

Parametric table of recordings for seismological studies at Mount Etna: implications for seismic microzonation studies

Gabriele Tarchini^{*1}, Giulio Brunelli², Giovanni Lanzano², Giuseppina Tusa³, Chiara Felicetta², Tommaso Castelbarco², Daniela Famiani⁴, Salomon Hailemichael⁴, Giuliano Milana⁴, Mario Paratore³, Daniele Spallarossa¹, Raffaele Azzaro³, Francesca Pacor²

(1) University of Genoa, Department of Earth, Environmental and Life Sciences, Genoa, Italy

(2) Istituto Nazionale di Geofisica e Vulcanologia (INGV), Sezione di Milano, Milano, Italy

(3) Istituto Nazionale di Geofisica e Vulcanologia (INGV), Osservatorio Etneo, Catania, Italy

(4) Istituto Nazionale di Geofisica e Vulcanologia (INGV), Sismologia e Tettonofisica, Roma, Italy

Article history: received April 29, 2026; accepted October 1, 2026

Abstract

We present a new dataset of earthquake ground-motion records for the volcanic area of Mount Etna (eastern Sicily, Italy), aimed at supporting seismological analyses and seismic microzonation studies. The dataset includes 200 volcano-tectonic earthquakes with local magnitude $M_L \geq 3.0$ that occurred between 2008 and 2023, recorded by a dense network of permanent and temporary stations up to 150 km. Waveforms were processed using a uniform workflow based on manual phase picking, signal-to-noise criteria, and instrument correction. From the processed signals, we derived peak parameters (Peak Ground Acceleration and Peak Ground Velocity), Fourier Amplitude Spectra, and 5% damped response spectra, all distributed through a parametric table integrating ground-motion measures with event, station, and record metadata. Site conditions were characterized through a multi-level approach combining existing databases, geological information, Horizontal-to-Vertical Spectral Ratio analyses, and dedicated geophysical surveys at selected stations. Dataset consistency was evaluated through residual analysis, allowing the identification of anomalous records while preserving statistical representativeness. To support future ground motion modeling in the area, we identified two clusters with specific seismological characteristics, corresponding to different depth ranges. The resulting dataset provides a unique resource for investigating ground-motion variability in volcanic environments, which are underrepresented in existing strong-motion archives, and for advanced applications in ground-motion modeling and seismic microzonation.

Keywords: Mount Etna; Ground Motion Dataset; Waveform processing; Site Characterization; Residual analysis

1. Introduction

Seismicity in volcanic areas poses distinctive challenges for seismic hazard assessment and engineering seismology, not only because it combines low-energy earthquakes with strong ground motion, but also because seismic modeling must account for complex source processes, strong lateral heterogeneity along propagation paths, and significant variability of the site effects over short spatial scales (e.g., Hirn et al., 1991; Giampiccolo et al., 2020; Totaro et al., 2024). Mount Etna (eastern Sicily, Italy) is an emblematic case, being one of the most active volcanoes in Europe and characterized by persistent volcano-tectonic seismicity, with episodic swarms or sequences often associated with intrusive and eruptive phases. These frequently culminate in damaging shallow events on the volcano flanks (Gambino and Distefano, 2022). These conditions recently motivated the study of the peculiar ground motion characteristics in this area by calibrating specific predictive models of the main relevant parameters of engineering interest (Tusa and Langer, 2016; Lanzano and Luzi, 2020; Tusa et al., 2020; Ramadan et al., 2023).

To characterize ground motion in a volcanic area such as Mt. Etna, the availability of a rich and high-quality dataset of local volcano-tectonic earthquake waveforms is essential. Over the past two decades, the availability of accelerometric and velocimetric seismic records has increased substantially, thanks to national and international archives and monitoring infrastructures. In Italy, processed strong-motion waveforms of events with $M > 3.0$ and their associated metadata are disseminated through resources such as the ITalian ACcelerometric Archive (ITACA, Felicetta et al., 2023). At the European scale, the Engineering Strong-Motion (ESM, Mascandola et al., 2026) database provides harmonized access to ground-motion records and metadata, fostering cross-regional analyses and the development and testing of engineering models. However, these databases are mainly designed for crustal seismicity of tectonic origin, and often assume rather high seismicity thresholds for volcanic seismicity characteristics (e.g., ESM assumes a minimum magnitude threshold of 4.0). On the other hand, volcanic areas have dense local networks, including broad-band and short-period seismometers, which are generally not available in strong motion repositories.

In this study we present a novel dataset of earthquake ground-motion records for the Mt. Etna region, released as a parametric table integrating parameters of engineering and seismological interest (5% damped acceleration response spectra – SA – ordinates and peak parameters and Fourier Amplitude Spectra – FAS – ordinates), together with record, event and station metadata enabling a range of applications, such as: i) empirical characterization of near-source shaking in the volcanic context; ii) assessment of intra-event and inter-event variability; iii) investigation and measurement of site and topographic effects; iv) development and validation of ground-motion models (GMMs) tailored to volcanic regions; and v) engineering analyses that require consistent spectral and peak descriptors for record selection, structural response estimation, and comparative studies. One of the first results within the third level seismic microzonation study at Mt. Etna is the estimation of the empirical amplification factors, described in detail in Laurenzano et al. (this issue).

This paper describes in detail the strategies and procedures for developing this dataset, namely: i) selection of cataloged volcano-tectonic events in the study area; ii) acquisition of records from public seismic networks; iii) semi-automated data processing; iv) calculation of relevant parameters; v) assignment of metadata (with particular reference to data characterizing station sites); vi) consistency analysis of the dataset for seismological and engineering applications; and, finally, vii) dissemination of the dataset in accordance with the FAIR (Findable, Accessible, Interoperable and Reusable) principles (Wilkinson et al., 2016).

2. Seismicity and earthquake catalog

The volcanic region of Mt. Etna is located within a highly complex seismotectonic setting, lying at the regional scale between two major seismic sources, capable of producing very large and destructive seismic events ($M > 7.0$), such as the 1693 Val di Noto earthquake (moment magnitude, $M_w = 7.32$ according to Rovida et al., 2022) in southeastern Sicily and the 1908 Strait of Messina earthquake ($M_w = 7.1$) in the northern sector. At the local scale, much more frequent but less energetic earthquakes affect the volcanic area of Mt. Etna, with different features depending on the volcano-tectonic system involved (Scarfi et al., 2025; Azzaro et al., this issue and Fig. 1). In general, the eastern flank is characterized by a very shallow seismicity (usually at focal depths $H < 3$ km) frequently related to the activity of different tectonic systems (Pernicana fault, Timpe faults, etc.), responsible for the high level of hazard of this densely urbanized area (Azzaro et al., 2013; D'Amico et al., 2025). The central sector of the volcano exhibits earthquakes typically related to volcanic processes (magma ascent, dyke intrusion, opening of eruptive fractures),

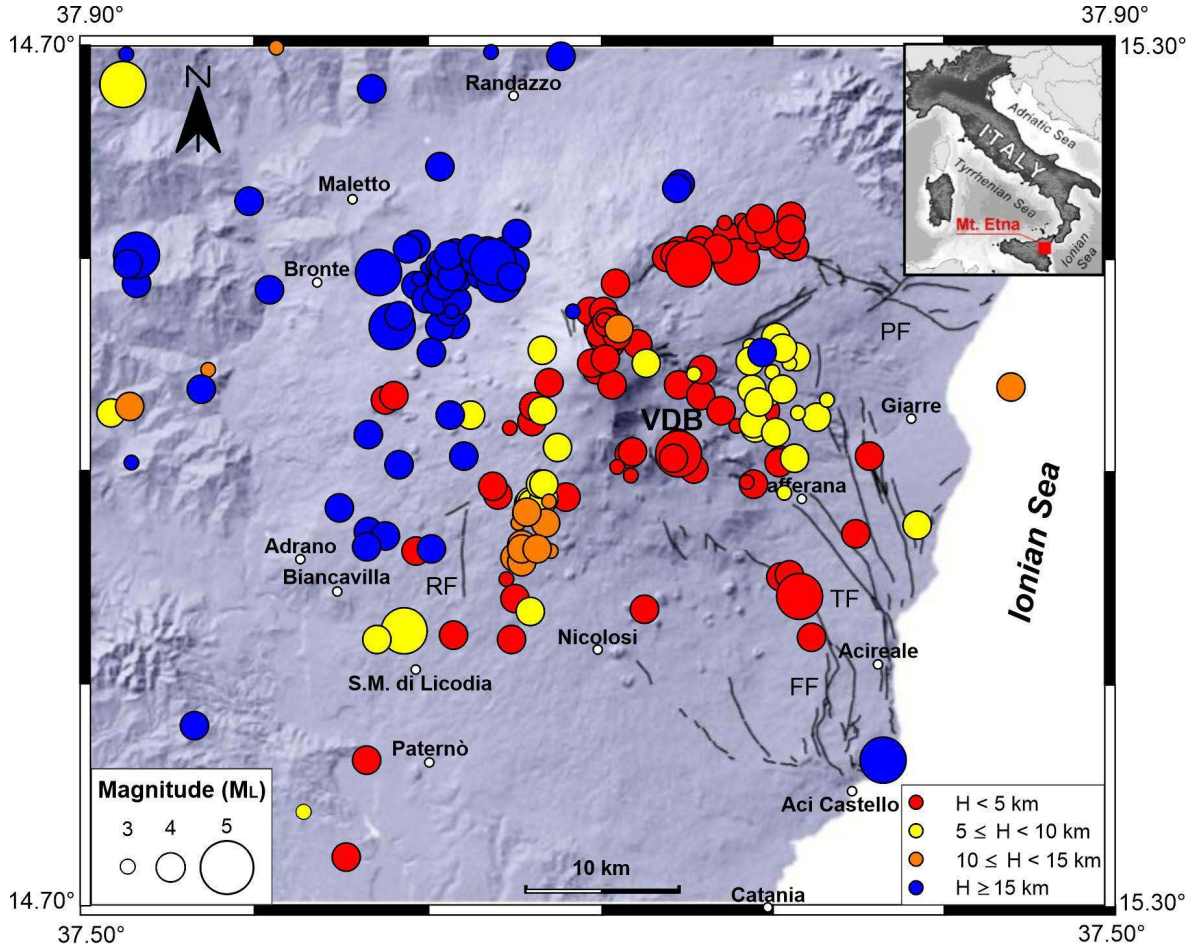


Figure 1. Epicenter distribution of earthquakes with magnitude $M_L \geq 3.0$ occurring between 2008 and 2023 in the volcanic area of Mt. Etna. The circle size indicates magnitude, while the colors indicate range depth. Keys for major fault systems: PF (Pernicana Fault), RF (Ragalna Fault), TF (Timpe Fault), FF (Fiandaca Fault). The position of Valle del Bove (VDB) is also shown.

often clustered in swarms and located at different levels in the crust ($1 < H < 15$ km) (Firetto Carlino et al., 2022). Finally, the northwestern sector of Etna is characterized by deep seismicity ($H > 15$ km) closely associated with the regional tectonic structures and controlled by a compressive stress field affecting the underlying sedimentary basement. Interestingly, the largest magnitude events (local magnitude, $M_L = 4.8$) occur on both the eastern and north-western flanks, but with substantially different implications for engineering seismology. On the eastern flank, a shallow $M_L 3.5$ event can cause damage, while earthquakes up to $M_L 4.8$ may reach macroseismic intensity of VIII-IX on the EMS scale. By contrast, events of similar magnitude occurring at greater depth beneath the north-western flank are generally not damaging.

The intensive seismic monitoring operated by the Istituto Nazionale di Geofisica e Vulcanologia – Osservatorio Etneo (INGV-OE) led to the compilation of specific earthquake catalogs for the Etna region, from which the data used in this study were selected (Alparone et al., 2020; Barberi et al., 2020). These include the Mt. Etna Revised and Concise Seismic Catalog from 1999 (EtnaRSC; Alparone et al., 2020), whose webpage also provides access to catalogs covering specific time intervals, including the Mt. Etna Revised Seismic Catalog from 2020 (EtnaRSC2020; Barberi et al., 2020). The latter is updated daily by the INGV-OE analyst team and covers seismicity from 2020 onward. The earthquake data used in this study were selected from these resources and extend through 2023. In detail, the earthquakes were located using the HYPOELLIPSE code (Lahr, 1989) and a seven-layer, one-dimensional velocity model that accounts for the difference in elevation among the seismic stations (Hirn et al., 1991). In this study, we selected 200 earthquakes, whose epicentral distribution is shown in Fig. 1. The events have M_L values between 3.0 and 4.8 and H up to 30 km, and occurred from February 2008 to July 2023 (Fig. 2). We chose this time span to ensure high-quality data and homogeneity in the instrumental chain.

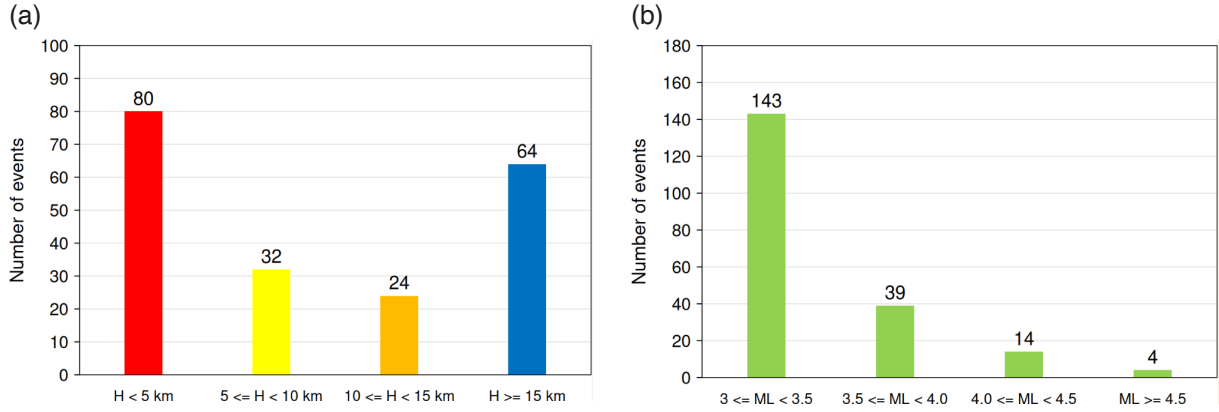


Figure 2. Statistical characteristics of the earthquake dataset selected for the Etna region. (a) Distribution of number of events as a function of focal depth; (b) number of events as a function of local magnitude.

The starting dataset was screened from the aforementioned earthquake catalogs by applying strict quality criteria: horizontal and vertical location errors (erh and erz) ≤ 2 km, root mean square of travel-time residuals (rms) ≤ 0.35 s, and a minimum of 10 phase readings. Within this subset, event locations are well-constrained: erh and erz are generally ≤ 1.1 km and ≤ 1.6 km, respectively, with only three events having $1.2 \leq erz \leq 1.6$ km. All earthquakes in this study exhibit high-frequency, impulsive P-wave onsets, and clearly identifiable S phases. They therefore conform to the classical definition of high-frequency (HF), or volcano-tectonic (VT), earthquakes provided by McNutt and Roman (2015).

The dataset used in this study includes four earthquakes with $M_L \geq 4.5$, representing the strongest events instrumentally recorded in the Mt. Etna volcanic area. The two largest events in the catalog ($M_L = 4.8$) were associated with major seismic sequences that occurred at Mt. Etna in late 2009 and late 2018, respectively. In particular, the first event, which occurred on 19 December 2009, represents the mainshock of a seismic sequence consisting of about 150 events. It was located on the north-western flank of Mt. Etna at a focal depth of approximately 30 km. This earthquake produced the strongest effects in the municipalities of Maletto (V EMS) and Bronte (IV-V EMS), and more generally across the northern flank of the volcano. The shock was widely felt by the population and, fortunately, did not cause significant damage; the felt area extended across the whole of eastern Sicily, up to distances exceeding 100 km from the epicenter (Azzaro et al., 2014). The second $M_L = 4.8$ event of the catalog occurred on 26 December 2018 and was the mainshock of a shallow seismic sequence consisting of about 50 events on the eastern flank of Mt. Etna. It was one of the most damaging shallow earthquakes in the recent history of the volcano. Surface faulting was clearly observed along the Fiandaca fault system, with visible ruptures extending several kilometers. The earthquake produced the strongest effects in the epicentral area, causing structural damage in the municipalities of Fleri, Pennisi, and nearby villages, with maximum intensity reaching VIII EMS. The shock was widely felt across eastern Sicily, including the city of Catania, up to distances exceeding 40–50 km from the epicenter (Azzaro et al., 2020).

The catalog contains two additional earthquakes with $M_L \geq 4.5$. The first ($M_L = 4.7$) occurred on 6 October 2018 near Santa Maria di Licodia, and represented the mainshock of a local seismic sequence of about 60 events. The earthquake had a focal depth of 6 km, and was widely felt across eastern Sicily, including Catania, Enna, and Messina. It produced notable effects in the epicentral area, such as the fall of cornices and plaster from historic buildings and minor structural damage in Santa Maria di Licodia, Biancavilla, Paternò, and Adrano (Azzaro et al., 2020). The second event ($M_L = 4.5$) occurred offshore of Aci Castello on 21 April 2023. It was a moderate earthquake, located at a depth of 20 km. The shock was widely felt across eastern Sicily, including Catania, Siracusa, and Ragusa.

3. Data collection and processing

Waveforms for the three components were retrieved from the European Integrated Data Archive (EIDA; see Data and Resources; Strollo et al., 2021) repository in MiniSEED format using the `fdsnws-datasetselect` service. For each event, the corresponding seismograms were extracted from continuous recordings and converted into Seismic Analysis Code (SAC) format. The waveforms were selected based on the following attributes: i) event origin time, based on the

reference earthquake catalog, described in the previous section; ii) epicentral distance (R), considering records with $R \leq 150$ km; iii) networks, including all open-access networks with both permanent and temporary stations; iv) high sampling rate velocimetric and accelerometric channels (i.e., velocimetric channels HH and EH; and accelerometric channels HN, HL, and HG). Data were requested with a sufficient pre-event and post-event time-window to capture all relevant parts of the signal, including noise and seismic phases. The dataset was subsequently expanded by integrating recordings from the Italian Civil Protection Department (DPC) strong-motion network (network code “IT”; Presidency of Council of Ministers-Civil Protection Department, 1972); this integration was partial, however, as DPC data were not available for the entire study period and were therefore incorporated only for the years 2020-2023.

In total, approximately 50,000 waveforms were collected from the seismic stations belonging to eight different seismic networks: 1J (Moretti et al., 2021), 3D (Alparone et al., 2021), 7C (Costanzo et al., 2022), IT (Presidency of Council of Ministers – Civil Protection Department, 1972), IV (Istituto Nazionale di Geofisica e Vulcanologia [INGV], 2005), IY (Università Della Calabria, Italy, 1981), MN (MedNet Project Partner Institutions, 1990), and TV (Istituto Nazionale di Geofisica e Vulcanologia [INGV], 2008; <https://terremoti.ingv.it/instruments/network/TV>). The main spatial coverage of the Mt. Etna region is provided by the networks most densely distributed across the volcano and its surroundings, namely, 1J, IT, and IV, while the remaining networks contribute additional recordings that improve the azimuthal coverage, increase the waveform availability, and enhance the overall robustness of the subsequent analyses.

3.1 Waveform processing

All waveforms were visually inspected to evaluate data quality and to identify problematic recordings, including clipped signals, spikes or glitches, waveform distortion, data gaps, and incomplete codas. Manual picks of P- and S-wave onsets were then performed to support subsequent window selection and spectral processing. Following phase picking, all velocimetric records were converted into acceleration waveforms. The acceleration traces were corrected through linear detrending – applied to the record to correct for any baseline trend and mean offset – and 5% Hann tapering (Harris, 1978) – applied at both ends of the record to ensure that it starts and ends at zeroes and transitions smoothly on both ends. The waveforms were then filtered using a second-order acausal zero-phase Butterworth band-pass filter. The processing scheme, first defined in Pacor et al. (2016), was adopted here with the following parametrization: the low-cut (high-pass) corner frequency was automatically determined based on the signal-to-noise ratio (SNR) and constrained within the range 0.04-0.6 Hz to satisfy $SNR > 2.5$ (Fig. 3a), while the

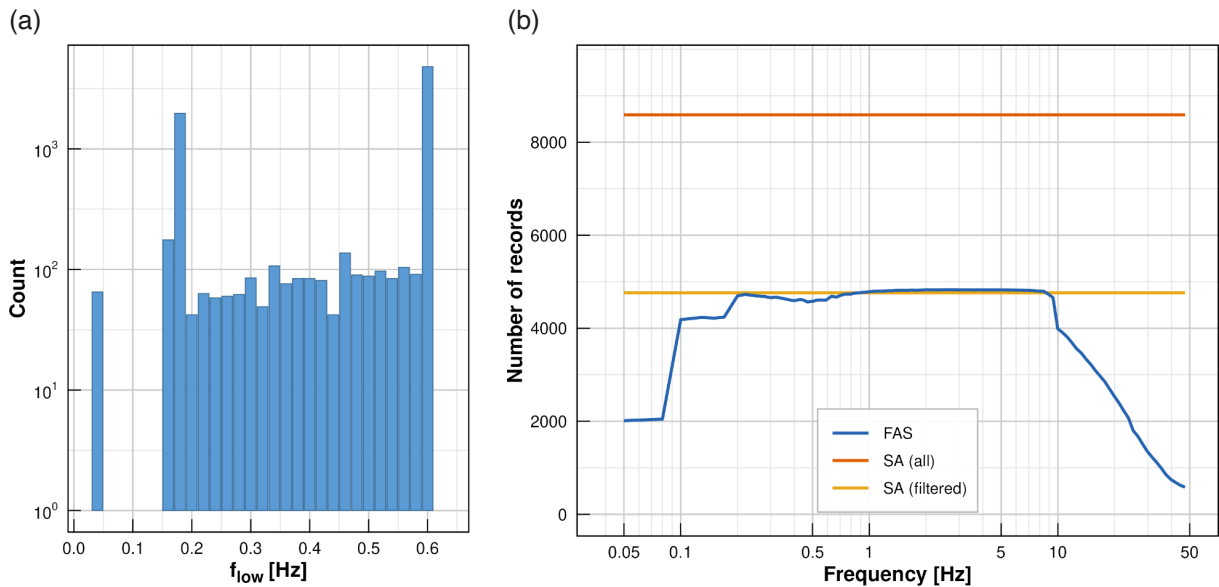


Figure 3. (a) Low cut corner frequency distribution in the dataset. (b) Number of usable data as a function of frequency for FAS and SA spectral ordinates. Horizontal lines indicate the total number of SA records (SA all) and the subset retained after discarding recordings not available in the FAS dataset (SA filtered).

high-cut (low-pass) corner frequency was fixed at 48 Hz. The distribution of low-cut corner frequencies (Fig. 3a) shows that the bulk of values falls within the range 0.15–0.6 Hz, with a pronounced concentration near 0.2 Hz and at the imposed upper bound of 0.6 Hz. Figure 3b shows the number of usable records as a function of frequency for both FAS and SA ordinates. To ensure consistency between the FAS and SA datasets, records present in the SA tables but persistently excluded from all FAS tables – indicating persistently low *SNR* – were discarded from the SA tables (Fig. 3b).

After filtering, the instrument response was removed to obtain waveforms in displacement physical units. The displacement time series were corrected through linear detrending, offset removal, and cosine tapering. Velocity and acceleration records were obtained by successive differentiation of the tapered displacement traces. For each three-component recording, Peak Ground Acceleration (PGA) and Peak Ground Velocity (PGV) were extracted directly from the processed waveforms as the maximum absolute amplitude observed on each component, expressed in physical units (PGV in cm/s and PGA in cm/s²).

3.2 Spectral intensity measures

FAS were computed using S-wave windows defined through a distance-dependent cumulative energy criterion. Time windows started 0.1 s before the S-wave onset and ended when a fraction of cumulative signal energy (Arias Intensity, *AI*) was reached as a function of source-to-site hypocentral distance R_{hyp} :

- $R_{hyp} \leq 25$ km: 90% *AI*;
- $25 < R_{hyp} \leq 50$ km: 80% *AI*;
- $R_{hyp} > 50$ km: 70% *AI*.

The distance-dependent *AI* fractions account for the increasing contribution of scattered coda and the progressively lower *SNR* at larger hypocentral distances: enforcing a fixed high fraction (e.g., 90%) at far distances would extend the window into noise-contaminated late arrivals, whereas the adopted reduction keeps the window centered on the high-*SNR* S-wave energy.

To avoid excessively long windows, the S-wave window length was capped at 25 s (i.e., windows longer than 25 s were truncated). Conversely, because windows for near-source records can be very short (<2 s), a minimum window length of 4 s was enforced to guarantee adequate spectral resolution below 1 Hz.

Extracted waveforms were tapered using Hanning windows, with taper length adjusted to the duration of the selected S-wave window. Spectral amplitudes were computed at 98 logarithmically equally spaced frequencies in the 0.05–47.20 Hz band and smoothed using the Konno and Ohmachi (1998) algorithm with smoothing parameter $b = 40$. The final FAS dataset was obtained by rejecting spectral amplitudes characterized by a *SNR* lower than 2.5, frequency by frequency. As a result, the number of usable records varies with frequency (Fig. 3b), reaching the maximum number in the 1–5 Hz range.

Response spectra were computed using fixed-length 30-s S-wave windows. Unlike the FAS windows, a fixed duration was adopted to provide a uniform and sufficiently long time interval for computing the single-degree-of-freedom oscillator response across the dataset. Using the distance-dependent energy criterion adopted for FAS would instead make the duration entering the response-spectrum calculation systematically dependent on distance and on the selected *AI* threshold. SA were evaluated at 98 vibration periods, corresponding to the same set of 98 logarithmically spaced frequencies adopted for the FAS computation, ensuring full consistency between the two spectral products.

No selection based on *SNR* was applied for SA and PGA at the initial computation stage. As described above, SA records with no usable FAS counterpart at any frequency were discarded from the SA tables, resulting in approximately 4,700 records available in the SA and PGA tables. Figure 4a shows the magnitude-distance distribution of the PGA dataset: the dataset includes a significant number of near-source recordings ($R < 10$ km), amounting to 331 records from 101 events. Figure 4b shows the focal depth vs. epicentral distance distribution, with hypocentral depths mostly confined within the upper 30 km of the crust.

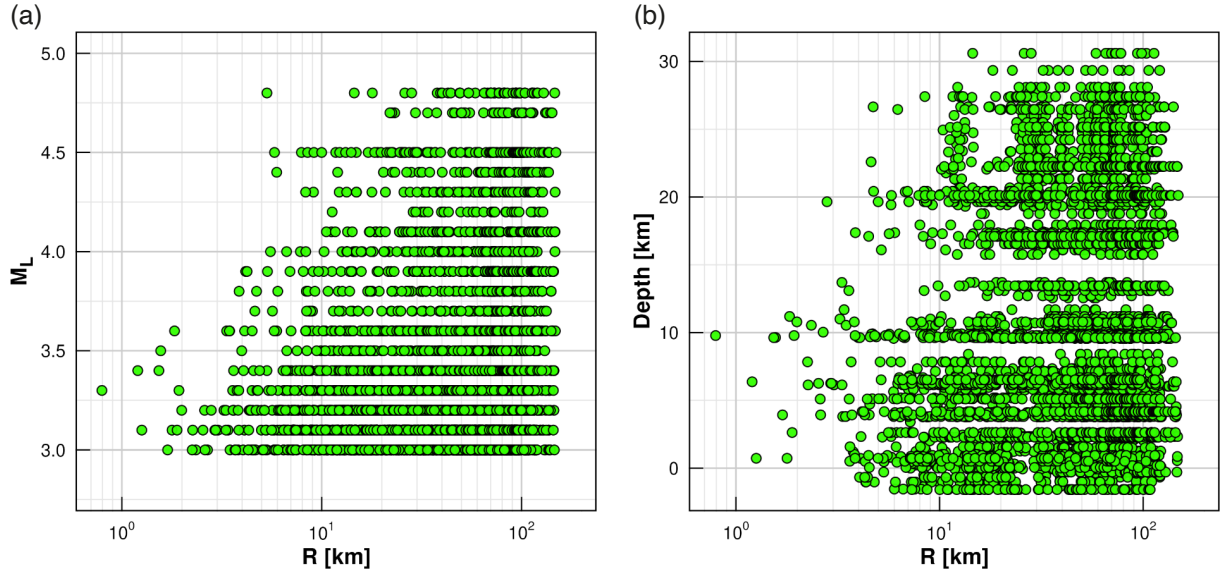


Figure 4. Data distribution of the PGA dataset: (a) local magnitude vs. epicentral distance; (b) focal depth vs. epicentral distance.

4. Site characterization of the recording stations

The characterization of sites of the recording stations in the dataset was conducted through an integrated approach combining data from several different sources, namely: i) site characterization data (time-averaged shear-wave velocity in the uppermost 30 m depth – $V_{S,30}$ – values, Eurocode 8 – EC8 – and Norme Tecniche per le Costruzioni – NTC – soil categories, NTC topographic categories) from existing databases, such as ITACA and ESM; ii) surface geology classification based on the ISPRA lithological map at the 1:100,000 scale (ISPRA, 2021), properly revised by Brunelli et al. (2026); iii) the result of ad hoc geophysical surveys for shear-wave velocity (V_S) profile reconstruction; iv) the assessment of Horizontal-to-Vertical Spectral Ratio (HVSr) parameters from noise measurements and signal analysis.

The analysis is structured into three levels of detail. The most detailed level concerns 6 recording stations located in the area of Mt. Etna that were specifically investigated through dedicated geophysical surveys to estimate the V_S profile with depth (Fig. 5). The second intermediate level includes all 40 stations in the Mt. Etna area (Fig. 5), for which the site natural vibration frequency f_0 and the associated amplitude $A(f_0)$, were estimated from HVSr, using the EIDA and RAN (Rete Accelerometrica Nazionale, network IT; Presidency of Council of Ministers-Civil Protection Department, 1972) data streams. The third level provides a basic site characterization for the 147 stations of the dataset (Fig. S1 of the Electronic Supplementary Material), located within a wider area, based on the available information in existing repositories.

4.1 Geophysical survey campaigns

Passive seismic surveys were conducted to obtain the V_S profiles at selected permanent seismic monitoring stations in the Mt. Etna area. The investigated sites include EPOZ, ESML, ENIC, ESAL, EVRN (IV network) and SVN (IT network) (Fig. 5). The stations were selected based on the availability of high-quality seismic recordings, whose preliminary analysis revealed peculiar features in the site response, motivating a dedicated investigation of local site conditions.

A series of ambient seismic noise surveys were performed using two-dimensional (2D) seismic arrays. The acquisition system consisted of multiple three-component nodal sensors (4.5 Hz eigen-frequency) equipped with high-dynamic-range recorders and GPS timing. Sensor positions were determined with millimetric accuracy using differential GPS, allowing precise reconstruction of the array geometry. The arrays were deployed in radial geometries with characteristics linked to the logistical conditions of each investigated site: the number of used

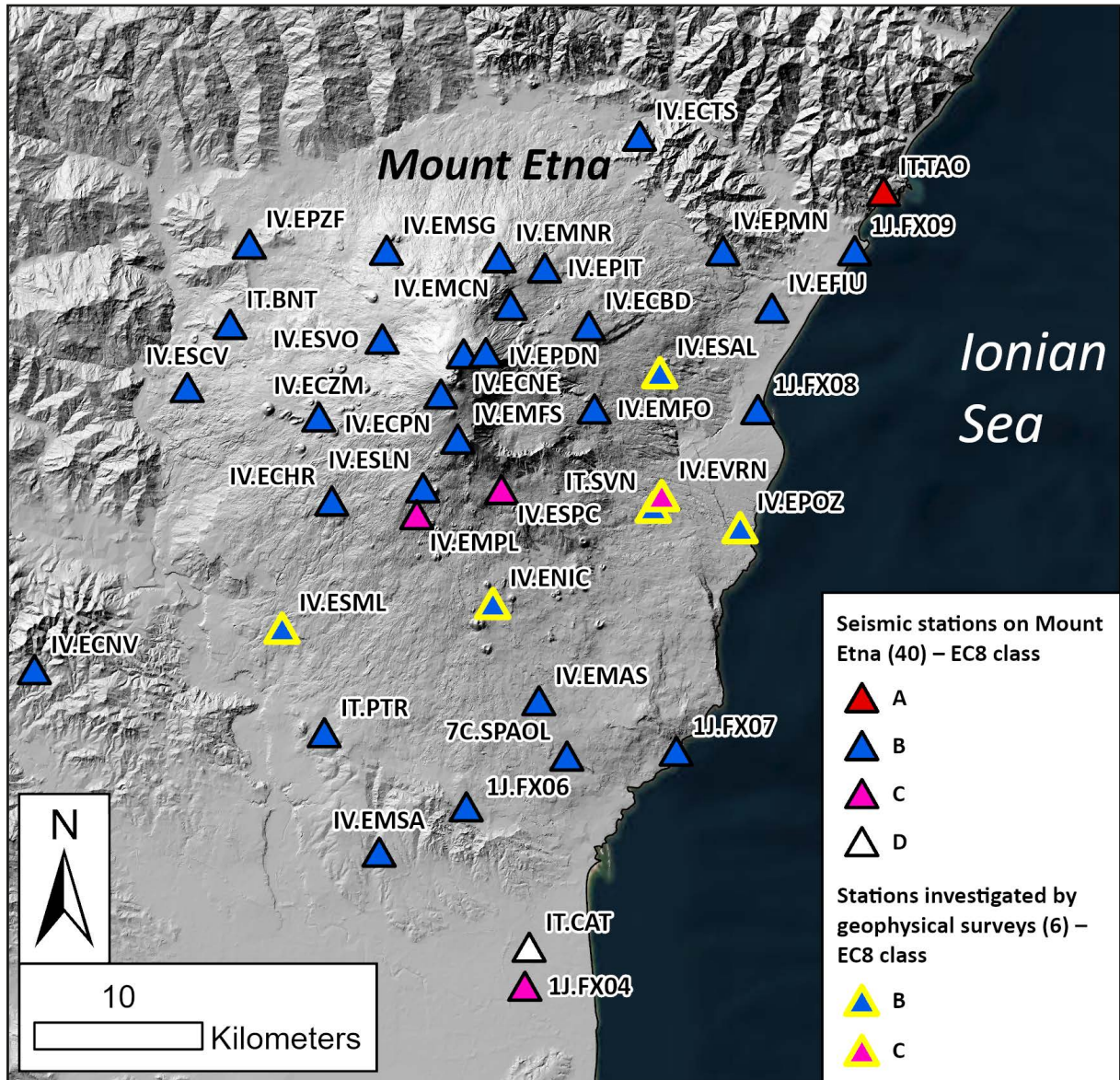


Figure 5. Spatial distribution of the 40 seismic stations in the Mt. Etna area, classified according to the EC8 site classes. The six stations investigated by geophysical surveys are highlighted with yellow-outlined symbols.

nodes varied between 18 and 46, with array maximum aperture ranging from 60 to 220 m and recording durations between approximately 1.5 and 15 hours.

For each seismic node, the HVSF function was determined from the noise recordings (N-HVSF) using a consistent processing workflow (SESAME, 2004). Specifically, the data were divided into 60-second time-windows. The spectra for each ground-motion component within these windows were smoothed using the Konno and Ohmachi (1998) algorithm, employing a smoothing parameter of $b = 40$. To obtain the horizontal spectra, the squared average of the North (N) and East (E) components was computed. Finally, the overall HVSF function was derived by averaging the functions obtained from each selected time window. The analyses were performed using the Geopsy software suite (Wathelet et al., 2010, 2020).

In general, the N-HVSFs, calculated for individual nodes at each investigated site, are highly consistent in the low-to-intermediate frequency range (below 4 Hz), while showing large variability at higher frequencies, which may reflect a marked subsurface heterogeneity at shallow depths (Fig. 6). In some sites, such as EPOZ and SVN (Figs. 6b and 6d), the N-HVSF curves of all seismic nodes are equivalent across the 0.5-10 Hz frequency range, suggesting a rather homogeneous subsurface structure. Overall, the investigated sites display different N-HVSF

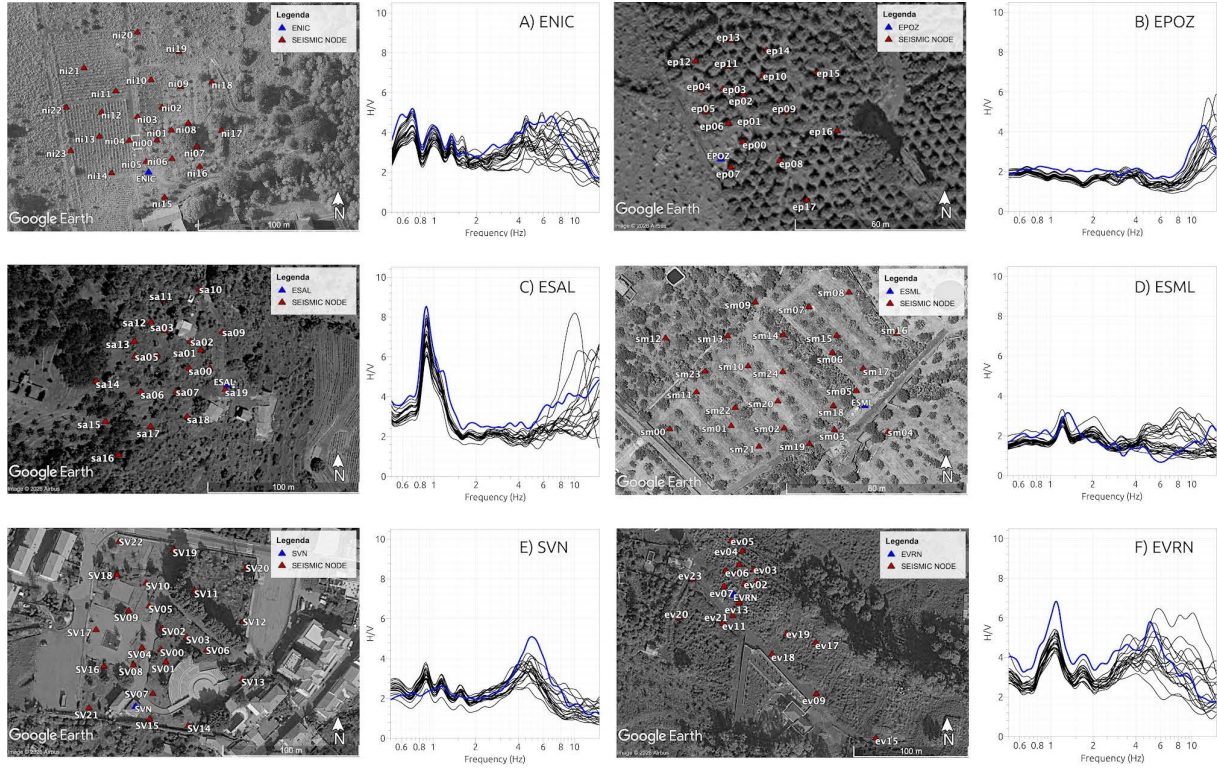


Figure 6. Map of the 2D array layouts (left panels) showing seismic nodes as red triangles and permanent stations as blue triangles), together with the N-HVSR curves calculated for each seismic node of the array (black curves in the right panels). The N-HVSR curves calculated for the permanent seismic stations (blue curves) are also shown (from ITACA and/or ESM).

characteristics, ranging from single high frequency resonance (EPOZ, Fig. 6b) to low frequency resonance (ESAL, Fig. 6c) and multiple peaks (ENIC, Fig. 6a).

With the exception of the SVN site, where neither RTBF nor MSPAC processing provided sufficiently clear and reliable dispersion images to extract an experimental dispersion curve and constrain the subsequent inversion, likely due to the lack of coherent signals in the acquired dataset, for all other sites, the shear-wave (V_S) velocity profile was obtained through analysis of the dispersive properties of surface waves (Bard et al., 2010). Indeed, assuming: i) 1D subsurface conditions for the investigated volume, ii) predominance of surface waves in the ambient seismic noise field, and iii) propagation approximable to plane waves relative to the recording array, it is possible to extract the Dispersion Curve (DC) of the surface waves using appropriate signal processing techniques and obtain the V_S profile by solving the inverse problem for the estimation of the subsurface model based on the experimental DC data. Among the most common techniques for processing surface wave data, the RTBF technique (Wathelet et al., 2018), proposed as a development of beamforming (Capon, 1969) applied to multi-component recordings, and the MSPAC technique (Bettig et al., 2001; Köhler et al., 2007), founded on the theory proposed by Aki (1957) for arrays arranged along fixed-radius circles and developed for acquisition geometries of any type, were used. In particular, the RTBF method was used for retrieving the DC of Rayleigh- and Love-waves fundamental mode as well as Rayleigh-waves ellipticity, whereas the MSPAC method was mainly used for estimating the autocorrelation curve of the array. For the inversion of surface wave data, a global search method based on the neighborhood algorithm (Wathelet, 2008) was chosen. The RTBF and MSPAC methods were implemented complementarily to establish the inversion targets and, in general, a joint inversion of multiple targets was performed for better constraining the inversion. Such targets were carefully selected on a site basis among DC, ellipticity and autocorrelation data depending on the experimental signal quality and spatial resolution limits. For an exhaustive treatment of the HVSR technique and the use of surface wave-based exploration methods, reference is made to the summary publications related to the SESAME (2004) and Interpacific (Foti et al., 2018) projects. All the methods described above are implemented in the software suite used, i.e., Geopsy (Wathelet et al., 2010, 2020).

As an example, the results for EPOZ are shown in Fig. 7. For this site, the fundamental mode Rayleigh- and Love-wave DCs and Rayleigh-wave ellipticity were used as targets of the joint inversion. These targets were defined within the 9-18 Hz frequency range. Considering the estimated maximum phase velocity, the investigation depth was limited to about half of the maximum investigated wavelength, i.e. the uppermost 25-30 m. The V_S profile resulting from the inversion showed a rapid increase in seismic velocity within the uppermost few meters, which is compatible with the very high-frequency resonance peak; then V_S increases linearly with depth. The resulting $V_{S,30}$ is 394 m/s. It is worth noting that the joint inversion of Rayleigh-wave DC and ellipticity data, including the ellipticity peak frequency around 14 Hz estimated through N-HVSR (Fig. 6B), allowed for obtaining a lower uncertainty in the determination of shallower V_S values, whereas a higher uncertainty was obtained for the V_S corresponding to the maximum investigated depths. This is a well-known limit of passive surface wave testing which is characterized by decreasing resolution with depth (Foti et al., 2014). While such non-uniqueness of the inversion solution introduces uncertainty into layer-specific velocity estimates at depth, integrated parameters such as $V_{S,30}$ remain robust across the acceptable model ensemble (Garofalo et al., 2016).

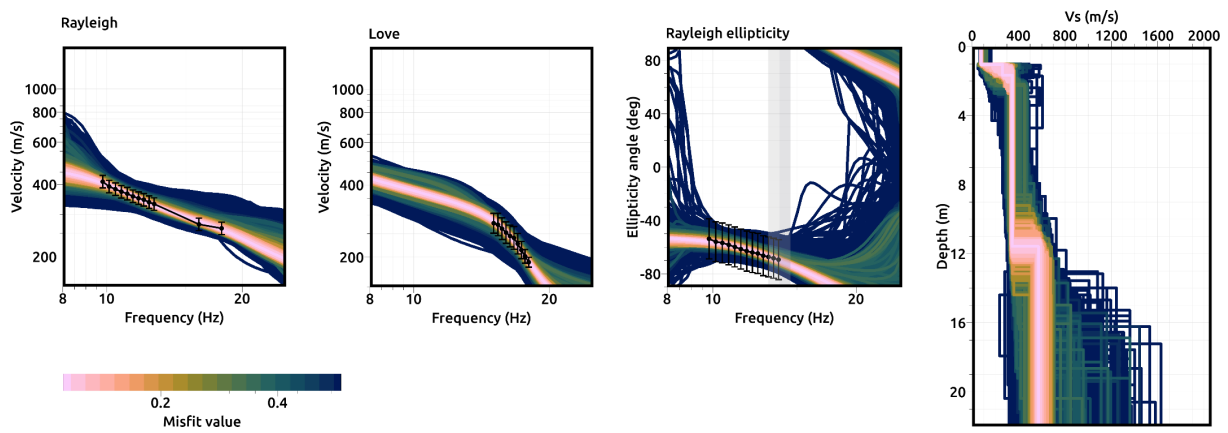


Figure 7. V_S profile resulting from the joint inversion of Rayleigh-, Love-wave DCs and Rayleigh-wave ellipticity data for the EPOZ site.

Applying a similar procedure, we estimated $V_{S,30}$ at the other investigated stations, resulting in 421 m/s (ESAL), 505 m/s (ESML), 386 m/s (ENIC), and 298 m/s (EVRN). Almost all sites can be classified as stiff-soil sites in category B according to EC8 or NTC, with the exception of EVRN, which is classified as soil category C. The inversion results for all characterized stations are provided in the Electronic Supplementary Material (Figs. S2-S5). In addition, the V_S profiles and ambient-noise measurements for the accelerometric stations were made publicly available through the ITACA and ESM websites. Since ESAL is a velocimetric station, its V_S profile is provided in the Electronic Supplementary Material (Fig. S3).

The V_S profiles were additionally used for 1D site-response modeling, and the resulting transfer functions were compared with the empirical amplification functions derived from the analysis of the ground-motion dataset, providing an independent validation of the latter (Laurenzano et al., this issue).

4.2 HVSR analyses

The HVSR functions were also computed following the approach by Tarchini et al. (2024) for both ambient noise recordings and earthquake signals (S-HVSR), using the data streams of seismic networks for all 40 recording stations in the Mt. Etna area. The procedure was applied separately to the two horizontal components (north-south and east-west), which were subsequently vectorially combined to obtain a single horizontal spectrum. For earthquake recordings, HVSRs were calculated using fixed-length time windows of 12 s, starting 0.1 s before the S-wave onset. For ambient noise analysis, windows of the same duration were selected to end 1 s before the P-wave onset, ensuring that only pre-event noise was included. All selected windows were tapered using the same Hanning function prior to spectral analysis and smoothed using the Konno and Ohmachi (1998) algorithm with smoothing parameter

$b = 40$. The spectral amplitudes were computed at the same frequencies as those of the FAS tables, ensuring internal coherence across the entire processing workflow. The resulting HVSR curves are publicly available through the STATION web service (<https://distav.unige.it/rsni/station2025.php>; Tarchini et al., 2024). Before calculating the resonance frequency of the site, we also checked these results with those obtained from the ambient noise analyses for the six stations described in Section 4.1, showing an overall good agreement.

The f_0 was estimated for a subset of 40 stations located in the Mt. Etna area (Fig. 5), providing additional information to be included in the parametric tables. The curves were analyzed following the ITACA procedure (Felicetta et al., 2023), i.e., using a local maximum search algorithm to identify the resonant frequencies of the site, whose reliability is verified using the SESAME (2004) criteria. According to Puglia et al. (2011), each identified frequency is associated with a CW index, ranging between 0 and 1, which defines the bandwidth of the peak. If the HVSR curve shows one or more well-defined peaks of amplitude greater than $2\sqrt{2}$ and $CW < 0.55$, it is classified as peaked or multi-peaked; if $CW \geq 0.55$ the curve is classified as broad-band; finally, if the amplitude never exceeds the value of $2\sqrt{2}$ the curve is classified as flat. The automatic classification was subsequently reviewed manually to ensure the reliability of the results. f_0 was determined by identifying a clear peak in the HVSR curve, evaluated

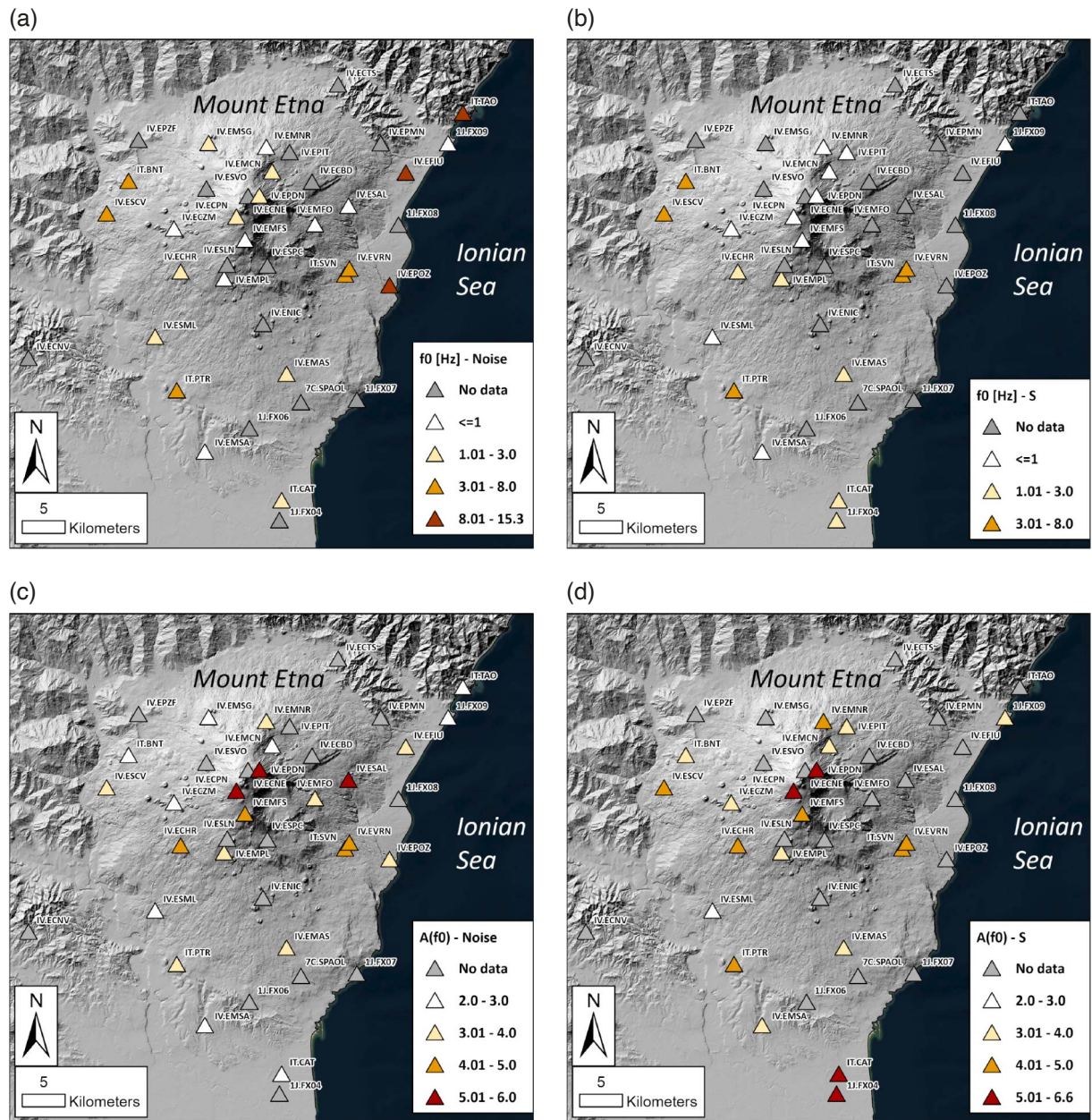


Figure 8. Maps of f_0 (a, b) and $A(f_0)$ (c, d) values for the 40 stations in the Etna area. (a, c) results from N-HVSR; (b, d) results from S-HVSR.

separately for noise and S-wave windows. Only stations exhibiting stable and well-defined peaks were considered reliable, while flat curves or those lacking a clear maximum were classified as having no identifiable f_0 .

Overall, the HVSR analysis reveals a heterogeneous behavior across the Etna stations: several sites show well-defined peaks while others exhibit flat or poorly defined curves, suggesting weak or absent site resonance within the analyzed frequency range (0.5-20.0 Hz). In general (Figs. 8a and 8b), f_0 appears to increase from the summit area of Mt. Etna to the peripheral borders of the volcano: at the summit, all available values are at frequencies below 3 Hz; further inland, the f_0 values range between 3 and 8 Hz; finally, along the coast, the resonance frequencies are higher, exceeding 8 Hz. When f_0 could be estimated from both ambient-noise and S-wave, the two methods generally provide comparable resonance frequencies. More evident differences are observed in the associated peak amplitudes $A(f_0)$ (Figs. 8c and 8d). The largest discrepancies ($\Delta A(f_0) = A(f_0)_{S-HVSR} - A(f_0)_{N-HVSR}$) occur at IT.CAT, 1J.FX09, IV.EMSA, IT.PTR, and IT.BNT, where S-HVSR systematically yields higher amplitudes than N-HVSR. At IT.CAT, $\Delta A(f_0)$ exceeds 3, as the amplitude for S-HVSR is 5.85, whilst N-HVSR is 2.44 (Fig. S6 of the Electronic Supplementary Material). Such a behavior has been previously reported in the literature (e.g., Lachet et al., 1996) who showed that earthquake-derived HVSR amplitudes tend to be systematically higher than those obtained from ambient noise due to differences in the underlying wavefield composition and the more efficient excitation of site resonance by S-waves. However, the relevance of the amplitude differences should be assessed by also considering the uncertainty associated with $A(f_0)$ estimate. Since the S-HVSR and N-HVSR amplitude estimates are independent, the uncertainty associated with $\Delta A(f_0)$ was computed as $\sigma_{\Delta A} = \sqrt{(\sigma^2 A(f_0)_{S-HVSR} + \sigma^2 A(f_0)_{N-HVSR})}$, and the normalized difference was evaluated as $|\Delta A(f_0)|/\sigma_{\Delta A}$. For IT.CAT, this ratio is about 1.4, indicating that the amplitude difference is appreciable with respect to the combined uncertainty, although the individual amplitude estimates are still affected by non-negligible uncertainties.

The high $A(f_0)$ values observed at the summit stations ECPN and EPDN, for both N-HVSR and S-HVSR, likely reflect the combined effects of strong near-surface heterogeneities, intense fracturing, and topographic amplification in the crater area. The consistency of these observations across both ambient noise and earthquake recordings suggests that the observed amplification represents a site-specific response rather than a source-dependent effect. Moreover, moving away from the summit, we observe a progressive decrease in peak amplitude which could suggest more homogeneous subsurface conditions and a reduced influence of topography, resulting in weaker local amplification effects.

4.3 Statistics of the site characterization results

In addition to the HVSR analysis, we assigned a NTC18 topographic category to the 40 stations in the Etna area, by using the ESM topography web service (Mascandola et al., 2021). Figure 9a shows that the topographic classification is dominated by class T1 (flat surface), which includes 29 stations (72.5%), followed by class T3 (hills with a crest width much smaller than the base and inclination $15^\circ \leq i \leq 30^\circ$) with 9 stations (22.5%) and class T4 (hills with a crest width much smaller than the base and $i > 30^\circ$) with 2 stations (5.0%), while class T2 (slope with $i > 15^\circ$) is not represented. The distribution indicates that most of the selected stations are located in topographic settings not associated with significant topographic amplification. Only two stations, IT.TAO and IV.EPMN, fall into topographic class T4, potentially associated with stronger topographic amplification; both are located outside the volcanic cone of Mt. Etna. In particular, IT.TAO is situated within the town of Taormina, where a steep limestone cliff slopes down toward the sea.

As mentioned earlier, all 147 monitoring stations in the dataset were characterized using information available from public databases. In particular, the following parameters were derived from ITACA: i) $V_{S,30}$ values derived from geophysical surveys other than those described in Section 4.1; and ii) $V_{S,30}$ obtained by correlation with topographic slope (Mascandola et al., 2021); EC8 subsoil category derived from iii) geophysical measurements, iv) topography and v) correlation with surface geology at large scale (generally 1:100,000).

Figure 9b shows the frequency distribution of the EC8 soil category for the 147 seismic stations and the subset of the 40 located in the Mt. Etna region. For plot representation, a hierarchy has been established for assigning a single EC8 class to the station, prioritizing the class derived from geophysical testing, followed by topography. For consistency, the same hierarchy was applied for $V_{S,30}$ assignments.

As a result, 27 sites (18.4%) are associated with an EC8 class derived from measured V_S profiles, whereas in the subset of 40 selected stations this applies to 7 sites (17.5%). The dataset is strongly dominated by classes A and B,

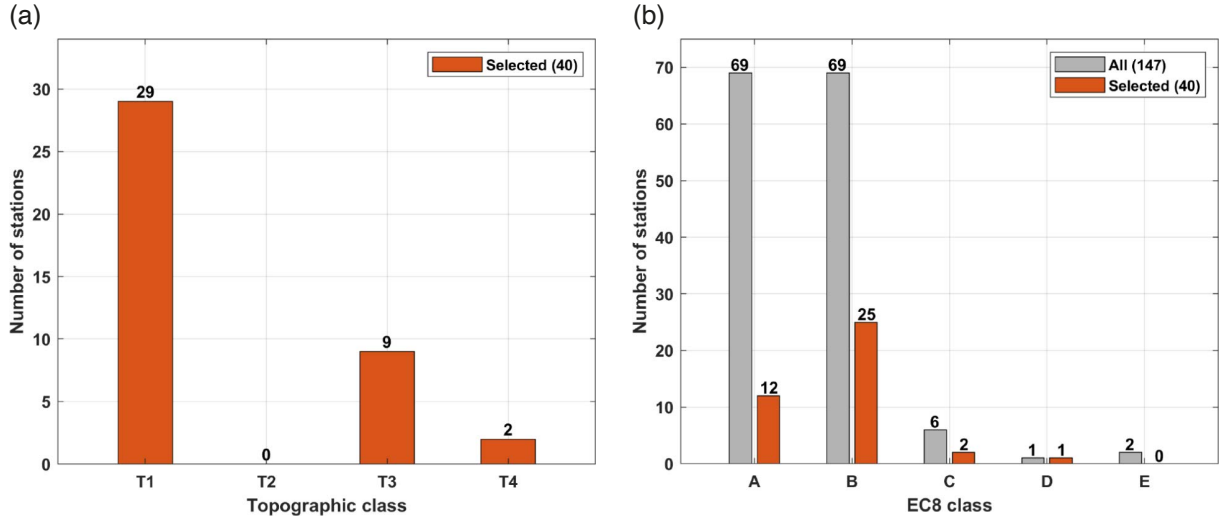


Figure 9. (a) Frequency distribution of the topographic classes for the selected 40 stations. (b) Frequency distribution of the EC8 site classes for the selected subset of 40 stations, compared with the full dataset of 147 stations.

which include 69 stations each and together account for 93.9% of the total. Conversely, classes C, D, and E are poorly represented, with only 6 (4.1%), 1 (0.7%), and 2 (1.4%) stations, respectively. Spatially, classes A and B are widely distributed across the study area, whereas the softer site classes occur only locally. The corresponding $V_{S,30}$ ranges are 804-1500 m/s for class A, 372.8-796 m/s for class B, and 208-353 m/s for class C. Class D is represented by a single station with $V_{S,30} = 159$ m/s, whereas class E includes two stations with $V_{S,30}$ values between 469 and 638 m/s.

A further investigation was carried out, since a classification based on topography can present issues in contexts where steep slopes (such as the flanks of a volcano) do not necessarily correspond to competent rocks. In particular, in order to identify the stations for which the attribution to site class A could derive from a possible overestimation of the $V_{S,30}$ values inferred from slope, the lithological classification proposed by Brunelli et al. (2026) was used. This classification derives from a reworking of the ISPRA (2021) lithological map at the 1:100,000 scale, obtained by aggregating the original 50 lithological classes into 11 broader groups. The aggregation was mainly guided by geophysical and lithological criteria, evaluating the similarity among the mean values of the original classes in terms of $V_{S,30}$ values; for classes represented by a limited number of measurements, the grouping was mainly carried out on a lithological basis.

Figure 10 shows the correspondence matrix between the EC8 classes according to ITACA/ESM and the lithological classification of Brunelli et al. (2026), for the entire dataset and for the subset of stations selected in the Etna area, respectively. With reference to the latter, the sites assigned to subsoil category A (Fig. 10b) fall within the ASM (Argillites, Siltstones, Marls), EMR (Effusive Magmatic Rocks), and LIDO (Limestone and Dolomites) classes. For the sites falling within the ASM and EMR classes, as well as for two additional stations originally classified as A but not covered by the lithological map, the reclassification was supported by the results of Gresta and Langer (2002), also used for the ground motion modeling in Tusa and Langer (2016) and Tusa et al. (2020). Thanks to a large geophysical and geotechnical database available for the province of Catania, derived mainly from down-hole and cross-hole measurements, the authors were able to relate the lithological units to the corresponding geotechnical parameters, in particular, elastic moduli and density, and to develop GIS-based digital maps useful for assigning EC8 subsoil classes.

As a result, starting from the 12 stations assigned to soil category A from topographic proxies, IV.EMPL and IV.ESPC were reclassified as category C, and other 9 stations were reclassified as category B; only the IT.TAO (Taormina) station, which belongs to the LIDO class and is located outside the volcanic cone, was retained in category A. The sites reclassified from A to B were assigned a $V_{S,30}$ of 550 m/s, whereas those reclassified as C were assigned a $V_{S,30}$ of 270 m/s, corresponding to the average values of their respective new soil categories. The $V_{S,30}$ and EC8 values, derived from the defined hierarchy and revised on the basis of detailed studies, are those provided in the parametric tables (the V_{S30_fin} and $EC8_fin$ fields in the flatfiles) and shown in Fig. 5.

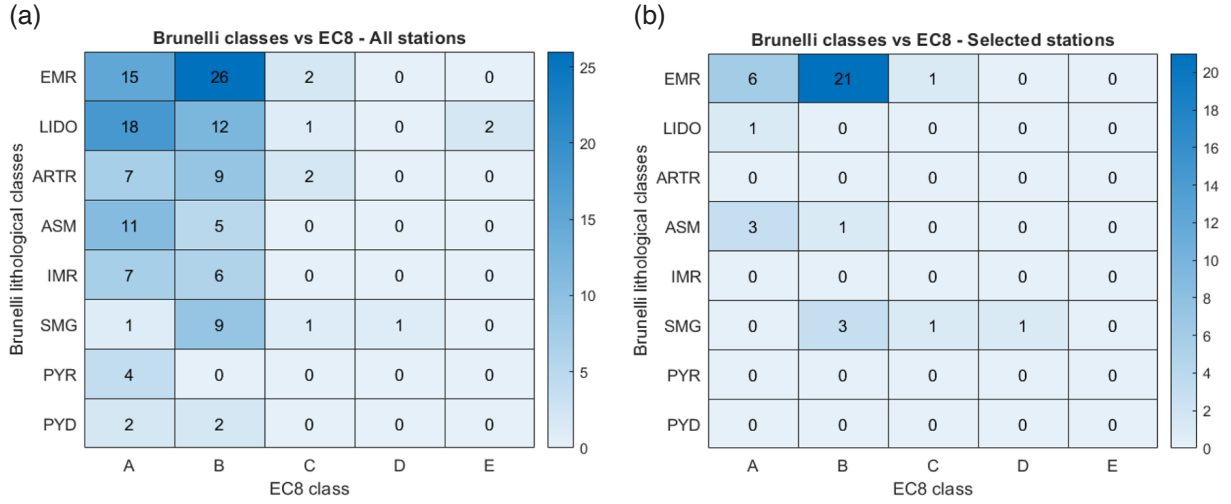


Figure 10. Matrix plot of Brunelli et al. (2026) geologic classification and EC8 soil categories for (a) all stations and (b) the 40 stations in Mt. Etna region (right). EMR: Effusive magmatic rocks; LIDO: Limestone and Dolomites; ARTR: Sandstones, pelitic-arenaceous complexes, calcareous arenaceous complexes, conglomerates, travertine, quartzarenites; ASM: Argillites, siltstones, marls; IMR: Magmatic rocks and metamorphic rocks; SMG: Coarse-to-medium-grained soils; PYR: Pyroclastic rocks; PYD: volcanic pyroclastic deposits.

5. Consistency check

In order to identify potentially problematic recordings, a consistency check based on the statistical analysis of regression residuals was performed, allowing the detection of outliers within the dataset. Similar approaches are commonly applied in the preparation of seismological datasets to detect events, stations, or individual recordings exhibiting anomalous behavior not attributable to physical features of ground motion, including persistent amplification or deamplification effects (e.g., Bindi et al., 2019; Traversa et al., 2020; Mascandola et al., 2022; Lanzano et al., 2025).

We calibrated a median prediction model directly on the dataset to remove any overall median bias in the residuals, thereby facilitating the identification of outliers in the tails of the residual distribution. The functional form was intentionally kept as simple as possible and was selected according to the most important explanatory variables (ML and R). The adopted functional form is:

$$\log_{10} Y = e_1 + F(M) + G(R) + Q(R) + \delta B_e + \delta S2S_s + \delta W_{0,es}, \quad (1)$$

where Y denotes the observed intensity measure (IM), computed as the geometric mean of the horizontal components of PGA and/or PGV. The magnitude scaling $F(M)$ is given by:

$$F(M) = b_1(M - M_{ref}) + b_2(M - M_{ref})^2, \quad (2)$$

and the distance-dependent attenuation scaling, $G(R)$, is defined by the geometrical spreading term:

$$G(c_1; \log_{10} R) = c_1 \log_{10}(R/R_{ref}), \quad (3)$$

and by the anelastic attenuation term:

$$Q(k_1; R) = k_1(R - R_{ref})/100, \quad (4)$$

in which the fixed effects coefficients e_1, b_1, b_2, c_1, k_1 , were obtained from calibration and allow prediction of the median IMs and the associated standard deviations for different M_L and R scenarios. We assigned the parameter-independent

coefficients $R_{ref} = 10$ km (reference distance) and $M_{ref} = 1.5$ (reference magnitude) from a preliminary analysis of the dataset distribution. The model parameters were estimated using a linear mixed-effect regression framework (Bates et al., 2015), introducing random effects for both stations and events, enabling the estimation of a partially non-ergodic variability term following Al Atik et al. (2010). In particular, δB_e in Eq. (1) is the between-event term and represents the average deviation of the recordings of a given earthquake from the median model prediction. The term $\delta S2S_s$ denotes the site-to-site component and describes the systematic bias of the ground motions recorded at a given station s relative to the median model prediction. Finally, the event- and station- corrected leftover residual $\delta W_{0,es}$ accounts for the aleatory variability, as it absorbs contributions to the residuals that are not event or station specific (e.g., path, source directivity or radiation pattern effects). The $\delta S2S_s$ and δB_e residuals approximate normal distributions with zero mean and standard deviations ϕ_s and τ , respectively. Because the random effects are independent distributions, the standard deviation σ of the complete residual distribution is given by:

$$\sigma = \sqrt{(\tau^2 + \phi_s^2 + \phi_0^2)}, \quad (5)$$

in which ϕ_0 is the standard deviation of $\delta W_{0,es}$.

The residuals were computed for both horizontal and vertical ground motion components. However, no significant differences emerged between vertical and horizontal residuals, and only the results for the latter are presented and discussed here (plots for vertical components are available in the Electronic Supplementary Material, Figs. S7 and S8).

Figure 11 shows the trends of the components of the residuals as a function of the explanatory variables for the horizontal component of PGA and PGV. In particular, δB_e and $\delta W_{0,es}$ are plotted as a function of M_L and R , respectively. Large variability in δB_e values is observed at short periods (i.e., high frequencies) and all magnitudes, indicating that part of the source-related variability is not fully captured by the adopted functional form. This behavior likely reflects differences in source properties, such as stress drop or rupture processes, which are not explicitly parameterized in the model and therefore remain absorbed in the δB_e term. This suggests that the adopted magnitude scaling is not sufficient to describe the high-frequency source variability. A relevant bias is also observed at very short distances, where residuals tend to deviate from zero. This suggests that the introduction of a pseudo-depth term in the functional form could improve the fit in the near-field (see ground motion modeling by Lanzano and Luzi, 2020 and Tusa et al., 2020). Since site effects are not explicitly modeled in the median GMM, differences in site response among stations are entirely captured by the $\delta S2S_s$ term. The $\delta S2S_s$ values for the horizontal components of PGA and PGV are also shown in Fig. 11, color-coded and represented with different symbols according to the EC8 site class. A significant dispersion of $\delta S2S_s$ values is observed within each EC8 class, suggesting that the EC8 classification alone is not sufficient to fully describe site effects in the dataset, especially at high frequency (i.e., PGA). This is particularly relevant in volcanic environments, where the stratigraphy can be highly complex due to the alternation and superposition of lava flows, pyroclastic deposits, and volcanic ash layers. Such configurations may produce velocity inversions, while volcanic rocks themselves can exhibit highly variable degrees of weathering and fracturing. These factors can lead to different site responses even for stations assigned to the same EC8 class. Therefore, this aspect requires further investigation before ground motion can be modeled for engineering purposes, both by improving the quality of the metadata associated with station sites and by introducing more advanced parameterizations to describe site effects.

Recurrent outliers with residuals several times larger than the standard deviation of the calibrated model were excluded. Specifically, recordings with absolute residuals $|\delta W_{0,es}|$ exceeding four times the standard deviation ϕ_0 were removed; earthquakes with $|\delta B_e|$ larger than three times τ were excluded; and stations with $|\delta S2S_s|/\phi_s$ greater than 3 were also discarded. It is worth noting that the consistency check was intentionally designed to be conservative, with the aim of removing only clearly anomalous observations while preserving as much of the original dataset as possible. As a result, a limited number of recordings were excluded, ensuring that the statistical representativeness and M_L - R coverage of the dataset (see Fig. 4) remain essentially unchanged. The equivalent residual plots for the post-consistency-check dataset are provided in the Electronic Supplementary Material (Fig. S9).

The final PGA-PGV dataset consists of 4,711 recordings from 200 earthquakes recorded at 147 stations, while the final FAS and SA datasets comprise 4,678 recordings from 193 earthquakes recorded at 134 stations. Overall, the consistency check resulted in the exclusion of 95 event-station recordings, corresponding to approximately 2% of the initial 4,806 recordings. The exclusions were not randomly distributed across the dataset, but involved 22 stations and

28 events. Approximately 50% of the excluded recordings were associated with the temporary MZ network (stations ET01-ET10), while a smaller concentration was observed at a few permanent stations, most notably IV.CMDO (12 recordings). The excluded recordings were also predominantly associated with events that occurred in 2022-2023. No systematic differences were observed between the horizontal and vertical components. The complete list of excluded event-station recordings is provided in Table S10 of the Electronic Supplementary Material.

In addition to the consistency check based on GMM residual values, we performed an additional analysis to further investigate the large variability of the δB_e term, observed in Fig. 11. The regression of Eq. (1) was performed on the FAS ordinates and the frequency dependent δB_e curves were estimated for each event. The resulting set of curves was then analyzed through an unsupervised clustering procedure aimed at identifying recurrent patterns in the frequency-dependent source-related behavior. In this study, we applied the k -means algorithm of Hartigan and

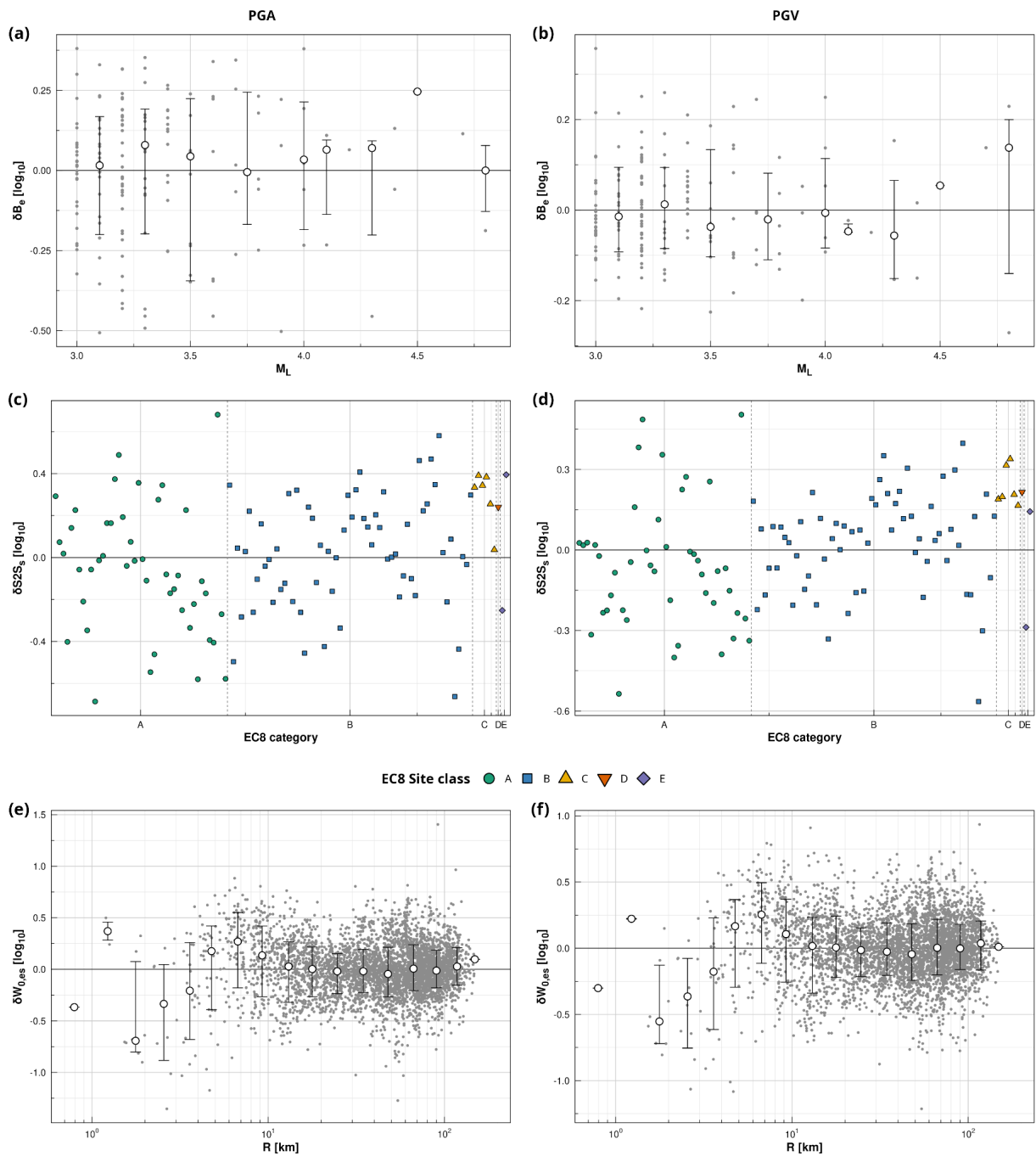


Figure 11. Distribution of residuals from consistency check: (a),(b) δB_e vs. local magnitude; (c),(d) $\delta S2S_s$ vs. EC8 soil categories; (e),(f) $\delta W_{0,es}$ vs. epicentral distance. (a),(c),(e) PGA; (b),(d),(f) PGV.

Wong (1979), as implemented in the R software environment (R Core Team, 2020), which requires only the array of observations and the predefined number of clusters, k , as input. The optimal number of clusters was determined by evaluating the mean silhouette width for values of k ranging from 2 to 6. The maximum mean silhouette value was obtained for $k = 2$, indicating that the dataset is best described by two main families of δB_e spectral shapes.

The mean spectral curves of the two clusters are shown in Fig. 12a, revealing clearly distinct frequency-dependent behaviors, especially at high frequencies. Cluster 1 is characterized by slightly positive residuals at the lowest frequencies, followed by a marked decrease in the intermediate-frequency range, where the average δB_e reaches a pronounced minimum of about -0.5 around 8-15 Hz. At higher frequencies, the residuals progressively increase again toward values close to zero. Cluster 2 exhibits the opposite tendency. After slightly negative values at low frequencies, δB_e increases with frequency and reaches positive values in the intermediate-frequency band (roughly 5-12 Hz), where the average residual attains values of about $+0.3$. At higher frequencies (>20 -30 Hz), the residuals gradually return toward values close to zero. This behavior suggests that the between-event term captures at least two recurrent spectral patterns within the dataset, which may reflect systematic differences in source properties such as stress drop, corner frequency, or rupture complexity.

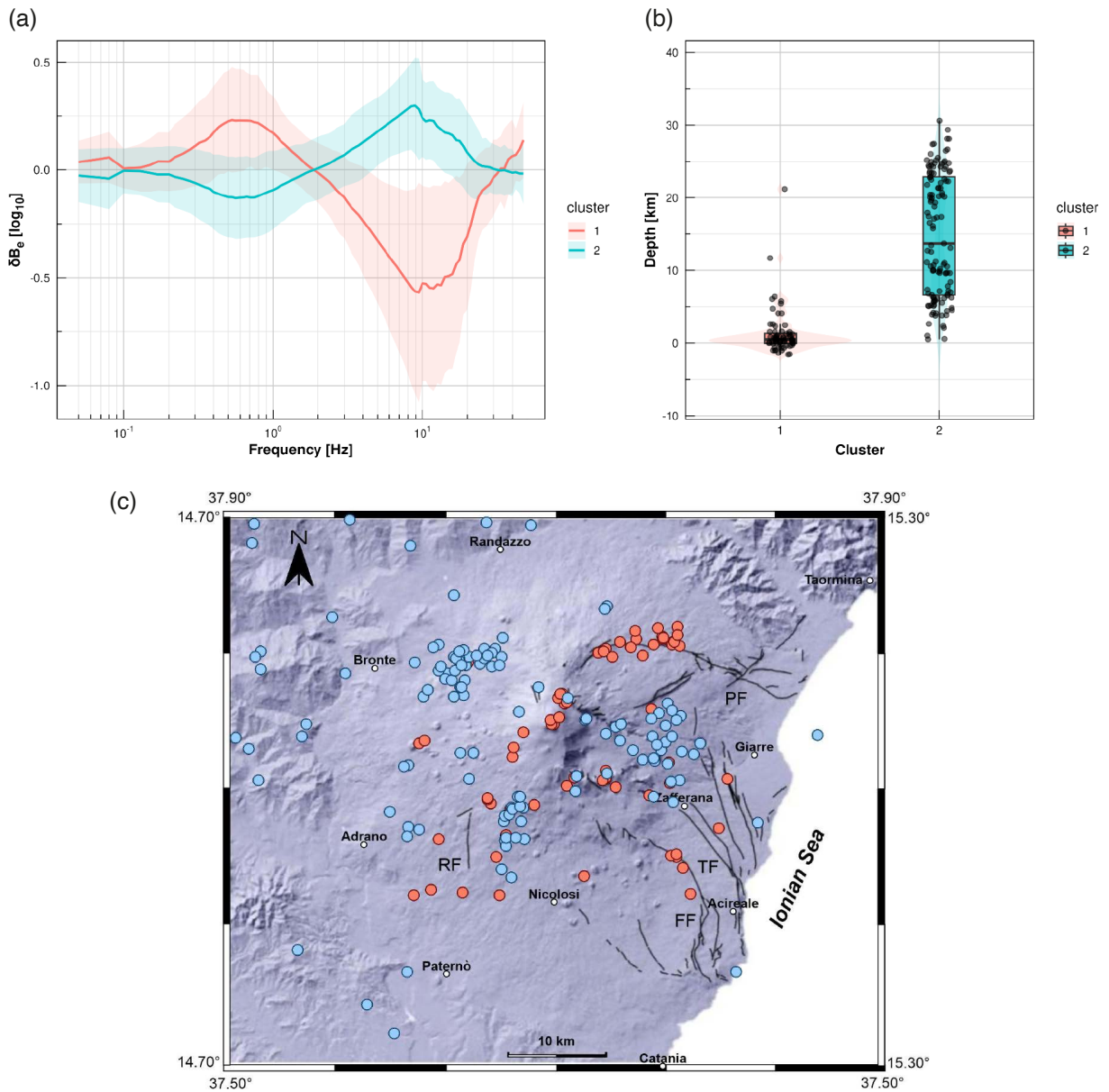


Figure 12. (a) δB_e clusters as a function of frequency; (b) δB_e boxplot of the two clusters as a function of focal depth; (c) map of seismic events belonging to cluster 1 (red circles) and cluster 2 (light blue circles). The major tectonic structures are also shown. Names of the main faults as in Fig. 1.

The resulting clusters were then examined as a function of focal depth in order to assess whether, and to what extent, earthquake depth influences the spectral characteristics of event-specific ground-motion residuals. A comparison with hypocentral depth reveals a clear separation between the two clusters (Fig. 12b): shallower events are preferentially associated with Cluster 1, whereas deeper earthquakes predominantly belong to Cluster 2. To quantify this relationship, a logistic regression model was fitted with cluster membership as the binary response and hypocentral depth as the main predictor, with M_L included as a covariate to account for possible magnitude effects. The fitted model shows a statistically significant dependence on focal depth, while magnitude is not relevant. A transition at ~ 3.8 km, emerging from the data, separates the two clusters, indicating that even small depth variations influence the spectral shape of the between-event residual term. This data-driven result is in line with previous ground motion modeling for volcano-tectonic earthquakes in the Mt. Etna area: Tusa and Langer (2016) defined a hinge depth of 5 km to distinguish between shallow and deep events, for which separate models were calibrated. The same approach was subsequently adopted by Lanzano and Luzi (2020).

The spatial distribution of the two clusters is shown in Fig. 12c and reveals clear geographic patterns of the ground motion characteristics across the study area. Cluster 1 (mainly within the first 3 km of depth) is mainly distributed along the eastern flank of Mt. Etna, with a notable concentration in the northern sector along the Pernicana Fault (PF) and in the southern sector along the Fiandaca Fault (FF). A few epicenters are also observed along the Timpe Fault (TF), whereas a scattered cloud of events is observed near the Ragalna Fault (RF) on the southwestern flank. Cluster 2 (mainly at depths greater than 3 km), by contrast, is more broadly distributed across the volcanic edifice and its surroundings. However, clusters of events are observed in the northwestern, central-eastern, and southeastern sectors of the volcano, where seismicity is generally associated with deeper crustal structures. This spatial partitioning is not random: events sharing similar spectral characteristics – and thus similar source properties – tend to cluster along the same seismogenic structures, suggesting that the observed differences in ground motion behavior are at least partially controlled by the local tectonic and volcanic setting. The shallow seismicity of Cluster 1 may reflect stress concentrations in the uppermost crust associated with volcano-tectonic processes, gravitational instability of the volcanic flanks, or fluid-induced seismicity, all of which are known to produce source characteristics distinct from those of deeper tectonic earthquakes. Conversely, the deeper events of Cluster 2 are more likely rooted in regional crustal tectonics, resulting in systematically different stress drop and rupture properties. These findings underscore the need to account for depth-dependent source variability in ground motion modeling for the Mt. Etna area, and suggest that spatially varying source terms – potentially linked to specific fault systems or volcanic domains – could meaningfully improve the predictive capability of ground motion models calibrated for this region.

6. Dissemination

The parametric tables for the dataset are provided in CSV format, with one file for each spectral ordinate of FAS and SA, plus one additional file for PGA and PGV. The metadata included in the tables are divided into three groups (event, station and waveform); the column headers for each metadata field are explained in a support file named README.txt. Ground-motion parameters are reported for each of the three components (Z, NS, and EW), in physical units (cm/s^2 for PGA and SA, cm/s for PGV and FAS). The tables are included in a zip file with a user manual. The compressed files are available via a dedicated web page (https://shake.mi.ingv.it/volcanic-earthquakes/mt_etna_flatfile.html), where all relevant product information is available, in accordance with the FAIR principles (Wilkinson et al., 2016). The products are citable using a Digital Object Identifier (DOI), with the full citation given in the reference list as Tarchini et al. (2026). The Etna ground-motion parametric tables are licensed under the terms of the Creative Commons Attribution 4.0 International (CC BY 4.0) License. This means that the user is free to share (copy and redistribute the material in any medium or format) and adapt (remix, transform, and build upon the material) for any purpose.

7. Conclusions

In this study, we presented a new dataset of ground-motion recordings for the Mt. Etna volcanic area, specifically designed to support seismological investigations and seismic microzonation applications. The dataset integrates several thousand high-quality three-component recordings from 200 volcano-tectonic earthquakes ($M_L \geq 3.0$) that

occurred between 2008 and 2023, processed through a consistent and transparent workflow and organized in the form of parametric tables including peak and spectral intensity measures together with comprehensive metadata.

A key strength of the proposed dataset lies in the integration of weak- and strong-motion recordings from dense local networks, which allows a robust characterization of near-source ground motion in a volcanic environment. The multi-level site characterization, combining information from existing databases, HVSR analyses, and dedicated geophysical surveys, further enhances the value of the dataset by providing constraints on local site response, which is a critical aspect for seismic microzonation studies.

The internal consistency of the dataset was assessed through a mixed-effects regression framework, enabling the identification of anomalous recordings while preserving the statistical representativeness of the data. Beyond quality control, the analysis highlighted a significant variability of the between-event residuals, which can be interpreted in terms of source-related processes not fully captured by simple predictive models. In this context, the identification of two distinct clusters of events, primarily controlled by focal depth, provides insights into the frequency-dependent characteristics of ground motion in the Etna region and suggests that depth-dependent parameterizations may be required in future ground-motion models for volcanic areas.

The dataset offers a wide range of potential applications. From a seismological perspective, it can be used to investigate source, path, and site effects in volcanic environments, as well as to explore the variability of ground motion at short distances. From an engineering standpoint, it provides a consistent basis for the development and calibration of ground-motion models tailored to volcano-tectonic seismicity, for the selection of records in structural analyses, and for the estimation of empirical amplification factors within seismic microzonation studies. The availability of spectral and peak parameters over a broad frequency range makes the dataset particularly suitable for both research and applied purposes.

Future developments will focus on further improving the characterization of site conditions, including the integration of additional geophysical measurements and advanced proxies for site response, as well as on refining ground-motion modeling approaches by incorporating additional explanatory variables, such as focal depth, source parameters, and alternative distance metrics. The dataset may also be progressively expanded as new recordings become available, ensuring its continuous update and long-term relevance. In this perspective, the dissemination of the dataset according to FAIR principles and through openly accessible platforms ensures its reusability and promotes its integration into broader national and international initiatives on ground-motion data sharing.

Overall, the proposed dataset represents a significant contribution toward bridging the gap between seismological observations and engineering applications in volcanic regions, providing a solid and flexible framework for future studies on ground-motion characterization and seismic hazard assessment at Mt. Etna and in similar tectonic environments.

Data and sharing resources. We used data and information retrieved from the Observatories and Research Facilities for European Seismology – European Integrated Data Archive (ORFEUS-EIDA; <https://www.orfeus-eu.org/data/eida/>) and the Italian Civil Protection Department (<http://ran.protezionecivile.it/EN/index.php>).

The bulletin of the Istituto Nazionale di Geofisica e Vulcanologia (INGV) serves as a reference for downloading the data (<http://webservices.ingv.it/fdsnws/event/1/>) and extracting the earthquake locations. The specifications of the International Federation of Digital Seismograph Networks (FDSN) can be found at <http://www.fdsn.org/> and the Standard for the Exchange of Earthquake Data (SEED) manual can be downloaded at http://www.fdsn.org/pdf/SEEDManual_V2.4.pdf.

We used data from networks with FDSN code 1J (Moretti et al., 2021), 3D (Alparone et al., 2021), 7C (Costanzo et al., 2022), IT (Presidency of Council of Ministers – Civil Protection Department, 1972), IV (Istituto Nazionale di Geofisica e Vulcanologia [INGV], 2005), IY (Università Della Calabria, Italy, 1981), MN (MedNet Project Partner Institutions, 1990), and TV (Istituto Nazionale di Geofisica e Vulcanologia [INGV], 2008; <https://terremoti.ingv.it/instruments/network/TV>). Plots were made using ArcGIS Pro 3.4.2 (Esri Inc., 2024), R (R Core Team, 2020), MATLAB version 9.11.0 (R2021b) (The MathWorks Inc., 2021), and CorelDRAW® Graphics Suite 2020 version 22.1.1.523.

Acknowledgements. This work was carried out within the project entitled “Attività di supporto al DRPC Sicilia per la prevenzione e riduzione del rischio sismico nell’ambito del Piano Regionale di Microzonazione Sismica per la realizzazione di studi Microzonazione Sismica, di cui al programma nazionale di prevenzione del rischio sismico (ex art. 11, Legge 77/2009)” coordinated by Centro di Microzonazione Sismica (CentroMS).

The Authors thank the Site effect (ESITO) laboratory of the INGV (<https://www.ingv.it/en/monitoraggio-e-infrastrutture/laboratori/laboratorio-effetti-di-sito>) for preparing, configuring and providing the seismic instrumentation, and for generating the output files and state-of-health files. The ESITO Laboratory is part of the Integrated Laboratories for Geosciences and Environment-ILGE of MEET Project (Monitoring Earth's Evolution and Tectonics, <https://meet.ingv.it>; National Recovery and Resilience Plan (PNRR), Mission 4, Component 2, Investment line 3.1 - project code IR0000025).

References

- Aki, K. (1957). Space and time spectra of stationary stochastic waves, with special reference to microtremors, *Bull. Earthquake Res. Inst.*, 35, 415-456.
- Al Atik, L., N. Abrahamson, J. J. Bommer, F. Scherbaum et al. (2010). The variability of ground-motion prediction models and its components, *Seismol. Res. Lett.*, 81, 5, 794-801, doi:10.1785/gssrl.81.5.794.
- Alparone, S. C., G. Barberi, S. D'Amico, G. Di Grazia et al. (2020). Mt. Etna Revised and Concise Seismic Catalog from 1999 (EtnaRCSC) [Data set]. Istituto Nazionale di Geofisica e Vulcanologia (INGV), doi:10.13127/ETNASC/ETNARCS.
- Alparone, S., O. M. G. Cocina, D. Contrafatto, G. Larocca et al. (2021). Seismic data acquired by the INGV temporary network for seismic monitoring of the Island of Vulcano to increase volcanic activity (Sicily, Italy) – 2021, Istituto Nazionale di Geofisica e Vulcanologia (INGV), doi:10.13127/sd/cfz7ftsw-y.
- Azzaro, R., S. D'Amico, L. Peruzza and T. Tuvè (2013). Probabilistic seismic hazard at Mt. Etna (Italy): the contribution of local fault activity in mid-term assessment, *J. Volc. Geotherm. Res.*, 251, 158-169, doi:10.1016/j.jvolgeores.2012.06.005.
- Azzaro, R., S. D'Amico, A. Mostaccio, L. Scarfi et al. (2014). Macroseismic surveys of earthquakes occurred in Eastern Sicily from January 2009 to December 2013, *Quaderni di Geofisica*, 120, doi:10.13127/qdg/120.
- Azzaro, R., S. D'Amico, T. Tuvè, L. Scarfi et al. (2020). Macroseismic surveys of earthquakes occurred in Eastern Sicily from January 2014 to December 2018, *Quaderni di Geofisica*, 160, 1-62, doi:10.13127/qdg/160.
- Azzaro, R. et al. (2026). Seismotectonic setting and seismic hazard in the Etna region: constraints for the definition of the seismic input for microzonation. *Ann. Geophys.*, 69, 6, this issue.
- Barberi, G., G. Di Grazia, F. Ferrari, E. Giampiccolo et al. (2020). Mt. Etna Revised Seismic Catalog from 2020 (EtnaRSC2020) [Data set]. Istituto Nazionale di Geofisica e Vulcanologia (INGV), doi:10.13127/etnasc/etnarsc2020.
- Bard, P.-Y., M. Campillo, F. Cotton, P. Guéguen et al. (2010). From non-invasive site characterization to site amplification: recent advances in the use of ambient vibration measurements, in M. Garevski and A. Ansal (Editors), *Earthquake Engineering in Europe, Geotechnical, Geological, and Earthquake Engineering*, 17, Springer, Dordrecht, 105-123, doi:10.1007/978-90-481-9544-2_5.
- Bates, D., M. Mächler, B. Bolker and S. Walker (2015). Fitting linear mixed-effects models using lme4, *J. Stat. Softw.*, 67, 1, 1-48, doi:10.18637/jss.v067.i01.
- Bettig, B., P.-Y. Bard, F. Scherbaum, J. Riepl et al. (2001). Analysis of dense array noise measurements using the modified spatial autocorrelation method (SPAC): application to the Grenoble area, *Boll. Geofis. Teor. Appl.*, 42, 281-304.
- Bindi, D., S. R. Kotha, G. Weatherill, G. Lanzano et al. (2019). The pan-European engineering strong motion (ESM) flatfile: consistency check via residual analysis, *Bull. Earthq. Eng.*, 17, 2, 583-602, doi:10.1007/s10518-018-0466-x.
- Brunelli, G., S. Sgobba, E. M. Thompson, D. J. Wald et al. (2026). Evaluation of proxies of the averaged shear-wave velocity to a depth of 30 m (Vs30) for site characterization in Italy, *Bull. Geophys. Oceanogr.*, 67, 2, 295-324, doi:10.4430/bgo00534.
- Capon, J. (1969). High-resolution frequency-wavenumber spectrum analysis, *Proc. IEEE*, 57, 8, 1408-1418, doi:10.1109/PROC.1969.7278.
- Costanzo, A., S. Falcone, C. La Piana, A. Caserta et al. (2022). Urban Seismic Network (RUSN-CAL-INGV) - Real time Urban Seismic Network managed by CALabrian headquarters of INGV for monitoring urban areas, critical infrastructures and cultural heritage sites, Istituto Nazionale di Geofisica e Vulcanologia (INGV), doi:10.13127/sd/o5nrhmp9n7.
- D'Amico, S., R. Azzaro, G. Tusa, T. Tuvè and E. Varini (2025). Volcano-tectonic seismicity and related hazard: a component of the multi-hazard assessment in the highly exposed region of Mt. Etna (Italy), *Ann. Geophys.*, 68, 1, V101, doi:10.4401/ag-9152.

- Di Capua, G., G. Lanzo, V. Pessina, S. Peppoloni et al. (2011). The recording stations of the Italian strong motion network: geological information and site classification, *Bull. Earthquake Eng.*, 9, 1779-1796, doi:10.1007/s10518-011-9326-7.
- Esri Inc. (2024). ArcGIS Pro, Version 3.4.2. Redlands, CA: Environmental Systems Research Institute Inc.
- Felicetta, C., E. Russo, M. D'Amico, S. Sgobba et al. (2023) Italian Accelerometric Archive v4.0 - Istituto Nazionale di Geofisica e Vulcanologia, Dipartimento della Protezione Civile Nazionale, doi:10.13127/itaca.4.0.
- Firetto Carlino, M., L. Scarfi, F. Cannavò, G. Barberi et al. (2022). Frequency-magnitude distribution of earthquakes at Etna volcano unravels critical stress changes along magma pathways, *Commun. Earth. Environ.*, 3, 68, doi:10.1038/s43247-022-00398-6.
- Foti, S., C. G. Lai, G. Rix and C. Strobbia (2014). Surface wave methods for near-surface site characterization. Boca Raton, FL, USA: CRC Press.
- Foti, S., F. Hollender, F. Garofalo, D. Albarello et al. (2018). Guidelines for the good practice of surface wave analysis: a product of the InterPACIFIC project, *Bull. Earthquake Eng.*, 16, 6, 2367-2420, doi:10.1007/s10518-017-0206-7.
- Gambino, S. and G. Distefano (2022). Intrusive Seismic Swarms as Possible Precursors of Destructive Earthquakes on Mt. Etna's Eastern Flank, *Int. J. Geophys.*, 8565536, doi:10.1155/2022/8565536.
- Garofalo, F., S. Foti, F. Hollender, P.-Y. Bard et al. (2016) Interpacific project: comparison of invasive and non-invasive methods for seismic site characterization, Part I: intra-comparison of surface wave methods, *Soil Dyn. Earthq. Eng.*, 82, 222-240, doi:10.1016/j.soildyn.2015.12.010.
- Giampiccolo, E., O. Cocina, P. De Gori and C. Chiarabba (2020). Dyke intrusion and stress-induced collapse of volcano flanks: the example of the 2018 event at Mt. Etna (Sicily, Italy), *Sci. Rep.*, 10, doi:10.1038/s41598-020-63371-3.
- Gresta, S. and H. Langer (2002). Assessment of seismic potential in south eastern Sicily, in C.A. Brebbia (Editor), *Risk Analysis III*, WIT Press, Southampton, 93-101.
- Hartigan, J. A. and M. A. Wong (1979). Algorithm AS 136: a k-means clustering algorithm, *J. R. Stat. Soc. Ser. C Appl. Stat.*, 28, 1, 100-108, doi:10.2307/2346830.
- Hirn, A., A. Nercissian, M. Sapin, F. Ferrucci et al. (1991). Seismic heterogeneity of Mt. Etna: structure and activity, *Geophys. J. Int.*, 105, 139-153, doi:10.1111/j.1365-246X.1991.tb03450.x.
- ISPRA (2021). Lithological map of Italy at 1:100,000 scale, https://inspire-geoportal.ec.europa.eu/srv/ita/catalog.search#/metadata/ispra_rm:Meta_Geo_SV000004_SVD.
- Istituto Nazionale di Geofisica e Vulcanologia (INGV) (2005). Rete Sismica Nazionale (RSN), Istituto Nazionale di Geofisica e Vulcanologia (INGV) (in Italian), doi:10.13127/sd/x0fxnh7qfy.
- Istituto Nazionale di Geofisica e Vulcanologia (INGV) (2008). INGV experiments network, International Federation of Digital Seismograph Networks.
- Köhler, A., M. Ohrnberger, F. Scherbaum, M. Wathelet and C. Cornou (2007). Assessing the reliability of the modified three-component spatial autocorrelation technique, *Geophys. J. Int.*, 168, 2, 779-796, doi:10.1111/j.1365-246X.2006.03253.x.
- Konno, K. and T. Ohmachi (1998). Ground-motion characteristics estimated from spectral ratio between horizontal and vertical components of microtremor, *Bull. Seismol. Soc. Am.*, 88, 1, 228-241, doi:10.1785/BSSA0880010228.
- Lachet, C., D. Hatzfeld, P.-Y. Bard, N. Theodulidis et al. (1996). Site effects and microzonation in the City of Thessaloniki (Greece) comparison of different approaches, *Bull. Seism. Soc. Am.*, 86, 1692-1703, doi:10.1785/BSSA0860061692.
- Lahr, J. C. (1989). Hypoellipse Version 2.0*: a computer program for determining local earthquake hypocentral parameters, magnitude, and first motion pattern, U.S. Geological Survey Open-File Report, 89-116, doi:10.3133/ofr89116.
- Lanzano, G. and L. Luzi (2020). A ground motion model for volcanic areas in Italy, *Bull. Earthq. Eng.*, 18, 57-76, doi:10.1007/s10518-019-00735-9.
- Lanzano, G., L. Vitrano, D. Bindi, C. Felicetta et al. (2025). ITACAext 2.0: A High-Quality Parametric Table of Strong-to-Weak Motion Recordings for Seismological and Engineering Research in Italy, *Seismol. Res. Lett.*, 96, 5, 3295-3316, doi:10.1785/0220240309.
- Laurenzano, G., G. Brunelli, G. Lanzano, C. Barnaba et al. (2026). Empirical amplification factors for seismic microzonation studies in volcanic regions: the study case of Etna mountain, *Ann. Geophys.*, 69, 6, this issue.
- Mascandola, C., L. Luzi, C. Felicetta and F. Pacor (2021). A GIS procedure for the topographic classification of Italy, according to the seismic code provisions, *Soil. Dyn. Earthq. Eng.*, 148, 106848, doi:10.1016/j.soildyn.2021.106848.
- Mascandola, C., G. Lanzano and F. Pacor (2022). Consistency Check of ITACAext, the Flatfile of the Italian Accelerometric Archive, *Geosci.*, 12, 9, 334, doi:10.3390/geosciences12090334.

- Mascandola, C., C. Felicetta, E. Russo, L. Luzi et al. (2026). Engineering Strong Motion Database (ESM), version 3.0, Istituto Nazionale di Geofisica e Vulcanologia (INGV), doi:10.13127/esm.3.
- McNutt, S. R. and D. C. Roman (2015). Volcanic Seismicity. In H. Sigurdsson, B. Houghton, H. Rymer, J. Stix and S. McNutt (Eds.), *The Encyclopedia of Volcanoes*, 1011-1034, Academic Press, doi:10.1016/B978-0-12-385938-9.00059-6.
- MedNet Project Partner Institutions (1990). Mediterranean very broadband seismographic network (MedNet), Istituto Nazionale di Geofisica e Vulcanologia (INGV), doi:10.13127/sd/fbbtdtd6q.
- Moretti, M., L. Margheriti, S. Alparone, A. Costanzo et al. (2021). Seismic data acquired by FocusX temporary land-network (FXland), Southern Italy, Istituto Nazionale di Geofisica e Vulcanologia (INGV), doi:10.13127/sd/o5qwm6wjcd.
- Pacor, F., D. Spallarossa, A. Oth, L. Luzi et al. (2016). Spectral models for ground motion prediction in the L'Aquila region (central Italy): evidence for stress-drop dependence on magnitude and depth, *Geophys. J. Int.*, 204, 697-718, doi:10.1093/gji/ggv448.
- Presidency of Council of Ministers-Civil Protection Department (1972). Italian strong motion network, International Federation of Digital Seismograph Networks, doi:10.7914/SN/IT.
- Puglia, R., D. Albarello, A. Gorini, L. Luzi et al. (2011). Extensive characterization of Italian accelerometric stations from single-station ambient-vibration measurements, *Bull. Earthq. Eng.*, 9, 1821-1838, doi:10.1007/s10518-011-9305-z.
- R Core Team (2020). R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria. <https://www.r-project.org/>.
- Ramadan, F., G. Lanzano and S. Sgobba (2023). Vertical seismic ground shaking in the volcanic areas of Italy: prediction equations and PSHA examples, *Soil Dyn. Earthq. Eng.*, 175, 108228, doi:10.1016/j.soildyn.2023.108228.
- Rovida, A., M. Locati, R. Camassi, B. Lolli et al. (2022). Catalogo Parametrico dei Terremoti Italiani (CPTI15), versione 4.0 [Data set]. Istituto Nazionale di Geofisica e Vulcanologia (INGV), doi:10.13127/cpti/cpti15.4.
- Scarfì, L., G. Barreca, C. Cassisi and H. Langer (2025). Earthquake clustering and structural modelling unravel volcano-tectonic complexity beneath Mount Etna, *Sci. Rep.*, 15, 18607, doi:10.1038/s41598-025-03321-z.
- SESAME (2004). Guidelines for the implementation of the H/V spectral ratio technique on ambient vibrations: measurements, processing and interpretation, SESAME European Research Project EVG1-CT-2000-00026, Deliverable D23.12, 62. Available at: https://sesame.geopsy.org/Delivrables/Del-D23-HV_User_Guidelines.pdf.
- Strollo, A., D. Cambaz, J. Clinton, P. Danecsek et al. (2021). EIDA: The European integrated data archive and service infrastructure within ORFEUS, *Seismol. Res. Lett.*, 92, 3, 1788-1795, doi:10.1785/0220200413.
- Tarchini, G., D. Scafidi, S. Parolai, M. Picozzi et al. (2024). The Seismic Station and Site Amplification Service: Continuously Updated Information on Italian Seismic Stations, *Seismol. Res. Lett.*, 96, 2027-2038, doi:10.1785/0220240291.
- Tarchini, G., G. Brunelli, G. Lanzano, G. Tusa et al. (2026). Mt_Etna_flatfile: Parametric table of the intensity measures and associated metadata of waveforms of the earthquake ($M_L \geq 3$) occurred in Mount Etna region in the time interval 2008-2023. [Data set]. Istituto Nazionale di Geofisica e Vulcanologia (INGV), doi:10.13127/volcanic_gm_data/mt_etna.1.0.
- The MathWorks Inc. (2021). MATLAB (R2021b). Natick, Massachusetts, United States.
- Totaro, C., M. Aloisi, C. Ferlito, B. Orecchio et al. (2024). 3D seismic velocity models from local earthquake tomography furnish new insights on the Mount Etna volcano (Southern Italy), *Sci. Rep.*, 14, doi:10.1038/s41598-024-74349-w.
- Traversa, P., E. Maufroy, F. Hollender, V. Perron et al. (2020). RESIF RAP and RLBP dataset of earthquake ground motion in mainland France, *Seismol. Res. Lett.*, 91, 4, 2409-2424, doi:10.1785/0220190367.
- Tusa, G. and H. Langer (2016). Prediction of ground motion parameters for the volcanic area of Mount Etna, *J. Seismol.*, 20, 1, 1-42, doi:10.1007/s10950-015-9508-x.
- Tusa, G., H. Langer and R. Azzaro (2020). Localizing ground-motion models in volcanic terranes: shallow events at Mt. Etna, Italy, revisited, *Bull. Seismol. Soc. Am.*, 110, 6, 2843-2861, doi:10.1785/0120190325.
- Università Della Calabria, Italy (1981). Rete Sismica Unical, International Federation of Digital Seismograph Networks (in Italian), doi:10.7914/SN/IT.
- Wathelet, M. (2008). An improved neighborhood algorithm: parameter conditions and dynamic scaling, *Geophys. Res. Lett.*, 35, L09301, doi:10.1029/2008GL033256.
- Wathelet, M., M. Anthymidis, P.-Y. Bard, H. Cadet et al. (2010). Geopsy online documentation. Available at: <https://www.geopsy.org> (last accessed December 2025).

- Wathelet, M., B. Guillier, P. Roux, C. Cornou et al. (2018). Rayleigh wave three-component beamforming: signed ellipticity assessment from high-resolution frequency-wavenumber processing of ambient vibration arrays, *Geophys. J. Int.*, 215, 1, 507-523, doi:10.1093/gji/ggy286.
- Wathelet, M., J.-L. Chatelain, C. Cornou, G. Di Giulio et al. (2020). Geopsy: a user-friendly open-source tool set for ambient vibration processing, *Seismol. Res. Lett.*, 91, 3, 1878-1889, doi:10.1785/0220190360.
- Wilkinson, M., M. Dumontier, I. Aalbersberg, G. Appleton et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship, *Sci. Data*, 3, 160018, doi:10.1038/sdata.2016.18.

***CORRESPONDING AUTHOR: Gabriele TARCHINI,**

University of Genoa, Department of Earth, Environmental and Life Sciences, Genoa, Italy

e-mail: gabriele.tarchini@edu.unige.it

© 2026 the Author(s).

Open Access. This article is licensed under a Creative Commons Attribution 4.0 International License