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

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)
- 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, 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.

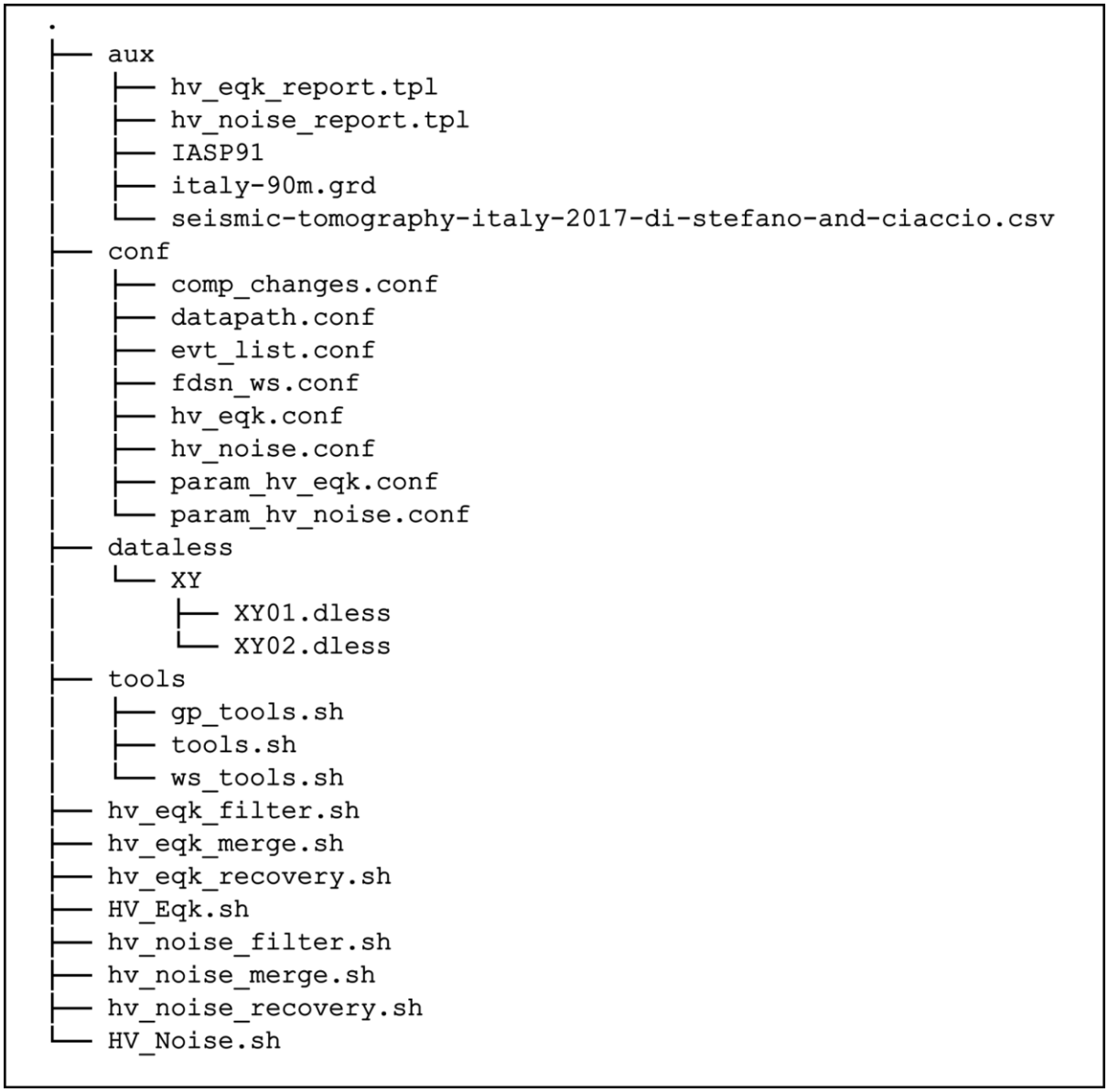

Figure S1. Contents of the package root directory.

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);
- <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

Table 1: parameters in hv_noise.conf

parameter name	Description
g_cfg_drt0 (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
g_cfg_step (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
g_cfg_perc (percentage)	Minimum number of windows that are to be selected by Geopsy so that a day is taken into account in the final average

g_cfg_plot_xscale, g_cfg_plot_yscale, g_cfg_plot_ymin, g_cfg_plot_ymax	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 g_cfg_plot_ymin and g_cfg_plot_ymax are not set, they fit the minimum and maximum amplitude respectively
g_cfg_plot_fscale, g_cfg_plot_cbscale, g_cfg_plot_cbmin, g_cfg_plot_cbmax	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. g_cfg_plot_cbmin g_cfg_plot_cbmax behave in the same manner as g_cfg_plot_ymin and g_cfg_plot_ymax when not set
g_cfg_plot_ftype	Controls the output images file type (pdf or png)
g_cfg_noise_rept	Full path of the report template

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

Table 2: parameters in hv_eqk.conf

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 g_cfg_fmodl3d	IASPEI file and 3D model file necessaries 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, g_cfg_plot_ymin, 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

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/). 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(1) (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(2): 0%
Composition of the horizontal components: Vectorial sum
f_0(3) (Hz): 1.32
A_0: 3.17

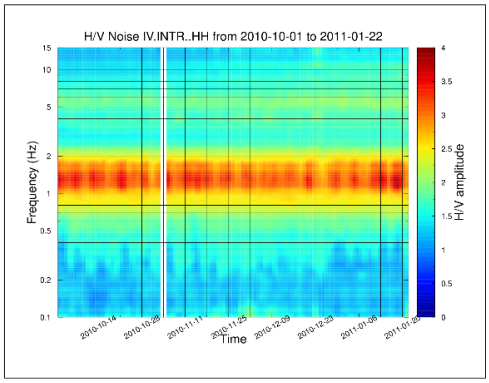

Figure 1

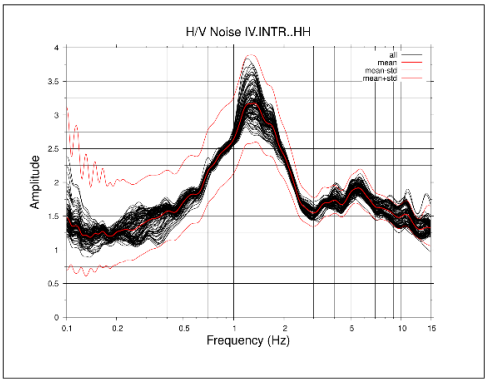

Figure 2

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 width = 10 ^(π/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 informaton 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(1): 0%
Minimum reliable frequency: 0.5
f_0(2) (Hz): 1.53
A_0: 2.84

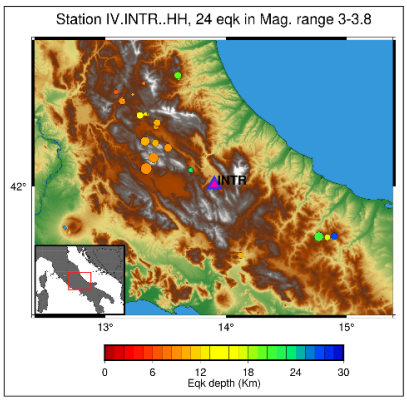

Figure 1

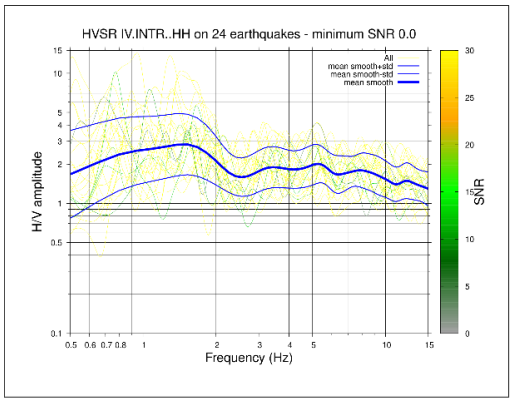

Figure 2

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 width = 10 ^(π/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_HV19 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).

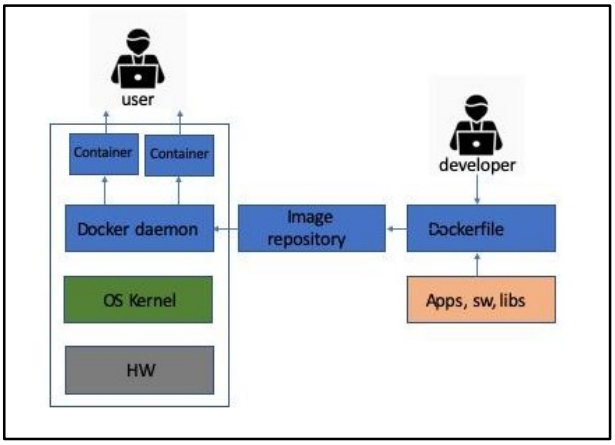

Figure S4. Graphic representation of docker technology.

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:

- geopsypack-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)
- 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.