SDQ (Seismic Data Quality): a Python project for seismo-accelerometric data quality check

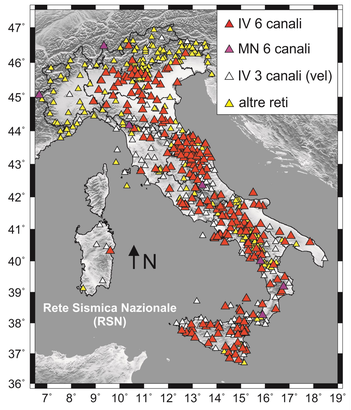
Fabio Varchetta[1],Marco Massa[1],Rodolfo Puglia[1],Peter Danecek[2],Sandro Rao[2],Alfonso Mandiello[2],Davide Piccinini[3]

[1] INGV, sezione di Milano, Italia , [2] INGV, ONT, Osservatorio Nazionale Terremoti, Roma, Italia , [3] INGV, sezione di PISA, Italia

Sharing is encouraged

Introduction

In this work, we introduce the Seismic Data Quality (SDQ) project. SDQ is an open-source Python-based tool package that verifies the functioning of accelerometric and velocimetric stations and assesses the records in terms of quality classification.
These verifications are conducted through analyses of seismic events and continuous data streams.
Regarding earthquake data, these verifications are based on comparing ground motion parameters derived from co-located accelerometers and velocimeters, subsequently classifying the waveforms into quality classes. For continuous data, the verifications rely on ambient instrumental noise.



SDQ project is focused on two main branches:

SDQ event which works on waveforms from earthquakes

SDQ stream which works on continous data stereams

Methods

Here we describe the algorithmic structures of two SDQ codes. In particular, the codes run through a command line in a terminal. When considering **SDQ events code** , the command line must contain the names of external input files, where the ID-event, netcode, and stacode are indicated, along with the columns containing this information.

./SDQ_event.py –wl FILE_INPUT –cwl 0 1 2

ACQUISITION
File Input (EventID, Stacode, Netcode) → Waveform and metadata download → EIDA Italia → Signal cutting and identification of event/noise windows

PRE-PROCESSING
Pre-processing →
- Baseline correction (detrend)
- Remove of mean
- Pass band [fmin = 0.001; fmax=50]
- Tapering of signal

Waveform exclusion conditions →
- Verification of the 6 components
- Clip verification on EH,HH [70% * (2^24-1)] (for digit 24 bits)
- Multiple Event Dector: T05 of Arias Intensity is within event window on accelerometer waveform
- Comparison of the event-to-noise RMS (Root Mean Square) ratio with a thresold value

Conversion of signals from counts to physical units

Determininig acceleration and velocity traces by deriving and integrating accelerometric and velocimetric waveforms

PROCESSING
Definition of Quality Class → Comparison of Integral ratio event on pre-event noise in the three frequency bands 0.3-1; 1-5 Hz; 5-15 Hz againts a threshold value →
- If 3 band > thr = class A
- If 2 band > thr = class B
- If 1 band > thr = class C
- If ratio < thr = class D

Other Checks →
- Comparison of horizontal components
- Comparison of acceleration shaking values with values estimated by attenuation law

Automatic Filtering → Band pass: [Fmin = SNR (Signal to Noise Ratio), Fmax = 0.80% Fnyquist]

Definition of shaking parameters for raw and filtered signals

Normalised cross-correlation of raw and filtered acceleration signals from accelerometer and velocimeter

Considering **SDQ code** based on **analysis of instrumental noise data** , the command lines must contain: the name of the external input file (where only the net code and station code are reported), the start date expressed in ymd format, and the analysis interval expressed in days, months, or years.

./SDQ_stream.py –wl FILE_INPUT –cwl 0 1 –ymd 2023-11-17 –i day –nu 1

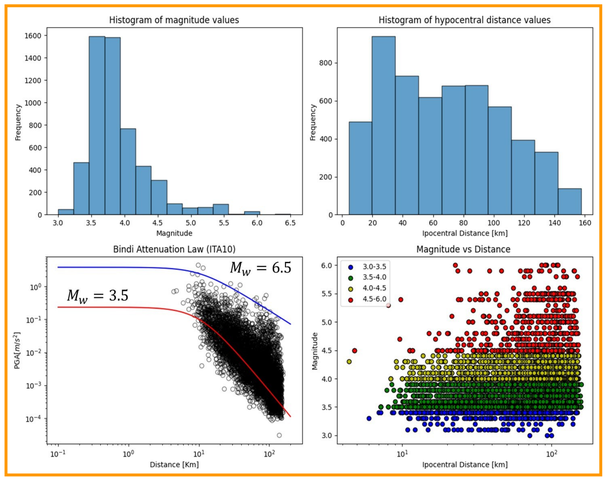
ACQUISITION
File Input (Netcode,Stacode) → Waveform and metadata download → EIDA Italia, SeisComP, database

Quality check
Evaluation of quality parameter →
- data gap or data availability (%);
- number of gaps;
- duration of the sum of the gaps (s);
- duration of the maximum temporal gap (s);
- mean root mean square (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)

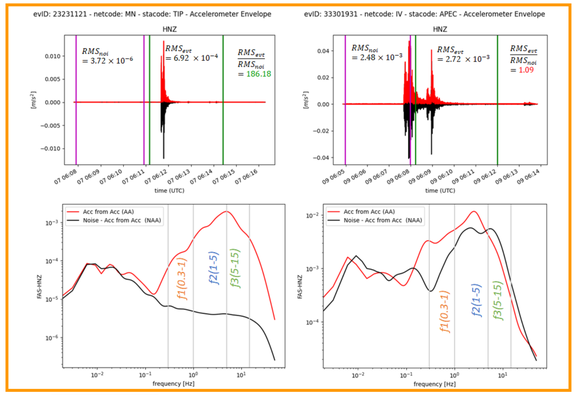
Definition of Quality class → Comparison quality parameter with empirical thresholds → Class A : High, Class B: Good, Class C: Low, Class D: Bad

Data

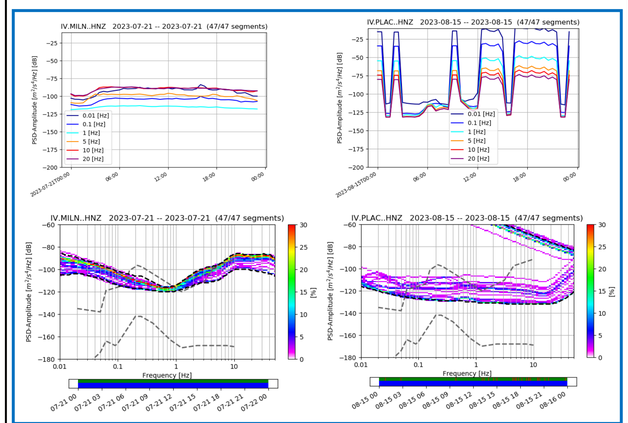
SDQ event were tested considering 200 stations from the Italian Seismic Networks IV and the Mednet Network MN (see **red triangle** ▲ and **purple triangle** ▲ on the Left Figure)
We evaluated **15.000** waveforms with $M > 3.5$ and **Ipocentral Distance** ($R_{ipo} < 150\ km$) from EIDA, covering the time period from **2012- 2022** (**see figures below**)



Each waveform must first pass certain exclusion conditions before the quality class is assigned and processing can begin. In the **figures below** , the root mean square ratio (RMS) conditions and classification of the waveforms into quality classes in the frequency domain are shown.



Regarding the **SDQ streams code** , we analyze the instrumental noise and evaluate quality metrics to compare them with empirical thresholds in order to classify the recordings into quality classes. In the **figure below** , we present an example of quality metrics: Power Spectral Density (PSD) and Probability Density Function (PDF) of PSD for stations MILN and PLAC.



Results

SDQ provides summary tables for both earthquake and continuous data, as well as summary images of stations, useful for monitoring the functioning of stations, along with explanatory text files (log and warning files), allowing users to better evaluate the results.
Considering **SDQ-based events** , we have summary images where we compare Peak Ground Acceleration (PGA) derived from co-located accelerometers and velocimeters. As observed in the **figures below** , we present four cases, considering the stations IV.ACER (Acerenza, PZ), IV.NRCA (Norcia,PG), IV.APRC (Apricena, FG), and IV.CTL8 (Castelleone, CR) from left to right.
We define correct functioning of stations when the value of PGA recorded for each component is equal and they align on a reference line (represented in black).

It is possible to distinguish specific patterns:
•correct functioning
•Pattern linked to set an uncorrect clipping threshold
•Pattern that enhanced problems

Summary

We propose SDQ project, a Python package, in which we assessed the functioning of seismic stations located throughout the national territory and assign the quality of recording in a specific **quality class**.

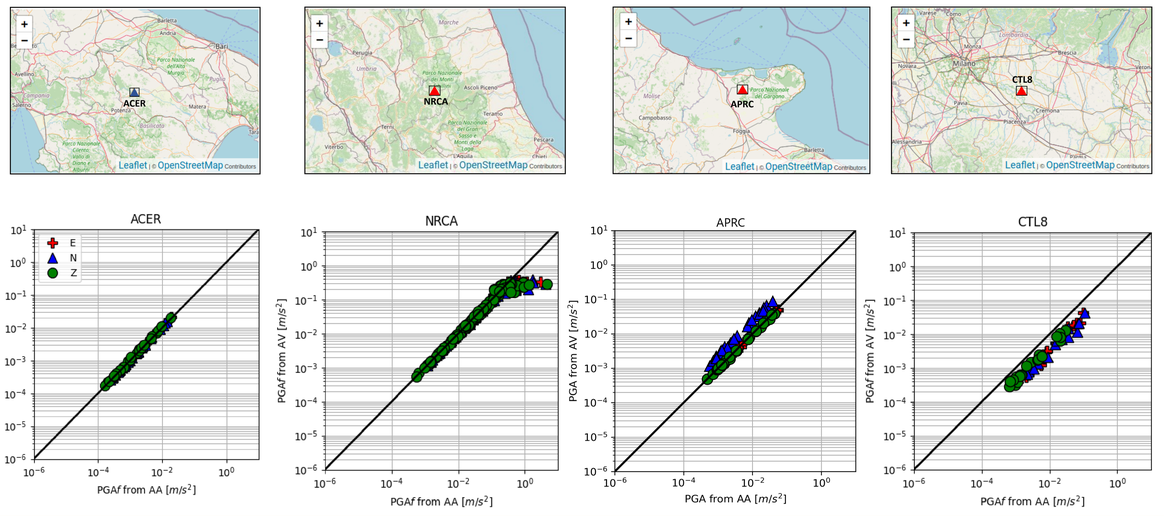
SDQ events code testing phase **results:**
•identifying correct fuctioning of seismic stations
•recognizing saturation of velocimeter during high-magnitude events at short hypocentral distances with incorrect threshold clipping
•identifying sensor misalignment and issues with one or more motion components
•identifying stations with temporary or persistent problems

SDQ streams code testing phase **results:**
- Identifying correct functioning of seismic stations observing constant patterns of temporal series from quality metrics

Outlook

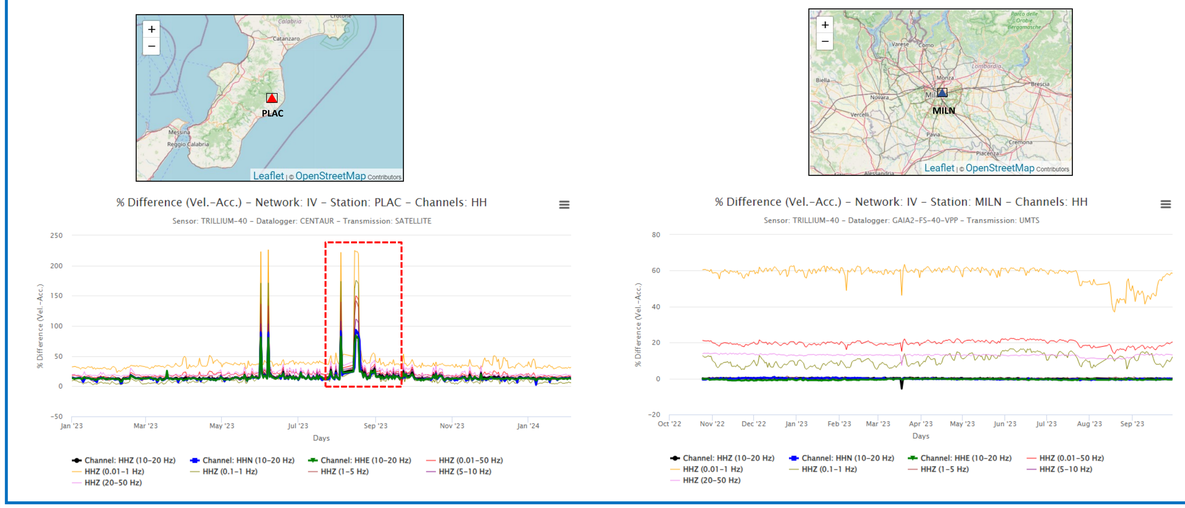
•Integrate SDQ packages into structured databases for automatic publication of data quality parameters
•Integration involves automatically updating the input file with data from new earthquakes or new stations
•Goal: Ensure dynamic monitoring of the health status of the national seismic network

SDQ_events: Summary images stations



Considering **SDQ streams code** , from the CSV table, it's possible to build temporal series of quality metrics and observe their behavior.
In the figures shown in the **panel below** , we present two cases for stations IV.PLAC (Placanica, RG), and IV. MILN (Milano, MI). Here, we plotted the percentage difference from the Power Spectral Density (PSD) of the velocimeter and PSD of the accelerometer for specific frequency bands.
We define the correct functioning of stations when the pattern of a temporal series is constant, as seen in the MILN case figures below on the right, and when anomalies are not observed in the signals, represented with a red rectangle in the PLAC case (figures below on the left).

SDQ_streams: Temporal series



References

For other informations this work is just published on Rapporti Tecnici INGV:

Varchetta F. et al., (2024). SDQ: ***un tool Python-based per la valutazione della qualità dei dati sismo-accelerometrici della Rete Sismica Nazionale dell'INGV.*** *Rapp. Tec. INGV, 482: 1-32, https://doi.org/10.13127/rpt/482*

Further information

Although SDQ is currently under development, it is freely available for download at https://gitlab.rm.ingv.it/EIDA/quality/sdq.
If you have a question or comment, please don't hesitate to contact me:
fabio.varchetta@ingv.it

EGU General Assembly 2024

ISTITUTO NAZIONALE DI GEOFISICA E VULCANOLOGIA