Separation of Multi-mode Surface Waves by Supervised Machine Learning Methods

Jing Li^{1}, Yuqing Chen², Gerard T. Schuster²*

¹*College of Geo-Exploration Science and Technology, Jilin University, Changchun
130026.*

²*Division of Physical Science and Engineering, King Abdullah University of Science
and Technology (KAUST), Thuwal, Saudia Arabic*

Corresponding author: Jing Li (inter.lijing@gmail.com)

ABSTRACT

Logistic regression, neural networks, and support vector machines are tested for their effectiveness in isolating surface waves in seismic shot records. To distinguish surface waves from other arrivals, we trained the algorithms on three distinguishing features of surface-wave dispersion curves in the $k - \omega$ domain: spectrum coherency of the trace's magnitude spectrum $D(k, \omega|x_s)$, local dip, and the frequency range for a fixed wavenumber k of the spectrum. Here, x_s is the coordinate of the shot along a recording line. Numerical tests on synthetic data show that the kernel-based support vector machines algorithm gives the highest accuracy in predicting the surface-wave window in the $k - \omega$ domain compared to neural networks and logistic regression. This window is also used to automatically pick the fundamental dispersion curve. The other two methods correctly pick the low-frequency part of the dispersion curve but failed at higher frequencies where there is interference with higher-order modes.

KEYWORDS

Multi-mode Surface Wave; Supervised Machine Learning Methods; Surface-wave dispersion inversion

INTRODUCTION

The dispersion curves for the fundamental mode (or higher-order modes) estimated from surface waves **can be used to** invert the subsurface S-velocity distribution $V_s(x, y, z)$ down to a depth of approximately 1/2 to 1/3 the dominant wavelength (Park et al., 1998; Li and Hanafy, 2016; Li et al., 2019). Such information is important for earthquake, seismology and near-surface engineering studies (Xia et al., 1999; Lin et al., 2008; Li and Hanafy, 2016). For example, earthquake seismologists can predict the horizontal-shaking intensity in a populated region by modeling the seismic responses with a known S-velocity model. These responses are used to assess the degree of earthquake risk at different sites. For highway construction, the estimated shear strength of the shallow subsurface is used to design the most durable highways. Soils with low shear velocity and shear strength will require a more robust road design.

Conventional 1D inversion of surface waves assumes a layered medium model (Xia et al., 1999; Socco et al., 2010). This is an efficient procedure that can be computed on **the laptop, but the inverted accuracy will be low when there are stronger** lateral variations in the near-surface velocity. To mitigate this problem, Li et al. (2017) presented the theory of inverting dispersion curves associated with handles arbitrary variations in velocity. The misfit function is the sum of the squared differences between the predicted and observed dispersion curves, and the inverted result is a moderate-to-high-resolution estimate of the near-surface S-velocity model. This procedure is denoted as wave-equation dispersion inversion (WD), which is less prone to getting stuck **into the local minima when** compared to full waveform inversion

(FWI) (Perez Solano et al., 2014; Schafer et al., 2014; Dokter et al., 2017). The inverted velocity model from WD can provide S-wave velocity models with much higher resolution than the conventional 1D inversion method (Maraschini et al., 2010).

Getting an accurate picking of dispersion curves is a key step in the WD method. This is difficult because the fundamental-mode arrivals **might be polluted by the** energy from higher-order modes and/or body waves. In the near-offset region, the surface waves are contaminated with strong near-field scattering and body-wave (Rix et al., 2001). This contamination severely distorts the surface wave dispersion curves at the low-frequencies (Bodeta et al., 2009).

To eliminate these unwanted interferences, a variety of methods have been developed for mode separation, such as adaptive-window muting and the linear Radon transform method (LRT). The LRT method is often effective in extracting reliable dispersion curves, but requires manual picking of the dispersion curves in the Radon domain (Luo et al., 2009). Manual picking is tedious and time-consuming **especially** **for** large amounts of data. In addition, the predicted dispersion curves must be picked at every iteration of inversion.

We now propose the use of machine learning methods to improve both the accuracy and efficiency in picking dispersion curves. In this paper, three machine learning (ML) algorithms are tested and compared, which are logistic regression, neural network (NN) and support vector machine (SVM) (Bishop, 2006). These methods are trained to predict a surface-wave window in the $k - \omega$ domain that will, theoretically, only accept surface waves and exclude any other types of energy. The mode of most interest

is the fundamental surface waves so the windowing parameters should be tuned to this mode. Once the surface-wave window is defined by the ML algorithms then the strongest amplitudes typically define the fundamental-mode dispersion curve.

The typical ML workflow for surface-wave separation is shown in Figure 1. The input image (on the left of Figure 1) consists of a raw shot gather, which includes surface waves, body waves and other scattered waves in the shot gather. Firstly, we extract the typical features of surface waves and build the training set. Then, we feed the training set into the supervised ML method to distinguish and classify the surface waves from other arrivals. Finally, we use the decision window to extract the surface waves in the $F - K$ domain and transform back to the time-domain to get a clean shot gather which only contains surface waves.

The next section presents an overview of the supervised machine learning methods of logistic regression, NN and SVM (Bishop, 2006). Then, the numerical test shows the results for isolating the surface-wave arrivals in CSGs. All the three ML methods are tested on the same data and the resulting dispersion curves are compared to those computed from a semi-analytic solution. A field data test is carried out on data recorded by a shallow seismic survey in Saudi Arabia. The last section presents a summary.

SUPERVISED MACHINE LEARNING METHODS

Figure 2a displays a 2D S-wave velocity model, and a corresponding vertical-component shot gather is shown in Figure 2b. The $x - t$ shot gathers $d(x, t|x_s)$ are transformed

into the $k - \omega$ domain by a 2D Fast Fourier Transformation (FFT) to get the magnitude spectra $D(k, \omega | \mathbf{s})$ (Figure 2c). To distinguish surface waves from any other arrivals, we trained the algorithms on three different features of dispersion curves in the $k - \omega$ domain. These features are the spectrum coherency of the magnitude spectra $D(k, \omega | x_s)$, the local-dip angle, and the frequency range of the spectrum. In our examples, each feature is computed by an analysis window with the size of 80×120 grid points centered at (k, ω) in the magnitude spectra.

The above three features are used to separate the surface waves from other events. For example, we recognize that the noise or other arrivals should have weak coherency, large dip angles, and low or high frequencies centered at the i^{th} pixel, while the surface-wave signals should mostly be characterized by strong coherency, modest dip angles, and a specified frequency range. More features in the data might be also used to improve the picking accuracy, but our results suggest that these three features are sufficient for surface wave separation.

1. **Coherency** x_1 . For a fixed wavenumber k , the magnitude spectra $D(k, \omega | x_s)$ is correlated in wavenumber with $D(k + \Delta k, \omega | x_s)$. Here, Δk is the wavenumber interval. The maximum correlation energy is assigned to the center pixel of the analysis window. The window is shifted by one pixel and the procedure is repeated for all pixels. As an example, Figure 2b displays a shot gather and a corresponding coherency image is displayed in Figure 3a. Each pixel is assigned a coherency feature value denoted by x_1 .
2. **Dip Angle** x_2 . The dominant dip angle in the analysis window is computed

by a local slant-stack method (Yilmaz, 2001). The angle associated with the strongest stacking energy is mapped to the center pixel of the analysis window. The window is shifted over by one pixel and the new slant-stack energy is computed. This procedure is repeated for all pixels. For the shot gather shown in Figure 2b, the dip-angle values are displayed in the color image shown in Figure 3b. Each pixel is assigned a dip-angle feature value denoted by x_2 .

3. **Frequency Range x_3 .** The surface waves only exist in a certain frequency range in the raw shot gather. This range of frequencies is identified and displayed in the color image shown in Figure 3c. Each pixel is assigned a peak frequency value denoted by x_3 .

The frequency spectrum image shown in Figure 2c is used for training. An interpreter assigns the target value $y = 1$ to an interpreted surface wave pixel, and the value $y = 0$ is labeled for a non-surface wave event at each point in the spectrum. 80 non-surface wave points are picked manually and 30 surface-wave points are selected randomly from this image. Only 10% of the pixels in this spectra are used as the training set and the others are used as the validation set. Each input at each pixel is a 3×1 vector where each element in the vector represents a feature value as explained above. Training is considered to be finished when we achieve an accuracy greater than 90% for the predicted values for the training data and the validation data. The well-trained support vector machine (SVM) model is then used to predict the class (either $y = 0$ or $y = 1$) of the untrained data at each pixel in the $k - \omega$ spectra. Any pixel in the $k - \omega$ spectra with the predicted value $y = 0$ is muted, while the pixels with $y = 1$

are preserved except near the edges of the surface-wave window which are tapered to prevent ringing in the $x - t$ domain.

The architecture of the neural network (NN) method is shown in Figure 4. For our tests, there is one input layer, one hidden layer, and one output layer. There are 10 nodes in the hidden layer, and 2 nodes in the output layer. A sigmoid function is used as the activation function (Bishop, 2006). This architecture can find a non-linear decision boundary that separates the surface-wave feature pixels from the others. The logistic regression architecture is similar to the NN method except there is no hidden layer. This indicates that the logistic regression cannot account for the highly non-linear decision boundaries.

The SVM has the same goal as the logistic regression and NN methods in classifying data, but SVM is sometimes preferred because it is simpler because there is no need to design the complicated network architecture. The non-linear SVM method can also be generalized to classify data which requires a non-linear separation boundary by using the kernel function (Bishop, 2006). SVM believes that the low-dimension non-separable dataset can be separated in a very high-dimension. The kernel function, such as Gaussian kernel, can mapping the low-dimension data into the infinite high-dimension space. The reason is that the Taylor expansion of Gaussian kernel is infinite high-order in theoretical (Figure 5). There are different types of kernel functions: linear kernel, polynomial kernel, and Gaussian kernel. In our work, we use the Gaussian kernel to achieve good linear separation in the higher dimension.

NUMERICAL TESTS

Synthetic data computed from the S-velocity model in Figure 2a is used to test the three machine learning methods for defining the surface-wave window in the $k - \omega$ domain. These methods are also tested on field data recorded over a known low-velocity zone.

Low-Velocity Layer Model Test

A finite-difference solution to the 2D the elastic wave equation is computed for sources and vertical-component receivers on the surface of the low-velocity layer (LVL) model shown in Figure 2a. The simulation data consisted of 30 shot gathers (Figure 2b) with a shot interval of 6 m . Each shot is recorded by 60 receivers on the free surface with a receiver spacing of 2 m , where the direct waves and surface waves interfere with one another at the near offsets. This is consistent with surface-wave propagation through a LVL type structure where there is multimode mixing in the dispersion curves (O'Neill and Matsuoka, 2005).

Figure 6 shows the decision boundaries predicted by the logistic regression, neural network (NN) and support vector machine (SVM) methods, respectively. These decision boundaries are able to separate most of the yellow and black symbols, which represent two different classes of the dataset. However, the decision boundary for the logistic regression method is a 2D plane and misclassifies some of the points. Fewer points are misclassified by the NN, and almost all of the points are correctly classified

by the SVM. By using the SVM or NN, we should be able to learn a non-linear decision boundary for separating the surface waves separation from other types of waves in the $k - \omega$ spectra.

Figure 7 shows the results of assigning surface-wave windows to the $k - \omega$ spectrum of a shot gather that was in the training set. Figure 7a is the original spectra, and Figures 7b-7d show the windows computed by logistic regression, NN, and SVM, respectively. The SVM window seems to be the most generous in allowing higher frequencies into the spectrum along with the same range of dip angles. The resulting CSGs after windowing and filtering are shown in Figure 8. Compared to the raw CSG, the three methods largely remove the body waves and scattered events (Figure 8b, 8c and 8d). However, the logistic regression and NN results lose some high-frequency information (Figure 9). In contrast, the SVM result in Figure 8d and the single trace spectrum in Figure 9 (red line) show a bandwidth more consistent with that of the raw shot gather.

Figure 10a depicts the dispersion-curve image of a raw shot gather. In order to test the accuracy of the filtered result, we use a layered-velocity model which has the same thickness and velocity as our low-velocity layer model in Figure 2a. Then, the layered-model dispersion curves are calculated by a Thomson-Haskell matrix method (Aki and Richards, 2002; Ke et al., 2011) for the fundamental mode (white dashed line in Figure 10). In Figures 10a-10c, the low-frequency part of the dispersion curve has a lower velocity than the analytical solution because of contamination with the near-field body-wave energy (McMechan and Yedlin, 1981; Roy and Jakkab, 2017).

In contrast, the SVM dispersion curves in Figure 10d are mostly in good agreement with the semi-analytical solution. In this example, the non-linear kernel SVM best separates the fundamental mode from the higher-order modes. In order to fairly compare the NN and SVM method, we use two hidden layer in NN method to separate the fundamental mode from the same shot gather shown in Figure 8a. The results of NN with two-hidden layer method are almost similar with the SVM method (Figure 11). However, we still choose to use SVM instead of the NN method. The reason is that the SVM is trying to find a separation line which has the fattest margin. This property is much more fascinating than the NN method because it could reduce the possibilities of over-fitting, especially when the data amount for the training set is small.

The dispersion curves can be automatically picked as following the peak energies in the SVM window. These curves are then inverted by the WD method (Li and Schuster, 2016) and the inverted S-velocity tomogram is shown in Figure 12. For the raw data, the near-offset part of the low-frequency dispersion curves is not accurately computed (Figure 10a). Consequently, WD cannot completely reconstruct the middle low-velocity layer, especially in the low-velocity layer area. However, the WD tomogram inverted from the SVM filtered data in Figure 12b is mostly consistent with the true model (Figure 2a).

Figure 13 shows the dispersion curves of raw shot gather and SVM processed data at $x=20\text{ m}$ and $x=100\text{ m}$. The SVM result can separate the high-order dispersion curve and reduce the artifacts from the scatter waves. We select the window offset

to be 20 m for each shot gather to compute the dispersion curves. Then, 30 shot gathers are used to generate 30 distinct dispersion curves, which can be concatenated together to get a pseudo-section of the phase-velocity distribution as a function of frequency in Figure 14 and the location of the source. This pseudo-section roughly approximates the lateral variations of the S-velocity structure. In addition, Figure 14 shows the phase-velocity contours computed from the recorded shot gathers (black dashed lines), which correlate well with the contours of the inverted phase-velocity values (white dashed lines).

In addition, after 49 iterations, the normalized data residual of the two datasets decreased to 0.2 and 0.4, respectively (Figure 15). It also demonstrates that the WD tomogram with SVM processed data has better inversion stability and higher accuracy than that of raw data WD tomogram.

Field Data Test

The SVM method is applied to field data recorded over a hidden low-velocity zone (LVZ) near the King Abdullah Economic City (KAEC) in Saudi Arabia. This low-velocity layer is interpreted to be related to the Qademah fault (Roobol and Kadi, 2008), where surface geology suggests a hidden fault. Li and Schuster (2016) used the WD S-velocity tomogram to infer a low-velocity zone in this area. The seismic data are recorded over the Qademah fault, where the shot spacing is 10 m, the vertical component receiver spacing is 5 m, and there are 120 traces per shot gather to give a total of 60 shot gathers. Figures 16a and 17a depict a typical shot gather and the

associated dispersion curves.

It can be seen that there are strong scattering noise and body wave arrivals in the raw data. Figures 16b and 17b show, respectively, a shot gather filtered by the SVM window and the associated dispersion curves. We can see that the filtered shot gather has a higher signal-to-noise ratio (SNR) than the raw data. In addition, the dispersion curves computed from the filtered data are more continuous and easier to pick than those of the raw data. The SVM tomogram from the filtered data is interpreted to provide a more reliable estimate of the S-velocity model than that obtained from the other filtered data.

SUMMARY

We tested three supervised ML methods in their ability to design a surface-wave window in the $k - \omega$ domain. Synthetic results suggest that the kernel SVM method provides the most effective window compared to the NN and logistic regression methods. Three features are used to distinguish the surface waves from other events in a local $k - \omega$ analysis window: coherency, dip angle, and peak frequency. Other features can be identified and used as well, but these three features work well for the examples in this paper.

To successfully use the SVM method to separate multi-mode dispersion curves, more practical details are discussed below:

- 1) A reliable training set should include the valid signal and artifacts in the surface

wave spectrum. This method is that manual labor is needed to classify points in the images of the training set. For our examples, no more than 0.5 hours were used to classify and train the ML methods. We need to build reliable classify points carefully.

2) The features for SVM or other machine learning algorithms greatly affect the training efficiency and classification accuracy. A good feature should be informative, discriminating and independent. We choose the coherency, dip angle, and peak frequency as our features because their value varies dramatically between the valid signals and artifacts very quickly.

The limitation of this method is that data recorded from different areas of the world might require the creation of new features that distinguish surface waves from other events. Also, we cannot give general formulas for the amount of data needed for effective training. Only experience will eventually determine how much data and training are needed.

ACKNOWLEDGEMENTS

The research reported in this publication was supported by the King Abdullah University of Science and Technology (KAUST) in Thuwal, Saudi Arabia. We are grateful to the sponsors of the Center for Subsurface Imaging and Modeling (CSIM) Consortium for their financial support. This work is supported by the Natural Science Foundation of China (41874134)

REFERENCES

- Aki, K., and P. G. Richards, 2002, Quantitative Seismology: *University Science Books*, **1**.
- Bishop, C., 2006, *Pattern Recognition and Machine Learning (Information Science and Statistics)*: Springer-Verlag New York, **1**.
- Bodeta, L., O. Abrahamb, and D. Clorennec, 2009, Near-offset effects on Rayleigh-wave dispersion measurements: Physical modeling: *Journal of Applied Geophysics*, **68**, 95–103.
- Dokter, E., D. Kohn, D. Wilken, D. D. Nil, and W. Rabbel, 2017, Full-waveform inversion of SH- and Love-wave data in near-surface prospecting: *Geophysical Prospecting*, **65**, 216–236.
- Ke, G., H. Dong, A. Kristensen, and M. Thompson, 2011, Modified Thomson-Haskell matrix methods for surface-wave dispersion-curve calculation and their accelerated root-searching schemes: *Bulletin of the Seismological Society of America*, **101**, 1692–1703.
- Li, J., Z. Feng, and G. Schuster, 2017, Wave-equation dispersion inversion: *Geophysical Journal International*, **208**, 1567–1578.
- Li, J., and S. Hanafy, 2016, Skeletonized inversion of surface waves: active source *vs* controlled noise comparison: *Interpretation*, **3**, 11–19.
- Li, J., F. C. Lin, Y. Ben-Zion, A. Alam, Z. Liu, and G. T. Schuster, 2019, Wave-equation dispersion inversion of surface waves recorded on irregular topography: *Geophysical Journal International*, **217**, 346–360.

- Li, J., and G. Schuster, 2016, Skeletonized wave equation of surface wave dispersion inversion: *SEG Technical Program Expanded Abstracts 2016*, 3630–3635.
- Lin, F.-C., M. P. Moschetti, and M. H. Ritzwoller, 2008, Surface wave tomography of the Western United States from ambient seismic noise: Rayleigh and Love wave phase velocity maps: *Geophysical Journal International*, **173**, 281–298.
- Luo, Y., J. Xia, R. Miller, Y. Xu, J. Liu, and Q. Liu, 2009, Rayleigh-wave mode separation by high-resolution linear Radon transform: *Geophysical Journal International*, **179**, 254–264.
- Maraschini, M., F. Ernst, S. Foti, and L. V. Socco, 2010, A new misfit function for multimodal inversion of surface waves: *Geophysics*, **75**, G31–G43.
- McMechan, G., and M. Yedlin, 1981, Analysis of dispersive waves by wave field transformation: *Geophysics*, **46**, 869–874.
- O'Neill, A., and T. Matsuoka, 2005, Dominant higher surface-wave modes and possible inversion pitfalls: *Journal of Environmental and Engineering Geophysics*, **10**, 185–201.
- Park, C. B., R. D. Miller, and J. Xia, 1998, Imaging dispersion curves of surface waves on multi-channel record: *SEG Technical Program Expanded Abstracts 1998*, 1377–1380.
- Perez Solano, C. A., D. Donno, and H. Chauris, 2014, Alternative waveform inversion for surface wave analysis in 2D media: *Geophysical Journal International*, **198**, 1359–1372.
- Rix, G., C. Lai, and S. Foti, 2001, Simultaneous measurement of surface wave dispersion and attenuation curves: *Geotech. Test. J.*, **24**, 350–358.

- Roobol, M., and K. Kadi, 2008, Cenozoic faulting in the Rabigh area, Central West Saudi Arabia (including the sites of King Abdullah Economic City and King Abdullah University for Science and Technology): *Saudi Geological Survey Technical Report SGS-TR-2008-6*, **1**, 1–17.
- Roy, N., and R. Jakkab, 2017, Near-field effects on site characterization using MASW technique: *Soil Dynamics and Earthquake Engineering*, **97**, 289–303.
- Schafer, M., L. Groos, T. Forbriger, and T. Bohlen, 2014, Line-source simulation for shallow-seismic data. part 2: Full-waveform inversion—a synthetic 2D case study: *Geophysical Journal International*, **198**, 1405–1418.
- Socco, L. V., S. Foti, and D. Boiero, 2010, Surface-wave analysis for building near-surface velocity models established approaches and new perspectives: *Geophysics*, **75**, 75A83–75A102.
- Xia, J., R. D. Miller, and C. B. Park, 1999, Estimation of near-surface shear-wave velocity by inversion of Rayleigh waves: *Geophysics*, 691–700.
- Yilmaz, Z., 2001, Seismic Data Processing: SEG Press, Tulsa, **2**.

Figure 1: The classical machine learning workflow consists of four steps: data pre-processing, feature extraction, model learning and model evaluation.

Figure 2: a) True S-velocity layer model, b) a shot gather and c) the frequency spectrum of raw shot gather.

Figure 3: Three feature images associated with the magnitude spectra of a shot gather: a) amplitude coherency, b) dip angle by slant stacking, and c) the frequency range.

Figure 4: Architecture of the neural network (NN) used in our tests.

Figure 5: Diagram of the higher dimensional representation associated with the SVM method.

Figure 6: Plots of decision boundaries computed by logistic regression, neural networks, and SVM.

Figure 7: a) The frequency spectrum of raw shot gather, b), c) and d) show the output mute-window delineated by the logistic regression, neural network and SVM windows, respectively.

Figure 8: a) Raw shot gather, b), c) and d) show the shot gathers filtered by the logistic regression, neural network and SVM windows, respectively.

Figure 9: Comparison of the frequency spectra of a single trace in the a) raw shot gather (black line) with those filtered by the logistic regression (green line), neural network (blue line) and SVM (red line) windows.

Figure 10: The dispersion curves of a) the raw shot gather and those filtered by the b) logistic regression, c) neural network, and d) kernel SVM windows.

Figure 11: a) The frequency spectrum of raw shot gather, b) show the output mute-window delineated by the multi-layer neural network (NN), c) show the shot gathers filtered by the multi-layer NN and d) The dispersion curves of multi-layer NN processed data.

Figure 12: WD tomograms computed from the traces obtained from the a) raw and b) SVM filtered data sets.

Figure 13: The dispersion curves of a) and b) are the raw shot gather at $x=20$ m and $x=100$ m; c) and d) are the SVM processed dispersion curves at $x=20$ m and $x=100$ m.

Figure 14: True (black dashed lines), initial (red solid lines) and inverted (white dashed lines) dispersion curves profiles as a function of frequency.

Figure 15: Plots of normalized a) normalized residual of raw data WD (black dashed line) and SVM processed WD (black line) against iteration number.

Figure 16: Surface-wave separation test. a) Raw shot gather, b) shot gather by the SVM window, c) the difference between the raw and SVM-filtered shot gathers.

Figure 17: The dispersion curves of the a) raw and b) SVM-filtered shot gathers.