Inversion of surface wave dispersion at European strong motion sites using a multi-model parameterization and an information-theoretic approach

G. Di Giulio, A. Savvaidis & N. Theodoulidis
Institute of Engineering Seismology & Earthquake Engineering, Thessaloniki, Greece

M. Ohrnberger & B. Endrun
University of Potsdam, Potsdam, Germany

M. Wathelet, C. Cornou, F. Renalier & P.Y. Bard
LGIT, Université Joseph Fourier, Grenoble, France

Ohrid
30.08-03.09
14 ECEE 2010

ABSTRACT
Within the scope of the EC-projects NERIES and ITSAK-GR we have applied a procedure able to combine a multi-model space parameterization and an information theoretic approach in analysis of dispersion curve inversion. In detail we considered the dispersion curve assessed at 14 strong motion European sites. At each site we investigated the model space through four different parameterization groups within the wavelength range estimated by actual dispersion curves. In order to explore the influence of model space we increased progressively the number of layers for each parameterization. We therefore addressed the model evaluation among a set of competing models obtained by inversion following the corrected Akaike's Information Criterion (AICc). By using such information-theoretic approach, we found an acceptable agreement between the inverted shear-velocity profiles of the best models and the available borehole results.

Keywords: surface waves, inversion of dispersion curves, Akaike's Information Criterion

1. INTRODUCTION

Surface wave methods are used in geotechnical engineering to estimate shear-velocity (Vs) profile that is the key parameter for site characterization. These methods are commonly adopted in near-surface studies and make use of both passive and active sources recorded by multi-receivers deployed in array configuration (1D linear or 2D layout). Independently of the array technique applied and of source used for retrieving a dispersion curve, a delicate task is the inversion of surface wave dispersion which is strongly dependent on the model space parameterization. Indeed the inversion of dispersion curve for deriving velocity profiles faces with a non-linear and non-unique problem between ground model parameters and observations. The final result of an inversion can be strongly affected -in addition to the performance of an inversion algorithm- by the starting model parameterization in terms of number of layers, range of velocities and depth, velocity law, Poisson's ratio and density allowed within each layer. Additionally to the model space exploration, also the selection of the most representative inverted models is often questionable when independent and reliable information are not available. In this study we focus on the two above issues: how to explore the influence of starting model parameterization and how to perform a ranking of the best class of models. We analyse several European strong motion sites where surface waves dispersion was inferred within the EC-Project NERIES (NEtwork of Research Infrastructures for European Seismology). We invert the dispersion curves exploring the model space through a multiple-model parameterization (Savvaidis et al., 2009). We adopt the Akaike's Information criterion (Burham and Anderson, 2001) in order to rank among models produced by inversion using different parameterization.

2. ESTIMATION OF DISPERSION CURVES AT NERIES SITES

We consider the phase-velocity dispersion inferred at 14 strong motion sites within the JRA4-TaskC of NERIES EC-Project. The purpose of this task is the developing of a low cost prototype tools for the geotechnical characterization of European strong motion sites and broad-band stations. In the following we show only a brief presentation of the selected sites as well as of data processing; details can be found in recent papers (Endrun et al., 2009; Renalier, 2010) and in the reports of NERIES project (deliverables JRA4-D2 and JRA4-D6). Five sites are located in Greece (Aigio, Knidi, Korinthos, Nestos, Volvi), three in Turkey (Bolu, Duzce, Sakarya) and six in Italy (Benevento, Buia, Colfiorito, Forli', Norcia, Sturno). The Vs structure at most of the sites is already known from available borehole data. The resonance frequency (f0) within sites was deduced by the peak of the H/V noise spectral ratios and is varying from 0.4 to 6 Hz. However, at some sites the estimate of f0 by spectral ratios is questionable because a clear peak of the H/V curves was not observed.

The dispersion curve at each site has been inferred combining 2D small-array measurements of ambient vibration as well as 1D active seismic surveys (i.e. MASW; Park et al., 1999). Ambient vibration 2D arrays consist of three-component seismological sensors (Le3d-5s with eigen-frequency of 0.2 Hz) recording from about 30 to 90 minutes. MASW experiments consist of 1D linear configuration of 24 vertical and horizontal 4.5 Hz geophones equally spaced between 1 and 5 m and recording on the average 5 shots. The surface wave dispersion at each site was inferred during the NERIES Project applying frequency-wavenumber (FK) and spatial autocorrelation methods (MSPAC; 3C-MSPAC) (Capon 1969; Lacoss et al., 1969; Aki, 1957; Kohler et al., 2007). In the aim of the present paper, we did not consider a direct inversion scheme of the autocorrelation curves (Bettig et al., 2001; Asten et al., 2004; Wathelet et al., 2005). The autocorrelation curves derived from MSPAC were translated in dispersion curves to address the inversion of phase-velocity curves only. To summarize, we use Rayleigh curves derived from FK, MSPAC and MASW, whereas Love curves were assessed by 3C-MSPAC and MASW methods. We include in the inversion process higher modes when they are clearly observed. The mean dispersion of each Rayleigh and Love mode were selected by averaging the dispersion curves estimated through the different array methods. Figure 1 shows an example of dispersion curves inferred at NESTOS site. The resolution limits were deduced in terms of wavenumber computing the theoretical array response function of the largest array, as explained by Wathelet et al. (2008). The upper frequency limits were usually given by MASW curves. Branches of dispersion curves that are ambiguous or deviating significantly from the others curves were neglected. The number (nf) of total points of final dispersion curves –including both Rayleigh and Love modes- ranges at each site from 55 to 129.

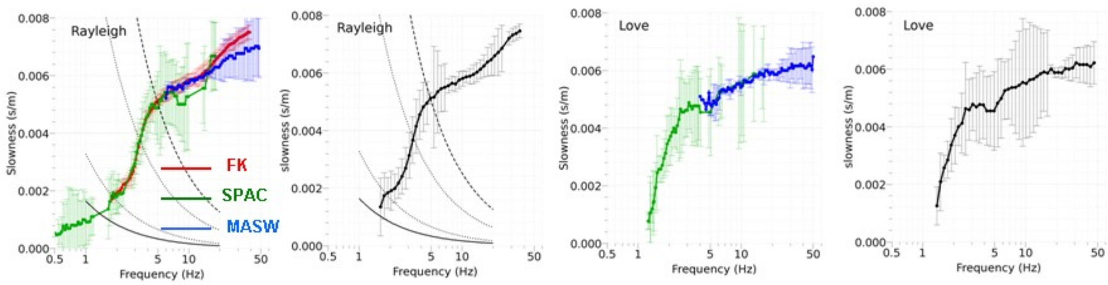

Figure 1. Dispersion curves inferred at NESTOS site for Rayleigh and Love waves using MSPAC, FK and MASW methods. The second and four panels show the mean dispersion after averaging the experimental curves.

3. MULTI-MODEL PARAMETERIZATION AND NEIGHBOOURHOOD INVERSION

We consider four distinct parameterization classes of the model space to invert the surface-wave dispersion. For each class we progressively increase the number of layers over the halfspace (hs). The four parameterization groups are:

a) uniform layering (1 to 9 layers over hs).

b) top-layer with a velocity-depth profile exhibiting power-law dependence. The Vs at a depth *z* is given by ($V_S(z) = V_0((1 + z - z_0)^{\alpha})$) where z_0 indicates the top of the layer and V_0 is the velocity at z_0.

This top-layer is approximated by five sub-layers and was overlaying 1 to 8 uniform layers (over hs).

c) top-layer composed of five sub-layers with a velocity-depth profile following a linear-law function (i.e., $\alpha = 1$). The top-layer is underlaid by 1 to 8 uniform layers (over hs).

d) Similar to a) but with a different definition of the thickness limits with the aim to increase the penetration depth. The distribution of thickness (h) is generated using a geometrical progression ($z_i = a \cdot h^i$) with the parameter a set to a quarter of the minimum measured wavelength.

We allow for each layer a Vs range from 50 to 2500 m/s that was increased to 150-3500 m/s for the hs. The compressional velocity (Vp) varies from 200 to 5000 m/s with the Poisson's Ratio uniform in the range 0.2-0.5. We fix the density to 2 t/m3 consistently with its low influence on surface wave dispersion. The thickness limits are defined through the wavelengths (lambda) derived from frequencies and phase-velocities of the actual dispersion curves. In the first three parameterization groups the bottom depth of each layer was varying between lambda_min/3 and lambda_max/3, where lambda_min and lambda_max are the minimum and the maximum measured wavelengths at the investigated sites (ranging typically from few meters to several hundred meters, respectively). In the last parameterization the thickness of first layer was fixed to lambda_min/4 and the final bottom depth to lambda_max/2.

We use the Neighbourhood Algorithm (Sambridge, 1999) as implemented by Wathelet (2008) in order to invert the experimental curves (*www.geopsy.org*). Neighbourhood Algorithm is a directed-search method for nonlinear inversion making use of Voronoi cells to investigate the multidimensional model space and to generate iteratively new random models inside the most promising cells. The tuning parameters are ni, ns and nr. A misfit function is first computed for the initial set of ni models. Within the nr cells with the lowest misfit a total of ns new models are added (ns/nr samples generated per cell). The last two steps are repeated N times resulting in a total of ni + N*ns models. The tuning parameters we used are ni = 100, ns = 50, nr = 50, and N = 4000 resulting in a total of 200100 models that are enough to reach the plateau-branch of the misfit trend. We repeat for each set of parameterization 5 runs (5 different seeds) in order to test the robustness of the results. The *Dinver* inversion software has produced on the average 80 Gb of binary report files at each site, with a machine time of about 24 hours on a Linux computer.

Inside each layer within the four parameterization groups the free parameters are Vs, the ratio Vs/Vp and the thickness in the limit range previously indicated. However, the most important parameter in surface wave inversion is Vs. In each layer, we link the Vp interface to the Vs interface, in order to have a large range of possible solutions without increasing excessively the number of degrees of freedom (dof). The dof of the model space of the four parameterizations ranges from 5 to 31 (Table 1).

Table 1. List of degrees of freedom (dof) associated to each model parameterization. UF, 1PL, 1L and GP indicates the model parameterization a), b), c), and d) described in the paragraph 3 of the text. The last number in the names of model parameterization indicates the number of uniform layers over the halfspace.

MODEL PARAMETERIZATION	UF_1 GP_1	UF_2 GP_2	UF_3 GP_3	UF_4 GP_4	UF_5 GP_5	UF_6 GP_6	UF_7 GP_7	UF_8 GP_8	UF_9 GP_9
DOF	5	8	11	14	17	20	23	26	29
MODEL PARAMETERIZATION	1L_1 1PL_1	1L_2 1PL_2	1L_3 1PL_3	1L_4 1PL_4	1L_5 1PL_5	1L_6 1PL_6	1L_7 1PL_7	1L_8 1PL_8	
DOF	10	13	16	19	22	25	28	31	

The misfit measure (m) between observed and theoretical dispersion curves is computed for each inverted model and is defined as

$$m=\sqrt{\sum_{i=0}^{n_f} \frac{(x_{di} - x_{ci})^2}{\sigma_i^2 \cdot n_f}} \tag{3.1}$$

where x_{di} and x_{ci} are the phase-velocity of observed and theoretical dispersion curve at frequency f_i, respectively. σ_i is the uncertainty of the datum at frequency i and n_f is the total number of samples. For sake of simplicity we generally do not allow the presence of low velocity zones (LVZ) except for a site (Korinthos) where borehole data indicate a strong velocity-inversion. We deliberately do not introduce the information from borehole data or resonance frequency indicated by the H/V curves. These a-priori data are used for comparison with the results obtained from inverted models.

4. MODEL RANKING BASED ON BIAS CORRECTED MISFIT (AKAIKE'S INFORMATION CRITERION)

The multi-model parameterization with an extensively inversion produce a large number of models showing similar value of misfit (Eqn. 3.1). Each model space within a parameterization class is then characterized by a different number of degrees of freedom (Table 1). For such problem of identification of the best estimate from an ensemble of acceptable models we use the Akaike Information Criterion (AIC). Akaike' idea was to relate the Kullback-Leibler information number, which indicates the information lost when an approximating model is used to explain the reality, to the maximum likelihood function (Kullback and Leibler, 1951; Akaike, 1974; Bozdogan, 2000). AIC estimator is indeed a measure of the lack-of-fit between approximating model and reality and penalizes the complexity of the model space. This is in according to the principle of parsimony which investigates the number of free parameters of possible fitting models in order to reach the best compromise between bias and variance. Models with lower value of AIC are considered to be better models. AIC or similar information-theoretic approaches do not require particular assumption on the experimental data and they can be used for problem of model decision in many applications (Burnham and Anderson, 2001).
In the case of least squares estimation with normally distributed errors as well as for small sample size adjustment, AIC is modified in AICc and expressed as

$$\mathrm{AICc} = nf \bullet \ln(m) + 2K + (2K(K+1)/(nf - K - 1)) \quad \text{(valid for } nf/K < 40\text{)} \tag{4.1}$$

where ln indicates the natural logarithm, nf is the total number of observations, m is the misfit (residual sums of squares divided by nf) and K indicates the number of free parameters. In our application m is defined by Eqn. 3.1, nf is the number of samples of the experimental dispersion curves, K is the degree of freedom of each considered model parameterization (listed in Table 1).

5. RESULTS

At each site we follow the Akaike criterion evaluating as a best set of models, among the results produced by multiple-model parameterization, the one with a lower value of AICc (Eqn. 4.1). Within each paramaterization class, we group the results of the five inversion runs (i.e. five seeds), then we plot the minimum AICc (and minimum misfit) as a function of the dof (i.e. of the number of layers progressively added over half-space) (Table 1). We select as best model parameterization the one showing the lowest AICc number. Each coloured curve in the top panel of Figure 2 indicates a parameterization class. The black vertical arrow of Figure 2 shows the lowest AICc that in this case is corresponding to a parameterization through 4 uniform layers (i.e. 14 dof). The colour of inverted models in the bottom panel of Figure 2 is proportional to the misfit carried out by the inversion. The black curves within the Vs profiles and the dispersion curves indicate the borehole and the input data, respectively. The orange curves show the forward modelled dispersion from the independent information. In Figure 2, the actual H/V noise spectral ratios (red curves) are also compared to the theoretical ellipticities (black curves) of the 100 models at lowest misfit within the "best" model-parameterization.

5.1 Sites with a Good Match between Model Profiles and A-Priori Independent Information

At three sites (Colfiorito, Nestos and Volvi) our inversion strategy provides a good fitting with the a-priori independent information (i.e. borehole data and f0). The minimum misfit is about 0.3 and the best model parameterization selected by the AICc minima is the uniform layering [(a) of paragraph 3] using 3 or 4 layers over halfspace depending by the site. The Vs30 from independent borehole information and inverted models is within the same soil class category (following the EC8 code). The actual H/V noise spectral ratios (red curves) are well matched by the theoretical ellipticities (black curves) of the best 100 models (Figure 2). These three sites, at the scale of the array measurements, are likely characterized by a 1D simple layering with a strong velocity contrast between the soft-to-bedrock interface. All the three sites are belonging to class C or D and the resonance frequency is clearly indicated by a strong peak of the H/V curves.

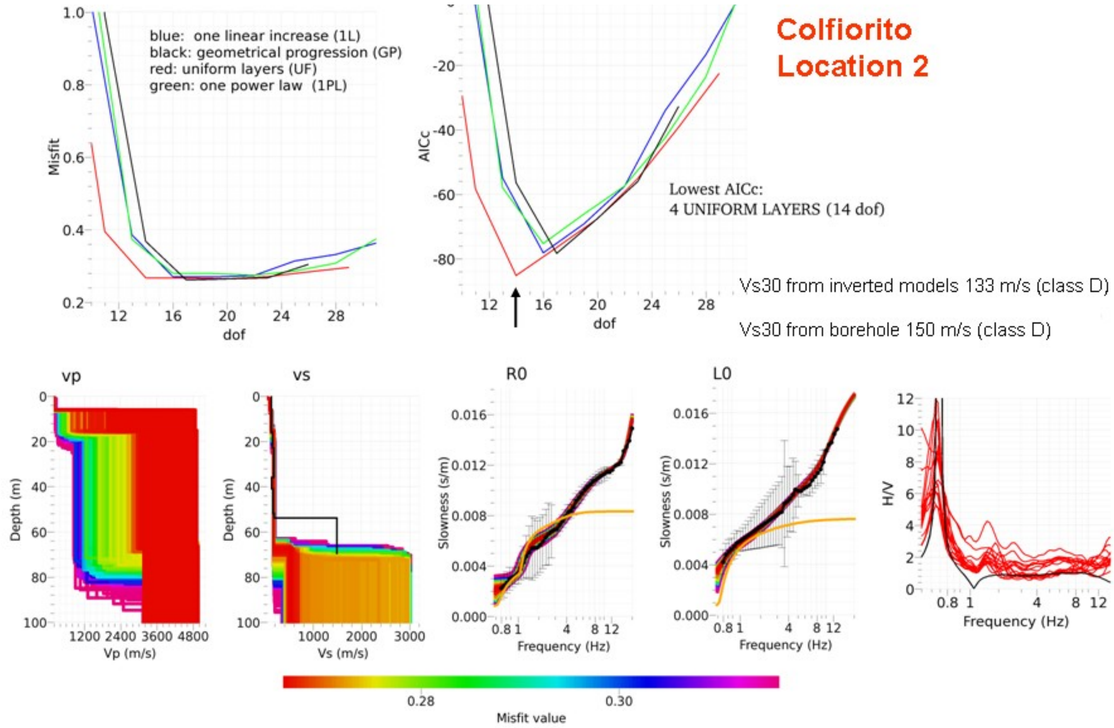

Figure 2. Top panel) Misfit (and AICc) trend versus the number of dof. Bottom panel) Inversion results considering the best model parameterization (4 uniform layers over half-space) following Akaike criterion; Vp and Vs model profiles, and fundamental Rayleigh and Love modes (R0 and L0). Note that the inversion at Colfiorito indicates a larger depth of the soft-to-bedrock interface than the depth by borehole data; this difference may be explained in terms of distance between borehole and array position (Di Giulio et al., 2006).

5.2 Sites with Ambiguity between the Model Profiles and A-Priori Independent Information

At some sites (Benevento, Buia, Norcia, and Sturno) the experimental dispersions are fairly well-fitted by the inversion. Our strategy provides a clear indication of the best parameterization using the AICc for model-selection ranking. Otherwise, the borehole profile and the experimental H/V curves are not completely matched from the inverted velocity models and from the theoretical ellipticites, respectively. Buia is an example of a significant mismatch between inversion results and a-priori borehole information (Figure 3). At this site there is a significant difference in terms of Vs30 because 320 m/s is indicated by inversion and 260 m/s is by borehole data (both values of Vs30 lead to the same soil class category considering the EC8 prescription; soil class C). The reasons of these discrepancies in terms of Vs30 are unknown. However, the shear-wave profiling from surface wave methods is more indicative of an enlarged area compared to borehole survey. Additionally the differences between inversion results and a-priori information could be ascribed to wrong mode identification or to the effects of a low velocity zone.

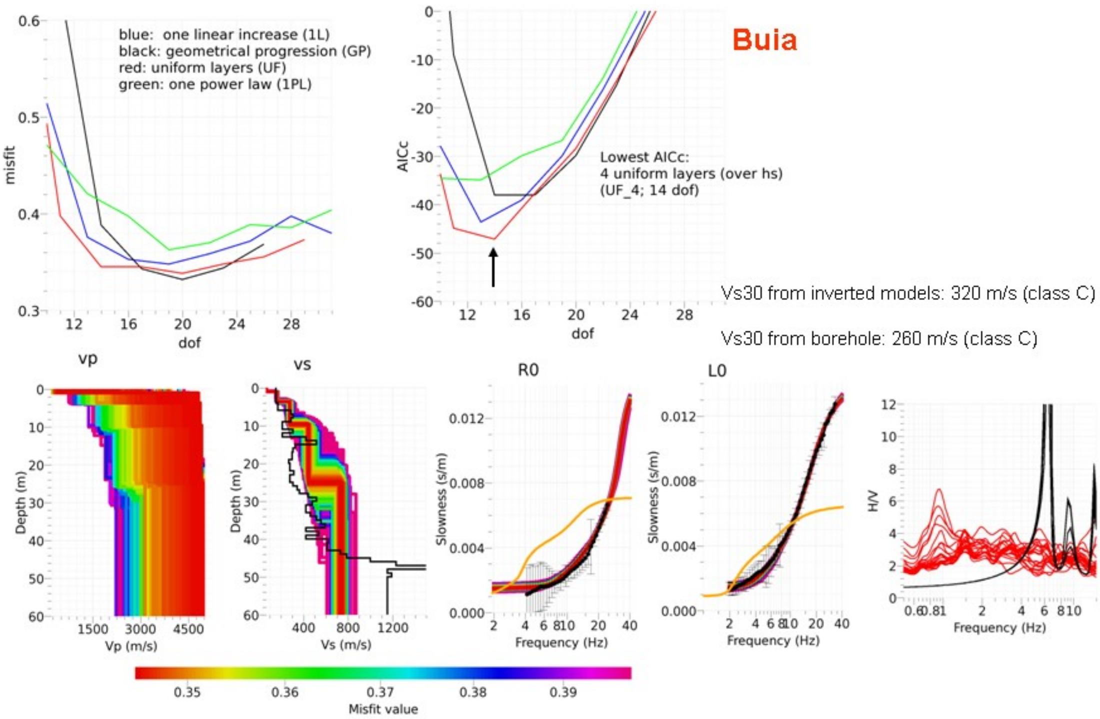

Figure 3. Top panel) Misfit (and AICc) trend versus the number of dof. Bottom panel) Inversion results considering the best model parameterization (indicated by the vertical arrow); Vp and Vs model profiles, and fundamental Rayleigh and Love modes (R0 and L0).

5.3 Sites with Low Velocity Zone

At Korinthos, the inversion without including low velocity zone returns not satisfactory results with a very high misfit (about 0.9) and without no clear indication of the AICc minima. The shape of the experimental R0 curve (Figure 4) can suggest an effect of Low Velocity Zone (LVZ), as indicated by the borehole measurements. In order to improve the results we repeat the inversion allowing LVZ. We use a model parameterization with uniform layering, from three to six uniform layers progressively added over hs. We allow LVZ for Vs in the third layer; this choice was driven by the borehole profile. The “best” parameterization including LVZ is with four uniform layers (14 dof, see Figure 4). The results including LVZ show a nice fit of the field dispersion curves with the theoretical one (misfit about 0.2). The Vs profiles from inversion indicate clearly a LVZ in fairly good agreement with borehole data. However, the H/V ratios are not fitted by the theoretical ellipticities mainly because we do no have estimates of phase velocities at low frequency (i.e < 0.8 Hz, in proximity to the peak of the H/V curves).

5.4 Sites with Misinterpretation of Modes

At Sakarya site, the inversion in terms of fundamental Rayleigh and Love modes (R0 and L0) provides a very large misfit (about 3) and no clear AICc minimum varying the model parameterization (see Figure 5). This is likely caused by an erroneous association of modes. Indeed the misfit decreases at about 0.3 and the AICc obtains a minimum using four uniform layers over hs (Figure 5) when the experimental Rayleigh dispersion is interpreted as first higher mode (R1). In this case the H/V curves are fairly fitted by the ellipticities and the Vs30 from available information approaches the Vs30 from inverted models (400 m/s and 449 m/s, respectively). Note that the reference Vs30 at this site was not provided by borehole data; it was derived from independent MASW measurements.

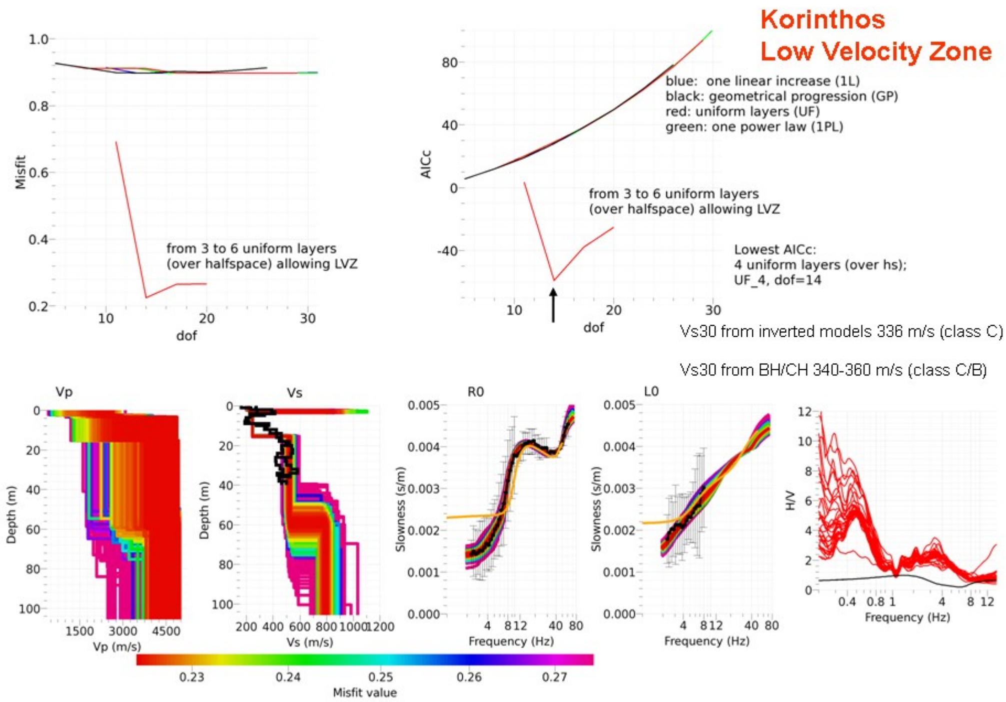

Figure 4. Top panel) Misfit (and AICc) trend versus the number of dof. The curves with highest misfit and highest AICc number refer to the inversion without considering LVZ; note the absence of AICc minima in these curves. Bottom panel) Inversion results in terms of Vp and Vs model profiles, and fundamental Rayleigh and Love modes (R0 and L0). The two black curves within Vs profile show the downhole and crosshole data.

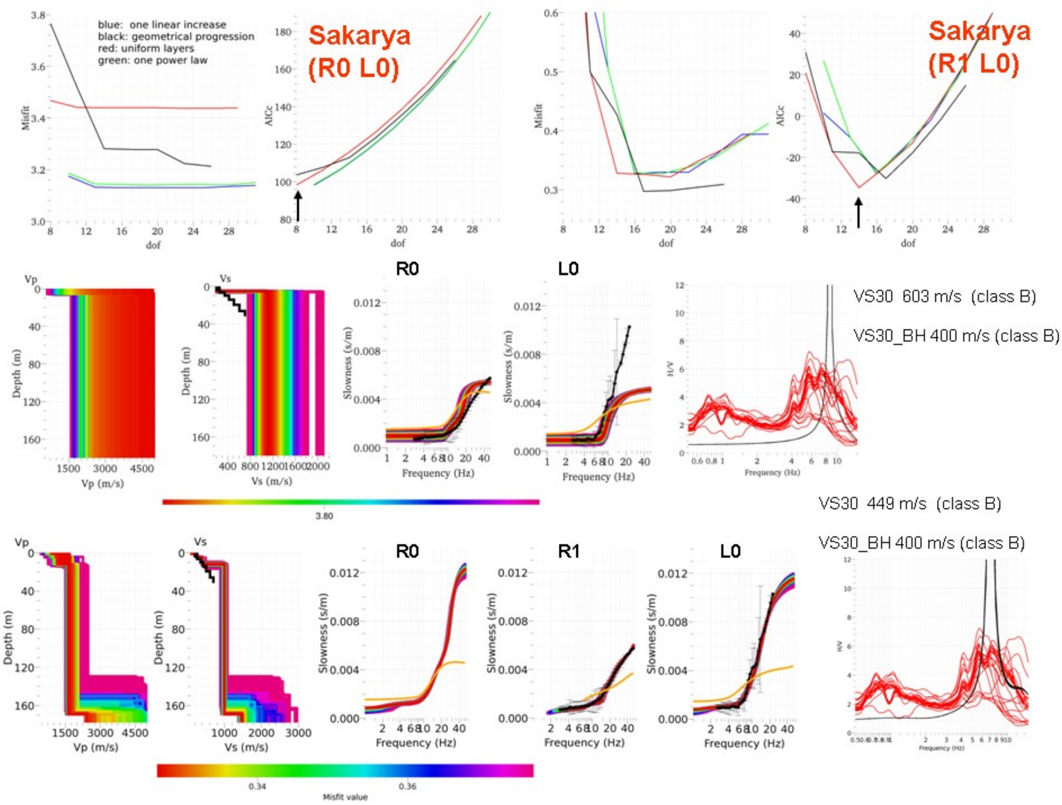

Figure 5. Top panel) Misfit (and AICc) trend versus the number of dof assuming two different modes interpretation (R0, L0 and R1, L0; on the left and right, respectively). Middle Panel) Inversion results assuming fundamental modes (R0 and L0). Bottom panel) Inversion results considering the R1 and L0 modes.

6. CONCLUSION

The strategy of surface wave dispersion that combines multi-model parameterization and the Akaike information-theoretic approach can be an effective tool for ranking best models among equivalent models in terms of misfit. The best match between inversion results and independent information are obtained at sites characterized by a strong soft-to-bedrock velocity contrast as suggested by the clear resonance frequency of the H/V noise spectral ratios. Otherwise the absence of clear minima of the AICc number varying the dof in conjunction with a large misfit can indicate wrong modes identification as well as effects related to low velocity zones.

REFERENCES

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Trans. Automat. Contr.* **19**, 716-723.

Aki, K. (1957). Space and time spectra of stationary stochastic waves, with special reference to microtremors. *Bull. Earth. Res. Inst. Tokyo Univ.* **25**, 415-457.

Asten, M. W., Dhu, T., and Lam, N. (2004). Optimized array design for microtremor array studies applied to site classification; comparison of results with SCPT logs. In Proc. of *13th World Conf. on Earthquake Engineering*, paper no. 2903, Vancouver, B.C., Canada.

Bettig, B., Bard, P.-Y., Scherbaum, F., Riepl, J., Cotton, F., Cornou, C., and Hatzfeld, D. (2001). Analysis of dense array noise measurements using the modified spatial auto-correlation method (SPAC). Application to the Grenoble area., *Boll. Geof. Teor. Appl.* **42**, 281–304.

Bozdogan, H. (2000). Akaike's Information Criterion and Recent Developments in Information Complexity, *Journal of Mathematical Psychology* **44**, 62–91.

Burham, K.P. and Anderson, D.R. (2001). Kullback-Leibler information as a basis for strong inference in ecological studies. *Wildlife Research* **28**, 111-119.

Capon, J. (1969). High-resolution frequency–wavenumber spectrum analysis, *Proc. IEEE* **57:8**, 1408–1418.

Di Giulio, G., Cornou, C., Ohrnberger, M., Wathelet, M. and Rovelli, A. (2006). Deriving wavefield characteristics and shear-velocity profiles from two-dimensional small-aperture arrays analysis of ambient vibrations in a small-size alluvial basin, Colfiorito, Italy. *Bull. Seism. Soc. Am.* **96:5**, 1915-1933.

Endrun, B., Ohrnberger, M. and Savvaidis, A. (2009). On the repeatability and consistency of three-component ambient vibration array measurements, *Bull. Earthquake Eng.*, doi 10.1007/s10518-009-9159-9.

Kohler, A., Ohrnberger, M., Scherbaum, F., Wathelet, M. and Cornou, C. (2007). Assessing the reliability of the modified three-component spatial autocorrelation technique, *Geophys. J. Int.* **168**, 779–796.

Kullback, S. and Leibler, R. (1951). On information and sufficiency, *Annals of Mathematical Statistics* **22**, 79-86.

Lacoss, R.T., Kelly, E.J. and Toksöz, M.N. (1969). Estimation of seismic noise structure using arrays. *Geophysics* **34**, 21–38.

Park, C. B., Miller, R. D. and Xia, J. (1999). Multi-channel analysis of surface waves (MASW), *Geophysics* **64**, 800–808.

Renalier, F. (2010). Caractérisation sismique de sites hétérogènes à partir de méthodes actives et passives: variations latérales et temporelles, *Ph.D. Thesis*, Joseph Fourier University, Grenoble, 224 pages.

Sambridge, M. (1999). Geophysical inversion with a neighbourhood algorithm: I. Searching a parameter space, *Geophys. J. Int.* **138**, 479– 494, doi:10.1046/j.1365-246X.1999.00876.x

Savvaidis, A., Ohrnberger, M., Wathelet, M., Cornou, C., Bard, P.-Y. and Theodoulidis, N. (2009). Variability analysis of shallow shear wave velocity profiles obtained from dispersion curve inversion considering multiple model parameterization. Abstract in *Seismological Research Letters* **80:2**, 354, SSA Meeting, April 2009, Monterey, USA.

Wathelet, M., Jongmans, D. and Ohrnberger, M. (2005). Direct Inversion of Spatial Autocorrelation Curves with the Neighborhood Algorithm. *Bull. Seism. Soc. Am.* **95**, 1787-1800.

Wathelet, M., Jongmans, D., Ohrnberger, M. and Bonnefoy-Claudet, S. (2008). Array performance for ambient vibrations on a shallow structure and consequences over Vs inversion. *J. Seism.* **12**, 1–19. doi:10.1007/s10950-007-9067-x

Wathelet, M. (2008). An improved neighborhood algorithm: parameter conditions and dynamic scaling. *Geophys. Res. Lett.* **35**, L09301, DOI:10.1029/2008GL033256