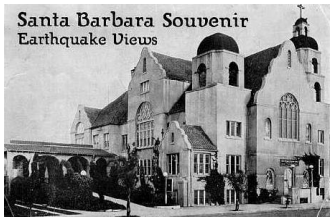

4th IASPEI / IAEE International Symposium:

Effects of Surface Geology on Seismic Motion

August 23–26, 2011 · University of California Santa Barbara

STATISTICAL ESTIMATION OF EARTHQUAKE SITE RESPONSE FROM NOISE RECORDINGS

Valerio De Rubeis
Istituto Nazionale di Geofisica e Vulcanologia
Roma
Italy

Giovanna Cultrera
Istituto Nazionale di Geofisica e Vulcanologia
Roma
Italy

Héloïse Cadet
ISTerre
Université de Grenoble
Grenoble
France

Pierre-Yves Bard
ISTerre, CNRS, IFSTAR
Université de Grenoble
Grenoble
France

Nikos Thedoulidis
Institute of Engineering Seismology and Earthquake Engineering (ITSAK)
Thessaloniki
Geece

ABSTRACT

Standard spectral ratio from earthquake recordings (SSR) is considered the reference empirical method for assessing site effects as a function of frequency. However, other estimates can be easily obtained from noise measurements (i.e., Horizontal-to-Vertical Spectral Ratio, HVN), even though their reliability in terms of amplitude is controversial.
In the framework of the ToK ITSAK-GR (2006-2010) EC project, Cultrera *et al.* (2010) analyzed recordings from 64 sites worldwide, founding that it is possible to have linear combinations of the HVN amplitudes significantly correlated to linear combinations of the SSR. In the present paper we show how to estimate the SSR spectral ratios when only noise measurements are available, using the results of the canonical correlation analysis between SSR and HVN recorded at several sites. The SSR evaluation has been tested by a cross validation procedure: the expected SSR at each validation site are in turn estimated by a weighted average of the SSR values measured at the other sites; the weights are properly set to account more for the sites with similar behavior in terms of the canonical correlation between HVN and SSR. To evaluate the goodness of the estimation, we compared all the inferred and original SSR, and we performed a critical analysis on the spectral characteristics of earthquake site response that can be easily recovered from noise measurements.

INTRODUCTION

There are several approaches to estimate how the earthquake ground motion would be modified by the surface geology. One of the best site effect estimator is the spectral ratio of earthquake recordings at one site compared to a reference site (Standard Spectral Ratio, SSR). However, it is not always feasible to have earthquake recordings available. In this case an alternative method involves the use of the Horizontal-to-Vertical spectral ratio using ambient noise recordings (HVN).
Till today the comparison between the Standard Spectral Ratio (SSR) and the Horizontal-to-Vertical Spectral Ratio from earthquake recordings (HVSR) and from ambient noise (HVN) has been mainly made through fundamental frequency, fo, and its amplitude, Ao, even though these two quantities do not represent an exhaustive picture of the site effect expected for an earthquake.
In general, the correlation between the results obtained with different spectral ratio techniques has always been a controversial task. A systematic short review of such comparisons can be found in Haghshenas *et al.* (2008); all the analyzed studies exhibit (a) a very good agreement between fundamental frequencies obtained with either techniques and (b) an absence of straightforward correlation between HVN amplitude and the actual site amplification measured by the SSR method.
Potentially, analysis of the whole amplification spectral curve could provide more information on the correlation of HVN with the other spectral ratios and on the capability to provide reliable and relevant information regarding site conditions. Cultrera *et al.* (2010)

investigate the whole bandwidth (0.2-10Hz frequency range) of each kind of spectral ratio for a subset of sites from Haghshenas *et al*. (2008), to find if and how the different techniques are mutually correlated. They applied the canonical correlation analysis (Davis, 2002; Cara *et al*., 2008; Theodoulidis *et al*., 2008) to verify and quantify the relationship between SSR and HVN measured at the same sites. Results show that there exists some significant correlation between SSR and HVN which can be statistically quantified, and that it is modulated in specific frequency range.
In this paper we formalize an application of the canonical correlation analysis to reconstruct the expected SSR spectral ratio from the recorded HVN curve at sites located in the investigated areas, through the application of a standard cross validation procedure.

DATA SET

We used a dataset of earthquake and noise recordings at 64 sites around the world (Table 1) collected within the SESAME European project (Theodoulidis *et al*. 2004; Cultrera *et al*., 2010). These data were homogeneously reprocessed by Haghshenas *et al*. (2008) to estimate the mean and standard deviation of spectral ratios (SSR, HVSR and HNV) at each site.

Table 1. List of the experimental areas (Haghshenas *et al*., 2008) and number of sites whose SSR and HVN were used in the statistical analysis.

Experiment area	**Number of sites**
Benevento (Italy)	4
Catania (Italy)	2
Ebron (France)	4
Fabriano (Italy)	4
Grenoble (France)	2
Lourdes (France)	2
Predappio (Italy)	12
Tehran (Iran)	8
Verchiano (Italy)	7
Volvi 1994 (Greece)	13
Volvi 1997 (Greece)	6
TOT	64

We conditioned the spectral ratios to reduce the number of variables: the logarithm of the mean ratio has been averaged within 7 frequency intervals (or bins), chosen with a constant logarithmic lag in the range 0.2-10Hz. This range is the most representative range of the resonance frequencies (fo) for all the sites in the database (50% of sites have fo<3Hz) and for the earthquake engineering as well (Fig. 1).

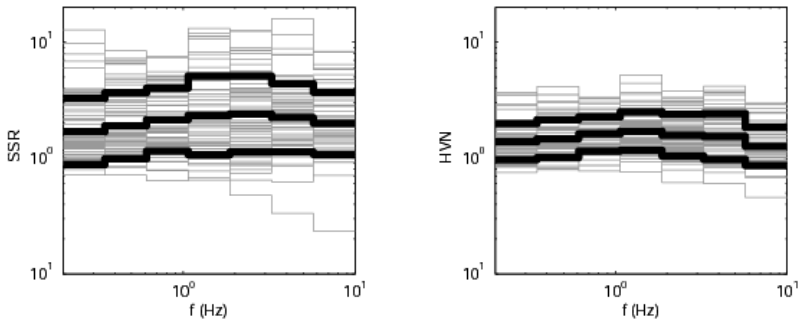

Fig. 1. Average spectral ratios (in 7 bins) for SSR and HVN at 64 sites from 10 experiment areas (Table 1). Black lines refer to the arithmetic mean and plus/minus 1 standard deviation. The bins are obtained subdividing the frequency range 0.2-10Hz in 7 intervals: 0.2-0.35, 0.35-0.61, 0.61-1.07, 1.07-1.87,1.87-3.27, 3.27-5.72, 5.72-10.0 (Hz).

APPLICATION OF CANONICAL CORRELATION ANALYSIS TO SPECTRAL RATIOS

The correlation between the results obtained with different spectral ratio techniques has always been a controversial task. We decided to face this topic using the canonical correlation analysis (Davis, 2002; Theodoulidis *et al*., 2008), which allows to verify and quantify the relationship between two different types of spectral ratios measured at the same sites.
The principle of the canonical correlation is to look for the linear combinations of one set of variables, i.e. frequency bins of SSR, offering the best correlation with the linear combinations of another set of variables, i.e. frequency bins of HVN (Fig.1). In synthesis, the 7 bins of each spectral ratio can be grouped in one single variable (X_{can} for HVN and Y_{can} for SSR), which are the linear combination of the original bins:

$$\begin{aligned} X_{can} &= \sum_{k=1}^{7} b_k HVN(k) \\ Y_{can} &= \sum_{k=1}^{7} c_k SSR(k) \end{aligned} \tag{1}$$

where b_k and c_k are the coefficients (or canonical loadings) obtained for HVN and SSR from the canonical correlation results, and k identify each of the 7 bins.
The coefficients of the linear transformations are found in order to give the highest correlation between the 2 canonical variables (X_{can} and Y_{can}). There can be more than one possible canonical couple and the best ones are chosen using the correlation coefficient *r* and the Rao's scoring test (Rao, 1973), which indicates the probability to have a fortuitous correlation with no meaning. For example, Fig.2a shows the X_{can}-Y_{can} plots of the 1st, 2nd and 3rd canonical couple obtained from the SSR and HVN spectral ratios of Fig.1, having correlation values from about 0.9 to 0.7 and a very low probability to have a correlation with no meaning ($p<10^{-4}$).
The coefficients of Eq. 1 transform the original variable values to the respective canonical variables. It is then interesting to quantify how the amplitudes at different frequencies are related between them and with the other kind of spectral ratio. This is represented through the correlation between each original variable (the 7 frequency bins of HVN and SSR) and their corresponding canonical variable X_{can} (Fig.2b) or Y_{can} (Fig.2c). For example, the correlation between the SSR bins and the Y_{can} (Fig.2c, 1st couple) shows that the 3 central bins (0.6-3.3Hz) are correlated, i.e. they go in the same direction: when one is amplified or deamplified, the others follow. In the case of HVN bins and X_{can}, the range 0.35-1.9Hz is anti-correlated with the 3.3-10Hz interval, that is the low-intermediate frequencies amplification is followed by a lowering of the high-frequencies amplitudes and vice-versa (Fig.2b, 1st couple). Even more, the canonical correlation analysis is telling us how the SSR values are related to the HVN; for example, the 1st couple of Fig.2 shows that the 3 correlated bins of SSR (0.6-3.3Hz; Fig.2c) are correlated with the 0.35-1.9Hz range of HVN (Fig.2b): this means that the amplification or deamplification of a broad range of intermediate frequencies of SSR correspond to the same behavior in HVN in a smaller frequency band.
Finally, the very low and very high frequencies in SSR do not contribute to the correlation with HVN. The low correlation of the first and the last bins in Fig.2c explains that the large variability of the ratios at those frequencies cannot be accounted for in the comparison between the two techniques.

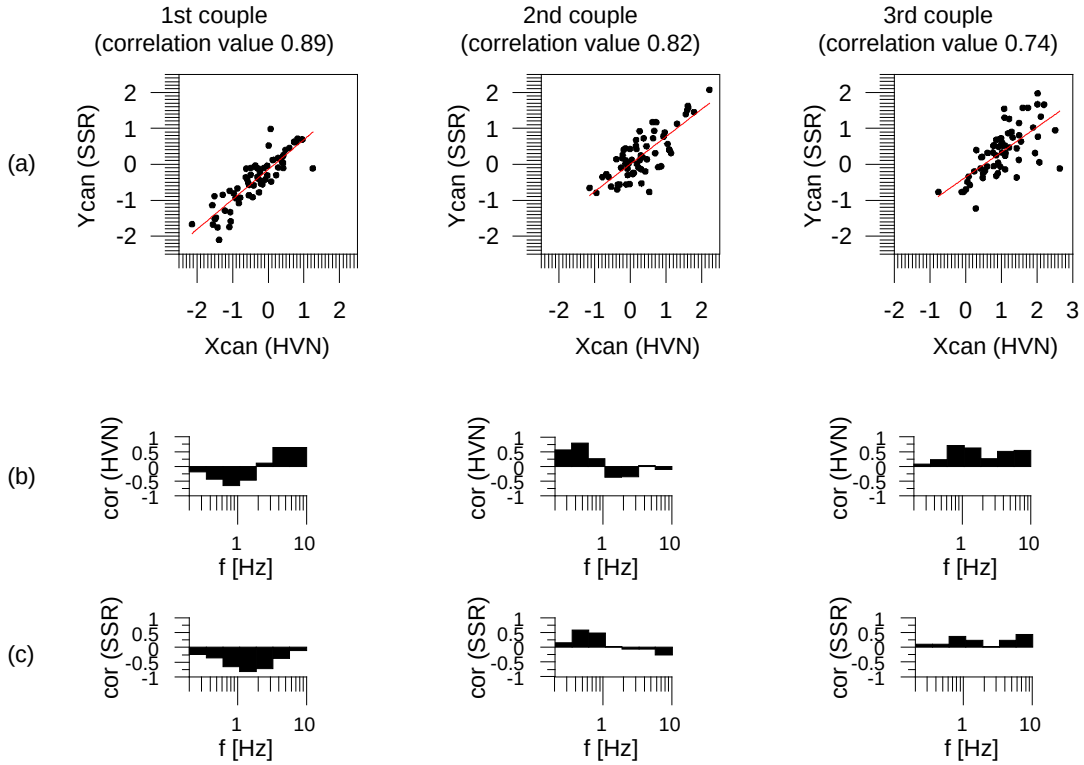

Fig.2. Canonical correlation results for the 1st (left), 2nd (middle) and 3rd (right) canonical couples of SSR and HVN: (a) plot of canonical couple for each site (Xcan from HVN and Ycan from SSR); (b) correlations between the HVN bins and the corresponding Xcan; (c) correlations between the SSR bins and the corresponding Ycan.

ESTIMATION OF EXPECTED SSR FROM HVN MEASUREMENTS

In this paper we show how it is possible to estimate the SSR spectral ratios where noise measurement only are available, using the results of the canonical correlation analysis between SSR and HVN recorded at several sites. The SSR evaluation has been tested by a cross validation procedure: in turn, each single site is extract from the database and we perform an independent canonical correlation analysis with data of all other sites, using such information to retrieve SSR at this site. After repeating the same procedure to all other sites, we performed a data comparison between inferred and original SSR.

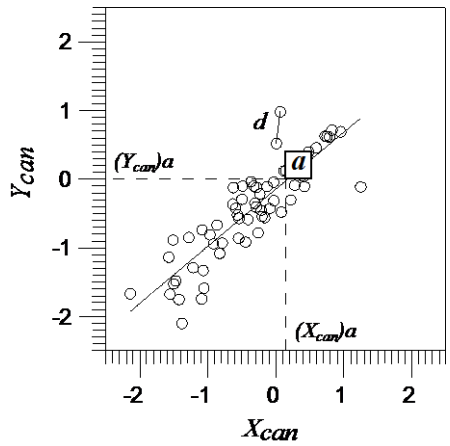

Fig. 3. Example of canonical correlation results of SSR and HVN bins. The plot display the values of the 1st canonical couple for each site (Xcan from HVN and Ycan from SSR) and at the target site a.

Let us suppose to have the 7 frequency bins of HVN at the target site *a* for which we wish to estimate the 7 frequency bins of SSR. Through the canonical coefficients b_k, (Eq. 1) it is possible to assign to this site the proper value of the canonical variable X_{can} from the linear combination of the original HVN(*k*) values. The best representation of the corresponding Y_{can} is estimated from the regression line (Fig.3).

Because the correlation results are statistically significant, we are confident that the true Y_{can} value at the target site is not too different from the value indicated by canonical regression line $(Ycan)_a$. However, we cannot directly recover the SSR curve as a function of frequency from its linear combination Y_{can}. We solved this problem by estimating each bin value expected for the SSR at the target site *a* as a weighted average of the values measured at the other sites.

The weights are properly set in order to account more of the sites with similar behavior in the X_{can}-Y_{can} canonical space for a chosen canonical couple. Moreover, a single bin contribute differently to the linear combination of Eq. 1, as shown by the correlation between each original variable (the 7 frequency bins of HVN and SSR) and their corresponding canonical variable X_{can} (Fig.2b) or Y_{can} (Fig.2c). We used the Moran's index (*MI*) to account of both characteristics, that is the similarity in the X_{can}-Y_{can} canonical space and the contribution of each bin to the canonical variable. *MI* measures the spatial correlation of the bin value at the sites and it is defined as the ratio between the covariance in a given distance range *d* on the X_{can}-Y_{can} plot (Fig.3) and the variance of all sites:

$$MI(k,d) = \frac{n}{\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} w_{ij}} \frac{\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} w_{ij}(SSR(k)_i - \overline{SSR}(k))(SSR(k)_j - \overline{SSR}(k))}{\sum_{i=1}^{n}(SSR(k)_i - \overline{SSR}(k))^2} \quad (2)$$

where $SSR(k)_m$ is the value for the considered bin *k* (from 1 to 7) at the site *m*, *n* is the total number of sites and w_{ij} accounts for the distance range ($w_{ij} = 1$ if the distance between the two sites *i* and *j* is inside the range $d-\Delta d<d_{ij}<d+\Delta d$, and $w_{ij} = 0$ otherwise). The Moran Index can assume values from +1 (perfect correlation) to -1 (anticorrelation) and it is obtained at discrete distances which are fitted with a second degree polynomial fit (MI_{fit}), as shown in Fig.4 for the 7 bins:

$$MI_{fit}(k,d) = a_k d^2 + b_k d + c \quad (3)$$

Figure 4 shows that for the central bins (3 to 5) there is a high correlation at close sites in the canonical space which quickly decreases for larger distances: the spectral ratios in the 0.6-3.3Hz frequency range are very similar when the sites have similar X_{can}-Y_{can} values, and diverge for sites located at distant positions on the canonical regression line.

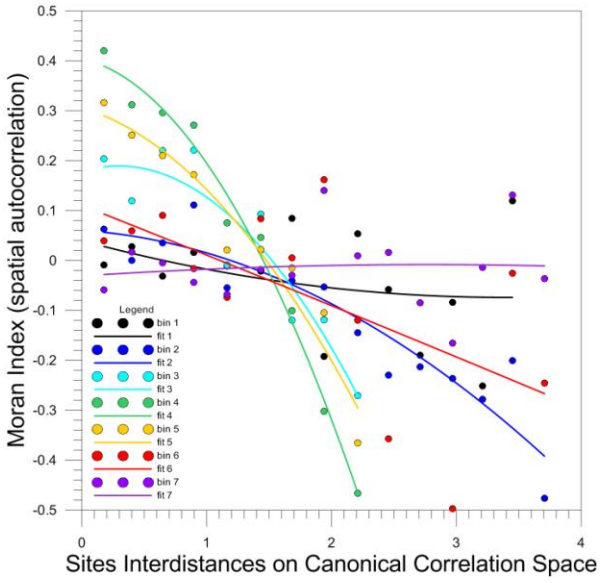

Fig.4. Moran Index behavior for the 7 SSR bins (colored dots) produced by the 1st canonical couple, as a function of the distance in the canonical correlation plane (Fig.3). The continuous lines are the respective 2nd order polinomial fits.

The $MI_{fit}(k,d)$ of Eq. 3 is used to build the weights of each site m (u_{km}, where k is the considered bin) in the estimate of SSR at the target site a:

$$SSR_{ai}(k) = \sum_{m=1}^{n} u_{km} SSR(k)_m,\ m \neq a \tag{4}$$

The computation is done for each canonical couple i. The weight u_{km} in Eq. 4 is a function of the distance d between the target site a and the other sites m on the canonical plane (Fig.3), and it is defined as follow:

$$\begin{aligned} u_{km} &= MI_{fit}(k,d),\ \text{if } MI_{fit} > 0 \\ u_{km} &= 0,\ \text{if } MI_{fit} \leq 0, \end{aligned} \tag{5}$$

In Eq. 5 we account for sites having positive correlation only, because we are looking for the most similar sites to reconstruct the SSR bin of the target site.
Finally, we applied the same procedure to all the other statistically significant canonical couples and we averaged the results to evaluate the SSR at the target site a:

$$SSR_a(k) = \sum_{i=1}^{c} \frac{SSR_{ai}(k)}{U(k)} \tag{6}$$

where SSR_{ai} is defined by Eq.4, c is the total number of all significant canonical couples and U is the normalization factor:

$$U(k) = \sum_{m=1}^{n} u_{k1} + \sum_{m=1}^{n} u_{k2} + \ldots \sum_{m=1}^{n} u_{kc},\ m \neq a \tag{7}$$

The use of the results of all the significative canonical couples (Eq. 6) allows to increase the reliability of each bin reconstruction, because it is observed that bins poorly represented by a single canonical couple are often well represented by one of the other couples. Couples are, by construction, mutually uncorrelated: this increases the ability of each couple to recover the behavior of bins that do not find correlation inside other couples (Fig.2). Note that the weight associated to a single site changes from a canonical couple to the other, because the position of the site (hence the distance range) in the canonical space varies accounting for the different correlation behavior.
The standard deviation of each bin reconstruction (Eq. 6) measures the differences among the SSR evaluated at the target site and the values recorded at the other sites, and it is defined in Eq. 7:

$$\sigma_a(k) = \sqrt{\sum_{i=1}^{c} \sigma_{ai}^2(k)} \tag{7}$$

where σ^2_{ai} is the variance of the evaluation from a single canonical couple i,

$$\sigma_{ai}^2(k) = \sum_{m=1}^{n} \frac{u_{km}}{U} \left(SSR(k)_m - SSR_{ai}(k) \right)^2,\ m \neq a \tag{8}$$

RESULTS

The proposed methodology allows to compute the standard spectral ratios estimated at each site (SSR_{est}) using (*i*) the HVN measurements at the site and (*ii*) the canonical correlation analysis between SSR and HVN at all the other sites. In the latter step (*ii*) we define the weights for the SSR estimate at the target site (Eq 2 to 5). The weights depend on the similarity between the X_{can}-Y_{can} values at the target site and the values at the other sites; in principle, the Y_{can} of the target site cannot be computed without the SSR measurements (Eq.1), but we used the canonical correlation results (Fig.3) to evaluate it from the correlation fit, as from the first step (*i*).

Using the results of the cross-validation procedure described in the previous section, we can compare all the estimated standard spectral ratios with the recorded ones to evaluate the goodness of the fit. We computed the differences between the recorded (SSR_{rec}) and the estimated (SSR_{est}, Eq.6) standard spectral ratios at each site *a*, normalized by the estimated standard deviation σ_a (Eq.7):

$$Res_a(k) = \frac{\left(SSR_{rec}(k) - SSR_{est}(k)\right)_a}{\sigma_a(k)} \tag{8}$$

where *k* indicates one of the 7 bins.

For each site *a*, the average of $Res_a(k)$ over the 7 bins (mean residual value) indicates how much the recorded SSR fall within the mean plus/minus 1 standard deviation of the estimated SSR (Fig.5d). This parameter allows to rank the sites in terms of the goodness of their reconstruction. In Fig.5a to Fig.5c we show examples of SSR from selected sites having different residual values.

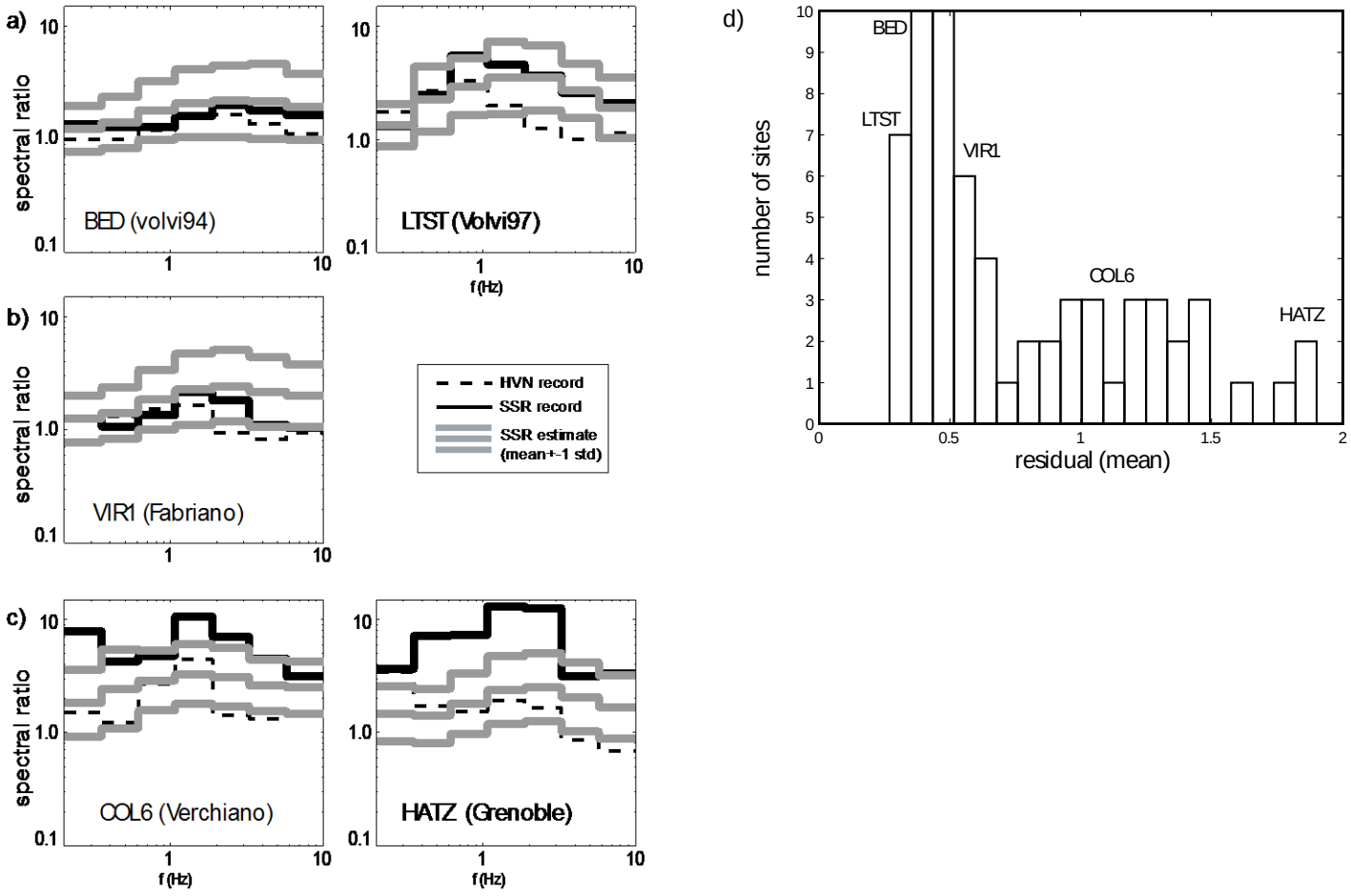

Fig. 5. Comparison of estimated standard spectral ratios (SSRest) from HVN measurements with the recorded ones at sites with increasing mean residuals: (a) lower than 0.3 (0.28 BED and 0.27 LTST), (b) around 0.50 for VIR1 and (c) larger than 1 (1.21 COL6 and 1.90 HATZ). The plot (d) represents the histogram of the mean residuals.

The first two plots in Fig.5a are chosen within the sites having the best fit between estimated and observed curves (BED and LTST have residual values of about 0.3, Fig.5d). The estimated SSR at BED increases the original HVN and reproduce the small amplification at 2-4Hz as observed in the recorded SSR; this behavior is also visible for LTST, where the amplified band in the 0.5-2.5Hz range of SSR (not shown in the HVN) is well fitted.
At larger residual values (0.50 for VIR1, Fig.5d), the reconstructed SSR better fits the recorded values in the 1-3Hz range, but it overestimates them at larger frequencies, increasing the original HVN ratio more than necessary (Fig.5b).
When the differences between SSR and HVN are too large, that is the HVN do not reproduce at all the high amplification of SSR, our method is not able to recover the all complexity of the recorded SSR which stay well above the mean + 1 standard deviation, as shown in Fig.5c and from the residuals increase (1.21 for COL6 and 1.90 for HATZ, Fig.5d).
The residuals of Fig.5d show the average fit for each site, that is how many times the differences are included in one standard deviation. To have an overview on how the proposed method is able to reproduce each bin separately, we plot, for each bin, the cumulative percentage of sites having a given residual value (Fig.6) as defined in Eq. 8. Every recorded bin is within the mean plus/minus one standard deviation of the evaluated SSR for about 70-80% of sites. The best performance is for the frequencies 1.0 to 5.7 Hz (central bins 4, 5 and 6), whereas the fit of the lower frequencies (0.2-0.35 Hz, bin 1) is less accurate.

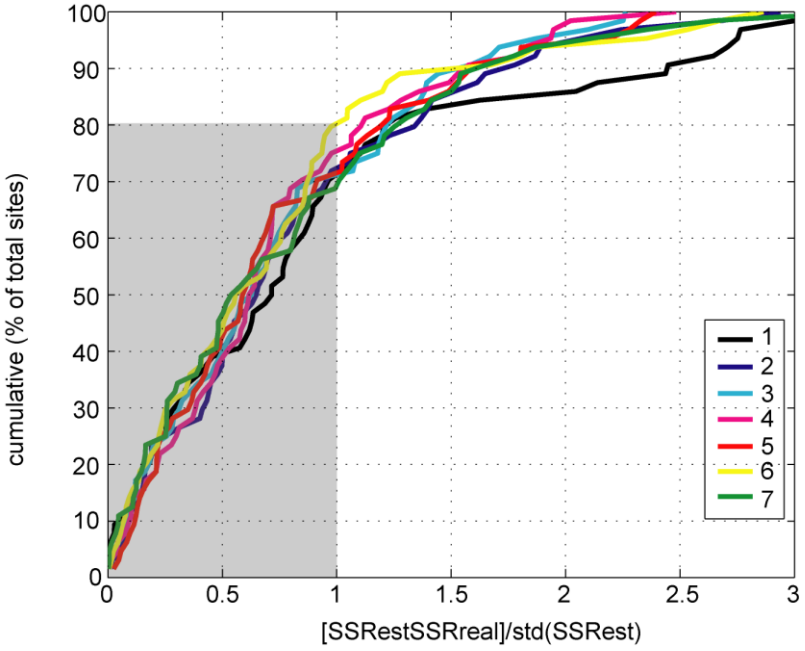

Fig. 6. Cumulative percentage of sites having a given residual value, for each bin separatelly. The gray area indicates the percentage of sites whose recorded SSR bins are within one standard deviation from the evaluated ones.

DISCUSSIONS AND CONCLUSION

In this paper we proposed a strategy based on statistical methods to reconstruct the standard spectral ratios (SSR) at a target site using the spectral ratios on noise measurements (HVN) at the same site.
The statistical analyses on spectral ratios from the different techniques (SSR and HVN) showed that: (i) the behaviour of the entire frequency range of the studied spectral ratios can be represented by few uncorrelated frequency intervals; (ii) the correlation is not limited to a single peaked frequency (fo) and its related amplification (Ao), but it concerns a frequency band which is broader for SSR and narrower for HVN; (iii) the very low frequency band (f<0.35-0.6Hz) does not contribute to the correlation of the other frequencies: the variability of the ratios at low frequencies does not affect the statistical relations. In particular, the presence of low-intermediate frequency peaks in the horizontal-to-vertical spectral ratios followed by a deamplification of the high frequencies is a possible evidence of the presence of the fundamental Rayleigh mode in the investigated wavefield (Fäh *et al*., 2003; Bonnefoy-Claudet *et al*., 2006, and references therein; Bonnefoy-Claudet *et al*., 2008).
The main question for a practical purpose is: how can we use these correlation results to estimate one type of spectral ratio (i.e. SSR)

from another type (i.e. HVN)? To answer it, we performed a cross-validation test on the dataset used for the correlation between SSR and HVN, i.e. we tried to recover the SSR values from the HVN measurements.
The SSR was estimated as a weighted average on the SSR recorded at other sites and the procedure was based on a statistical analysis on a large data set; however, our results are certainly constrained to the specific database we used and cannot be used for sites outside the areas of the selected experiments. More efforts are needed to enrich our database with additional sites representative of a wider geologic variety. For this specific database, we are able to reproduce sufficiently well the recorded SSR for a large number of sites, specifically in the intermediate frequency range (1-6 Hz). In general, the worst cases are related to HVN and SSR recorded at the same site which are too different, i.e. the estimate is not able to reach high real SSR values starting from a recorded HVN too low.
However, the proposed methodology is promising and can be used to extend the SSR results from a group of stations to other sites in the same area.

REFERENCES

Bonnefoy-Claudet, S., Cotton F., and Bard P.-Y. [2006], "The nature of the seismic noise wave field and its implication for site effects studies: a literature review", Earth Sci. Rev. 79, nos. 3–4, 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", Bull. Seism. Soc. of Am, 98: 288 – 300, Feb 2008, doi: 10.1785/0120070063

Cara F., Cultrera G., Azzara R. M., De Rubeis V., Di Giulio G., Giammarinaro M. S., Tosi P., Vallone P., and Rovelli A. [2008], "Microtremor Measurements in the City of Palermo, Italy: Analysis of the Correlation with Local Geology and Damage", Bull. Seism. Soc. of Am., Vol. 98, No. 3, pp. 1354–1372, June 2008, doi: 10.1785/0120060260.

Cultrera G., V. De Rubeis, N. Theodoulidis, P. Bard, H. Cadet [2010], "Statistical investigation of site effects with emphasis on sedimentary basins, using earthquake and ambient noise recordings", SH4/We/O3. European Seismological Commission 32nd General Assembly, Sept 6-10 2010, Montpellier, France.

Davis J.C., [2002], "Statistics and data analysis in geology", 3rd edn. Wiley, New York

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", J. of Seism., Volume 7, Number 4, 449-467, DOI: 10.1023/B:JOSE.0000005712.86058.42

Haghshenas E., Bard P.-Y., Theodulidis N. and SESAME WP04 Team (Atakan K., Cara F., Cornou C., Cultrera G., Di Giulio G., Dimitriu P., Fäh D., de Franco R., Marcellini A., Pagani M., Rovelli A., Savvaidis A., Tento A., S. Vidal, Zacharopoulos S.) [2008], "Empirical evaluation of microtremor H/V spectral ratio", Bull Earthquake Eng, 6:75–108, DOI 10.1007/s10518-007-9058-x

Theodulidis N., G. Cultrera, A. Tento, D. Faeh, K. Atakan, P.-Y. Bard, A. Panou and the SESAME-Team [2004], "Empirical evaluation of the horizontal-to-vertical spectral ratio technique: results from the SESAME project", 13th World Conference on Earthquake Engineering,Vancouver, August 2004.

Theodoulidis N., G. Cultrera, V. De Rubeis, F. Cara, A. Panou, M. Pagani, P. Teves-Costa [2008], "Correlation between damage distribution and ambient noise H/V spectral ratio", Bull Earthquake Eng, 6:109–140. DOI 10.1007/s10518-008-9060-y.