Introduction

A proper characterization of the near-surface geophysical properties is a major task for oil and gas exploration. A proven methodology for estimating the shear-wave velocity model (S-model) of the shallow subsurface is Rayleigh surface wave inversion. The dispersive behaviour of the propagation modes associated with these waves can be described by phase velocity changes with frequency (dispersion curves), which is the information used to derive the S-model through inversion (Socco et al., 2010).

A critical aspect of the methodology is picking the dispersion curves from the seismic measurements. A common approach is to transform the time-offset seismic gathers to frequency-phase velocity spectra and to manually pick the magnitude maxima in the transformed domain. For each frequency, the picked maxima correspond to the phase velocities of the surface wave fundamental and higher modes. This manual picking step is time consuming, highly subjective and represents a cumbersome task for modern large seismic surveys.

Recent efforts have been made to implement dispersion curve picking through machine learning (ML) techniques. Zheng and Miao (2014) presented an automatic method to pick dispersion curves in the frequency-phase velocity spectrum based on binarization and thinning. Their approach makes two important assumptions that typically do not hold for field data, namely: i) propagation modes are not masked by high amplitude noise, and ii) samples of the fundamental mode have higher magnitudes than those of the other modes. To overcome these limitations, Alyousuf et al. (2018) proposed a dispersion curve picker based on ML and in particular on a special class of deep neural network, called a deep belief network. This method combines initial unsupervised learning with human expertise on dispersion curve picking for a subsequent supervised phase. The method proved to be successful in a field data application in an arid environment. Supervised learning methods require training, and labelling the training dataset needs to be done manually. Neural network based solutions often do not generalize well, requiring additional training when applied to new datasets. Maselet et al. (2019) investigate the use of an unsupervised learning algorithm for near-surface characterization in the South of Oman. The described technique was used to improve the quality of already picked surface wave dispersion curves and, as such, it does not solve the problem in a fully automatic way.

In this paper we present a novel application of fully automatic surface wave dispersion curve picking using a clustering algorithm. The accuracy of the results and robustness of the approach are proved on synthetic data for the fundamental mode.

Method

Clustering algorithms can handle the problem of identifying surface wave propagation modes in the frequency-phase velocity spectrum. Ideally, when applied to big datasets, these algorithms need to be able to discover clusters of arbitrary shape with good efficiency and should have a low number of input parameters. Density-based spatial clustering of applications with noise (DBSCAN) is a well-known clustering algorithm, which is designed to deal with arbitrary shaped clusters, it is highly efficient on large datasets, it requires only one input parameter which is automatically derived (Ester et al., 1996). These properties, together with full automation and robustness in the presence of noise (due to its density-based clustering approach), make this algorithm a perfect candidate for solving the dispersion curve picking problem.

We integrated DBSCAN in our picking workflow, which can be summarized as follows: 1. The input seismic gathers are sorted in the domain, which shows least aliasing for the surface waves (in this example, the CMP domain) and transformed into frequency-phase velocity spectra; 2. windowing is applied to isolate specific frequencies and phase velocities, followed by low-pass filtering to remove high frequency noise; 3. for each frequency the maximum magnitude of the phase velocity spectrum is picked; 4. DBSCAN is applied to the cloud of 3D points representing the maxima previously picked for all the analysed frequencies and CMPs (or a predetermined portion of them); 5. each of the clusters

EAGE2020: Annual Conference Online
December 2020
discovered by DBSCAN represents points belonging to a surface corresponding to a surface-wave propagation mode. It is likely that some points are missing for some particular frequencies or CMPs. We therefore apply 3D linear interpolation to fill these gaps in and to be able to extract the dispersion curves at all the required frequencies for each CMP. Optionally, the results can be further refined by repeating steps 3 to 5 multiple times, using the output of step 5 as an initial guess for picking the magnitude maxima at step 3.

Example

We tested the new workflow on the SEAM Arid model synthetic dataset (Oristaglio, 2015), which has been specifically designed to be a realistic representation of the desert environment, including complex near-surface features, such as karsts, wadis, sand and other unconsolidated sediments, outcropping bedrock, and highly variable topography. The model extends laterally for 10 km and vertically for 3.75 km with a spatial sampling of 6.25 m in each of the three spatial dimensions.

We targeted the first 50 m of subsurface and Figure 1 is a cross-section of the S-model extracted at x=3.75 km, on top of which a high-resolution 2D survey, with 6.25 m shot spacing, has been simulated. The theoretical dispersion curves were generated (based on the forward modelling described by Lai et al., 2005) for a frequency range 5-50 Hz as a reference and they are shown in Figure 2. The seismic data has been sorted by CMP and transformed into phase velocity spectra. Figure 3 shows two such spectra with fundamental and higher order propagation modes, modelling noise and some artefacts, including aliasing. For the left hand spectrum, dispersion curve picking is less complicated due to less noise and more energy coherence for the different modes. Despite the fact that the proposed approach can also be applied to higher order modes, in this example we focused on picking the dispersion curve of the fundamental mode. We performed steps 3-5 followed by one re-picking iteration, obtaining the dispersion curves shown in Figure 4, which closely resembles the theoretical ones of Figure 2. A similar conclusion can be drawn by overlaying the estimated and the theoretical dispersion curves on the frequency-phase velocity spectra (Figure 3).

The final goal of the exercise is to derive an accurate S-model from the inversion of the picked dispersion curves. To this end, we first converted the picked (frequency, phase velocity) pairs into (pseudo-depth, shear velocity) pairs, which can be seen as samples of an approximate shear velocity distribution. We used an empirical relationship, referred to as the Lambda/3 approximation, also used in the steady-state Rayleigh wave method: assuming $\lambda$ is the wavelength associated with a particular frequency, the corresponding pseudo-depth is given by $\lambda/3$. The shear velocity is calculated by applying a multiplicative factor of 1.1 to the corresponding phase velocity (Brown et al., 2000). The S-model resulting from such a conversion is shown in Figure 5 and is a very good approximation of the true model of Figure 1. Finally, following Rovetta et al. (2018), we performed an inversion of the picked dispersion curves of Figure 4 using the model of Figure 5 as the initial model. The inversion converged after 10 iterations with a data RMS misfit reduction of about 50%. Figure 6 is the final result, which showed improved agreement with the true S-model relative to the starting model. Similar conclusions were reached by running additional inversion tests, such as using different initial models (i.e., velocity gradient) or using the theoretical dispersion curves as input data for the inversion.

Conclusions

A fully automatic method to pick surface wave dispersion curves in the frequency-phase velocity domain has been successfully developed and tested on synthetic data for the fundamental mode. The method integrates a density-based spatial clustering algorithm in the picking workflow, which is designed to handle arbitrary shaped clusters and that has been proved to be efficient when dealing with big datasets contaminated by noise. We are currently testing this approach on field data. If these tests prove successful, this work could help in the further adoption of surface wave inversion (or joint inversion) as a viable method for near-surface characterization for large seismic surveys.
Figure 1 A shallow cross-section of SEAM Arid shear-wave velocity model ($v_s$).

Figure 2 Dispersion curves (phase velocity $v_{ph}$ versus frequency $f$) associated with the Rayleigh wave fundamental mode, computed from the true model of Figure 1.

Figure 3 Examples of phase velocity ($v_{ph}$) - frequency ($f$) spectra of CMP gathers extracted by the seismic data generated from the true model of Figure 1. The black solid lines represent the theoretical dispersion curves associated with the Rayleigh wave fundamental mode; the stars correspond to their estimation through the ML approach.

Figure 4 ML estimation of the dispersion curves from the seismic data.
Figure 5 Estimation of the shear-wave velocity model through phase velocity-shear velocity conversion of the dispersion curves of Figure 4.

Figure 6 Estimation of the shear-wave velocity model through the inversion of the dispersion curves of Figure 4 and with the initial model of Figure 5.

References


