Bulletin of Geophysics and Oceanography Vol. XX, n. X, pp. XXX-XXX; Xxxxx 20xx
DOI 10.4430/bgo00484

SDQ: a new tool for the evaluation of seismic-accelerometric data quality

F. Varchetta[1], M. Massa[1], R. Puglia[1], P. Danecek[2], S. Rao[2], A. Mandiello[2] and D. Piccinini[3]

[1] *INGV, Istituto Nazionale di Geofisica e Vulcanologia, Sezione di Milano, Italy*
[2] *INGV, Istituto Nazionale di Geofisica e Vulcanologia, Osservatorio Nazionale Terremoti, Roma, Italy*
[3] *INGV, Istituto Nazionale di Geofisica e Vulcanologia, Sezione di Pisa, Italy*

(Received: 13 June 2024; accepted: 20 November 2024; published online: 28 January 2025)

ABSTRACT This work introduces the Seismic Data Quality (SDQ) project, an open-source Python tool package designed to evaluate the performance of co-located accelerometric and velocimetric stations, assess data quality, and support technicians in the seismic surveillance room at the Rome headquarters. SDQ conducts these verifications through analyses of seismic events and continuous data streams. For earthquake data, SDQ compares ground motion parameters from co-located sensors, subsequently classifying waveforms into quality classes. Continuous data verification relies on ambient noise streams and quality metrics, such as percentage gap, data availability, root-mean-square, and power spectral density. The SDQ tool was tested using data from 200 stations of the National Institute of Geophysics and Volcanology (Italy), each with six channels, within the National Seismic Network, analysing over 15,000 waveforms from Italian earthquakes with local or moment magnitudes (M_L and/or M_W) ≥ 3.5 and hypocentral distances (R_{ipo}) ≤ 150 km, recorded from 2012 to 2023. Earthquake data and station metadata are sourced from the International Federation of Digital Seismograph Networks web services (https://terremoti.ingv.it/webservices_and_software). Continuous data analyses utilise daily miniSEED recordings from each station, classifying waveforms into quality categories based on various metrics. SDQ generates summary tables, images, and explanatory text files for both seismic events and continuous data. SDQ is available for download at https://gitlab.rm.ingv.it/EIDA/quality/sdq.

Key words: seismic data quality check, seismic station, monitoring network.

1. Introduction

Italy faces medium to high seismic activity, resulting in notable social and economic repercussions from recent significant earthquakes such as Aquila in 2009 (Ameri *et al*., 2009), Emilia in 2012 (Luzi *et al*., 2013), and Amatrice-Norcia in 2016 (Morasca *et al*., 2019). Consequently, in recent years, the National Institute of Geophysics and Volcanology [INGV (www.ingv.it)], has intensified its efforts to improve the coverage of the National Seismic Network (RSN, https://eida.ingv.it/it/networks/network/IV) in the Italian territory. This initiative aims to enhance monitoring capabilities, providing essential data for both research and civil protection purposes. Ensuring data reliability is crucial for the effective functioning of these stations. Consequently, recent efforts have focused on improving seismic data processing and verification methods.

These refinements aim to enhance the effectiveness and accuracy of analysing strong-motion data, which is particularly crucial during significant earthquakes.

To date, two tools are available in Italy and are structured in such a way as to be able to daily analyse all RSN data for preliminary data quality estimation, namely the European Integrated Data Archive in Italy [EIDA Italia (Danecek *et al.*, 2021; https://eida.ingv.it/it/getdata)] and INGV Strong Motion Data quality [ISMDq (Massa *et al.*, 2022) http://ismd.mi.ingv.it/quality.php]. The EIDA Italia node is an infrastructure from which selected portions of seismic data (raw format, counts unit) can be downloaded from the continuous stream at each station. EIDA Italia uses a tool for data quality assessment developed within the ORFEUS (https://www.orfeus-eu.org/data/eida/quality/) framework, which is capable of providing information on root-mean-square (*RMS*) values and any signal gaps for each available station with a delay of approximately seven days. ISMDq represents a recent quasi-real-time data quality analysis system developed for events with magnitude >3 recorded by RSN seismo-accelerometric stations and for continuous data recorded by all six-channel RSN stations and associated partners managing regional seismic networks. ISMDq daily publishes, for all available working stations, quality parameters including *RMS*, power spectral density (PSD), probability density function (PDF) (McNamara *et al*., 2004), as well as information on data availability and gaps.

This study aims to provide an additional tool to support seismic monitoring network operators in the seismic surveillance room at the Rome headquarters (Margheriti *et al*., 2021) by enabling rapid and effective assessments of the proper functioning of seismic stations. These verifications are based on the comparison between accelerometer and velocimeter data from seismic events recorded at the same stations and on the observation of the behaviour of time-series of specific quality metrics.

The proposed approach consists in the use of two semi-automatic Python codes currently available on the GitLab platform of the INGV internal network (https://gitlab.rm.ingv.it/EIDA/quality/sdq): Seismic Data Quality_ (SDQ_) event, regarding analysis based on seismic events, and SDQ_stream, regarding analysis based on ambient noise streams.

In detail, the SDQ_event code is structured in subsequent phases (Varchetta *et al*., 2024):

1) data and metadata downloading from the EIDA Italia node (https://www.eida.ingv.it/);
2) pre-processing of the available earthquake waveforms;
3) application of waveform selection criteria to accept or exclude selected data;
4) evaluation of compatibility between accelerometer and velocimeter;
5) evaluation of the quality class for each waveform;
6) results in terms of summary tables, text files, and images.

The proposed code has been tested using a data set of 640 seismic events that occurred in Italy between 2012 and 2023, resulting in about 15,000 waveforms related to earthquakes with a magnitude greater than 3.5 and hypocentral distance smaller than 150 km.

The SDQ_stream code is structured in subsequent phases:

1) downloading continuous data streams and related metadata from the EIDA Italia node, a SeisComP remote or local data archive (https://www.seiscomp.de/doc/), or from an owner data archive;
2) evaluation of quality metrics and evaluation of data quality class;
3) results in terms of summary tables, text files, and daily images.

The proposed code has been tested considering a continuous 24-hour data stream recorded from RSN seismo-accelerometric stations.

2. National Seismic Network

RSN is the permanent INGV seismic network (Michelini *et al.*, 2016). Its purposes range from the seismic monitoring of the national territory to the recording of geophysical parameters for basic scientific research. RSN provides a fundamental support task for the Department of Civil Protection, in adopting measures for the prevention and mitigation of seismic risk, particularly in case of seismic emergencies. RSN currently comprises approximately 500 seismic stations throughout the national territory, including regional networks managed by other institutes or universities. Data are transmitted in real time to the acquisition centre of the National Earthquake Observatory (https://www.ont.ingv.it/) at the INGV headquarters in Rome (Margheriti *et al.*, 2021). Data stored at the main acquisition centre are redundantly stored at other INGV sites, which perform specific services such as backup or analyses of available data quality. Fig. 1 shows the current distribution of the RSN along with the stations of the Mediterranean Very Broadband Seismographic Network [MedNet, or MN according to the FDSN code (https://eida.ingv.it/it/network/MN)], also owned by INGV.

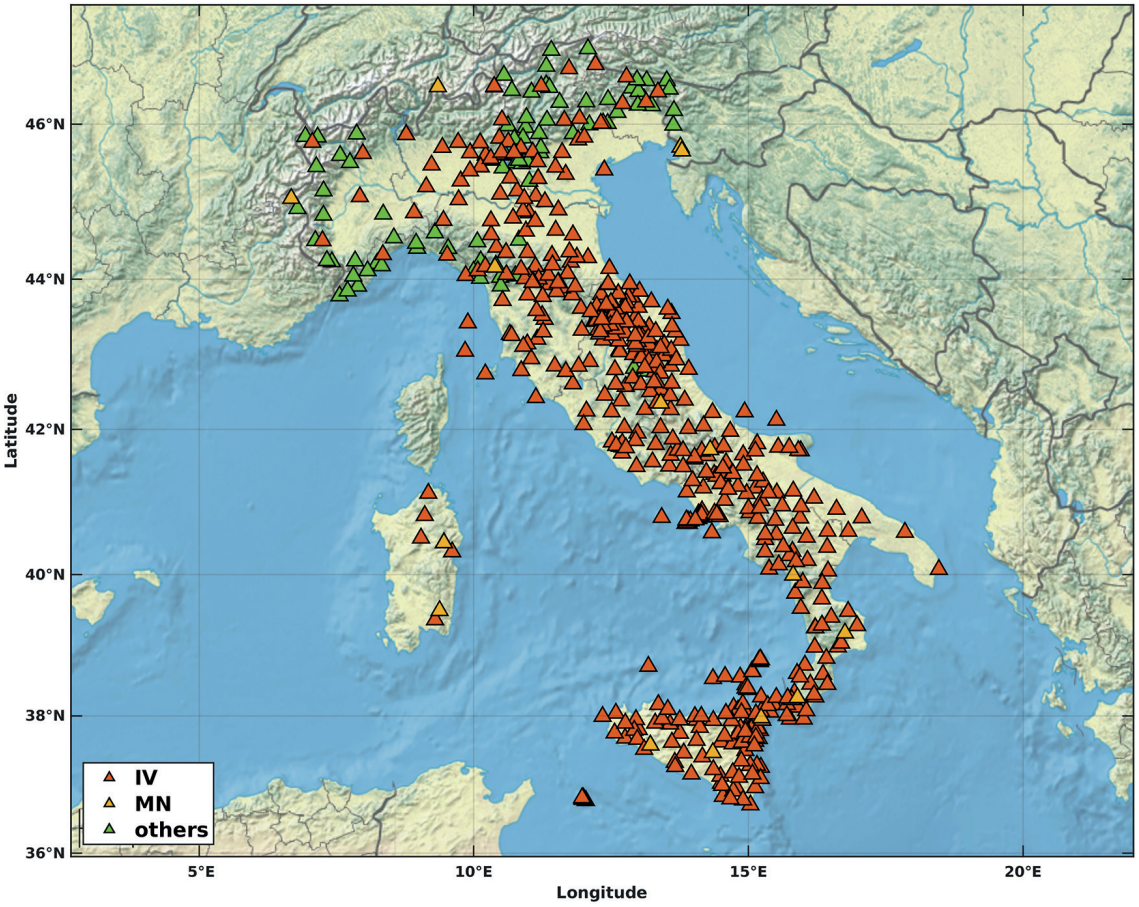

Fig. 1 - Distribution of real-time stations within the Italian territory. The red and orange triangles indicate the seismo-accelerometric IV and MN stations, respectively, following the FDSN (https://www.fdsn.org/) codification. The green triangles indicate the stations managed by other Italian INGV partners for real time data exchange.

The RSN is further contributed by seismic stations managed by other government agencies, institutions or universities, including, for example, the National Accelerometric Network (RAN, or IT as per the FDSN code) managed by the Civil Protection Department (Presidency of the Council of Ministers - Civil Protection Department, 1972), the seismic network in north-eastern Italy (https://eida.ingv.it/it/network/OX) managed by the Centre for Seismological Research of the National Institute of Oceanography and Experimental Geophysics [OGS (https://www.ogs.it/it)], the seismic network in north-western Italy and Lunigiana Garfagnana (https://eida.ingv.it/it/network/GU) managed by the University of Genova (https://distav.unige.it/), the Irpinia network (IsNet, https://eida.ingv.it/it/network/IX), as well as temporary networks installed for scientific experiments or during seismic emergencies by INGV operational groups (https://www.ingv.it/monitoraggio-e-infrastrutture/emergenze/gruppi-di-emergenza).

The distribution of all stations contributing to the RSN across the Italian territory is facilitated through the EIDA Italia portal (https://eida.ingv.it/it/getdata; Danecek *et al*., 2021).

Most of the installation sites of the RSN are equipped with six-channel seismic recorders coupled with both a velocimetric and an accelerometric sensor. Co-locating both sensors at a site enables covering the entire frequency range of the deformation process leading to earthquakes, thus avoiding the loss of recordings sometimes due to velocimeter saturation for nearby events of significant magnitude [usually with local magnitude or Richter (M_L) or moment magnitude (M_W) > 3.5]. At the same time, it ensures the proper recording of local events of low magnitude, useful for characterising micro-seismicity.

3. SDQ_event: software architecture

SDQ_event is a semi-automatic Python code (https://www.python.org/), primarily utilising the Obspy module (Beyreuther *et al*., 2010; Megies *et al*., 2011; Krischer *et al*., 2015), designed to rapidly assess the quality of seismic event recordings, single seismic station recordings, or data sets containing thousands of recordings.

The code can be executed, for example, in a Linux environment, via a terminal using a command line that must include the name of the external input file to be imported. This input file is a .txt text file and must contain at least three columns of data including the event identification code (eventID variable), obtained from the INGV seismic bulletin (http://terremoti.ingv.it/), the station code (stacode variable), and the network code (netcode variable).

Once the external input file is imported, the code is launched. This code has been designed and written to process a large amount of data in the shortest time possible through parallel computing processing through the Python multiprocessing library, by fully exploiting the technical capabilities of available computational machines.

Here follows an example of a command line through which the SDQ code can be executed from the terminal:

test@test:~$ **./SDQ.py -wl file_MILN -cwl 1 2 3** -pp 8

where the bold part is necessary to execute the code, namely the input file name (i.e. file_MILN) to be imported (-wl) and the numerical identifier for the data columns to be used (-cwl), containing, respectively, the eventID, stacode, and netcode information. The other part of the command line indicates the possible non-mandatory conditions without which the code would still be executed correctly. In this case, the non-mandatory condition is given by the -pp

parameter, which indicates, for example, the use of eight processors for parallel computation on the test machine.

Waveform processing in the SDQ_event code occurs in three phases (Fig. 2):

1. data selection and data download;
2. pre-processing for data verification;
3. processing for data quality estimation.

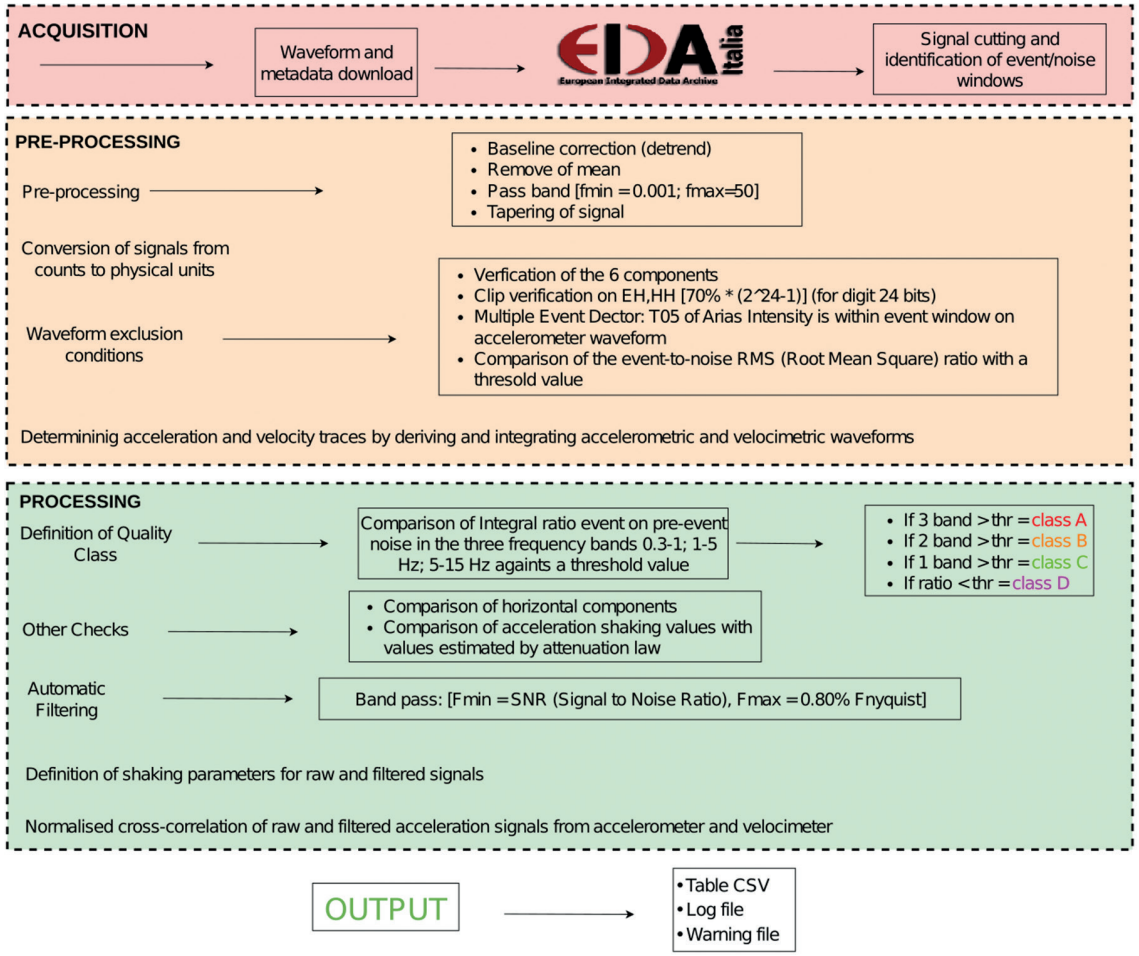

Fig. 2 - Flowchart of the SDQ_event code.

As shown in Fig. 2, the code is organised in such a way that a waveform must pass several conditions before being processed.

During the acquisition phase, after verifying the existence of the event on the INGV fdsnws-event web service (https://webservices.ingv.it/fdsnws/event/1/), the waveform data is downloaded using the EIDA Italia node. Subsequently, the traces are analysed to identify event and pre-event noise windows.

The signal cutting and the subsequent identification of the event and pre-event anthropic noise windows occur using a moving window dependent on the magnitude, hypocentral distance, and P-phase arrival time selected for the target event at each station. If the P-phase arrival time

metadata is not downloadable from the web service, the theoretical P-phase time is estimated using the procedures reported in Puglia *et al*. (2018) and developed for the Engineering Strong Motion (https://esm-db.eu/, Luzi *et al*., 2016) database service.

After signal cutting, the pre-processing phase begins for each event waveform recorded at each available station.

The conversion of downloaded signals into physical units is performed using the station's dataless files, which, moreover, also enable the operator to perform instrumental deconvolution to account for the frequency response of different sensors.

Subsequently, signal pre-filtering is performed, including baseline removal (detrend), mean removal (demean), by applying a band-pass filter (f_{min} = 0.001 Hz, f_{max} = 50 Hz) and applying a cosine taper to the entire downloaded signal trace.

Before waveforms can be processed, they must satisfy certain conditions, including:

1. all velocimetric and accelerometric components (N-S, E-W, vertical) must function: if only one of the six analysed components fails to record the seismic event, the station is discarded;
2. velocimetric signals must be saturated: waveforms are processed if their amplitudes remain below a specified saturation threshold, defined as 90% of the full scale of a 24-bit digitiser following the formulation:

$$FS = 2^{24-1}. \tag{1}$$

   If the raw signal amplitude in counts exceeds this threshold, the waveform is discarded;
3. multiple events must be checked for, as they are especially useful in the case of seismic sequences. If the signals are recorded correctly and no saturated velocimetric signals are identified, the detection of multiple events is determined by considering the T05 [i.e. the 5% of the cumulative function of the Arias Intensity calculated on the event window (Arias, 1970)] value of the cumulative Arias Intensity function. If the T05 value is outside the cutting window of the event signal, there is likely an event preceding the target event (or the background noise has very high values) and, thus, the waveform is discarded. The validity of this condition is confirmed through the triggering algorithm of the short time average over long time average (STA/LTA) function (Withers *et al*. 1998; Trnkoczy, 2009) implemented in Obspy;
4. the signal amplitude variation over time must be evaluated and the *RMS* ratio between event and noise to a predetermined threshold should be compared. For this action, the signal envelope in the time domain is first obtained through the application of the Hilbert transform (Kanasewich, 1981), and subsequently, the *RMS* value is calculated considering both the pre-event noise window and the analysed event window, using:

$$RMS_{(evt,noi)} = \sqrt{\left[\left(x_1^{\,2} + x_2^{\,2} + ... + x_n^{\,2}\right)/n\right]}. \tag{2}$$

Afterwards, once the *RMS* is calculated for both the event window and the pre-event noise window, their ratio is computed. If the resulting value is below a specific threshold set *a priori*, the code excludes the waveform.

The threshold value selection was made empirically (Table 1), based on the visualisation of the signals and the distribution of the values obtained from the ratio considering all events and stations in the available data set.

Once these conditions are met, the waveform cannot be excluded. Acceleration and velocity are calculated by integrating and differentiating the accelerometric and velocimetric waveforms.

After these conversions, the final processing phase begins.

In this phase, a quality class is assigned to each analysed waveform. This assignment is made by calculating the ratio between the integrals of the amplitude Fourier spectra of the event and the pre-event anthropic noise for three frequency bands and comparing the value with an empirically determined threshold for each selected frequency band (Table 1).

Through a graphical inspection of hundreds of target earthquakes, based on comparisons between the fast Fourier transforms (FFTs) of events and ambient noise typical of stations located in different geological and anthropic contexts, three representative frequency bands of typical seismic phenomena in the Italian peninsula were chosen: 0.3-1 Hz, 1-5 Hz, and 5-15 Hz. The quality classes used to represent the data quality range from A (excellent) to D (to be discarded).

Specifically, a waveform belongs to class A, B, or C when the values of the ratio between the integrals of the FFT amplitudes of the event and pre-event noise are higher than the empirically fixed minimum threshold values for all frequency bands (A, excellent quality), for only two out of the three intervals (B, good quality), or for only one frequency band (C, low quality).

Finally, a waveform is classified in class D (poor quality data) when in no case do the integral ratio values exceed the threshold values. Waveforms in class D are not discarded *a priori*; however, the information is entered into an external log file containing warnings indicating that the data is not reliable for statistical analysis.

Once the quality class is assigned to the waveform, additional checks are performed. These are not exclusive, suggesting to review the data in an external file if the outcome of the check performed is not positive.

The final checks include:

1. verification of the peak ground motion values of the two horizontal components. If the ratio of the peak ground motion values calculated on the two horizontal components exceeds a specific threshold (Table 1), a descriptive warning will be present in the log file;
2. peak ground motion value comparison between the actual data and the values estimated from a reference ground motion prediction equation (GMPE). For this study, the ITA10 relationship was used (Bindi *et al*., 2011). If the ground motion of the actual data is not within the range of +/- 3 standard deviations of the reference attenuation relationship, a descriptive warning will be reported in the log file.

Table 1 - Empirical threshold values for exclusion conditions, quality classes and additional controls. Error and warning messages are reported when conditions are not met.

Condition	Threshold values	Message
RMS ratio	10	Error-waveform not processed
Ratio between the integrals of the FFT amplitudes of the event and pre-event noise for frequency bands f1 (0.3-1) Hz, f2 (1-5) Hz, and f3 (5-15) Hz	f1 (0.3-1 Hz) = 5 f2 (1-5 Hz) = 10 f3 (5-15 Hz) = 7	Signal quality class identification: Class A Class B Class C Class D (Warning)
Verification of the peak ground motion values of the two horizontal components	5	Warning
Peak ground motion value comparison between the actual data and the values estimated from a reference GMPE (ITA10)	+/- 3 σ	Warning

Following these final conditions, the frequency range for automatic waveform filtering is defined. The automatic filter used is a third-order Butterworth filter, where the minimum frequency is chosen based on the signal-to-noise Ratio (SNR), while the maximum frequency is defined as 80% of the Nyquist frequency (Bormann, 2012).

After various tests applying variable SNR thresholds between 3 and 10, a conservative approach was chosen to use an SNR value of 5. The search for the minimum filtering frequency is based on the ratio of the event and background noise acceleration FFTs: the minimum cutoff frequency corresponds to the first value for which the FFT ratio exceeds the pre-set threshold value. However, in cases where the low frequency is characterised by significant amplitude values in the background noise, this approach can be ineffective, often resulting in unrealistic filter values. To overcome this, minimum and maximum frequency thresholds are set based on which the filter value must fall within a predefined range to avoid gross filtering errors.

These limitations are based on the following conditions:

1. if the minimum frequency value is ≥0.4 Hz, the frequency value is set to 0.4 Hz;
2. if for events with magnitude <4.5 the minimum frequency value is <0.2 Hz, a frequency value of 0.2 Hz is set;
3. if for events between magnitude 4.5 and 5.5 the minimum frequency is <0.1 Hz, a minimum frequency value of 0.1 Hz is set;
4. if for events with magnitude ≥5.5 the minimum frequency is <0.05 Hz, a minimum frequency value of 0.05 Hz is set.

Finally, for each raw and filtered signal, the following are calculated:

1) peak ground acceleration (*PGA*) and peak ground velocity (*PGV*) parameters, both directly recorded by the sensors and obtained through the integration of acceleration to velocity or velocity to acceleration conversions;
2) normalised cross-correlation coefficients (CCs) to assess the similarity degree of the entire waveform, considering both acceleration signals (raw and filtered) and the velocity signal derived from the velocimetric signal;
3) ratios between *PGA* and *PGV* shaking values directly from the accelerometer and velocimeter and their analytically obtained values through derivative and integral.

The product between the CCs and the values of the ratios obtained in point 3 is useful for a more robust estimation of waveform coherence.

At the end of the processing operations, the code outputs provide:

1) a summary table in CSV format containing all the parameters of interest for each processed waveform;
2) a text format log file indicating whether the waveform has been processed or discarded due to a specific condition;
3) a text format log file containing warnings for processed waveforms that may still present issues identified during supplementary checks to the processing phase.

Lastly, to facilitate the visualisation of the results, comparative graphs are optionally generated with the shaking values calculated and inferred from both instruments.

In Table 2, an example table is shown where the terms listed in each column of the summary CSV table obtained at the end of processing are specified, with each record indicating the results for event/station/component.

Table 2 - Parameters calculated for each processed waveform that will form the final output CSV table.

ID	Event ID
Date time	Earthquake origin time expressed in UTC date and time
Netcode	Network code
Stacode	Station code
Stream_acc.	Accelerometer band and instrument code
Sensitivity_acc.	Accelerometer sensitivity
Stream_vel.	Velocimeter band and instrument code
Channel	Channel code
PGA_AA	*PGA* value from accelerometer [m/s^2]
PGA_AA_F	*PGA* value from filtered accelerometer [m/s^2]
PGA_AV	*PGA* value from velocimeter [m/s^2]
PGA_AV_F	*PGA* value from filtered velocimeter[m/s^2]
PGV_VV	*PGV* value from velocimeter [m/s]
PGV_VV_F	*PGV* value from filtered velocimeter [m/s]
PGV_VA	*PGV* value from accelerometer [m/s]
PGV_VA_F	*PGV* value from filtered accelerometer [m/s]
RPGA_AA/PGA_AV	*PGA* ratio of accelerometer to velocimeter
RPGA_AAF/PGA_AVF	*PGA* ratio of accelerometer to velocimeter (filtered)
RPGV_VV/PGV_VA	*PGV* ratio of velocimeter to accelerometer
RPGV_VVF/PGV_VAF	*PGV* ratio of velocimeter to accelerometer (filtered)
CC	Cross-correlation coefficient
CC_F	Cross-correlation coefficient of filtered signals
CC/RPGA	Ratio of CC to *RPGA*
CC_F/RPGA_F	Ratio of CC (filtered) to *RPGA* (filtered)
CC/RPGV	Ratio of CC to *RPGV*
CC_F/RPGV_F	Ratio of CC (filtered) to *RPGV* (filtered)
Repi	Epicentral distance [km]
Ripo	Hypocentral distance [km]
Mag	Magnitude
S/N_RMS	Signal to noise *RMS*
RINT_0.3_1	Integral ratio of event to noise of the FFT between 0.3-1 Hz
RINT_1_5	Integral ratio of event to noise of the FFT between 1-5 Hz
RINT_5_15	Integral ratio of event to noise of the FFT between 5-15 Hz
Qletter	Signal quality index (A-D)
Fmin	Minimum pass-band filter frequency [Hz]
Fmax	Maximum pass-band filter frequency [Hz]

4. SDQ_event: data set and processing

SDQ was tested considering more than 200 stations belonging to IV and MN networks equipped with co-located accelerometric and velocimetric sensors (Fig. 1). Overall, we processed approximately 15,000 waveforms characterised by magnitude (M_L and/or M_W) greater than 3.5 in the January 2012 to June 2023 period and hypocentral distance below 150 km. Given this

broad observation period, we were able not only to assess the current operation of each station, but also to identify periods of temporary malfunction caused by various problems. These can range from problems at the installation site to equipment failures or errors in the metadata due to incorrect settings. Fig. 3 shows several representations of the data set of this study, offering a visual overview of the data collected over time.

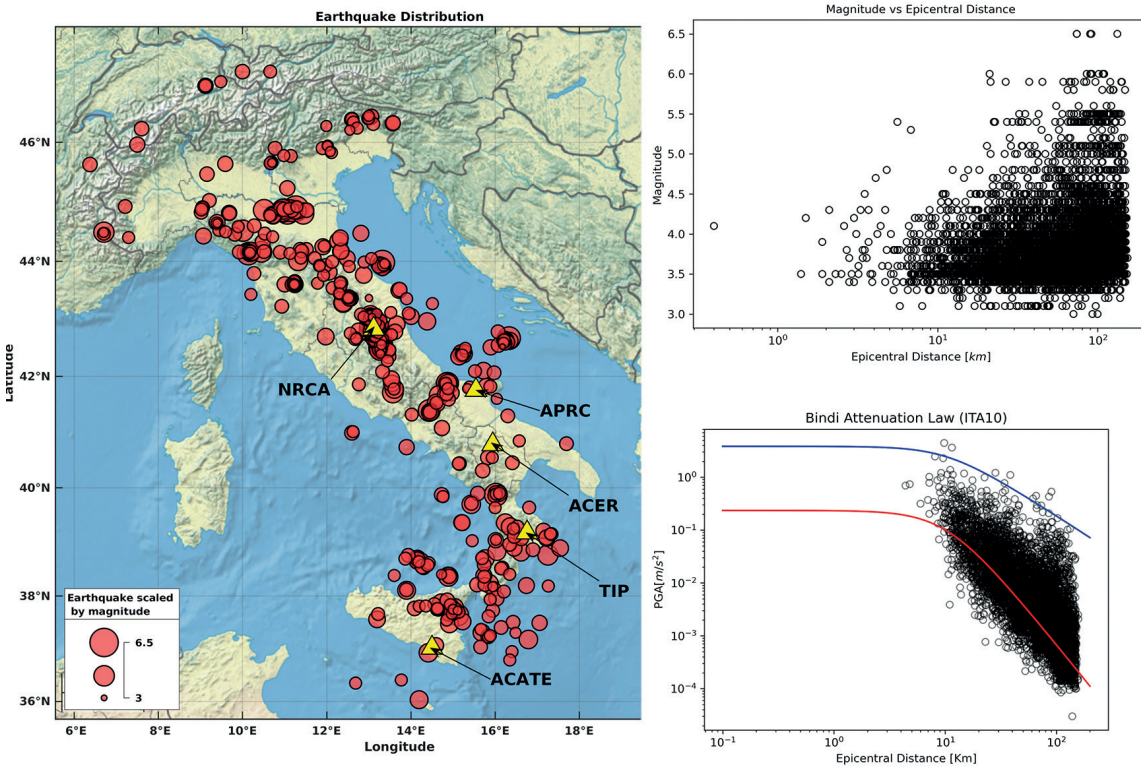

Fig. 3 - Data set used for the SDQ_event analysis. In the left panel, the red circles represent the earthquakes considered for the analysis, with different sizes indicating their magnitudes. Various yellow triangles represent example stations used to illustrate the results. In the right panel, the data set is presented in terms of magnitude and epicentral distance (top), and in terms of *PGA* and epicentral distance (bottom). The red and blue lines denote the medians of the GMPE calibrated by Bindi *et al*. (2011) for magnitudes 3.5 and 6.5, respectively.

As shown in Fig. 3, the left panel illustrates earthquakes across the national territory, with point sizes (correlated to magnitudes) and stations represented by triangles. These stations are employed to demonstrate the phases of the algorithm and provide examples of results regarding the functioning of stations.

The upper-right panel indicates that events with significant magnitudes (>5) are predominantly recorded by stations situated more than 20 km away.

In the bottom-left panel, the *PGA* value range available is compared to estimates derived from the ITA10 attenuation relationship (Bindi *et al*., 2011), representing the lower (M_L and/or M_W = 3.5, red) and upper (M_w = 6.5, blue) magnitude limits of our data set.

Next, we provide a step-by-step description of the processing of an individual waveform by SDQ. For this example, we consider recordings from the TIP station (MN network) for an event with a magnitude M_L of 4.0 and hypocentral distance of 41.5 km, which occurred on 7

October 2019 at 06:11:32 UTC (INGV ID https://terremoti.ingv.it/event/23231121). This event is compared with an excluded waveform from a different seismic event (INGV ID https://terremoti.ingv.it/event/22632151), which occurred on 8 July 2019 at 04:38:28 UTC, with a magnitude M_w of 3.8 and a hypocentral distance of 79.4 km, recorded at the ACATE station.

The first step involved checking the existence of the event on the INGV FDSN web service, followed by the downloading of the corresponding waveforms through the EIDA Italia node. Subsequently, the code selected the analysis trace and identified the pre-event noise windows, delineated by blue dashed lines, and the event window, delineated by grey dashed lines (Fig. 4), using the previously described cutting algorithm.

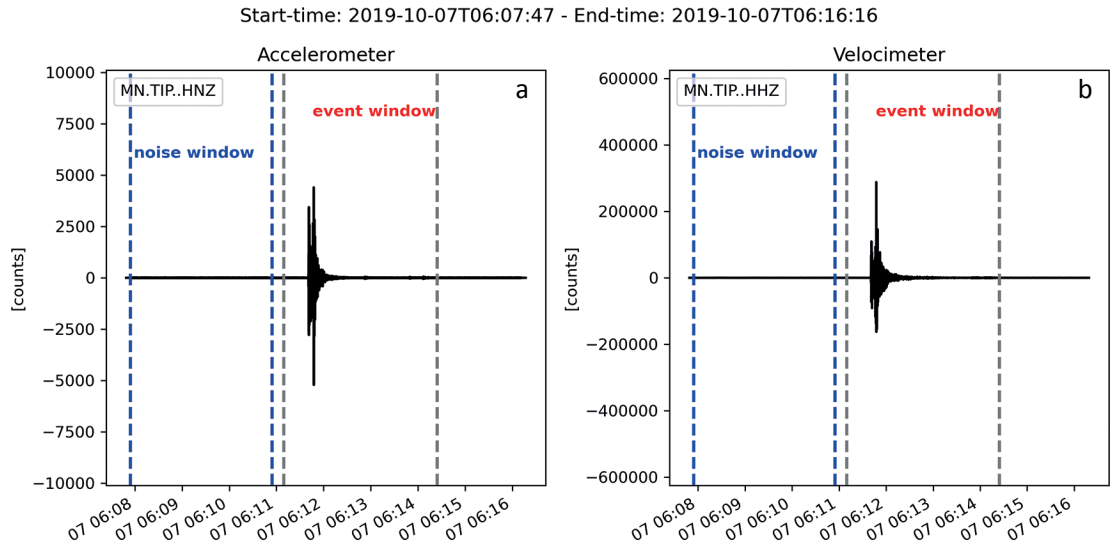

Fig. 4 - Analysis for the event at the target station. The blue and grey dashed lines represent the pre-event and event windows, respectively. Panel a shows recordings from the vertical accelerometer component of the TIP station (Kinemetrics Episensor-FBA), while panel b displays recordings from the vertical component of the broadband velocimeter (STS2-120s).

Pre-processing operations were conducted due to the presence of all six motion components and the absence of clipping issues in the velocimetric traces. This involved removing the signal mean and baseline, by applying a Butterworth band-pass filter (0.001 Hz to 50 Hz), and converting signals from digital units to physical units.

Subsequently, we excluded the possibility of multiple events within the downloaded trace. The exclusion condition is based on the normalised Arias function (Arias, 1970) that represents the energy developed by the waveform over time. We considered the 5% threshold of the normalised Arias function, which must lie within the event window. Additionally, on the same trace, we calculated the STA/LTA function to observe the signal amplitude behaviour, especially in the pre-event window. Error messages were generated when this condition was not met.

An example is illustrated in Fig. 5, showing the Z component of the accelerometer trace for an event recorded at the TIP station, the STA/LTA function and the related normalised Arias intensity.

In Fig. 5, at station TIP, the 5% Arias intensity cumulative time falls within the event window, indicating no significant events preceding the analysed event and lower pre-event noise amplitude compared to the seismic transient.

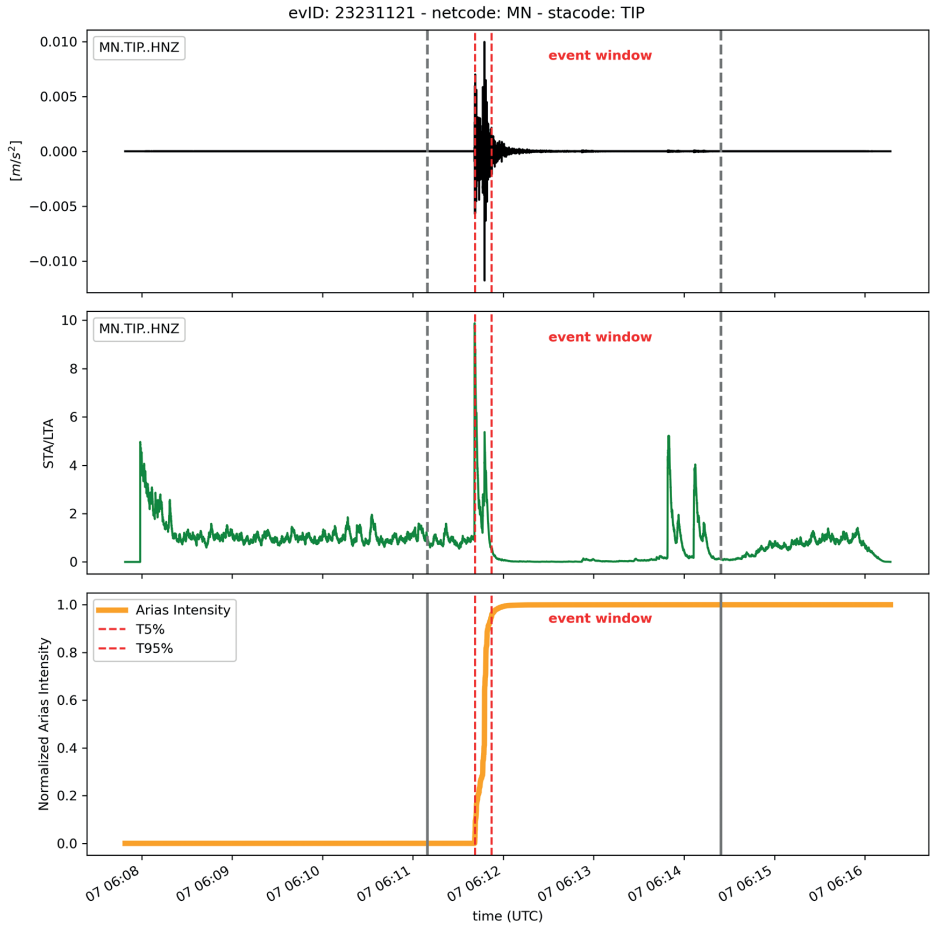

Fig. 5 - Event of 7 October 2019 at 06:11:32 UTC (ID-INGV 23231121) recorded at the TIP accelerometer station, in the top panel. The grey dashed lines represent the event windows, while the red dashed lines represent 5% and 95% Arias intensity. The middle panel displays the STA/LTA ratio function (in green) and the bottom panel shows the cumulative Arias intensity (in orange).

For comparison in Fig. 6, we represented the same condition for the event on 8 July 2019 at 04:38:28 UTC (INGV ID 22632151) recorded at station ACATE. In this example, we observed that the 5% Arias intensity cumulative time precedes the start time of the target event window, suggesting either a very noisy event or multiple events within the same analysis window.

The presence or absence of multiple signals can be observed from the shape of the cumulative Arias intensity and the amplitude of the STA/LTA function in the pre-event noise window.

The last exclusion condition is based on *RMS* analysis. Specifically, it evaluates the *RMS* ratio between the event window and the pre-event window. Recordings with an *RMS* below a certain threshold are excluded from further processing due to excessive noise. The threshold was empirically determined through initial testing and fixed at 10 for subsequent runs.

Once the exclusion conditions are verified, the seismic waveform is classified into a quality class and processed according to the scheme previously described.

After running the code, we consider the output CSV table, where the shaking parameters can be compared in a summary image to verify the functioning of stations during the recording of the earthquake.

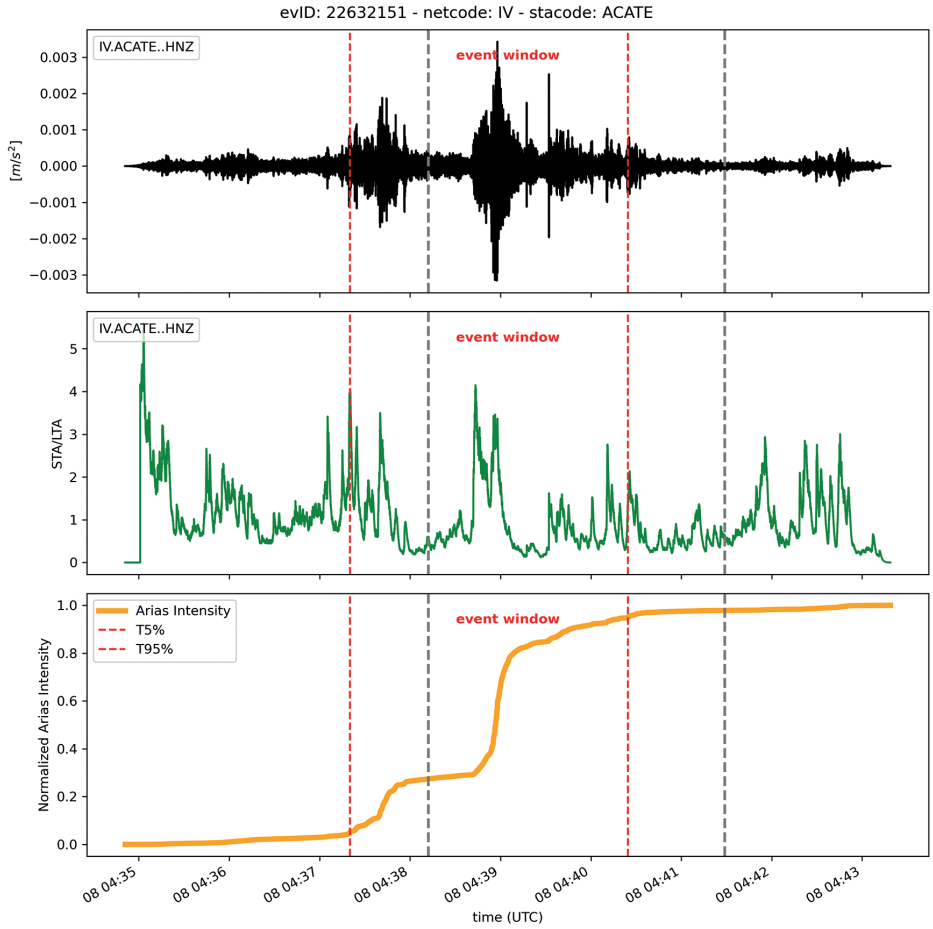

Fig. 6 - Event of 8 July 2019 at 04:38:28 UTC (INGV ID 22632151) recorded at ACATE accelerometer station (top). The grey dashed lines represent the event windows, while the red dashed lines represent 5% and 95% Arias intensity. The middle panel displays the STA/LTA ratio function (in green) and the bottom panel shows the cumulative Arias intensity (in orange).

5. SDQ_event code: results

After conducting the data processing for each waveform of the data sets, summary images are produced to observe the functioning of the stations.

Specifically, the images include:

1) the comparison between the acceleration shaking value (raw and filtered) recorded by the accelerometer and the acceleration shaking value obtained from the co-located velocimeter (raw and filtered);
2) the ratio of the acceleration shaking value (*RPGA*) between recorded acceleration and the acceleration obtained from the velocimeter (raw and filtered) as a function of both acceleration shaking value and time.

Some examples of summary images for stations ACER (Acerenza), NRCA (Norcia) and APRC (Apricena) are shown in Fig. 7.

In Fig. 7, the grey, green, and red symbols represent the results for the vertical, N-S, and E-W motion components, respectively, for each earthquake recorded at the analysed station.

Specifically, the top panels compare the recorded (filtered) *PGAs* with those inferred from the velocimeter, while the bottom panels depict the ratios between real and velocimeter-inferred *PGAs* as a function of the recorded *PGA* value.

Proper functioning of the station, characterised by full consistency between recorded and velocimeter-derived acceleration, is indicated when the symbols in the top panels align along the diagonal line or when, in the bottom panels, they align along the horizontal line, indicating a ratio of one.

In Fig. 7, three different scenarios are depicted:

1) proper functioning is observed for the ACER station, showing full consistency between recorded accelerometer values and those inferred from the velocimeter. All shaking values for each component align along the diagonal (Fig. 7, top-left panel) and correspond to a horizontal line indicating a ratio of one (Fig. 7, bottom-left panel);
2) a temporal variation in behaviour is evident for the NRCA station, where part of the data set shows proper functioning, while values deviate from the diagonal (and the ratio of one) with increasing shaking values. This trend highlighted saturation issues in velocimeter sensors during the recording of events with magnitudes >4 and hypocentral distances <20 km during the Norcia-Visso-Amatrice 2016 sequence (Fig. 7, central panels). This trend revealed that relying solely on 90% of the available dynamic range of a 24-bit digitiser may not always recognise clipped data and ensure its exclusion;
3) a probable issue with the N-S component is indicated for the APRC station, where values deviate from references throughout the installation period (see green triangles in the right panels of Fig. 7).

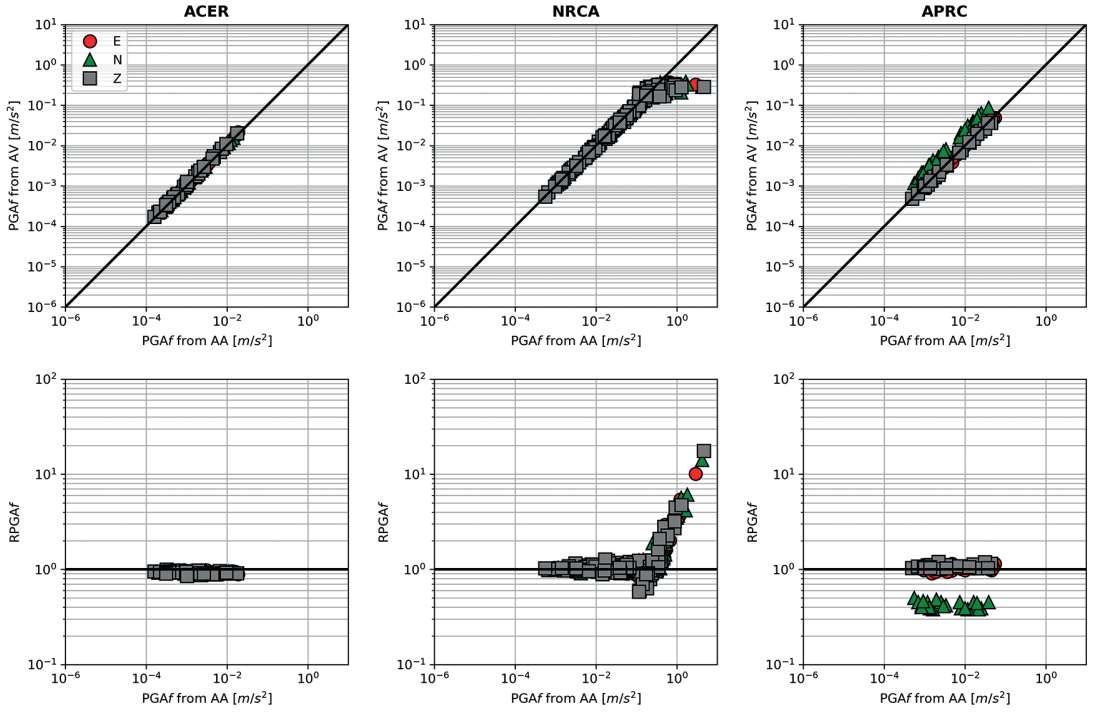

Fig. 7 - Image obtained from the SDQ-event code for the ACER, NRCA and APRC stations, comparing filtered *PGA* values recorded from the accelerometer (AA) and from the velocimeter (AV). The vertical, N-S, and E-W components are represented with grey, green, and red symbols, respectively.

6. SDQ_stream code: software architecture

SDQ_ stream, like the SDQ_event, is a semi-automatic Python code. It determines the quality class of seismic continuous recordings by assessing quality metrics and comparing them to empirical thresholds.

This code can also be executed via a terminal using a command line that must include the following parameters:

1) the name of the external input text file (.txt) to be imported, which must contain the network code and the station code;
2) the start date, indicating when the analysis is performed;
3) the time interval for analysis, which can be specified in hours or days;
4) the number of time intervals, quantifying the specific duration for analysis.

An example of a command line through which the SDQ_stream code can be executed from the terminal is:

test@test:~$ **./SDQ_stream.py -wl FILE_INPUT –cwl 0 1 –ymd 2023-11-17 -i day -nu 1**

In this command line, the parameters -wl, -cwl, -ymd, -i, and -nu are used to specify the input file name, identify the columns containing the network code and station code, indicate the start date for the daily analysis, specify the analysis interval (in this case, daily), and set the number of intervals (in this case, one).

The waveform processing in the SDQ_stream code occurs in two phases:

1. data selection and data download;
2. processing: evaluation of daily quality metrics and definition of quality class.

In Fig. 8, the flowchart of the SDQ_stream code is displayed.

As observed in Fig. 8, both the waveform and metadata can be downloaded from either the EIDA Italia node or from a local archive like SeisComP.

In the processing phase, records are not discarded as SDQ events and quality parameters are evaluated.

The following are calculated:

- data gap or data availability (%);
- number of gaps;
- duration of the sum of the gaps (s);
- duration of the maximum temporal gap (s);
- mean *RMS*, without filter application (counts);
- mean *RMS*, with Butterworth band-pass filter between 0.01 and 50 Hz (counts);
- mean PSD, with Butterworth band-pass filter between 0.01 and 50 Hz (dB).

Finally, the quality class is determined by comparing the calculated quality metrics with empirical thresholds, similar to those used in ISMDq (http://ismd.mi.ingv.it/images/documentazionev1_eng.pdf).

The code produces in output:

1) a log file containing the processing time for quality analysis;
2) daily images of quality metrics;
3) a summary table in CSV format containing all relevant quality metrics for continuous ambient noise data streams.

During the testing phase, the analysis is performed considering a daytime interval. An illustration of the daily images generated by the code is shown in Fig. 9. These images depict the

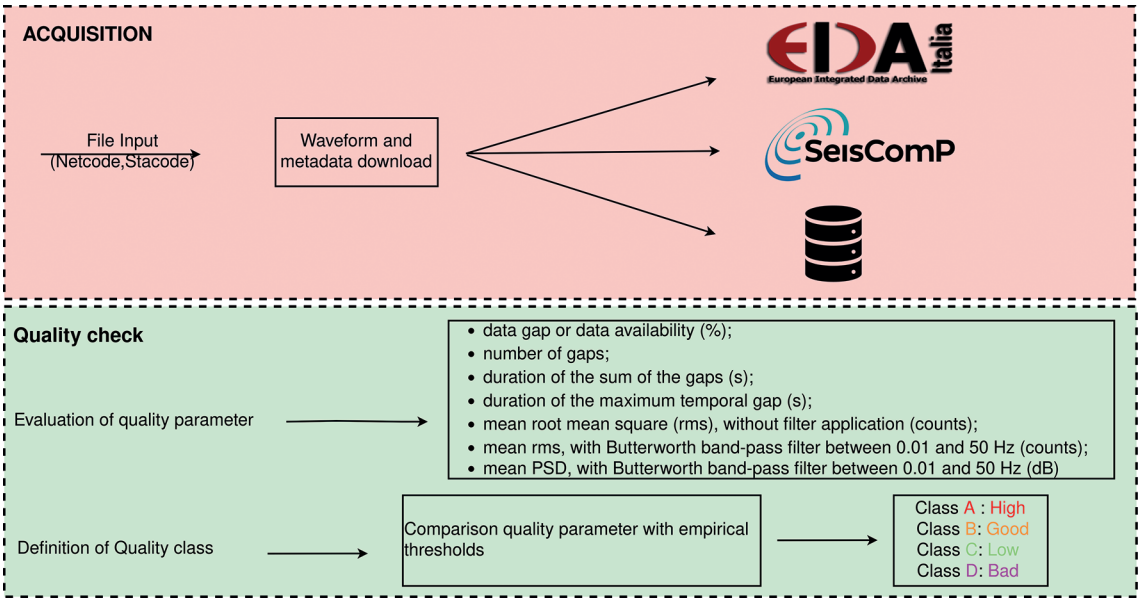

Fig. 8 - Flowchart of the SDQ_stream code.

PDFs computed for the PLAC and MILN stations recorded on 15 August 2023 and 21 July 2023, respectively.

Regarding the CSV table, the SDQ_stream code is designed to continuously update the table for the analysed stations and daily processing. This makes it possible to construct time series for each metric, similar to those calculated on ISMDq, enabling their behaviour to be studied. An example of such time series, calculated by ISMDq, is shown in Figs. 10 and 11, with a PSD time series calculated in a 10-20 Hz frequency band for the PLAC and MILN stations.

In Figs. 10 and 11, we show the PSD calculated in the 10-20 Hz band for each sensor component (vertical, E-W, and N-S) and represented with different colours (black, green, and blue). When anomalous behaviours are observed, as indicated by the red box in the PLAC case, it is likely that the station has some issues. In such cases, one can retrieve daily images (e.g. Fig. 9 left

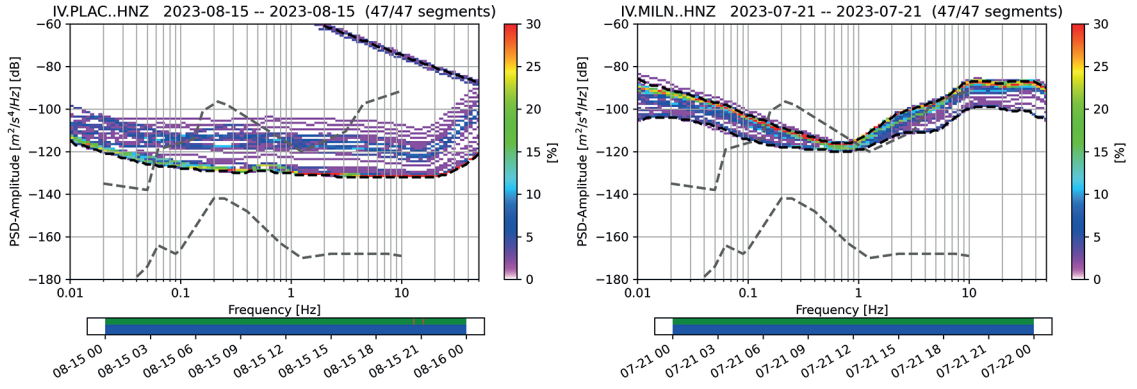

Fig. 9 - Probability density function of the PSD computed for recording on 15 August 2023, at the PLAC station (left panel); probability density function of the PSD computed for recording on 21 July 2023, at the MILN station (right panel).

panel) to verify if there are any problems. By contrast, the MILN case in Fig. 11 shows a regular pattern, and as observed in the daily images (e.g. Fig. 9 right panel), the PDFs do not indicate any problems.

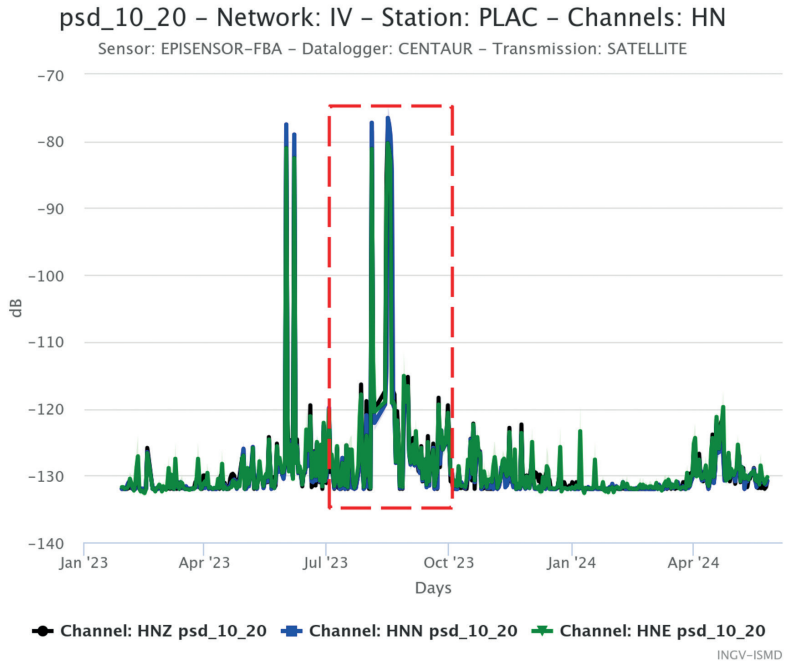

Fig. 10 - Time series showing the PSD calculated by ISMDq in the 10-20 Hz frequency band from the accelerometer at the PLAC station. Black, blue, and green colours correspond to the PSD calculated for the vertical, N-S, and E-W components, respectively. The red box highlights anomalous behaviour observed in the series.

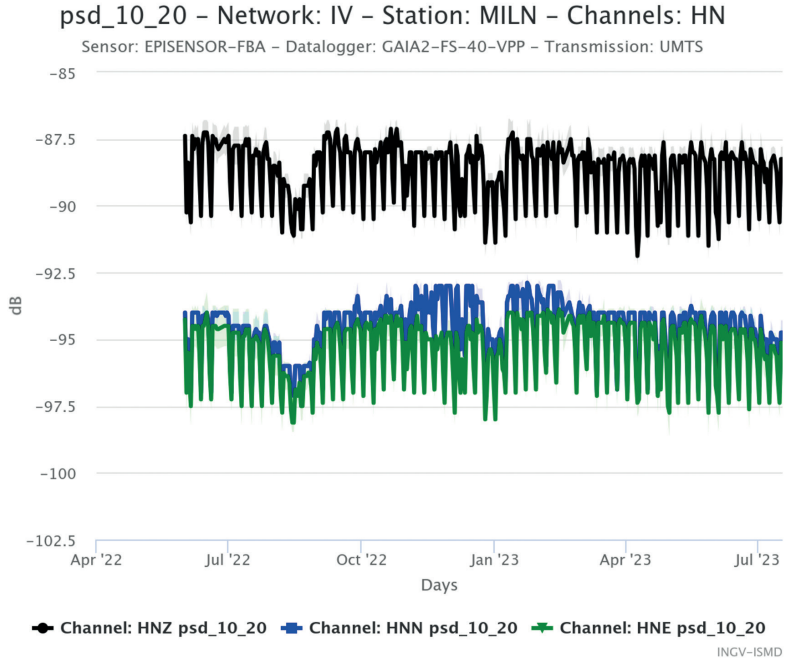

Fig. 11 - Time series showing the PSD calculated by ISMDq in the 10-20 Hz frequency band from the accelerometer at the MILN station. Black, blue, and green colours correspond to the PSD calculated for the vertical, N-S, and E-W components, respectively.

7. Malfunctioning of seismic stations

During the testing phase of the SDQ code, issues were identified with some stations from the RSN and MN networks. For earthquake data, SDQ_event highlighted the importance of co-located sensors. In particular, it is crucial to compare ground motion parameters using the CSV tables described in sections 2 and 4 and to study these values over time to assess the proper functioning of the stations.

Two types of issues were observed during testing: persistent and temporary problems. Persistent problems were evident from the initial installation, where acceleration and velocity signals lacked coherence. This issue was often related to incorrect station configurations, resulting in discrepancies between the accelerometric and velocimetric data. Temporary problems were associated with station malfunctions that were resolved after technician interventions. These issues typically involved one motion component or incorrect sensor configurations.

Anomalous station behaviour was also observed when comparing clipped velocimetric data. This occurred particularly during earthquakes with magnitudes >4.0 and hypocentral distances <20 km. Such comparisons can lead to incorrect conclusions, making it essential to select an appropriate clipping threshold and exclude affected traces.

For noise analysis, the study of temporal patterns of quality metrics is crucial for diagnosing station functionality. For example, percentage gap analysis can indicate issues related to data transmission, instrument temperature, or internal station problems, while the study of PDF or PSD can indicate issues related to internal station problems, or temporary noise sources. Understanding and analysing these temporal patterns of noise data quality metrics is crucial for assessing whether a station is functioning properly before an earthquake occurs. This knowledge can guide the selection of reference stations, thereby reducing epistemic uncertainty in subsequent research analyses.

8. Conclusions

In this study, we proposed the SDQ project, which enables assessing the functioning of seismic stations by analysing seismic events (with the SDQ_event code) or ambient noise (with the SDQ_stream code) and classifying seismic recordings into quality classes. These codes were tested using data from RSN and MN networks, with the goal of providing additional support to seismic monitoring network operators at the seismic surveillance room in the Rome headquarters.

For the SDQ_event code, recordings must pass specific exclusion conditions before being classified into a quality class: presence of the event on the web service, availability of both accelerometric and velocimetric co-located data, signal saturation, multi-event detection, and *RMS* condition.

In the end, the SDQ_event code provides a summary table that collects all relevant parameters for each processed waveform, along with an explanatory log file to aid users in evaluating the results. Additionally, the code can generate various additional images: plots related to individual waveforms to examine the step-by-step analysis of waveform data and plots for the analysed stations to monitor their correct operation and state of health.

During the testing phase, the additional summary images for stations, comparing ground motion parameters calculated and observed at two sensors, revealed that some stations showed no problems as their values aligned with reference lines. However, issues such as velocimeter saturation during high-magnitude events at short hypocentral distances (e.g. the Norcia case), due

to incorrect threshold clipping, and issues with one or more motion components were identified.

For the SDQ_stream code, ambient noise recordings do not need to pass any conditions. The code can perform analysis over short time periods (hours) or over long time periods (days, months, and years).

In the testing phase, we focused on daytime intervals. Daily quality metrics were calculated and compared to empirical thresholds to establish a quality class. Like the SDQ_event code, the SDQ_stream code provides a summary CSV table and daily images for each quality metric. The continuous update mechanism for the CSV table and the ability to generate time series for each metric enable long-term monitoring and in-depth analysis of seismic station performance, as demonstrated in the example cases of stations PLAC and MILN.

Overall, the SDQ project is still under development, with the aim of further reducing machine processing time to accommodate future data sets characterised by increasing amounts of data. The codes are designed to run continuously, updating the input files with data from new earthquakes or newly installed stations, enabling a dynamic monitoring of the health status of the seismic network.

Acknowledgments. The production of data from the National Seismic Network involves many colleagues at INGV who perform various tasks in the data production chain, from ensuring the proper functioning of the stations, to data acquisition, processing, archiving, and subsequent distribution. We thank all INGV colleagues who, on a daily basis, contribute to the management of the National Seismic Network and the mobile seismic networks of INGV.

REFERENCES

Ameri G., Massa M., Bindi D., D'Alema E., Gorini A., Luzi L., Marzorati S., Pacor F., Paolucci R., Puglia R. and Smerzini C.; 2009: *The 6 april 2009, Mw 6.3, L'Aquila (central Italy) earthquake: strong-motion observations.* Seismol. Res. Lett., 80, 951-966, doi: 10.1785/gssrl.80.6.951.

Arias A.; 1970: *A measure of earthquake intensity*. In: Hansen R.J. (ed), Seismic design for nuclear power plants, Mit Press, Cambridge, MA, USA, pp. 438-483.

Beyreuther M., Barsch R., Krischer L., Megies T., Behr Y. and Wassermann J.; 2010: *ObsPy: a Python toolbox for seismology.* Seismol. Res. Lett., 81, 530-533, doi: 10.1785/gssrl.81.3.530.

Bindi D., Pacor F., Luzi L., Puglia R., Massa M., Ameri G. and Paolucci R.; 2011: *Ground motion prediction equations derived from the Italian strong motion database.* Bull. Earthquake Eng., 9, 1899-1920, doi: 10.1007/s10518-011-9313-z.

Bormann P.; 2012: *New manual of seismological observatory practice (NMSOP-2).* Borman (ed), IASPEI, GFZ German Research Centre for Geosciences, Postdam, Germany, doi: 10.2312/GFZ.NMSOP-2.

Danecek P., Pintore S., Mazza S., Mandiello A., Fares M., Carluccio I., Della Bina E., Franceschini D., Moretti M., Lauciani V., Quintiliani M. and Michelini A.; 2021: *The Italian node of the European integrated data archive*. Seismol. Soc. Am., 92, 1726-1737, doi: 10.1785/0220200409.

Kanasewich E.R.; 1981: *Time sequence analysis in geophysics, 3rd ed*. University of Alberta Press, Edmonton, AB, Canada, 480 pp.

Krischer L., Megies T., Barsch R., Beyreuther M., Lecocq T., Caudron C. and Wassermann J.; 2015: *ObsPy: a bridge for seismology into the scientific Python ecosystem.* Comput. Sci. Discovery, 8, 014003, doi: 10.1088/1749-4699/8/1/014003.

Luzi L., Pacor F., Ameri G., Puglia R., Burrato P., Massa M., Augliera P., Franceschina G., Lovati S. and Castro R.; 2013: *Overview on the strong-motion data recorded during the May-June 2012 Emilia seismic sequence.* Seismol. Res. Lett., 84, 629-644, doi: 10.1785/0220120154.

Luzi L., Puglia R., Russo E., D'Amico M., Felicetta C., Pacor F., Lanzano G., Çeken U., Clinton J., Costa G., Duni L., Farzanegan E., Gueguen P., Ionescu C., Kalogeras I., Özener H., Pesaresi D., Sleeman R., Strollo A. and Zare M.; 2016: *The engineering strong-motion database: a platform to access Pan-European accelerometric data*. Seismol. Res. Lett., 87, 987-997, doi: 10.1785/0220150278.

Margheriti L., Nostro C., Cocina O., Castellano M., Moretti M., Lauciani V., Quintiliani M., Bono A., Mele F. M., Pintore S., Montalto P., Peluso R., Scarpato G., Rao S., Alparone S., Di Prima S., Orazi M., Piersanti A.,

Cecere G., Cattaneo M., Vicari A., Sepe V., Bignami C., Valoroso L., Aliotta M., Azzarone A., Baccheschi P., Benincasa A., Bernardi F., Carluccio I., Casarotti E., Cassisi C., Castello B., Cirilli F., D'Agostino M., D'Ambrosio C., Danecek P., Cesare W.D., Bina E.D., Di Filippo A., Di Stefano R., Faenza L., Falco L., Fares M., Ficeli P., Latorre D., Lorenzino M.C., Mandiello A., Marchetti A., Mazza S., Michelini A., Nardi A., Pastori M., Pignone M., Prestifilippo M., Ricciolino P., Sensale G., Scognamiglio L., Selvaggi G., Torrisi O., Zanolin F., Amato A., Bianco F., Branca S., Privitera E. and Stramondo S.; 2021: *Seismic surveillance and earthquake monitoring in Italy.* Seismol. Res. Lett., 92, 1659-1671, doi: 10.1785/0220200380.

Massa M., Scafidi D., Mascandola C. and Lorenzetti A.; 2022: *Introducing ISMDq-A web portal for real-time quality monitoring of Italian strong-motion data.* Seismol. Res. Lett., 93, 241-256, doi: 10.1785/0220210178.

McNamara D.E. and Buland R.P.; 2004: *Ambient noise levels in the continental United States.* Bull. Seismol. Soc. Am., 94, 1517-1527, doi: 10.1785/012003001.

Megies T., Beyreuther M., Barsch R., Krischer L. and Wassermann J.; 2011: *ObsPy-What can it do for data centers and observatories?.* Ann. Geophys., 54, 47-58, doi: 10.4401/ag-4838.

Michelini A., Margheriti L., Cattaneo M., Cecere G., D'Anna G., Delladio A., Moretti M., Pintore S., Amato A., Basili A., Bono A., Casale P., Danecek P., Demartin M., Faenza L., Lauciani V., Mandiello A.G., Marchetti A., Marcocci C., Mazza S., Mele F.M., Nardi A., Nostro C., Pignone M., Quintiliani M., Rao S., Scognamiglio L. and Selvaggi G.; 2016: *The Italian National Seismic Network and the earthquake and tsunami monitoring and surveillance systems*. Adv. Geosci., 43, 31-38, doi: 10.5194/adgeo-43-31-2016.

Morasca P., Walter W.R., Mayeda K. and Massa M.; 2019: *Evaluation of earthquake stress parameters and its scaling during the 2016-2017 Amatrice-Norcia-Visso sequence-Part I*. Geophys. J. Int., 218, 446-455, doi: 10.1093/gji/ggz165.

Presidency of Counsil of Ministers - Civil Protection Department; 1972: *Italian Strong Motion Network*. International Federation of Digital Seismograph Networks, doi: 10.7914/SN/IT.

Puglia R., Russo E., Luzi L., D'Amico M., Felicetta C., Pacor F. and Lanzano G.; 2018: *Strong-motion processing service: a tool to access and analyse earthquakes strong-motion waveforms.* Bull. Earthquake Eng., 16, 2641-2651, doi: 10.1007/s10518-017-0299-z.

Trnkoczy A.; 2009: *Understanding and parameter setting of STA/LTA trigger algorithm.* In: Bormann P. (ed), New manual of seismological observatory practice (NMSOP-2)), IASPEI, GFZ German Research Centre for Geosciences, Postdam, Germany, doi: 10.2312/GFZ.NMSOP-2_IS_8.1.

Varchetta F., Massa M., Puglia R., Danecek P., Rao S., Mandiello A. and Piccinini D.; 2024: *SDQ: un tool Python-based per la valutazione della qualità dei dati sismo-accelerometrici della Rete Sismica Nazionale dell'INGV*. INGV, Roma, Italy, Rapp. Tec., 482, 32 pp., doi: 10.13127/rpt/482.

Withers M., Aster R., Young C., Beiriger J., Harris M., Moore S. and Trujillo J.; 1998: *A comparison of select trigger algorithms for automated global seismic phase and event detection.* Bull. Seismol. Soc. Am., 88, 95-106, doi: 10.1785/BSSA0880010095.

Corresponding author: Fabio Varchetta
INGV, Istituto Nazionale di Geofisica e Vulcanologia, Sezione di Milano
Viale delle Rimembranze di Lambrate 18, 20134 Milano, Italy
Phone: +39 3483789322; e-mail: fabio.varchetta@ingv.it