

ISMDq (INGV Strong-Motion Data quality)

The beta version of the *STATION QUALITY* session of the *ISMDq* web portal is devoted to the evaluation of the data recorded in continuous mode by the accelerometric stations of the *INGV* National Seismic Network (<http://terremoti.ingv.it/instruments/network/IV>) and partners that have assured real-time data exchange with the *INGV* through active agreements between public authorities and Universities (see also <http://terremoti.ingv.it/instruments>; <http://ismd.mi.ingv.it/notes-net.php>).

The link for *STATION QUALITY* is located in the right framework (at the bottom) of the home page of the *ISMDq* website (<http://ismd.mi.ingv.it/>), leading users to enter in the station-quality web page (<http://ismd.mi.ingv.it/quality.php/>).

The results for the accelerometric stations are available on-line and archived in the *ISMDq* database from October 2020, while from April 2021 the results for the velocimetric sensors coupled to the accelerometers are also available.

The screenshot displays the ISMDq web portal interface. At the top, the logo reads "ISMDq v.2.1 INGV STRONG MOTION DATA" with the tagline "Rapid Response and quality check". To the right is the logo for the "ISTITUTO NAZIONALE DI GEOFISICA E VULCANOLOGIA" and a "REAL TIME" banner. Below the header is a navigation bar with tabs: HOME PAGE, THE PROJECT, EARTHQUAKES, STATIONS, WAVEFORMS, QUALITY, and INFORMATION. The main content area is divided into two sections. On the left, under the heading "Archive: last 6 months", there are four panels: "DATA METRICS" (line graph), "DATA AVAILABILITY" (line graph), "DAILY PLOTS" (seismic waveforms), and "INFORMATION" (document stack). On the right, under "Last daily update (2021-05-14)", there is a map of Italy showing station locations. A legend indicates quality levels: High (green triangle), Good (yellow triangle), Low (orange triangle), Bad (red triangle), and No data (black triangle). The map includes search filters for "HN" and "Z", an "Availability (%)" dropdown, and a "GO" button. A "TABLE view" button is also present. The map is powered by Leaflet and OpenStreetMap.

The *quality.php* page is composed by two main frameworks

1) QUALITY OF THE LAST AVAILABLE DAY

On the right there is a real time stations map with the updated parameters analyzed considering the last available day. The current beta version, due to technical reasons related to the loading of the large amount of data (about 1000 streams/day), currently has a delay of two days compared to the present (e.g. on December 15th the user will view the data of the day 12). The analyzes of the 24 hours carried out on each stream of each available accelerometric and velocimetric station will be displayed.

The map, in relation to the reference date (viewable in the upper part), shows with triangles all the accelerometric stations that send data in real time to the INGV-ONT (<http://www.ont.ingv.it/>) seismic room for surveillance. Data are then redundant to the INGV Milano (<http://www.mi.ingv.it/>) acquisition center for the presented analysis.

In the map at the top right you can see the box for choosing the parameters, the SEED channels (accelerometric: HN and HG; velocimetric HH and EH; see also <https://scedc.caltech.edu/data/station/seed.html>) and the ground motion component (Z: vertical; N: North-South; E: East-West).

Currently the following parameters, calculated on 24 hours of signals, are available:

- Gap or availability (i.e. indicated in the map as Gap/Availability%);
- Number of gaps (Gap count);
- Duration (s) of the sum of the gaps (Σ gap);
- Duration (s) of the maximum time gap (max gap);
- quadratic mean (RMS, Goldstain et al., 2003) without filter application (RMS no-filt.);
- RMS with Butterworth band-pass filter application (RMS filt. 0.01-50 Hz);
- Power Spectral Density (PSD) with Butterworth band-pass filter (PSD filt. 0.01-50 Hz);
- general quality index.

In the map at the bottom left there is the key to access the same results in tabular form where each color of the cell has the same meaning as the relative triangle in the map. Both in the map and in the table the colors green, yellow, orange, red respectively indicate a decreasing data quality from excellent (good, sufficient) to insufficient, defined according to the intervals below (missing data are indicated in black). The selected values for quality class are reported in the table below.

SEED channels *HN, HG*

Indice di Qualità	A	B	C	D
disponibilità (%)	>90	90:75	75:50	<50
gap (%)	<10	10:25	25:50	>50
num. gap	<50	50:100	100:300	> 300
Σ gap (sec)	<1800	1800:3600	3600:(3600*3)	> 3600*3
max gap (sec)	<1800	1800:3600	3600:(3600*3)	> 3600*3
rms-no fil (count)	<50k	50k:100k	100k:200k	> 200k
rms-fil (count)	<10k	10k:20k	20k:30k	> 30k
PSD-fil (dB)	(-120:-100)	(-125:-120)(-100:-95)	(-135:-125) & (-95:-85)	<-135 & >-85

SEED channels *HH, EH*

rms-no fil (count)	<5k	5k:15k	15k:40k	> 40k
rms-fil (count)	<3k	3k:10k	10k:20k	> 20k
PSD-fil (dB)	(-130:-160)	(-130:-110)	(-110:-100)	<160 & >-100

Following the results obtained in the estimation of the RMS (filtered between 0.01 and 50 Hz), of the PSD (filtered between 0.01 and 50 Hz) and considering the % of daily data availability, each station, channel and component are associated with a general quality index indicated with the alphanumeric codes A (green: high quality), B (yellow: good quality), C (orange: low but sufficient quality), D (red: bad quality, to be removed).

The general quality index is expressed both in the map and in the corresponding table by means of a color (triangles in the map and cell in the table) associated with the lower value selected considering the % of data availability and the average PSD in the interval 0.01-50 Hz, for each selected channel and component (ex: availability > 90 & PSD -97 dB = general quality index B, yellow color).

2) STATION QUALITY ARCHIVE (6 MONTHS)

On the left side of the web page *quality.php* are showed the results for the the archived data on db-quality for the last 6 months. The data is available starting from October 2020 for SEED channels *HN* and *HG* and from April 2021 for the SEED channels *HH* and *EH*.

The archive includes 3 main sessions: Metrics, Availability, Daily plots.

DATA METRICS

The graphs for the data metric show the quality parameters for the last 6 months. The requests can be made for each network, station, channel and component relating to all stations that flow in real time to the INGV seismic room for surveillance. The graphs allow to build the time history of the recording stations through the average root mean square (RMS), the average Power Spectral Density (PSD) and a comparison between accelerometers and coupled velocimeters in terms of percentage difference between averaged PSD (% diff acc-vel). All analyzes are carried out for successive filtering intervals between 0.01 and 50 Hz (i.e., 0.01-0.1; 0.1-1; 1-5; 5-10; 10-20; 20:50 Hz) capable of covering the entire range of seismological interest and highlight the nature of possible problems at the registration site. Each parameter is displayed associated with its relative degree of uncertainty ($\pm 1\sigma$). For each selected parameter, the graphs display all the motion components represented in different colors. All dynamic graphs can be enlarged by the user for details.

DATA AVAILABILITY

The data availability graphs allow users to build the time history of the last six months for each network, station, channel and component, in relation to the daily gap (%) in the continuous recorded data stream, to the maximum daily quantity of gaps (number), or to the sum (sec) of all gaps and the maximum duration in seconds of the maximum daily gap. For each selected parameter, the graphs display all the motion components represented in different colors. All dynamic graphs can be enlarged by the user for details.

DAILY PLOTS

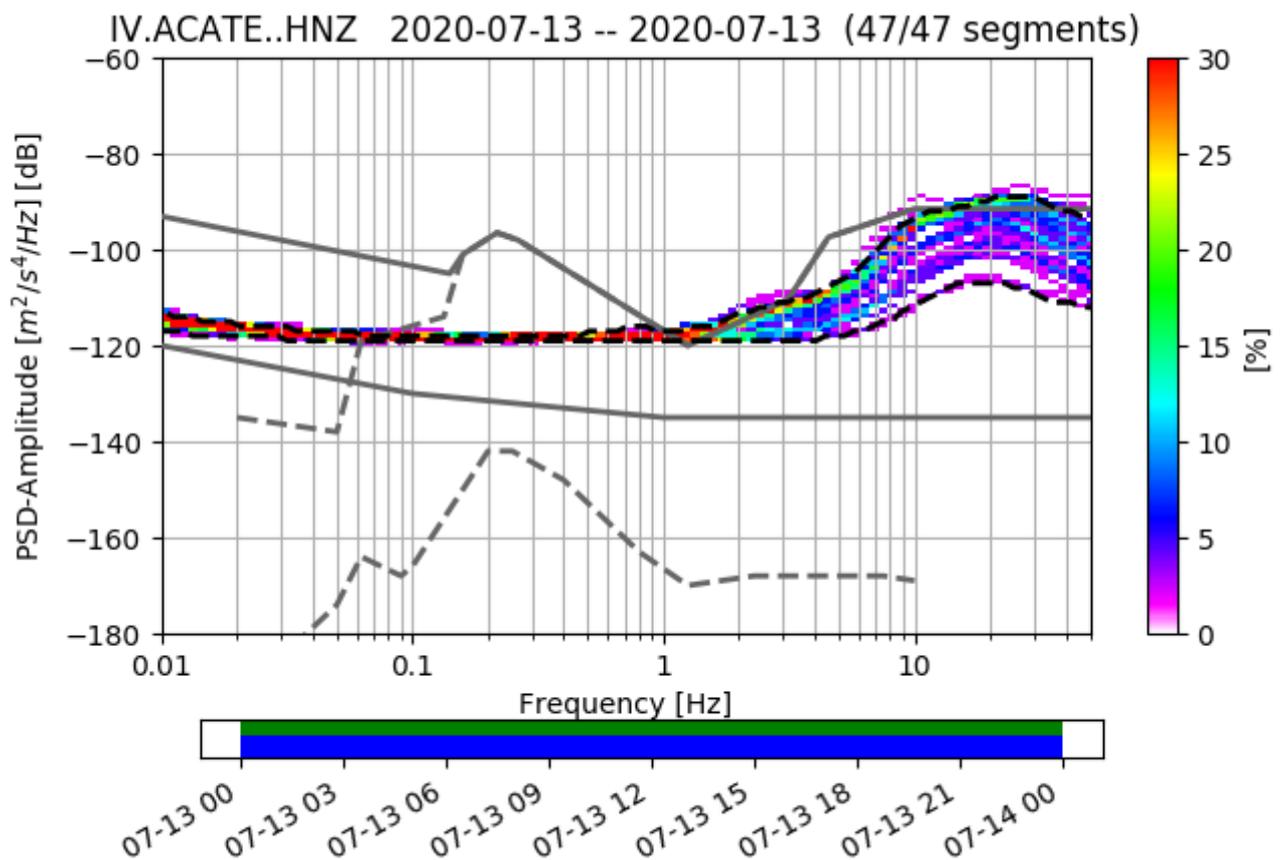
For each day (24 hours) of the last 6 months you can view the static graphs that can be selected by network and station (all channels) relating to the following analyses:

- PDF: Probability Density Function;
- PSD: Power Spectral Density at different frequency intervals;
- comparison RMS (quadratic mean) vs. PSD at different frequency intervals;
- daily graphs for waveforms (24 hours);
- PSD spectrograms;
- gaps.

In detail:

PDF

The analyzes are made considering 24 hours of data in miniSEED format (<http://ds.iris.edu/ds/nodes/dmc/data/formats/miniseed>). The title of the graph shows the network (e.g. IV), the station code (e.g. ACATE), the accelerometric channel (e.g. HNZ), the date considered for the analysis and the number of windows (segmets, e.g. 47 / 47) used for the calculation of PSDs.



The figure above shows the noise models (minimum and maximum) by Cauzzi et al. (2013) (solid gray lines) and Peterson's (1993), converted to acceleration (dashed gray lines).

The black dotted lines show the $\pm 2\sigma$ values relating to the cumulative PDF from the start of the analysis (for a maximum period of 6 months) and currently represent a first approximation for a noise model at the single station. They will soon be replaced by maximum and minimum noise models calculated off-line for each INGV station included in the db-quality.

The green bar below the graph indicates the input data for the PDF calculation. Any gaps are indicated in red (absent in the example). The blue bar indicates the single PSD measurement that goes into the frequency distribution histogram. As these are accelerometers, the conversion from count to cm/s^2 is carried out by considering the channel sensitivity for each station, setting a flat

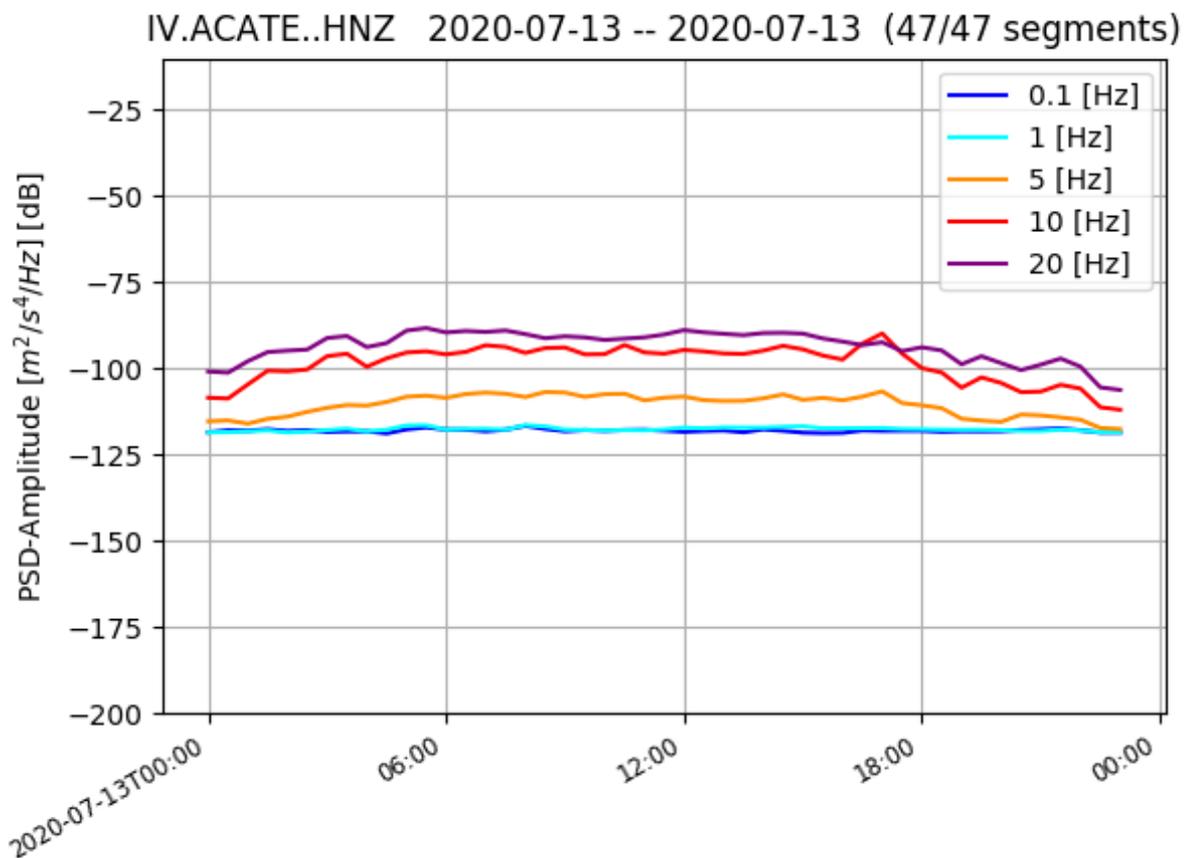
instrumental response. The analysis windows considered have a time length of 1 hour, overlapping by 50%. Each 1h window is further subdivided into 13 segments (overlapping by 75%; McNamara, 2004). For each of these the following processing is carried out:

- linear trend and offset removal
- cosine taper 10%
- PSD calculation

The PSD for each hour is obtained from the average of the PSDs of the 13 segments. For the calculation of the PDF, only the hourly PSDs are considered to calculate the frequency distributions at each period.

PSD (frequency dependent)

The Power Spectral density (PSD) is reported as a function of time for each analyzed day and for different frequency intervals. The title shows the network, the station code, the channel, the date considered and the number of segments used for the calculation of the PSD. The graph shows for each interval the average of the values of the same interval (indicated with different colors for each frequency interval).

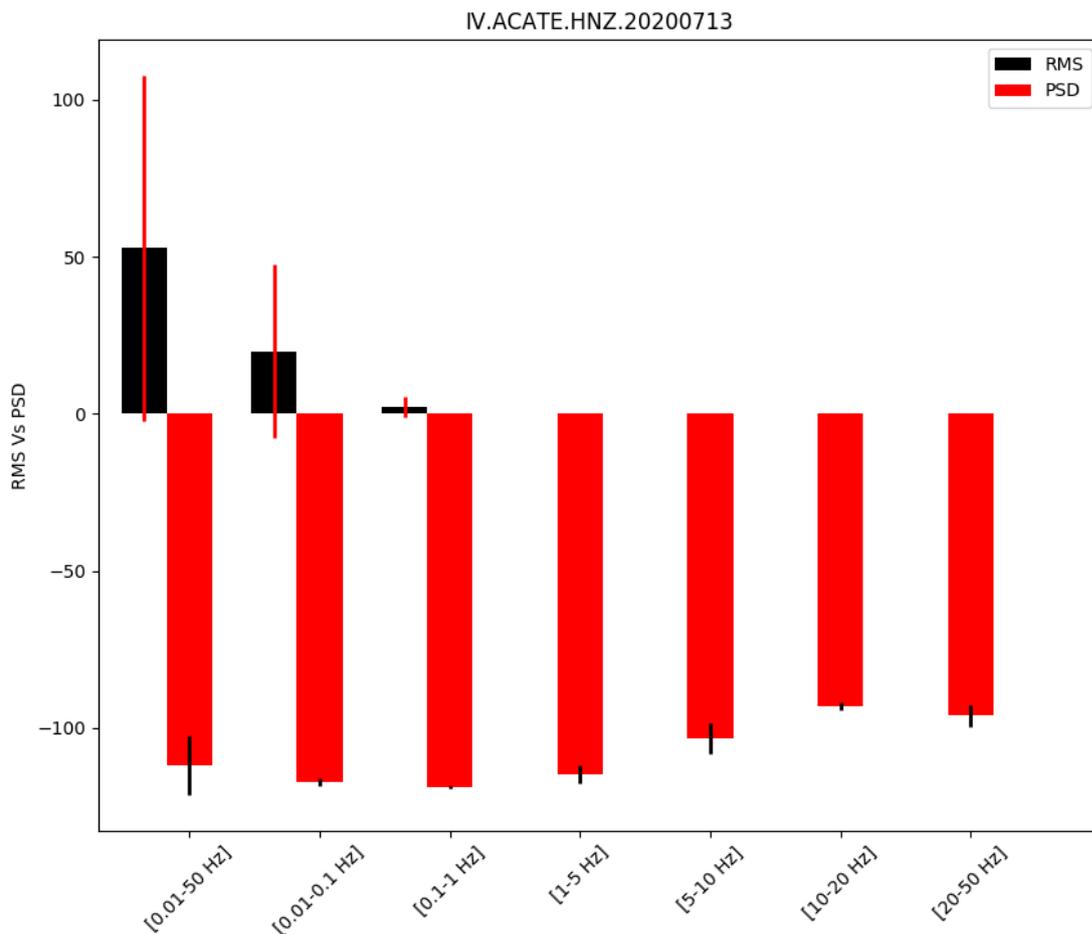


RMS vs. PSD

The values of the root mean square (RMS) and of the PSD calculated for the target day and for different frequency intervals are shown. The bars indicate the average values (black for RMS, red for PSD). The lines on the bars indicate their respective standard deviations (the colors of the lines are reversed from the bars to aid visualization). The RMS is calculated in count according to the formulation:

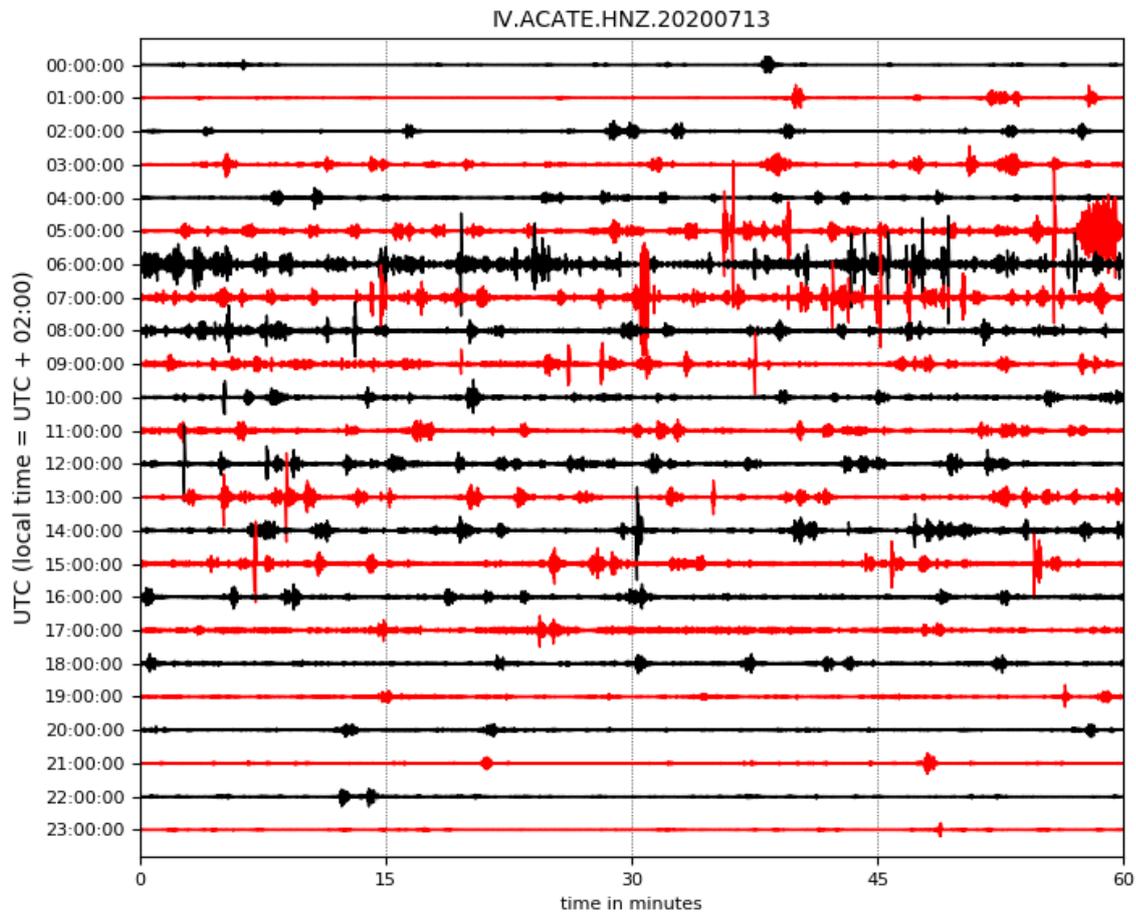
$$x_{\text{rms}} = \sqrt{[1/n (x_1^2 + x_2^2 + \dots + x_n^2)]}$$

where x is the amplitude value (count) of the single sample and n the number of samples of the considered trace.



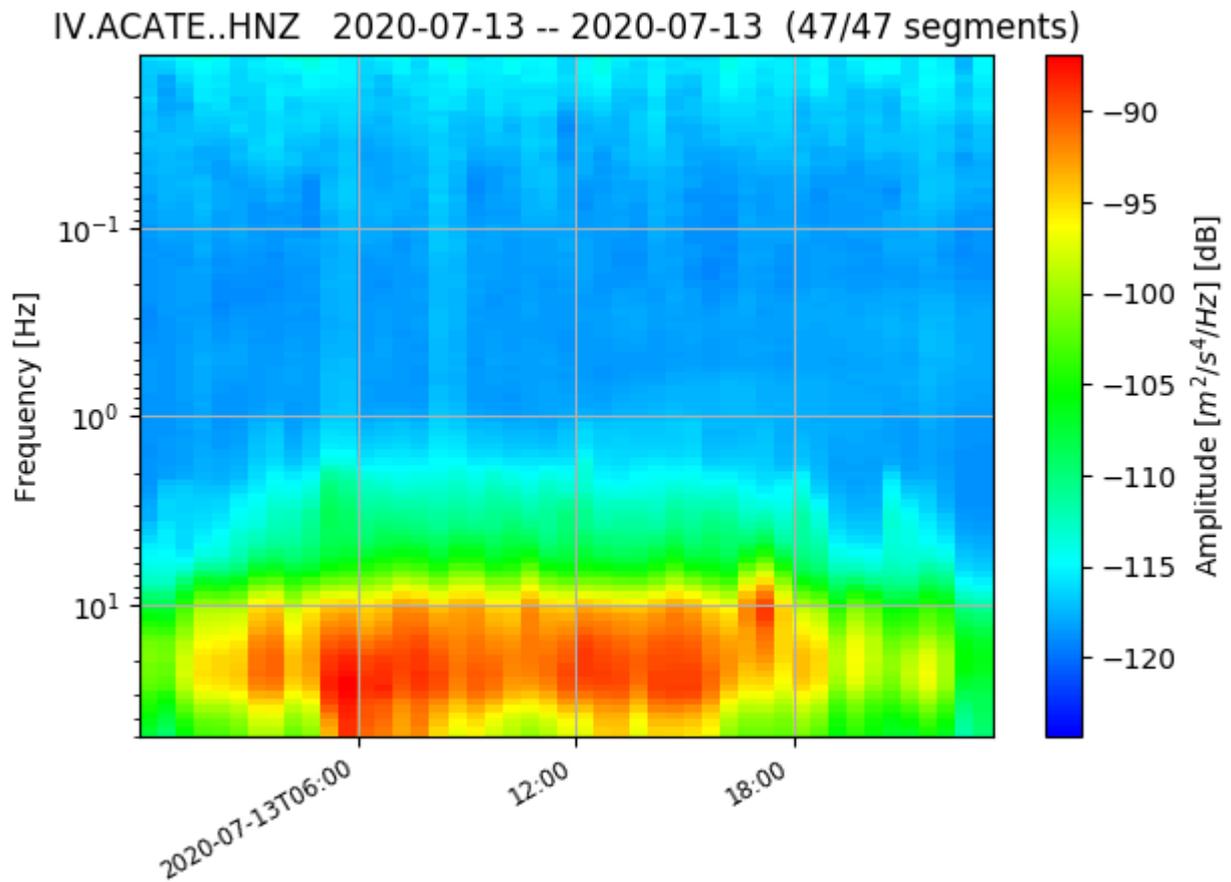
24 hours of waveforms plot

The title of the graph shows the network, the station code, the channel and the date of the day analyzed. The signal is represented by hourly intervals (abscissa in minutes). The alternating colors simply aid in reading the graph.



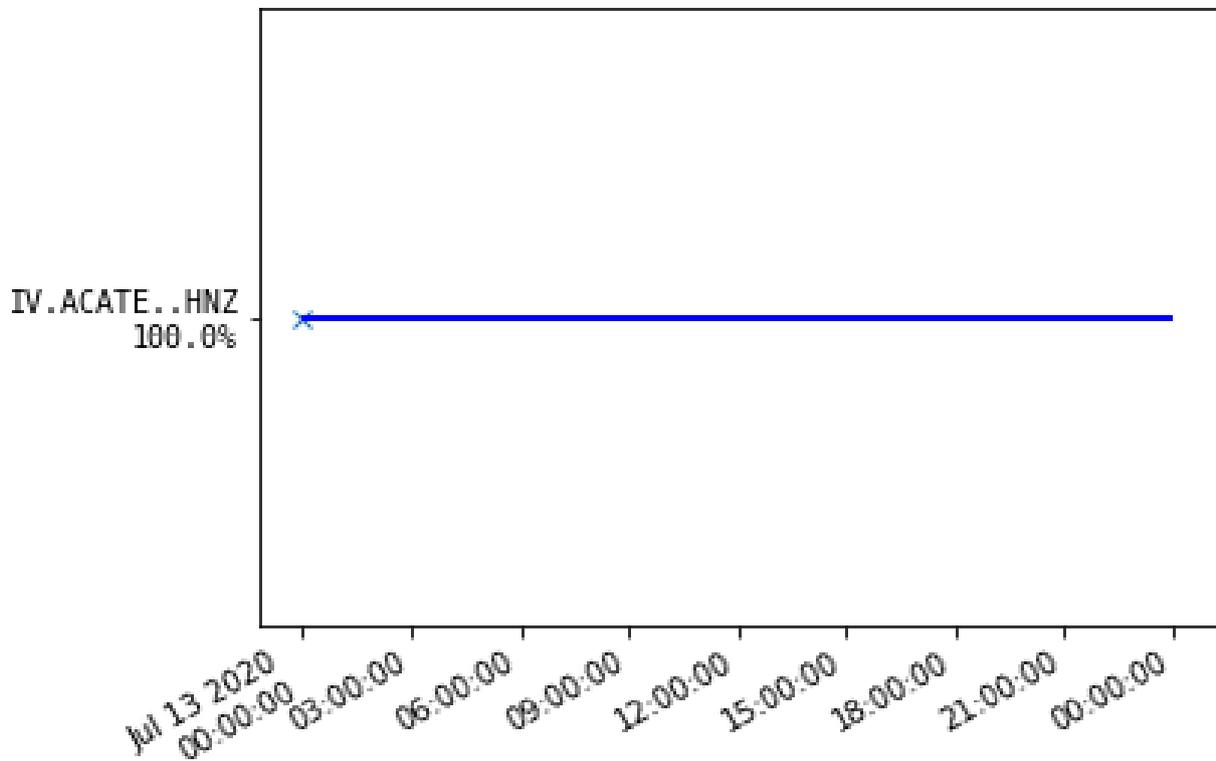
Spectrogram of PSD

The title of the graph shows the network, the station code, the channel, the date of the day analyzed and the number of segments used for the calculation of the PSD.



Gaps

The percentage shown on the left together the station code indicate the availability of data (in this case 100% and gaps 0%) for the analyzed day (see the abscissa). Any vertical red and blue bars (both absent in the example) indicate the presence of gaps and overlaps respectively.



References

Clinton J. and Cauzzi V. (2013). A High- and Low-Noise Model for High-Quality Strong-Motion Accelerometer Stations, *Earthquake Spectra*, 29(1), 85-102, doi:10.1193/1.4000107.

Goldstein, P., D. Dodge, M. Firpo, Lee Minner (2003) “SAC2000: Signal processing and analysis tools for seismologists and engineers, Invited contribution to “The IASPEI International Handbook of Earthquake and Engineering Seismology”, Edited by WHK Lee, H. Kanamori, P.C. Jennings, and C. Kisslinger, Academic Press, London.

McNamara D.E. and Buland R.P. (2004). Ambient Noise Levels in the Continental United States, *Bulletin Seismological Society of America*, 94, 4, 1517–1527. <https://doi.org/10.1785/012003001>.

Peterson J.R. (1993). Observation and modeling of seismic background noise, *U.S. Geol. Surv. Tech. Tech. Rept.*, 93-322, 1-95.