# Automatic wave mode identification using machine learning

*Ohad Barak and Fantine Huot*

## ABSTRACT

In addition to reflection data, seismic recordings contain many different wave modes that are either unwanted or unneeded, degrading overall data quality. We use support vector machines (SVM), a type of supervised learning algorithm, for automatic wave mode classification. We decompose multicomponent translational and rotational seismic data from a field survey into polarization vectors, by applying continuous wavelet transforms (CWT) followed by singular value decomposition (SVD). We train an SVM classifier to distinguish surface waves from body waves from these polarization vectors, and show classification results on different portions of the field data. Our method does not rely on spatial continuity, and can therefore be applied to spatially aliased data.

## INTRODUCTION

In addition to reflection data, seismic surveys are cluttered with many other seismic responses that are either unwanted or unneeded, resulting in recorded data containing many different wave modes (Yilmaz, 2001). In the case of land acquisitions, high amplitude ground roll noise can obscure signal, degrading overall data quality.

Ground roll is the main type of coherent noise in land seismic surveys and is characterized by low frequencies and high amplitudes. Common processing techniques for attenuating ground roll include frequency filtering (Yilmaz, 2001), Radon transform (Liu and Marfurt, 2004), wavelet transforms (Deighan and Watts, 1997), and the curvelet transform (Yarham and Herrmann, 2008). However, these techniques can be limited when ground roll is spatially aliased and has non-linear moveout, and it is notoriously difficult to model ground roll with sufficient generality. Therefore, ground noise removal remains a tedious task.

Barak and Ronen (2016) demonstrate how to use combined translational and rotational data to identify and separate particular wave modes. They apply continuous wavelet transforms (CWT) followed by singular value decomposition (SVD) to identify the polarization signature of the particular wave modes associated with ground roll. A filter is then applied to attenuate sections of the data with similar polarization signatures.

Nonetheless, this methodology requires selective manual picking of events on the data section. Moreover, it is based on the underlying assumption of stationarity of wave modes along offset, although there is no guarantee that the polarization signature of a certain wave mode at a particular offset should remain similar at other offsets. Therefore, we extend this process by incorporating pattern recognition algorithms as described by Huot and Clapp (2016). We use support vector machines (SVM), a set of supervised learning methods used for classification, to identify the waves modes associated with ground roll, at all times, offsets and azimuths.

We train an SVM classifier to identify specific wave modes from the polarization vectors computed on field data, and then test our classification on different portions of the data where the wave modes are clearly distinguishable to assess the quality of our classification.

# PREPARING DATA FEATURES FOR MACHINE LEARNING

## Continuous Wavelet Transform (CWT)

The continuous wavelet transform (CWT) is defined as:

$$C_i\left(a,b;g_i(t),\psi(t)\right)=\int_{-\infty}^{\infty}g_i(t)\frac{1}{\sqrt{a}}\psi^*\left(\frac{t-b}{a}\right)dt, \tag{1}$$

where $g_i(t)$ is the input signal of the $i^{th}$ data component, $\psi(t)$ is a mother wavelet, $\psi^*$ is a daughter wavelet, which is the complex conjugate of the mother wavelet stretched by scale $a$ and time-shifted by $b$. It is common to use the Morlet wavelet as a mother wavelet, and we do so in this paper. For brevity, we will use $C_i\left(a,b\right):=C_i\left(a,b;g_i(t),\psi(t)\right)$

The continuous wavelet transform effectively shows how correlated our time-series is with a particular daughter wavelet. Since the correlation is done in running time windows (shifted by $b$), the transform retains the temporal sense of the data and yet decomposes it to wavelet scales, which are in essence similar to frequency. We use this time-frequency decomposition to identify wave modes of particular frequencies that appear at particular times in multicomponent data.

## Singular Value Decomposition (SVD)

We apply singular value decomposition to a time slice of a single continuous-wavelet transformed multicomponent trace $C_i(a,b_k)$, where $b_k$ is the time index of the slice, $a$ ranges through the wavelet scales and $i$ represents the data component. Therefore, we have an $N_a\times N_c$ data matrix $\mathbf{D}$ where the rows are the wavelet scales and the columns

are the data components. SVD is a method of finding the waveform $\mathbf{u}$, magnitude $\sigma$, and polarization $\mathbf{v}$ of the signal that is present in the data $\mathbf{D}$. The SVD of the data $\mathbf{D}$ is given by

$$\mathbf{D} = \mathbf{U}\Sigma\mathbf{V}^T, \tag{2}$$

where $\mathbf{D}$ is the product of an $N_a \times N_c$ matrix $\mathbf{U}$, an $N_c \times N_c$ diagonal matrix $\Sigma$, and the transpose of an $N_c \times N_c$ matrix $\mathbf{V}$. The unit left and right singular vectors $\mathbf{u}_i$ and $\mathbf{v}_i$ are the column-vectors of $\mathbf{U}$ and $\mathbf{V}$. The singular values $\sigma_i$ are the diagonal elements of $\Sigma$. They are ordered such that $|\sigma_1|$ is the greatest and $|\sigma_{N_c}|$ the smallest. The left and right singular vectors are mutually orthogonal, such that $\mathbf{U}^T\mathbf{U} = \mathbf{I}$ and $\mathbf{V}\mathbf{V}^T = \mathbf{I}$.

The right singular vectors $\mathbf{v}_i$ display the polarization of the data within the particular frequency window along the data axes. We transpose and multiply the matrix $\mathbf{V}$ by the singular value matrix $\Sigma$, to obtain the scaled polarization vectors:

$$\mathbf{S} = \mathbf{\Sigma}\mathbf{V}^T. \tag{3}$$

## Support Vector Machines (SVM)

The polarization vectors obtained after applying CWT and SVD are used as input for classifying the different wave modes. From here, we set up the classification problem according to the methodology described in Huot and Clapp (2016).

Support vector machines (SVM) are a supervised learning algorithm that have been shown to perform well in a variety of settings, and are often considered one of the best "out of the box" classifiers. To train the classifier, we build a training dataset of $N$ sample pairs, $(x_1, y_1), (x_2, y_2), \cdots, (x_N, y_N)$, where the $x_i \in \mathbb{R}_p$ represent the features, which are the $p$ values of the polarization vectors, and the $y_i \in \{-1, 1\}$ are the binary labels indicating to which wave mode the sample belongs, namely, whether it is a surface wave or a body wave.

The support vector classifier determines the optimal hyperplane separating the two classes in the features space, given a certain amount of slack $\xi = (\xi_1, \xi_2, \cdots, \xi_N)$ to account for the fact that the samples are not perfectly separable. If we define a hyperplane by the following:

$$\{x : f(x) = x^T\beta + \beta_0 = 0\}, \tag{4}$$

where $\beta$ is a unit vector and $\beta_0$ a constant, the classification rule induced by $f(x)$ can be expressed as:

$$G(x) = \text{sign}[x^T\beta + \beta_0]. \tag{5}$$

With these notations, the support vector classifier for the nonseparable case is commonly expressed as:

$$\min_{\beta,\beta_0} \|\beta\| \quad \text{subject to} \begin{cases} y_i({x_i}^T\beta + \beta_0) \geq 1 - \xi_i\,, \ \forall i \\ \xi_i \geq 0, \ \forall i \quad \text{and} \quad \sum \xi_i \leq K. \end{cases} \tag{6}$$

By bounding the sum $\sum \xi_i$, we bound the total proportional amount by which predictions fall on the wrong side of their margin. Misclassifications occur when $\xi_i > 1$, so bounding $\sum \xi_i$ at a value $K$ bounds the total number of training misclassifications at $K$.

However, the support vector classifier subsequently described is limited to finding linear boundaries in the input feature space. To achieve better training-class separation, we introduce the idea of enlarging the feature space using basis expansions. Linear boundaries in the enlarged space translate to nonlinear boundaries in the original space, thereby making the problem more flexible. Once the basis functions $h_m(x)$, $m = 1, \cdots, M$ are selected, the procedure is the same as before. We fit the support vector classifier using input features $h(x_i) = (h_1(x_i), h_2(x_i), \cdots, h_M(x_i))$, $i = 1, \cdots, N$, and produce the (nonlinear) function $\hat{f}(x) = h(x)^T\hat{\beta} + \hat{\beta}_0$. The classifier is $\hat{G}(x) = \text{sign}[\hat{f}(x)]$ as before.

In practice, we need not specify the transformation $h(x)$ at all, but require only knowledge of the following kernel function:

$$K(x, x') = \langle h(x), h(x') \rangle, \tag{7}$$

that computes the inner products in the transformed space. For particular choices of $h$, these inner products can be computed very efficiently.

Three popular choices for $K$ in the SVM literature are shown in the following:

$$d^{\text{th}}\text{—degree polynomial:} \quad K(x, x') = (1 + \langle x, x' \rangle)^d, \tag{8}$$

$$\text{Radial basis:} \quad K(x, x') = \exp(-\gamma \|x - x'\|^2), \tag{9}$$

$$\text{Neural network:} \quad K(x, x') = \tanh(\kappa_1 \langle x, x' \rangle + \kappa_2). \tag{10}$$

A more complete overview of the implementation of support vector machines is described in Huot and Clapp (2016).

## SVM TRAINING AND TESTING WITH KETTLEMAN DATA

The 2D Kettleman survey comprised multiple types of sources and multiple types of receivers, both on the surface and at depth. The shot line length was 1.6 km long. The seismic sources used were a vibroseis truck, an accelerated weight-drop, and buried dynamite charges. At one end of the shot line there were five 3 component (3C) linear

accelerometers, which were closely spaced at a 2.1 m interval inline. Additionally, near the center of the shot line there were two adjacent 3C geophones buried at 1 m depth, spaced at a 2 m interval inline.

Although the survey also had 3C rotation sensors, the signal to noise ratio of these sensors was very low for offsets greater than 300 m. Therefore, we derived the pitch rotational component by differencing adjacent vertical geophones and accelerometers at each receiver station. Derivation of rotational components from translational data is described in Barak and Ronen (2016), Barak et al. (2015), Muyzert et al. (2012) and Brokesova and Malek (2015).

Since the sensors were arrayed only in the inline direction, we could not derive rotational data for the roll and yaw components. Consequently, our analysis here includes only the vertical and radial translational components and the pitch rotational component. Figure 1(a) is the vertical component receiver gather at station 335 shot with the vibroseis source, Figure 1(b) is the radial component, and Figure 1(c) is the pitch component. AGC has been applied for display.

The P body-wave reflections are visible at earlier times and longer offsets, but most of the section is dominated by various modes of surface waves, which is typical in land data. There is a very slow Rayleigh wave mode propagating at around 250 m/s, and a faster mode propagating at 400 m/s. Henceforth we will refer to these wave modes as "slow" and "fast" ground roll. There is yet another, faster mode, propagating at 600 m/s - 700 m/s, which is not part of the analysis here.

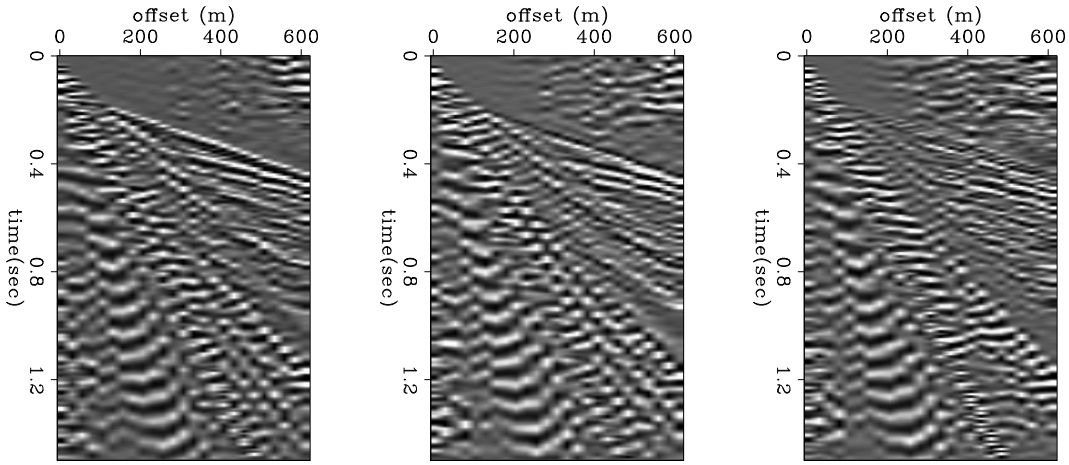



Figure 1: Receiver gathers at station 335 for the vibroseis source. a) Vertical accelerometer component. b) Radial accelerometer component. c) Pitch component derived from differencing two adjacent inline vertical accelerometers. AGC has been applied for display. Note the various types of surface waves present. [**ER**]

Standard seismic processing requires that the surface waves ("ground-roll") be removed, as the data we are commonly interested in are the P body-wave reflections

that are obscured by the surface waves. However, as a first step, the unwanted energy must be identified.

## SVM training data for classification of surface waves

In order to separate the surface waves from the body waves, we train an SVM classifier on the data from receiver station 335. We decompose the training data into scaled continuous-wavelet polarization vectors by applying CWT followed by SVD, as in equations 1 and 3. Since we are using three data components, for each offset and each time sample we obtain three polarization vectors. We use only the first polarization vector as our input feature vector for the SVM classification, since most of the energy is contained within the first vector.

The feature vectors are illustrated in Figure 2(a). We can clearly distinguish high amplitude parcels corresponding to the specific seismic signature of the different wave modes. Figure 2(b) represents the same feature vectors plotted by color code, red for slow ground roll and blue for all other wave modes. We can see that the wave modes have distinct behaviors in the feature space.

In order to train the classifier, we label the training data manually by selecting particular wave modes in the data from receiver station 335. The selection was done using a linear mute function that windowed only the ground-roll energy. Figures 3(a), 3(b) and 3(c) illustrate which regions were labeled as slow ground roll, fast ground roll and total ground roll. For each of these selected wave modes, we trained an SVM classifier using the aforementioned polarization feature vectors as input. The ground roll is labeled as "Class 1", while all the other wave modes are labeled as "Class 0".

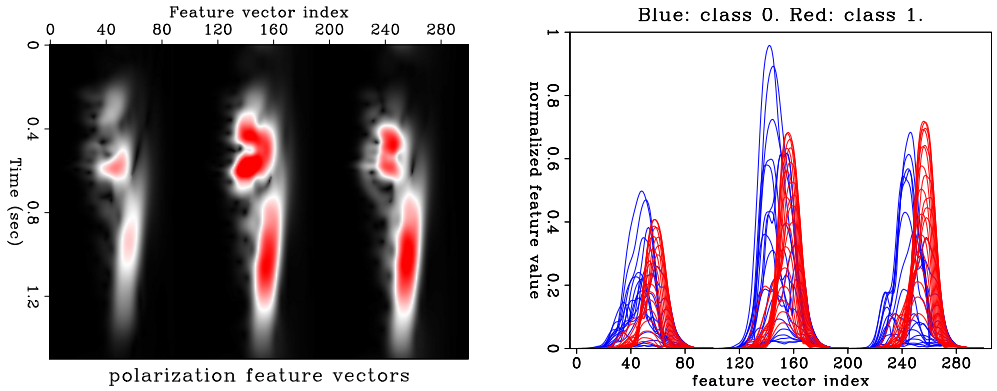
We then test the SVM's fitting capability. The classification prediction results obtained on the training data are illustrated in Figures 3(d), 3(e) and 3(f). They are highly similar to our hand-picked labels, indicating that our classifiers fit the training data well, with a low training error.

## SVM prediction of surface waves on test data

### Testing on other receiver stations

We then tested our previously trained classifiers on data from other receiver stations for the same vibroseis source type.

Stations 336 and 337 are adjacent to stations 335 (the training data station) near the end of the shot line, and therefore the data at these receiver stations should be similar to the training data. This enables us to estimate how well the classifier predicts the ground roll wave mode on these test data. Figure 4(a) is the vertical




(a) (b)

Figure 2: The training data at receiver station 335 for the SVM. a) Feature vectors for each time sample at offset 180m. b) The same feature vectors color coded by their respective classification, 'Class 1' being the slow ground roll mode and 'Class 0' representing everything else. [**ER**]

component at station 336. Figures 4(b), 4(c) and 4(d) are the predicted slow ground roll, fast ground roll and combined slow+fast ground roll wave modes at this station, respectively. Recall that this prediction is done using the classification generated by labeling the data at station 335. Observe that the classifier has managed to identify the ground roll in these test data relatively well, and it does not misclassify the body wave energy as ground roll.

The prediction seems to be better when both ground roll modes are combined, as show in Figure 4(d). It seems that it is more difficult for the classifier to differentiate between the two types of ground roll (slow vs fast) than between the ground roll and body waves. The same conclusions can be drawn for the prediction results for station 337, shown in Figures 4(f), 4(g) and 4(h).

Station 191 is about 1 km away from station 335. At station 191 there were geophones installed, rather than accelerometers as at stations 335, 336 and 337. Furthermore, the geophones were buried at 1 meter depth, whereas at stations 335, 336 and 337 the accelerometers were on the surface.

Figures 4(j), 4(k) and 4(l) are the predicted slow ground roll, fast ground roll and combined slow+fast ground roll wave modes at station 191, respectively. The classifier seems to have more difficulty differentiating between the slow and the fast ground roll modes at this station. Additionally, it has misclassified a small portion of the body wave energy as ground roll, as can be seen on the left Figure 4(l) at $t = 0.4$ s. However, by and large, the classification does enable the identification of ground roll energy, despite the differences between the data at station 191 vs station 335, on which we trained our SVM classifier.

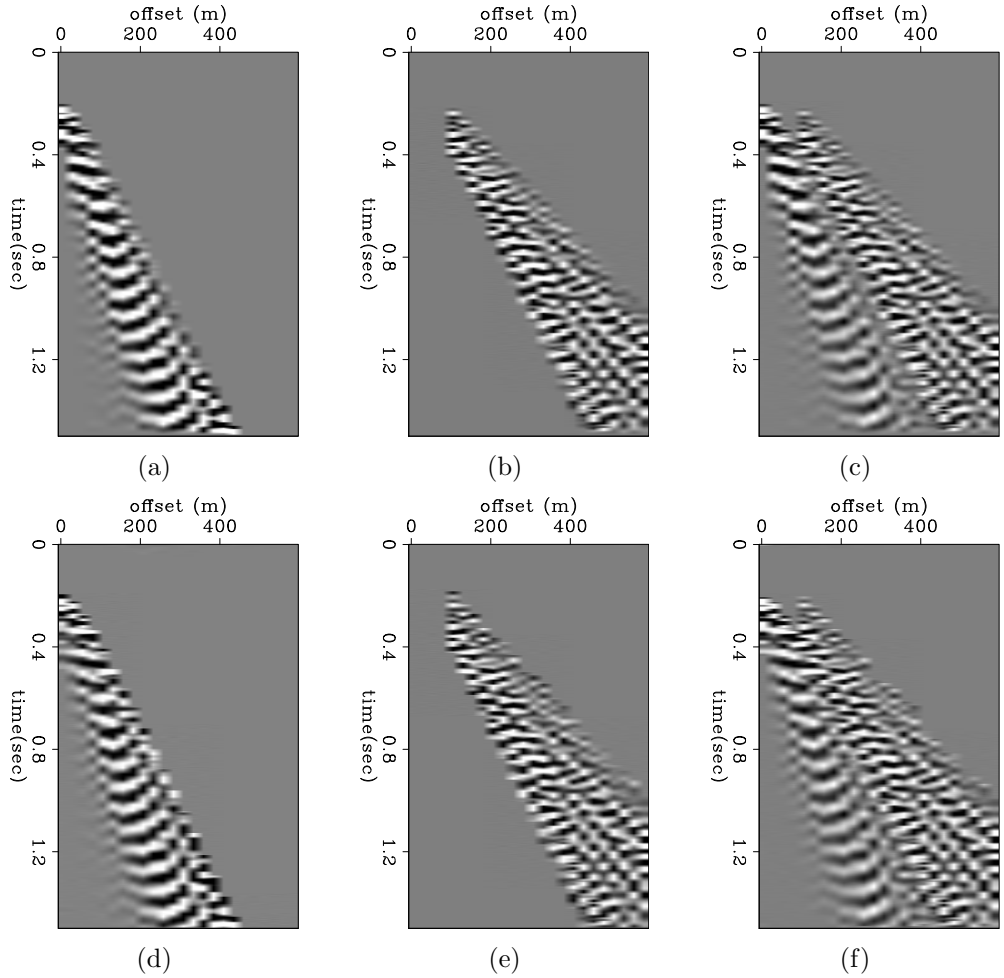



Figure 3: SVM training data and fitting results from station 335, for the vibroseis source. a) Slow ground roll mode training data. b) Fast ground roll mode training data. c) Combined (slow+fast) ground roll modes training data. d) SVM fitting of slow ground roll mode on training data. e) SVM fitting of fast ground roll mode on training data. f) SVM fitting of combined (slow+fast) ground roll modes on training data. [**ER**]

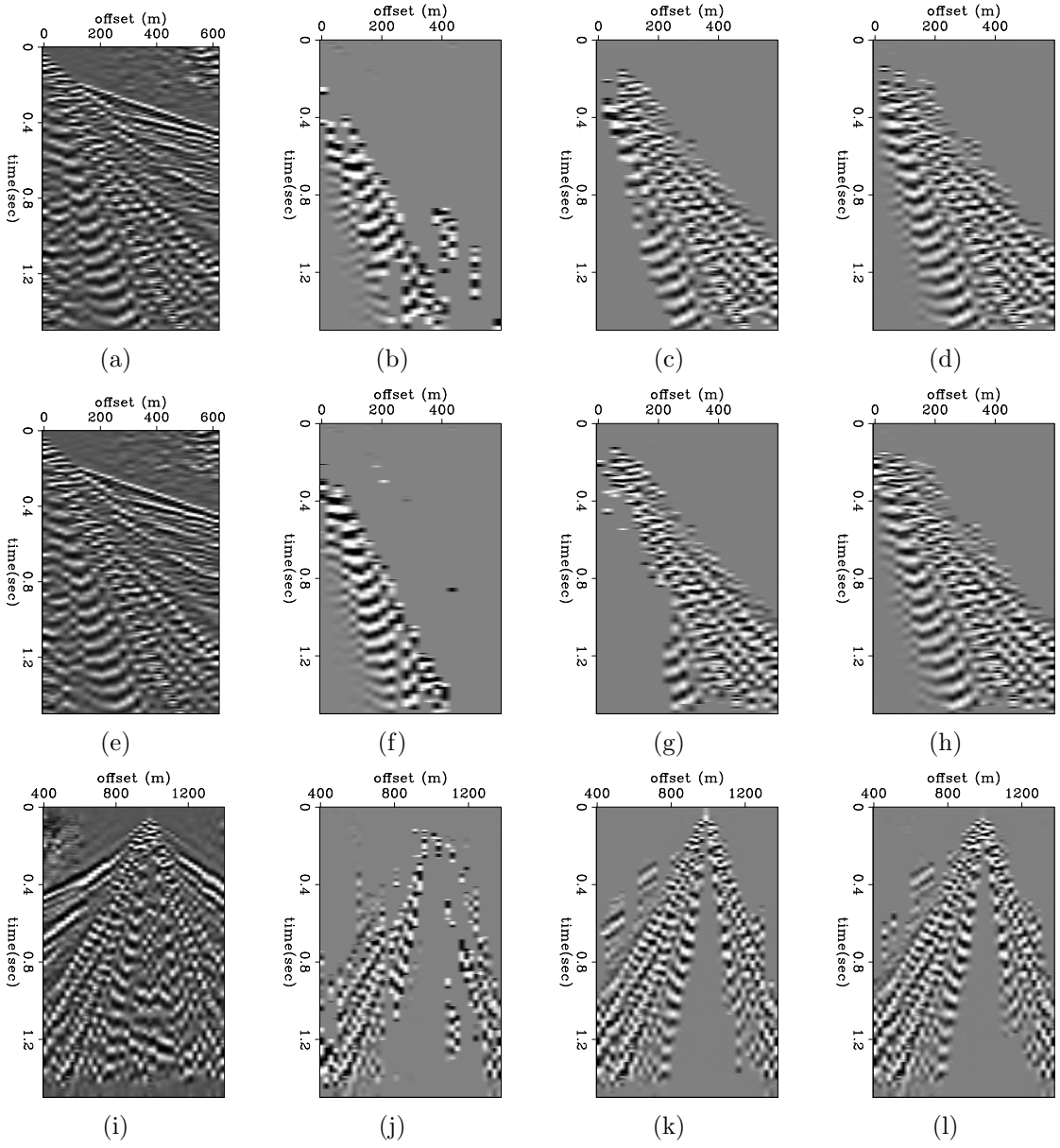



Figure 4: SVM prediction of surface wave modes on test data, for the vibroseis source. a,e,i): Vertical component at receiver stations 336, 337 and 191. b,f,j): prediction of slow ground roll mode for stations 336, 337 and 191. c,g,k): prediction of fast ground roll mode for stations 336, 337 and 191. d,h,l): prediction of combined (slow+fast) ground roll modes for stations 336, 337 and 191. [**ER**]

### Testing on other source types

We then tested the classifier on the same station that provided the training data (335), but for alternate seismic source types.

Figures 5(a) and 5(d) are again the receiver gather of the vertical component at station 335, and portion of combined (slow+fast) ground roll labeled as "Class" 1 for the training data.

Figure 5(b) is the receiver gather of the vertical component at station 335 for an accelerated weight-drop source. Note that the sampling for this source type was much denser than the vibroseis source. The ground roll wave modes visible in this section are similar to those in Figure 5(a), however they are not aliased. Figure 5(e) shows the SVM's prediction of the combined (slow+fast) ground roll modes for the accelerated weight drop data. Despite the fact that the source type is different, the classification is able to correctly predict the ground roll in this section. Figure 5(e) is remarkably similar to Figure 5(d), albeit with a better sampling.

Both the vibroseis and the accelerated weight-drop source were on the surface. Figure 5(c) is the receiver gather at station 335, but for a dynamite source buried at 25 m depth. The ground roll energy has a different appearance in this section. Figure 5(f) is the SVM prediction of the slow+fast ground roll wave mode for the dynamite source. Note that the energy classified as ground roll does indeed seem to have the same linear moveout as the ground roll in the vibroseis section. The classifier even picks up on a portion of a faster surface wave mode that does not appear in the vibroseis data.

In both test cases shown in Figures 5(e) and 5(f), there is no misclassification of the body wave energy as ground roll.

## DISCUSSION

Standard seismic processing may utilize some known attributes of ground roll to identify it. For example, ground roll has a much slower moveout than body waves in common shot or receiver gathers. Additionally, in multicomponent data the elliptical polarization associated with Rayleigh waves may be employed to identify ground roll.

The approach of machine learning is very different however. Standard methods use an a priori, analytical model of ground roll, such as moveout or elliptical polarization. A machine learning approach does away with any predetermined model, and substitutes a model learned from the data themselves. The advantage of such an approach is that it may work in cases where a useful, representative model is beyond our capability to represent analytically.

The case of surface waves (ground roll) is of particular interest in this respect,

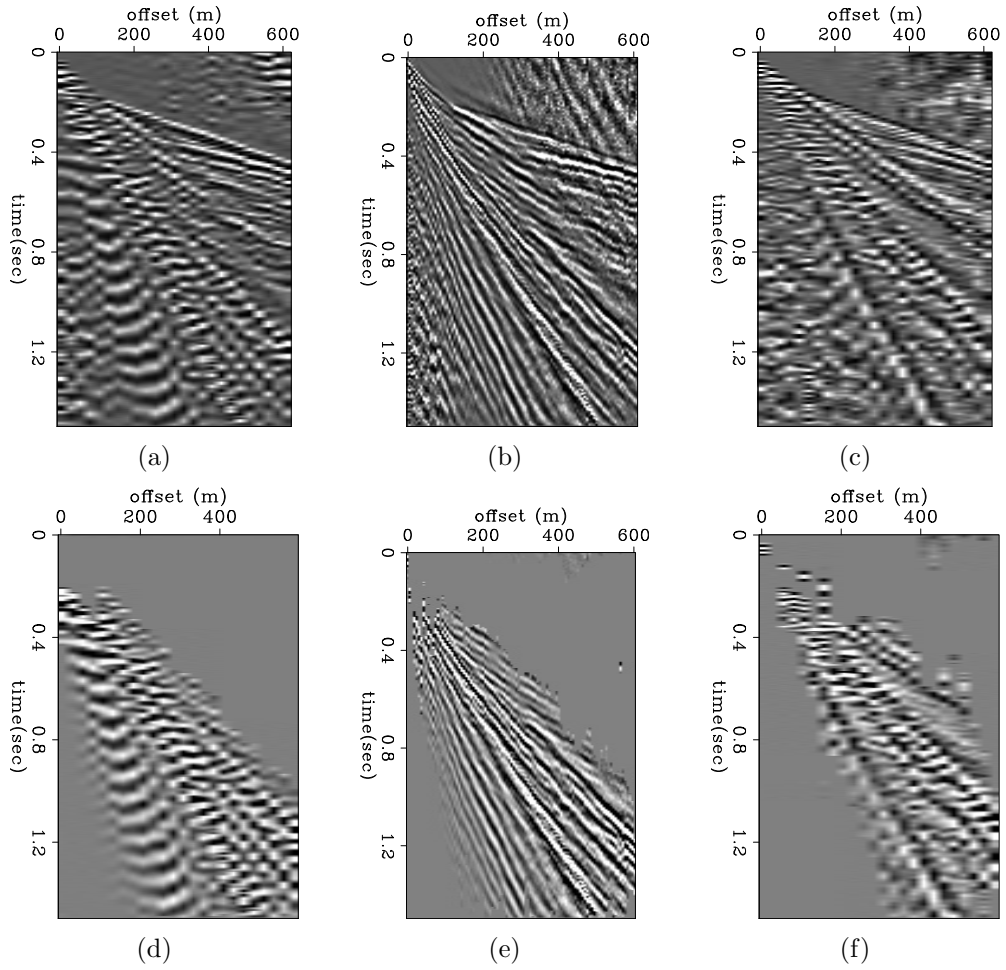



Figure 5: SVM prediction of surface wave modes at station 335, for different seismic source types. a) Vertical component of vibroseis source. b) Vertical component of accelerated weight-drop source. c) Vertical component of dynamite source. d) The portion of the combined (slow+fast) ground roll mode labeled as class 1 from the vibroseis source data. e) prediction of combined (slow+fast) ground roll mode for the accelerated weight-drop source. f) prediction of combined (slow+fast) ground roll mode for the dynamite source. [**ER**]

since the near surface of the Earth is commonly very complex, and in general it is difficult to accurately model wave propagation in the near surface. For example, Rayleigh waves do not always have elliptical polarization, nor are their moveouts necessarily predictable. Surface waves may also be generated by body waves incident on near surface scatterers, in addition to being radiated directly away from the source position. Therefore, the prospect of enabling an algorithm to learn from the data what ground roll may look like in multiple scenarios is compelling.

We have used the continuous-wavelet polarization vectors of multicomponent seismic data to train an SVM algorithm to identify ground roll. The classification results using the test data indicate that the SVM is indeed able to identify ground roll based on this (rather minimal) training.

There remain several open questions with respect to practical application of this machine learning algorithm:

1. In order to have a good representation of the various ways each wave mode may be polarized at multiple times/offsets/azimuths (given a set of data components), massive amounts of seismic data are required. How can we classify massive quantities of data?
2. Can an SVM trained on a particular dataset be used for classification of another dataset?
3. Are more components necessarily better for the SVM's classification of wave modes? If not, which minimal set of components would be the most useful?

The Kettleman dataset is instrumental in showing the possibility of identifying wave modes in land data using machine learning, since it has multiple components, both translational and rotational, which enable identification by polarization. However, the Kettleman dataset is very small, and therefore we cannot answer the questions stated above with it. For that, we would require a very large, 3D multicomponent land or OBS dataset.

## ACKNOWLEDGEMENTS

We thank Chevron for permission to publish the Kettleman dataset.

## REFERENCES

Barak, O., K. Key, S. Constable, P. Milligan, and S. Ronen, 2015, Acquiring rotation data on the ocean bottom without rotation sensors: SEP-Report, **158**, 311–322.

Barak, O. and S. Ronen, 2016, Wave-mode separation in the complex wavelet domain using combined translational and rotational data : SEP-Report, **163**, 1–19.

Brokesova, J. and J. Malek, 2015, Six-degree-of-freedom near-source seismic motions II: Examples of real seismogram analysis and S-wave velocity retrieval: Journal of Seismology, **19**, 511–539.

Deighan, A. J. and D. R. Watts, 1997, Ground-roll suppression using the wavelet transform: Geophysics, **62**, 1896–1903.

Huot, F. and R. Clapp, 2016, Detecting karst caverns by pattern recognition : SEP-Report, **163**, 297–307.

Liu, J. and K. J. Marfurt, 2004, 3D high resolution Radon transforms applied to ground roll suppression in orthogonal seismic surveys: SEG Technical Program Expanded Abstracts, 2144–2147.

Muyzert, E., A. Kashubin, E. Kragh, and P. Edme, 2012, Land seismic data acquisition using rotation sensors: 74th Conference and Exhibition, EAGE, Extended Abstracts.

Yarham, C. and F. J. Herrmann, 2008, Bayesian ground-roll separation by curvelet-domain sparsity promotion: SEG Technical Program Expanded Abstracts, 2576–2580.

Yilmaz, O., 2001, Seismic data analysis: Society of Exploration Geophysicists, Tulsa, **1**.