Exploring nonlinearity and nonuniqueness in surface-wave inversion for near-surface velocity estimation

HUBB DOUMA, ION Geophysical/GXT Imaging Solutions
MATTHEW M. HANEY, U.S. Geological Survey

With the rapid increase in prospecting for unconventional oil and gas, a large part of which remains on land, the demand for land seismic data processing has increased substantially and is expected to further increase in the future. It is known that land seismic data are often of much poorer quality than marine seismic data. This is to a large extent caused by the presence of unconsolidated rock in the near surface with often complex velocity structure, which is absent in many marine settings. Such near-surface variations cause the wavefield to scatter or even lose its coherence as it propagates. This makes it difficult to accurately image the deeper-lying targets in land seismic data.

Moreover, the data contain surface waves that have substantially larger amplitudes than the reflected body waves and as such are considered noise that hides part of the reflections needed for accurate imaging. Besides attempts to increase the data quality from a data-acquisition point of view, one can try to improve the imaging with a data-processing point of view and use the surface waves to invert for the near-surface (shear-) wave velocity structure (e.g., Xia et al., 1999). Once this velocity structure is known, it can be used to model the surface waves and subtract them from the data to increase the signal-to-noise ratio or, e.g., to calculate shear-wave receiver statics for converted-wave seismic imaging. In that way, the hampered imaging of the deeper lying targets can be improved.

Surface waves are sensitive to the Earth's properties up to a depth of roughly one wavelength. Therefore, with observed frequencies typically somewhere between 3 and 30 Hz, assuming typically observed velocities, they can be used to invert for the (shear-wave) velocity up to depths of 100–150 m (e.g., Ivanov et al., 2006; Muyzert, 2007). Recently Haney and Douma (2012) inverted group- and phase-velocity Rayleigh-wave dispersion curves for the near-surface shear-wave velocity using a perturbational approach applied to the forward method known as the thin-layer method (Lysmer, 1970; Kausel, 1999).

The forward problem of modeling dispersion curves for surface waves, however, remains nonlinear. In linearized inversions, this inherent nonlinearity is evident because the sensitivity kernels are dependent on the model parameters, making the inversion dependent on the starting model. Through the mere acceptance of uncertainties in the data by fitting the data up to a certain tolerance, the problem also becomes inherently nonunique. Even though linearized inversions are in the daily practice of exploration geophysics often the only practical option, they cannot deal with nonlinearity and nonuniqueness. To get a feeling for the nonlinearity and nonuniqueness of surface-wave inversion in an exploration geophysical context, we compare the results obtained from a linearized inversion with that of a nonlinear search technique. For this article, we illustrate our methods and results by inverting a single dispersion curve obtained from field data. We focus on the inversion of the phase velocity only.

Data

Figure 1a shows a receiver gather that was band-pass filtered between 2 and 15 Hz, with offsets between 50 and 300 m. By limiting the offsets to a narrow region around the receiver gather, spatial averaging or path effects are minimized, such...
that the final inverted shear-wave velocity profile is indeed mostly associated with the local area around the receiver gather. The surface (Rayleigh) wave is clearly visible and can be used in a slant-stacking (or beamforming) procedure as a function of frequency to obtain the phase-velocity spectrum (Van der Kruk, 2007). Because of the irregular distribution of offsets and azimuths, the surface wave has a somewhat “ragged” appearance.

Figure 1b shows the associated velocity spectrum and the picked dispersion curve (white line). The dispersion curve was picked through finding the global maximum in the velocity spectrum and following the maximum in a search window in the direction of both increasing and decreasing frequency. In this way, mode jumping can be minimized, although this receiver gather does not contain higher modes beyond the fundamental mode. A good quality dispersion curve could be obtained only over a narrow bandwidth (here 4–8.25 Hz). The dispersion curve was picked at a frequency interval of 0.25 Hz.

**Linearized inversion results**

For the inversion, we chose a layer thickness for the forward modeling of 3 m. With a maximum frequency of 8.25 Hz in the dispersion curve and minimum observed phase velocity of about 350 m/s, this layer thickness is less than a tenth of the estimated shortest wavelength of 350/8.25 = 42 m. This ensures the accuracy of the thin-layer method used in the forward modeling (Kausel, 1999). We imposed an exponential smoothness constraint on the inversion using a model-covariance matrix with a 1/e length of 9 m (about a quarter of the smallest wavelength) and a model standard deviation of 34 m/s. The data standard deviation was set to 10 m/s for all frequencies.

The Rayleigh-wave phase-velocity is primarily sensitive to the shear-wave velocity. Therefore, in practice, only the shear-wave velocity can be reliably estimated in phase-velocity dispersion-curve inversion. The sensitivity to the P-wave velocity and the density are taken into account by assuming a constant $V_P/V_S$ ratio and using Gardner’s relation to relate density to the P-wave velocity.

To set up a linearized inversion based on the thin-layer method, a linearized perturbation analysis is used. Because the matrix in the obtained linear inverse relation is sparse, we use the LSQR algorithm of Paige and Saunders (1982). Convergence is established when $\chi^2 < 1$.

We used two different starting models for our linearized inversion: a linearly increasing shear-wave velocity with depth (Figures 2a and 2b), and a smoothed version of a low-velocity layer on top of a half-space (Figures 2c and 2d). As can be seen in Figure 2e, both final models fit the data well (within one standard deviation because convergence was found for both models with $\chi^2 < 1$). When comparing both models (Figure 2f), we note that the models differ mostly in the upper 20 m and below 50 m. When looking at the update of the models as a function of iteration (Figures 2b and 2d), we notice that below 70–80 m both final models are not much different from their respective initial models.

Figure 3 shows the mode shapes for the fundamental-mode Rayleigh wave for both the horizontal and vertical particle velocities, using the linear initial velocity model. It is clear from Figure 3 that below 70–80 m there is little sensitivity
Nonreflection seismic and inversion of surface and guided waves

to the model parameters. This lack of sensitivity is caused by the absence of low frequencies (< 4 Hz).

The results from this linearized inversion emphasize the nonlinear character of the surface-wave dispersion-curve inversion; the inversion can get stuck in a local minimum of the objective function, making the final model dependent on the starting model. Without knowledge of the uncertainties in the model parameters, there is no way of knowing which model to choose, as both models fit the data equally well (within one standard deviation, $\chi^2 < 1$). To try and get an idea of the uncertainty in the model parameters, we employed a nonlinear search method to search the model space for an ensemble of models that fit the data equally well.

**Nonlinear search and the neighborhood algorithm**

There are many different methods to search a model space, amongst which are genetic algorithms, neural networks, and simulated annealing, to name but a few. Here we chose the neighborhood algorithm (NA) by Sambridge (1999). The choice of this algorithm was motivated by its simplicity in both parameterization and implementation as well as its previous use in surface-wave dispersion-curve inversion (e.g., Huang et al., 2010).

Figure 4 explains how the NA works for the simple case of a two-parameter model space. The algorithm is initialized with 10 random models (a), and $N_r = 2$ regions of lowest misfit are selected as shown by the green regions in (b) and each repopulated with $N_s = 1$ new model as shown by the red dots in (c). The algorithm proceeds by calculating the misfits of the newly added models and continues to again select the $N_r = 2$ regions of lowest misfit as shown by the orange regions in (d) and repopulate each of these regions with $N_s = 1$ new model as shown by the green dots in (d). The algorithm proceeds in the same fashion until a (user-defined) maximum number of iterations is reached.
neighborhood regions repopulated with \( N_i \) new models each. The algorithm proceeds onward in the same fashion until a maximum number of iterations is reached.

The NA thus zooms in on the regions of model space that most reduce the misfit. It is able to avoid getting trapped in a local minimum because it resamples multiple nearest neighborhood regions with the lowest misfits. As such, it has the potential to find a global minimum, although convergence to this minimum, as with most search algorithms, is not guaranteed.

To reduce the number of parameters, we enforced a regularized inversion grid where three consecutive layers have the same shear-wave velocity (and thus the same density and compressional-wave velocity). This gives an effective layer thickness of \( 3 \times 3 = 9 \) m which is about 1/4 of the smallest (approximate) wavelength of 42 m. In this way, we reduce the size of the model space considerably while maintaining the necessary resolution. The forward modelling grid still has a smallest layer thickness of 3 m, ensuring the accuracy of the thin-layer method.

The NA was initialized with 2500 uniformly distributed random models. At each iteration, we decided to repopulate the \( N_r = 100 \) best nearest neighborhood regions with two additional models, leading to a total of 200 added models per iteration. The total number of iterations was set to 50. This resulted in 12,500 models, shown by the gray lines in Figure 5c. The bounds of the search space are clearly visible. Figure 5a shows the velocity spectrum and the measured dispersion curve (white dashed line), as well as all the dispersion curves (gray lines) associated with each model in the total ensemble. It is clear that most models have associated dispersion curves that do not fit the data. For comparison, the models and their associated dispersion curves resulting from both linearized inversions are shown.

We chose a subset of the generated ensemble that best fit the data. This resulted in an ensemble of 75 models shown by the black lines in Figure 5c. Their associated dispersion curves are shown in Figure 5b, and it is clear that these models fit

![Figure 5](image-url)

**Figure 5.** Ensemble of shear-wave velocity models (gray lines) generated by the NA (c) and their associated dispersion curves as shown by the gray lines in (a), and the ensemble of best 75 models as shown by the black lines in (c) and their associated dispersion curves as shown by the black lines in (b). Note that the ensemble of best shear-wave velocity models fit the data well (b). The white dashed line in (a) and (b) indicates the measured dispersion curve. For comparison, both models from the linearized inversion and their associated dispersion curves are shown by the blue and magenta curves in (b) and (c).

![Figure 6](image-url)

**Figure 6.** Cartoon illustrating that the variation of models in an ensemble (green ellipse) is largest in the direction of the long axis of the ellipse (a), and smallest in the direction of the short axis (b). To obtain the most robust features in the ensemble (i.e., the ones that vary the least), the ensemble needs to be projected onto the short axis of the ellipse (c).
the data well (cf. the dispersion curves related to both models obtained using the linearized inversion shown by the magenta and blue curves in Figure 5b). The individual models exhibit a trade-off relation between velocity and depth because no a-priori smoothness constraint was imposed on the models, as opposed to the linearized-inversion case. Note that both models obtained using the linearized inversion appear to be contained within the bounds of the ensemble of 75 best models.

**Ensemble inference**

Once we have an ensemble of models that fit the data to a prescribed tolerance, we can ask what all these models have in common. Hence, instead of looking for an optimum solution as in the linearized inversion case, we now ask what robust information we can infer from the ensemble. To do so, we proceed in the same way as Douma et al. (1996).

Suppose that, for the sake of argument, we can depict our ensemble of models as an ellipse in a two-dimensional model space (Figure 6). In that case, it is clear that the models vary the most in the direction of the long axis of the ellipse, and the least in the direction of the short axis. That is, the most robust pattern that the models have in common is defined by the short axis of the ellipse. Hence to extract these patterns from the ensemble we must project the ensemble onto the short axis of the ellipse.

The major axes of the ellipse are found by the eigenvectors of the covariance matrix of the models, where the shear-wave velocities as a function of depth are treated as the independent parameters and the different model realizations as different measurements of these parameters. The eigenvector with the largest eigenvalue is then related to the long axis of the ellipse, while the eigenvector with the smallest eigenvalue is related to the short axis. The eigenvalue itself is a direct measure of the variability of the models in the direction of the associated eigenvector.

In the simple case of the cartoon in Figure 6, it is easy to determine which eigenvector to keep because there are only two such vectors: one with a bigger eigenvalue and one with a smaller value. We would keep the small eigenvalue and project the ensemble onto its associated eigenvector only. In case we have more than two parameters, such as in our inversion case, we need to look at the eigenvalue spectrum and see if we can identify a knee-point or a break in the eigenvalue spectrum, above which the eigenvalues are substantially larger or become larger more rapidly as a function of the eigenvector index. In that case we project the ensemble onto the eigenvectors with eigenvalues below the knee-point.

We can perform this ensemble inference on both the whole ensemble generated using the NA (gray lines in Figure 5c) and on the ensemble of best models only (black lines in Figure 5c). At first sight, doing this for the whole ensemble might not seem meaningful, because there are only two such vectors: one with a bigger eigenvalue and one with a smaller value. We would keep the small eigenvalue and project the ensemble onto its associated eigenvector only. In case we have more than two parameters, such as in our inversion case, we need to look at the eigenvalue spectrum and see if we can identify a knee-point or a break in the eigenvalue spectrum, above which the eigenvalues are substantially larger or become larger more rapidly as a function of the eigenvector index. In that case we project the ensemble onto the eigenvectors with eigenvalues below the knee-point.

We can perform this ensemble inference on both the whole ensemble generated using the NA (gray lines in Figure 5c) and on the ensemble of best models only (black lines in Figure 5c). At first sight, doing this for the whole ensemble might not seem meaningful, because this ensemble contains mostly models that do not fit the data (Figure 5a). However, knowing that the NA focuses its model space sampling on those regions that reduce the misfit the most, it follows that the most robust features in this ensemble are related to the parameters that are best resolved by the data.

Plotting the eigenvalues of the covariance matrix of the whole ensemble identifies a knee-point in the eigenvalue spectrum (Figure 7a) around eigenvector index 17 or 19. Projecting the whole ensemble onto these first 17 or 19 eigenvectors gives the projected ensemble shown in blue in Figure 7b and Figure 7d, respectively. The projected ensembles indicate that the shear-wave velocity from 15 to 75 m is best resolved by the data. From the linearized inversion, we already knew that
there was little resolution below 70–80 m, but the ensemble inference applied to the whole ensemble generated using the NA search method seems to further reveal that the shallow shear-wave velocity is poorly constrained by the data.

To confirm this somewhat surprising observation in the context of the linearized inversion, we calculated the resolution matrices for the linearized inversions for both final models (Figure 8). If resolution would be perfect, the resolution matrix would be the identity matrix. For both models, however, it is clear that at shallow depths (< 20 m) the resolution matrices contain substantial side lobes away from the diagonal. This is consistent with the above conclusion drawn from the ensemble inference on the whole ensemble of models (Figure 7) and explains why for the linearized inversion the models obtained with different starting models differ considerably at shallow depths (Figure 8c).

![Resolution matrices for the linearized inversion with both the smoothed layer over a half-space initial model (a) and the linear initial model (b).](image)

**Figure 8.** Resolution matrices for the linearized inversion with both the smoothed layer over a half-space initial model (a) and the linear initial model (b). Above 20 m, the resolution matrices contain strong side lobes away from the diagonal and below 70 m the matrices are near zero. At the depths between 20 and 70 m, the resolution matrices are reasonably diagonal, indicating good resolution at this depth range only. The final models obtained using the linearized inversion for both starting models are shown in (c) for comparison.
We attribute the lack of resolution at the shallow depths (< 20 m) to the maximum frequency in our dispersion curve data being limited to 8.25 Hz. It appears that without higher-frequency surface waves the shallow shear-wave velocity cannot be accurately determined. This is important, since the majority of the problems with the near surface often stem from the complexity of the Earth in the first 10–20 m. Therefore, when using dispersion-curve surface-wave inversion to try and find this shallow complexity, it is important to focus on methods of dispersion-curve estimation that maximize the largest usable frequency in the data. Possible measurement and inversion of other physical quantities (such as the H/V ratio) could be considered when inverting for the shallow shear-wave velocity.

By applying the ensemble inference to the ensemble of 75 best models, we obtain the eigenvalue spectrum shown in Figure 9a. The only clear break in the eigenvalue spectrum appears to be at eigenvector index 23 (Figure 9c). By projecting the ensemble on the first 23 eigenvectors, we get the projected ensemble shown by the green lines in Figure 9d. This projection reduces the model parameter uncertainty for depths shallower than about 70 m when comparing the range of values of the shear-wave velocity in the projected ensemble (green lines) with the original ensemble of best 75 models (black lines) but not much for depths below 70 m. When, using instead the first 14 eigenvectors for the projection (Figure 9b), we see that the uncertainty is further reduced, while the final models obtained from the linearized inversion remain contained within the bounds of the projected ensemble for the well-resolved depth range of 15–75 m. Hence, both models obtained from the linearized inversion seem to be consistent with the robust information in the ensemble. This confirms that the linearized inversion seems to perform well in the well-resolved depth range of 15–75 m. As such, it seems both models are both good estimates of the shear-wave velocity at those depths.

We observe that in the well-resolved part of the model space, for the projections in Figure 9b and Figure 9d, the models resulting from the linearized inversion tend to be on the low end of the shear-wave velocity range indicated by the projected ensemble. This is likely caused by the smoothness constraint that was imposed on the linearized inversion, because this constraint is known to include a minimization of the norm of the model (Yanovskaya and Ditmar, 1990). This causes the linearized inversion to tend toward the model with the minimum shear-wave velocity. No smoothness constraint was imposed on the nonlinear search.

The filtered ensemble shows a range of possible shear-wave velocities at each depth. Therefore, the filtered ensemble can be used to estimate the uncertainty of the shear-wave velocity. However, the range of estimate shear-wave velocities depends on the number of eigenvectors used to filter the ensemble. When there is a clear knee-point or break in the eigenvalue spectrum, it is easy to find the number of eigenvectors on which to project. However, when no clear knee-point or break is present, it becomes harder to determine how many eigenvectors to use for the projection. To avoid ambiguity in choosing the number of eigenvectors to use and thus determining the uncertainty in the model parameters, a more objective criterion would be desirable.

**Conclusion**

We have explored the nonlinearity and nonuniqueness of the inversion of fundamental-mode Rayleigh-wave dispersion curves by comparing linearized inversion results based on the finite-element thin-layer method to the results obtained from
ensemble inference of an ensemble of models generated through a nonlinear search method (the neighborhood algorithm). The ensemble inference applied to the whole ensemble highlighted the depths that were well resolved, and emphasized the need for high frequency dispersion measurements to be able to resolve the near-surface (<10–20 m) shear-wave velocity. In the well-resolved depth-range, the results from two linearized inversions with different starting models compared well with the results obtained from the nonlinear search.

**References**


**Acknowledgments:** We thank ION Geophysical/GXT Imaging solutions for permission to publish these results.

**Corresponding author:** hdouma@iongeo.com

**OMNI 3D 2013**

With OMNI 3D 2013 seismic survey design & modeling software you can plan, execute, monitor, and analyze multiple land, marine, transition zone, vertical seismic profile, and multicomponent surveys in the same project simultaneously on a Windows® operating system.

**Visit us at EAGE stand no. 960**

www.GEDCO.com/OMNI3D