

# **HV Noise and Earthquake Automatic analysis (*HVNEA*)**

**PREPRINT of paper <https://doi.org/10.1785/0220220115>**

**Maurizio Vassallo<sup>1</sup> (corresponding author, [maurizio.vassallo@ingv.it](mailto:maurizio.vassallo@ingv.it)), Gaetano Riccio<sup>2</sup> ([gaetano.riccio@ingv.it](mailto:gaetano.riccio@ingv.it)), Alessia Mercuri<sup>3</sup> ([alessia.mercuri@ingv.it](mailto:alessia.mercuri@ingv.it)), Giovanna Cultrera<sup>3</sup> ([giovanna.cultrera@ingv.it](mailto:giovanna.cultrera@ingv.it)), Giuseppe Di Giulio<sup>1</sup> ([giuseppe.digiulio@ingv.it](mailto:giuseppe.digiulio@ingv.it))**

<sup>1</sup> Istituto Nazionale di Geofisica e Vulcanologia, Sezione Roma1, viale Francesco Crispi, 43, 67100, L'Aquila Italy

<sup>2</sup> Istituto Nazionale di Geofisica e Vulcanologia, Sezione Roma1, C.da Ciavolone, 83035, Grottaminarda (AV) Italy.

<sup>3</sup> Istituto Nazionale di Geofisica e Vulcanologia, Sezione Roma1, via di Vigna Murata, 605, 00143, Rome Italy

***Cite as:** Vassallo, M., G. Riccio, A. Mercuri, G. Cultrera, and G. Di Giulio (2022). HV Noise and Earthquake Automatic Analysis (*HVNEA*), *Seismol. Res. Lett. Seismological Research Letters* 2022;; 94 (1): 350–368. doi: <https://doi.org/10.1785/0220220115>*

## **Declaration of Competing Interests**

The authors acknowledge there are no conflicts of interest recorded



## Abstract

We present *HVNEA* tool (HV Noise and Earthquake Automatic analysis), a software package to automatically compute the horizontal-to-vertical spectral ratios (HV) on continuous months-to-years recordings, including both earthquake and ambient noise vibration. It is designed for processing large datasets of seismic signals, e.g. from long-term permanent or temporary stations belonging to monitoring networks, by using free softwares and packages that are commonly adopted in the engineering-seismology community. *HVNEA* is in the form of Linux/Unix shell scripts and packaged in a Docker image to run the code in any environment. It is designed to automatically download data from the main international repositories or from a local archive. The main achievement consists of the average HV as a function of time and saved in ascii format, along with attractive and readable graphic plots in order to easily inspect the results. The analysis highlights the variation of the HV over a long span of time. *HVNEA* software package can be of great interest for big data analysts and seismic network operators, with regard to the site characterization analysis and for studies based on temporal variability of HV ratios over time.

## 1. Introduction

The number of seismic stations in the national or local networks, together with the availability of seismic data within the seismological and engineering community, are nowadays increasing very rapidly. Modern networks are generally equipped with six-channels digitizers (both accelerometer and velocimeter) recording in continuous mode, and the improvement of technologies makes it easy the retrieval of a large amount of seismic data from dedicated open web-services (Cauzzi et al. 2021).

Therefore it has become important to develop automatic tools to carry out extensive and routinary analyses on both ambient vibrations and earthquake data, for assessing the data quality, or indicating possible local effects related to rheological or morphological properties of the ground.

Among the simple single-station methods of analysis, HV (Horizontal-to-Vertical spectral ratio) is widely used for studies of site effects and seismic microzonation applications. It consists in calculating the ratio between the Fourier amplitude of horizontal and vertical components (HV) of the ground motion recorded by a single station. The purpose of HV

analysis is to get a raw estimate of the transfer function of the site under investigation, at least in terms of the fundamental resonance frequency ( $f_0$ ).

HV computed on earthquakes can be considered as a first-order level of analysis to characterise a site, and is usually followed by more accurate seismological studies to evaluate the empirical transfer functions, as for example the standard spectral analysis using a reference site (Borcherdt 1970). The hypothesis underlying the HV on earthquake recordings is that the vertical component of motion is not affected by the properties of the local geology and composition of the more superficial layers. Consequently, by deconvolving the vertical component from the horizontal component of the motion (i.e. by calculating the ratio between spectra), a raw estimate of the site transfer function is obtained. A fairly exhaustive treatment of the method, also from the analytical point of view, can be found in Field et al. 1995, where comparisons between results deriving from analyses with various types of techniques are also presented. The HV method on earthquakes has been applied for various purposes, for example to derive some properties of the Earth's crust using teleseismic recordings (Dünder et al. 2016), or for the estimation of the site transfer function, by means of the HV ratios calculated on windows containing the S-wave phase of earthquakes (Bonilla et al. 1997; Theodoulidis et al. 2018). Although some observations indicate that the vertical components is often amplified during seismic sequences (Felicetta et al. 2021), HV ratios using earthquake records is largely used for site characterization analysis, together with standard-to-spectral ratios using a reference site (Borcherdt 1970) or combined to numerical modelling or ad-hoc correction (Luo et al. 2020), to obtain an estimate of the fundamental resonance frequency  $f_0$  and indication of site response.

The HV method applied to ambient seismic noise (Nogoshi and Igarashi 1971), so-called Nakamura method (Nakamura 1989 and 2019), has drawn particular attention in the last decades because it is very convenient and fast (therefore also inexpensive) to apply especially in areas with low seismicity (see SESAME 2004 for very comprehensive guidelines on the method and data processing). It is widely adopted in many studies focused on the estimation of site resonance frequencies, such as: seismic microzonation activities (Priolo et al. 2019), characterization of sedimentary basins (Parolai et al. 2002; Guéguen et al. 2007), investigation of the buildings response (Bindi et al. 2015; Riccio et al. 2017), verification of the suitability of a seismic station to act as a reference site (Milana et al. 2019; Lanzano et al. 2020) among many other applications. However, it is not reliable for estimating the ground motion amplification factors (see e.g. Lachet et al. 1994, Cultrera et al. 2014). Another

drawback of the method is that currently there is not a full agreement about its theoretical justifications (Bonney-Claudet et al. 2006; Lunedei and Malishevsky 2015; Piña-Flores et al. 2017; Molnar et al. 2022). Nakamura (2019) interprets the HV curves on noise in terms of the theoretical transfer function of SH-wave, whereas many other authors explain it as Rayleigh-wave ellipticity (Tokimatsu 1997, Ferreira & Woodhouse 2007). A contribution of the Airy-phase of Love wave in increasing the amplitude of the HV peak was also found by Bonney-Claudet et al. (2008) especially in presence of strong-velocity seismic contrast. Many recent studies (e.g. Sánchez-Sesma 2017) relate the HV noise curves to the imaginary part of the Green's function emerging from the auto-correlation function under the assumption of ambient diffuse field (Weaver 1982). The diffuse field concept was extended in the interpretation of HVs using earthquake data by Kawase et al. (2018), which proposed an empirical translation from HV computed on microtremors to HV computed on earthquakes. The latter mentioned article shows a correlation between HV computed on earthquakes and noise, with the possibility to obtain a direct estimation of S-wave amplification by means of a correction, combining HV ratios on microtremor and earthquake recordings at the same site (as also exploited in Zhu et al. 2020a).

Nowadays, permanent stations of many networks record data continuously, but the analysis of spectral ratios is still performed using a limited number of noise windows and a limited selection of earthquakes recorded at the station. Specific studies on temporal variations of the HV spectral ratios (considering long-term time scale) are limited in literature to few peculiar cases: for monitoring seasonal permafrost active layer variability (Köhler and Weidle, 2018); for monitoring the stability of rock masses (Colombero et al., 2018, Lotti et al. 2018); to study and quantify long-term temporal changes of site response after the largest earthquake in Japan (Wu and Peng, 2011, 2012). In these articles, the authors used the processing software developed by themselves (and not distributed to the community) in order to extend the analysis of spectral ratios over long time windows and many earthquake records. There is currently no software available in the seismological community to facilitate these operations. The aim of this work is to fill this gap by proposing a free software package to calculate and monitor variations in HV spectral ratios using both ambient seismic noise and earthquakes extracted from long time windows data (up to several years) acquired at a single station.

In this paper we describe *HVNEA*, a software package based on two programs that implement the computation of HVs: one operates on earthquake recordings, the other on ambient seismic

noise. They are developed for the main purpose of having a robust and reliable tool to quickly compute the HV ratios from large datasets of ambient noise and earthquake data. The availability of HV ratios on a large number of stations over a long time makes it possible to observe geographical and spatial variations of HV curves, and to look for their relationships with the human activities, meteorological conditions, or any other variable that may influence the site response over time. After a detailed description of *HVNEA*, we will show their potentiality through some applications on data acquired by two permanent seismic stations.

## **2. Description of *HVNEA* procedures**

*HVNEA* is implemented in the form of two main Bash shell scripts, making extensive use of third-party programs. In addition, we made the whole analysis package also available through a Docker image (Anderson, 2015), as described in *HVNEA* user manual in Supplemental Material. The advantage of it is that the running container is autonomous and completely independent from the host environment; and therefore the package can be used on any platform where the Docker Engine is available (Linux, Windows, MacOS, cloud).

*HVNEA* tools meet the needs to automatically perform HV ratios computed on both seismic noise and earthquake signals by managing a large amount of data. They exploit the capability to retrieve seismic records from the main waveform repositories, and automatically perform quick analyses on massive data collected in months or years by limiting the interaction with the user during the analysis process to what is strictly necessary. The main core of *HVNEA* is based on the Geopsy software (Wathelet et al. 2020), which is a free and robust software widely adopted in the engineering-seismology community. Geopsy is designed for the analysis of ambient noise data, and here it is adapted to work also on earthquake data. *HVNEA* was initially designed for a massive study on the seismic networks whose data are distributed through FDSN web services. Then the procedures were also adapted to work on seismic data locally available. In this latter case, the data must be filed according to the Seedlink Data Structure, and the necessary information about the station configuration must be provided by dataless files.

The two main scripts of *HVNEA* are named *HV\_Noise.sh* and *HV\_Eqk.sh* for noise and earthquake HV analysis, respectively. In both scripts, the Geopsy package is used in non-interactive mode, via the command line. Most of the parameters used by Geopsy for the

computation of the spectral ratios (number of frequency points of the output HV files, type of tapering and smoothing of the individual spectral ratios, etc.) can be customised by the user by acting on the configuration files ‘param\_hv\_noise.conf’ and ‘param\_hv\_eqk.conf’. Instead, the configuration files for the main scripts (hv\_eqk.conf and hv\_noise.conf) contain the file paths necessary for the scripts to work properly (for example the files needed to build the 1D velocity model for the HV\_Eqk.sh script, or the report templates for both scripts), the default values of some input arguments (e.g. the record length to be analyzed for each day for the HV\_Noise.sh script), some settings for creating the output figures (e.g. file type, linear or logarithmic scale, and so on). See *HVNEA* user manual in Supplemental Material.

For both the scripts, data files are downloaded through the FDSN web services, or read from a local archive, based on data directory paths set in ‘datapath.conf’ file. When the last configuration file is present and correctly point versus an archive, the scripts work on local data. Otherwise, the data download happens through FDSN web services (for more information, see the Supplemental Material).

## **2.1 HV Noise Spectral Ratios**

The script *HV\_Noise.sh* is used for the automatic computation of the HV spectral ratios on ambient seismic noise. Its flow diagram is shown in Figure 1. The HV ratios of single windows are averaged for each day by using the Geopsy program (geometric mean) and finally the daily mean HV ratios are averaged over the whole period using the geometric mean.

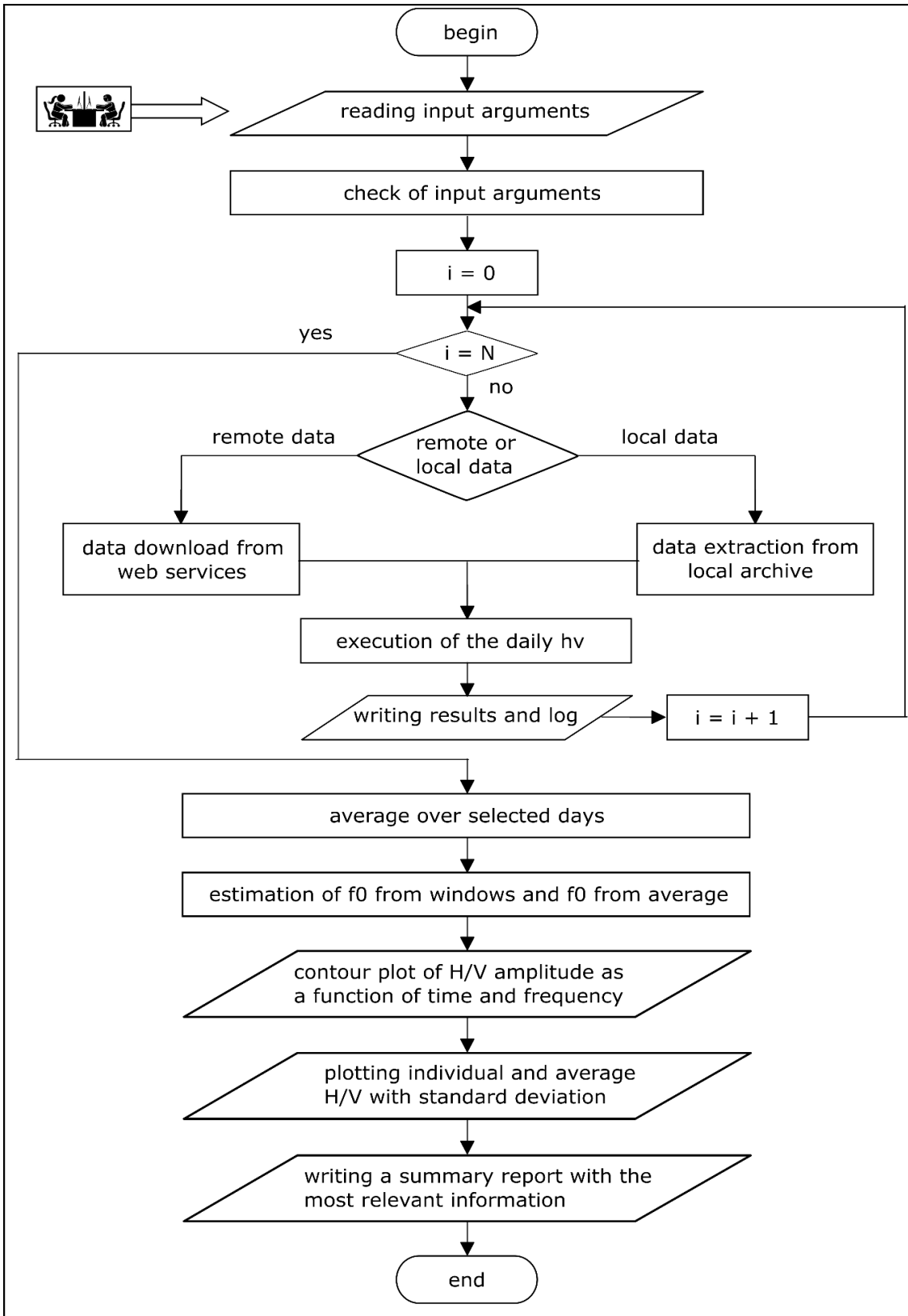


Figure 1. Simplified flow chart of HV\_Noise.sh.



The script execution is made through a simple command line by specifying as arguments all the parameters necessary for the analysis i.e.: name of the network and station, components to be analysed, start date and number of days on which the analysis is to be carried out.

As example, the command line

```
HV_Noise.sh -n IV -s INTR -c HH -f 2010-10-01 -d 114 -o OUT/INTR_HV10
```

gives the results shown in Figure 2; the analysis is performed for the stream HH (indicating a broadband high-gain seismometer) of station IV.INTR on a time interval of 114 days and starting time 2010-10-01 (see *HVNEA* user manual in Supplemental Material for details).

The main output files are:

- The ascii file with the final HV curve over the whole period and its standard deviation;
- an image showing the contour plot of the HV amplitude as a function of time and frequency, where the color scale is proportional to the spectral ratios amplitude (Fig. 2a);
- an image with the daily HV curves together with the overall average  $\pm 1$  standard deviation (Fig. 2b), evaluated by statistical analysis of the results produced by each individual data windows;
- a synthetic report in pdf format containing all the relevant information and the two produced images.

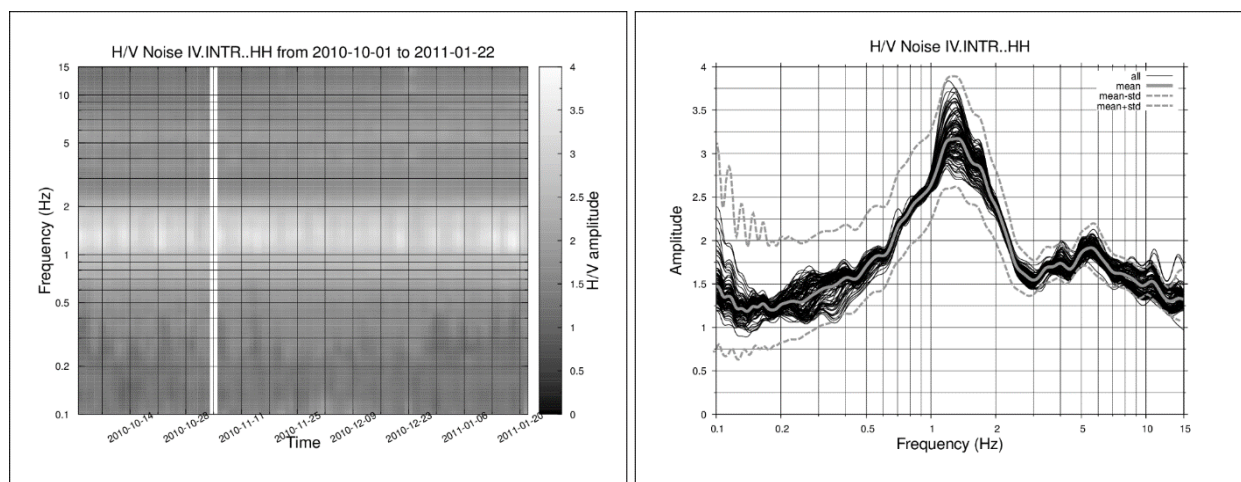


Figure 2. Output figures of `HV_Noise.sh` for HH stream of station IV.INTR, obtained from the analysis carried out on 114 days starting from 2010-10-01: (a) contour plot of the daily HV amplitude as a function of time and frequency, where the colour scale is proportional to the spectral ratios amplitude; (b) daily HV curves (in black) together with the overall average  $\pm 1$  standard deviation (in red).

The full list of input arguments and the complete output are explained in detail in the HVNEA user manual in Supplemental Material.

## 2.2 HV Earthquake Spectral Ratios

The Script `HV_Eqk.sh` performs the automatic computation of HV spectral ratios on earthquakes. It allows the selection of a list of earthquakes from a FDSN web service request tool (tool *event*) provided by the data center specified by the user in the configuration file, or from a ready-made list; note that this list does not take into account the real availability of the event recordings at the target station.

The script execution is made through a command line with the parameters required for the analysis as: network, station and data stream, and the parameters used to build the list of events or, alternatively, the name of a local file containing the ready-made list.

The flow diagram of `HV_Eqk.sh` is shown in Figure 3. The script initially shows a map with the selected events according to the search parameters provided on the command line, or alternatively with the events read from a text file, and asks the user whether to proceed in the analysis or not (Figure 3a). This map provides a first indication on the number, magnitude and azimuthal distribution of the events selected by the script.

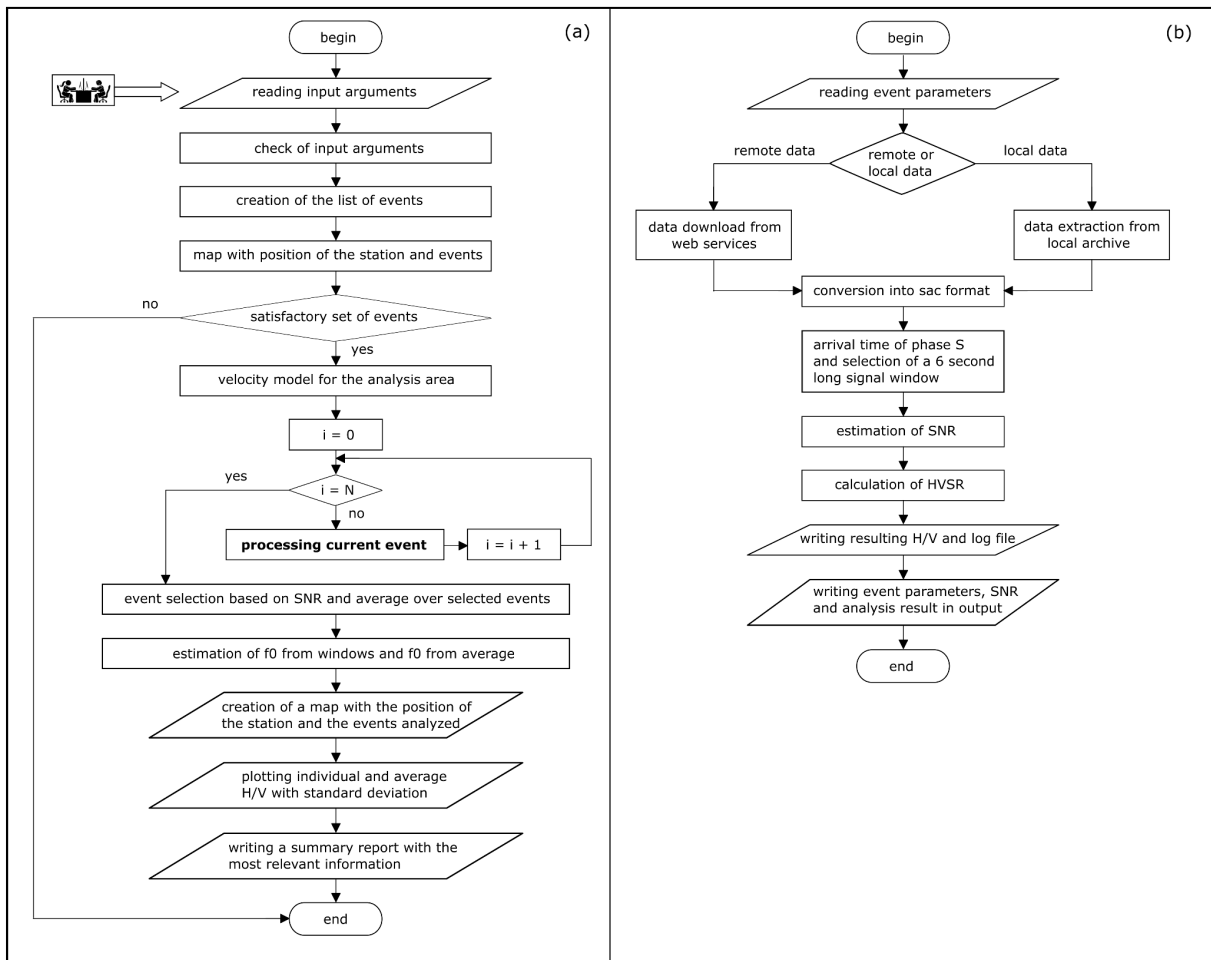


Figure 3. Simplified flow chart of HV\_Eqk.sh (a) and detail describing the processing of a single event (b), corresponding to the block labelled "processing current event" formatted in bold in (a).

In case of a negative answer (when the selection of events is not satisfactory), the script stops. In case of positive answer, the spectral ratio for each individual event is calculated on a signal window of 6 seconds starting from the theoretical arrival of the S-wave phase according to the seismic ray theory (Figure 3b). The script also computes the signal-to-noise ratio (SNR), defined as the ratio between the maximum amplitude in the selected time-window and the rms value on a noise time window preceding the origin time of the event. Finally, the geometric mean and standard deviation are computed on selected spectral ratios having SNR larger than a fixed threshold (Figure 3a).

For the estimation of the S-wave arrival times, the script *HV\_Eqk.sh* relies on the classical theory of the seismic ray for 1D plane-parallel subsoil model (Červený 2005). The 1D model corresponding to the area of interest, defined as the minimum rectangular area containing all the selected events and the station, is built by averaging a 3D  $V_s$  velocity model provided by

the user. For Italy, we used the Italian 3D model by Di Stefano & Ciaccio (2014), updated in 2017 with the inversion of the layer at 0 km (Di Stefano, personal communication); its resolution limit in depth is 60 km and, for larger depths, the 1D IASP91 model is used. For each event, the theoretical arrival time of S-wave phase is then estimated with the *cake* program available in the pyrocko package (Heimann et al. 2017), by means of the build 1D model and the coordinates of the hypocenter and of the station.

We consider the automatic picking accuracy of S-waves good enough for a quick and massive HV analysis (see Figure 4 as an example): after the automatic picking of S-waves (Fig. 4a), the waveform is selected and cut (Fig. 4b), pre-processed with a removal of the mean and linear trend followed by a tapering (Fig. 4c), and finally zero padded to obtain a fixed window length of 60 seconds that gives a suitable number of points for Geopsy's spectral processing (Fig. 4d).

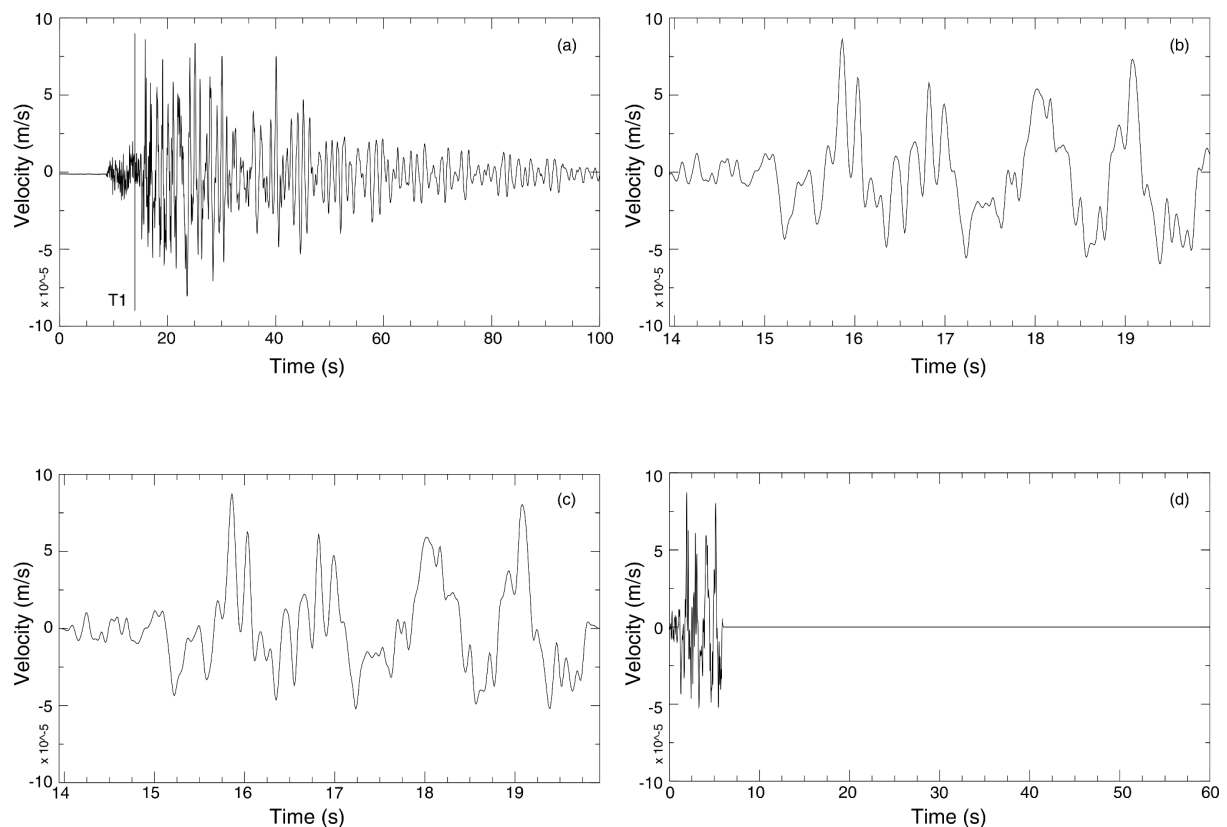


Figure 4. Example of automatic picking of S phase for the E-W component of station IV.INTR. The event occurred on 2010-07-04 11:38:58.260 (ML 3.7) with the following hypocentral parameters: latitude 42.177, longitude 13.4, depth 9 km. (a) initial time window, with indication of the theoretical arrival time  $T_1$  of the S phase automatically computed (13.94 s from the origin time of the event); (b) time window obtained by selecting 6 seconds starting from  $T_1$ ; (c) waveform after removing mean and linear trend, and applying a 5% width taper; (d) final waveform after zero-fill to get the 60 seconds window length.

As example, the command line

```
HV_Eqk.sh -n IV -s INTR -c HH -f 2010-01-01 -t 2012-12-31 -m 3 -M 7 -R 100 -o OUT/INTR_EQHH
```

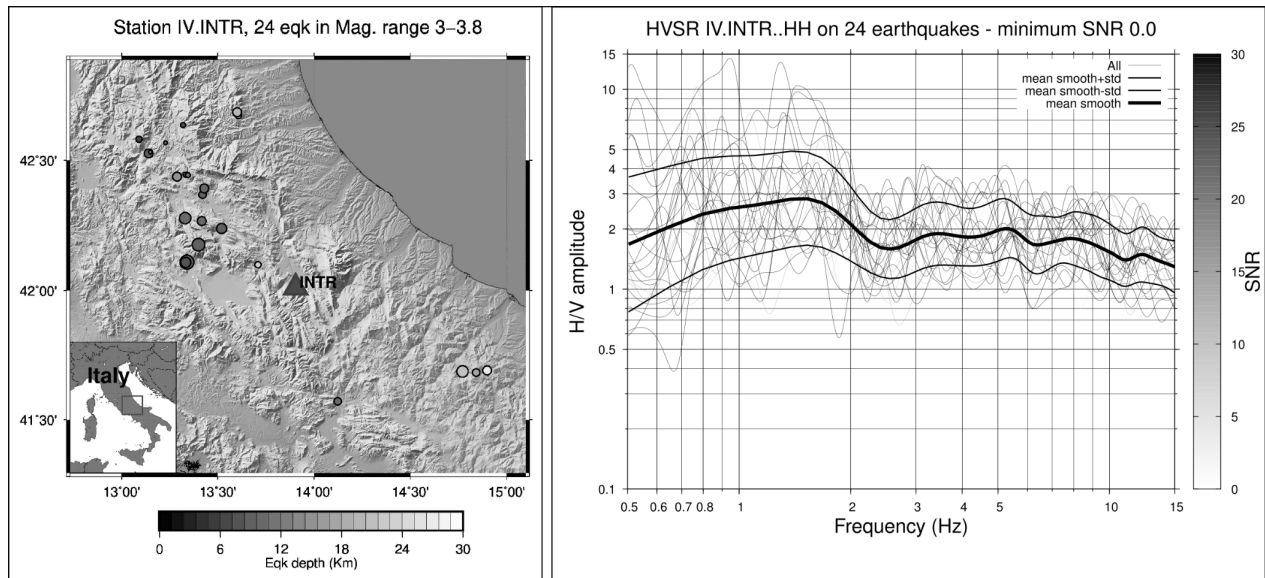


Figure 5. *HV\_Eqk.sh* for station IV.INTR (channels HH), obtained by using a time interval from 2010-01-01 to 2012-12-31, minimum magnitude 3, and maximum distance from the station 100 km. (Left) map with the distribution of selected earthquakes (circles) around the target station (triangle), with the scale proportional to the event depth. (Right) HV curves for each selected event with scale proportional to the SNR value, together with the smoothed average curve  $\pm 1$  standard deviation (thick curve).

gives the results shown in Figure 5, obtained for the stream HH (velocimeter) of the station IV.INTR, selecting events on a time interval from 2010-01-01 to 2012-12-31 with magnitude in the range 3 to 7 and maximum distance of 100 km from the target station (see *HVNEA* user manual in Supplemental Material for details). The output figures display the map with the distribution of selected earthquakes for the target station and the HV curves of the selected events, with colour scale proportional to the signal-to-noise ratio (SNR), together with their smoothed average and standard deviation.

In addition to the figures, the script also provides in the output directory:

- a text file containing the event parameters (origin time, Latitude, Longitude, magnitude...), the station specification (i.e. network, station, location, stream), the calculated SNR and a binary flag indicating whether the analysis was successful or not.

- a cumulative text file containing all the individual HV curves.
- a subdirectory named “SNR-<snr>-<n>” (where <snr> is the chosen minimum SNR and n is a sequential number) containing two text files, analogous to those present in the main output directory, taking into account only the events that are above the SNR threshold, the two images shown in Figure 5, a synthetic report containing all the relevant information concerning the performed analysis.

The complete output is described in the *HVNEA* user manual in Supplemental Material.

### 2.3 Auxiliary scripts

*HVNEA* package provides six auxiliary scripts for post-processing, results recovery and results merging for earthquake (*hv\_eqk\_filter.sh*, *hv\_eqk\_recovery.sh*, *hv\_eqk\_merge.sh*) and noise analysis (*hv\_noise\_filter.sh*, *hv\_noise\_recovery.sh*, *hv\_noise\_merge.sh*).

The script *hv\_eqk\_filter.sh* is dedicated to earthquake analysis and it allows to select a-posteriori the events to be averaged, by changing SNR threshold, range of amplitude, magnitude, distance, depth and back azimuth. It is also possible to reduce the analysis to defined time intervals within the original event list. Similarly, the script *hv\_noise\_filter.sh* allows to refine an analysis taking into account only days that fall within a chosen time interval and with HV amplitude in a range defined by the user.

The two auxiliary scripts *hv\_eqk\_recovery.sh* and *hv\_noise\_recovery.sh* are able to finalise the analysis of *HV\_Eqk.sh* and *HV\_Noise.sh* in case of unexpected interruption of the script execution, taking into account the already computed individual HV curves.

Finally, *hv\_eqk\_merge.sh* and *hv\_noise\_merge.sh* join several output directories which are already produced by the scripts. The directories to be merged can refer to different analysis periods, or, in case of earthquakes, may be obtained with different parameters for selection of events; for these two scripts the output directory must be specified by the user. The scripts perform basic network, station, and stream consistency checks between input folders before performing their merge. See also *HVNEA* user manual in Supplemental Materials for further details.

### 3. Application to permanent seismic stations

To test the potentiality of proposed procedures, we ran the *HVNEA* scripts on more than 15 years of data acquired by two stations of the Italian Seismic Network (RSN) maintained by INGV (network code IV; Margheriti et al., 2021): IV.INTR and IV.MRLC. All data acquired from RSN stations are archived and distributed in EIDA (European Integrated Data Archive) and are accessible via the INGV data center. IV.INTR and IV.MRLC are equipped with accelerometer and velocimeter sensors and can be considered as representative of the stations of the Italian network. Both of them have been operating for approximately 20 years, and are located in seismically active regions of the national territory, that is Central (IV.INTR) and Southern (IV.MRLC) Apennines mountain chain, respectively (Chiarabba et al., 2009; Festa et al., 2021).

Station book of both stations and geological and geophysical information on site characterization can be found in ESM-Engineering Strong Motion seismic archive (Luzi et al. 2020), in ITACA-Italian Accelerometric Archive (devoted to Italian and European accelerometric stations), and in CRISP archive (<http://crisp.ingv.it/>), dedicated to all the permanent stations of the RSN network (Cultrera et al. 2022).

The IV.INTR station is equipped with a broad-band velocimeter (Nanometrics Trillium 40s; stream HH) and an accelerometer (FBA Kinematics Episensor; stream HN) connected to a GAIA2 24-bit digitizer (Rao et al. 2010). It has been operating since 2003-09-03 and its station book is available on CRISP and on both ESM and ITACA together with the collection of the earthquake records. Moreover, in the framework of a national project on site characterization of the accelerometric stations belonging to the Italian networks, IV.INTR was recently investigated by geological and geophysical surveys (Working group INGV, 2020) revealing a clear resonance peak at about 1.3 Hz and polarised along a N-S direction, which is found at least for 200 m in the area surrounding the station.

The IV.MRLC station is equipped with sensors of the same type as the previous site, but connected to a Nanometrics Trident 24-bit digitizer. It has been operating since 2003-03-01 and its station book is available on CRISP and on both ESM and ITACA.

We analysed the continuous recordings at both stations with the *HVNEA* package over their entire operational period (from 2005-01-04 to 2021-11-25 for IV.INTR and from 2005-03-25

to 2021-12-30 for IV.MRLC), by using both HH and HN streams for earthquake analysis and HH only for noise. Figure 6 shows the earthquake and noise HV curves with their corresponding standard deviations for IV.INTR and IV.MRLC obtained with the *HVNEA* package compared with the HV spectral ratios already available in ESM/ITACA and CRISP archives, comparison comments are given in the next section.

### 3.1 HV earthquake analysis

For HV from earthquakes we analysed two groups of event signals: one with magnitude smaller than 3 within a maximum distance of 100 km, the other with magnitude in range 3.1-6.1 within a distance of 385 km for IV.INTR and magnitude in 3.0-6.5 within a distance of 310 km for IV.MRLC. The magnitude and distance selection is the same of the ones used for the HV computation in ESM/ITACA archives, with the aim of facilitating the comparison. The results for the analyzes on events in 3.1-6.1 range are shown in Fig 6, while results from smaller magnitude events are shown in the figures in the appendix.

Figures 6a and 6b show the HV of S-phase acceleration Fourier spectra from ESM archive, compared with *HVNEA* earthquake spectral ratios with SNR=30, which allows us to select the events having high signal level with respect to background noise. *HVNEA* tool selected 369 and 172 events on HH stream for IV.INTR and IV.MRLC respectively, whereas ESM used 75 and 47 events for the same stations; the selection on HN stream selected a lower number of events. A good correspondence is observed for both velocimetric and accelerometric streams, although the results obtained with the *HVNEA* analysis include a higher number of records because they include the events with low magnitude, not recorded by the accelerometric stations available in ESM/ITACA. The standard deviation associated with our continuous analysis is slightly greater than S-phase curves from the ESM archive; this could be explained by the large amount of data selected automatically by *HVNEA* compared to the number of good-quality earthquakes manually selected in ESM.

As previously mentioned the *HVNEA* package provides the auxiliary script *hv\_eqk\_filter.sh* for applying various selection criteria to earthquakes. In particular, we performed a test by varying the ranges of magnitude and SNR in order to evaluate the different contributions in the obtained HV curve. This test was also performed to provide useful indications to the users on the choice of SNR and minimum magnitude for obtaining reliable results. Table A1 shows for both stations the numbers of earthquakes automatically selected by gradually



increasing the minimum value of the magnitude and SNR threshold. The obtained results show, within their standard deviations, comparable trends and amplitudes of ESM curves for all the selected magnitude intervals, with a slightly better overlap for the higher magnitude and SNR values (see Figure A1a,b,c,d, and Figure A2a,b,c,d in Appendix). The comparison of the curves obtained with the contribution of only small events ( $M=0-3$ ,  $M=0-4$ ) and  $SNR=30$  still show comparable results with the curves obtained with the largest events (Figure A1e,f and Figure A2e,f in Appendix).

The results of this test show that the automatic analysis of selected events with a high SNR (SNR higher than 10 or 20) ratio provides robust results also on small-magnitude events (events with magnitude smaller than 3 or 4). This is most likely due to the large number of small-magnitude earthquakes (from several hundred to a few thousand) used in the final HV spectral ratio estimate, which provides HV spectral ratios very close to those obtained from a manual selection of high-magnitude events. This result could be particularly useful for the site characterization of stations located in areas where there are no energetic earthquakes instrumental records (earthquakes with magnitude greater than 3 or 4). In this case, using *HVNEA*, a robust estimate of the HV spectral ratio could be obtained by extending the analysis to all low-energy earthquakes located near the analysed station.

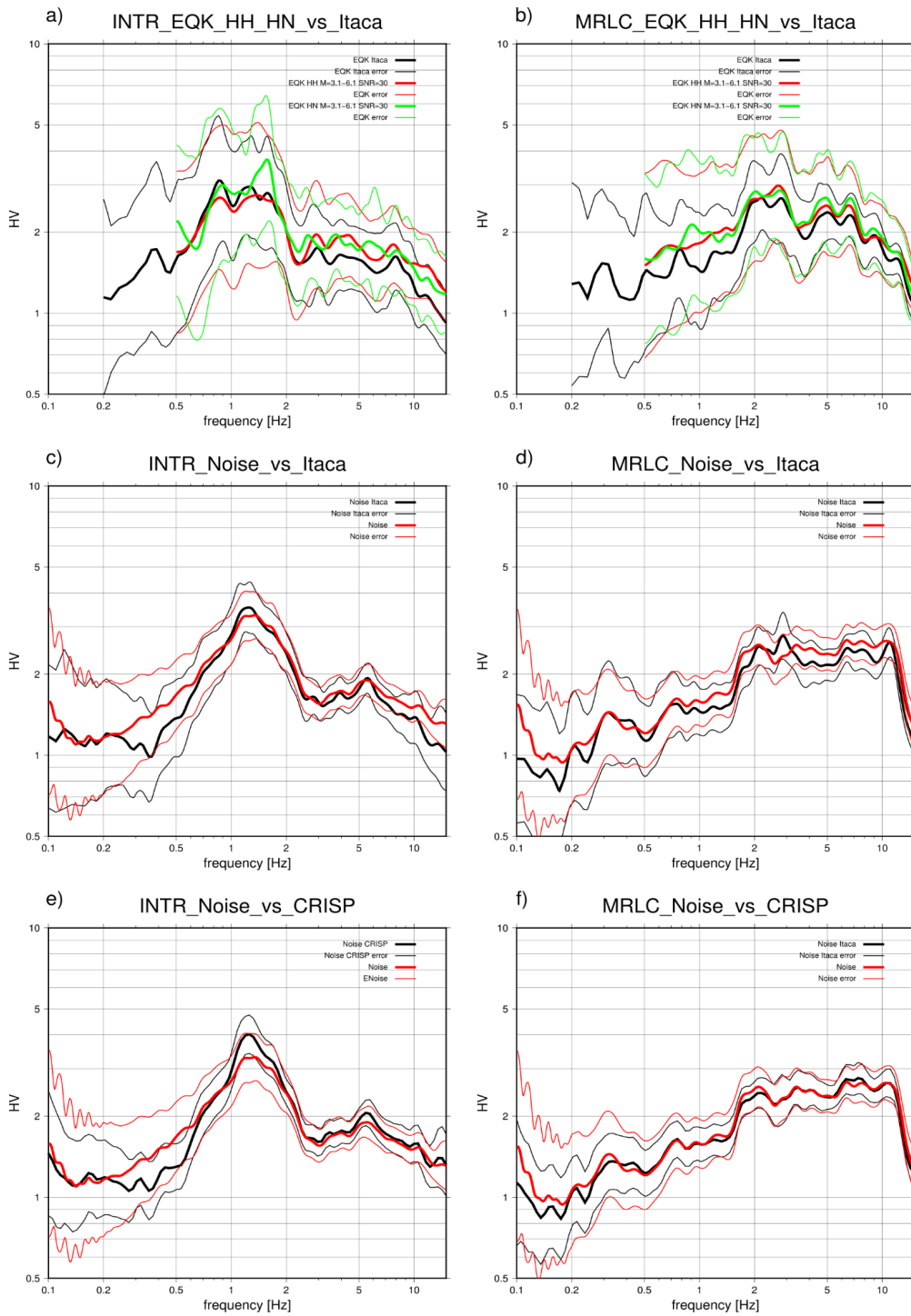


Figure 6. HV for IV.INTR and IV.MRLC: Comparison between *HVNEA* analysis on earthquake (a and b) and noise with HV results published on ESM archive (c and d) or on CRISP database (e and f). Black lines represent ESM or CRISP results. Red lines represent the output of *HVNEA* analysis on HH stream using noise and earthquakes while green lines show the *HVNEA* results performed on HN stream by using only the earthquakes. For both the stations, *HVNEA* earthquake curves refer to SNR=30. Each curve is represented with its standard deviation and in the frequency range allowed by the methodology.

### 3.2 HV noise analysis

The HV noise curves were calculated with the *HVNEA* package using about 14 years of recordings for IV.INTR and about 17 years for IV.MRLC. The results show a good correspondence of shape and amplitude with HV noise curves from ESM (Figures 6c and 6d) and CRISP (Figures 6e and 6f), that compute them on a few hours of recording. The HV noise on ESM uses a noise record length of 90 min for both stations; moreover, it computes a vectorial sum of horizontal components that was divided by  $\sqrt{2}$  to be compared with our results. Instead, the noise curves of CRISP are calculated with a similar procedure of *HVNEA* (using Geopsy) but on a 24 h signal length (May 31, 2014 for INTR and May 31, 2015 for MRLC). The noise curves from CRISP have a greater similarity with our results on the MRLC station than on the INTR station. This can be explained due to the large time length used in Crisp (24 hours versus few hours of ESM).

The above comparison suggests that the analysis over decades is likely redundant if the final aim is to get an average HV curve. At the two selected stations, the average HV curves obtained over years are very similar to the one computed using a limited number of hours (from Figures 6c to 6f). However, the evolution of HV spectral ratios over time (Figures 7b and 7d) emphasises several temporal variations involving fluctuations of both amplitude and peak frequencies that could provide important information on the elastic properties of the shallow layers as well as on the mechanisms that can produce the observed temporal variations.

### 3.3 Coherence between air temperature and frequency peaks from noise

Investigating such temporal variations can be of great interest in the seismological community to better understand the origin of the HV peaks and their relation with other physical parameters (like the temperature, pressure, water table and so on). Variations of HV peaks versus time are usually performed in fairly peculiar contexts as studies of permafrost layer, rock masses, sites characterised by topographic effects (Köhler and Weidle 2018; Kula et al., 2018; Colombero et. al, 2018; La Rocca et al. 2020). Only recently, studies performed by more than 10 years of continuously acquired seismological data from permanent seismic stations in Central Apennine (Italy; Vassallo et al. 2022), have shown the presence of

seasonal variations and peak frequency changes at occurrence of strong earthquakes in the spectral HV analysis.

For both stations analysed in this work, and in several frequency ranges, HVNEA results show trends of peak frequency and peak amplitude with temporal variations that appear to be characterised by yearly cyclical patterns.

With the aim to better characterise such variations observed at IV.INTR and IV.MRLC, we extracted the temporal evolution of the main peak frequency and associated amplitude by selecting HV values from different frequency ranges: [0.9-1.9 Hz], [4.9-6.1 Hz], and [9.9-11.9 Hz] for IV.INTR (indicated as 1, 2 and 3, respectively, in Figure 7d); [0.27-0.38 Hz], [2.0-2.5 Hz], [7.2-8.8 Hz] and [9.5-12 Hz] for IV.MRLC (indicated as 1, 2, 3 and 4, respectively, in Figure 7b). We selected the amplitude peaks mainly on the basis of the analysis summarized in figure 7, selecting peaks that showed cyclic variations and trying as much as possible to isolate HV peaks with similar temporal variations. We did not always take the criteria of the SESAME guidelines into account when selecting HV peaks, e.g. we selected several spectral peaks characterized by amplitude lesser than 2 as they appear to show a clear cyclic trend over time. The trend of amplitudes and frequency peaks versus time are shown in Figures 8 and A3. For the INTR station, in the band between 4.9-6.1 Hz, there are two frequencies associated with the maximum HV peak, each of which is predominant on different days in the analyzed time window. The first one frequency peak appears to be between 5 and 5.4 Hz, and the second between 5.5 and 6 Hz. Such a distinction was impossible to highlight from the simple plot of the HV trend over time in Figure 7. This observation further emphasizes the role that the frequency extraction analysis can also play in distinguishing the various spectral contributions defining a single HV spectral ratio peak.

For the two stations the amplitude and frequency of the peak extracted show a yearly cyclical character that we compare with seasonal meteorological variations, in particular with the temperature measured at meteorological stations. The idea is to use the recorded air temperature as a proxy for the seasonal weather variation at the investigated sites, and to study the possible correlation with the temporal variability of the different HV peaks extracted.

We compared changes in amplitudes and peak frequencies with the daily average of air temperature acquired by two meteorological stations installed close to the seismic ones:

BELLA (Bella, Potenza; Lat. 40°46'6"N, Lon. 15°32'11"E, Alt. 693m) at about 6 km from IV.MRLC, and AQ01 (Coppito, L'Aquila; Lat. 42° 22' 00" N, Lon. 13° 21' 00" E, Alt. 680m) at about 60 km from IV.INTR. Although far from the seismological station, AQ01 recorded for a larger time span than the closest weather station SIAP (Anversa degli Abruzzi, L'Aquila, at about 9 km from IV.INTR), whose temperature values are very similar in the common recording period (Figure 8a). We selected the time span where both air temperature (T) and seismological parameters (peak frequency,  $f$ , or its amplitude, A) are available; the sparse missing values, if any, were substituted with the previous closest measured value. We then computed the correlation between the two signals as magnitude-squared wavelet coherence in the time-frequency plane using the *wcoherence* Matlab function. Let's consider 2 time series, one representing temperature data (x) and the other the HV frequency peak or its amplitude (y).

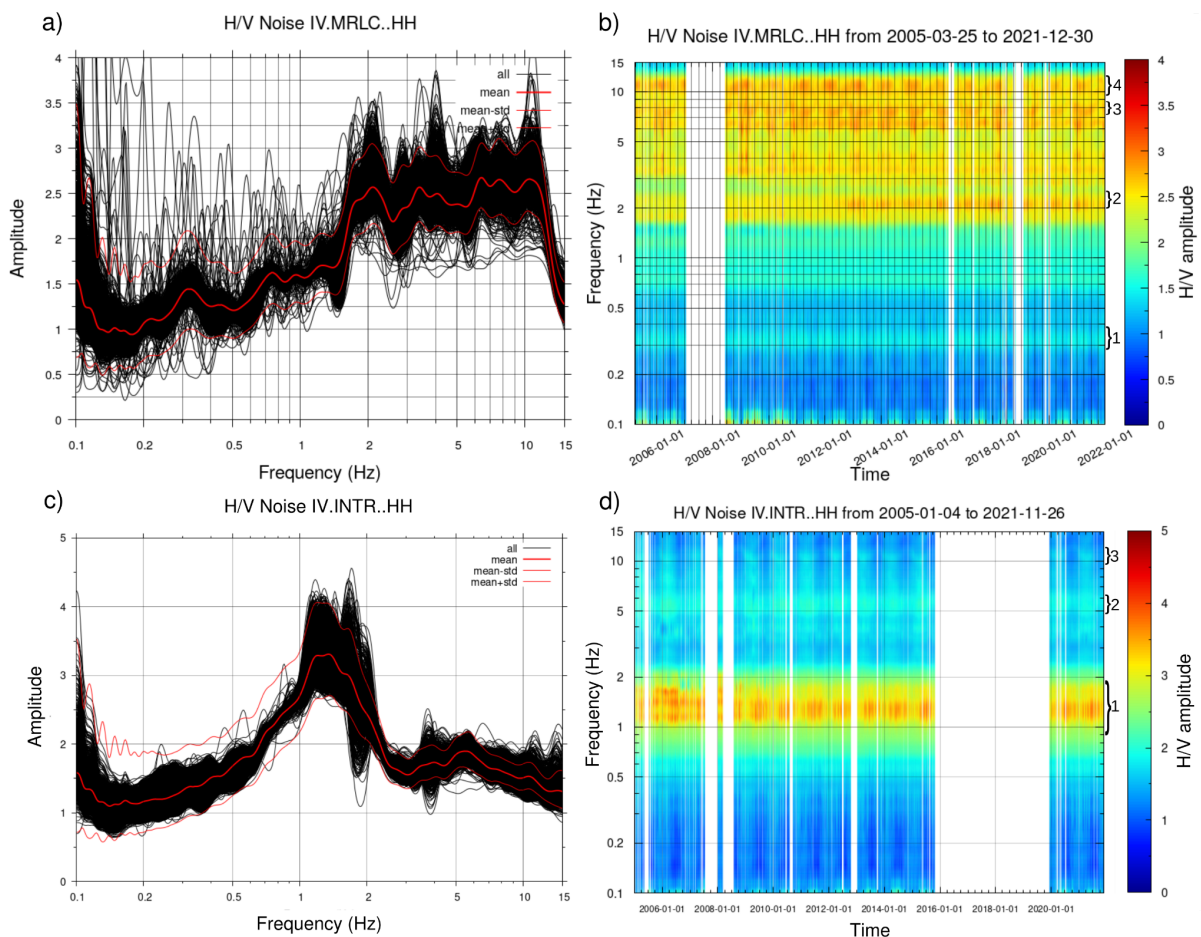


Figure 7. HV noise analysis performed for IV.INTR (a and b) and IV.MRLC (c and d) stations. The analysis was performed using all the data acquired by seismic stations starting from 2005 to the end of 2021. The white

rectangles in the contour plot (b and d) indicate data gaps or instrumental malfunctions. The numbered brackets on the right side of the contour plots (b and d) indicate the frequency ranges used to analyse the temporal variability of the amplitudes and peak frequencies.

The magnitude-squared wavelet coherence (MSWC) is defined as:

$$MSWC = |S(C_x^* \cdot C_y)|^2 / [S(|C_x|^2) \cdot S(|C_y|^2)],$$

where  $C_x$  and  $C_y$  are the continuous wavelet transforms of  $x$  and  $y$  by using the analytic Morlet wavelet over logarithmic scales, the asterisk denotes the conjugation operator, and  $S$  is a smoothing operator in time and scale (number of scales to smooth = 12, number of voices per octave = 12). The effect of the smoothing operator is such that the increased smoothing causes reduced low frequency resolution, whereas the number of voices per octave controls the discretization (see Vassallo et al. 2022 for further details).

We performed the coherence analysis for the air temperature and the HV parameters (frequency peak and related amplitude) extracted in different frequency ranges, as shown in Figure 8 for IV.INTR. Results for IV.MRLC are displayed in Figure A3.

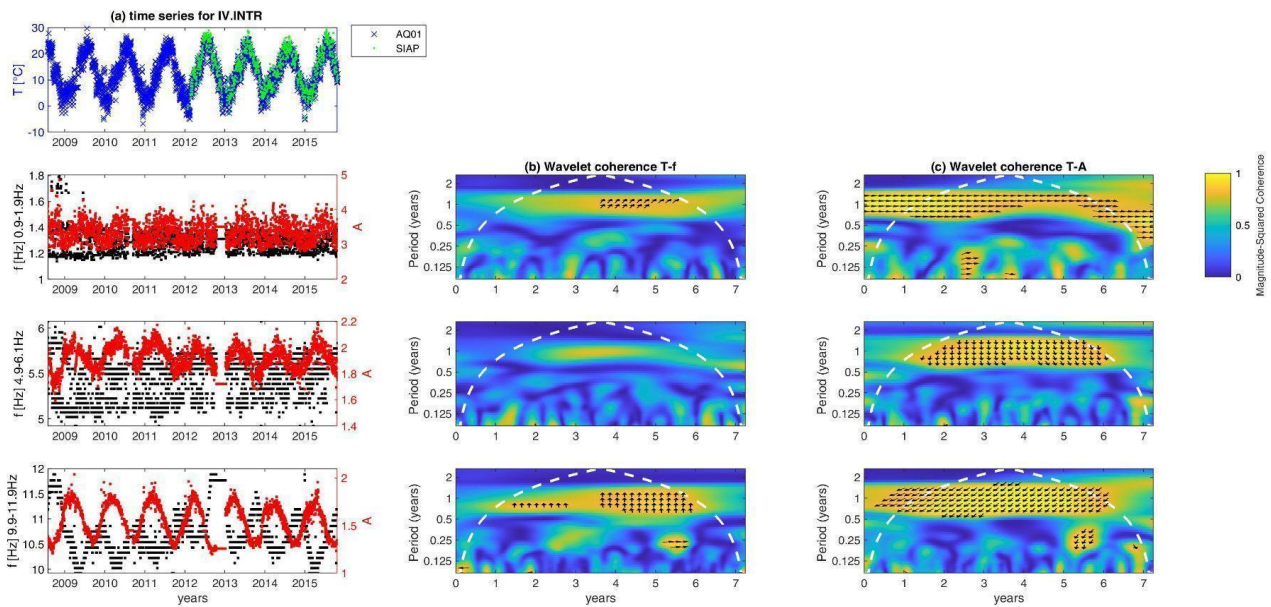


Figure 8. Correlation between temperature measured at AQ01 ( $T$ ) and HV parameters for IV.INTR (peak frequencies,  $f$ , and its amplitude,  $A$ ) for the period 30/07/2008-27/10/2015. (a) From top to bottom: time series of temperature at AQ01 and SIAP weather stations; temporal variation of different peak frequencies (black dots, left axis) and their amplitudes (red dots, right axis): 0.9-1.9 Hz, 4.9-6.1 Hz, 9.9-11.9 Hz. (b) Magnitude-squared wavelet coherence between  $T$  and  $f$  in the same frequency range: arrows indicate the phase lag of  $f$  with respect to  $T$  only where the coherence is greater than or equal to 0.8. The direction of the arrows corresponds to the

phase lag on the unit circle, which is in this case 1 year (for example, an E direction is 0 lag whereas a vertical arrow indicates a quarter-cycle phase lag). White dashed line is the cone of influence for the wavelet coherence (in cycles per unit time), i.e. areas outside or overlapping the cone of influence are less reliable. (c) Magnitude-squared wavelet coherence between T and A, i.e. the same as (b) but with amplitude in place of frequency.

For IV.INTR we selected the 3 frequency ranges which include the main peaks of Figure 7d: 0.9-1.9 Hz (resonance frequency), 4.9-6.1 Hz and 9.9-11.9 Hz. Figure 8 shows that frequency peaks (f) and their amplitudes (A) are well correlated with the seasonal variation of air temperature (T) in all the investigated HV frequency ranges. Indeed, we observe that the time series of temperature T versus both f and A have a coherence larger than 0.8 at a period of about 1 year, except for the correlation of 4.9-6.1 Hz whose coherence is slightly lower. The corresponding phase-lag is different between f and A and varies for the different frequency ranges: the positive shift of the first and third frequency peak respect to temperature is around 19%-25% of one-year-cycle (70-90 days), whereas the negative shift of amplitude respect to temperature increases with the frequency: -4% (first peak), -24% (second) and -33% (third) of one-year-cycle, that is -15, -70 and -120 days respectively.

The coherence analysis for IV.MRLC has been performed at 4 frequency ranges which include the main peaks of Figure 7d: 0.27-0.38 Hz, 2-2.5 Hz, 7.2-8.8 Hz, 9.5-12 Hz. Figure A3 shows a clear correlation with the seasonal variation of air temperature (T) for some of the investigated HV frequency ranges at a period of about 1 year: the peak frequencies have a coherence value larger than 0.8 only for the high frequencies [7.2-8.8 Hz] and [9.5-12 Hz], whereas the amplitudes show the same correlation also for the lower frequency ranges. The corresponding phase-lag is highly variable, with a shift of frequency peak with respect to temperature around 14% and -44% of one-year-cycle (50 and -160 days) for the third and fourth peak. The amplitudes show a comparable negative shift with respect to temperature for the second, third and fourth peak (-170, -125, -150 days respectively, which correspond to -47%, -34% and -41% of one-year-cycle), whereas the lowest peak has a positive shift of 170 days (46% of one-year-cycle).

In synthesis, at both seismic stations we observe a correlation of the frequency-peak variability in time with seasonal changes measured by air temperature. We cannot exclude correlation at smaller periods (days, weeks) that are probably more sensitive to the distance between weather and seismic stations. A similar behaviour has been already observed by

other authors (see Vassallo et al., 2022, and references therein) and it seems to be related to the changes in the soil itself (e.g. changing water level, etc.) rather than on large scale effects such as the moving of the microseismic noise source over the year.

## 4. Conclusions and perspectives

The *HVNEA* scripts presented in this paper are designed to work on very long continuous data (months to years) to easily compute the HV ratios from earthquake and noise. In terms of computational performances, the time needed to run the scripts mainly depends on the data download speed which is strongly conditioned by the internet connection of local PCs and by the capabilities of websites distributing seismic data. To give an indication on the computational time of *HVNEA* procedure, to run HV noise analysis on a whole year of data (one stream with data downloaded from EIDA) the script needs about 25 min, whereas it takes 65 min to select and analyse 1066 events for HV earthquake on a pc with a good network connection.

More specifically, the developed procedures can be used for the following purposes:

- automatic analysis of large amounts of seismic data, and thus providing robust results in relatively short time, which could hardly be obtained through manual or semi-manual analyses;
- individuation of the resonance frequency ( $f_0$ ) of the investigated site by analysing long-time data, and by comparing the HV ratios from ambient noise and earthquake data;
- study of the temporal variations of HV on the ambient seismic noise to evaluate the stability of the spectral ratios (and therefore of  $f_0$ ) at a single station over an extended period of time;
- Correlation analysis between the variations in frequency and amplitude of the HV peaks with some other physical parameters (as for the air temperature used as proxy of seasonality);
- calculation of the HV spectral ratios on earthquakes recorded from seismic networks even located in low-seismicity or low-magnitude areas;
- analysis of the behaviour of the HV earthquake ratios during seismic sequences.



The analysis using *HVNEA* tool can help in a better characterization of the seismic stations, as the shape of the HV curve and the resonance frequency ( $f_0$ ) can be an important proxy in predicting the seismic site response (Ghofrani and Atkinson 2014; Kwak et al. 2017; Kawase et al., 2019; Felicetta et al. 2021). The site characterization of the permanent stations belonging to seismic networks has recently been investigated by several authors (Hollender et al. 2018; Cultrera et al. 2021; Di Giulio et al. 2021; Hobiger et al. 2021; Mital et al. 2021) but only a few seismological national networks expose site condition characteristics with very detailed information on site topography, morphology, geology and geophysics (Cornou and Bard 2019). For example the ESM-Engineering Strong Motion database or the ITACA-Italian Accelerometric Archive display information on geological and geophysical data, when available. In the Italian context, the site-specific geophysical and geological analyses are stored in the CRISP archive, dedicated to site characterization of the stations for the Italian seismic network run by INGV.

Unfortunately, the site classification of a seismic permanent station is often limited to the estimate of EC8 soil class, which is derived from the  $V_{S30}$  (average travel-time velocity in the uppermost 30 m) or  $V_{s,eq}$  (average travel-time velocity down to the substrate having  $V_s \geq 800$  m/s and depth smaller than 30m) in case of the Italian building code (NTC18 2018). A massive computation of HV on the whole seismic network, together with the  $f_0$  estimation, can increase the level of information on site response and possible local effects. In addition, recent studies (e.g. Bergamo et al. 2021, Zhu et al. 2020b) indicate that a combination of both  $f_0$  and  $V_{S30}$  give a more complete information on site response. HV curves can be also inverted under specific assumptions and used as constraints for deriving the shear-wave velocity profile (Fäh et al. 2003; Arai and Tokimatsu 2004; Molnar et al. 2018; García-Jerez et al. 2016; Kawase et al. 2019).

Furthermore, a very interesting topic is the investigation of the temporal variations of the HV ratios; temporal variations that can be easily evidenced by the *HVNEA* analysis. The behaviour of the noise fluctuations as a function of the weather conditions or anthropic activities is well-documented in literature especially for noise recordings (Ringdal and Bungum 1977; Bonnefoy-Claudet et al. 2006; Bottelin et al. 2013; Hillers and Ben-Zion 2011). More recently, sharp variation in the HV curves (and therefore of the  $f_0$  value) at a single station are observed during strong earthquakes and interpreted in correlation to soil nonlinearity effects of the shallow subsoil (Bonilla et al. 2021; Vassallo et al. 2022). These

fluctuations during seismic sequence are very similar to the ones derived from interferometric techniques based on autocorrelation of seismic signals (Minato et al. 2012). Probably HV analysis over time and interferometric results could provide similar results when applied to the same frequency bands. This can be observed when new analyzes will be carried out by comparing the results from the two methodologies for the same site.

Several parameters used in HVNEA were fixed to provide robust results on the Italian national seismic network data. For example, the duration of window for noise analysis (daily based) and the length of the window around the S phase for the earthquake analysis (6 s) were selected to work on databases of continuous recordings composed by many years of data and on earthquakes at distances up to a few hundred km from the investigated station and when a reasonably accurate 1D Vs velocity model is available for the area. When these criteria are verified, such parameters remain valid for obtaining robust HV spectral ratios regardless of the investigated area in which both station and earthquake fall. We are planning to modify the HVNEA package by inserting a user-editable parameter that will enable the user to select the analysis window that best suited their purposes.

Our next effort will be to include the results of *HVNEA* analysis in the Italian national seismic databases, aimed at providing more complete information on the site in terms of  $f_0$  and HV curves. Finally, *HVNEA* can be upgraded without too much effort for other types of common analysis, for example investigation of the single component spectra or directional HV ratios, or the spectral ratios using a reference site.

## 5. Data and resources

The *HVNEA* package is available from the link: <https://github.com/INGV/hvnea>

A user Manual for HVNEA software is available in the Supplemental Material of this article. *HVNEA* can work on seismic networks whose data are distributed through FDSN web services (<https://www.fdsn.org/webservices/>, last accessed on 14 April 2022). The proposed *HVNEA* scripts use FDSN datacenters (<https://www.fdsn.org/datacenters/>, last accessed on 14 April 2022) for downloading seismological data, creating earthquake catalogues (<https://www.fdsn.org/webservices/fdsnws-event-1.2.pdf>, last accessed on 14 April 2022) and obtaining information on seismic stations (<https://www.fdsn.org/webservices/fdsnws-station-1.1.pdf>, last accessed on 14 April 2022).

When the procedures work on seismic data locally available, the data must be filed according to the Seedlink Data Structure (<https://www.seiscomp3.org/doc/applications/slarchive/SDS.html>, last accessed on 14 April 2022), and it is necessary to provide the dataless files ([http://www.fdsn.org/pdf/SEEDManual\\_V2.4.pdf](http://www.fdsn.org/pdf/SEEDManual_V2.4.pdf), last accessed on 14 April 2022).

The two seismic stations belong to the Italian permanent network (hereinafter RSN; INGV Seismological Data Centre, 2006, Margheriti et al. 2021) operated by the Istituto Nazionale di Geofisica e Vulcanologia (INGV). Their data were downloaded from the European Integrated Data Archive (EIDA, <https://eida.ingv.it/en/>, <http://www.orfeus-eu.org/data/eida/> last accessed on 14 April 2022; Strollo et al. 2021), using the INGV data center (<https://www.fdsn.org/datacenters/detail/INGV/>, last accessed on 14 April 2022)

To analyze the earthquakes data at the two seismic stations, we used the Italian 3D model by Di Stefano & Ciaccio (2014), <https://gitlab.rm.ingv.it/raffaele.distefano/seismic-tomography-italy-2017-di-stefano-and-ciaccio/>

The comparison of the results obtained using *HVNEA* with those previously existing, was carried out by retrieving the HVs available for the stations IV.INTR and IV.MRLC from ESM-Engineering Strong Motion database (<https://esm-db.eu>, last accessed on 14 April 2022), ITACA database (D'Amico et al. 2020, <http://itaca.mi.ingv.it>, last accessed on 14 April 2022) and CRISP archive (<http://crisp.ingv.it>, last accessed on 14 April 2022)

Temperature data are provided by different Institutions: the Center of Excellence Telesensing of Environment and Model Prediction of Severe events (CETEMPS; <http://cetemps.aquila.infn.it/>, last accessed on 14 April 2022) for AQ01 weather station, the Territory - Environment Department of the Abruzzo region (Civil Protection Activities Planning Service, Hydrographic and Oceanographic Office) for SIAP (<https://www.regione.abruzzo.it/content/servizio-idrografico-mareografico>, last accessed on 14 April 2022), the "Functional Center for Forecasting meteorological and monitoring hydro pluviometric weather and landslides" of the Campania Region for BELLA (<http://centrofunzionale.regione.campania.it/>, last accessed on 14 April 2022). The data at both meteorological stations are raw and the providers are relieved of all responsibility for their use.

The programs proposed in this work are based on extensive use of third-party softwares required for both data analysis and plots creation. The main software are listed below:

- Geopsy 3.4.1 (<http://www.geopsy.org/>, last accessed on 14 April 2022) for HV spectral ratios computation;
- Seismic Analysis Code 101.6a (SAC, Goldstein et al., 2003) for data analysis (<http://ds.iris.edu/ds/nodes/dmc/forms/sac/>, last accessed on 14 April 2022);
- GSAC, an open source seismic analysis code ([https://www.eas.slu.edu/eqc/eqc\\_cps/CPS/CPS330/cps330g.pdf](https://www.eas.slu.edu/eqc/eqc_cps/CPS/CPS330/cps330g.pdf)), part of Computer Programs in Seismology ([https://www.eas.slu.edu/eqc/eqc\\_cps/CPS/CPS330.html](https://www.eas.slu.edu/eqc/eqc_cps/CPS/CPS330.html));
- Pyrocko (<https://git.pyrocko.org/pyrocko/pyrocko.git>, last accessed on 14 April 2022) for computation of S-phase travel-times using ray theory;
- Generic Mapping Tool (GMT) 5.4.5 for some plots and for data analysis (<https://github.com/GenericMappingTools/gmt/releases/tag/5.4.5>, last accessed on 14 April 2022);
- Gnuplot 5.2 for the creation of figures (<https://sourceforge.net/projects/gnuplot/files/gnuplot/5.2.8/>, last accessed on 14 April 2022)

## Acknowledgements

This study has benefited from funding provided by the Italian Presidenza del Consiglio dei Ministri – Dipartimento della Protezione Civile (DPC), DPC-INGV 2019-21 Agreement - Annex B2 - WP1-TASK 2 “Accelerometric site characterization”. This paper does not necessarily represent DPC official opinion and policies.

We thank professor R. B. Herrmann, Department of Earth and Atmospheric Sciences Saint Louis University, who kindly allowed us to use GSAC software.

We thank Chiara Felicetta for sharing the analysis stored in ESM.

## References

- C. Anderson, (2015). Docker [Software engineering]. *IEEE Software*, vol. 32, no. 3, pp. 102-c3, doi: 10.1109/MS.2015.62.
- Arai, H. and Tokimatsu, K., (2004). S-wave velocity profiling by inversion of microtremor HV spectrum. *Bulletin of the Seismological Society of America*, 94(1), pp.53-63
- Bergamo, P., Hammer, C. and Fäh, D., (2021). On the relation between empirical amplification and proxies measured at Swiss and Japanese stations: Systematic regression analysis and neural network prediction of amplification. *Bulletin of the Seismological Society of America*, 111(1), pp.101-120
- Bindi, D., Petrovic, B., Karapetrou, S., Manakou, M., Boxberger, T., Raptakis, D., Pitilakis, K.D. and Parolai, S., (2015). Seismic response of an 8-story RC-building from ambient vibration analysis. *Bulletin of Earthquake Engineering*, 13(7), pp.2095-2120
- Bonilla, L. F., H. J. Steidl, G. T. Lindley, A. G. Tumarkin and R. Archuleta (1997). Site amplification in the San Fernando Valley, California: Variability of the Site effect Estimation Using the S-Wave, Coda and H/V methods. *Bulletin of Seismological Society of America* 87 (3), 710-730
- Bonilla L.F., Guéguen, P., Gélis, C., (2021). Contribution of k-net and kik-net data to the monitoring of nonlinear properties of the shallow crust, 6th IASPEI / IAEE International Symposium: Effects of Surface Geology on Seismic Motion, 1 August 2021, Kyoto, Japan
- Bonnefoy-Claudet S, Cotton F, Bard P-Y (2006). The nature of noise wavefield and its applications for site effects studies: a literature review. *Earth Sci Rev* 79:205–227
- Bonnefoy-Claudet, S., Köhler, A., Cornou, C., Wathelet, M. and Bard, P.Y., (2008). Effects of Love waves on microtremor H/V ratio. *Bulletin of the Seismological Society of America*, 98(1), pp.288-300.
- Borcherdt R. D. (1970). Effects of local geology on ground motion near San Francisco Bay. *Bull. Seismol. Soc. Am.* 60:29–61.
- Bottelin, P., Levy, C., Baillet, L., Jongmans, D. and Gueguen, P., (2013). Modal and thermal analysis of Les Arches unstable rock column (Vercors massif, French Alps). *Geophysical Journal International*, 194(2), pp.849-858.
- Cauzzi, C., Custódio, S., Evangelidis, C.P., Lanzano, G., Luzi, L., Ottemöller, L., Pedersen, H. and Sleeman, R., (2021). Preface to the focus section on European seismic networks and associated services and products. *Seismological Society of America*, 92(3), pp.1483-1490.
- Červený V. (2005). *Seismic Ray Theory*, Cambridge University Press, 2005, ISBN 0-521-01822-6.

Chiarabba, C., A. Amato, M. Anselmi, P. Baccheschi, I. Bianchi, M. Cattaneo, G.P. Cecere, L. Chiaraluca, M.G. Ciaccio, P. De Gori, G. De Luca, M. Di Bona, R. Di Stefano, L. Faenza, A. Govoni, L. Improta, F.P. Lucente, A. Marchetti, L. Margheriti, F. Mele, A. Michelini, G. Monachesi, M. Moretti, M. Pastori, N. Piana Agostinetti, D. Piccinini, P. Roselli, D. Seccia and L. Valoroso (2009). The 2009 L'Aquila (central Italy) Mw 6.3 earthquake: mainshock and aftershocks, *Geophys. Res. Lett.*, 36, L18308; doi:10.1029/2009GL039627

Cornou C, Bard P-Y (2019) D7.3 European strong ground motion characterization road map (Work package WP7: Networking databases of site and station characterization). Submission date 23.04.2019. [http://www.sera-eu.org/export/sites/sera/home/.galleries/Deliverables/SERA\\_D7.3\\_European\\_strong\\_ground\\_motion\\_characterization\\_road\\_map.pdf](http://www.sera-eu.org/export/sites/sera/home/.galleries/Deliverables/SERA_D7.3_European_strong_ground_motion_characterization_road_map.pdf)

Cultrera G., L. Faenza, C. Meletti, V. D'Amico, A. Michelini, A. Amato (2014). Shakemaps uncertainties and their effects in the post-seismic actions for the 2012 Emilia (Italy) earthquakes. *Bull Earthquake Eng.*, 12, 5, 2147-2164; DOI: 10.1007/s10518-013-9577-6

Cultrera G., Bordoni P., Mercuri A., Quintiliani M., Minarelli L., Famiani D., Casale P., Pischiutta M., Ladina C., Cara F., Di Giulio G., Pucillo S., Monti G., and CRISP Working Group (2022). Database for site characterization of permanent seismic stations (CRISP). Istituto Nazionale di Geofisica e Vulcanologia (INGV). <https://doi.org/10.13127/crisp>

Cultrera, G., Cornou, C., Di Giulio, G. et al. (2021). Indicators for site characterization at seismic station: recommendation from a dedicated survey. *Bull Earthquake Eng* 19, 4171–4195. <https://doi.org/10.1007/s10518-021-01136-7>

D'Amico M, Felicetta C, Russo E, Sgobba S, Lanzano G, Pacor F, Luzi L (2020). Italian Accelerometric Archive v 3.1 - Istituto Nazionale di Geofisica e Vulcanologia (INGV). <https://doi.org/10.13127/itaca.3.1>

Di Giulio, G., Cultrera, G., Cornou, C. et al. (2021). Quality assessment for site characterization at seismic stations. *Bull Earthquake Eng* 19, 4643–469. <https://doi.org/10.1007/s10518-021-01137-6>

Di Luccio, F., Chiodini, G., Caliro, S., Cardellini, C., Convertito, V., Pino, N.A., Tolomei, C. and Ventura, G., 2018. Seismic signature of active intrusions in mountain chains. *Science advances*, 4(1), DOI: 10.1126/sciadv.1701825

Colombero, C, Baillet, L., Comina, C., Jongmans, D., Larose, E., Valentin, J. and Vinciguerra S. (2018). Integration of ambient seismic noise monitoring, displacement and meteorological measurements to infer the temperature-controlled long-term evolution of a complex prone-to-fall cliff, *Geophysical Journal International*, Volume 213, Issue 3, Pages 1876–1897, <https://doi.org/10.1093/gji/ggy090>

Cultrera G., Bordoni P., Mercuri A., Quintiliani M., Minarelli L., Famiani D., Casale P., Pischiutta M., Ladina C., Cara F., Di Giulio G., Pucillo S., Monti G., and CRISP Working Group (2022). Database for site

characterization of permanent seismic stations (CRISP). Istituto Nazionale di Geofisica e Vulcanologia (INGV). <https://doi.org/10.13127/crisp>

Di Stefano R. and M.G. Ciaccio (2014). The lithosphere and asthenosphere system in Italy as inferred from the Vp and Vs 3D velocity model and Moho map. *Journal of Geodynamics*, Volume 82, Pages 16-25, ISSN 0264-3707, <https://doi.org/10.1016/j.jog.2014.09.006>.

Dündar, S., Dias, N.A., Silveira, G. et al. (2016). Estimation of the Crustal Bulk Properties Beneath Mainland Portugal from P-Wave Teleseismic Receiver Functions. *Pure Appl. Geophys.* 173, 1949–1970 . <https://doi.org/10.1007/s00024-016-1257-4>

Fäh, D., Kind, F. and Giardini, D., (2003). Inversion of local S-wave velocity structures from average H/V ratios, and their use for the estimation of site-effects. *Journal of Seismology*, 7(4), pp.449-467.

Felicetta, C., Mascandola, C., Spallarossa, D., Pacor, F., Hailemichael, S. and Di Giulio, G., (2021). Quantification of site effects in the Amatrice area (Central Italy): Insights from ground-motion recordings of the 2016–2017 seismic sequence. *Soil Dynamics and Earthquake Engineering*, 142, p.106565.

Festa, G.; Adinolfi, G.M.; Caruso, A.; Colombelli, S.; De Landro, G.; Elia, L.; Emolo, A.; Picozzi, M.; Scala, A.; Carotenuto, F.; Gammaldi, S.; Iaccarino, A.G.; Nazeri, S.; Riccio, R.; Russo, G.; Tarantino, S.; Zollo, A (2021). Insights into Mechanical Properties of the 1980 Irpinia Fault System from the Analysis of a Seismic Sequence. *Geosciences* 2021, 11, 28. <https://doi.org/10.3390/geosciences11010028>

Ferreira, A.M. & Woodhouse, J.H., (2007). Source, path and receiver effects on seismic surface waves, *Geophys. J. Int.*, 168(1), 109–132.

Field, E. H., K. A. Jacob (1995). Comparison and test of various site-response estimation techniques, including three that are not reference-site dependent. *Bulletin of Seismological Society of America* 85, 1127-1143, August 1995

García-Jerez, A., Piña-Flores, J., Sánchez-Sesma, F.J., Luzón, F. and Pertou, M., (2016). A computer code for forward calculation and inversion of the H/V spectral ratio under the diffuse field assumption. *Computers & geosciences*, 97, pp.67-78.

Ghofrani, H. and Atkinson, G.M., (2014). Site condition evaluation using horizontal-to-vertical response spectral ratios of earthquakes in the NGA-West 2 and Japanese databases. *Soil Dynamics and Earthquake Engineering*, 67, pp.30-43.

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.

Guéguen, P., Cornou, C., Garambois, S. and Banton, J., (2007). On the limitation of the H/V spectral ratio using seismic noise as an exploration tool: application to the Grenoble valley (France), a small apex ratio basin. *Pure and applied geophysics*, 164(1), pp.115-134.

Heimann, S., Kriegerowski, M., Isken, M., Cesca, S., Daout, S., Grigoli, F., Juretzek, C., Megies, T., Nooshiri, N., Steinberg, A., Sudhaus, H., Vasyura-Bathke, H., Willey, T.; Dahm, T. (2017). *Pyrocko - An open-source seismology toolbox and library*, Potsdam: GFZ Data Services. <https://doi.org/10.5880/GFZ.2.1.2017.001>

Hillers, G. and Ben-Zion, Y., (2011). Seasonal variations of observed noise amplitudes at 2–18 Hz in southern California. *Geophysical Journal International*, 184(2), pp.860-868.

Hobiger, M., Bergamo, P., Imperatori, W., Panzera, F., Lontsi, A.M., Perron, V., Michel, C., Burjánek, J. and Fäh, D., (2021). Site Characterization of Swiss Strong-Motion Stations: The Benefit of Advanced Processing Algorithms. *Bulletin of the Seismological Society of America*.

Hollender, F., Cornou, C., Dechamp, A., Oghalaei, K., Renalier, F., Maufroy, E., Burnouf, C., Thomassin, S., Wathelet, M., Bard, P.Y. and Boutin, V., (2018). Characterization of site conditions (soil class, V S30, velocity profiles) for 33 stations from the French permanent accelerometric network (RAP) using surface-wave methods. *Bulletin of Earthquake Engineering*, 16(6), pp.2337-2365.

INGV Seismological Data Centre (2006, January 1). *Rete Sismica Nazionale (RSN)*. Istituto Nazionale di Geofisica e Vulcanologia (INGV), Italy. <https://doi.org/10.13127/SD/X0FXNH7QFY>

Köhler, A. and Weidle, C. (2019). Potentials and pitfalls of permafrost active layer monitoring using the HVSR method: a case study in Svalbard, *Earth Surf. Dynam.*, 7, 1–16, <https://doi.org/10.5194/esurf-7-1-2019>.

Kula, D., Olszewska, D., Dobiński, W. and M. Glazer (2018). Horizontal-to-vertical spectral ratio variability in the presence of permafrost, *Geophysical Journal International*, Volume 214, Issue 1, July 2018, Pages 219–231, <https://doi.org/10.1093/gji/ggy118>

Kawase, H., Mori, Y. and Nagashima, F., (2018). Difference of horizontal-to-vertical spectral ratios of observed earthquakes and microtremors and its application to S-wave velocity inversion based on the diffuse field concept. *Earth, Planets and Space*, 70(1), pp.1-32.

Kawase, H., Nagashima, F., Nakano, K. and Mori, Y., (2019). Direct evaluation of S-wave amplification factors from microtremor H/V ratios: Double empirical corrections to “Nakamura” method. *Soil Dynamics and Earthquake Engineering*, 126, p.105067. <https://doi.org/10.1016/j.soildyn.2018.01.049>

Kwak, D.Y., Stewart, J.P., Mandokhail, S.U.J. and Park, D., (2017). Supplementing VS30 with H/V spectral ratios for predicting site effects. *Bulletin of the seismological Society of America*, 107(5), pp.2028-2042.



Lachet, C., P.-Y. Bard (1994). Numerical and theoretical investigations on the possibilities and limitations of the Nakamura's technique. *J. Physics of the Earth*, 42-4, 377-397

Lanzano G, Felicetta C, Pacor F, Spallarossa D, Traversa P (2020). Methodology to identify the reference rock sites in regions of medium-to-high seismicity: an application in Central Italy. *Geophys J Int* 222(3):2053–2067. <https://doi.org/10.1093/gji/ggaa261>

La Rocca M, G D Chiappetta, A Gervasi, R L Festa (2020). Non-stability of the noise HVSR at sites near or on topographic heights, *Geophysical Journal International*, Volume 222, Issue 3, Pages 2162–2171, <https://doi.org/10.1093/gji/ggaa297>

Lotti, A., Pazzi, V., Saccorotti, G., Fiaschi, A., Matassoni, L., Gigli, G. (2018). HVSR Analysis of Rockslide Seismic Signals to Assess the Subsoil Conditions and the Site Seismic Response. *International Journal of Geophysics*, <https://doi.org/10.1155/2018/9383189>

Lunedei, E. and Malischewsky, P., (2015). A review and some new issues on the theory of the H/V technique for ambient vibrations. *Perspectives on European earthquake engineering and seismology*, pp.371-394.

Luo, Y., X. Fan, R. Huang, Y. Wang, A.P. Yunus, and H.B. Havenith (2020). Topographic and near-surface stratigraphic amplification of the seismic response of a mountain slope revealed by field monitoring and numerical simulations, *Eng Geol* 271. doi: 10.1016/j.enggeo.2020.105607

Luzi L., Lanzano G., Felicetta C., D'Amico M. C., Russo E., Sgobba S., Pacor, F., & ORFEUS Working Group 5 (2020). Engineering Strong Motion Database (ESM) (Version 2.0). Istituto Nazionale di Geofisica e Vulcanologia (INGV). <https://doi.org/10.13127/ESM.2>

Margheriti L, Nostro C, Cocina O, Castellano M, Moretti M, Lauciani V, et al. (2021). Seismic Surveillance and Earthquake Monitoring in Italy. *Seismological Research Letters* 2021, 92 (3): 1659–1671. doi: <https://doi.org/10.1785/0220200380>

Milana, G., G. Cultrera, P. Bordoni, A. Bucci, F. Cara, R. Cogliano, G. Di Giulio, D. Di Naccio, D. Famiani, A. Fodarella, A. Mercuri, M. Pischiutta, S. Pucillo, G. Riccio, M. Vassallo (2019). Local site effects estimation at Amatrice (Central Italy) through seismological methods. *Bulletin of Earthquake Engineering*. <https://doi.org/10.1007/s10518-019-00587-3>

Minato, S., Tsuji, T., Ohmi, S. and Matsuoka, T., (2012). Monitoring seismic velocity change caused by the 2011 Tohoku-oki earthquake using ambient noise records. *Geophysical Research Letters*, 39(9).

Mital, U., Ahdi, S., Herrick, J., Iwahashi, J., Savvaidis, A., Yong A. (2021). A Probabilistic Framework to Model Distributions of VS30 (2021). *Bulletin of the Seismological Society of America* 111(4): 1677–1692, doi: <https://doi.org/10.1785/0120200281>

Molnar, S., Cassidy, J.F., Castellaro, S., Cornou, C., Crow, H., Hunter, J.A., Matsushima, S., Sánchez-Sesma, F.J. and Yong, A., (2018). Application of microtremor horizontal-to-vertical spectral ratio (MHVSR) analysis for site characterization: State of the art. *Surveys in Geophysics*, 39(4), pp.613-631.

Molnar, S., Sirohey, A., Assaf, J., Bard, P.Y., Castellaro, S., Cornou, C., Cox, B., Guillier, B., Hassani, B., Kawase, H. and Matsushima, S., (2022). A review of the microtremor horizontal-to-vertical spectral ratio (MHVSR) method. *Journal of Seismology*, pp.1-33.

Nakamura, Y. (1989). A method for dynamic characteristics estimation of subsurface using microtremor on the ground surface. *Quarterly Reports of Railway Technical Research Institute (RTRI)*, 30, 25-33

Nakamura, Y., (2019). What is the Nakamura method?. *Seismological Research Letters*, 90(4), pp.1437-1443.

Nogoshi, T. Igarashi (1971). On the amplitude characteristics of microtremor (part 2), *J. Seismol. Soc. Japan*, 24 (1971), pp. 26-40, [In Japanese with English abstract]

NTC18: Ministero delle Infrastrutture e dei Trasporti (2018). Aggiornamento delle Norme Tecniche per le Costruzioni. Part 3.2.2: Categorie di sottosuolo e condizioni topografiche, *Gazzetta Ufficiale* n. 42 del 20 febbraio 2018.

Parolai, S., Bormann, P. and Milkereit, C., (2002). New relationships between  $V_s$ , thickness of sediments, and resonance frequency calculated by the H/V ratio of seismic noise for the Cologne area (Germany). *Bulletin of the seismological society of America*, 92(6), pp.2521-2527.

Piña-Flores, J., Perton, M., García-Jerez, A., Carmona, E., Luzón, F., Molina-Villegas, J.C. and Sánchez-Sesma, F.J., (2017). The inversion of spectral ratio H/V in a layered system using the diffuse field assumption (DFA). *Geophysical Journal International*, <https://doi.org/10.1093/gji/ggw416>

Strollo, A., Cambaz, D., Clinton, J., Danecek, P., Evangelidis, C.P., Marmureanu, A., Ottemöller, L., Pedersen, H., Sleeman, R., Stammler, K. and Armbruster, D., (2021). EIDA: The European integrated data archive and service infrastructure within ORFEUS. *Seismological Society of America*, 92(3), pp.1788-1795. <https://doi.org/10.1785/0220200413>

Wessel, P., W. H. F. Smith, R. Scharroo, J. Luis, and F. Wobbe, *Generic Mapping Tools: Improved Version Released*, *EOS Trans. AGU*, 94(45), p. 409–410, 2013. doi:10.1002/2013EO450001.

Working group INGV "Agreement DPC-INGV 2019-21, All.B2- WP1, Task 2", (2020). Site characterization report at the seismic station IV. INTR–Introdacqua (AQ). Report, <http://hdl.handle.net/2122/14054>

Poggi, V., J. Burjanek, C. Michel and D. Fäh (2017). Seismic site-response characterization of high-velocity sites using advanced geophysical techniques: application to the NAGRA-Net. *Geophysical Journal International*, 210(2), pages 645–659, <https://doi.org/10.1093/gji/ggx192>

Priolo, E., Pacor, F., Spallarossa, D., Milana, G., Laurenzano, G., Romano, M.A., Felicetta, C., Hailemichael, S., Cara, F., Di Giulio, G. and Ferretti, G., (2019). Seismological analyses of the seismic microzonation of 138 municipalities damaged by the 2016–2017 seismic sequence in Central Italy. *Bulletin of Earthquake Engineering*, pp.1-41.

Rao S., Salvaterra L., Acerra C., (2010). Software per l'installazione e la configurazione della stazione sismica GAIA2. *Rapporti Tecnici INGV*, 130, 52.

Riccio, G., Cogliano, R., Di Giulio, G., Fodarella, A., Pucillo, S. and Rovelli, A., (2017). A study of building vibrations induced by weak motions: effects of earthquake excitation, ambient noise and wind speed. *Annals of Geophysics*. 60(4), S0446, doi: 10.4401/ag-7149

Ringdal F. Bungum H., (1977). Noise level variation at NORSAR and its effect on detectability, *Bull. seism. Soc. Am.*, 67(2), 479–492.

Ringler, A. T., J. R. Evans, and C. R. Hutt (2015). Self-noise models of five commercial strong-motion accelerometers, *Seismol. Res. Lett.* 86,1143–1147, doi:10.1785/0220150027

Sánchez-Sesma, F.J., (2017). Modeling and inversion of the microtremor H/V spectral ratio: physical basis behind the diffuse field approach. *Earth, Planets and Space*, 69(1), pp.1-9.

SESAME (2004). Guidelines for the implementation of the H/V spectral ratio technique on ambient vibrations measurements, processing and interpretation, WP12 – Deliverable D23.12, SESAME EU research project (“Site EffectS assessment using AMbient Excitations”, European Commission – Research General Directorate Project No. EVG1-CT-2000-00026). [http://sesame.geopsy.org/Papers/HV\\_User\\_Guidelines.pdf](http://sesame.geopsy.org/Papers/HV_User_Guidelines.pdf) (last accessed on 15 July 2022)

Tarquini S., I. Isola , M. Favalli , F. Mazzarini, M. Bisson, M.T. Pareschi, E. Boschi (2007). TINITALY/01: a new Triangular Irregular Network of Italy, *Annals of Geophysics*, 50, 407-425.

Theodoulidis, N., Cultrera, G., Cornou, C. P.-Y. Bard, T. Boxberger, G. Di Giulio, A. Imtiaz, D. Kementzetzidou, K. Makra, and The Argostoli NERA Team (21018). Basin effects on ground motion: the case of a high-resolution experiment in Cephalonia (Greece). *Bull Earthquake Eng* 16, 529–560. <https://doi.org/10.1007/s10518-017-0225-4>

Tokimatsu, K., (1997). Geotechnical site characterization using surface waves, in *Earthquake Geotechnical Engineering*, pp. 1333–1368, ed. Ishihara, K., Balkema

Vassallo, M., Cultrera, G., Di Giulio, G., Cara, F., & Milana, G. (2022). Peak frequency changes from HV spectral ratios in Central Italy: Effects of strong motions and seasonality over 12 years of observations. *Journal of Geophysical Research: Solid Earth*, 127, e2021JB023848. <https://doi.org/10.1029/2021JB023848>

Wathelet, M., Chatelain, J.-L., Cornou, C., Di Giulio, G., Guillier, B., Ohrnberger, M. and Savvaidis, A. (2020). Geopsy: A User-Friendly Open-Source Tool Set for Ambient Vibration Processing. *Seismological Research Letters*, 91(3), 1878--1889, doi: 10.1785/0220190360

Weaver RL (1982). On diffuse waves in solid media. *J Acoust Soc Am* 71:1608–1609

Wu, C., Peng, Z. Temporal changes of site response during the 2011 Mw 9.0 off the Pacific coast of Tohoku Earthquake. *Earth Planet Sp* 63, 51 (2011). <https://doi.org/10.5047/eps.2011.06.011>

Wu, C. and Z. Peng (2012). Long-Term Change of Site response after the Mw9.0 Tohoku Earthquake in Japan, *Earth Planets Space*, 64, 1259-1266

Zhu, C., Pilz, M. and Cotton, F., (2020a). Evaluation of a novel application of earthquake HVSR in site-specific amplification estimation. *Soil Dynamics and Earthquake Engineering*, 139, p.106301.

Zhu, C., Pilz, M. and Cotton, F., (2020b). Which is a better proxy, site period or depth to bedrock, in modeling linear site response in addition to the average shear-wave velocity?. *Bulletin of Earthquake Engineering*, 18(3), pp.797-820.

## List of authors

**Maurizio Vassallo** (corresponding author, [maurizio.vassallo@ingv.it](mailto:maurizio.vassallo@ingv.it)), Istituto Nazionale di Geofisica e Vulcanologia, Sezione Roma1, viale Francesco Crispi, 43, 67100, L'Aquila Italy

**Gaetano Riccio** ([gaetano.riccio@ingv.it](mailto:gaetano.riccio@ingv.it)), Istituto Nazionale di Geofisica e Vulcanologia, Sezione Roma1, C.da Ciavolone, 83035, Grottaminarda (AV) Italy

**Alessia Mercuri** ([alessia.mercuri@ingv.it](mailto:alessia.mercuri@ingv.it)), Istituto Nazionale di Geofisica e Vulcanologia, Sezione Roma1, via di Vigna Murata, 605, 00143, Rome Italy

**Giovanna Cultrera** ([giovanna.cultrera@ingv.it](mailto:giovanna.cultrera@ingv.it)), Istituto Nazionale di Geofisica e Vulcanologia, Sezione Roma1, via di Vigna Murata, 605, 00143, Rome Italy

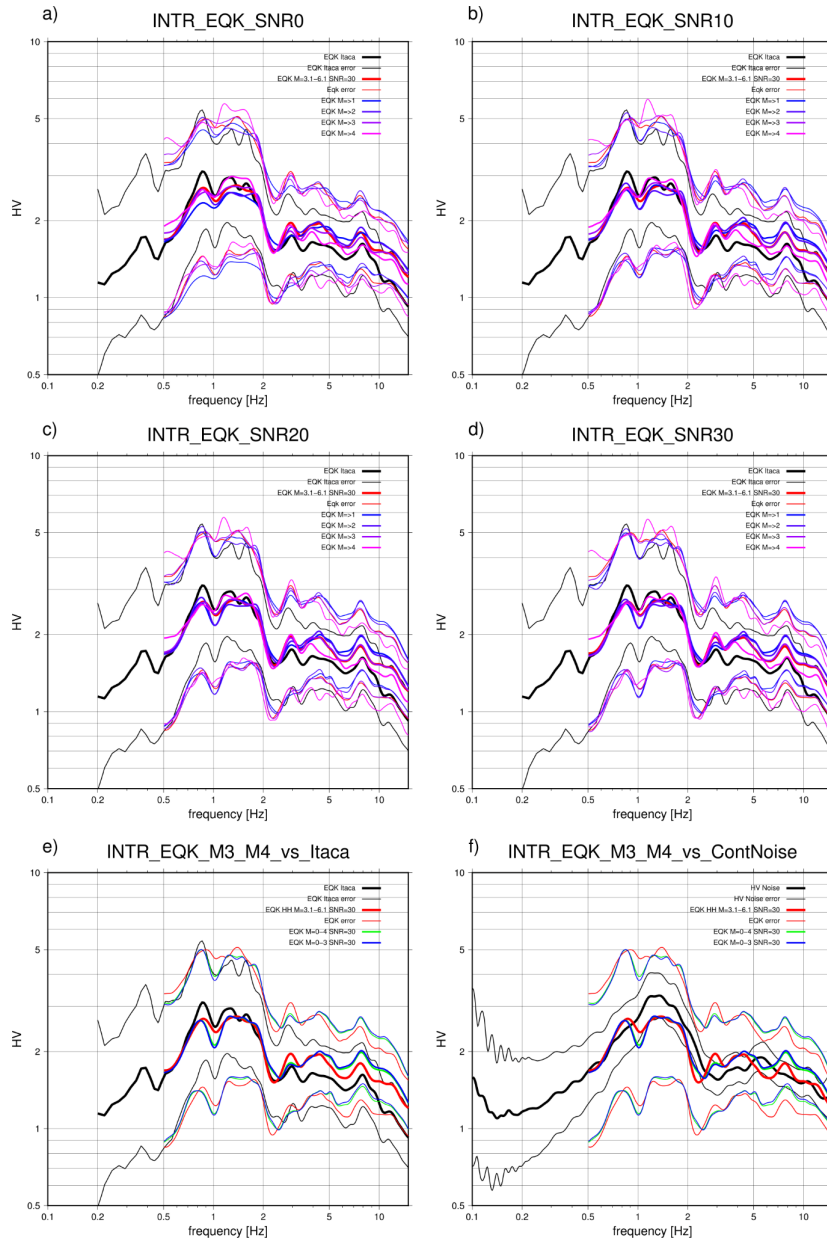
**Giuseppe Di Giulio** ([giuseppe.digiulio@ingv.it](mailto:giuseppe.digiulio@ingv.it)), Istituto Nazionale di Geofisica e Vulcanologia, Sezione Roma1, viale Francesco Crispi, 43, 67100, L'Aquila Italy

## APPENDIX

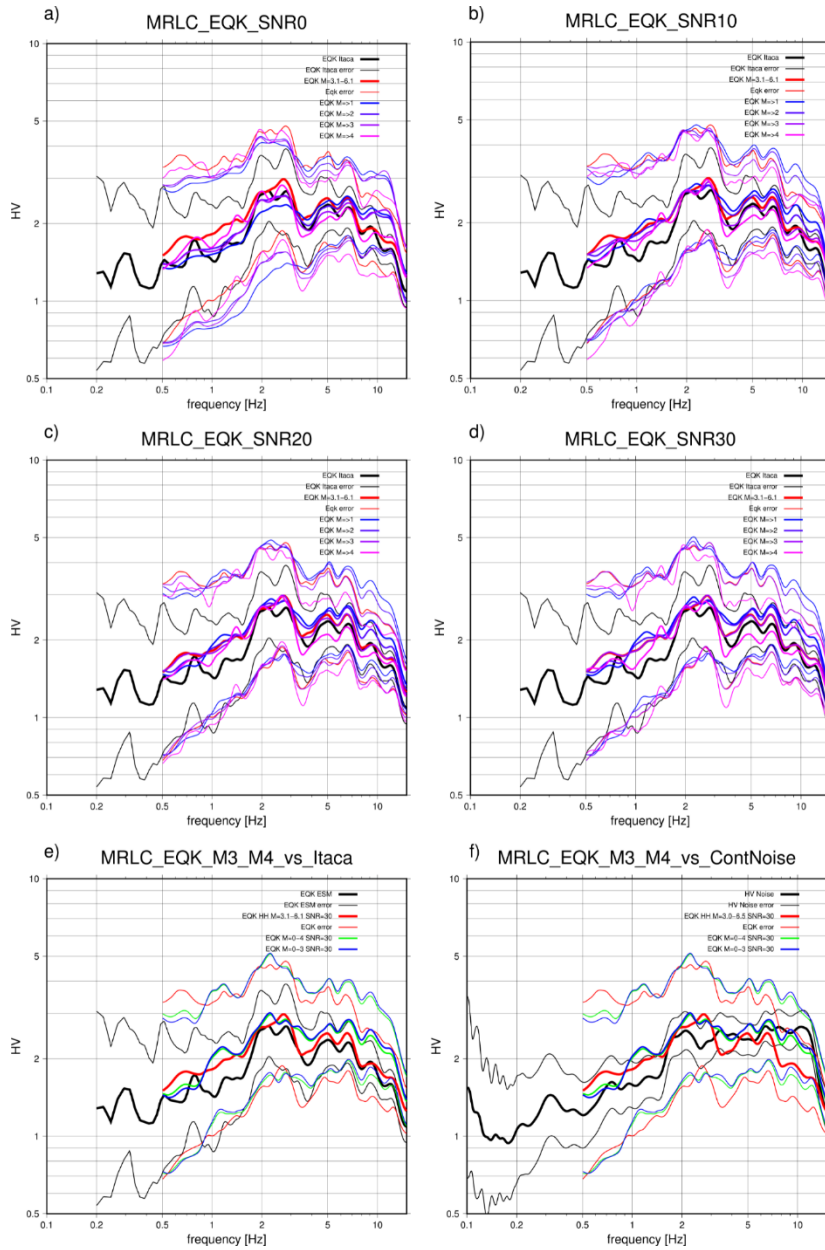
This appendix provides a table and three figures that are not shown in the main article. There is 1) the material related to a test performed using *HVNEA* on IV.INTR and IV.MRLC stations by varying the ranges of magnitude and SNR in order to evaluate the different contributions in the obtained HV curve, and 2) one figure with the results obtained by correlation analysis between air temperature and HV peaks of IV.MRLC station.

Table A1. Number of events processed for IV.INTR and IV.MRLC (HH stream) changing Magnitude and SNR selection. For events with magnitude in the range 1-3, the distance range was fixed at 0-100 km in order to limit the number of events to be analysed; for events higher than magnitude 3, we used the same distance ranges used in ESM database (0-385 km for IV.INTR and 0-310 km for MRLC). The large number of low-magnitude events recorded by the IV.INTR station belong to the 2009 L'Aquila sequence, while the 2016 Amatrice-Visso-Norcia sequence wasn't used due to instrumental problems at the station. The HV curves in ESM archive are obtained using 75 events for IV.INTR and 47 events for IV.MRLC.

|            | IV.INTR |        |        |        | IV.MRLC |        |        |        |
|------------|---------|--------|--------|--------|---------|--------|--------|--------|
|            | SNR=0   | SNR=10 | SNR=20 | SNR=30 | SNR=0   | SNR=10 | SNR=20 | SNR=30 |
| $M \geq 1$ | 13677   | 4366   | 2711   | 1924   | 9602    | 1384   | 865    | 651    |
| $M \geq 2$ | 5868    | 3056   | 2056   | 1512   | 3083    | 718    | 493    | 394    |
| $M \geq 3$ | 1026    | 598    | 490    | 426    | 1917    | 327    | 222    | 172    |
| $M \geq 4$ | 102     | 83     | 74     | 70     | 130     | 95     | 70     | 56     |
| $M \geq 5$ | 7       | 7      | 7      | 7      | 16      | 16     | 16     | 16     |

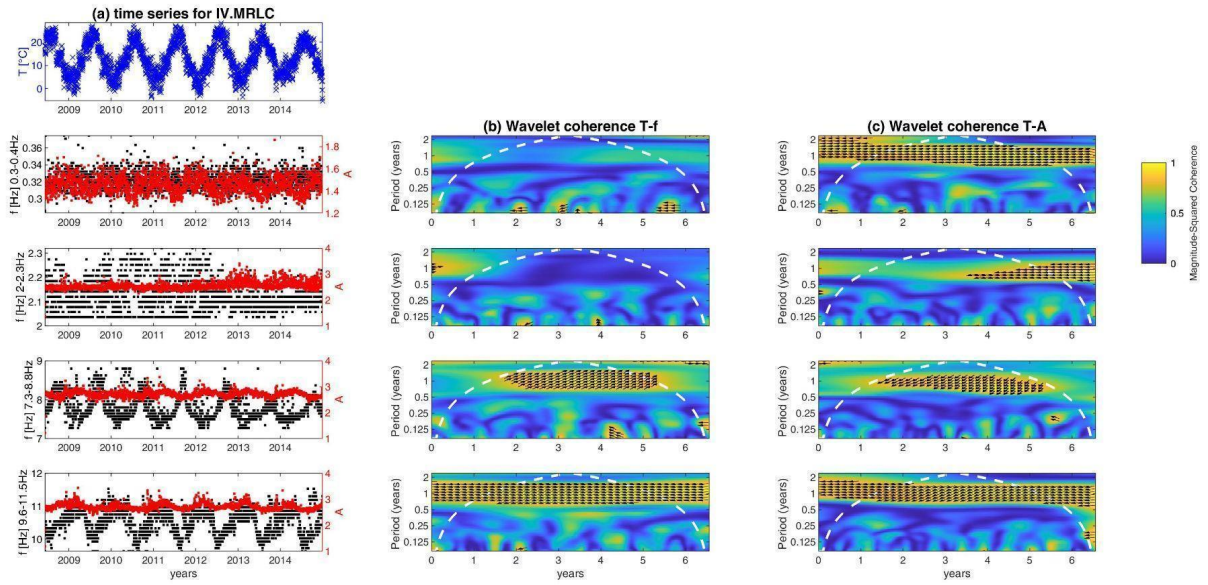


**Figure A1-** IV.INTR station. Comparison between the HV curves from ESM (black line) and HV from automatic analysis on earthquakes selected with parameters  $M=3.1-6.1$ ,  $R=385$  Km,  $SNR=30$  (red line), and by gradually increasing the minimum magnitude value  $M \geq 1$  (Blue),  $M \geq 2$  (light blue)  $M \geq 3$  (purple)  $M \geq 4$  (lilac), and the SNR parameter fixed at value (a)  $SNR=0$ , b)  $SNR=10$ , c)  $SNR=20$ , d)  $SNR=30$ . To estimate the events' number contributing to the curve refer to Table A1. Fig e) shows the comparison between HV curves from ESM (black line) and the HV from HVNEA on earthquakes selected with the following parameters:  $M=3.1-6.1$ ,  $R=385$  Km,  $SNR=30$  (red line);  $M=0-4$ ,  $R=385$  Km and  $SNR=30$  (green lines);  $M=0-3$ ,  $R=385$  Km and  $SNR=30$  (blu lines). Fig f) shows the comparison of HV curves obtained by HVNEA performed on noise (continuous data from 2005-01-04 to 2021-11-26) and on earthquakes selected with the same criteria as in figure e).



**Figure A2:** IV.MRLC station. Comparison between the HV curves from ESM (black line) and HV from analysis on earthquakes selected with parameters  $M=3.1-6.1$ ,  $R=385$  Km,  $SNR=30$  (red line), and by gradually increasing the minimum magnitude value  $M \geq 1$  (Blue),  $M \geq 2$  (light blue)  $M \geq 3$  (purple)  $M \geq 4$  (lilac), and the SNR parameter fixed at value (a)  $SNR=0$ , b)  $SNR=10$ , c)  $SNR=20$ , d)  $SNR=30$ . To estimate the events' number contributing to the curve refer to Table A1. Fig e) shows the comparison between HV curves from ESM (black line) and the HV from HVNEA on earthquakes selected with the following parameters:  $M=3.1-6.1$ ,  $R=310$  Km,  $SNR=30$  (red line);  $M=0-4$ ,  $R=310$  Km and  $SNR=30$  (green lines);  $M=0-3$ ,  $R=310$  Km and  $SNR=30$  (blu lines). Fig f) shows the comparison of HV curves obtained by HVNEA performed on noise (continuous data from 2005-03-25 to 2021-12-30) and on earthquakes selected with the same criteria as in figure e).





**Figure A3.** Correlation between temperature measured at BELLA ( $T$ ) and HV parameters for IV.MRLC (peak frequencies,  $f$ , and its amplitude,  $A$ ) for the period 09/06/2008-31/12/2014 and for four frequency range (0.3-0.4 Hz, 2-2.3 Hz, 7.3-8.8 Hz, 9.6-11.5 Hz). (a) From top to bottom: time series of temperature at BELLA; temporal variation of different peak frequencies (black, left axis) and their amplitudes (red, right axis). (b) Magnitude-squared wavelet coherence between  $T$  and  $f$  in the same frequency range: arrows indicate the phase lag of  $f$  with respect to  $T$  only where the coherence is greater than or equal to 0.8. The direction of the arrows corresponds to the phase lag on the unit circle, which is in this case 1 year (for example, an E direction is 0 lag whereas a vertical arrow indicates a quarter-cycle phase lag). White dashed line is the cone of influence for the wavelet coherence (in cycles per unit time), i.e. areas outside or overlapping the cone of influence are less reliable.

# Supplemental Material

## HV Noise and Earthquake Automatic analysis (HVNEA)

Maurizio Vassallo, Gaetano Riccio, Alessia Mercuri, Giovanna Cultrera, Giuseppe Di Giulio

### User Manual for *HVNEA* software

last update Jul 2022

*HVNEA* is a software package developed to automatically compute the horizontal-to-vertical spectral ratios (HV) on continuous recordings, including both earthquakes and ambient noise vibrations.

In the following we describe the necessary information to a proper use of the package:

|   |    |
|---|----|
| 1. Dependencies .....   | 1  |
| 2. Structure of the HVNEA package, input arguments, configuration files ... | 2  |
| 3. Output of HV_Noise.sh .....  | 7  |
| 4. Output of HV_Eqk.sh .....  | 13 |
| 5. Auxiliary scripts .....  | 18 |
| 6. Operation with Docker .....  | 20 |

## 1. Dependencies

The scripts proposed in this work are based on extensive use of third-party software required for both data analysis and plot creation. In order to ensure that shell scripts work properly, the following packages and commands are required (the version of the packages is suggested as it is the one tested for the latest script analyses):

- Geopsy 3.4.1 (<http://www.geopsy.org/download.php>)
- Gmt 5.4.5 (<https://github.com/GenericMappingTools/gmt/releases/tag/5.4.5>)
- Sac 101.6a (<http://ds.iris.edu/ds/nodes/dmc/forms/sac/>)
- Gsac 1.1.51 ([https://www.eas.slu.edu/eqc/eqc\\_cps/CPS/CPS330.html](https://www.eas.slu.edu/eqc/eqc_cps/CPS/CPS330.html))
- Qt 5.14.1

- Qlib2 2018.157, qmerge 2014.329, ms2sac 1.4.6, qedit 2013.260 (<http://www.ncedc.org/qug/software/ucb/>)
- rdseed (<https://ds.iris.edu/ds/nodes/dmc/software/downloads/rdseed/>)
- leapseconds file (in /usr/local/etc, actual version expires 28-06-2022)
- pyrocko (<https://git.pyrocko.org/pyrocko/pyrocko.git>)
- 
- msi 3.8 (<https://github.com/iris-edu/msi/releases>)
- gnuplot 5.2 (<https://sourceforge.net/projects/gnuplot/files/gnuplot/5.2.8/>)
- awk
- wget
- gv
- Tex Live
- jq 1.5

## 2. Structure of the *HVNEA* package, input arguments, configuration files

In this section we describe in detail the input arguments and the configuration files for the various scripts.

The two main scripts of *HVNEA* are for noise and earthquake HV analysis respectively, and can be found in the package root directory, along with the auxiliary scripts and the files and folders needed for correct operation. The content of the directory is as follows (Figure S1):

- the directory ‘aux’ containing the two templates used by the scripts to generate final reports and some auxiliary files necessary for the correct operation of HV\_Eqk.sh. Here, if desired, you can put the grd file for the topography (for example, italy-90m.grd for Italy, constructed from the DEM downloaded from [https://tinitaly.pi.ingv.it/Download\\_Area2.html](https://tinitaly.pi.ingv.it/Download_Area2.html), Tarquini et al., 2007) even if *HVNEA* still works without this file;
- the directory ‘conf’, which contains the two “main” configuration files used by the scripts (‘hv\_noise.conf’ and ‘hv\_eqk.conf’ for noise and earthquake, respectively), the two parameters files used by Geopsy (‘param\_hv\_noise.conf’ and ‘param\_hv\_eqk.conf’ for noise and earthquake, respectively), a file specifying the

event list structure ('evt\_list.conf') used by HV\_Eqk.sh, a file containing information about FDSN data centers from which the main scripts retrieve data and events ('fdsn\_ws.conf'), a file containing the data paths of the networks available locally, which is also used by the scripts to decide if the data of the chosen network are available locally ('datapath.conf'), finally a file containing any component changes to be made on the input files ('comp\_changes.conf'), necessary in order to still perform the H/V with Geopsy in cases where the channel names are not standard. For both scripts the path of the directory containing the configuration files can be overridden from the command line using option -g, to allow for customization in a multi-user environment;

- the directory 'dataless' needed for operation on local data;
- the two main scripts along with six auxiliary scripts ('hv\_eqk\_filter.sh', 'hv\_noise\_filter.sh', 'hv\_eqk\_recovery.sh', 'hv\_noise\_recovery.sh', hv\_eqk\_merge.sh and hv\_noise\_merge.sh) for output merging and/or post processing;
- the directory 'tools' containing files with functions used by the scripts.



Figure S1. Contents of the package root directory.

The command line input arguments for the two main scripts are used to identify the station and the group of channels (stream) to be analysed, to determine, as appropriate, the portions of noise or the events to be considered in the analysis, and to specify the output directory. For both scripts a directory containing the configuration files other than the default one can be specified. If the scripts are launched without or with insufficient arguments, information on the inputs required for execution appears on the screen.

A typical command line for *HV\_Noise.sh* is as follows:

```
HV_Noise.sh -n <net> -s <station> -c <stream> -f <start_date> -d <days> -o <output_dir>
```

where the meanings of the arguments are:

- `<output_dir>` is the name of the output directory (for example `OUT/INTR_HV10`);
- `<net>`, `<station>` and `<stream>` indicate network code, station code, and stream name (for example `IV`, `INTR` and `HH`). The stream names are defined according to SEED specifications (Standard for the Exchange of Earthquake Data, [http://www.fdsn.org/pdf/SEEDManual\\_V2.4.pdf](http://www.fdsn.org/pdf/SEEDManual_V2.4.pdf));
- `<start_date>` is the starting date (specified with `-f`) and `<days>` the number of days to be analysed (`-d`).

Optionally the user can specify the duration of a single analysis time window with the `-w` option (that must be between 3600 and 86400 sec), and the time interval between two successive days to be analysed (that must be between 1 and 30) with the `-h` option, overwriting the default values present in the configuration file (see later in this section).

Finally, the `-r` option instructs the script to retry the H/V computation using sac files when the computation on miniseed files produces a segmentation fault for some reason.

The generic input command line for *HV\_Eqk.sh* can be of the form:

```
HV_Eqk.sh -n <net> -s <station> -c <stream> -f <start_date> -t <end_date> -m <Mmin> -M <Mmax> -R <distance> -o <output_dir>
```

where the arguments have the following meanings:

- `<net>`, `<station>` and `<stream>` indicate network code, station code and stream name, just as in the case of noise;
- `-m <Mmin>`, `-M <Mmax>`, `-R <distance>`, `-f <start_date>` and `-t <end_date>` specify the minimum and maximum magnitude, the maximum distance from the station, the period of time in which to select events. All this information is used in input to the appropriate web service to build the list of events to be analysed; alternatively, the user can specify the name of a local file containing the ready-made list of events with option `-l`. After building the list of events (or alternatively read it from a local file), a preliminary map of the selected events is shown on the screen and the user is asked if he wants to proceed with the analysis (unless the `-y` option is specified, which implies execution of the analysis without asking for confirmation). This map is preliminary because it is generated only from the existing list and does not take into account the real availability of the event

recordings at the target station, but still important to make a decision whether to continue with the data analysis or to refine the parameters for a new events selection.

- `<output_dir>`, specified with option `-o`, is the output directory.

Optionally the user can specify the minimum signal-to-noise ratio (SNR) for each event to be taken into account in the analysis, thus overwriting the default value present in the configuration file. The SNR is defined as the ratio between the maximum amplitude of the signal in a 6 s time-window including the S arrival and the rms value on a noise time window of fixed length preceding the origin time of the event. The filter based on SNR can also be applied afterwards using the auxiliary script `'hv_eqk_filter.sh'` that works on the output directory.

The parameters in `HV_Noise.sh` main configuration file `'hv_noise.conf'` are explained in Table 1

| parameter name                    | description   |
|-----------------------------------|---|
| <code>g_cfg_drt0</code> (seconds) | Record length to be analysed for each day, with allowed range 3600 - 86400. It can be useful in cases where you want to differentiate the results over the course of the day, for example to analyse day/night variations to highlight the effects of any anthropogenic sources |
| <code>g_cfg_step</code> (days)    | Time interval between two consecutive days to be analysed. It can be useful to quickly get a rough idea of the HV noise trend over a long period of time, before carrying out the actual analysis   |

|   |   |
|---|---|
| <code>g_cfg_perc</code> (percentage)  | Minimum number of windows that are to be selected by Geopsy so that a day is taken into account in the final average  |
| <code>g_cfg_plot_xscale</code> ,<br><code>g_cfg_plot_yscale</code> ,<br><code>g_cfg_plot_ymin</code> ,<br><code>g_cfg_plot_ymax</code>    | These parameters are used to set some properties of the figure containing the individual HV curves, i.e. linear or logarithmic scale for frequencies and amplitudes, and amplitude range displayed on the y axis. When <code>g_cfg_plot_ymin</code> and <code>g_cfg_plot_ymax</code> are not set, they fit the minimum and maximum amplitude respectively |
| <code>g_cfg_plot_fscale</code> ,<br><code>g_cfg_plot_cbscale</code> ,<br><code>g_cfg_plot_cbmin</code> ,<br><code>g_cfg_plot_cbmax</code> | These are used to set in the contour plot linear or logarithmic scale for frequency axis and colorbar, and the range of values displayed on the colorbar. <code>g_cfg_plot_cbmin</code> and <code>g_cfg_plot_cbmax</code> behave in the same manner as <code>g_cfg_plot_ymin</code> and <code>g_cfg_plot_ymax</code> when not set                         |
| <code>g_cfg_plot_ftype</code>   | Controls the output images file type (pdf or png)   |
| <code>g_cfg_noise_rept</code>   | Full path of the report template  |

Table 1: parameters in `hv_noise.conf`

For the sake of convenience, the first two parameters can be overwritten from the command line by using the options `-w` and `-h`, respectively.

Parameters in `HV_Eqk.sh` configuration file '`hv_eqk.conf`' are explained in Table 2



| parameter name  | description   |
|---|---|
| g_cfg_fgrid   | Path of the grid dem file used to make the maps with the selected events. If this path is not or not correctly set, the maps are created on a simple two colour background  |
| g_cfg_fiasp91 and<br>g_cfg_fmodl3d                        | IASPEI file and 3D model file necessities for the theoretical computation of S-arrival. The user must set appropriately these paths in order for the script to work properly  |
| g_cfg_plot_yscale,<br>g_cfg_plot_ymin,<br>g_cfg_plot_ymax | These are used to set linear or logarithmic scale and amplitude range for y axis in the output figure containing individual HV curves. g_cfg_plot_ymin and g_cfg_plot_ymax if not set fit minimum and maximum amplitude of the actual ratios respectively |
| g_cfg_plot_filetype                                       | It is identical to that of the noise case   |
| g_cfg_sn0   | It is the default minimum signal to noise ratio, and is overwritable by the user from command line using option -N  |
| g_cfg_eqk_rept  | Full path of the report template  |

Table 2: parameters in hv\_eqk.conf

As for the Geopsy configuration files param\_hv\_eqk.conf and param\_hv\_noise.conf, both are editable by the user to meet specific needs. Obviously in the case of HV\_Eqk.sh the length set for the sliding time window of analysis is ignored by the script, which instead uses a single window containing phase S. Values of parameters contained in the version of

param\_hv\_noise.conf coming with the package were set using our experience and according to the standard values proposed by Sesame guidelines on data processing (Sesame 2004).

Both scripts make use of the ‘datapath.conf’ configuration file to decide whether the data for a given network should be downloaded from the FDSN web services or are available locally, and in the latter case to retrieve the storage path. An example of line in the file is:

```
XY 2010/01/01 2010/12/31 /data/XY
```

and means that data for network ‘XY’ are available locally limited to year 2010, and are filed in directory ‘/data/XY’ according to the Seedlink Data Structure. Note that the case of a network whose data is partly available locally and partly through the web services is not managed by the scripts, i.e. they consider a network available locally if it is present in the datapath.conf, remotely otherwise.

In the file ‘comp\_changes.conf’ the user specifies any changes to the channel names to be made on the input files; an example of a file with only one line besides the header is:

```
# net | sta | loc | comp | new_comp
```

```
IV | TB01 | 18 | DH1 | DHE
```

and instructs the scripts to change the channel name from DH1 to DHE in the input file when the triad network, station, location is IV, TB01, 18.

The two remaining configuration files ‘evt\_list.conf’ and ‘fdsn\_ws.conf’ contain the event list data structure (i.e. field separator and field names), and the information needed by the scripts to correctly use FDSN web services (i.e. the names of the data centres from which the user wants to retrieve data and events, in addition to the url from which the scripts retrieve information on individual data centres) respectively.

### **3. Output of HV\_Noise.sh**

Before describing the output of the two scripts in detail, it is worth mentioning that the data stream of a seismic station is uniquely identified by the group of four parameters: network, station, location and channel (see Standard for the Exchange of Earthquake Data

Manual at [www.fdsn.org/publications/](http://www.fdsn.org/publications/)). A seismic station can include more than one location, such as in the case of two sensors, one of which could be installed in a borehole, but more generally different locations are distinguished when the measurement points cannot be considered coincident. In our context any different locations of the station are kept distinct because the output file names contain the location name, as well as network, station, and stream (here defined by convenience as the string consisting of the first two letters of the channel name).

A HV\_Noise.sh output directory contains, in addition to the actual output files (i.e. images, files with the average HV curve, etc.), a subdirectory ‘HV-’, with all the daily HV curves stored in text files produced by Geopsy, and a subdirectory ‘files’, containing some text files with various information on the input and configuration parameters. When using the following command line

```
HV_Noise.sh -n IV -s INTR -c HH -f 2010-10-01 -d 114 -o OUT/INTR_HV10 (1)
```

the resulting output directory OUT/INTR\_HV10 (with the image type in the configuration file set to pdf) looks like:

OUT/INTR\_HV10

```
|— files
|
| |— args.txt
| |— cmdline.txt
| |— coords.txt
| |— param_hv_noise.conf
|— HV-
| |— IV.INTR..HH.2010.274.hv
| |— IV.INTR..HH.2010.274.log
| |— ...
| |— ...
| |— IV.INTR..HH.2011.022.hv
```

```

| └─ IV.INTR..HH.2011.022.log
|
|─ hv-noise.IV.INTR..HH.pdf
|
|─ IV.INTR..HH.all.hv.bw.pdf
|
|─ IV.INTR..HH.all.hv.pdf
|
|─ IV.INTR..HH.all.hv.txt
|
|─ IV.INTR..HH.list.txt
|
|─ mean-hv-noise.IV.INTR..HH.txt
|
└─ noise_info.IV.INTR..HH.txt

```

More specifically:

- the directory named ‘files’ contains the following files: ‘coords.txt’ with the station coordinates (latitude, longitude and elevation); ‘cmdline.txt’ reporting the command line used; ‘args.txt’ also reports the input parameters in form of list (name of the station, starting time, selected days, length of time analysis, and so on) together with some information deduced from them (e.g. whether the data is local or remote); ‘param\_hv\_noise.conf’ with the parameters used by Geopsy (actually from the geopsy-hv tool);
- the directory ‘HV-’ contains the spectral ratios on the individual days calculated by Geopsy and saved in text files ‘IV.INTR..HH.<yyyy>.<jjj>.hv’ together with the corresponding log text files ‘IV.INTR..HH.<yyyy>.<jjj>.log’, where <yyyy> and <jjj> indicating year and julian day, respectively;
- ‘IV.INTR..HH.all.hv.txt’ is a cumulative text file containing all the individual HV curves in 5 columns (the day of analysis, the frequency value and the resulting daily HV and standard deviation values, number of windows used). An extract of the file (obtained showing only the first and the last day of analysis) is

```

2010-10-01 0.100774 1.52503 2.1547 1439
...
2010-10-01 14.9654 1.23971 1.19942 1439
...
2011-01-22 0.100774 2.38865 2.57485 1421
...

```

2011-01-22 14.9654 1.4232 1.23681 1421

- ‘mean-hv-noise.IV.INTR..HH.txt’ is the HV curve obtained by averaging all the daily HV (geometric mean and standard deviation). After an initial header section, subsequent rows show for each frequency (column 1) the values of the average HV (column 2) and the same average decreased and increased by one standard deviation (columns 3 and 4). An example of the head of the file is as follows

```
# Start date: 2010-10-01
```

```
# Number of days used: 113
```

```
# Daily recording duration: 86400 sec
```

```
# Window length: 60 sec
```

```
# Total number of windows: 160616
```

```
# f0 from average: 1.32
```

```
# A0 from average: 3.17
```

```
0.100774 1.47687 0.698043 3.12464
```

```
0.101789 1.47211 0.703082 3.08231
```

```
0.102815 1.4623 0.713205 2.99817
```

- ‘IV.INTR..HH.all.hv.pdf’ is the image file showing the contour plot of the HV amplitude as a function of time and frequency;
- ‘IV.INTR..HH.all.hv.bw.pdf’ is the image file with the individual HV curves together with the overall average and standard deviation;
- ‘hv-noise.IV.INTR..HH.pdf’ is a synthetic report containing all the relevant information about output and the two images described above (Figure S2);
- the text file ‘noise\_info.IV.INTR..HH.txt’ contains some basic information about output such as: number of days for which the analysis was successfully carried out; starting date; record duration for each day;  $f_0$  value calculated as the geometric mean over the frequencies (calculated by Geopsy on the individual days) at which the maximum amplitudes occur, with corresponding standard deviation ( $f0\_wins$ );  $f_0$  value obtained from the average curve as the lowest frequency where a peak with amplitude at least 2 occurs ( $f1\_avg$ ) and the corresponding amplitude ( $a1\_avg$ ). The content of this file is

```
# general information
```

```
net_code IV
```

sta\_code INTR  
loc\_code  
min\_rel\_frequency 0.1  
stream HH  
nslc IV.INTR..HH  
start\_date 2010-10-01  
begin\_time 00:00:00  
days 113  
num\_windows 160616  
record\_length 86400  
f0\_wins 1.28 1.24 1.31  
f1\_avg 1.32  
a1\_avg 3.17  
data\_source 1  
data\_from INGV Seismology data centre - Italian EIDA node

# geopsy parameters  
gp\_window\_length 60  
gp\_freq\_min 0.1  
gp\_freq\_max 15  
gp\_freq\_scale\_type Log  
gp\_freq\_step\_type Count  
gp\_freq\_num\_samples 500  
gp\_taper\_win\_type Tukey  
gp\_taper\_win\_rev n  
gp\_taper\_win\_width 5  
gp\_sm\_method Function  
gp\_sm\_width\_type Log  
gp\_sm\_width 20  
gp\_sm\_scale\_type Linear

gp\_sm\_win\_type KonnoOhmachi

gp\_horiz\_comp Vectorial sum

- the file 'IV.INTR..HH.list.txt' contains the list of the days for which the analysis was successful, and, for each day, the  $f_0$  value calculated by Geopsy, the number of windows used to calculate the HV, and minimum and maximum amplitude of the curve.

## HVNEA - Noise HVSR analysis on station IV.INTR..HH

Creation date 2022-06-30 07:31:15

Start date: 2010-10-01 00:00:00

Number of days used: 113

Daily recording duration<sup>(1)</sup> (sec): 86400

Window length (sec): 60

Minimum reliable frequency: 0.1

Total number of windows: 160616

Taper type: Tukey

Taper width: 5%

Smoothing type: KonnoOhmachi

Smoothing width<sup>(2)</sup>: 0%

Composition of the horizontal components: Vectorial sum

$f_0$ <sup>(3)</sup> (Hz): 1.32

$A_0$ : 3.17

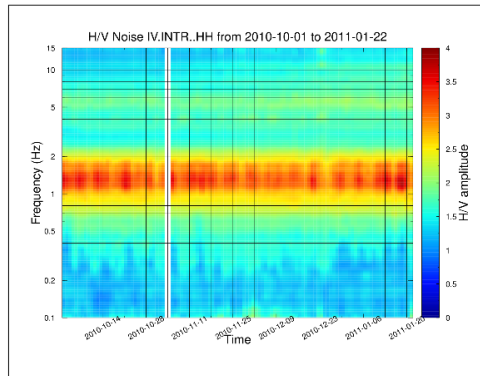


Figure 1

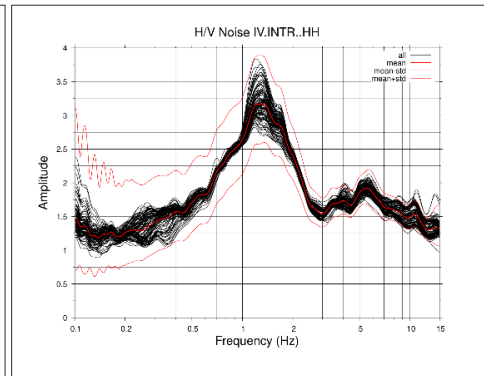


Figure 2

Figure 1: contour plot of noise HVSR amplitude as a function of time and frequency

Figure 2: H/V curves for the selected days (black); mean H/V curve, mean H/V curve minus and plus 1 standard deviation (red)

Note

(1) Daily recording duration is the length of the recording for each analysis day

(2) In the case of Konno-Ohmachi smoothing, smoothing width is related to the parameter  $b$  present in the previous versions of Geopsy by the relation  $\text{width} = 10^{(\pi/b) - 1}$

(3)  $f_0$  is the frequency of the first peak of the H/V curve with amplitude at least equal to 2

Vassallo M., Riccio G., Mercuri A., Cultrera G., Di Giulio G., 2022. HV Noise and Earthquake Automatic Analysis (HVNEA).

<https://github.com/INGV/hvnea>

Disclaimer: The results come from a semi-automatic analysis and could include some data anomalies

Data from INGV Seismology data centre - Italian EIDA node

Figure S2. Synthetic HV\_Noise.sh output report when using command line (1).



In cases when there are two or more locations for a station, the script produces the set of output files described above for each location (except files present in directory 'files', that are inclusive of all locations).

## 4. Output of HV\_Eqk.sh

Referring to the following command line to run HV\_Eqk.sh:

```
HV_Eqk.sh -n IV -s INTR -c HH -f 2010-01-01 -t 2012-12-31 -m 3 -M 7 -R 100 -o OUT/INTR_EQHH (2)
```

the contents of the output directory OUT/INTR\_EQHH, are (with the image type in the cfg file set as pdf) :

OUT/INTR\_EQHH

```
|— evts
| |— ...
| |— 2147989
| | |— HV-
| | | |— IV.INTR..HH.hv
| | | |— IV.INTR..HH.log
| | |— sac
| | |— IV.INTR..HHE.2010.01.17.11.44.11.SAC
| | |— IV.INTR..HHE.2010.01.17.11.44.11.SAC.cut
| | |— IV.INTR..HHN.2010.01.17.11.44.11.SAC
| | |— IV.INTR..HHN.2010.01.17.11.44.11.SAC.cut
| | |— IV.INTR..HHZ.2010.01.17.11.44.11.SAC
| | |— IV.INTR..HHZ.2010.01.17.11.44.11.SAC.cut
| |— ...
| |— ...
| |— 2466739
| | |— HV-
| | | |— IV.INTR..HH.hv
```

- | | └─ IV.INTR..HH.log
- | └─ sac
- | └─ IV.INTR..HHE.2011.12.17.03.03.53.SAC
- | └─ IV.INTR..HHE.2011.12.17.03.03.53.SAC.cut
- | └─ IV.INTR..HHN.2011.12.17.03.03.53.SAC
- | └─ IV.INTR..HHN.2011.12.17.03.03.53.SAC.cut
- | └─ IV.INTR..HHZ.2011.12.17.03.03.53.SAC
- | └─ IV.INTR..HHZ.2011.12.17.03.03.53.SAC.cut
- └─ files
- | └─ args.txt
- | └─ cmdline.txt
- | └─ coords.txt
- | └─ evt\_list.txt
- | └─ model\_1D.txt
- | └─ param\_hv\_eqk.conf
- └─ SNR-0.0-001
- | └─ files
- | | └─ args.txt
- | | └─ coords.txt
- | | └─ param\_hv\_eqk.conf
- | └─ all.IV.INTR..HH.hv
- | └─ eqk\_info.IV.INTR..HH.txt
- | └─ evts\_resp.txt
- | └─ hv-eqk.IV.INTR..HH.pdf
- | └─ map.IV.INTR..HH.pdf
- | └─ mean-hv-eqk.IV.INTR..HH.txt
- | └─ mean-hv-eqk.IV.INTR..HH.txt.S
- | └─ mean.IV.INTR..HH.pdf
- | └─ params.txt
- └─ all.IV.INTR..HH.hv

└─ evts\_resp.txt

└─ map.0.ps

More specifically:

- A file named ‘all.IV.INTR..HH.hv’ containing all the single HV curves together with corresponding SNR.
- A text file named ‘evts\_resp.txt’ containing, for each event/station pair, the event parameters (origin time, latitude, longitude and so on), the station specification (i.e. network, station, location, stream), the  $f_0$  computed by Geopsy for each event ( $f0w$ ) as the frequency at which the maximum amplitude occurs, the minimum and maximum amplitude of the HV curve, the calculated SNR, and a binary flag (res) indicating whether the analysis was successful or not for the given pair. In the current example the first few rows are

```
#id nsls dat lat lon dep mag baz dst f0w min max snr res
```

```
2147989 IV.INTR..HH 2010-01-17T11:44:34.310000 42.63500 13.32000 7.5 3.1 145 84.4 1.31271 0.570394  
7.66529 38.1 0
```

```
2149399 IV.INTR..HH 2010-01-22T12:30:50.780000 42.23900 13.52000 9.1 3.5 128 40.6 1.94767 0.997143  
4.871 191.5 0
```

where the first field is the event identifier, the value of the last field indicates the result of analysis (with 0 and 1 indicating success and failure, respectively). Failure of analysis happens when data are not present for the event under examination, or when components of the motion are not three (NS, EW and Z), or when the configuration of the channels changes in the time interval of interest, or frequency sampling is different for the three-components, and so on.

- A directory ‘files’ containing some intermediate files necessary for the analysis, such as the 1D model for the investigated area ‘model\_1D.txt’, a file ‘evt\_list.txt’ with parameters of the events, a file with the station coordinates ‘coords.txt’, a file ‘cmdline.txt’ reporting the command line used, a text file ‘args.txt’ with the input parameters in form of name/value pairs together with the information on whether the data is local or remote, and the parameter file ‘param\_hv\_eqk.conf’ used by Geopsy.
- A directory ‘evts’ containing, for each event, a subdirectory named with the event identifier with a subdirectory ‘sac’ collecting the seismic traces in sac binary format, the text

files with the HV spectral ratios and the relevant Geopsy log files in subdirectory 'HV-'. We save two kinds of time series: \*.SAC, starting 20 s before the origin time and ending 100 s after it, and \*.SAC.cut, the same traces but cut around S-wave arrival time;

- A directory "SNR-<snr>-<n>", where <snr> is the chosen minimum Signal-to-Noise ratio (SNR) and <n> is a sequential number, containing the analysis results limited to events that are above the threshold value. In the current example the directory is named 'SNR-0.00-001' because no SNR limitation has been applied. This directory contains a cumulative text file 'all.IV.INTR..HH.hv' identical to that present in the main directory, a text file 'mean-hv-eqk.IV.INTR..HH.txt' with the geometric mean of HV ratios and relative standard deviation. There is an additional file with the smoothed geometric mean and relative standard deviation indicated by suffix '.S', the image file 'map.IV.INTR..HH.pdf' containing the map with the events involved and station position, and the image file 'mean.IV.INTR..HH.pdf' showing the single HV ratios together with the smoothed average HV curve with relative standard deviation. Furthermore, it contains a synthetic report 'hv-eqk.IV.INTR..HH.pdf' (Figure S3) with the most relevant information including the two images, a list of the event/station pairs used for the analysis 'evts\_resp.txt' analogous to that in the main output directory but containing only the events for which the analysis has been carried out successfully and that meet the criterion on minimum SNR value (0 in this case), a text file 'params.txt' containing the parameters used to select the events, a text file 'eqk\_info.IV.INTR..HH.txt' containing some basic informations about output. Finally, note that the SNR-0.0-001 directory is made completely equivalent to a main output directory by the fact that it contains a directory 'files' with the appropriate files inside (args.txt, coords.txt, param\_hv\_eqk.conf); so it can be filtered and merged with others output directories. The image file 'map.0.ps' with the preliminary map of the events shown to the user in order to decide whether to continue with the analysis or not; this map is not produced when the -y option is present in the command line.

Note that file 'all.IV.INTR..HH.hv' in this case is identical in the main directory and in the effective output directory 'SNR-0.0-001' because minimum SNR is set to 0; instead, files 'evts\_resp.txt' could be different between the two directories because the version present in 'SNR-0.0-001' contains only events for which the analyses has been carried out successfully.

## HVNEA - Earthquake HVSR analysis on station IV.INTR..HH

Creation date 2022-06-30 07:25:28

Number of earthquakes used for the analysis: 24  
Minimum signal to noise ratio: 0.0  
Time range: 2010-01-17 11:44:34 - 2011-12-17 03:04:16  
Magnitude range: 3.0 - 3.8  
Range of distances from station: 18.9 - 99.1  
Smoothing type: KonnoOhmachi  
Smoothing width<sup>(1)</sup>: 0%  
Minimum reliable frequency: 0.5  
 $f_0$ <sup>(2)</sup> (Hz): 1.53  
 $A_0$ : 2.84

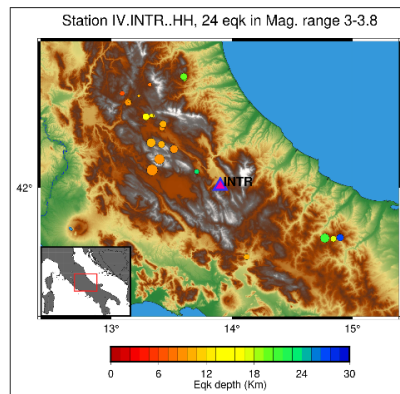


Figure 1

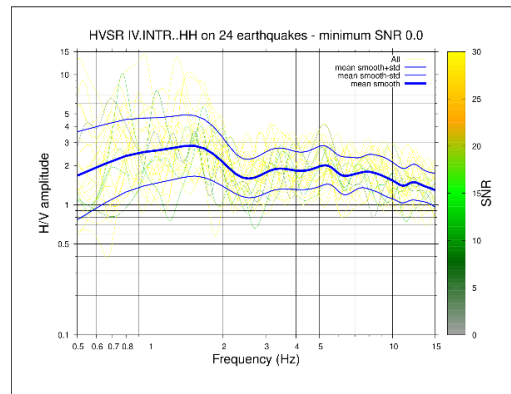


Figure 2

Figure 1: map with the events used in the analysis; the circles representing the events are coloured according to the depth colorbar, the diameters are proportional to the magnitude

Figure 2: individual H/V curves for the selected events, coloured according to the SNR(3) colorbar; smoothed mean, smoothed mean minus and plus one standard deviation (blue)

### Note

(1) In the case of Konno-Ohmachi smoothing, smoothing width is related to the parameter  $b$  present in the previous versions of Geopsy by the relation  $\text{width} = 10^{(\pi/b) - 1}$

(2)  $f_0$  is the frequency of the first peak of the H/V curve with amplitude at least equal to 2

Vassallo M., Riccio G., Mercuri A., Cultrera G., Di Giulio G., 2022. HV Noise and Earthquake Automatic analysis (HVNEA).

<https://github.com/INGV/hvnea>

Disclaimer: The results come from a semi-automatic analysis and could include some data anomalies

Data from INGV Seismology data centre - Italian EIDA node

Events from INGV Seismology data centre - Italian EIDA node

Figure S3. Synthetic HV\_Eqk.sh output report when using command line (2).

## 5. Auxiliary scripts

The *HVNEA* package includes three pairs of auxiliary scripts; one is for retrieving the results (`hv_noise_recovery.sh` and `hv_eqk_recovery.sh`), possibly partial of previous analyzes, the second to filter and/or refine the results according to the user's needs (`hv_noise_filter.sh` and `hv_eqk_filter.sh`), the last to merge several output directories (`hv_noise_merge.sh` and `hv_eqk_merge.sh`). The first pair is useful in cases where you need to recover executions that have not been successful for some reason (for example, interruptions in the power supply or Internet connection during the run). The second allows users to quickly filter results on the basis of various parameters (e.g. minimum SNR value), without having to repeat the download (or extraction) of the waveforms and the relative analysis. The third is used to combine results from different executions.

More specifically, the script '`hv_eqk_filter.sh`' is used to refine an analysis on earthquakes carried out in a previous run of `HV_Eqk.sh` (or `hv_eqk_merge.sh`, or `hv_eqk_recovery.sh`, or itself). It allows you to select from an output directory a subset of events based on the minimum SNR value, on the ranges of magnitude, epicentral distance from the station, depth, back azimuth of the events. The options of the script are:

-s minimum SNR ratio

-p minimum depth

-P maximum depth

-m minimum Magnitude

-M maximum Magnitude

-b minimum back-azimuth

-B maximum back-azimuth

-d minimum distance

-D maximum distance

- a minimum amplitude
- D maximum amplitude
- t text file containing time ranges to take into account
- i valid HV\_Eqk.sh output dir
- g possible configuration directory other than the default one

The resulting output directory is created within the specified output directory, and contains in its name the specified minimum SNR value as well as a sequential number.

For example, after an analysis carried out by running a command line like:

```
HV_Eqk.sh -o OUT/INTR_EQHH -c HH -n IV -s INTR ...
```

if you run `hv_eqk_filter.sh` in the following way uses option `-B`:

```
hv_eqk_filter.sh -i OUT/INTR_EQHH -B 180 (3)
```

you get a selection of events based on the maximum back-azimuth. In this case the output directory is named 'OUT/INTR\_EQHH/SNR-0.0-002', and contains all the appropriate output files. Since the back azimuth is defined here as the angle, measured clockwise starting from North, of the oriented segment that goes from the event to the station, all the events that are located east of the station are not taken into account. The script `hv_noise_filter.sh` is perfectly analogous to `hv_eqk_filter.sh`, but of course it allows selections only on time and amplitudes.

The two scripts "`hv_eqk_recovery.sh`" and "`hv_noise_recovery.sh`" perform the recovery of several output directories coming from `HV_Eqk.sh` and `HV_Noise.sh`, respectively. The directories to recover and complete, specified on the input command line, may be resulting from an incomplete analysis. The output files are generated within the original output directory.

As an example on how to use `hv_eqk_recovery.sh`, to recover some `HV_Eqk.sh` executions that have stopped for some reasons, the appropriate command line is of the type

```
hv_eqk_recovery.sh <dirs_to_recover> (4)
```

where `<dirs_to_recover>` are the directories to retrieve. The script produces the complete output within each directory specified on the command line.

At last, `"hv_eqk_merge.sh"` and `"hv_noise_merge.sh"` merge results from different executions, checking in advance the consistency of network, station, stream and other parameters between the directories to merge. The scripts discard the events or days that are duplicated.

As an example of use of `hv_noise_merge.sh`, let us suppose to have the results obtained from two executions of `HV_Noise.sh` on station IV.INTR, stream HH, on two different period of times (but with same record length, interval between days, and so on), with output directories `OUT/INTR_HV18` and `OUT/INTR_HV19` respectively. The appropriate command line to merge the two set of results into the `"hv_2018-19"` output directory is

```
hv_noise_merge.sh -o hv_2018-19 OUT/INTR_HV18 OUT/INTR_HV19    (4)
```

After this run, the output files (average HV, images, etc.) are produced consistently in the final output directory `hv_2018-19`.

## 6. Operation with Docker

Docker allows users to create, test and run applications, even if they have complex configurations and dependencies, in a simple and replicable way. It collects software into standardised units called *Images* that offer everything needed for a proper execution, including system libraries, system tools, settings, code and routines. Using Docker, the process of deploying an application is reduced to the set-up of a Docker Image, in our case a Linux environment complete with geo-tools, applications for creating graphs or viewing results, libraries, dependencies and configurations. Once an image is built it is possible to execute its functions instantiating a *container* that is a running instance of the *Image*. The performance footprint of the containerization layer is negligible.

The advantage of this approach is that the running container is autonomous and completely independent from the host environment, this means that the same image could be used to run the container and therefore the HVNEA package on any platform where the Docker Engine is



available (Linux, Windows, MacOS, cloud). The Docker Engine is freely available and can be downloaded from: <https://docs.docker.com/get-docker/>.

The definition of a Docker Image is specified in a blueprint file (Dockerfile) that contains all the instructions to build the image, the image itself can be made publicly or privately available storing it in an Image Repository or distributing it as a compressed archive file.

It is evident that using a containerization technology the distribution of a complex analysis tool is particularly fast and efficient, as all issues related to the installation, configuration and adaptation of a specific package are solved (Figure S4).

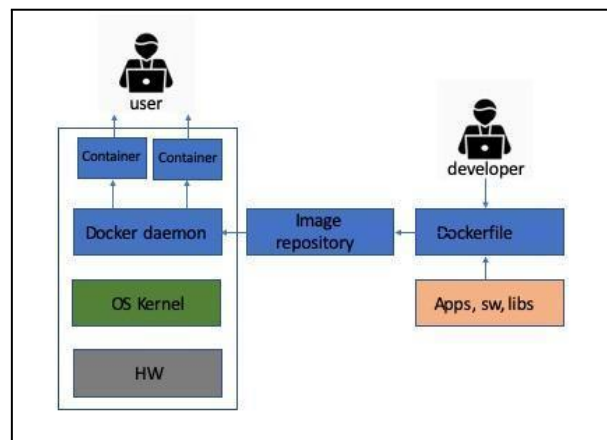


Figure S4. Graphic representation of docker technology.

In our specific case, the analysis package is stored in an image called "hvnea:2.0", that contains all the programs needed for HV analysis. Programs with their version are:

- geopsyack-src-3.4.1 (<http://www.geopsy.org/download.php>)
- gmt-5.4.5 (<https://github.com/GenericMappingTools/gmt/releases/tag/5.4.5>)
- 
- gsac 1.1.51 ([https://www.eas.slu.edu/eqc/eqc\\_cps/CPS/CPS330.html](https://www.eas.slu.edu/eqc/eqc_cps/CPS/CPS330.html))
- Qt.5.14.1
- qlib2.2018.157, qmerge.2014.329, ms2sac.1.4.6, qedit 2013.260 (<http://www.ncedc.org/qug/software/ucb/>)
- rdseed (<https://ds.iris.edu/ds/nodes/dmc/software/downloads/rdseed/>)
- leapseconds file (in /usr/local/etc, current version expires 28-06-2022)
- pyrocko (<https://git.pyrocko.org/pyrocko/pyrocko.git>)
- 
- msi 3.8 (<https://github.com/iris-edu/msi/releases>)
- gnuplot.5.2 (<https://sourceforge.net/projects/gnuplot/files/gnuplot/5.2.8/>)
- awk, wget, gv, Tex Live, jq.

In the following we describe the simple instructions for getting started with Docker. First, install the docker on your machine (<https://docs.docker.com/get-docker/>), then pull images from DockerHub <https://hub.docker.com/r/ingv/hvnea> using the command :

```
#> docker pull ingv/hvnea:2.0
```

where the 2.0 tag corresponds to the current version.

:

This command will create a Docker image called “ingv/hvnea:2.0” in the local repository.

Run a container called “DUM”:

```
#> docker run -e DISPLAY=host.docker.internal:0 --name DUM -it -v  
<path>/OUT:/root/HVNEA/OUT ingv/hvnea:2.0
```

where <path> is the host working directory and the “-e DISPLAY=host.docker.internal:0” option exports the DISPLAY.

The “-v <path>/OUT:/root/HVNEA/OUT” option mounts the “<path>/OUT” dir into the container and it allows to share the existing directory in the container “/root/HVNEA/OUT” with one of the host “<path>/OUT”. The files held in the OUT directory will be shared and visible in both the docker container and the host computer.

For Windows machines with WSL2 (Windows Subsystem for Linux), the sharing of OUT directory is automatic, while for machines with Hyper-V you must allow Docker to access the directory by setting sharing in "Resources menu/file sharing" .

Also, to launch a container named DUM with directory sharing on a Windows machine you need to quote the path as in the example

```
#> docker run --name DUM -it -v "/c/HV/OUT:/root/HVNEA/OUT" ingv/hvnea:2.0
```

where /c/HV/OUT is the path for the example directory (C:\HV\OUT)

Once the container is on, the user is in the HVNEA directory with root privileges, where all the scripts and files described in the paper are present. To re-access an existing container use the command start on an existing container

```
#> docker start -i DUM
```

The container “DUM” will exist until it is deleted using the command:

```
#> docker rm DUM
```

For further information on the commands, please refer to the official Docker documentation:

<https://docs.docker.com/get-started/overview/>

The continuous analysis package requires a graphical interface therefore the graphical terminal must be active. For macOS systems the graphical terminal needs network access permissions, for instance in XQuartz though Preferences/Security put ‘Authenticate connections’=OFF, and ‘Allow connections from network clients’=ON. To allow the channel macOS and docker communicate, the socat program has to be installed on the host. Socat is a relay for bidirectional data transfer between two independent data channels. Install <http://macappstore.org/socat/> and from command line only the first time:

```
#>socat TCP-LISTEN:6000,reuseaddr,fork UNIX-CLIENT:"$DISPLAY"
```

For Windows machines when you use a graphic terminal, for instance VcXsrv, start just the xserver, with the extra settings disabled for native opengl.