



Site Effects Assessment Using Ambient Excitations

SESAME

**European Commission – Research General Directorate
Project No. EVG1-CT-2000-00026 SESAME**

Final report

WP13

**Recommendations for quality array
measurements and processing**

Deliverable D24.13

March 2005

List of Contents

1. Introduction
2. Dispersion curve estimate
3. Inversion of dispersion/autocorrelation curves
4. Conclusions

List of Contributors (in alphabetical order after the last name)

Denis Jongmans	ULg, Liège, Belgium
Matthias Ohrnberger	UP, Potsdam, Germany
Marc Wathelet	ULg, Liège, Belgium

Project coordinator:	Pierre-Yves Bard	LGIT, Grenoble, France
WP13 Leader:	Denis Jongmans	ULg, Liège, Belgium

1. Introduction

Array measurements of ambient vibrations have been increasingly used during the last decade, particularly in Japan (Tokimatsu, 1995, Chouet et al. 1998, Bettig et al. 2001, Satoh et al. 2001, Asten et al. 2004). Given the assumptions that the site structure is horizontally stratified and that the ambient vibrations are predominantly made of surface waves, array measurement analysis allows retrieving the dispersion curves of surface waves (Tokimatsu, 1995). Major advantages of these techniques are that they can be applied in urban areas and that they are able to investigate deep soil properties (down to several hundred meters or more), according to the low frequency content of the seismic noise (Horike 1985, Ishida et al. 1998, Miyakoshi et al. 1998, Yamamoto 1998, Scherbaum et al. 2003). These two features are of particular interest in site effect assessment, as numerous big cities in seismic areas (Mexico City, Los Angeles, Caracas, Tokyo...) are built on thick soil layers.

Although the noise array measurements appear to be a very attractive tool for site characterisation, very little work has been done on the reliability of the method to obtain a correct estimate of the dispersion curve. This is a major issue as the V_s structure is obtained from the inversion of this dispersion curve. Uncertainties on the V_s values are directly linked to the errors on the phase velocity values and to the frequency range on which the dispersion curve is retrieved. One of the main purposes of the SESAME project was to provide recommendations to enable high-quality dispersion curve estimates from ambient vibration array measurements.

The first part of this paper thoroughly investigates the influence of the experimental conditions (instrumental behaviour, array geometry) and of the processing techniques (frequency wavenumber methods and the spatial autocorrelation technique) on the determination of the dispersion curve, identifying sources of possible errors and corresponding uncertainties. A measurement and processing strategy is proposed to overcome some of the problems involved

The second part is dedicated to the inversion process itself. Key elements to consider during this step to obtain reliable V_s values are discussed. Of major importance is to correctly identify and interpret the dispersion curves in terms of surface wave modes. Any misinterpretation (confounding a higher mode with the fundamental one, for instance) will result in erroneous models, which may have a dramatic influence on site effect assessment. New software based on the Neighbourhood Algorithm (Sambridge, 1999) was developed to invert the dispersion curve outcoming from the array analysis of ambient vibrations. The derived V_s profiles have to be validated as much as possible by other geophysical or geological data. Geotechnical investigations are usually numerous in urban areas and these data can be used as prior or posterior information. In this paper, a particular attention is paid to the introduction of a prior information in the inversion of the dispersion curves, which contributes to the reduction of the number of possible solutions.

2. Dispersion curve estimate

In order to allow a reliable determination of the velocity structure itself by any inversion algorithm, we require a dispersion curve estimate of acceptable quality within a broad frequency range. It is important to mention, that the estimation of the dispersion characteristics is not just related to technical issues. The detectability of phase velocity curves within specific frequency bands is inherently coupled with the propagation characteristics at the site and the nature of ambient vibration sources. Furthermore, significant influences of the experimental conditions (instrumental behaviour, instrumental layout) and of the analysis procedure (wave-number methods vs. autocorrelation) on the final outcome of the dispersion curve estimate have to be taken into account. Although the interplay of all these components and their effects on phase velocity estimates are complex, in the following we outline specific problems of dispersion curve estimation from ambient vibration measurements. The purpose of this procedure is to identify sources of possible errors, corresponding uncertainties and to give suggestions how to overcome some of the problems involved by proposing an appropriate measurement and processing strategy.

Applied analysis methods

For the analysis of ambient vibration wavefields, we have considered several frequency wavenumber techniques as well as a modified form of the spatial autocorrelation autocorrelation technique (Aki, 1957). The modification developed in Bettig et al. (2003) regard the relaxation of the strict geometrical requirements in Aki's original work and allows the computation of autocorrelation curves for arbitrary array geometries. Below, we distinguish the following methods:

CVFK: a conventional semblance based frequency-wavenumber method after Kvaerna and Ringdahl (1986), evaluated in sliding time window manner and narrow frequency bands around some center frequency. The coherence estimate is given by:

$$\text{RP}(\omega, \vec{k}) = \frac{\sum_{l=1}^L \left| \sum_{i=1}^N X_i(\omega_l) e^{i\omega_l \vec{k} \vec{r}_i} \right|^2}{\sum_{l=1}^L \sum_{i=1}^N \left| X_i(\omega_l) e^{i\omega_l \vec{k} \vec{r}_i} \right|^2} \quad \text{Eq. 1}$$

where $X_i(\omega_l)$ are the complex Fourier coefficients of the observed signals at stations i , ($i = 1, \dots, N$), at discrete frequencies ω_l , ($l = 1, \dots, L$). The phase shifts $e^{i\omega_l \vec{k} \vec{r}_i}$ account for the delay times related to the horizontal wavenumber \vec{k} from which the direction θ and the horizontal slowness s can be derived as:

$$\theta = \arctan\left(\frac{k_y}{k_x}\right) \text{ and } s = \frac{|\vec{k}|}{\omega} \quad \text{Eq. 2}$$

In order to obtain the propagation characteristics of the most coherent plane wave arrival, a grid search over the wavenumber plane is performed. For all f-k approaches we use a wavenumber grid layout, which is sampled equidistantly in slowness and azimuth.

CAPON: the high-resolution frequency wavenumber approach after Capon (1969) is based on the evaluation of the cross spectral matrix $R(\omega) = E \langle X(\omega)X^H(\omega) \rangle$ (CSM), where $E \langle \cdot \rangle$ denotes the expectation value. The CSM is estimated by a block-averaging technique of the observed signal contributions at the array stations $X(\omega) = [X_1(\omega), \dots, X_N(\omega)]$ for a given target frequency. The estimator is constructed to minimize the spectral leakage in the wavenumber domain and is given by:

$$P(\omega, \bar{k}) = \frac{1}{A(\omega, \bar{k})^H R(\omega)^{-1} A(\omega, \bar{k})} \quad \text{Eq. 3}$$

$A(\omega, \bar{k}) = [e^{jk_1 \bar{r}_1} \dots e^{jk_N \bar{r}_N}]^T$ are the steering vectors for wavenumbers \bar{k} at frequencies ω and the superscript H denotes the Hermitian conjugate. Due to its improved resolution capabilities, this f-k technique is the most widely used within the context of microtremor analysis (Tokimatsu 1997, Kind et al., 2005).

SPAC and modified SPAC (MSPAC): The frequency wavenumber techniques are based on the assumption of the validity of the plane wave signal model. In contrast, the spatial autocorrelation method (SPAC, Aki, 1957) bases its theoretical foundation on the precondition of a stochastic wavefield which is stationary in both time and space. Aki (1957) showed, that, given this assumption, the existing relation between the spectrum densities in space and time can be used to derive the following expression:

$$\bar{\rho}(r, \omega) = J_0\left(\frac{\omega r}{c(\omega)}\right) \quad \text{Eq. 4}$$

$\bar{\rho}(r, \omega) = \int_0^\pi \rho(r, \omega, \theta) d\theta$ represents the azimuthally averaged spatial autocorrelation $\rho(r, \omega, \theta)$ for station pairs separated by distance r and the inter-station direction θ , and J_0 denotes the Bessel function of the first kind and zero-th order. The above relation allows deriving the single valued phase velocity $c(\omega)$ at a given frequency ω by inverting the observed averaged spatial autocorrelation coefficients. Aki (1957) suggested the use of dense semicircular array deployments to readily obtain these autocorrelation coefficients for various radii and target frequencies and applied this technique for the analysis of surface wave dispersion characteristics from microtremor recordings.

Bettig et al. (2001) suggested a modification of Aki's original SPAC formula which allows applying the spatial autocorrelation method for less ideal experimental array configurations. The modification concerns the evaluation of the averaged spatial autocorrelation coefficients from station pairs taken from rings of finite thickness $r_1 \leq r \leq r_2$ instead of using a fixed radius r . The modified formula for the averaged autocorrelation coefficient is:

$$\bar{\rho}(r_1, r_2, \omega) = \frac{2}{r_2^2 - r_1^2} \int_{r_1}^{r_2} r J_0\left(\frac{\omega r}{c(\omega)}\right) dr = \frac{2}{r_2^2 - r_1^2} \frac{c(\omega)}{\omega r} \left[r J_1\left(\frac{\omega r}{c(\omega)}\right) \right]_{r_1}^{r_2} \quad \text{Eq. 5}$$

Practically, the values for minimal and maximal radii r_1 and r_2 are determined from displaying the co-array configuration for the array geometry and selecting station pairs with similar inter-station distances and good azimuthal coverage for the computation of the averaged autocorrelation coefficients.

Error analysis

Potential sources of error can be attributed to one or a combination of the following three issues: recording, processing and nature of ambient vibration wavefield. Considering the use of common array analysis techniques (f-k or SPAC) for the determination of dispersion curves, the main source of uncertainties and/or bias in the estimation procedure can be associated with relative arrival time errors between any pair of sensors within the array setting. Thus, the occurrence of timing errors may be related to different sources, such as: instrumentation, phase determination, violation of the plane wave signal model. Furthermore, we can distinguish between systematic and random time shifts at individual stations or between stations.

Instrumentation effects on dispersion curve estimates (recording)

Systematic time shifts can stem from unlocked GPS time synchronization, from internal clock drifts of digitizers, from station positioning errors or can be caused by differences between phase delays of individual sensors. Time synchronization errors or station positioning deviations clearly can be avoided, however, even in case of occurrence those timing errors are presumably small (a few milliseconds at most). Still the difference of phase delays between sensors due to small relative calibration inaccuracies can be unexpectedly large, even for a homogeneous set of instruments, as shown here after.

In order to demonstrate this effect, we show the results of a simulation experiment. We have evaluated the frequency dependence of phase delay differences for a pair of sensors for small perturbations of the individual seismometer constants f_0 and h . For the numerical experiment we chose a typical short-period instrument used for ambient vibration measurements with $f_0=1\text{Hz}$ and critical damping $h=0.707$. The perturbations in f_0 and h were drawn from a normal distribution around the ideal parameters and a relative standard deviation of 1% (well calibrated sensors) and 5% (fair calibrated sensors). The simulation results are depicted in Fig. 1. Especially around the corner frequency of the instruments relative phase delay differences of up to 15 ms (1% case) and 70 ms (5% case) can be observed.

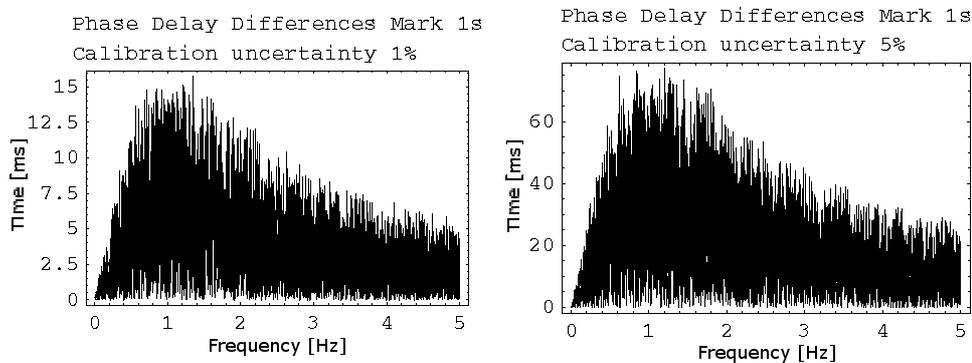


Figure 1: Phase delay differences between sensor pairs caused by slight variation of seismometer constants. Left panel: 1% standard deviation in f_0 and h , right panel: 5% standard deviation in f_0 and h .

The impact of these potential relative (and constant) time delays between sensors on the final phase velocity estimates is difficult to predict, but certainly they are not negligible. The amount of bias introduced depends on the array geometry (inter-station distances and number of sensors involved) and the direction of the incident wavefield.

In order to keep the bias introduced by instrumental effects to a minimum, we recommend using only well-calibrated instruments and a sufficient number of sensors (at minimum 5) for ambient vibration array measurements. Additionally, it is preferable to use sensors with corner frequencies well below the frequency band of interest at a given site, as the amount of phase delay differences will diminish for frequencies well above the corner frequency of the instruments (compare Fig. 1).

Random arrival time delays and related uncertainties (recording and processing)

Besides systematic timing errors as described above, we also expect the occurrence of random relative time shifts. Those timing errors will be inevitably caused during the process of estimating the relative travel time differences (usually performed in frequency domain). Indeed these uncertainties occur independent of the applied analysis method (f-k or spatial autocorrelation methods), whereas their magnitudes are related to the signal to noise ratio. We define this in the context of ambient vibration measurements as energy ratio of coherent vs. non-coherent wavefield components. In order to assess the impact of random phase errors at individual stations of the array onto the phase velocity estimates, we used a bootstrapping approach.

For a fixed virtual array setup (5 stations and a radius of 30 m), we first constructed an ideal plane wave signal by superposing phase shifted sine waves for a target slowness vector at each individual sensor. The forward modelling has been realized in the frequency domain within a narrow frequency band and a small number of discrete Fourier coefficients. This ideal wave front has then been disturbed by adding random phase shifts to the signals at each station. The random time delays were drawn from a normal distribution with zero mean and varying standard deviations. Finally, we analysed the perturbed signal with the conventional frequency-wavenumber algorithm and picked the maximum of the resulting slowness map as estimate of the true slowness vector (1.2 s/km).

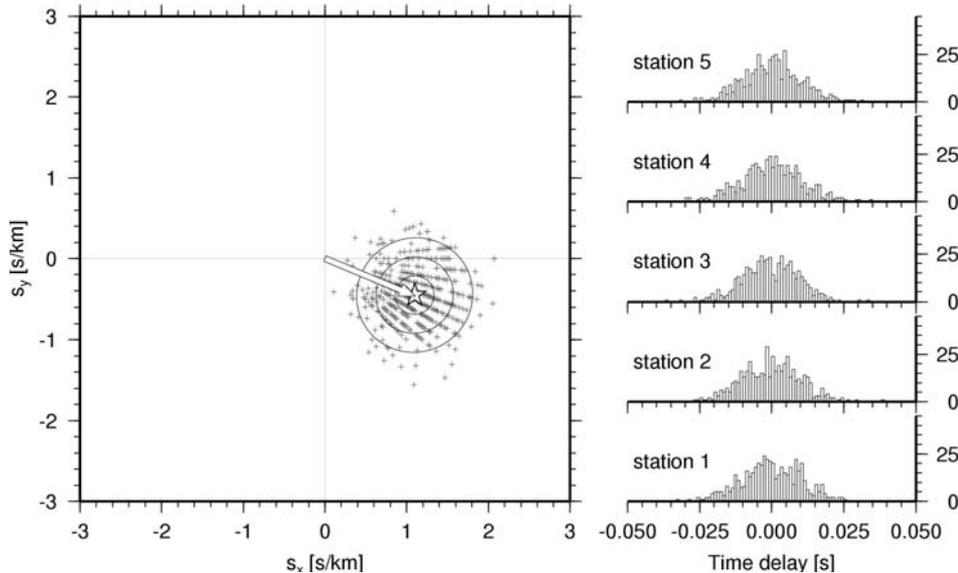


Figure 2: Scatter plot of estimated slowness vectors (cross symbols) obtained by a semblance based conventional frequency-wavenumber algorithm from 500 perturbed plane wave fronts. The time delays imposed at each of the five stations were drawn from a normal distribution with zero mean and 10 ms standard deviation. The distribution of the samples is shown for each station to the right. Due to the circular array geometry (5 stations, radius 30m) the distribution of slowness estimates resembles closely a two-dimensional Gaussian with nearly equal variances along the principal axis. Error ellipses for 1, 2 and 3 standard deviations are shown as black contour lines. The average slowness vector estimate is given by the white star. The input slowness vector (indicated by the white arrow) has an absolute slowness value of 1.2 s/km.

The overall distribution of the bootstrap samples resembles the shape of the array response pattern. Using a highly symmetric configuration, the distribution of estimates has a nearly circular shape and can be approximately described by a two-dimensional Gaussian.

Thus, in order to summarize the tests, for each bootstrap run with fixed aperture and target slowness we fitted a Gaussian distribution and recorded the average of the standard deviations along the principal axes from these. In this way, we quantify the uncertainty of slowness estimates due to the defined standard deviation of the imposed random time delays and its dependence on array size and

slowness of the arriving wave front. In Figure 3, we display the obtained results as shaded contour plots.

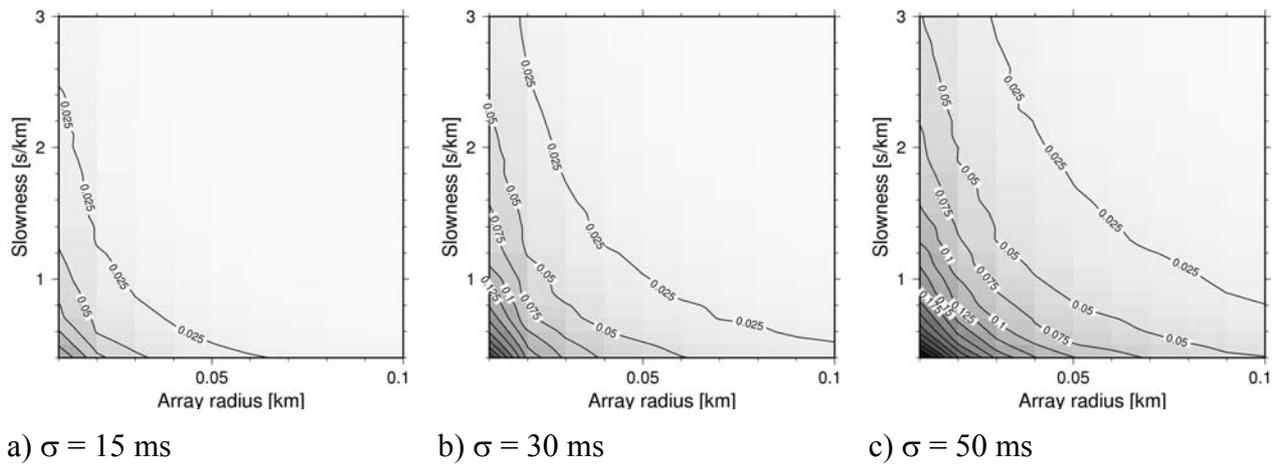


Figure 3: Summary of bootstrap tests linking the random time delay error at each station to the resulting uncertainty in slowness estimates. Standard deviations of random time delays were chosen as a) 15ms, b) 30ms, and c) 50ms. Shaded maps depict the value of uncertainties obtained from a bootstrap experiment for a specific array radius and target slowness combination (contour lines at 0.025 s/km intervals).

From the simulations we can draw the following conclusions: i) the mean of the slowness estimate distribution is an unbiased estimate of the true slowness; ii) the uncertainty of estimates scales directly with the array aperture and the target slowness, in particular the error increases with decreasing array size and faster propagating wave arrivals (smaller slownesses).

It must be noted, that in the simulation we have only considered a single wave arrival. Thus, the evaluated effect of random time delays on uncertainties in the slowness determination seems to be an optimistic outcome compared to real situations. Nevertheless, we consider the results as a lower bound of uncertainties which have to be taken into account in any field experiment. From the conclusions drawn above, we can derive some recommendations in order to improve the phase velocity estimates: i) for stationary time series, the amount of phase delay errors introduced from estimating the signal phase in the frequency domain can be reduced by considering the usage of long analysis windows (order 25 to 50 times the centre period); ii) for a reliable estimate of the slowness, a large number of analysis windows should be considered allowing for a consistent average phase velocity measure and additionally to provide rough uncertainty estimates.

Curved wavefronts and related uncertainties (wavefield, recording, processing)

In practical terms, the situation investigated above can be considered as the observation of a wavefront arrival showing small undulations (e.g. due to phase estimation errors or small lateral heterogeneities of the site). By assessing the impact on slowness estimates, we were actually investigating the stability of the algorithm to provide correct results for those randomly perturbed wavefronts. In addition, if the sources of ambient vibrations are close to the array, a more deterministic distortion of the wavefront is expected from its curvature. In this case, additional predictable time delays are encountered at individual stations. As all standard frequency-wavenumber analysis algorithms are based on the plane wave model, in this case the estimated slowness will be biased (Almendros et al., 1999).

Again, we tried to quantify the corresponding effects on phase velocity estimates by a numerical experiment. Compared to the previous experiment, only a slight modification of the forward modelling part is needed. The exact curved wave fronts are calculated using the arrival times from individual sources located at given distances and the backazimuths measured relative to the array centre. No extra random time delays are added, but for each run all source azimuths are considered

(5° resolution). The narrowband Fourier spectra of the curved wavefronts are then analysed either using the semblance based conventional frequency wavenumber algorithm or by computing the average autocorrelation coefficient according to Aki (1957). To allow for a better comparison between the results obtained from the CVFK and the SPAC methods, the autocorrelation values are further converted to slowness estimates (Eq. 4) using a simple grid search technique. For each simulation, we can compare the true slowness with the estimated one.

In the numerical experiment we used the same pentagon shape array configuration as before for the CVFK analysis, whereas for the SPAC analysis method an 18-station virtual semi-circular array setting is utilised (Fig. 4).

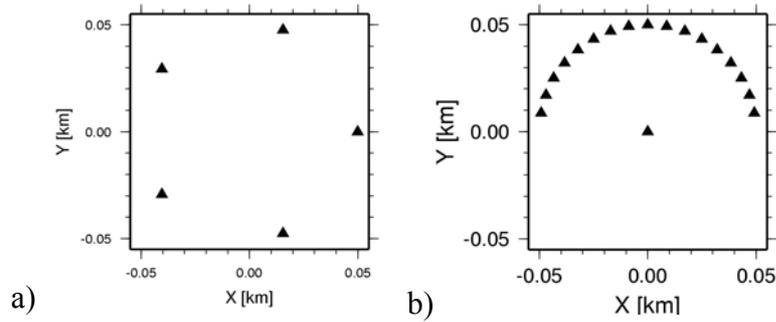


Figure 4: Array geometries used in the numerical experiment for assessing the bias introduced by the occurrence of curved wavefronts. a) pentagon shape for CVFK analysis. b) 18-station semicircular shape for SPAC analysis. During the experiment, the array sizes and source distances have been varied.

In Figure 5 we depict the analysis results for incident curved wavefronts. Each box represents the summary of one test run for a fixed array size and a given source distance but variable azimuths. The true slowness is plotted vs. the estimated slowness value for each individual sample of the experiment (source azimuth). The colour shading of cross symbols has been selected to be proportional to the relative slowness deviation. The ratio between source distance and array radius (d/r) is given in the lower right corner of each box.

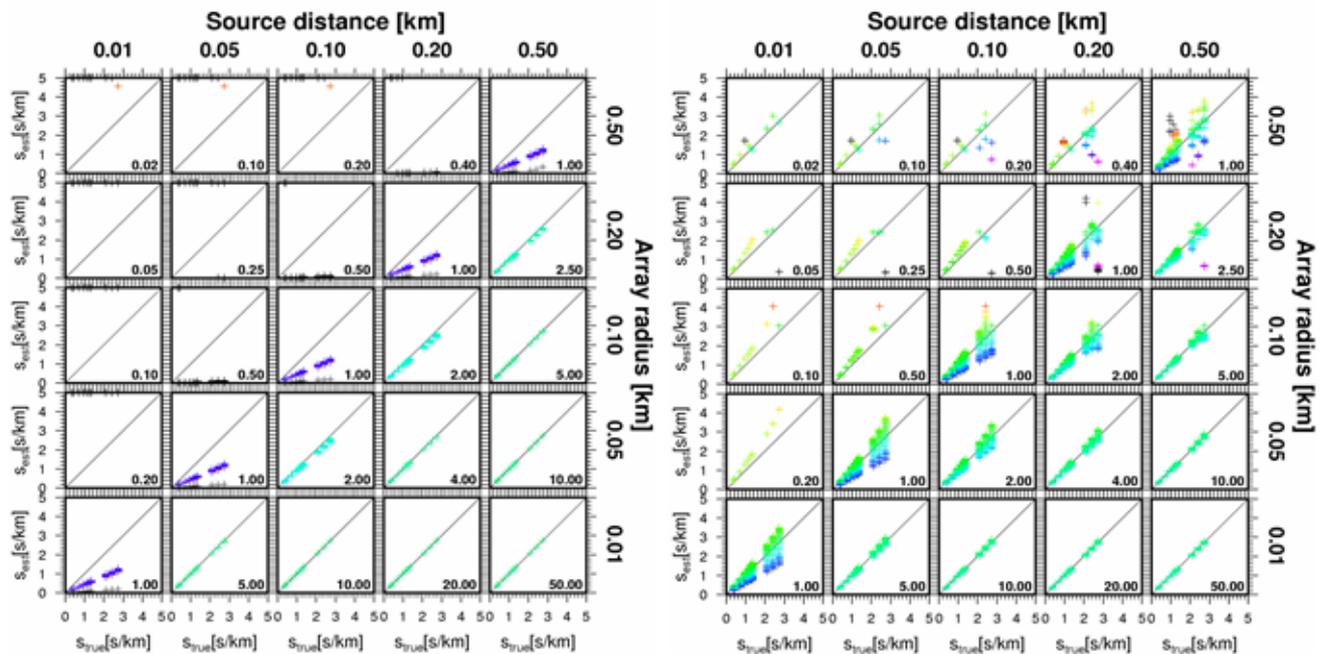




Figure 5: Bias of slowness estimates due to the existence of curved wavefronts in the wavefield. In the individual boxes, the true slowness is plotted vs. the estimated slowness value for different array sizes and source distances. Left panel: results of CVFK analysis. Depending on the ratio of source distance to array radius (value given in the lower right corner of each box), a consistent bias to lower slowness values can be observed. Right panel: results from SPAC analysis. Until a source distance to array radius ratio of 1, the SPAC estimates of the slowness show a symmetric scatter around the true slowness value. For details see text.

From Figure 5, it is clearly recognized that the CVFK results show a significant bias which is introduced by mapping the original curved wavefronts to the underlying plane wave model. The estimated slowness is consistently lower than the true slowness and its magnitude is related to the ratio d/r . For ratios d/r up to 2.5, the deviations are as large as 20%. Ratios below or equal to one (sources inside the array configuration) do not allow any reasonable estimate from the f-k analysis. In comparison, the results of the SPAC analysis show a spread of estimated slowness values in dependence of the azimuth of the incident wavefronts. This spread increases for decreasing values of d/r . However, down to ratios of one, the average of computed slowness values in this experiment still allows a good estimate of the true propagation characteristics. Finally, sources inside the array configuration ($d/r < 1$) apparently give too high estimates of the true slowness value.

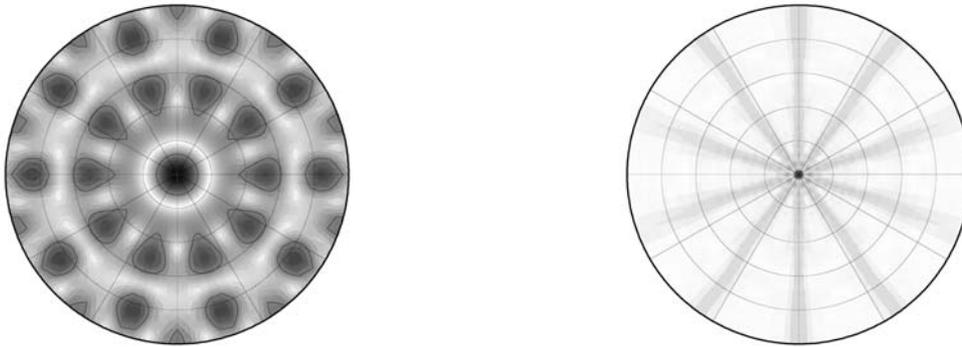
Array geometry, resolution and aliasing (wavefield, recording, processing)

An important issue in any array measurement is the question of optimal geometry (size, shape, number of sensors, etc.) regarding the resolution limits and spatial aliasing effects. This topic has been treated both for large aperture arrays in the context of earthquake/explosion monitoring (Haubrich, 1968, Harjes, 1990) as well as in the field of small aperture array for ambient vibration processing (Woods and Lintz, 1973, Asten and Henstridge, 1984, Kind, 2002). The main differences between these different fields of application can be attributed to the signal components of interest which are to be analysed. Whereas in the field of earthquake detection, arrays are usually optimized for undispersed broadband transient signal arrivals (body wave detection) from relatively distant sources (plane wavefront), for ambient vibration surveys we have to consider a mostly random wavefield caused by nearby superficial sources which has to be analysed within narrow frequency bands. Therefore, within the SESAME project, the matter of optimal array layouts has been re-evaluated focusing on the specific task of determining the dispersion curve characteristics from microtremor recordings.

Traditionally, the performance of a given array geometry is evaluated by the array response pattern (also known as beam pattern) commonly depicted for a vertically incident plane wave (see Figure 6). Several (weak) criteria can be given to judge the quality of an array layout from its response. First, for a good resolution of waves arriving at close wavenumbers, the central peak should be “as sharp as possible”. Secondly, the aliasing peaks should be “as separated as possible” from the central peak when evaluating the array response for the frequency band for which the analysis will be performed. Finally, the ideal response pattern for an array should show similar suppression capabilities for all directions of arrivals (DOAs).

On basis of these considerations, Capon (1969) derived his well-known high-resolution f-k estimator (HRFK) by optimizing the windowing function such, that the wavenumber response approaches a delta-function in the f-k domain. Capon’s estimator has proven to be not only theoretically appealing, but also practically well-suited for array analysis tasks. Compared to conventional f-k algorithms, the gain in slowness resolution is traded against the lack of time resolution caused by the estimation procedure (termed block averaging by Capon, 1969) of the cross spectral matrix (CSM) and the computational complexity involving the inversion of the CSM. Stabilization of the matrix inversion is generally required to avoid numerical difficulties. Unfortunately, the construction of the wavenumber window in Capon’s HRFK involves the signal

itself. Thus, the high resolution response can not be studied without the knowledge of the signal, and one is left to study the beam pattern instead to derive a lower limit of slowness resolution (Kind, 2002).



a) narrowband: 0.9-1.1 Hz, $N = 5$, $r = 400\text{m}$

b) broadband: 0.1-10.0 Hz, $N = 5$, $r = 400\text{m}$

Figure 6: Array response pattern for narrowband and broadband estimates for a pentagon shaped array with radius $r = 400\text{m}$. Maximum slowness has been set to 5 s/km , the radial resolution is 0.025 s/km and the azimuthal resolution 5° . The frequency bands for the evaluation of the beam patterns are given below the plots. The semblance values are depicted as gray scale map with a linear scale from 0 (white) to 1 (black). Black lines depict the -3dB contour relative to the main lobe maximum.

Due to the aim of investigation, i.e. the determination of frequency dependent phase velocity curves, array methods have to be employed for a narrowband analysis of the ambient vibration wavefield. Compared to broadband processing, the resolution of narrowband array responses is significantly reduced (Fig. 6) and aliasing peaks are fully developed. The improved response behaviour of broadband estimators stems from the superposition of beam patterns for different frequencies. The stacking of those response patterns, which differ with respect to the positions of the corresponding side lobes, leads to a smearing of the power contributions for all slownesses except for the main lobe.

Unfortunately, the use of broader frequency bands is prohibitive for the determination of dispersion characteristics. Thus, an enhancement of the resolution/aliasing capabilities for narrowband analysis can only be achieved by improving the spatial sampling of the wavefield. In turn, this requires the use of a large number of stations in field experiments. Clearly, viewed from the economical perspective, this option is prohibitive in most cases considering the initial equipment cost, the increased logistical effort and additional man power required for field experiments. Given the economical and logistical constraints, there is little one can do with respect to the resolution/aliasing issue for dispersion curves, except choosing an appropriate array size suitable for the analysis of a narrow wavelength range. The use of a single array deployment for the determination of well-resolved, non-aliased phase velocity estimates over a large frequency band seems unfeasible under common circumstances.

The question remains, how to optimize then the array geometry for a narrow target wavelength range. The resolution capability of some given station geometry is determined by the largest inter-station distance effectively seen by the incident wave arrivals from a certain direction as $\lambda^{\text{max}} \sim d_{i,j}^{\text{max}}$. This condition is a conservative approximation according to the Rayleigh criterion and for conventional beamforming techniques. For Capon's estimator, the resolution limit has been determined empirically as $\lambda^{\text{max}} \sim 3d_{i,j}^{\text{max}}$ (e.g. Tokimatsu, 1997). Just as well, the aliasing condition for 1D-station configurations with equidistant spatial station intervals can be given as $\lambda^{\text{min}}/2 \geq d_{i,j}^{\text{min}}$. Whenever no a priori information about the direction of the incoming wavefield is available, the variation of the maximal (minimal) inter-station distance when considering different DOAs should be minimized to obtain an azimuthally constant resolution/aliasing criterion. Geometries

showing a good azimuthal sampling are therefore the preferable choice for ambient vibration experiments. Particularly helpful for judging the azimuthal coverage of any array configuration is the co-array concept (Haubrich, 1968), which graphically depicts the distribution of all station pair difference vectors.

The criteria given above for the aliasing limits are approximate values, which strictly just apply for equidistant 1D station deployments. For a 2D-array setting the aliasing condition varies strongly with the direction of the incoming wave arrival, as the effective lattice seen by the wavefield changes significantly. In addition, the number of the shortest inter-station distances $d_{i,j}^{\min}$ may be small compared to the total number of larger $d_{i,j}$ involved in a particular observation direction. In order to visualize and quantify the effective aliasing behaviour for a general 2D array deployment, we have taken the following approach. For discrete azimuth values, we computed the projections of inter-station distances $d_{i,j}$ onto the observation direction. The effective lattice was derived by counting the number of $d_{i,j}$ in 10m intervals. The density plot in Fig. 6 shows the corresponding azimuthal variation of $d_{i,j}$ distributions. Considering the existence of a random wavefield (uniform distribution of DOAs), we averaged these lattices to obtain the mean effective inter-station density which can be interpreted as the average spatial filter for the wavefield components (gray curve in right panel, Fig. 7). Finally, we can derive the percentage of inter-station distances which fails to fulfil the aliasing criterion for a given wavelength by computing the cumulative distribution and normalizing to the total number of inter-station distances involved (black curve in right panel, Fig. 7). We called this the cumulative aliasing criteria curve (CACC), which allows the estimation of a relative amount of observed aliasing features for a particular wavelength. In the example shown in Fig. 7 (pentagon shaped array with 100m radius), we determine a critical effective half wavelength $\lambda^{\min}/2$ of approx. 90m, for which 50% of the involved inter-station distances show an insufficient spatial sampling of the wavefield.

It is clear, that the requirements on the array geometry setup in terms of resolution and aliasing lead to different conclusions. For high resolution capabilities, particularly at long wavelengths, the array size should be as large as possible. However, in order not to violate the aliasing criterion for even the shortest wavelengths of interest, the inter-station distances should be as small as possible. Some compromise is therefore needed to optimize the array geometry for a given narrow wavelength range.

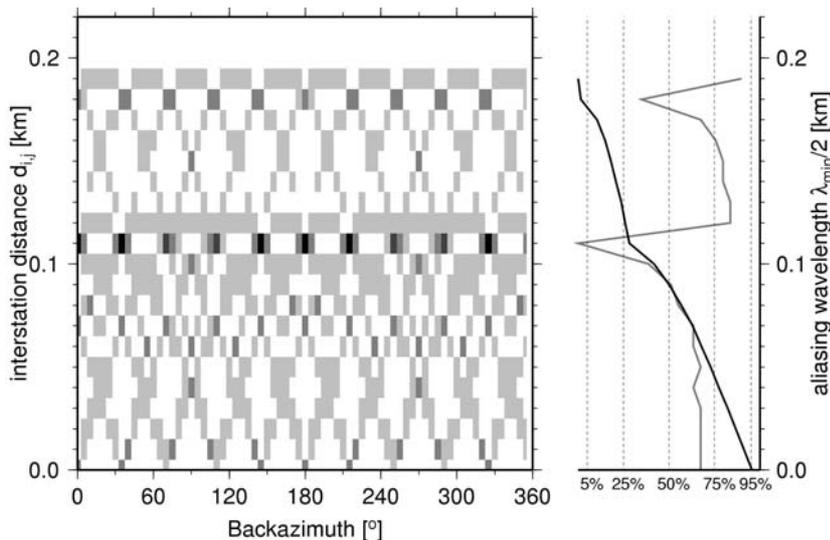


Figure 6: Density plot of inter-station distances $d_{i,j}$ for a pentagon shaped array with 100m radius. For different viewing directions (resolution 5°), the length of the projected inter-station difference vectors is binned in 10m intervals. The effective lattice for waves propagating from a single direction will determine the aliasing and resolution capabilities along the viewing direction. The average effective lattice for a large number of wave arrivals from different directions (as expected for ambient vibration wavefields) is given in the right panel as gray curve. The cumulative aliasing criteria curve (black curve, right panel) estimates the percentage of inter-station distances $d_{i,j}$ which fail on average to fulfil the aliasing criterion.

We suggest the usage of circular deployments with an odd number of stations. These particular array shapes show the best possible azimuthal suppression capabilities for a given number of sensors. Further, the number of smallest inter-station distances is always large and equals the number of deployed sensors. Thus, the aliasing criterion is indeed well approximated by $\lambda^{\min}/2 \geq d_{ij}^{\min}$ and the overall aperture D of the array is close to the largest involved inter-station distances d_{\max} . As a rule of thumb, one can then derive the wavelength limits of the array configuration from the chosen aperture alone, by using $d_{\max} \sim D$ and $d_{\min} \sim \pi D/N$, with N being the number of sensor, we get: $\lambda_{\max} \sim D > \pi D/N > \lambda_{\min}/2$. As an example, we give the approximate wavelength ranges resulting from applying the classic resolution/aliasing limits as specified above to this kind of array geometries with 5, 7 and 9 stations and radius $r \sim D/2$ in Table 1.

N	exact $d_{ij}^{\min} - d_{ij}^{\max}$	$\lambda^{\min} - \lambda^{\max} - 3*\lambda^{\max}$	approximate $d_{ij}^{\min} - d_{ij}^{\max}$	$\lambda^{\min} - \lambda^{\max} - 3*\lambda^{\max}$
5	1.17*r - 1.90*r	2.34*r - 1.90*r - 5.70*r	1.26*r - 2*r	2.52*r - 2.*r - 6.*r
7	0.87*r - 1.95*r	1.74*r - 1.95*r - 5.85*r	0.90*r - 2*r	1.80*r - 2.*r - 6.*r
9	0.68*r - 1.97*r	1.36*r - 1.97*r - 5.91*r	0.70*r - 2*r	1.40*r - 2.*r - 6.*r

Table 1: Resolution and aliasing limits for circular array layouts with N stations.

The usage of circular array layouts yields an additional advantage: the spatial autocorrelation curves can be computed in their original form (Aki, 1957, Eq. 4), which is easily recognized by considering the derived co-array configurations. Deviations from the ideal layout due to logistical constraints (feasibility of occupying station positions in urban environments) can be captured by the modified computation scheme suggested by Bettig et al. (2001, Eq. 5).

Theoretically, the SPAC method shows better resolution capabilities than the f-k methods. The resolution limit can be derived from considering the rate of change of the spatial autocorrelation curves. For small values of the argument of the Bessel function ($2\pi fr/c(f) < 0.4$), the sensitivity of the autocorrelation curve to changes of the phase velocity ($c(f)$) is low compared to the variation of the observed uncertainties. Nevertheless, using 0.4 as lower limit of the argument for the Bessel function, the longest wavelengths which should be resolved with the SPAC method are in the order of $\lambda^{\max} \sim 15-16*r$. The aliasing limitations of the SPAC method are expressed by the oscillating nature of the Bessel function, showing multiple solutions for arguments larger than ~ 3.6 . Thus, given a dense azimuthal sampling of the wavefield, the aliasing wavelength is given by approximately $\lambda^{\min} \sim 1.8*r$.

Influence of wavefield situation

One of the main difficulties involved in the analysis of ambient vibrations arises from our lack of knowledge about the nature of the wavefield origin. The spatio-temporal distribution of sources, the source types and excitation depths as well as the type and energy contents of the radiated waves are usually unknown to us. It is therefore difficult to assess the reliability of array analysis results from real field measurement data sets. Thus, we have chosen to analyse simulated data sets of ambient vibration data (Bonney-Claudet et al., 2004, Cornou et al., 2004, Ohrnberger et al., 2004a, 2004b), which allowed us to compare the phase velocity results to the theoretical input models used for the forward calculation.

Data sets have been computed for a variety of 1D-velocity structures for distinct geological settings (shallow vs. deep alluvial plains, variation of impedance contrasts and poisson ratios, high vs. low attenuation). The forward computation has been performed using the modal summation technique

and the code of Herrmann (1996), which allowed us to simulate wavefields containing either fundamental mode surface waves only or all higher mode contributions. For all tests, we assumed sources located at the surface, having a delta-like excitation function and randomly variable force orientation. We analysed pre-dominantly the vertical component of the resulting wavefields by both f-k and autocorrelation approaches. Among the tests performed, we will show specific examples to summarize the main findings.

Influence of spatio-temporal source distribution

For investigating the effect of multiple signal arrivals on the phase velocity estimates, we considered variable spatial source distributions. Starting from the simplest case consisting of a repeatedly excited single source location, the complexity of the source distribution was increased until reaching a configuration characterised by a dense random distribution (5000-10000 sources).

Interestingly, even for the simplest case, it was sometimes not possible to determine the complete dispersion characteristics over the frequency band of interest, regardless the employed analysis method. This unexpected result can be explained by the existence of spectral energy holes in the waveform spectra of the vertical component, which are caused by the filter characteristics of the layered site structure. The lack of coherent energy contribution within a specific frequency band prohibits the correct estimation of the narrowband signal phases. For site structures exhibiting a sharp impedance contrast, this effect will be always observed at least around the H/V peak frequency, which is the expression of the vanishing vertical signal energy (Scherbaum et al., 2003).

Concentrating therefore only on those frequency bands containing sufficient energy contribution, the analysis results of these experiments lead to the following conclusions (Ohrnberger et al., 2004b): i) for f-k techniques, the existence of a single source, or an azimuthally constrained dominant source region, is a favourable wave field situation for the determination of Rayleigh wave dispersion characteristics. Random source configurations lead to underestimation of the estimated slownesses for lower frequency bands. This effect can be attributed to the insufficient resolution capabilities of a specific array configuration for fast travelling, low frequency waves. The superposition of array responses from variable directions results in a single broad peak located close to the origin of the wavenumber map. ii) the results for the modified spatial autocorrelation technique are, in accordance with the underlying assumption of a stationary stochastic wavefield in space and time, best for random source distributions. Single source or dominant source contributions from an azimuthally constrained region may lead to biased results in case that the azimuthal distribution of station pairs does not cover the full azimuthal range. It is noteworthy, that this effect could be observed in a real data example (Fig. 7).

Clearly, f-k and spatial autocorrelation techniques show opposite behaviour with respect to the spatial distribution of sources. Without knowledge of the real source distribution characteristics, it is therefore difficult to argue, which of the methods are to be preferred. Thus, in real field experiments, we rather favour the idea to use the complementary information of these techniques to obtain correct phase velocity estimates for any situation. The results of the f-k technique permits to judge the wavefield direction characteristics within certain frequency bands. Depending on the existence of dominant propagation azimuths, either the results of the f-k or spatial autocorrelation method can be taken into account.

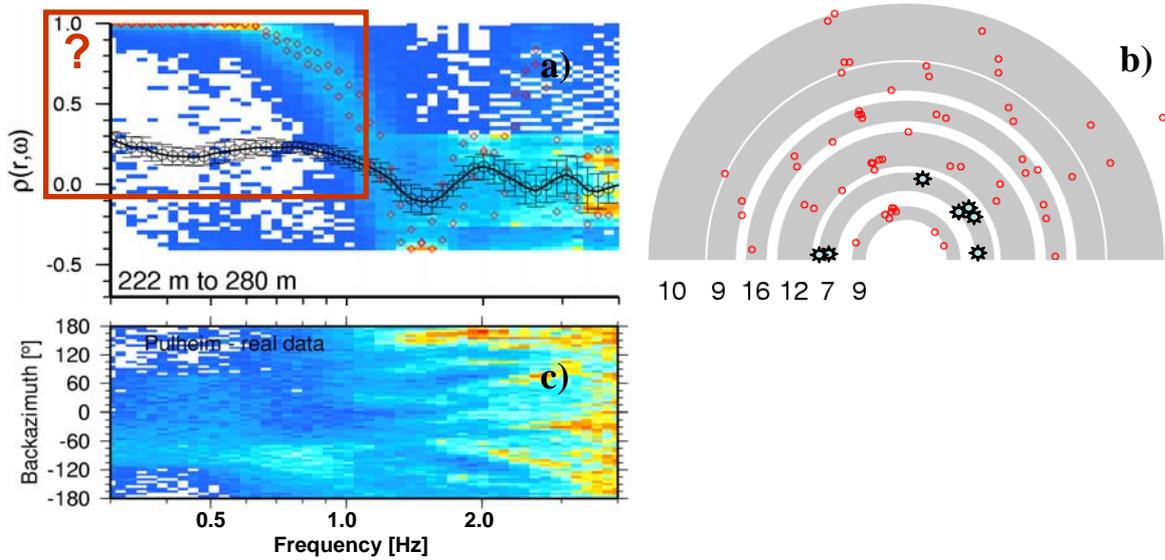


Figure 7: Possible bias of autocorrelation results caused by insufficient azimuthal sampling. a) Autocorrelation curve (black circles with error bars, MSPAC) computed for the second ring of the co-array distribution shown in panel b) (blue stars). The underlying density map shows for comparison the results of individual f - k estimates (CVFK) converted to autocorrelation values by Eq. 5 for r_1 and r_2 (f - k mean values depicted as red diamonds). An unexpected distortion of the autocorrelation curve is observed for frequencies below 1 Hz. The effect is explained, by the existence of a dominant wavefield direction as determined from the f - k analysis in c) in combination with only partially sampled azimuth range by the contributing station pairs as shown in b). The wavefield propagation directions are concentrated along wave paths in W or WSW azimuth and the station pair directions cover only a restricted azimuth range (lacking directions from N to WNW).

Influence of attenuation structure

While analysing the simulated datasets for fundamental mode Rayleigh waves and the simplest source configuration (single source location), one peculiar feature was observed. The phase velocity curves derived from “far-distant” (>1 km) sources showed, especially for higher frequency bands, poor quality when compared to the results obtained from closer located excitation points of the wavefield. This effect was contradictory to our expectations regarding the influence of curved wavefront arrivals as discussed above. The spectral analysis of the individual time series revealed, that the cause of this outcome is related to a lack of coherent energy which can be attributed to the intrinsic attenuation properties of the medium. In order to confirm this observation, we tried to derive the influence of the site’s attenuation properties on the array analysis results by comparing simulated datasets for two velocity models which just differed in their respective Q-factors in the individual layers. A single source at a distance to the array center of 500 m only was used for the simulation and all Rayleigh wave modes have been included in this computation.

The phase velocity estimation results obtained from the conventional and high resolution f - k algorithms (CVFK and Capon) are shown in Fig. 8 together with the corresponding spectral energy contributions of the individual Rayleigh wave modes. The comparison of the spectra indicates the appearance of pronounced holes for the fundamental and higher Rayleigh wave modes for the stronger attenuating model. Interestingly, the higher mode energy contributions dominate for certain frequency bands, whereas for the less attenuating structure, the fundamental mode is the strongest wavefield component for large portions of the analysed frequency band. As a consequence, the estimated phase velocities jump between the theoretical dispersion curves of fundamental and higher mode branches, depending on their modal energy contributions of the wavefield.

Intermediate phase velocities are found for comparable energy levels between different modes of propagation.

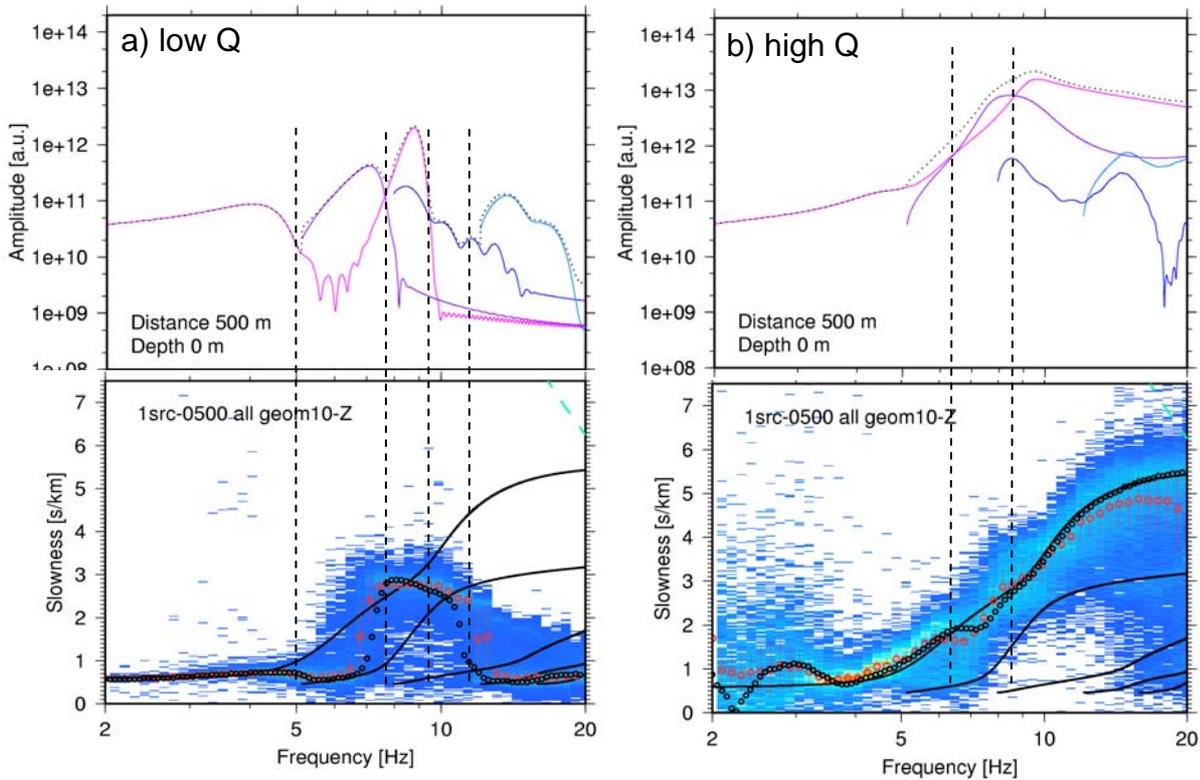


Figure 8: Influence of attenuation structure on the phase velocity estimates. a) Dispersion curve analysis by CVFK method for a shallow sedimentary site with low Q -factors (high intrinsic attenuation) and corresponding spectral energy contributions of individual Rayleigh wave modes on top. The simulated wavefield is composed of repeated excitation of a single source located at 500m distance from the array center. The colored density map depicts the histogram of slowness estimates from all analysed time windows within narrow frequency bands. Black curves indicate the theoretical Rayleigh wave dispersion curve relations for the velocity structure. b) as a), but for model with high Q -values (low intrinsic attenuation). Vertical dashed lines show the location of cross-over points, where the dominating energy contribution in the wavefield changes from one mode branch to another.

The existence of intermediate phase velocity estimates has been previously described by Tokimatsu et al. (1992a,b) and Tokimatsu (1997). The authors termed the observed dispersion curve results “apparent phase velocities” and suggested to include the media response of higher mode contributions into the forward computation problem when inverting for the velocity model.

Our simulations suggest that the attenuation structure heavily influences the outcome of the estimated dispersion curves. This implies an additional dependence of the results on the involved source-receiver distances which would rule out any attempt to find a correct treatment in the inversion task. However, from experimental results in real environments, we know that ambient vibration array analysis results are reproducible and remain temporally relatively stable. Without further investigations about the true nature of the ambient vibration wavefield, it seems difficult to bring these conflicting observations into agreement. However, for the moment, we prefer the following interpretation: The stability of real ambient vibrations measurements is a consequence of the randomness of the source distribution and the attenuation properties of the media. For uniform spatial random distribution of source locations, the number of contributing sources increases proportional to the areal increase around the receiver array, i.e. approximately quadratically with the distance. On the other hand, geometrical spreading for surface waves show a root square dependence of the amplitudes with distance. Finally, due to the intrinsic attenuation, distant source

contributions are further diminished. Thus, the overall distance range which contains sources contributing significantly to the observed microtremor wavefield is probably rather limited.

Summary of observations

The following Table 2 summarizes the observations from the numerical experiments and concludes on the possible action required for improved phase velocity estimates.

<i>Experiment</i>	<i>Observation</i>	<i>Conclusions/Recommendations</i>
Plane wavefronts (randomly perturbed)	FK: - mean is unbiased estimate - variance scales with array aperture - variance scales with slowness	- larger time windows needed to improve phase delay estimate. - large number of windows necessary to obtain good statistics - larger array sizes perform better
Curved wavefronts	FK: - biased estimate of true slowness (lower) - bias scales with 'curvedness' - 'curvedness' related to ratio of source distance to array radius d/r SPAC: - directional dependent deviation - average of estimates unbiased until $d/r \sim 1$ - $d/r < 1$ biased estimates (larger)	- sources inside array geometry give unreasonable results (CVFK) or overestimate the slowness (SPAC). - nearby sources $d/r < 2.5$ must be avoided → semblance based estimate outperforms power-based algorithms - very small arrays preferred (allows certain control of inside sources / d/r increases fast from site of operation).
2D-Aliasing (narrow-band estimate)	FK: - multiple plane wave arrivals lead to biased estimates (always lower) - bias scales with resolving capability of the array (related to array size) - aliasing patterns related to peaks in cumulative criteria failure curve SPAC: - similar aliasing criterion as for FK methods. - better resolution capabilities for longer wavelengths.	- preferable symmetric arrays showing similar array response behaviour for all directions and favourable cumulative criteria failure curve. - circular array setups allow combined usage of f-k and SPAC techniques.
Wavefield simulation: single vs. multiple arrival directions	FK: - dominant source directions are favourable situation, multiple signal directions lead to biased results (s.a. 2D-Aliasing) SPAC: - dominant direction of wavefield propagation may introduce biased results due to insufficient directional sampling using unfavourable array layouts.	- complementary use of both f-k and spatial autocorrelation techniques is required to resolve peculiar wavefield situations.
Wavefield simulation: low attenuating vs. high attenuating site structure	All methods: - high attenuating structures may favour the existence of higher mode contributions in the wavefield. Apparent phase velocity curves are observed and interpretation is complicated.	Kinks in dispersion curve estimates at certain frequencies obtained for different analysing methods indicate the existence of higher mode contributions in the wavefield. An appropriate treatment in the inversion procedure should be considered.

Table 2: Summary of observations and corresponding conclusions/recommendations.

Improved measurement strategy for ambient vibration array analysis

From the above experiments and argumentation it seems clear that, due to the complex interaction between measurement layout, analysing method and wavefield situation, the determination of highly reliable, multi-modal dispersion curve estimates with small error bounds in general remains a difficult task. Nevertheless, we could identify sources of error and the limiting factors of the estimation procedure. Most important herein seems the peculiar observed wavefield situation, which is a result from both excitation, and propagation properties:

- source location and excitation: space-time density (single source, dominant source area, multiple sources) of ambient vibration sources and their spectral energy content (source time function, source orientation and source depth).
- wave propagation: filter effect of layered velocity structure and Q-structure, resulting in changes of the energy partitioning of surface wave modes).

In order to improve the dispersion curve determination from microtremor recordings, we can at best react to the given ambient wavefield situation by improving the measurement and analysis scheme. Due to the absence of a priori knowledge about the specific site conditions in terms of source excitation and wave propagation characteristics (i.e. the site structure), it is necessary to acquire the necessary information from analysis of the recordings in the field. Then, we are able to judge the wavefield situation on site and adapt our measurements (array layout / processing) accordingly. In particular, we propose the following general work flow:

Pre-experimental steps:

- (1) Select adequate equipment for measurements ($f_{\text{seis}} \ll$ lowest frequency of interest)
- (2) Huddle test – verify calibration of instruments (i.e. estimate possible instrument related delay times)

Measurement/Processing/Analysis scheme:

- (3) Set up very small array, e.g.: circle with N sensors with radius $r = 2 \dots 5$ [m], the number of sensors N has to be chosen as a compromise between cost and resolution capabilities. A minimum of $N=5$ sensors should be considered in any situation.
- (4) Estimate amplitude spectra and H/V ratios at individual array stations. By doing so, some information of the analyzable frequency bands can be acquired. Furthermore, a detection of a possible 2D- or 3D-site condition can be inferred!
- (5) Apply narrowband array analysis methods for the full frequency band which has been determined as appropriate (4). Make use of competing algorithms (CVFK, CAPON, (M)SPAC). Use sufficiently large number of long time windows for the analysis.
- (6) Derive frequency band(s) which show(s) consistent phase velocity estimates for all methods. Contradicting results may indicate complex wave field situations which may not be resolvable (mixed mode contributions) and mark those results as critical information which may require extra treatment in the inversion process.
 - derive corresponding wavelength range
 - cross check results with capabilities/limitations of array geometry in terms of aliasing and resolution
 - derive (first) partial branch of dispersion curve
- (7) Define next narrow target wavelength range which extends the previously analyzed wavelength interval to longer wavelengths and seek the array aperture being optimal for this wavelength range.
- (8) Rearrange array geometry according to (7)

- (9) Repeat steps (4) to (8) until a well estimated, consistent dispersion relation is obtained for the frequency band of interest (i.e. necessary frequency band for the inversion of velocity models).

Figure 9 summarizes the proposed measurement scheme.

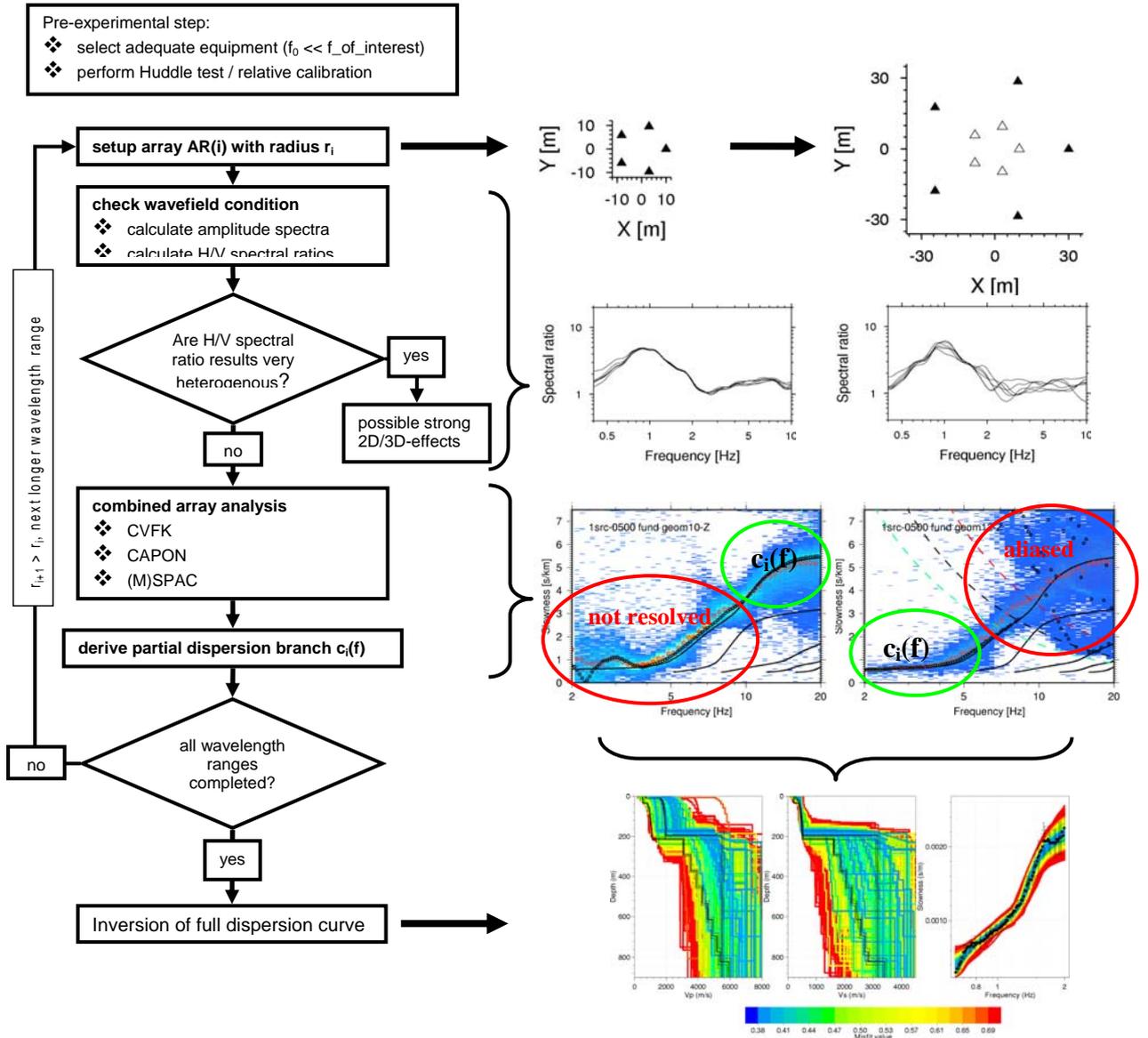


Figure 9: Proposed measurement scheme for determination of improved broadband dispersion curve characteristics by iterative deployment of adapted array configurations for narrow wavelength ranges.

3. Inversion of dispersion/autocorrelation curves

In the first part of this paper, a measurement and processing strategy is proposed to estimate the dispersion curve (usually, the fundamental mode of Rayleigh wave) with the lowest uncertainty over the widest frequency range. A new software was developed to invert the dispersion data with

their uncertainty. Based on the neighbourhood algorithm (Sambridge, 1999), it is fully integrated with the array processing system, making the whole processing chain easy to use and very flexible. The neighbourhood algorithm (NA) is a direct search inversion method randomly generating models inside a parameter space. To efficiently find the solutions of the problem, only the best promising parts of the parameters are densely sampled. As it is a random process the robustness of the results is estimated by launching several inversions with differing random seeds (Wathelet et al., 2004). The software has been designed to invert different kinds of dispersion curves (Rayleigh or Love waves, fundamental and/or higher modes) as well as autocorrelation curves coming out from the SPAC analysis (Wathelet et al (a), submitted). The misfit is given by:

$$misfit = \sqrt{\sum_{i=1}^{n_F} \frac{(x_{di} - x_{ci})^2}{\sigma_i^2 n_F}}$$
Eq. 6

where x_{di} is the velocity of the data curve at frequency f_i , x_{ci} is the velocity of the calculated curve at frequency f_i , σ_i is the uncertainty of the frequency sample considered, and n_F is the number of samples.

Classically, obtaining the Vs profile at one site with the SPAC method is a two-stage processing involving the derivation of the dispersion curve from the auto-correlation curves with a least-square scheme (e.g. Bettig et al. 2001) and the inversion of the dispersion curve. Recently, Asten et al. (2004) proposed to merge these two steps into a single inversion based on least-square optimisation (Herrmann 1994), allowing the determination of Vs(z) directly from the auto-correlation curves. The approach we propose is conceptually the same except that we make use of the neighbourhood algorithm (3.3), which allows an exploration of nearly all equivalent minima in terms of the misfit function. The different processing methods (CVFK, Capon, SPAC) have been thoroughly compared on synthetic data and one real site (Wathelet et al (b) and (c), submitted). As the main conclusions are very similar when considering the inversion of dispersion or autocorrelation curves, we only present the processing of the dispersion curve in the following.

Synthetic ground model

The efficiency of the neighbourhood algorithm is tested on a synthetic ground model which is shown in figure 10. The model structure is made of three layers including the bottom half space, the characteristics of which are given in table 3. Figure 10 displays the Vp and Vs profiles of this model as well as the dispersion curves for Rayleigh and Love waves (fundamental and first higher modes), and the Rayleigh ellipticity curve for the fundamental mode. The frequency corresponding to the curve peak (5.63 Hz in this case) is usually close to the fundamental frequency of the ground structure (Bonnefoy-Claudet, 2004), which is mainly controlled by the seismic contrast at 10 m in this case.

Layer #	Thickness (m)	Vp (m/s)	Vs (m/s)	Poisson's ratio
1	10	375	200	0.3
2	90	1750	1000	0.25
3	∞	4500	3000	0.1

Table 3: Characteristics of the synthetic model. Density is 2 t/m^3 in all layers.

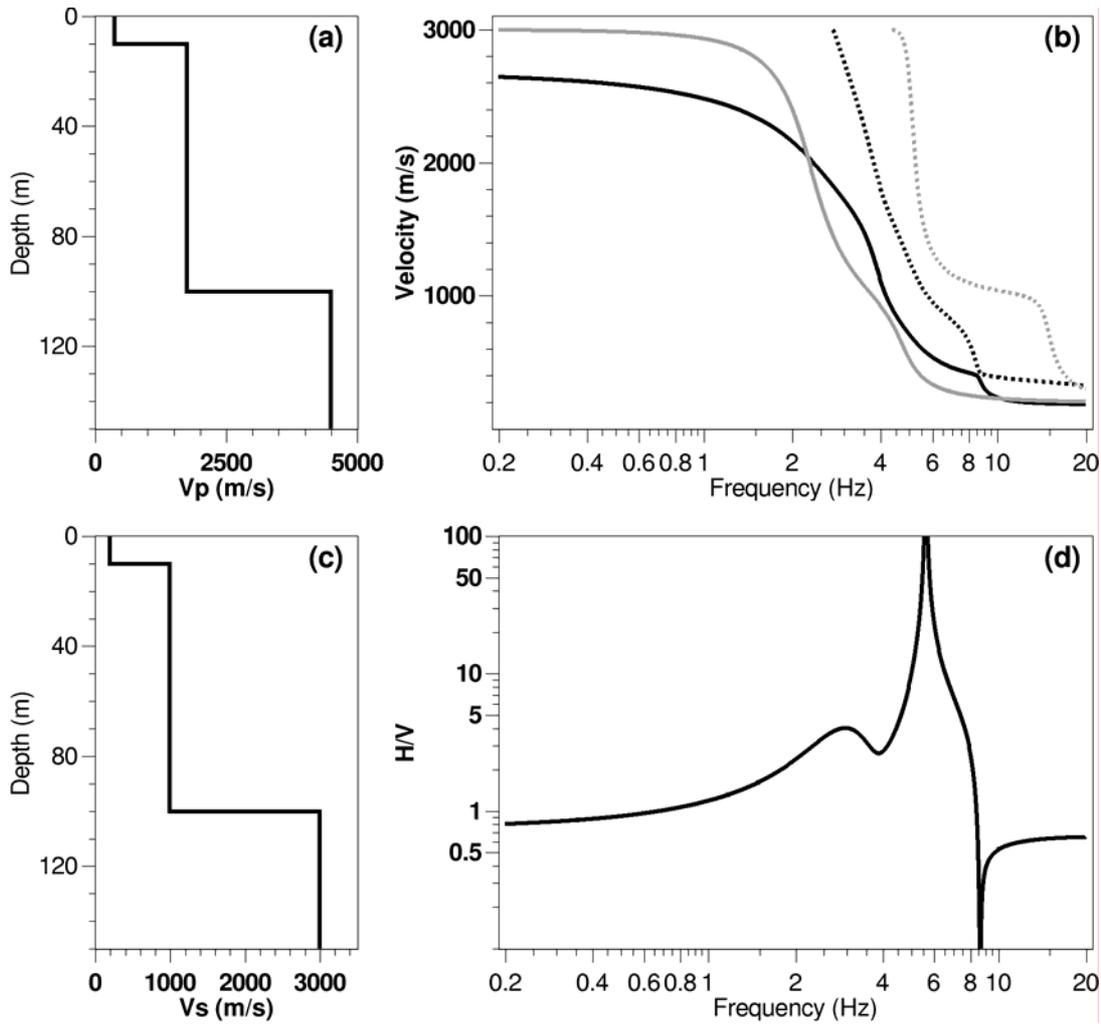


Figure 10: Synthetic model. (a) V_p profile. (b) Dispersion curve for the fundamental (solid) and the first higher mode (dots) of Love (grey) and Rayleigh (black) waves. (c) V_s profile. (d) Rayleigh fundamental ellipticity.

Parameterization of the ground model

The fundamental Rayleigh dispersion curve shown in figure 10 is sampled with 50 points on a wide frequency range from 0.2 to 20 Hz. It is then inverted using the neighbourhood algorithm with a three-layer structure. The number of layers in the inversion fits the reality, which is the most favourable case for inverting the dispersion curve. The influence of the number of layers will be examined afterwards. The next step is to fix the intervals of parameter variations, which can be determined by prior knowledge about the geological structure or seismic properties. Table 4 gives the list of parameters and their prior intervals which are very large (no a priori information) in this case. V_p profile is guaranteed to be monotonous by setting the velocity variations as parameters rather than the absolute value, while V_s is kept monotonous by the penalization technique (Wathelet, 2005).

Layer #	Thickness (m)	V_p (m/s)	V_s/V_p	Density (t/m^3)
1	1 to 50 m	200 to 2,000	0.01 to 0.707	2
2	1 to 200	+10 to 2,000	0.01 to 0.707	2
3	∞	+10 to 3,000	0.01 to 0.707	2

Table 4: Parameterized model for three-layer inversions. The "+" sign stands for incremental velocity when the parameter is the velocity gap between the first and the second layer.

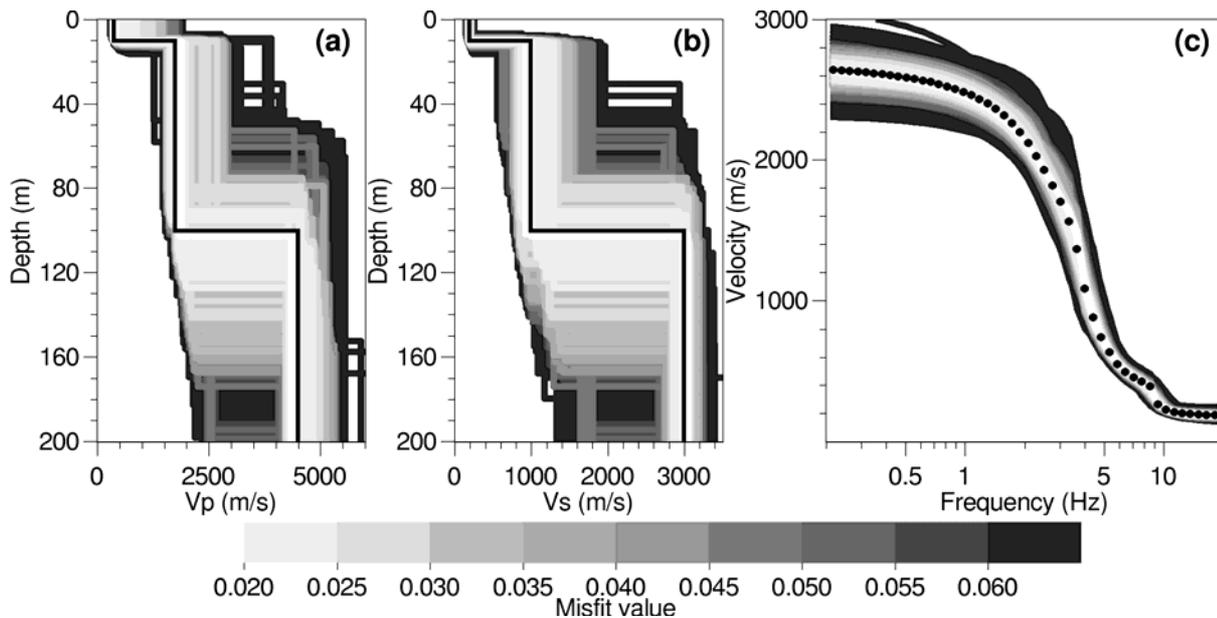


Figure 11: Inversion with a three-layer model over a broad frequency range. (a) Resulting V_p profiles. (b) Resulting V_s profiles. The black lines are the theoretical velocity profiles. (c) Dispersion curves corresponding to models of figures (a) and (b). The black dots are the theoretical dispersion curve used as the target curve during inversion.

Five independent runs are started generating 75500 model, with a minimum misfit around 0.02. The V_p and V_s profiles of models for which the misfit is less than 0.1 are plotted in figure 11, as well as the corresponding dispersion curves. The number of plotted models is about 8900.

The V_s profiles with a good misfit (lower than 0.03) more or less match the theoretical one. V_s values in the first layer and in the bottom half-space, as well as the depth of the first interface, are well constrained, while uncertainties on the velocity in the intermediate layer and on the depth of the second interface are higher. These results show that a large number of models are able to explain the dispersion data and illustrate the problem of non uniqueness of the dispersion curve inversion. V_p profiles show similar results with a higher uncertainty on all layer properties.

In this inversion, the chosen number and properties (homogeneity) of layers perfectly match the synthetic model. In practice, this information is unknown, unless prior information is given on the site (borehole description, for instance). This latter effect is studied in a further section.

In order to assess the influence of the parameterization choice, V_s profiles obtained for 3 homogeneous layers (identical to the one of figure 11), 2 heterogeneous layers over an half-space and 11 layers are compared on figure 12. In heterogeneous layers, a velocity gradient (power law) is allowed in order to gain flexibility in the matching of the dispersion curve shape (Wathelet et al., 2004). For the 10 layer case, the thickness values are fixed and V_s and V_p values are inverted. The number of parameters to invert is then 8 for 3 homogeneous layers (V_s , V_p and thickness), 10 for heterogeneous layers (V_s , V_p , thickness and the exponent of the power law) and 20 (V_p and V_s values) for the 10 homogenous layers. Low velocity layers are allowed in the last case. The introduction of vertically heterogeneous layers (figure 12b) does not change too much the V_s values in the three layers for the best fitting models. On the other hand, the uncertainty on the bedrock depth strongly increases. With ten homogeneous layers, a lot of models may explain the data with a good misfit. This example illustrates the non-uniqueness problem and the necessity to choose an adequate parameterisation at the beginning of the process. To this regard, any inversion has to be launched with the minimum number of layers and limiting low velocity layers, compatible with the prior geological information available for the site.

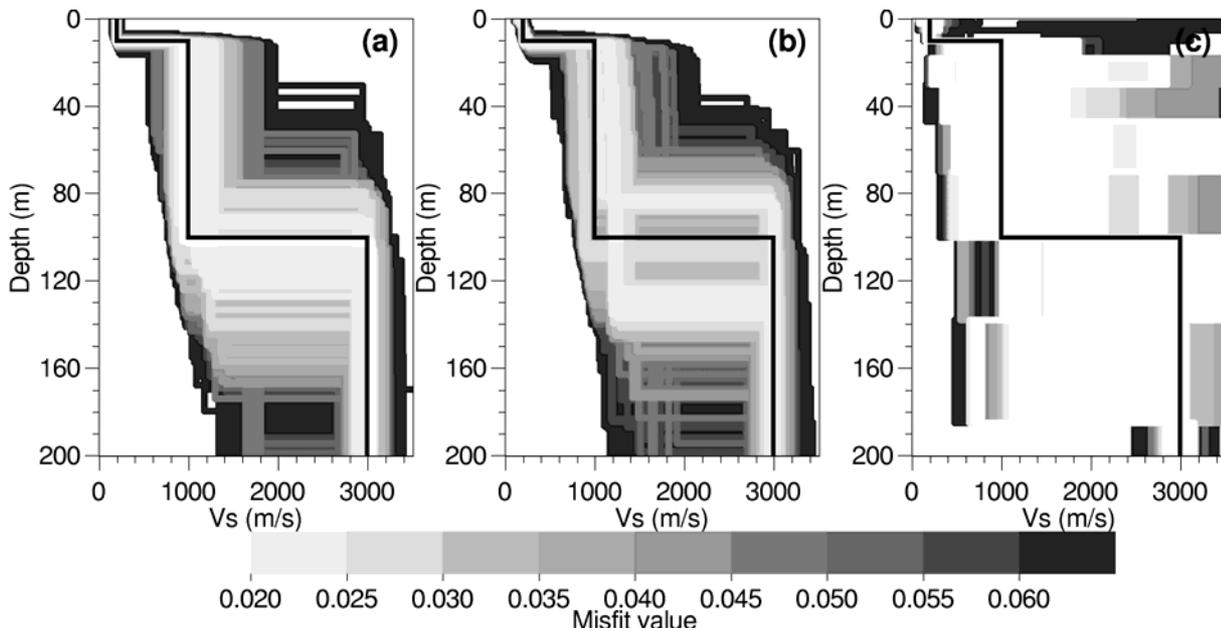


Figure 12: Comparison between the inversion results with (a) 3 homogeneous layers, (b) 2 heterogeneous layers overlying a half space (c) 10 homogeneous layers

Frequency range

Contrary to synthetic dispersion curves that can be calculated on the whole frequency interval (figure 10c), experimental curves are generally obtained on a restricted frequency band, which depends on the source and sensor characteristics. Even in an ideal case (Dirac source and broad band sensors), Scherbaum et al. (2003) showed that the energy on the vertical component drastically decreases in the vicinity and below the fundamental frequency of the soil structure. As Rayleigh waves are mostly measured on the vertical components, the dispersion curve is generally only available above the natural frequency of the ground structure. In our synthetic model, this condition is particularly severe as the dispersion curve will be unknown below 5.5 Hz. To simulate this high pass effect, the same dispersion curve as in figure 12 is sampled with 30 points on a restricted frequency range between 5.5 and 15 Hz. The parameterisation is exactly the same as in table 4. The model with misfits lower than 0.1 (4800 matches) are displayed in figure 13. The minimum misfit is around 0.02. The Vs profile is correctly retrieved down to 10 m. Below that depth, most of the models (with a misfit lower than 0.04) explain the data and there is no constrain on the characteristics of the second and third layer. This results from the loss of information at depth due to the removal of the low frequency part of the dispersion curve.

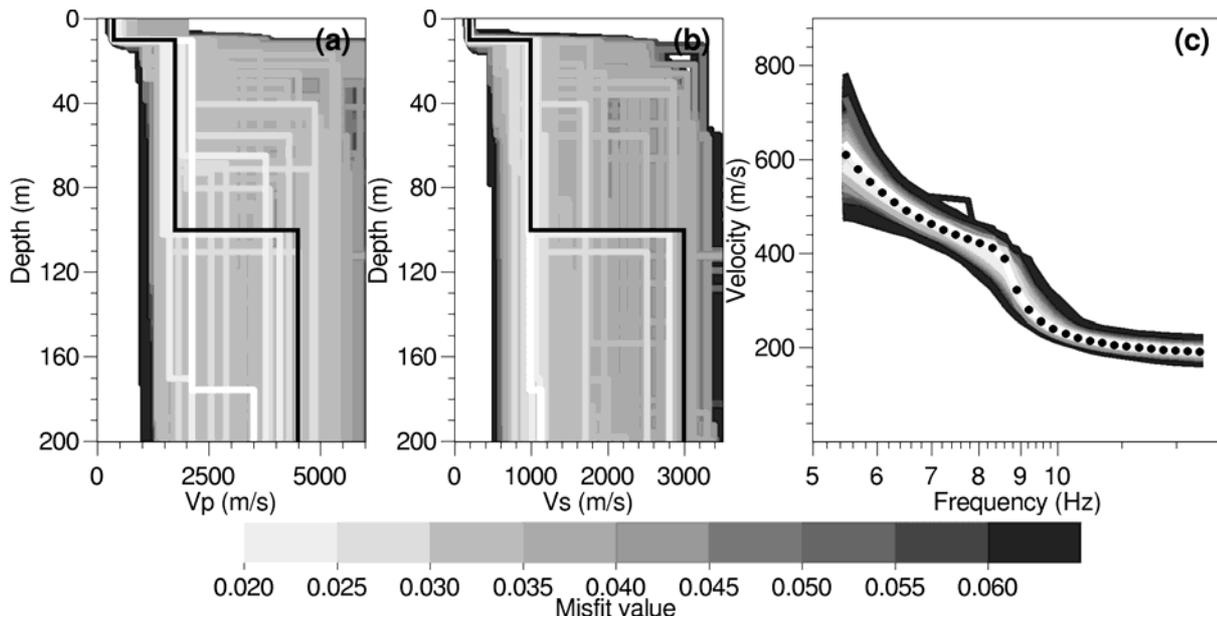


Figure 13: Inversion with a three-layer model over a restricted frequency range. (a) Resulting V_p profiles. (b) Resulting V_s profiles. The black lines are the theoretical velocity profiles. (c) Dispersion curves corresponding to models of figures (a) and (b). The black dots are the theoretical dispersion curve used as the target curve during inversion.

In practice, the high frequency part of the dispersion curve can also be difficult to obtain from array measurements (Nguyen et al., 2004), due to the lack of coherency of ambient vibrations for small aperture arrays or to the presence of higher modes. An experiment with only large aperture arrays is simulated by removing the dispersion information above 8 Hz. The fundamental Rayleigh dispersion curve in figure 14(c) is sampled with 30 points from 0.2 to 8 Hz. The inversion is run with five distinct processes, keeping the parameterization of table 4. The retrieved V_p and V_s profiles are shown in figures 14(a) and 14(b), while the corresponding dispersion curves are plotted in figure 14(c). Whereas the velocities (V_p and V_s) of the basement are obtained with approximately the same accuracy as with the full frequency range (see figure 12), a strong bias is observed for the properties of the first two layers. On another hand, the uncertainty on the basement depth is higher without the high frequency part of the dispersion curve. These results highlight the need for a good definition of the dispersion curve at high frequency (above 8 or 10 Hz in this case). Active methods, which allow a better control on the source parameters, can usually provide this complementary information at high frequency (Nguyen et al, 2004)

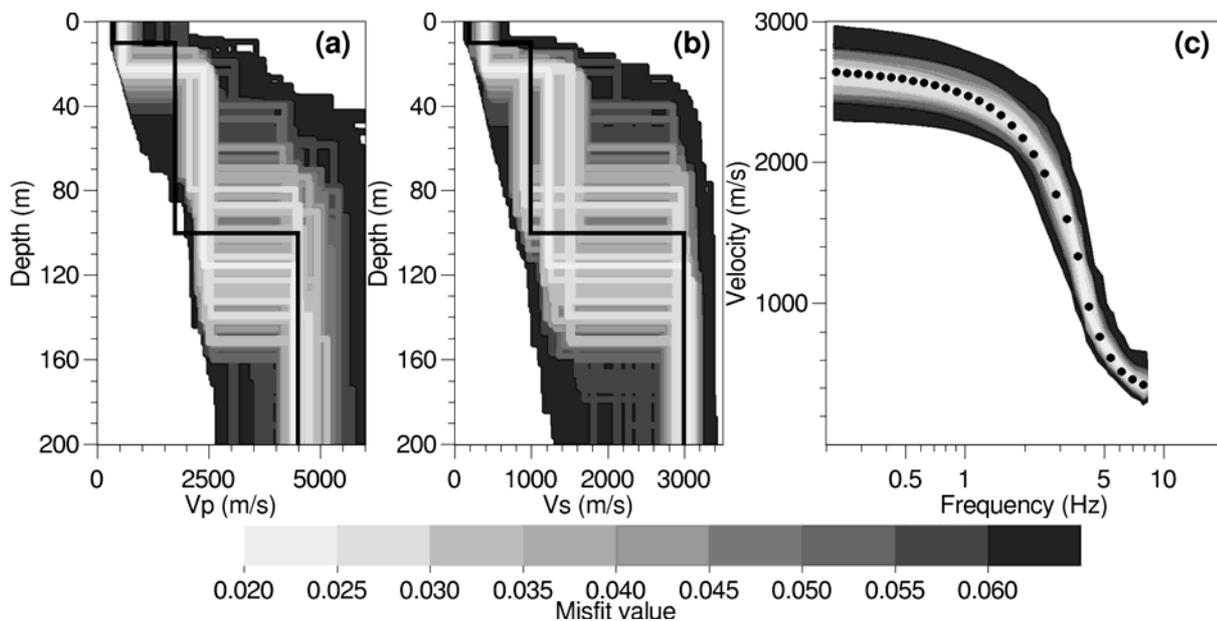


Figure 14: Inversion with a three-layer model over a low frequency range. (a) Resulting V_p profiles. (b) Resulting V_s profiles. The black lines are the theoretical velocity profiles. (c) Dispersion curves corresponding to models of figures (a) and (b). The black dots are the theoretical dispersion curve used as the target curve during inversion.

A priori information

The developed software allows prior information to be easily incorporated in the parameterisation or in the misfit computation during inversion. Such prior information can come from variable sources, including geological data (borehole logs), geotechnical results (penetration tests, density measurements), applied geophysical measurements (V_p and V_s profiles from refraction or borehole tests, vertical electrical soundings, electrical images....), or H/V curves from single station ambient vibration measurements. Here, we show the influence of prior knowledge about the layer depth, about V_p values and about the H/V curve on the inversion process, in the case of the 3 layer synthetic model.

Depth data

If the depth of a velocity contrast is known from other investigations, it can be introduced in the parameterization. Inversion is performed on the same dispersion curve as in figure 11 with the parameter ranges given in table 5. The depth of the basement is supposed to be known with an uncertainty of plus or minus 5 metres.

Layer #	Depth (m)	V_p (m/s)	V_s/V_p	Density (t/m^3)
1	1 to 90	200 to 2,000	0.01 to 0.707	2
2	95 to 105	+10 to 2,000	0.01 to 0.707	2
3	∞	+10 to 3,000	0.01 to 0.707	2

Table 5: Parameterized model for three-layer inversions with prior depth. The "+" sign stands for incremental velocity: the parameter is the velocity gap between the first and the second layer.

Five runs are launched generating the models displayed in figure 15 and the misfit values can be compared directly to the ones of figure 11. Comparing the velocity profiles in figures 11 and 15, the prior depth information has obviously a positive influence on the inversion results, with a strong decrease of the uncertainty on the intermediate layer velocities.

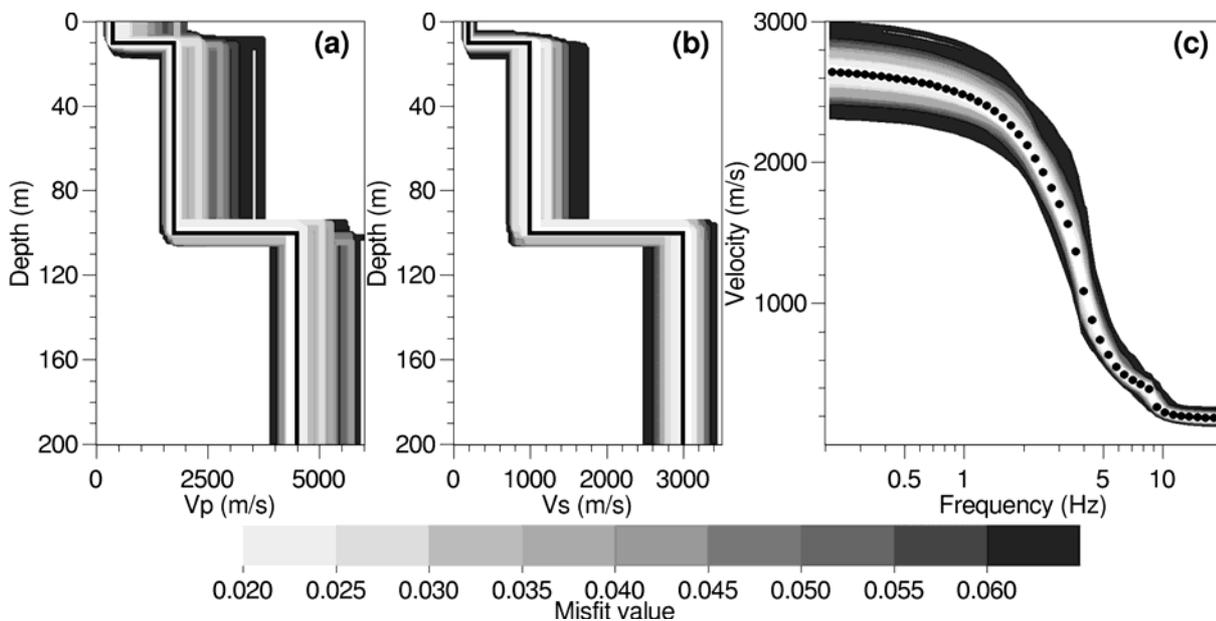


Figure 15: Inversion with a three-layer model with prior depth. (a) Resulting V_p profiles. (b) Resulting V_s profiles. The black lines are the theoretical velocity profiles. (c) Dispersion curves corresponding to models of figures (a) and (b). The black dots are the theoretical dispersion curve used as the target curve during inversion.

V_p values

V_p profiles may be measured by other methods, which are not related to surface wave properties, as refraction tests, borehole logging and cross-hole tests. Like the depth, prior information about V_p can be introduced in the parameterization. In order to investigate this effect, V_p values are fixed in the three layers during the inversion.

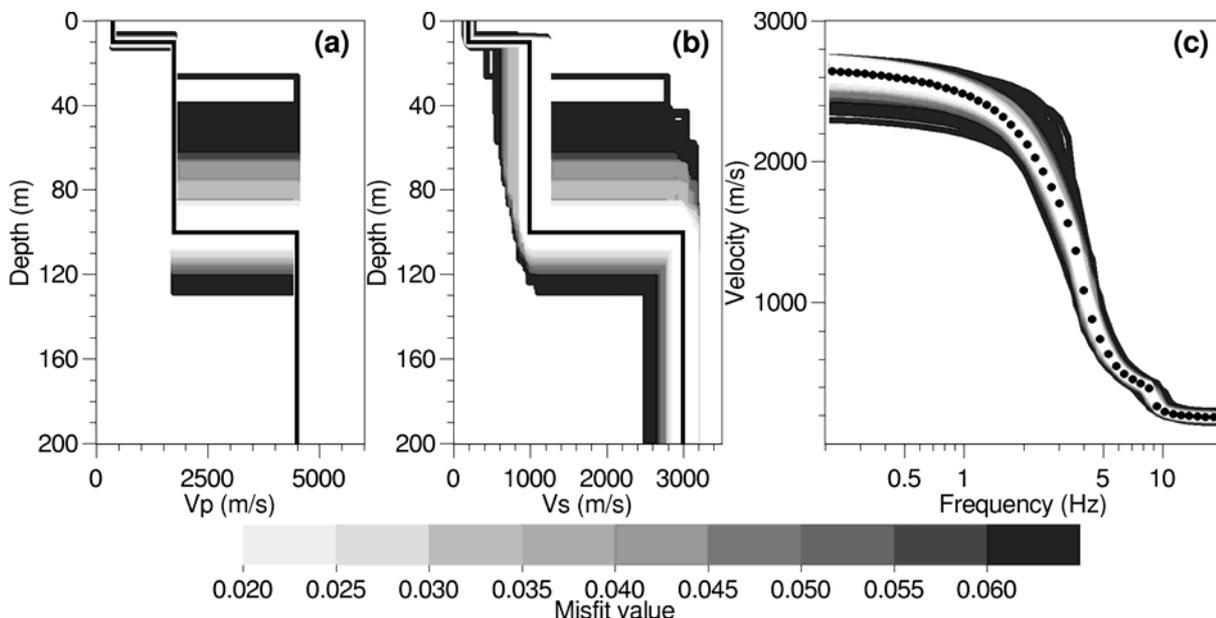


Figure 16: Inversion with a three-layer model with prior V_p . (a) Resulting V_p profiles. (b) Resulting V_s profiles. The black lines are the theoretical velocity profiles. (c) Dispersion curves (whole frequency range) corresponding to models of figures (a) and (b). The black dots are the theoretical dispersion curve used as the target curve during inversion.

The dimension of the parameter space then reduces from 8 to 5. Inversion results are shown in figure 16 for five runs. 3100 models have a misfit lower than 0.1, with a minimum around 0.0025.

Compared to figure 11, the uncertainty on the intermediate layer properties (V_s and depth) is again greatly reduced. Fixing V_p automatically limits the maximum V_s value to $0.7 V_p$. In addition, the V_p information also diminishes the depth uncertainty on the basement. Other tests (not shown here) with wrong prior V_p profiles show that the final V_s results are moderately affected by overestimated V_p values. In contrast, any underestimation of V_p completely ruins the inversion results in terms of V_s . Consequently, V_p values can only be fixed if the reliability of the external data can be proved. When possible, the best solution is to introduce prior information with its uncertainty.

H/V peak

The frequency of the H/V peak, which is known to match the resonance frequency of the site (Bonnefoy-Claudet, 2004), depends on the seismic stratigraphy of the site and may add a valuable constrain in the inversion, as far as the frequency is stable over the site.

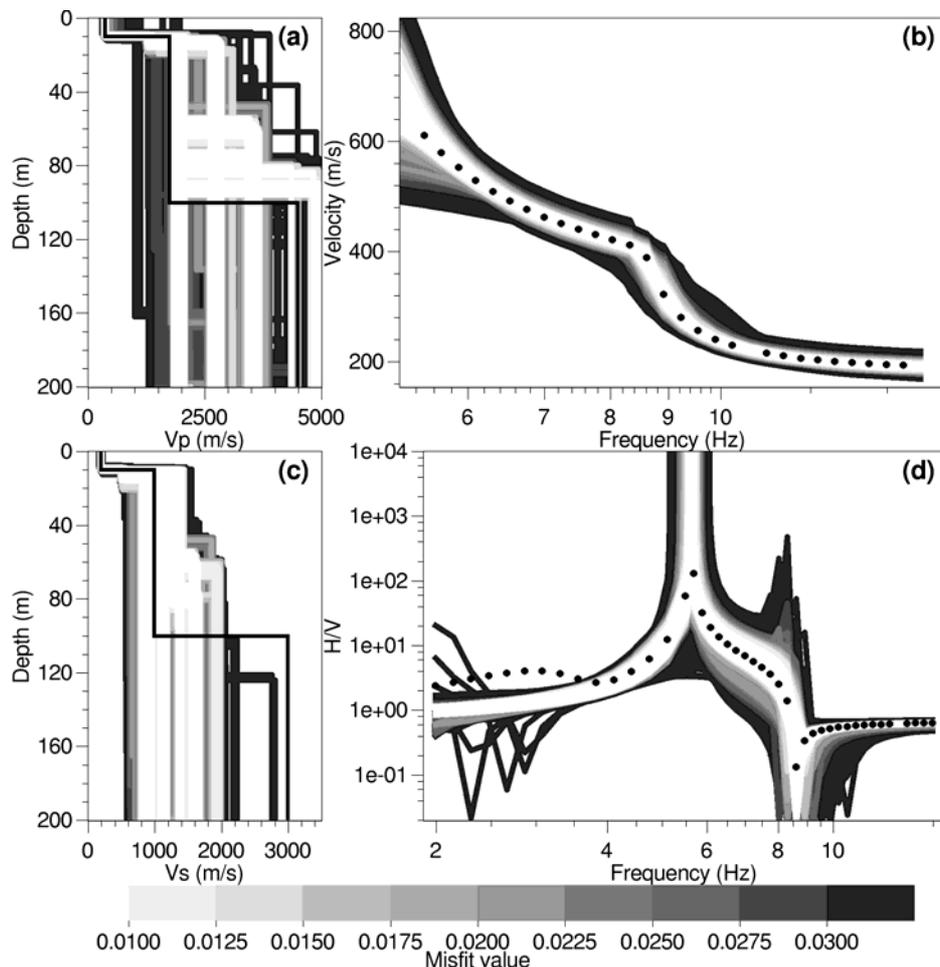


Figure 17: Joint inversion of the dispersion curve (limited frequency range) and the H/V peak frequency. (a) and (c): V_p and V_s profiles. (b) Dispersion curves. (c) Ellipticity curves.

Figure 17 presents the results of the joint inversion of the dispersion curve (limited frequency range) and of the peak frequency value (5.63 Hz). The misfit is calculated by a weighted sum of the dispersion and the ellipticity misfits, with a weight of 50% for each of them. This ensures that nearly all generated models are complying with an ellipticity peak at 5.63 Hz. Consequently, to achieve the same fit of the dispersion curve as in figure 5.4, the misfit scale is divided by 2. Five runs are launched with the parameters described in table 4. The total ensemble of models with a misfit less than 0.01 is plotted in figure 6.15. In this case, the major benefit of considering the natural frequency is to reduce the uncertainty on the shear wave velocity in the second layer, as

shown by figure 18 which compares the Vs profiles over the first 20 m with and without the joint inversion. For the chosen model, no improvement has been found on the bedrock depth (figure 17), as the peak frequency is mainly controlled by the shallow seismic contrast at 10 m.

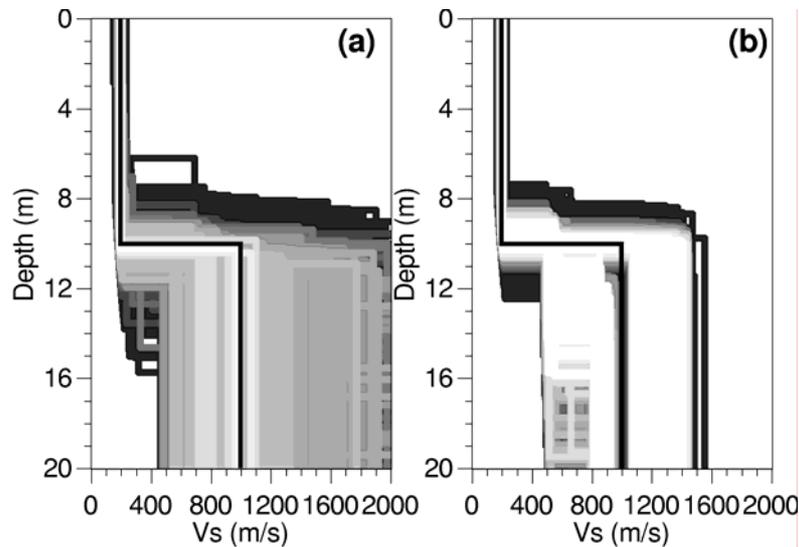


Figure 18 : Comparison between Vs profiles over the first 20 metres resulting from (a) the inversion with a three-layer model over a restricted frequency range (figure 13) and (b) the joint inversion of the dispersion curve and the H/V peak frequency (figure 17).

In summary, the introduction of any prior and reliable information has a positive effect on the inversion results, by eliminating models fitting the data and reducing the parameter uncertainty. The neighbourhood algorithm allows external data to be introduced easily in the inversion process, either in the parameterization of the model or in the inversion itself. The reliability of this prior information has to be carefully checked and the data uncertainty should be known or estimated before inversion.

Mode jumping

For both ambient vibration and active source experiments, the dispersion curves of higher Rayleigh modes are sometimes measured. The presence of higher modes depends upon the depth and the type of acting sources, in relation with the stratigraphy (Aki and Richards 2002, Xia et al. 2003, Socco and Strobbia 2004). Even if higher modes can be inverted in a way similar to the fundamental one, a correct dispersion curve inversion assumes the different modes to be clearly identified. As shown by Socco and Strobbia (2004), the measured apparent dispersion curve can jump from the fundamental mode to higher modes in real conditions. In our 3 layer synthetic model, the fundamental and the first higher mode for Rayleigh waves are very close to each other around 9 Hz (figure 10(b)). Depending on to experimental conditions and on the energy of the two modes, the apparent dispersion curve can be determined with a branch below 9 Hz corresponding to the fundamental mode and another branch above 9 Hz following the first higher mode. This situation is depicted in figure 19 where the observed apparent velocity is marked by black dots.

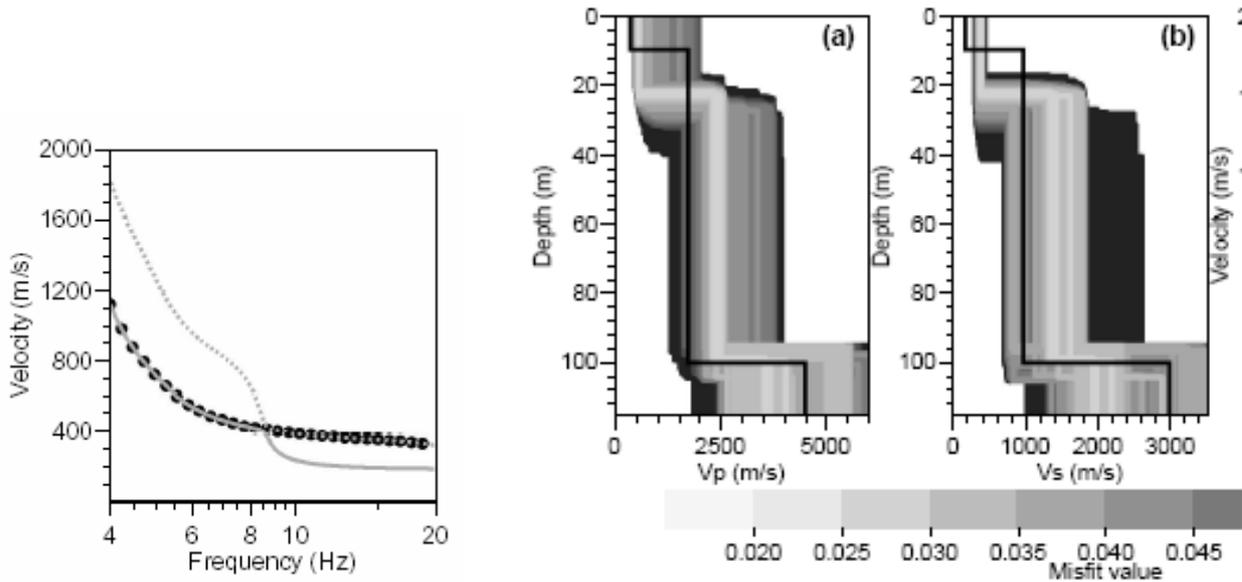


Figure 19: (top) Composite dispersion curve. The black dots represent the dispersion as it can be observed. The grey line are the theoretical dispersion curves of the fundamental (plain line) and the first higher (dotted line) modes.(bottom, (a) and (b)) Vp and Vs profiles resulting from the inversion of the composite dispersion curve.

At first glance, the obtained curve may be interpreted as a single fundamental mode. This curve is inverted as the fundamental Rayleigh mode with a prior information that the depth of the basement is situated between 95 and 105 m like in the inversion plotted in figure 15. The results of five runs are summarized in figure 19. The black lines in figures 19(a) and 19(b) are the theoretical ground model. The difference is especially strong on the first 20 meters where the velocity profiles are usually well retrieved. There are more than 50% of bias in the obtained results. For real sites, this phenomenon can be detected only if external data of a prior knowledge are also available, because there is no argument to reject the interpretation of figure 18 from the dispersion curve itself. If the results of the inversion with the fundamental Rayleigh mode are far from the expected profiles, the inversion with other Rayleigh modes can be tested with an inversion algorithm we developed to automatically identify higher modes (Wathelet, 2005). The inversion with this option requires only one data curve and the estimate of the number modes (nm) that are encountered by the data curve. For each generated model and for each frequency sample of the data curve, nm modes are simultaneously calculated. Compared to usual inversions, the misfit is computed in a completely different way. The velocity difference ($dvi = v_{di} - v_{ci}$) at each frequency between the data velocity and the theoretical Rayleigh velocities of each mode (up to nm) are calculated. Only the minimum value is kept in the summation of equation 6.

Virtually, the best fitting mode may be different for each frequency sample. However, these kinds of oscillations are never observed, as the curves are usually smooth enough to naturally restrict the

number of mode changes to one or two on the available frequency range.

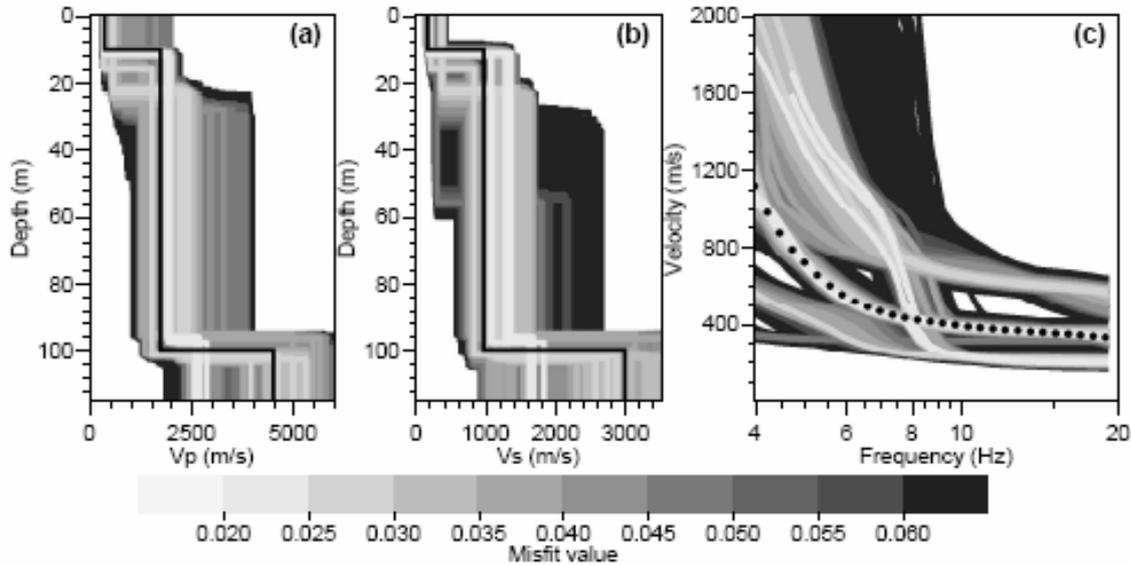


Figure 20: Inversion of the composite curve with mode identification. (a) Vp profiles. (b) Vs profiles. The black lines are the theoretical velocity profiles. (c) Fundamental and first higher mode dispersion curves corresponding to models of figures (a) and (b). The black dots are the composite dispersion curves used as the target curve during inversion.

The inversion method is tested on the dispersion curve displayed in figure 19 with the assumption that two modes may be present in the experimental curve. The frequency range of the dispersion curve (between 4 Hz and 20 Hz) is similar to the range used in figure 11 where little information below 10 m is recovered. The results of the five inversion runs (5×15100 models) are plotted in figure 19 where the two first modes are drawn for each model. Despite the messy aspect of figure 19c, several families of dispersion curves can be distinguished and the ensemble of models is split into four groups shown individually in figure 20. In the first group (figures 20a to 20c), the entire data curve is considered as being the first higher mode. The minimum misfit is higher than for the other groups, which does not mean that models are to be discarded. Valid arguments to reject these models would be that they exhibit a low superficial Vs value or a strong contrast between 40 and 60 m, which are not geologically admissible. In the second category (figures 20d to 20f) all models consider the data as the fundamental mode. Here again, complementary acquisitions about the superficial Vs or depth criteria could help to discard those models. In the third family of models, the data curve is also linked to the first higher mode but in a different way than in the first group. Finally, in the last group (figures 20j to 20l), the mode jump is detected around 9 Hz and the velocity profiles match better with the theoretical ground model.

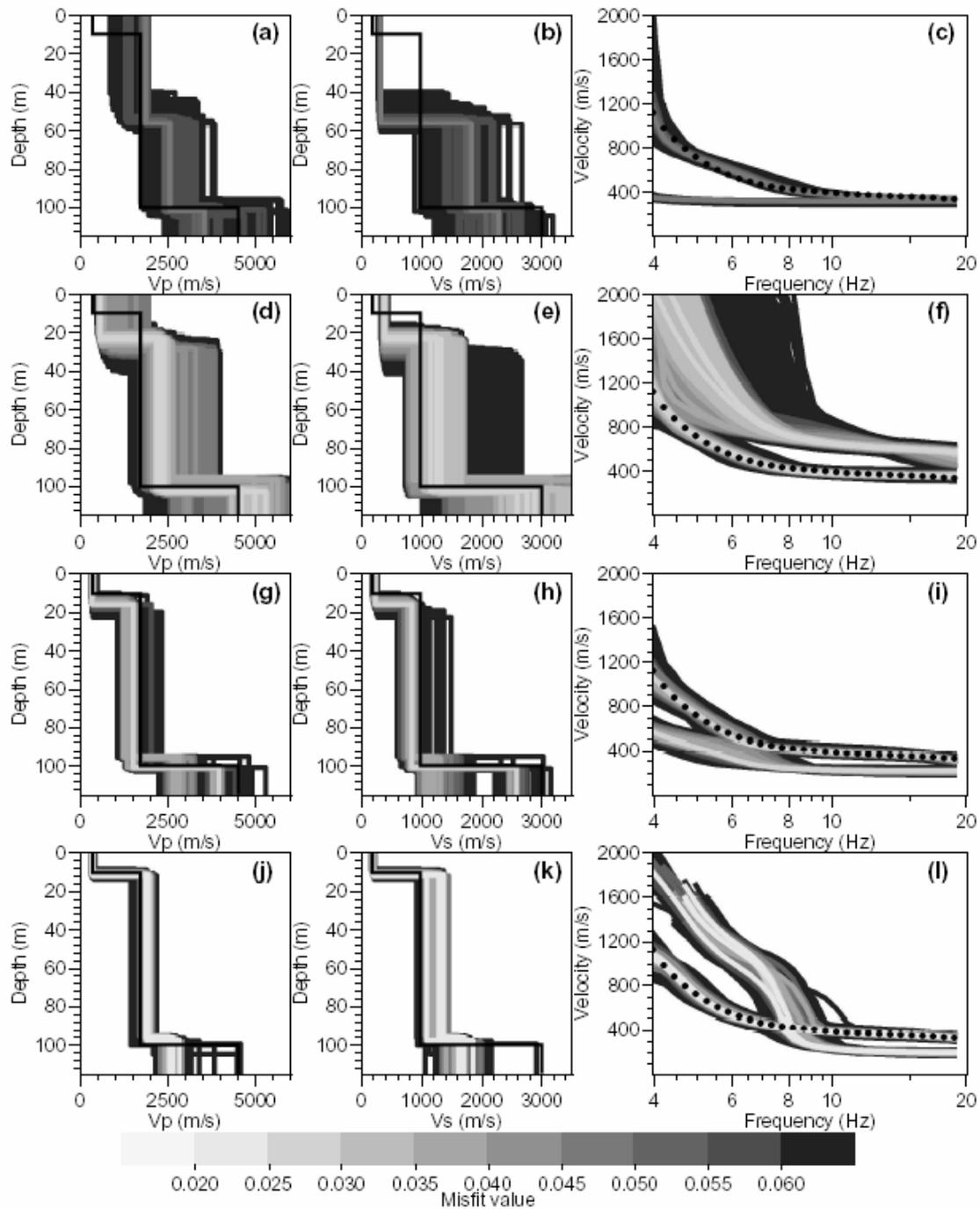


Figure 21 : Composite dispersion curve. The black dots represent the dispersion as it can be observed. The grey line are the theoretical dispersion curves of the fundamental (plain line) and the first higher (dotted line) modes. See details in the text.

4. Conclusions

Array measurements of ambient vibrations were extensively used and tested during the SESAME project, both on synthetic signals and on real data (five sites in Europe). Two types of processing techniques were tested: the f-k method and the spatial autocorrelation technique. The results and the experience gained during this 3-year project allows to give recommendations for quality array measurements and processing. First, the processing methods used during this project assume that the investigated site is made of horizontal layers with velocity values varying with depth. Before any deployment of a seismic array, the one-dimensional structure of the site has to be assessed from existing geological data or from the spatial stability of the H/V curves. The frequency content of the ambient vibrations has also to be checked and the sensor type adapted to the required frequency range. During acquisition, large time windows (duration greater than 25 cycles) are needed to improve the phase delay estimate and a large number of windows are necessary to obtain good statistics. No simple relation between the array geometry, the penetration depth and the usable frequency range of the dispersion curve was found. To address this problem, we computed the theoretical array response which yield two wave number limits (k_{\min} : resolution, k_{\max} : aliasing) defining a validity range for the dispersion curve. Tests on synthetic signals have shown that the dispersion curve estimations from both processing techniques were reliable within this wave number range. For a better control of the wavefield situation, the combination of both f-k and autocorrelation techniques are recommended. On the basis of the main conclusions, a field measurement methodology has been proposed for the in-field determination of the dispersion curve at a given site. The method is based on the deployment of a small array, the size of which is progressively increased in order to determine the dispersion curve piece by piece on narrow wavelength ranges. For the fundamental mode of Rayleigh waves, the lower frequency limit of the dispersion curve is given by the resonance frequency of the site (Scherbaum et al., 2003). The combination of two processing methods (the conventional f-k method and the SPAC technique) is highly recommended to increase the confidence about the results.

The inversion of a dispersion curve, truncated at the resonance frequency of the site for the fundamental mode of the Rayleigh wave, is a non-unique problem. To address this issue, prior information on velocity values or on the layer depth can be easily input in the inversion process. Also, the dispersion curve can be jointly inverted with dispersion data coming out from active measurements or with the frequency of the H/V peak. Introduction of external data as a priori information contributes to limit, sometimes drastically, the number of models fitting the data.

References

- Aki, K., 1957: *Space and time spectra of stationary stochastic waves, with special reference to microtremors*, Bull. Earthquake Res. Inst. Tokyo Univ., 35, 415-456.
- Aki, K. and Richards, P.G., 2002: *Quantitative Seismology, 2nd Edition*, University Science Books, Sausalito, California.
- Asten, M.W. & Henstridge, J.D., 1984: *Array estimators and the use of microseisms for reconnaissance of sedimentary basins*, Geophysics, 49(11), 1828-1837.
- Bettig B., Bard P.-Y., Scherbaum F., Riepl J., and Cotton F., 2003: *Analysis of dense array noise measurements using the modified spatial auto-correlation method (SPAC). Application to the Grenoble area*, Bolletino di Geofisica Teorica ed Applicata, 42(3/4), 281-304.
- Bonnefoy-Claudet S., Cornou C., Kristek J., Ohrnberger, M., Wathelet M., Bard P.-Y., Moczo P., Faeh D., Cotton F., 2004: *Simulation of seismic ambient noise: I. Results of H/V and array techniques on canonical models*, Paper No. 1120, XIII World conference on Earthquake Engineering, Vancouver, B.C., Canada, August 1-6, 2004.
- Cornou C., Kristek J., Ohrnberger, M., Di Giulio G., Schissele E., Guillier B., Bonnefoy-Claudet S., Wathelet M., Faeh D., Bard P.-Y., and Moczo P., 2004: *Simulation of seismic ambient noise: I. Results of H/V and array techniques for real sites*, Paper No. 1130, XIII World conference on Earthquake Engineering, Vancouver, B.C., Canada, August 1-6, 2004.
- Capon, J., 1969 : *High-resolution frequency-wavenumber spectrum analysis*, Proceedings of the IEEE, 57(8), 1408-1418.
- Forbriger, T., 2001: *Inversion flachseismischer Wellenfelder*, Dissertation, Institut für Geophysik, Universität Stuttgart.
- Forbriger, T., 2003a: *Inversion of shallow-seismic wavefields: I. Wavefield transformation*. Geophys. J. Int., 153, 719-734.
- Forbriger, T., 2003b, *Inversion of shallow-seismic wavefields: II. Inferring subsurface properties from wavefield transforms*, Geophys. J. Int., 153, 735-752.
- Harjes, H.-P., 1990: *Design and siting of a new regional array in Central Europe*, Bull. Seism. Soc. America, 80, 1801-1817.
- Haubrich, R.A., 1968: *Array design*, Bull. Seis. Soc. America, 58, 997-991.
- Herrmann, R.B., 1996: *Computer programs in Seismology, Version 3.0*.
- Horike M.; 1985: Inversion of phase velocity of long-period microtremors to the S-wave velocity structure down to the basement in urbanized areas, J. Phys. Earth., 33, 59-96.
- Horike M., "Geophysical exploration using microtremor measurements". Xth WCEE, Acapulco, 1996: Paper no. 2033, Elsevier Science Ltd.
- Ishida H., Nozawa T. and Niwa M.; 1998: Estimation of deep surface structure based on phase velocities and spectral ratios of long-period microtremors, Second International Symposium on the Effects of Surface Geology on Seismic Motion, Yokohama, Balkerna, 2, 697-704.
- Kind, F., 2002: *Development of Microzonation Methods: Application to Basle, Switzerland*, Dissertation ETH No. 14548, Swiss Federal Institute of Technology, Zurich, 110 p.
- Kind, F., Fäh, D. and Giardini, D., 2005: Array measurements of S-wave velocities from ambient vibrations, Geophys. J. Int., 160, 114-126. doi: 10.1111/j.1365-246X.2005.02331.x

- Kvaerna, T., and Ringdahl, F., 1986: *Stability of various f-k estimation techniques*, Semmiannual technical summary, 1 October 1985 – 31 March 1986, NORSAR Scientific Report, 1-86/87, Kjeller, Norway, 29-40.
- Miyakoshi K., Kagawa T. and Konioshita S., 1998, Estimation of geological structures under the Kobe area using the array recordings of microtremors, In: Irikura K., Kudo K., Okada H., and Satasini T. (eds), *The Effects of Surface Geology on Seismic Motion*, Balkema, Rotterdam, pp. 691-696.
- Köhler A., Ohrnberger, M., Scherbaum, F., Stange, S., and Kind, F., 2004: *Ambient vibration measurements in the southern Rhine Graben close to Basle*, *Annals of Geophysics*, accepted for publication, August 2004.
- Lacoss R.T., Kelly E.J., and Toksöz, M.N. "Estimation of seismic noise structure using arrays". *Geophysics* 1969: 34, 21-38.
- Matsushima T. and Okada H. "Determination of deep geological structures under urban areas using long-period microtremors", *BUTSURI-TANSA* 1990: 43(1), 21-33.
- Nguyen F., Rompaey G., Teerlynck H., Van Camp M., Jongmans D. and Camelbeeck T., 2004, Use of microtremor measurement for assessing site effects in Northern Belgium - interpretation of the observed intensity during the ms=5.0 June 11 1938 earthquake, *Journal of Seismology*, 8, 41-56.
- Ohrnberger, M., F. Scherbaum, F. Krüger, R. Pelzing and Sh.-K. Reamer, 2004a: *How good are shear wave velocity models obtained from inversion of ambient vibrations in the Lower Rhine Embayment (NW-Germany)*, *Boll. Geof. Teor. Appl.*, 45(3), pp. 215-232.
- Ohrnberger, M., Schissele E., Cornou C., Wathelet M., Savvaidis A., Scherbaum F., Jongmans D., and Kind F., 2004b: *Microtremor array measurements for site effect investigations: comparison of analysis methods for field data crosschecked by simulated wavefields*, Paper No. 0940, XIII World conference on Earthquake Engineering, Vancouver, B.C., Canada, August 1-6, 2004
- Ohrnberger, M., Schissele E., Cornou, C., Bonnefoy-Claudet, S., Wathelet, M., Savvaidis A., Scherbaum, F., and Jongmans, D., 2004c: *Frequency wavenumber and spatial autocorrelation methods for dispersion curve determination from ambient vibration recordings*, 13th WCEE, August 1-6, 2004, Vancouver, B.C., Canada, paper no. 946.
- Scherbaum F., Hinzen K.-G. and Ohrnberger M.; 2003: Determination of shallow shear wave velocity profiles in the Cologne, Germany area using ambient vibrations, *Geophys. J. Int.*, 152, 597-612.
- Tokimatsu, K., Tamura, S. and Kojima, H., 1992: Effects of multiple modes on Rayleigh wave dispersion. *Journal of Geotechnical Engineering*, ASCE, 118(10): 1529-1543.
- Tokimatsu, 1997: *Geotechnical site characterization using surface waves*, in: Ishiliara, Editor, *Earthquake Geotechnical Engineering*, Rotterdam, Balkema, 1333-1368.
- Wathelet M., Jongmans D., Ohrnberger M., 2004, Surface wave inversion using a direct search algorithm and its application to ambient vibration measurements, *Near surface geophysics*, 2, 211-221.
- Wathelet, M., D. Jongmans, and M. Ohrnberger (accepted in *Bulletin of Seismological Society of America*). Direct inversion of spatial auto-correlation curves with the neighborhood algorithm.
- Wathelet, M., D. Jongmans, M. Ohrnberger, and S. Bonnefoy-Claudet (submitted to *Geophysical Journal International*). On the determination of Vs profiles from ambient vibration-array recording: A. Synthetic case

Wathelet, M., M. Ohrnberger, and D. Jongmans (submitted to Geophysical Journal International). On the determination of Vs profiles from ambient vibration-array recording: B. Meuse alluvial plain, Liège, Belgium.

Woods, J.W. and Lintz, P.L., 1973: *Plane waves at small arrays*, Geophysics, 38, 1023-1041.

Yamamoto, H., 1998: An experiment for estimating S-wave velocity structure from phase velocities of Love and Rayleigh waves in microtremors, In: Irikura K., Kudo K., Okada H., and Satahara T. (eds), *The Effects of Surface Geology on Seismic Motion*, Balkema, Rotterdam, pp. 705-710.