OPTIMIZATION OF ANALYSIS OF AMBIENT NOISE INSTANTANEOUS POLARIZATION FOR SITE RESPONSE INVESTIGATION

V. Del Gaudio
Dipartimento di Scienze della Terra e Geoambientali, Università degli Studi "Aldo Moro", Bari, Italy

Introduction. The most commonly used technique employing ambient noise to investigate site response is the Nakamura's method (Nogoshi and Igarashi,1971; Nakamura, 1989), also known as HVSR (Horizontal-to-Vertical Spectral Ratio) or HVNR (Horizontal-to-Vertical Noise Ratio). According to this technique, site resonance properties can be inferred calculating the mean ratios H/V between spectral amplitudes of horizontal and vertical components of ambient noise recordings. There is a general agreement that, at least in simple geological conditions, this method allows a reliable identification of site resonance frequencies. Furthermore, if site response presents directional variations, an analysis of azimuthal variations of H/V ratios can reveal resonance directivity and maximum amplification directions (Del Gaudio *et al.*, 2013). However, amplification factors appears more problematic to be derived. In principle, one could invert curves of H/V as function of frequency in terms of subsoil models (see Castellaro and Mulargia, 2009), which, in turn, can be used to calculate amplification factors. However, H/V curve interpretation requires the definition of ambient noise composition among different types of waves (body waves, Rayleigh, Love) that, according to site conditions and source characteristics, can differently contribute to the observed signals (cf. Bonnefoy-Claudet *et al.*, 2006).

In general, analyzing ambient noise to characterize site response, it would be desirable to isolate Rayleigh waves within the recorded signal, since they can potentially provide more information, i.e. resonance frequency, amplification directivity and (indirectly, from particle motion ellipticity) amplification factors. Indeed, the presence of a significant or even predominant proportion of Love waves in the horizontal component of ambient noise (cf. Bonnefoy-Claudet *et al.*, 2008) can considerably alters the H/V ratios compared to what would be observed for Rayleigh waves only.

Recently, I proposed a method to derive site response properties from a different kind of ambient noise processing, based on analysis, instant by instant, of ground motion polarization to identify Rayleigh wave packets over which to calculate mean H/V ratios (Del Gaudio, 2013). The implementation of this method requires the definition of a series of procedural and parametric choices. In this study, in order to define some guidelines in such choices, a series of tests were carried out on synthetic signals. These were also used to compare the performance of the new method with that of the classical Nakamura's technique.

Methodology. Instantaneous polarisation properties of an ambient noise recording $u(t)$ can be obtained from its analytic representation, given by

$$u_c(t) = u(t) + j\hat{u}(t) = A(t)e^{j\Phi(t)} \tag{1}$$

where j is the imaginary unit and $\hat{u}(t)$ is the Hilbert transform of $u(t)$. Applying this transformation to all the signal components, ground motion can be represented as the real part of a complex vector, which describes a time-variant elliptical trajectory. Morozov and Smithson (1996) provided a simple method to calculate the semi-axes $\vec{a}(t)$ and $\vec{b}(t)$ of these instantaneous ellipses. Their vectorial product $\vec{p}(t)$ (named planarity vector: see Schimmel and Gallart, 2003), allows defining the attitude of the trajectory plane, whereas the quantity

$$rl = 1 - \frac{|\vec{b}(t)|}{|\vec{a}(t)|} \tag{2}$$

defined "rectilinearity" (Schimmel and Gallart, 2004), can be used as an index of how much instantaneous signal is close to a linear (rl =1) or circular (rl =0) polarization.

The methodology of ambient noise analysis proposed in Del Gaudio (2013), consists of passing the noise recording through narrow-band filters with varying central frequencies ν_c and then calculating instantaneous polarization characteristics from the analytic representation of the filtered time series in order to obtain:

1) instantaneous values of ratios H_{max}/V between the amplitudes of the maximum horizontal component of ground motion and the vertical one;

2) recognition of a preferential signal polarization, possibly reflecting directional site amplification;

3) identification of single data sample whose polarization is compatible with Rayleigh wave particle motion;

4) identification of data sample whose polarization is compatible with Love or SH-waves;

5) calculation of the H_{max}/V average over data samples of Rayleigh type.

The last calculation, carried out for different ν_c values, produces mean H/V ratios derived from instantaneous polarization (named HVIP values), which can be assumed as an estimate of the ratios between horizontal and vertical component of Rayleigh waves at difference frequencies.

The implementation of this procedure requires the choice of a filtering type and of some "threshold" parameters, to identify Rayleigh/Love waves, i.e.: i) the maximum admissible deviations of the planarity vector from horizontality and of vectors $\vec{a}(t)$ and $\vec{b}(t)$ from vertical/horizontal directions (*diplim*); ii) the maximum rectilinearity *rllim* to distinguish elliptical (Rayleigh waves) from linear (Love waves) polarization; iii) the minimum number *nmin* of consecutive data samples that identifies a wave packet of coherent type. The last criterion is justified considering that, within a large number of samples, single isolated cases satisfying the Rayleigh/Love identification criteria could be purely casual. Thus, a more reliable identification of Rayleigh or Love waves require the presence of a certain number of consecutive samples with coherent type of polarization, so that "packets" of Rayleigh/Love waves are identified, rather than just single data samples.

Test implementation. In order to optimize the analysis results, a series of tests were arranged using synthetic signals consisting of time series of 1000 s, sampled with a frequency of 100 Hz. These time series were generated synthetizing signals simulating Rayleigh and Love waves arriving at a recording station from sources randomly located around the station at distances between 100 m and 1 km. For each source, signal includes 50000 harmonic components spaced in frequency by 0.001 Hz, whose spectral amplitudes and phases, assumed equal at the source, are modified as function of wavelength, simulating an anelastic attenuation through a medium with a low quality factors (25-30). To give a transient character to these signals, each of them is modulated through a cosine window having a duration variable between 0.5 to 5 s. These transient signals emerge from a casual background noise of Gaussian type of mean 0 and standard deviation varied, among different synthetics, to obtain different signal-to-noise ratios. This choice was motivated by the results of preliminary tests conducted on real noise recordings (see Del Gaudio, 2013), which showed that only a small fraction of the recordings has a well defined polarization of Rayleigh or Love type. This probably occurs because, for most of the noise recording, signals of different polarization overlap and a signal with a specific kind of polarization can be identified only when it has much more energy than the others.

Rayleigh wave ellipticity of each harmonic component was attributed according to an H/V curve simulating the presence of 40 m thick soft layer with an S-wave velocity of 300 m/s overlying a stiffer bedrock, of 800 m/s. This velocity contrast generates an H/V peak value equal to 3.37 at 1.9 Hz. With regard to ground motion direction, two kinds of situations were simulated, one characterized by polarization controlled by wave propagation direction (i.e. with horizontal component parallel and transversal to this direction for Rayleigh and Love waves,

respectively) and the other characterized by a fixed polarization direction, simulating a site with directional resonance.

Using a purposely written code, six synthetic signals were generated, named: 1) surf100, derived from 50 sources of Rayleigh type and 50 of Love type, both with a ground motion polarized along an azimuth of 37°, summed with an incoherent background noise kept at a very low amplitude level (of the order of 1/1000 of Rayleigh waves); 2) surf100i, differing from surf100 for the variability of the azimuth of surface wave polarizations, depending on the wave propagation direction; 3) surf100sn3, obtained from surf100 by increasing the level of random noise so to reduce the signal-to-noise amplitude ratio (*snr*) to 3; 4) surf100sn3i, derived from surf100i by increasing noise as for surf100sn3; 5) surf100sn1, obtained from surf100 by further increasing noise so that *snr* is equal to 1; 6) surf100sn1i, derived from surf100i with the same noise as surf100sn1.

These signals were analyzed according to the procedure outlined in the previous section to obtain HVIP values calculated at 23 frequencies, spaced by 0.25 Hz, between 0.5 and 6 Hz. The HVIP values were compared with the actual H/V ratios of the Rayleigh waves present in the synthetic signal. During these tests, two aspects were mainly examined, i.e. accuracy and precision of the H/V estimates derived from HVIP values. Accuracy was evaluated from the root mean square of errors rms_{err} (HVIP deviations from actual H/V ratios of Rayleigh waves), precision from the scatter, expressed through the root mean square rms_{sc} of deviations of instantaneous H_{max}/V values from the mean HVIP values.

The code used for the instantaneous polarization analysis calculates both HVIP values averaged among samples polarized in any direction and among samples grouped into azimuth bins spaced by 10°. For signals with equally polarized surface waves (surf100, surf100sn3, surf100sn1), the actual H/V values were compared with HVIP averaged on samples polarized within the azimuth bin 30°- 40°, whereas, for randomly polarized signals (surf100i, surf100sn3i, surf100sn1i), with the HVIP averaged over all the azimuths. These two situations are representative of the cases of directional and isotropic site response, respectively.

Results. A preliminary series of test was aimed at selecting an optimal filtering type to reconstruct the H/V curve as function of frequency. The best results in terms of accuracy of H/V curve estimation were obtained by adopting a Gaussian filtering multiplying spectral amplitudes by a function

$$G(v) = e^{-\frac{(v - v_c)^2}{2\beta^2}}$$ (3)

where v_c is the central frequency and β is a fixed parameter governing the filtering bandwidth.

Using this kind of filtering, each of the six synthetic signals was analyzed adopting different combinations of values for: 1) the filtering parameter β; 2) the threshold *diplim* of angular deviations of $\vec{a}(t)$, $\vec{b}(t)$ and $\vec{p}(t)$ from horizontal/vertical directions; 3) the rectilinearity limit *rllim* distinguishing Rayleigh from Love waves; 4) the minimum number *nmin* of consecutive data samples that identifies a coherent wave packet.

The results of these tests show that the accuracy in H/V curve estimate is better for the directional case than for the isotropic one and deteriorates as *snr* decreases (more rapidly for the isotropic than for the directional one). Error minimization was generally obtained adopting a relatively small threshold for *diplim* (5°) and for *rllim* = 0.80, combined with the highest *nmin* values among those tested (15-20).

All these thresholds corresponds to more restrictive criteria in Rayleigh wave packet identification, taking however into account that a too restrictive criterion can excessively limit the number of Rayleigh type signals over which HVIP values are calculated. This limitation can have a negative effect on the relation between accuracy and precision.

In general, it is desirable to have a good correlation between accuracy and precision so that, in field observations, the latter can give indication about the former. From the tests carried out,
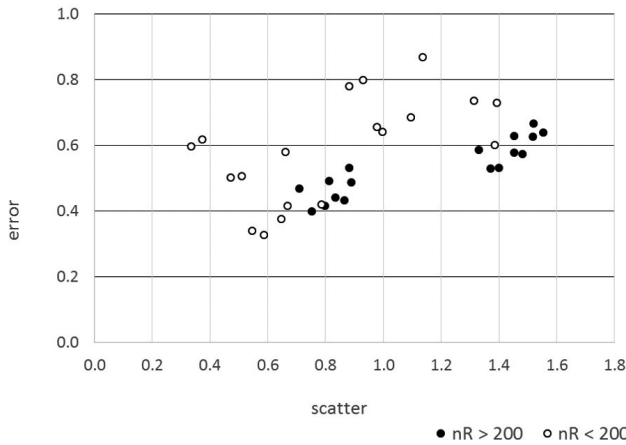
Fig. 1 – Comparison between scatters (root mean square of deviations of instantaneous H_{max}/V values from mean HVIP values) and root mean square of errors (HVIP deviations from the actual H/V values) relative to instantaneous polarization analyses conducted with different parameter combination on the synthetic signal surf100sn1. Open and full circles represent data relative to analyses that classified less or more than 200 samples of Rayleigh type, respectively, at the peak frequency of 2 Hz.

it was found that the rms_{err} values are mostly lower than rms_{sc}, but this may not be true when the number nR of samples identified as Rayleigh-type is extremely low. In such cases, correlation between accuracy and precision can worsen and HVIP curves characterized by smaller scatter may not be those also affected by smaller errors. As an example, Fig. 1 shows the relation between scatter and error values for the case of the signal surf100sn1. If one considers the results of all the analyses, independently on the number nR_{2Hz} of samples used to calculate the HVIP peak value (i.e. for azimuth=30°- 40° and frequency = 2.00 Hz), the correlation between errors and scatters appears rather weak, with minimum scatter found for analysis results affected by relatively large errors. On the contrary, excluding the results obtained with $nR_{2Hz} < 200$, the correlation is much better and the choice of the estimate affected by the lowest scatter would result in an H/V curve close to the optimal one from the point of view of accuracy.

On the other hand, excessively loose criteria of wave identification, which would generate a very large number of samples identified as Rayleigh-type, is not an effective approach. Indeed, this would cause the inclusion in the calculation of HVIP of a considerable number of estimates of H/V strongly scattered around the average, which make the precision a poor estimator of accuracy.

These results suggest that a good criterion to define the analysis parameters is to try different combinations, selecting the one providing the minimum scatter rms_{sc} among the combinations that classify a significant number of Rayleigh-type data samples (at least 200 for the peak frequency). Using this criterion, the percentage of classified samples on the total is correlated to the signal-to-noise ratio characterizing the analyzed recording and the value of scatter rms_{sc} provides an upper bound for the root mean square of errors rms_{err}.

Fig. 2 shows the HVIP curves corresponding to the best estimates, in terms of accuracy and precision, obtained analyzing the six synthetic signals according to the aforementioned criterion. These curves are compared with the actual values of H/V ratio of Rayleigh waves and with the curves obtained from analyses conducted with the classical Nakamura's technique. The latter was conducted subdividing the synthetic time series into 20 s windows, smoothing the spectra of horizontal and vertical components through a triangular average on frequency intervals of ±5% of the central frequency, and averaging the spectral ratios of different time windows after having discarded those showing abnormally high or low H/V ratios throughout the frequency band analyzed. The resulting "HVNR" curves were calculated at azimuth intervals of 10°: for directionally polarized signals (surf100, surf100sn3, surf100sn1) the comparison was made with the curve obtained along the direction N35°E, which showed the largest H/V values, whereas, for the case of isotropic signals (surf100i, surf100sn3i, surf100sn1i), both euclidean and geometric averages between north and east components were calculated.

For those signals whose background noise has amplitude much lower than the signal (surf100, surf100i), the HVIP provides very good estimates of the real H/V ratios of Rayleigh
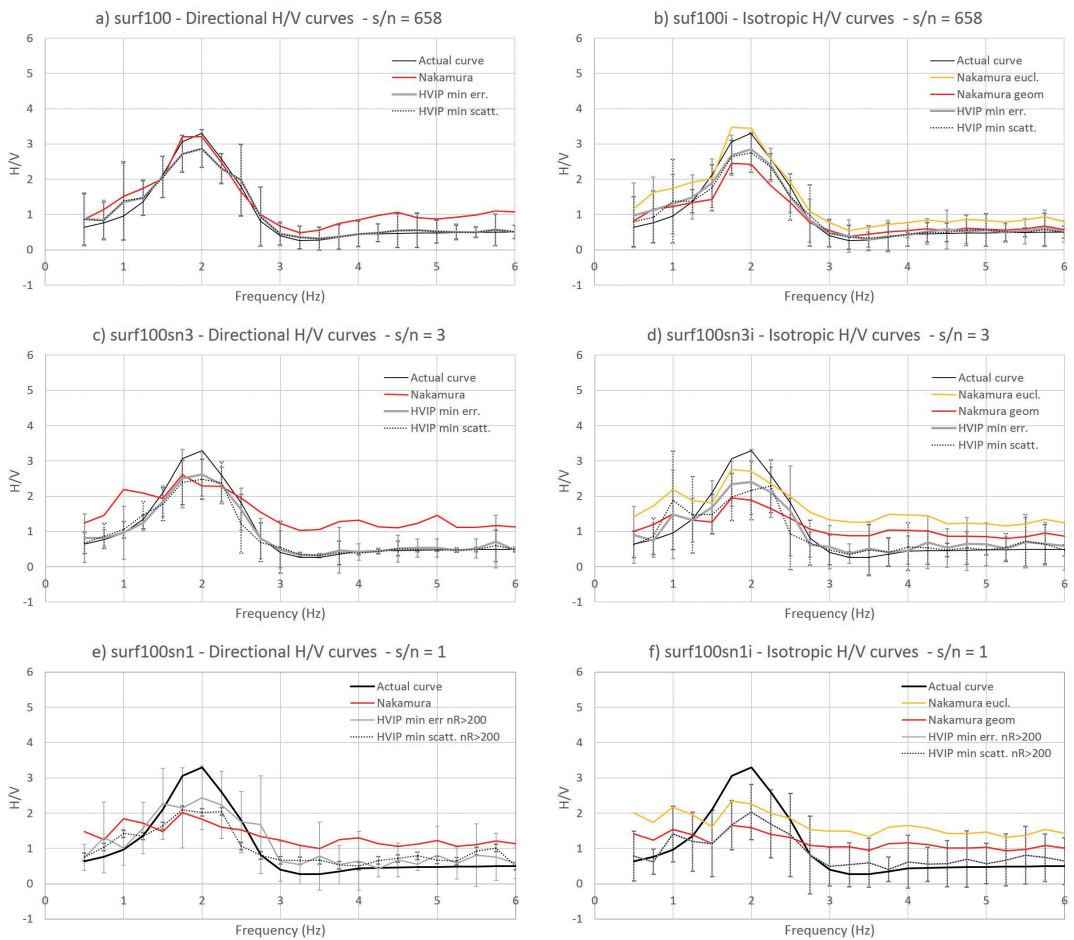
Fig. 2 – Comparison of actual H/V ratio curve of Rayleigh waves present in the synthetic signals (black solid line) with estimates obtained from HVIP analysis and from Nakamura's technique, for different synthetic signals: a) surf100, b) surf100i, c) surf100sn3, d) surf100sn3i, e) surfsn1, f) surf100sn1i. Grey and dotted lines represent the best HVIP estimate in term of accuracy and precision, respectively; red line for the case of directionally polarized signal, represent the HVNR curve obtained for the azimuth (35°) characterized by the highest spectral ratio values; red and yellow curves, for the case of isotropic signals, represent the HVNR values obtained from North and Est signal components through euclidean and geometric averages, respectively. Vertical bars represent the scatter of HVIP estimates at each analysed frequency.

waves both for the case of directionally and isotropically polarized signals. Selecting the analysis parameters that maximize precision (minimizing the scatter rms_{sc}), the resulting HVIP curves are coincident with (for surf100) or very close to (for surf100i) that minimizing errors. The rms_{err} value obtained for these curves are 0.18 and 0.21, respectively for the directional and isotropic cases, whereas the scatter rms_{sc} is 0.42 and 0.41, respectively. The HVIP curves correctly identifies the peak frequency, estimates H/V peak value with a 13-17% error and reproduces well the general shape of the H/V curve, including the lower part where the H/V ratios are below 1. Comparatively, the outcome of Nakamura technique provides a better estimate of the peak value for the directional case (3% error), but larger deviations at higher frequencies, so that rms_{err} is higher (0.39). In the case of isotropic signal (surf100i), the HVNR calculated through the Euclidean average gives a better estimate of the peak value (5% error), whereas geometric average better fits the lower part of the H/V curve, but both show overall a higher rms_{err} than HVIP (0.41 and 0.35, respectively).

For these signals, the HVIP analysis identifies a very high number nR_{2Hz} of Rayleigh-type samples at the peak frequency of 2 Hz: it is in the order of 20000, i.e. about 20 % of the entire recording. Their analysis points out very clearly the directional polarization of the signal surf100, since more than 98% of these samples is concentrated within the azimuth bin (30°-40°) containing the correct value (37°).

Increasing the noise amplitude, the discrepancies between HVIP and real H/V increase, but for $snr = 3$ the agreement is still good, especially for the case of directional polarization (surf100sn3). It shows an increase to 20-25% of the error in the estimate of the H/V peak value, but the estimate maximizing precision ($rms_{sc} = 0.36$) is still very close to that minimizing errors ($rms_{err} = 0.27$ and 0.22, respectively). More large discrepancies are found in case of isotropically polarized signals (surf100sn3i). Indeed the most accurate estimate, even correctly identifying the peak frequency, gives an error of 27% in the estimate of H/V peak value and $rms_{err} = 0.33$, whereas the curve having minimum scatter ($rms_{sc} = 0.51$) shows a maximum at 2.25 Hz, an error of 34% in the H/V peak estimate and $rms_{err} = 0.46$. In such cases nR_{2Hz} is in the order of 1000-1500, i.e. about 1-1.5% of the entire recording, which is more close to the classification percentages found in preliminary tests on real noise recordings, thus these cases can be considered more representative of conditions actually occurring in the field applications. In the case of directionally polarized signal, the identification of site response directivity orientation is still very clear, since a pronounced maximum, larger than 50%, is found in the polarization distribution for the 30°- 40° azimuth bin.

Comparatively, HVNR values derived from Nakamura's method provide considerably worse estimates, both in the directional case ($rms_{sc} = 0.74$) and in the isotropic one ($rms_{sc} = 0.80$ and 0.62, using Euclidean and geometric average, respectively). In general, in the HVNR curve, the peak at 2 Hz is not very clear as effect of the concomitant underestimate of the peak value and overestimate of the lower part of the H/V curve (which, on the contrary, the HVIP fits very well)

If the noise amplitude is comparable to that of the Rayleigh signals, the number of samples identified as Rayleigh-type decreases and the estimates undergo a further deterioration, especially with regard to the H/V peak values, whose estimate errors reach values of 30-40% and 40-50% for the directional (surf100sn1) and isotropic (surf100sn1i) signals, respectively. However, although with underestimated peak values, some general characteristics of the H/V curve can still be recovered. Indeed, the HVIP curves show a single major peak around 2 Hz and, although underestimating the maximum by up to 40%, reproduce quite well the rest of the curves. With regard to site response directivity, even in this less favorable noise conditions, it can be recognized from the sample polarization distribution: a maximum of concentration (up to 20-30%) of polarizations around azimuths differing by not more than 10° from the actual direction of Rayleigh wave ground motion is found at almost all the examined frequencies.

Comparatively, using the Nakamura's technique, the peak at 2 Hz is practically undistinguishable from the HVNR curves. Apart from the considerable underestimate of the H/V ratio maximum, these curves systematically fail in recognizing the presence of H/V ratios < 1 at higher frequencies.

Conclusions. A series of tests carried out on synthetic signals simulating ambient noise recordings and including a mix of transient Rayleigh and Love waves of known characteristics together with a casual Gaussian noise of different amplitude, showed that the analysis of instantaneous polarization allows extracting Rayleigh wave properties with a good level of approximation. For this purpose, signals are first passed through band-pass filters with different central frequency. Then, on the resulting time series, Rayleigh wave packets are identified when a minimum number of consecutive data samples are found to exceed optionally defined thresholds of angular deviation of the plane of instantaneous elliptical trajectories from verticality and of ellipse major/minor axes from horizontal/vertical direction. For each sample of these wave packets, an estimate of the instantaneous ratio H_{max}/V can be obtained, whose averages HVIP

provide an estimate of the curve of Rayleigh wave H/V ratios as function of frequency with an accuracy level depending on the signal-to-noise ratio *snr*.

Test results suggest that, since the estimate accuracy changes with the choice of filtering and identification parameters, one should make trials with different parameter combinations, selecting the one providing the minimum scatter of H_{max}/V around HVIP values, among the combinations that identifies a sufficiently high number of Rayleigh-type data samples (at least 200 for H/V peak values). Analyzing signals with *snr* = 3 or more, it was possible to correctly identify peak frequency and, when present, site response directivity orientation. The estimate of the H/V peak turned out affected by an error increasing with the noise level from 13% to 34%, the results being better for the case of directionally polarized signals in comparison to isotropic signals. Even when the background noise has an amplitude comparable to that of Rayleigh waves, it is still possible to recognize the presence of an H/V peak, its frequency and (if present) directivity, although the H/V peak ratio was considerably underestimated.

Comparatively, applying the classical Nakamura method to the same signals, the estimate of the H/V curve was always found less accurate and, in case of very high noise level, the H/V peaks were totally unrecognizable. The lower performance of the Nakamura technique was mainly due to the poor capacity of the HVNR curve to reproduce the lower part of the H/V curve (especially for H/V < 1), which, on the contrary was always outlined by the HVIP curve with good or excellent approximation.

References

Bonnefoy-Claudet S., Cornou C., Bard P.-Y., Cotton F., Moczo P., Kristek J. and Fäh D.; 2006: *H/V ratio: a tool for site effects evaluation. Results from 1-D noise simulations*, Geophys. J. Int., **167**, 827–837.

Bonnefoy-Claudet S., Köhler A., Cornou C., Wathelet M. and Bard P.-Y.; 2008: *Effects of Love Waves on Microtremor H/V Ratio*, B. Seismol. Soc. Am., **98**, 288–300.

Castellaro S. and Mulargia F.; 2009: *VS_{30} Estimates Using Constrained H/V Measurements*. B. Seismol. Soc. Am., **99**, 761–773.

Del Gaudio V.; 2013: *Analysis of ambient noise instantaneous polarisation for site response characterisation*. Proceedings XXXII Congress GNGTS, Trieste, 19-21 November, 2013, 226-234.

Del Gaudio V., Wasowski J. and Muscillo S.; 2013: *New developments in ambient noise analysis to characterise the seismic response of landslide prone slopes*. Nat. Hazards Earth Sys. Sci.*,* **13**, 2075-2087, doi:10.5194/nhess-13-2075-2013, 2013.

Morozov I. B. and Smithson S. B.; 1996: *Instantaneous polarization attributes and directional filtering*. Geophysics, **61**, 872-881.

Nakamura Y.; 1989: *A method for dynamic characteristics estimation of subsurface using microtremors on the ground surface*. Q. Rep. Rail. Tech. Res. Inst., **30**, 25–30.

Nogoshi, M. and Igarashi T.; 1971: *On the amplitude characteristics of microtremor* (part 2) (in japa-nese with english abstract). J. Seismol. Soc. Jpn., **24**, 26-40.

Schimmel M. and Gallart J.; 2003: *The use of instantaneous polarization attributes for seismic signal detection and image enhancement*. Geophys. J. Int., **155**, 653–668.

Schimmel M. and Gallart J.; 2004: *Degree of Polarization Filter for Frequency-Dependent Signal Enhancement Through Noise Suppression*. B. Seismol. Soc. Am., **94**, 1016–1035.