic nature of the preparatory phase of large earthquakes, which also in case of nearby earthquakes can result be dominated by different driving mechanisms.

Although not very large in magnitude, the Mw 6.3 L'Aquila 2009 (hereinafter, AQU), Italy, earthquake is very well known within the scientific community and one of the most studied recent earthquakes in the literature (International Seismological Centre 2022, On-line Event Bibliography, <https://doi.org/10.31905/EJ3B5LV6><sup>19</sup>). We find AQU interesting because previous studies proposed its preparatory phase being characterized by both the presence of foreshocks clustered near the nucleation area of the mainshock<sup>20–23</sup> and by a slow-slip event<sup>24</sup>, despite the latter is still debated<sup>25</sup>.

Studies of this preparatory phase from different perspectives have highlighted a foreshocks migration towards the mainshock nucleation point<sup>26</sup>, *b*-value changes associated to change in the stress level on the fault<sup>27</sup>, changes in the elastic properties of the medium<sup>28</sup>, and different source properties between foreshocks and aftershocks<sup>29</sup>. A more detailed analysis on the spatio-temporal evolution of foreshocks source properties approaching the mainshock was made possible thank to an innovative service for the Rapid Assessment of Seismic Moment and Radiated Energy in Central Italy (RAMONES<sup>30</sup>), which innovatively allows studying the small magnitude seismicity by direct estimates of their seismic moment and seismic radiated energy. Picozzi et al.<sup>17</sup> analyzed the temporal evolution of radiated energy and size of small magnitude earthquakes preceding AQU and showed that during the final activation phase preceding the mainshock the foreshocks had dynamic characteristics distinct from those of normal rate (background) seismicity.

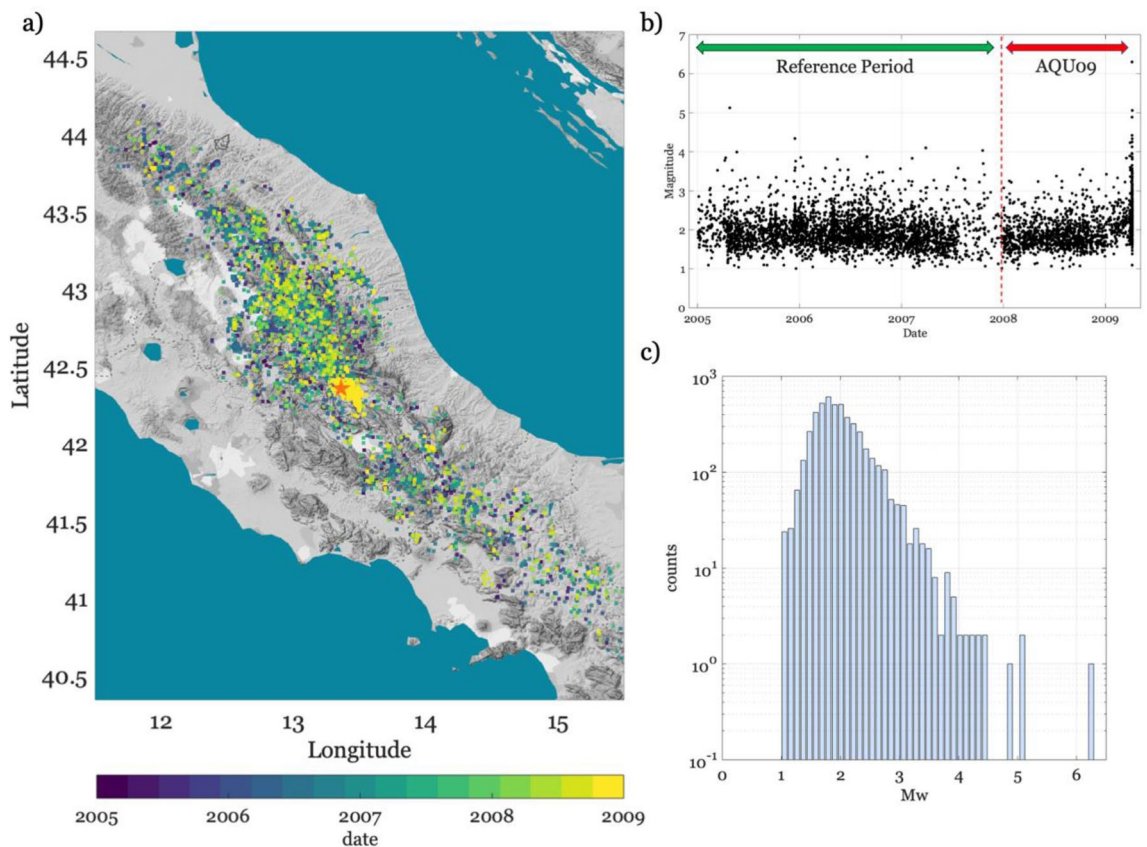
The key idea here is to retrospectively study the earthquakes that have preceded AQU to explore our capability to outline the evolution of fault loading processes. The main scientific question that we face is if we can identify trends in the spatio-temporal evolution and dynamic properties of seismicity that highlight changes with respect to the crustal background activity. In other words, can we catch the preparatory phase of large magnitude events by looking at tiny earthquakes?

We reprocessed thousands of earthquake recordings that have occurred in the Apennine region since 2005, creating an innovative high-resolution seismicity catalog (see supplemental material) with information on both the spatial and dynamic properties of seismic sources. These pieces of information are transformed into features representing the spatio-temporal clustering and the dynamic characteristics of the earthquakes. Then, we analyze and exploit the estimated features to comprehensively characterize the evolving pattern of seismicity. To this purpose, we set up a probabilistic framework by which we explore the differences between the spatio-temporal and dynamic source properties of microseismicity approaching the mainshock with respect to those of the background and clustered seismicity that occurred in the past.

## Results

**Overview of seismicity features.** We retrieved data from publicly available databases and identified 4820 earthquakes occurred between the 1 of January 2005 until the occurrence of AQU on 6 April 2009 (Supplementary material). The earthquakes range between magnitude Mw 1 and Mw 6.3, hypocentral depths between 0.3 and 29 km, and are distributed along the central Apennines (Fig. 1). Uncertainties in event location are mostly within 1 km both horizontally and vertically (Fig. S1). We first focus on a subset of data (i.e., between 2005 and 2007) to define reference models for the spatio-temporal evolution of seismicity. Hereinafter, we refer to this subset as ‘reference period’. Considering that seismicity can be seen as composed by two main populations, that is the background seismicity associated to tectonic stress field and the clustered one having different origin (e.g., foreshocks, aftershocks, swarms), we apply a clustering analysis<sup>31,32</sup> and we model the distribution with a sum of a log-Gaussian function<sup>33</sup> for discriminating the two populations (hereinafter, defined as background, B, and clustered, C, Figs. S2 and S3).

We characterize the seismicity in terms of a set of twelve physically based features varying in time and describing different aspects of the temporal and spatial evolution of seismicity: the *b*-value of the Gutenberg–Richter law<sup>34</sup>; the fractal dimension of hypocenters, *D<sub>c</sub>*<sup>35</sup>; the generalized distance between pairs of earthquakes,  $\eta$ , and its space and time components ( $R\eta$ ,  $T\eta$ , respectively), rate,  $\rho$ , and moment rate,  $\dot{M}_0$ <sup>36</sup>, the Shannon's information entropy,  $H$ <sup>37</sup>; the effective stress,  $\Delta\sigma_e$ <sup>38</sup>, the volume,  $V$ , by the 3D convex Hull of the hypocenters in a given time window, the Kostrov strain,  $\Delta\varepsilon$ <sup>39</sup>, and the Energy Index, EI<sup>17</sup> (see “Materials and methods” section for details about the computation of seismic features). We estimate the uncertainty associated with the features by applying a bootstrap approach<sup>40</sup>, repeating at each time instant the features computation with 200 random sampling realizations of the original dataset with replacement. This analysis is carried out for *b*-value, *D<sub>c</sub>*,  $\dot{M}_0$ ,  $H$ ,  $\Delta\sigma_e$ ,  $\Delta\varepsilon$ , and EI (Fig. 2), while we exclude from the bootstrap analysis the remaining five features that are function of the earthquake location and origin time only ( $\eta$ ,  $R\eta$ ,  $T\eta$ ,  $\rho$ , and  $V$ ). The features are computed also for the seismicity

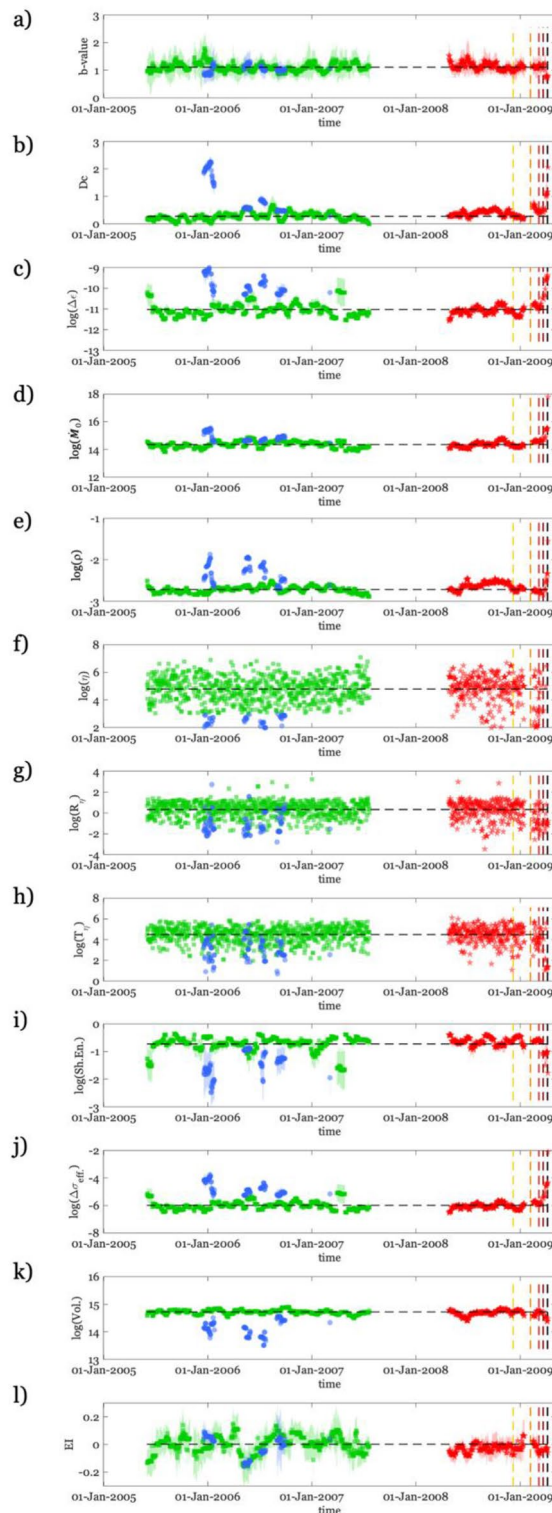


**Figure 1.** Overview of the 2005–2009 seismicity in central Italy. (a) location of the earthquakes colored per date of occurrence, and epicenter of the Mw 6.3 LAquila 2009 shown as red star. (b) distribution of magnitude in time. The reference and the AQU09 periods are highlighted and separated by a vertical dashed red line. (c) Histogram showing the distribution of Mw.

occurred from the 1 January 2008 until AQU (the latter period is indicated as AQU09). For AQU09, we do not discriminate between B and C seismicity. Indeed, we consider AQU09 as a ‘testing dataset’ and, after having characterized the B and C populations, we look for deviations in AQU09 with respect to B and C that could hint for the preparation process of the mainshock.

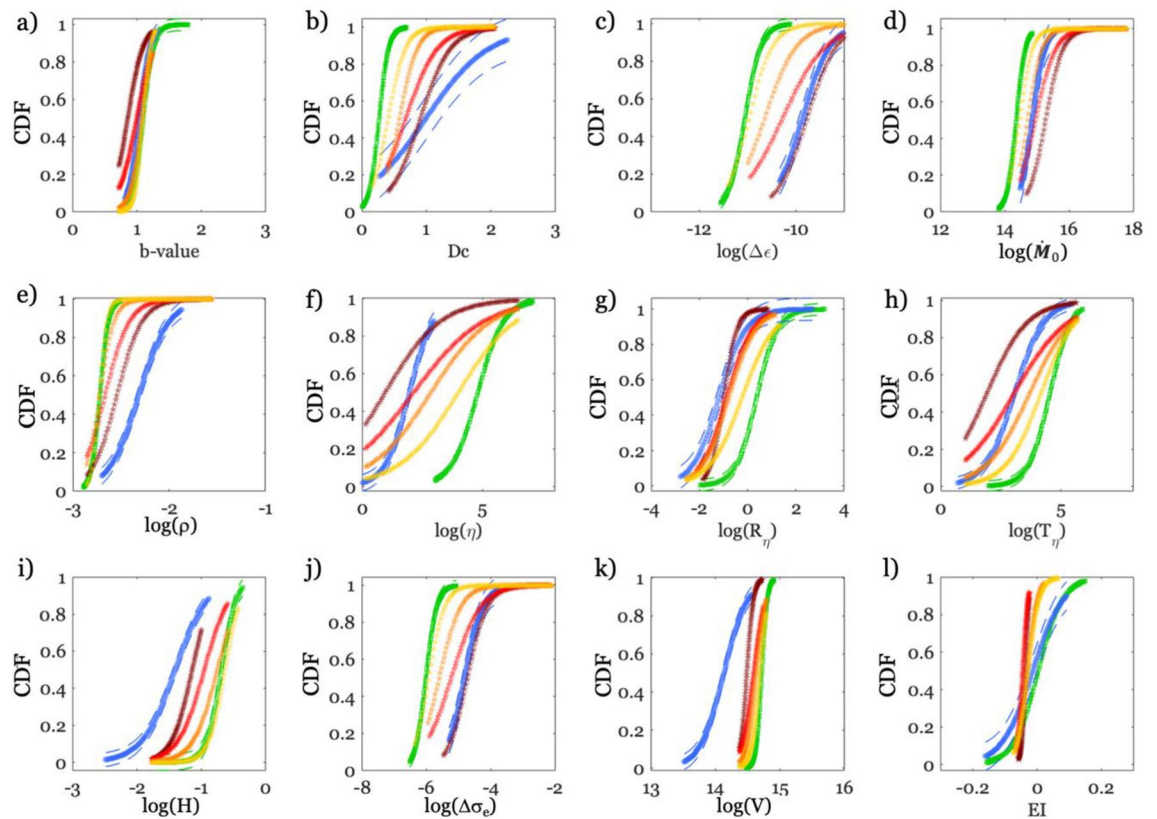
The temporal evolution of the features shown in Fig. 2 is obtained by considering events with magnitude equal or larger than  $M_w$  1.5. We also investigated the effect of different cutoff magnitudes ( $M_{cut}$ ) in computing the features considering the AQU09 period (i.e.,  $M_w$  1.5 Fig. S4,  $M_w$  1.6 Fig. S5,  $M_w$  1.7 Fig. S6,  $M_w$  1.8 Fig. S7,  $M_w$  1.9 Fig. S8). The gradual decrease in the number of events does not seem to influence the main trends in the temporal evolution of the features, in particular the rapid changes in the patterns during the activation phase. We compute for all the features and for different cutoff magnitudes the empirical cumulative density function, ECDF (Fig. S9) and we fit them with logistic functions (CDF in Fig. S10). The ECDF and CDF seem not affected in their trend and statistical properties with the change in  $M_{cut}$ . Nevertheless, comparing the features for  $M_{cut}$  1.5 (Fig. S4) and  $M_{cut}$  1.9 (Fig. S8), we see that lowering the number of data (i.e., higher  $M_{cut}$ ) would make any conclusion about the temporal evolution of the preparatory process less robust. We also verify the implications of different area cut-offs. We thus split the background seismicity in three subsets according to the geographical distribution of the earthquakes along the Apennines (Fig. S11a, with epicenters as black points in the Northern sector, as red points in the Central one, and as blue points in the Southern one). The ECDF and PDF for the three subsets are shown in Figs. S12, S13, S14. Their overlapping indicates that for the considered areas, the features do not present statistically significant differences. The same kind of analysis for the clustered seismicity is not possible due to the lower numerosity of data in the considered period (Fig. S11b), but it will be carried out in the next future for the period 2009–2023. Concerning the reference period, looking at Fig. 2, we find that a high fraction of seismicity is classified as background (green), with respect to which many features for both the clustered one (blue) and the seismicity preceding AQU09 (red) show distinct values and trends (Fig. 2). To highlight the existence of differences between the three populations, we compute their ECDF (Fig. S15) and CDF (Fig. 3) fitting them with logistic functions (Figs. S16 and S17). For the AQU09 period, we compute the CDF considering different time periods preceding the mainshock (i.e., 120, 60, 30, 15 days before AQU that are shown as yellow, orange, red and dark red in Fig. 3, respectively).

The B and C populations show distinct CDFs trend for most of the features (Fig. 3), except for the b-value and EI for which the CDFs appear overlapping. Considering that both EI and b-value have been put in relation to the stress level on the fault<sup>18,26</sup>, the CDF similarity for these two features suggest that background seismicity



**Figure 2.** Overview of the seismic features in time. From (a) to (l) we show the evolution in time of each feature, where for  $b$ -value,  $D_c$ ,  $M_w$ ,  $H$ ,  $\Delta\sigma$ ,  $\Delta\epsilon$ , and EI we show the mean value  $\pm$  the standard error (vertical bar). We highlight with different colors the Background events (green), the Clustered ones (blue), and the AQU09 seismicity considered for identifying the preparatory process (red). The colored vertical dashed lines indicate 120 days yellow, 60 days orange, 30 days light red, 15 days dark red before the mainshock.





**Figure 3.** Overview of the Cumulative Density Function (CDF) computed for the different features and seismicity belonging to different periods. From (a) to (l) we show for each feature the CDF with solid lines and  $\pm 1$  std. with dashed lines for the background seismicity in green, the clustered one in blue, and those of different time periods before the Mw 6.3 L'Aquila earthquake (120 days yellow, 60 days orange, 30 days light red, 15 days dark red).

and past seismic sequences, which did not culminate in large magnitude earthquakes, occur on small scale faults with similar properties, and that both static and dynamic stresses should reach higher values over larger spatial scales to sustain the development of large ruptures. We also find interesting the way the CDFs for AQU09 change with different time periods preceding the mainshock (from yellow to dark red in Fig. 3). We note cases where the AQU09-CDFs initially resemble the B-CDFs but progressively become like the C-CDFs (i.e., for  $\Delta\epsilon$ ,  $\eta$ ,  $R\eta$ ,  $T\eta$ ,  $\Delta\sigma_e$ ). This result highlights a progressive clustering of the events approaching the activation phase. All that features are indeed related to the events time and space occurrence.

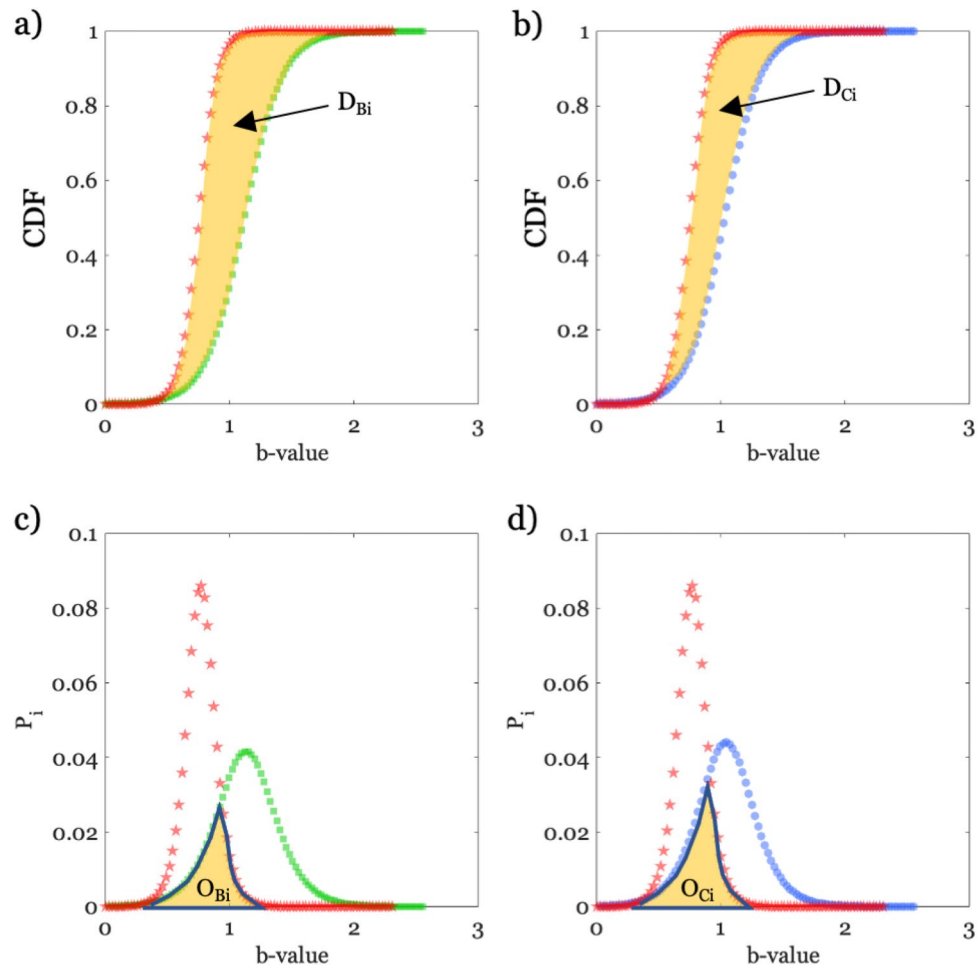
Furthermore, we also see other AQU09-CDFs that are initially similar to the B-CDFs, but then becoming distinct from both B and C CDFs (i.e.,  $D_c$ ,  $\hat{M}_0$ ,  $r$ ,  $H$ ,  $V$  and  $EI$ , which are related to the temporal evolution of dynamic source properties).

**Temporal evolution of features.** We aim to measure the temporal evolution of differences among CDF for different event populations. Our goal is to verify if deviations in source and spatial characteristics for events belonging to the AQU09 series with respect to those of background and clustered earthquakes are measurable. We thus track the relation between AQU09's CDFs with respect to those for B and C in time.

We explore two ways for quantifying the differences among CDFs: (1) the two sample Cramer-von Mises criterion used in statistic test<sup>41</sup>; (2) a distribution-free overlapping measure<sup>42</sup> (as example, Fig. 4 shows the two difference measures for the b-value). In few words, by the first approach we measure the cumulative distance (D) between CDFs, while by the second one we measure the overlapping (O) between PDFs derived by CDFs (see “Materials and methods” section).

The parameters D and O are computed for all the features and considering: (i) the whole population of events for B and C; (ii) a moving window with 30 days width that moves of 1 day at time for AQU09. We thus obtain time series of D and O values with respect to both B and C (i.e., we get  $D_B$  and  $D_C$ , Fig. 4a and b, and  $O_B$  and  $O_C$ , Fig. 4c and d).

The temporal evolution of  $D_B$  for all features is shown in Fig. 5a, where we plot only the last part of the reference period for better representing the preparatory phase during AQU09. During the reference period (i.e., before 2009), the features have both variable  $D_B$  amplitudes and incoherent trends. Conversely, during AQU09, we observe a progressive increase in  $D_B$  for all features while approaching the mainshock. The  $D_B$  trends are compliant with what we expected. At the initial stage of AQU09, the CDFs are like those of the reference period



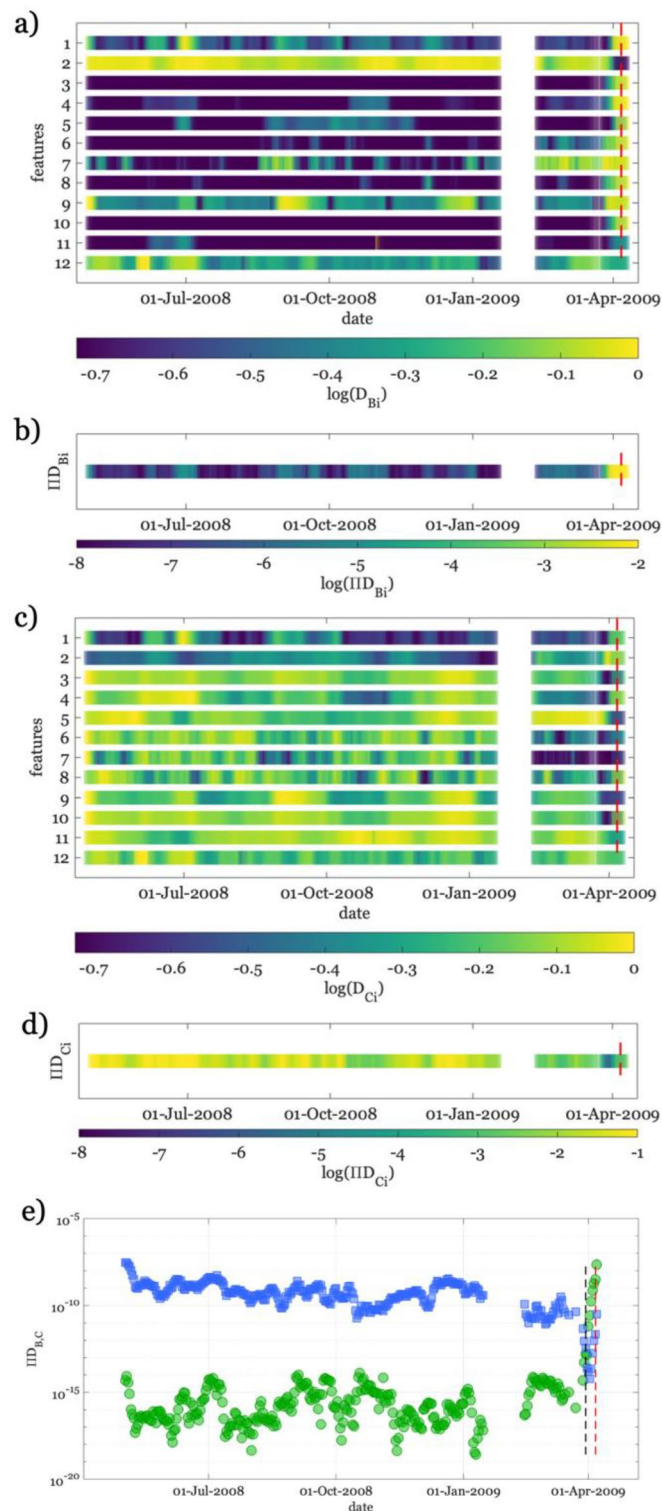
**Figure 4.** Outline of the criteria adopted to measure the distance between CDFs. (a) CDFs distance according to the Cramer-von Mises criterion ( $D_B$ ) with CDF relevant to the b-value for background in green and that of the AQU09 period in red at the generic time window  $i$ . (b) The same as (a), but with the CDF for clustered seismicity in blue ( $D_C$ ). (c) similar to (a), but considering the overlapping measure ( $O$ ) between probability density functions (PDF). (d) The same as (c), but for clustered seismicity.

( $D_B$  is very low). Conversely,  $D_B$  progressively increases approaching the mainshock's activation phase because during the latter phase the CDFs become dissimilar from those of the reference period.

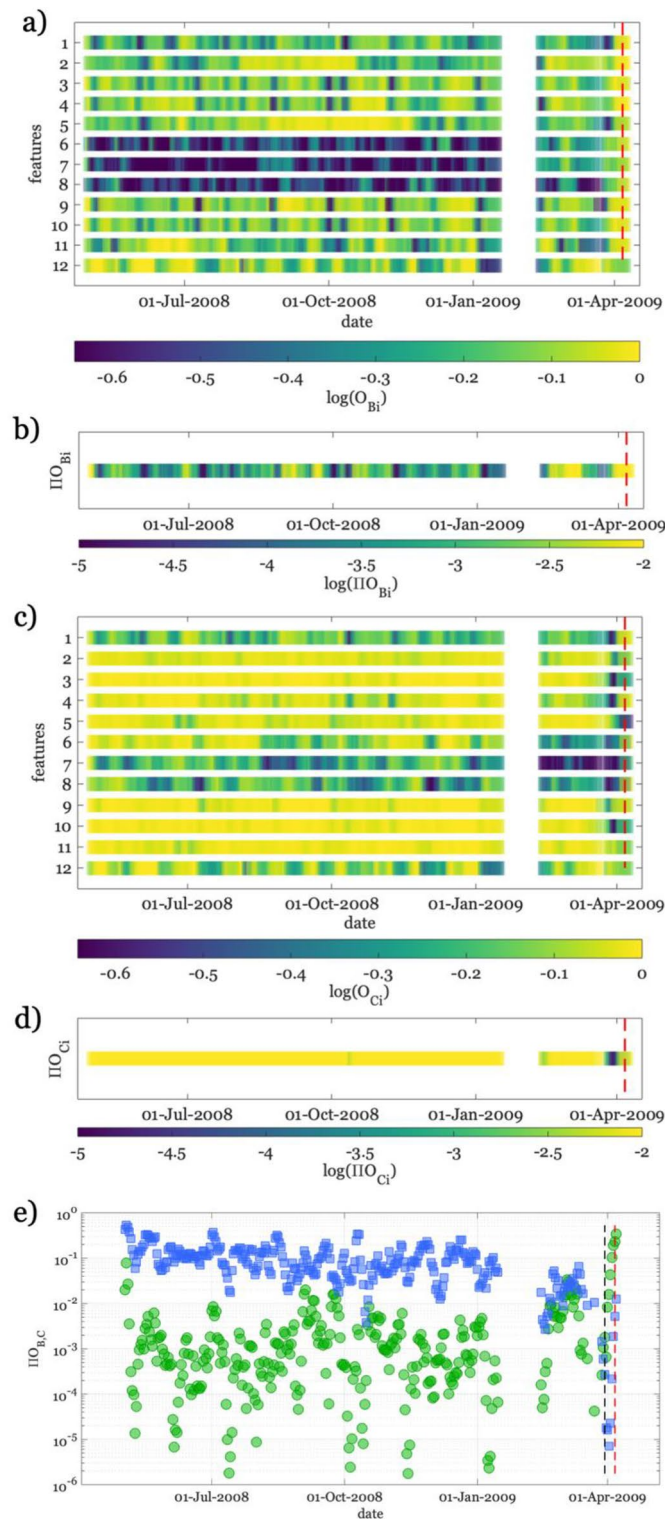
The ensemble of all features,  $\prod D_B$ , shows low values for the whole reference period and an increase before the mainshock (Fig. 5b). A closer look at the temporal evolution of  $\prod D_B$  highlights that during the activation phase (i.e., from the end of March 2009),  $\prod D_B$  increases of six orders of magnitude with respect to the range of values observed for the reference period (i.e., green dots in Fig. 5e).

Results for  $D_C$  (Fig. 5c) show trends complementary to those of  $D_B$ .  $D_C$  trends are in fact characterized by high values for both the reference period and the initial part of 2009. On the contrary, we observe a decrease in  $D_C$  for most of the features when approaching the activation phase. The high  $D_C$  values during the initial stage of AQU09 are clearly due to the similarity between that seismicity with the reference period one (small  $D_B$ ). Conversely, approaching the activation phase the CDFs for AQU09 become progressively similar to the clustered seismicity ones (small  $D_C$ ). Again, the ensemble of all features,  $\prod D_C$ , emphasizes the drop at the activation phase with respect to the previous period (Fig. 5d). We find intriguing the complementary trend of  $\prod D_B$  and  $\prod D_C$  (Fig. 5e), with the former increasing and the latter decreasing at the activation phase. We highlight that in the months preceding the activation phase neither an increase in  $\prod D_B$ , nor a decrease in  $\prod D_C$  of a similar amount is observed. Similar results are obtained considering  $O_B$  and  $O_C$  (Fig. 6), but using them we observe a less clear distinction between the activation and the previous periods. The ensemble of all features with the overlapping parameter (i.e.,  $\prod O_B$  and  $\prod O_C$ ) appears noisier than  $\prod D_B$  and  $\prod D_C$  and less efficient in discriminating the activation phase from the background seismicity (Fig. 6e). For these reasons, in the following, we will focus on  $\prod D_B$  and  $\prod D_C$  only.

We apply a Monte Carlo approach to assess the robustness of observed trends<sup>17,43</sup>. We consider the AQU09 period, and we generate for each feature 1000 random time series of values extracted from their PDF (random samples and time series are shown in Figs. S18, S19, and S20). In comparison with results of Fig. (5), we check how often: (1) the random time series have a slope during the short activation phase equal or larger than that for



**Figure 5.** Temporal evolution of the Cramer-von Mises distance criterion for the CDFs relevant to seismic features. **(a)** For each feature it is shown the base-10 logarithm of the distance ( $D_B$ ) between the CDF for background seismicity and that for AQU09. The features are ordered as follows: 1.  $b$ , 2.  $D_c$ , 3.  $\Delta\epsilon$ , 4.  $\dot{M}_0$ , 5.  $\rho$ , 6.  $\eta$ , 7.  $R\eta$ , 8.  $T\eta$ , 9.  $H$ , 10.  $\Delta\sigma_v$ , 11.  $V$ , 12.  $EI$ . The red dashed line represents the origin time of the mainshock. **(b)** Similar to **(a)**, but showing the ensemble of the CDF distances ( $\prod D_B$ ). **(c)** and **(d)** the same as **(a)** and **(b)**, respectively, but for the clustered seismicity. **(e)** Evolution of  $\prod D_B$  in green and  $\prod D_C$  in blue. The start of the activation phase is represented as black dashed line, while the origin time of the mainshock is shown as red dashed line.



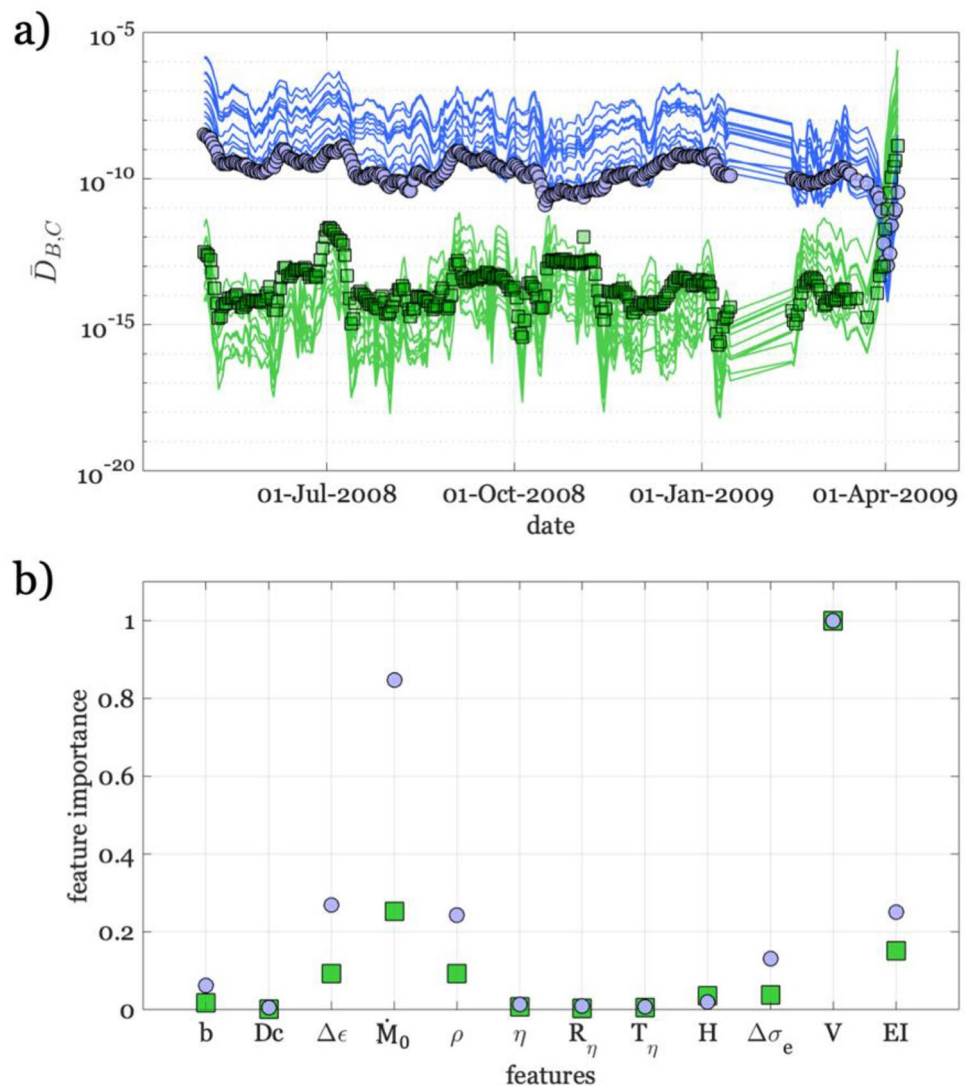
**Figure 6.** The same as Fig. 5, but for the overlapping (O) distance criterion between PDFs.

the reference time series; (2) if the last value is an extreme point (i.e., a maximum or a minimum). In this way, we assess the likelihood that the slope of features during the activation phase was the largest of the time series by chance, and similarly that the last value was an extreme point by chance. Since the activation phase is identified by the ensemble of features, we have also computed the mean of all computed rates (see Table S1). Table S1 shows that it is very unlikely that the observed results arise from noise, suggesting that our analysis results are robust.



**Features importance and best-set selection.** We try next to understand which, among the considered features, is dominating over the others in discriminating the preparation phase. Understanding the features importance would indeed allow us to tailor workflows for the detection of the preparation phase of large earthquakes. Hence, we set up a simple feature importance analysis by measuring the root mean squared error (RMSE) between  $\prod D_B$  for all the  $n$  features (i.e.,  $n=12$ ) and its counterpart obtained excluding one feature at time ( $\prod_{n-1} D_B$ ) (Fig. 7a). Intuitively, when we exclude a feature with low impact on  $\prod D_B$ , the two curves ( $\prod D_B$  and  $\prod_{n-1} D_B$ ) are very similar and the RMSE is low. On the contrary, excluding an important feature would determine different curves and high RMSE. Figure 7b shows the results of this analysis for both  $D_B$  and  $D_C$ , where the RMSEs for the two datasets are normalized to their maximum for facilitating the comparison. Concerning  $D_B$ , we observe that the five most important features are, in descending order  $V$ ,  $M_0$ ,  $EI$ ,  $\rho$ , and  $\Delta\epsilon$ . When we consider  $D_C$ , the most important features are the same, despite it changes their relative importance. Observing that the same features play a significant role for both  $\prod D_B$  and  $\prod D_C$  confirms us that both the characteristics with which the clustering of seismicity occurs and the change in dynamic properties of the foreshocks with respect to the background activity are key elements of the activation phase of the LAquila earthquake.

Besides a classification of the features importance, we also search for a set of features that allows to maximize the difference between the activation phase (i.e., up to 6 days before the mainshock) and the reference period ( $\Delta_{A,RP}$ ) in terms of  $D_B$ , and, at the same time, to minimize the one between the same time periods in terms of  $D_C$ . This analysis aims to verify if it is possible to separate the features in two groups with different sensitivity

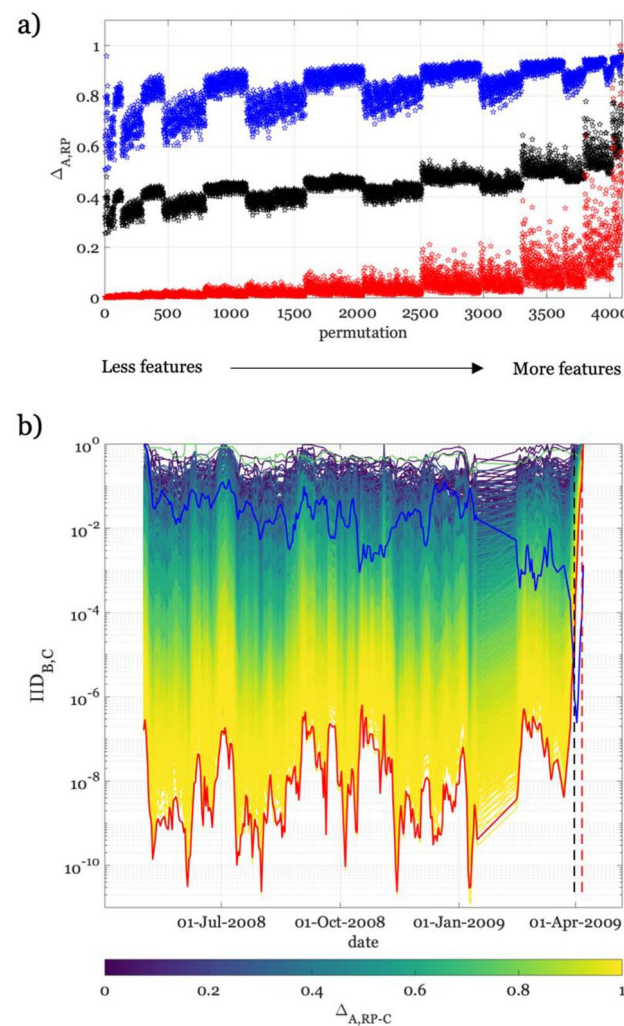


**Figure 7.** Feature importance analysis. (a) Comparison between  $D_B$  curves, whereas the reference one obtained by the ensemble of all features is shown as green squares, while those obtained neglecting one feature at time are represented as green lines. Similarly, the  $D_C$  for the ensemble of all features is shown as blue dots, while those eliminating one feature at time are shown as blue lines. (b) Normalized feature importance for  $D_B$ , shown as green squares, and  $D_C$ , blue dots.

to the clustering phenomena and to the change in dynamic properties. One fascinating hypothesis is that features independent from the spatial organization of the seismicity but sensitive to the activation phase would be informative of the stress level on the fault, and thus about the future size of the main event. Figure 8a shows the two distances  $\Delta_{A,RP}$  (i.e., red for  $D_B$  and blue for  $D_C$ ) with respect to the permutations (i.e., a different number and combination of features), where with the increase of permutation number does also increase the number of features considered. We combine the two  $\Delta_{A,RP}$  measures assigning to them the same importance (i.e., we use their mean,  $\Delta_{A,RP-C}$ , shown as black dots in Fig. 8a). Figure 8b shows all the curves colored per  $\Delta_{A,RP-C}$ , where  $D_B$  (red) and  $D_C$  (blue) for the best combination of features are highlighted (i.e., the case where ten over twelve features are used). Only the fractal dimension (2) and the energy index (12) are excluded from this optimal set of features. These results also suggest that there is no feature able to intercept the activation phase alone. Of course, this conclusion is related to the features considered here; while, hopefully, looking at other foreshocks properties would unveil new information.

## Discussion

The short-term activation of the 2009 L'Aquila earthquake falls into a well-known precursory pattern of large earthquakes<sup>44</sup>. This earthquake seems to comply rather well with the model proposed by Kato and Ben-Zion<sup>14</sup>. Despite our features-based methodology seems potentially useful to intercept the preparatory phase of events



**Figure 8.** Analysis for identifying the set of features maximizing the difference between activation phase and background seismicity. (a) Here, we represent, as function of different number and combination of features, the difference between the ensemble of features during  $D_B$  in the activation phase and in the reference-period is shown as red stars. We also plot the distance for  $D_C$  considering the clustered seismicity and the activation phase, which is shown as blue stars. The average of the two measures is shown as black stars. (b)  $D_B$  curves colored per the difference parameter  $\Delta_{A,RP-C}$ . The curves corresponding to the set of features that at the same time maximize the distance between the activation phase and the background seismicity in  $D_B$  and minimize that one between the activation phase and the clustered seismicity in  $D_C$  are shown as red and blue curves, respectively.

behaving as the 2009 L'Aquila earthquake, it is worth mentioning that different precursory patterns have been observed in the past<sup>8, 45</sup> (e.g., quiescence, accelerating seismic release, doughnut, and event migration inward and outward the nucleation area). Differences in the preparatory phase for the 2009 L'Aquila and the nearby 2016 Amatrice earthquakes highlight that crustal processes leading to large earthquakes are influenced by unforeseeable combination of heterogeneous fault properties, stress interactions and crustal fluids. For instance, it was proposed that the high-pressure deep CO<sub>2</sub>-dominated fluids along the central and southern Apennines<sup>46,47</sup> is responsible for the nucleation of large earthquakes in Italy (e.g., the 1997 Colfiorito, the 2009 L'Aquila, and the 2016 seismic sequences<sup>48,49</sup>).

We cannot generalize our results and it is certainly crucial to carry out further also considering different tectonic context (e.g., among others the Kumamoto, 2016, Valparaiso, 2017, and Ridgecrest, 2019 earthquakes).

Our statistical approach, which is still merely a preliminary workflow, can be easily extended for integrating features derived from other geophysical measurements repeated in time (e.g., geodetic deformation, geochemical parameters). Therefore, we believe that it could become a perfect tool for the multi-disciplinary study of active fault systems at near fault observatories. A similar principle was applied to forecast strong aftershocks in earthquake clusters in northeastern Italy and western Slovenia<sup>50</sup>. Our understanding of the processes occurring in the crust is inevitably hampered by the impossibility of collecting measurements directly within the seismogenic volumes<sup>51</sup>. This is certainly one of the most important limitations we must face while attempting to identify the preparatory process of large earthquakes. To overcome this obstacle, besides establishing denser seismic networks nearby active faults, we must boost up the use of microearthquakes as indicators of the mechanical state of the crust<sup>18,52</sup>. A key advantage of microearthquakes is that they are too small to interact with each other and therefore contribute little to crustal deformation<sup>53</sup>, but their properties and distribution in time and space are sensitive to stress changes. Therefore, in the framework of the Kato and Ben-Zion's model<sup>14</sup>, microearthquakes can help to intercept and to better understand processes occurring within the crust. The increasing availability of augmented seismic catalogs, especially along deep transition zones of megathrusts, or acoustic emissions in case of laboratory stick-slip experiments, is pushing the scientific community in mining data across wide spatio-temporal scales, from which it emerges that the spatio-temporal evolution of microseismicity/acoustic emissions provide information on seismic friction and coupling. The latter pieces of information resulted useful to also predict the occurrence time of laboratory earthquakes, seismic tremor, and slow slip events<sup>54–56</sup>.

Previous retrospective studies investigated specific aspects of the preparation phase of the L'Aquila earthquake, and they highlighted: b-value changes<sup>25</sup>; foreshocks migration towards the nucleation area of the mainshock<sup>24,25</sup>; different source properties between foreshocks and the preceding background seismicity<sup>17</sup>. Differently from previous studies, we have set up a workflow that allows to both characterize the evolution of the microseismicity spatio-temporal characteristics and source properties and then to compare their statistical properties with those of background and clustered seismicity. Instead of focusing on the properties of each single events, we use a probabilistic framework working on subgroups of events evolving in time. Indeed, we believe that only studying the collective spatial patterns and properties of earthquakes we will be able in future to get hints of the proximity to a large rupture (i.e., as recently done in stick-slip laboratory experiments<sup>57,58</sup>).

Our results highlight that not only a feature like the b-value, which is well-known being representative of the stress conditions associated to ruptures<sup>59</sup>, but also other features (e.g., the fractal dimension, the 3D convex Hull volume of the hypocenters, the seismic rate and moment rate, just to mention a few of them) seem able to mark the deviation from the features trend observed during the interseismic period, especially when their ensemble is considered. While the activation phase start (~ 1 week before AQU) is well intercepted, it is worth mentioning that our feature-based probabilistic framework seem rather blind to the long-term variations in seismicity occurrence and spatial organization observed by Sugan et al.<sup>26</sup>. When we consider the 2 months before the mainshock, some of our features seem to delineate a population distinct from that of background seismicity (e.g., the Shannon's information entropy,  $H$ , the effective stress,  $\Delta\sigma_e$ , and the Kostrov strain,  $\Delta\varepsilon$ , in Fig. 3). However, further studies are necessary to clarify if one of these features, or their ensemble, can highlight the presence of such long-term deviation from the background seismicity.

An issue limiting the power of the adopted approach is the minimum magnitude considered in the event selection (i.e.,  $M_w = 1.5$ ), which limits the number of events usable to characterize the spatio-temporal distribution and properties evolution of the microseismicity. The used threshold in magnitude is also due to difficulties in estimating the seismic radiated energy for small magnitude events (used for computing the energy index and the Shannon's entropy). Hence, further efforts will focus on strategies to characterize smaller magnitude events.

We have shown that it is possible to identify deviations of seismic activity from a background level. Hence, studying features describing different physical processes and monitoring their spatio-temporal evolution might represent a powerful tool for intercepting hints of large earthquake preparatory processes. However, we must recognize that our results suggest that we are not able to foresee towards which state the system is evolving (large earthquake vs. seismic sequence with low magnitude earthquakes). In respect to this issue, our level of knowledge seems still limited. For this reason, we believe that it is necessary to carry out systematic studies of source properties for foreshocks and seismic sequences to identify if differences among these two seismic populations exist, and to investigate if some of the features bring information on the size of the future event.

## Materials and methods

**Dataset and features computation.** To characterize the spatio-temporal and source properties evolution of seismicity, we use information extracted from a seismic catalog including 4820 earthquakes occurred in central Italy between the 1 January 2005 until the occurrence of AQU on 6 April 2009 (supplementary material S1). The catalogue includes for each earthquake information about the origin time, the hypocentral location, local magnitude, seismic moment and radiated energy (supplementary material S1). The latter parameters are

estimated following the RAMONES procedure (<http://www.distav.unige.it/rsni/ramones.php><sup>30</sup>), which exploits continuous data streams stored in free repositories (i.e., ORFEUS-EIDA, IRIS, DPC).

**b-value.** The b-value is estimated analyzing the frequency-magnitude distribution by the Gutenberg–Richter law<sup>34</sup>

$$\log N = a - b \cdot M_w, \quad (1)$$

where N is the cumulative number of earthquakes, a and b values are parameters describing the productivity and relative event size distribution). The b-value is obtained by the maximum likelihood approach<sup>60</sup>. Together with the b-value, we retrieve the simultaneous estimate of the completeness magnitude  $M_c$ , which is useful to estimate some of the following features.

**Fractal dimension.** The fractal dimension of earthquake hypocenters,  $D_c$ , is computed applying the correlation integral method<sup>35</sup>:

$$D_c = \lim_{r \rightarrow 0} \frac{\log C_r}{\log(r)}, \quad (2)$$

where r is the radius of a sphere of investigation and  $C_r$  is the correlation integral:

$$C_r = \lim_{n \rightarrow \infty} \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n H(r - |x_i - x_j|), \quad (3)$$

with n indicating the number of data in the analysis window (i.e.,  $n = 200$  events), x the hypocenter coordinates, and H the Heaviside step function  $H(x) = 0$  for  $x \leq 0$  and  $H(x) = 1$  for  $x > 0$ . Finally, the fractal dimension  $D_c$  is estimated as the slope of the best-fit straight line of  $C_r$  versus the distance r in a bi-logarithmic diagram.

**Moment rate.** We compute the moment rate  $\dot{M}_0$ <sup>61</sup> as follows:

$$\dot{M}_0 = \rho M_0 A \frac{b}{1.5 - b} \left[ 10^{(1.5-b)(m_{\max} - m_0)} - 1 \right], \quad (4)$$

where  $\rho$  is the seismic rate of events larger than  $M_c$ ,  $M_0$  and  $m_0$  are the seismic moment and magnitude corresponding to  $M_c$ , A is the area of finite extension including the events (in  $\text{km}^2$ ),  $m_{\max}$  is the largest magnitude in the catalogue.

**Seismic rate.** We compute the seismic rate  $\rho$ <sup>57</sup> considering the number of events  $\Delta N$  with magnitude larger than the completeness magnitude,  $M_c$ , that occurred in a time window  $\Delta T$  in areas of finite extension A

$$\rho = \frac{\Delta N}{(\Delta T \cdot A)}, \quad (5)$$

where  $\rho$  represents the events per day per square kilometers (eqks./ $(\text{day} \cdot \text{km}^2)$ ).

**Effective stress drop of earthquake clusters.** Following Fisher and Hainzl<sup>38</sup>, we compute the effective stress drop of earthquake clusters as follows:

$$\Delta \sigma_e = \frac{7}{16} \frac{\sum M_0}{R^3}, \quad (6)$$

where R is the radius of the 3D convex Hull of the hypocenters and  $\sum M_0$  is the sum of seismic moments in a given time window.

**Nearest-neighbor distance,  $\eta$ , rescaled distance,  $R_\eta$ , and time,  $T_\eta$ .** The nearest-neighbor approach<sup>31,32</sup> computes the generalized distance between pairs of earthquakes,  $\eta$ , from the analysis of the time–space distances between pairs of earthquakes. The parameter  $\eta$  is derived computing the distances in time (i.e., rescaled time,  $T_\eta$ ) and space (i.e., rescaled distance,  $R_\eta$ ) between an event i and its parent j normalized by the magnitude of the parent event as follows:

$$T_{ij} = t_{ij} 10^{-bm_i/2}, \quad (7)$$

$$R_{ij} = (r_{ij})^{D_c} 10^{-bm_i/2}, \quad (8)$$

where m is the magnitude ( $M_w$ ), b is the parameter of the Gutenberg–Richter law, t is the earthquake inter-occurrence time, r is the earthquake distance, and  $D_c$  is the fractal dimension. The values of b and  $D_c$  are changed according to the estimates obtained for the considered window of events.

Finally,  $\eta$  is defined as:



$$\log \eta_{ij} = \log R_{ij} + \log T_{ij}, \quad (9)$$

In our work,  $\eta$ ,  $R_{ij}$  and  $T_{ij}$  are considered as features.

**Shannon's information entropy.** The Shannon entropy<sup>37</sup>, also known as information entropy, provides a measure of the disorder level in a system. We compute the Shannon entropy using a regular 2-D grid (each 1.5 km  $\times$  1.5 km).

We compute the Shannon entropy as

$$H = - \sum_{k=1}^m \frac{e_k}{E_R} \left[ \ln \frac{e_k}{E_R} \right], \quad (10)$$

where  $e_k$  represents a fraction of the total seismic energy  $E_R$  radiated within the  $k$ th cell. We rely here on seismic radiated energy from RAMONES. The  $e_k/E_R$  ratio is assumed to represent an empirical approximation of the probability of the seismic energy radiated in the  $k$ th cell,  $P_k(E_R)$ , with respect to the total radiated seismic energy, conditioned on the total energy radiated.

Equation (10) can therefore be rewritten as

$$H = - \sum_{k=1}^m P_k(E_R) [\ln P_k(E_R)], \quad (11)$$

Therefore, computing  $H$  at a given temporal interval consists of summing up the  $e_k/E_R$  ratio for the entire grid. To allow comparison between different time intervals and to ensure unity total probability,  $H$  is generally normalized to the equipartition entropy  $HE$ , which corresponds to the case where  $E_R$  is uniformly distributed in the cells (i.e., given by the sum of  $E_R$  divided by the number of cells). The normalized information entropy  $h = H/HE$  ranges between 1 and 0, which correspond to the total disorder of the system and the extreme concentration of events, respectively.

The Shannon entropy concept provides hence a useful quantification of the system predictability; where  $h = 0$  suggests the highest level of predictability and  $h = 1$ , on the contrary, suggests high disorder and low predictability.

**Volume.** We estimate the volume,  $V$ , as the 3D convex Hull of the hypocenters in a given time window.

**Kostrov strain.** The Kostrov strain  $\Delta \varepsilon$ <sup>39</sup> is computed as follows:

$$\Delta \varepsilon = \frac{\sum M_0}{2\mu V}, \quad (12)$$

where  $\mu$  is the rigidity module,  $V$  the volume defined by the hypocenters and  $\sum M_0$  is the sum of seismic moments of earthquakes in a given time window.

**Energy Index.** The Energy Index<sup>17</sup>,  $EI$ , is derived considering seismic moment,  $M_0$ , and radiated energy,  $E_R$ , estimates obtained by RAMONES. The seismicity that occurred over the period 2005–2007 and consisting of 461 earthquakes has been used to calibrate a reference scaling model between the base-10 logarithm (indicated as 'log') of  $M_0$ , and  $E_S$ <sup>17</sup>.

$EI$  is then computed for new earthquakes as:

$$EI = \log(E_S) - \log(E_{St}) \quad (13)$$

where  $E_S$  is the experimental estimate for new earthquakes and  $E_{St}$  is the energy value derived from the median  $E_S$ -to- $M_0$  reference scaling model for the  $M_0$  of the experimental earthquakes. Positive  $EI$  values indicate that an earthquake has radiated with respect to the reference model more energy per unit-slip and unit-area (i.e., per seismic moment,  $M_0$ ) than expected. On the contrary, negative  $EI$  values are associated with earthquakes showing an excess of slip or larger rupture area and lower stress drop with respect to what expected from the reference model.

**Outline of the analyses.** We divide the seismic catalog including 4820 earthquakes (supplementary material S1) occurred in central Italy between 1 January 2005 until the occurrence of AQU on 6 April 2009 as follows: (i) we consider earthquakes occurred between 2005 and 2007 (i.e., referred as 'reference period') for calibrating the spatio-temporal and source dynamic properties of events belonging to the background, B, and clustered, C, seismicity (supplementary materials S2 and S3). Therefore, we split the data in two subsets (B and C); (ii) a third subset of data includes the events from 1 January 2008 and until the Mw 6.3, 2009 LAquila earthquake (supplementary material S4).

The discrimination between B and C seismicity in the reference period is carried out by computing the generalized distance,  $\eta$ , of the events, and modeling the  $\eta$  distribution with the sum of two log-Gaussian functions<sup>29</sup> (Fig. S2). By using a threshold  $\eta$  value (i.e.,  $\eta = 3$ ) for discriminating between B from C (Fig. S3), we implicitly accept that we will include a small portion of events belonging to one population into the other.

The third population of events includes all the earthquakes occurred after the 1 January 2008. For the latter, we do not discriminate between background and clustered seismicity because we are interested to analyze all the data for assessing the information provided by features about the preparatory phase of the mainshock.

We characterize the three populations (B, C and AQU09) computing the features on windows of events with fixed length (i.e.,  $l = 30$  days). For each window, the feature values are assigned to the end time of the last day in the window. Windows move of 1 day at time, hence each feature represents a time series. Windows with less than 50 earthquakes are discarded.

We then compute for each feature the Empirical Cumulative Density Function, ECDF, for both the B and C populations (Fig. S15), and we use a generalized linear regression model to fit them with a logistic function (Figs. S16 and S17). The latter analysis is also repeated for the features belonging to the AQU09 period, but in this case considering limited sets of data. For instance, for the results shown in Fig. 3, we adopted time windows of different lengths (i.e., 120, 60, 30, 15 days preceding the mainshock). Differently, during the temporal analysis of features, and their ensemble, shown in Figs. 5 and 6, we used windows with length 15 days moving 1 day at time.

## Data availability

We used data and information retrieved from ORFEUS-EIDA (<https://www.orfeus-eu.org/data/eida/>), IRIS (<https://www.iris.edu/hq/>) and DPC (<http://ran.protezionecivile.it/EN/index.php>). We used data mainly from networks IV (<https://doi.org/10.13127/SD/X0FXnH7QfY>), IT (<https://doi.org/10.7914/SN/IT>) and MN (<https://doi.org/10.13127/SD/fBBBtDtd6q>). Supplemental material includes 8 Figures.

Received: 19 May 2023; Accepted: 29 August 2023

Published online: 01 September 2023

## References

- Rikitake, T. Earthquake precursors. *Bull. Seismol. Soc. Am.* **65**, 1133–1162 (1975).
- Dieterich, J. H. Preseismic fault slip and earthquake prediction. *J. Geophys. Res.* **83**, 3940 (1978).
- Jones, L. M. & Molnar, P. Some characteristics of foreshocks and their possible relationship to earthquake prediction and premonitory slip on faults. *J. Geophys. Res. Solid Earth.* **84**, 3596–3608 (1979).
- Ohnaka, M. Earthquake source nucleation: A physical model for short term precursors. *Tectonophysics* **211**, 149–178. [https://doi.org/10.1016/0040-1951\(92\)90057-D](https://doi.org/10.1016/0040-1951(92)90057-D) (1992).
- Abercrombie, R. E. & Mori, J. Occurrence patterns of foreshocks to large earthquakes in the western United States. *Nature* **381**, 303–307 (1996).
- Dodge, D. A., Beroza, G. C. & Ellsworth, W. L. Detailed observations of California foreshock sequences: Implications for the earthquake initiation process. *J. Geophys. Res. Solid Earth.* **101**, 22371–22392 (1996).
- Lapusta, N. & Rice, J. R. Nucleation and early seismic propagation of small and large events in a crustal earthquake model. *J. Geophys. Res. Solid Earth.* <https://doi.org/10.1029/2001jb000793> (2003).
- Mignan, A. Seismicity precursors to large earthquakes unified in a stress accumulation framework. *Geophys. Res. Lett.* **39**, 21308 (2012).
- Petrillo, G. *et al.* The influence of the brittle-ductile transition zone on aftershock and foreshock occurrence. *Nat. Commun.* **11**, 3010. <https://doi.org/10.1038/s41467-020-16811-7> (2020).
- Socquet, A. *et al.* An 8 month slow slip event triggers progressive nucleation of the 2014 Chile megathrust. *Geophys. Res. Lett.* **44**, 4046–4053. <https://doi.org/10.1002/2017GL073023> (2017).
- Dresen, G., Kwiatek, G., Goebel, T. & Ben-Zion, Y. Seismic and aseismic preparatory processes before large stick-slip failure. *Pure Appl. Geophys.* **177**, 5741–5760 (2020).
- Yamashita, F. *et al.* Two end-member earthquake preparations illuminated by foreshock activity on a meter-scale laboratory fault. *Nat. Commun.* **12**, 4302. <https://doi.org/10.1038/s41467-021-24625-4> (2021).
- Cattania, C. & Segall, P. Precursory slow slip and foreshocks on rough faults. *J. Geophys. Res. Solid Earth* **126**, e2020JB020430. <https://doi.org/10.1029/2020JB020430> (2021).
- Kato, A. & Ben-Zion, Y. The generation of large earthquakes. *Nat. Rev. Earth Environ.* **2**(1), 26–39 (2020).
- Mignan, A. The debate on the prognostic value of earthquake foreshocks: A meta-analysis. *Sci. Rep.* **4**, 4099. <https://doi.org/10.1038/srep04099> (2014).
- Picozzi, M. & Iaccarino, A. G. Forecasting the preparatory phase of induced earthquakes by recurrent neural network. *Forecasting.* **3**, 17–36 (2021).
- Picozzi, M., Spallarossa, D., Iaccarino, A. G. & Bindi, D. *Geophys. Res. Lett.* **49**, e2021GL097382. <https://doi.org/10.1029/2021GL097382> (2022).
- Picozzi, M., Spallarossa, D., Bindi, D., Iaccarino, A. G. & Rivalta, E. *J. Geophys. Res. Solid Earth* **127**, e2022JB025100. <https://doi.org/10.1029/2022JB025100> (2022).
- di Giacomo, D. *et al.* A new ISC service: The bibliography of seismic events. *Seismol. Res. Lett.* **85**, 354–360 (2014).
- Boncio, P. *et al.* Coseismic ground deformation of the 6 April 2009 L'Aquila earthquake (central Italy, Mw6.3). *Geophys. Res. Lett.* <https://doi.org/10.1029/2010GL042807> (2010).
- Terakawa, T., Zoporowski, A., Galvan, B. & Miller, S. A. High-pressure fluid at hypocentral depths in the L'Aquila region inferred from earthquake focal mechanisms. *Geology* **38**, 995–998 (2010).
- Chiaraluce, L., Valoroso, L., Piccinini, D., di Stefano, R. & de Gori, P. The anatomy of the 2009 L'Aquila normal fault system (central Italy) imaged by high resolution foreshock and aftershock locations. *J. Geophys. Res. Solid Earth.* **116**, 12311 (2011).
- Malagnini, L., Lucente, F. P., de Gori, P., Akinci, A. & Munafo, I. Control of pore fluid pressure diffusion on fault failure mode: Insights from the 2009 L'Aquila seismic sequence. *J. Geophys. Res. Solid Earth.* **117**, 5302 (2012).
- Borghi, A., Aoudia, A., Javed, F. & Barzaghi, R. Precursory slow-slip loaded the 2009 L'Aquila earthquake sequence. *Geophys. J. Int.* **205**, 776–784 (2016).
- Amoruso, A., Crescentini, L. & Chiaraluce, L. Surface temperature and precipitation affecting GPS signals before the 2009 L'Aquila earthquake (Central Italy). *Geophys. J. Int.* **210**(2), 911–918. <https://doi.org/10.1093/gji/ggx210> (2017).
- Sugan, M., Kato, A., Miyake, H., Nakagawa, S. & Vuan, A. The preparatory phase of the 2009 Mw 6.3 L'Aquila earthquake by improving the detection capability of low-magnitude foreshocks. *Geophys. Res. Lett.* **41**, 6137–6144 (2014).
- de Gori, P., Lucente, F. P., Lombardi, A. M., Chiarabba, C. & Montuori, C. Heterogeneities along the 2009 L'Aquila normal fault inferred by the b-value distribution. *Geophys. Res. Lett.* <https://doi.org/10.1029/2012GL052822> (2012).
- di Luccio, F., Ventura, G., di Giovambattista, R., Piscini, A. & Cinti, F. R. Normal faults and thrusts reactivated by deep fluids: The 6 April 2009 Mw 6.3 L'Aquila earthquake, central Italy. *J. Geophys. Res. Solid Earth.* <https://doi.org/10.1029/2009JB007190> (2010).

29. Calderoni, G., Rovelli, A. & di Giovambattista, R. Stress drop, apparent stress, and radiation efficiency of clustered earthquakes in the nucleation volume of the 6 April 2009, Mw 6.1 L'Aquila earthquake. *J. Geophys. Res. Solid Earth*. **124**, 10360–10375 (2019).
30. Spallarossa, D. *et al.* The RAMONES service for rapid assessment of seismic moment and radiated energy in Central Italy: Concepts, capabilities, and future perspectives. *Seismol. Res. Lett.* **92**, 1759–1772 (2021).
31. Zaliapin, I., Gabrielov, A., Keilis-Borok, V. & Wong, H. Clustering analysis of seismicity and aftershock identification. *Phys. Rev. Lett.* <https://doi.org/10.1103/PhysRevLett.101.018501> (2008).
32. Zaliapin, I. & Ben-Zion, Y. Discriminating characteristics of tectonic and human-induced seismicity. *Bull. Seismol. Soc. Am.* **106**, 846–859 (2016).
33. Aden-Antônio, F. *et al.* *J. Geophys. Res. Solid Earth (in Press)*, <https://doi.org/10.1029/2019JB019337>.
34. Gutenberg, B. & Richter, C. F. Earthquake magnitude, intensity, energy, and acceleration. *Bull. Seismol. Soc. Am.* **32**, 163–191 (1942).
35. Grassberger, P. & Procaccia, I. Characterization of strange attractors. *Phys. Rev. Lett.* **50**, 346–349 (1983).
36. Catalli, F., Cocco, M., Console, R. & Chiaraluze, L. Modeling seismicity rate changes during the 1997 Umbria-Marche sequence (central Italy) through a rate- and state-dependent model. *J. Geophys. Res. Solid Earth*. **113**, 11301 (2008).
37. Shannon, C. E. A mathematical theory of communication. *Bell Syst. Tech. J.* **27**, 623–656 (1948).
38. Fischer, T. & Hainzl, S. Effective stress drop of earthquake clusters. *Bull. Seismol. Soc. Am.* **107**, 2247–2257 (2017).
39. Kostrov, B. V. Crack propagation at variable velocity. *J. Appl. Math. Mech.* **38**, 511–560 (1974).
40. Efron, B. Bootstrap methods: Another look at the jackknife. *Ann. Stat.* **7**, 1–26 (1979).
41. von Mises, R. *Wahrscheinlichkeit Statistik und Wahrheit* 2nd edn. (Springer, 1936).
42. Pastore, M. & Calcagni, A. Measuring distribution similarities between samples: A distribution-free overlapping index. *Front. Psychol.* **10**, 1089 (2019).
43. Bletery, Q. & Nocquet, J. M. The precursory phase of large earthquakes. *Science* **381**(6655), 297–301. <https://doi.org/10.1126/science.adg2565> (2023).
44. Jones, L. M. & Molnar, P. Some characteristics of foreshocks and their possible relationship to earthquake prediction and premonitory slip on faults. *J. Geophys. Res.* **84**, 3596–3608. <https://doi.org/10.1029/JB084iB07p03596> (1979).
45. Gentili, S., Peresan, A., Talebi, M., Zare, M. & Di Giovambattista, R. A seismic quiescence before the 2017 Mw 7.3 Sarpol Zahab (Iran) earthquake: Detection and analysis by improved RTL method. *Phys. Earth Planet. Inter.* **290**, 10–19. <https://doi.org/10.1016/j.pepi.2019.02.010> (2019).
46. Chiarabba, C. & Chiadini, G. Continental delamination and mantle dynamics drive topography, extension and fluid discharge in the Apennines. *Geology* **41**, 715–718 (2013).
47. Chiadini, G. *et al.* Correlation between tectonic CO<sub>2</sub> earth degassing and seismicity is revealed by a 10-year record in the Apennines, Italy. *Sci Adv.* [https://doi.org/10.1126/SCIADV.ABC2938/SUPPL\\_FILE/ABC2938\\_SM.PDF](https://doi.org/10.1126/SCIADV.ABC2938/SUPPL_FILE/ABC2938_SM.PDF) (2020).
48. Miller, S. A. *et al.* Aftershocks driven by a high-pressure CO<sub>2</sub> source at depth. *Nature* **427**(6976), 724–727 (2004).
49. Collettini, C., Barchi, M. R., de Paola, N., Trippetta, F. & Tinti, E. Rock and fault rheology explain differences between on fault and distributed seismicity. *Nat. Commun.* **13**(1), 1–11 (2022).
50. Gentili, S. & Di Giovambattista, R. Forecasting strong aftershocks in earthquake clusters from northeastern Italy and western Slovenia. *Phys. Earth Planet. Interiors* **303**, 106483. <https://doi.org/10.1016/j.pepi.2020.106483> (2020).
51. Ben-Zion, Y. A critical data gap in earthquake physics. *Seismol. Res. Lett.* **90**, 1721–1722 (2019).
52. Seeber, L. & Armbruster, J. G. Earthquakes as beacons of stress change. *Nature* **407**(6800), 69–72 (2000).
53. Stein, R. S. & Lisowski, M. The 1979 Homestead Valley Earthquake Sequence, California: Control of aftershocks and postseismic deformation. *J. Geophys. Res. Solid Earth*. **88**, 6477–6490 (1983).
54. Goebel, T. H. W. *et al.* Identifying fault heterogeneity through mapping spatial anomalies in acoustic emission statistics. *J. Geophys. Res. Solid Earth*. **117**, 3310 (2012).
55. Rouet-Leduc, B. *et al.* Estimating fault friction from seismic signals in the laboratory. *Geophys. Res. Lett.* <https://doi.org/10.1002/2017GL076708> (2018).
56. Rouet-Leduc, B. *et al.* Machine learning predicts laboratory earthquakes. *Geophys. Res. Lett.* **44**, 9276–9282 (2017).
57. Miyazaki, S., Segall, P., Fukuda, J. & Kato, T. Space time distribution of afterslip following the 2003 Tokachi-oki earthquake: Implications for variations in fault zone frictional properties. *Geophys. Res. Lett.* <https://doi.org/10.1029/2003GL019410> (2004).
58. Wang, K., Johnson, C. W., Bennett, K. C. & Johnson, P. A. Predicting fault slip via transfer learning. *Nat. Commun.* **12**(1), 1–11 (2021).
59. Scholz, C. H. On the stress dependence of the earthquake b value. *Geophys. Res. Lett.* **42**, 1399–1402 (2015).
60. Aki, K. Maximum likelihood estimate of b in the Gutenberg-Richter formula and its confidence limits. *Bull. Earthq. Res. Inst.* **43**, 237–239 (1965).
61. Chiaraluze, L., Valoroso, L., Anselmi, M., Bagh, S. & Chiarabba, C. A decade of passive seismic monitoring experiments with local networks in four Italian regions. *Tectonophysics* **476**, 85–98 (2009).

## Acknowledgements

We would like to thank the Editor A. Amoroso, G. Petrillo and another anonymous reviewer for their comments and suggestions that allowed us to significantly improve the manuscript content and form.

## Author contributions

Conceptualization: M.P. Methodology: M.P., A.G.I., D.S., D.B. Investigation: M.P., D.S. Visualization: M.P., A.G.I., D.S., D.B. Supervision: M.P., A.G.I., D.S., D.B. Writing—original draft: M.P., A.G.I., D.S., D.B.

## Funding

This research was partly carried out in the frame of Programme STAR PLUS (project DRAGON), financially supported by UniNA and Compagnia di San Paolo.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-023-41625-0>.

**Correspondence** and requests for materials should be addressed to M.P.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023