Multicomponent land data pre-processing for FWI: a benchmark dataset

Raul Cova, Bernie K. Law and Kris Innanen
CRWES/University of Calgary

Summary

Successful full-waveform inversion (FWI) studies using multicomponent marine data have been often reported in the literature. However, FWI applications to multicomponent land data remain limited. Among the challenges for a successful FWI in this setting we can find source repeatability, receiver coupling, rough topography, near-surface heterogeneities and strong elastic and attenuation effects. Due to the difficulty of including all these effects during the inversion process, it is a common practise to minimize their imprint on the data by conditioning the data before the inversion. Here we present a framework to study the effect of this conditioning on the FWI output. We first address the convenience of using a finite difference algorithm to include the topography during the forward modelling and its implications on the modelling of surface-waves. We compared this with the output of a spectral element method. A benchmark dataset was also created to understand the effect of the data conditioning processes on the FWI output. The goal was to obtain a synthetic dataset that would include some of the effects that are observed in the real multicomponent land data. This dataset will be used for understanding to what extent the conditioning of the data affects the inversion output and what strategies can be used to minimize their imprint in the inversion process.

Introduction

One of the features that difficult FWI on land data is the presence of complex irregular topography. In conventional processing, the effect of topography is removed by applying static corrections. However, this is usually done under the assumption of surface-consistency which might be invalid when near-surface velocities change gradually or where there exist strong variations in elevation. Including the topography leads to a second challenge in the application of FWI to land data, which is the introduction of surface-waves. This wave mode accounts for a significant part of the energy recorded on real datasets. Therefore, FWI practitioners are faced with the decision of either removing them from the data and ignore them during the FWI process or use them and exploit the information they carry.

Other characteristics of actual land data that difficult FWI are the poor signal-noise ratio, unknown and variable source signature and poor source/receiver coupling to the ground. In this study, we present our initial approximations to address some of these challenges. We first compare the use of a finite difference algorithm versus a spectral element method to incorporate the topography in the modelling of multicomponent seismic data. We pay special attention to the modelling of surface waves in each case. We then computed a set of synthetic data using well log data and a tomographic near-surface velocity model. Surface-consistent short-wavelength static corrections, deconvolution operators and amplitude balance corrections were extracted from real seismic data recorded nearby the wells. The inverse of these operators was then applied on the synthetic data to approximate some of the characteristics present on the real data. This dataset can be used for further understanding of the effects of data conditioning on the FWI output.

Data modelling: Finite difference vs spectral element method

To study the convenience of using a finite difference (FD) algorithm to simulate elastic wave propagation from the topography we built the velocity and density models displayed in Figure 1a, 1b and 1c. The actual topography from the Hussar 2D-3C seismic line (Margrave et al., 2011) was used to separate the air layer with parameters Vp = 310 m/s, Vs = 0 m/s and ρ = 1.25 Kg/m^3, from the near-surface layers. The code used for the elastic forward modelling was developed by Bohlen (2002). To avoid dispersion effects and to honour the CFL condition (Courant et al., 1967), we used a grid size of 1x1m for the finite difference modelling and time steps of $1x10^{-4}$ s to simulates seismic records of 1 s.

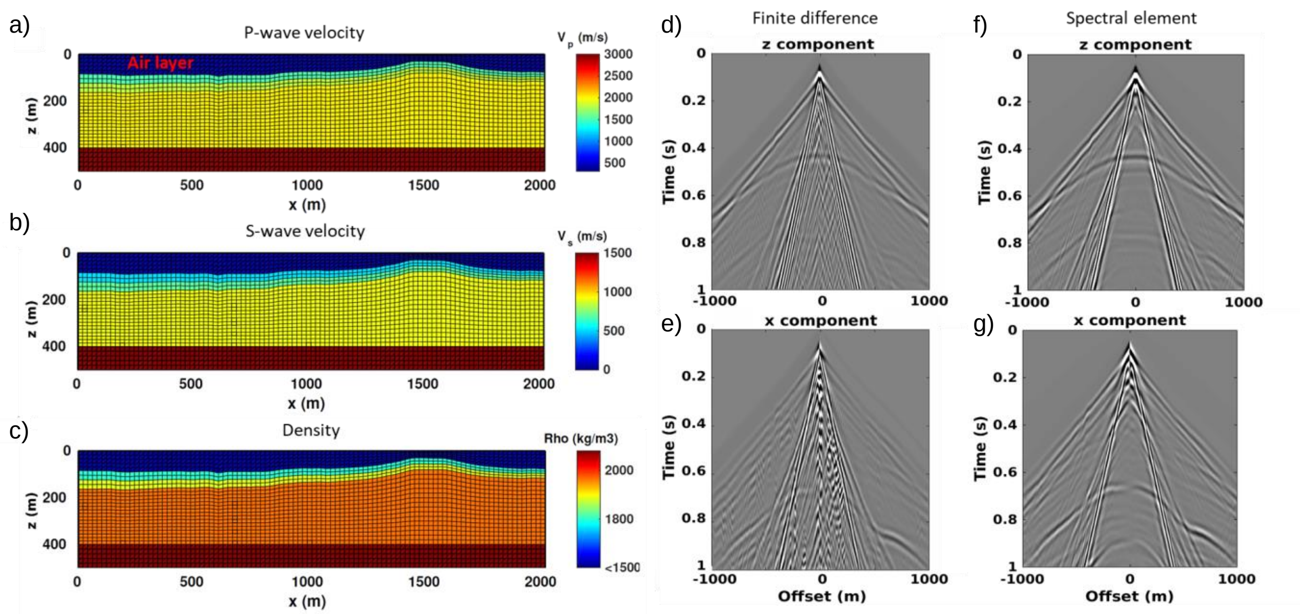

Figure 1. a) P-wave velocity, b) S-wave velocity and c) density models used to compute synthetic data. d) Vertical and e) horizontal component data computed using a finite difference algorithm. f) Vertical and g) horizontal component data computed using a spectral element algorithm.

Similar models were meshed and input into a spectral element algorithm (Ampuero, 2012). The spectral element method (SEM) offer the possibility to introduce deformed grids for the simulation of wave propagation. The ability to mesh the model conforming defined structural features allows for a better representation of the topography. As a result, a more accurate placement of the receivers over the topographic profile can be achieved and free-surface conditions along the topography can be better honoured.

Figures 1d to 1g display the horizontal and vertical component data obtained using the finite difference and the SEM algorithm. Even though the models used for the FD modelling were smoothed using a radial Gaussian smoothing operator with a half width of 5 m, there is still significant backscattered surface-wave energy within the surface-wave cone on both components. Moreover, the moveout of the events on the horizontal component data are distorted by short-wavelength static effects. These are the results of the discretization of the model in squared cells. Given the very low velocity in the S-wave velocity model after smoothing, a 1 m elevation change is translated into a S-wave static effect of 4 ms. On the other hand, the data computed using the SEM algorithm present a very clean character. No significant back-scattered S-wave energy is present in the data. Moreover, the moveout of the events is gradually deformed following the topography.

Despite the SEM algorithm is computationally more expensive than the FD algorithm, the flexibility provided by the SEM mesh allowed us to obtain results at a reasonable computation cost. The mesh we defined had a maximum distance between nodes of 10x10 m with 4 elements between nodes. After reshaping the mesh by including the topography and the geometry of the near-surface layers, the maximum and minimum distances between nodes obtained were of 3.42 m and 0.51 m, respectively. The time step was set at $2.4x10^{-4}$ s to satisfy the CFL stability condition. The average computational time we obtained was of 119 s per source record, which contrasts with the 1140 s it took for the FD algorithm to complete a simulation.

Although smoothing the FD grids helps to reduce the amount of backscattered surface-wave energy, the near-surface profile is altered in the process. This has an important effect on the dispersion spectrum of the surface-waves. Figure 2 shows the results of the modelling of surface wave data using the FD and the SEM algorithms. Of particular interest are the dispersion spectra displayed on Figures 2g and 2h. Notice how both dispersion spectra differ at frequencies larger than 15 Hz. The dashed line represents the analytical solution (Lai, 1998) for the original two-layers near-surface model. In Figure 2h the analytical solution falls very close to the maximum energy of the spectrum. However, when the model is smoothed the dispersion spectrum is significantly changed (Figure 2g). The shorter wavelengths (higher frequencies) now travel at lower velocities than before. This effect will difficult the proper inversion of the surface-wave data if velocity models are smoothed during the inversion process.

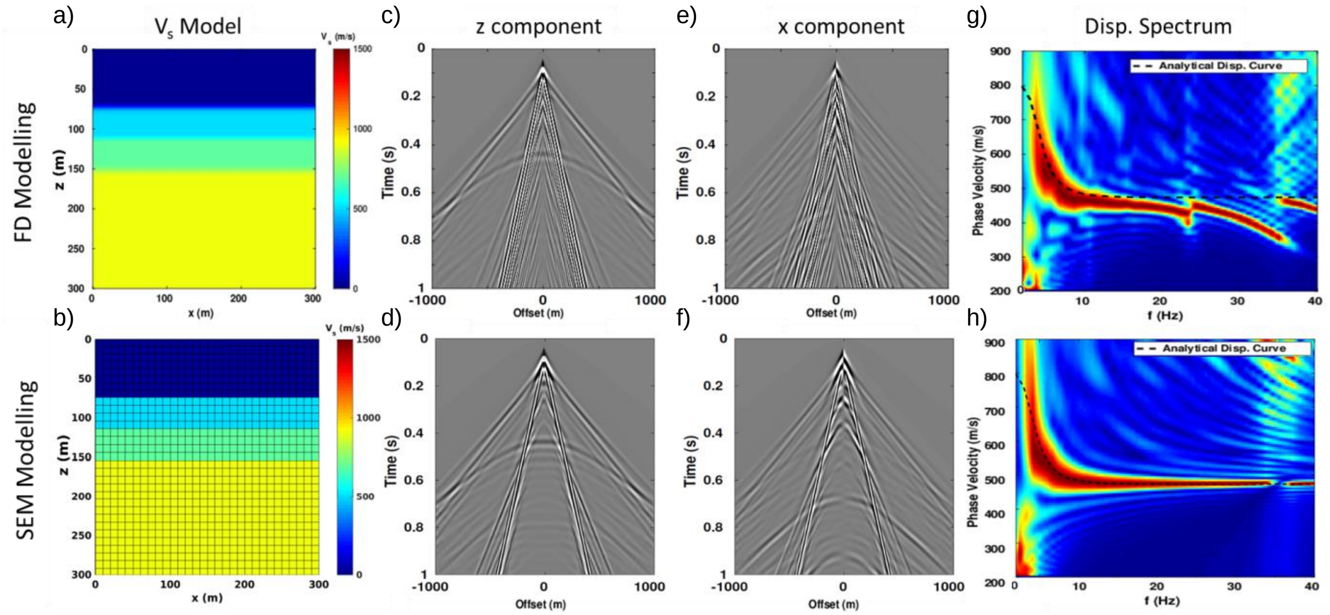

Figure 2. From left to right, near-surface velocity model used to compute synthetic data, vertical and horizontal component data obtained and the dispersion spectrum computed from the vertical component data. The top row are the results after using a FD algorithm and the bottom row shows the results with a SEM algorithm.

Pre-processing benchmark dataset computation

To study the effect of the conditioning of the data before performing a FWI we designed a controlled experiment using modified synthetic data. Figure 3a shows the velocity and density models created from the extrapolation of the well log data available along the Hussar 2D-3C seismic line (Margrave et al, 2011). The near-surface P-wave velocity model was computed using the results from a tomographic inversion of first arrivals (Law and Trad, 2017). This provided velocity information from the topography up to a depth of 150 m, approximately. A fixed Vp/Vs ratio of 4 was assumed to compute the S-wave velocities in the near-surface while the density was set constant at 2000 Kg/m^3 for this interval.

Surface-consistent corrections were extracted from both components of the field data and their effects were imposed on the synthetic data. Since the long-wavelength component of the static effects should be included in the near-surface velocity models we used to compute data, only the short-wavelength component of the statics solution was imposed after modelling the data. Surface-consistent amplitude balancing scalars were also used to descale the synthetic data introducing amplitude variations similar to the ones observed on the field data. Finally, a set of minimum phase filters were computed from the surface-consistent deconvolution operators obtained from the field data. We then applied these filters to the synthetic data in order to include the spectral variability from the field data on the synthetic traces.

The horizontal and vertical component source gathers obtained after applying this series of modifications are displayed on Figure 3e. There we can see how the frequency content of the original synthetic data (Figure 3d) has been degrade and the amplitudes of the traces has been altered. The effect of the residual statics is also manifested in the presence of jitter in the moveout of the events. Figure 3f displays the horizontal and vertical component data from the actual seismic line. An amplitude scaling of $t^{1/2}$ was applied to the field data to approximate the amplitudes modelled in our 2D simulation. Notice that the surface-wave data on the actual records show larger amplitudes than in the modelled data. To properly model the surface-wave data we will need to compute a more accurate S-wave velocity model for the near-surface. In terms of the moveout of the reflections and frequency content of the data, the simulated land data resemble the field data very closely. There are certainly many physical effects that were not included during the modelling. Particularly, anelastic attenuation may have an important effect on the decay of the recorded amplitudes, specially on the horizontal component data. Overall, the simulated land data now contains some of the characteristics that we can find in the actual data and they can be used to challenge our FWI algorithms.

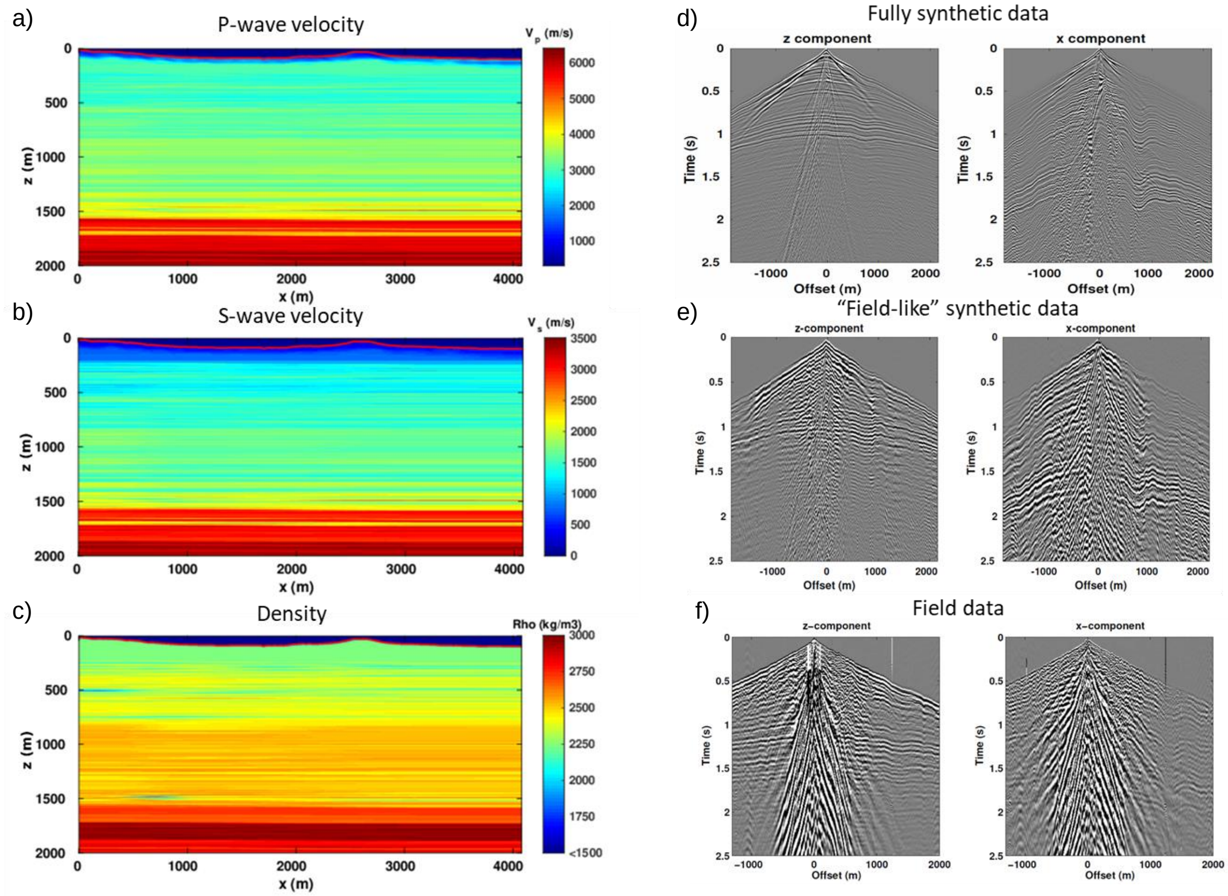

Figure 3. (a) P-wave velocity, (b) S-wave velocity and (c) density models generated with the well log data available along the Hussar 2D-3C seismic line and a tomographic near-surface velocity model. d) Fully synthetic source gathers and (e) "field-like" synthetic gathers obtained after imposing the surface-consistent effects observed in the real data. (f) Horizontal and vertical component data recorded on the field.

Conclusions

Using rectangular grids for the finite difference modelling of elastic waves demands of very fine cell sizes that may render the FWI process prohibitive in terms of computational cost. Smoothing the model parameters partly avoids the need for a finer grid but it introduces artificial velocity changes into the modelling. This introduces significant changes in the modelling of the surface-wave data. The SEM algorithm used in this study provided accurate elastic modelling while honouring the shape of the topography, without introducing significant numerical noise in the data. Moreover, the ability of the SEM algorithm to adapt to the geometry of the interfaces included in the model allowed us to optimize the computational burden of the modelling. Therefore, the spectral element method seems to be an attractive alternative for the modelling and inversion of multicomponent land data. We also computed a set of synthetic data that resembles more closely the character of actual land data. We achieved this by extracting short-wavelength static corrections, amplitude balancing scalars and deconvolution operators from the field data and applying their inverse on the synthetic data. The result was a dataset that contains some of the features that are usually removed during the pre-processing of the data. This dataset can be used to design optimal strategies for multicomponent data conditioning in elastic FWI projects.

Acknowledgements

The authors thank the sponsors of CREWES for continued support. This work was funded by CREWES industrial sponsors and NSERC (Natural Science and Engineering Research Council of Canada) through the Canada First Research Excellence Fund.

References

Ampuero, J. P., 2012, SEM2DPACK - A Spectral Element Method for 2D wave propagation and fracture dynamics, with emphasis on computational seismology and earthquake source dynamics. Version 2.3.8. URL https://sourceforge.net/projects/sem2d

Bohlen, T., 2002, Parallel 3-D viscoelastic finite-difference seismic modelling: Comput. and Geosci., 28, No. 8, 887–899.

Courant, R., Friedrichs, K., and Lewy, H., 1967, On the partial difference equations of mathematical physics: IBM Journal, 215–234.

Lai, C., 1998, Simultaneous inversion of Rayleigh phase velocity and attenuation for near-surface site characterization.

Law, B., and Trad, D., 2017, Robust refraction statics solution using feedback from reflection data: SEG Technical Program Expanded Abstracts 2017, 2666–2670.

Margrave, G. F., Mewhort, L., Phillips, T., Hall, M., Bertram, M. B., Lawton, D. C., Innanen, K., Hall, K.W., and Bertram, K., 2011, The hussar low-frequency experiment: CREWES Research Report, 23.