First break picking with machine learning

Bernard K. Law and Daniel O. Trad

ABSTRACT

First break (FB) picking is a laborious task for land data processing. In this report, we experiment with two Machine learning approaches 1) unsupervised automated editing of outlying picks by clustering. 2) supervised deep learning by training the networks with manually edited FB and classifying the first arrival energy waveforms as pre-FB and post-FB. The first approach is easier to apply but more limited. The second approach requires a catalogue of images and their first break picking for training. With enough training samples, the deep neural works will be able to classify the first arrival energy waveforms of new datasets as pre-FB and post-FB as accurately as the trained technicians.

INTRODUCTION

One of the most laborious and problematic tasks in refraction tomography is the first arrival traveltime or first break (FB) picking. Many automated FB picking methods determine the arrival time by the difference in amplitude, phase, or frequency characteristics between the data before and after the FB and are often done on a trace-by-trace basis. Spatial correlation between adjacent traces is only used for subsequent editing of mis-picks. The final step in FB picking is to confirm or manually modify the FB picks by trained technicians. With experiences from a large number of datasets with different topography and near-surface geological setting, experienced technicians can recognize the relationship between the FB and the complex waveform of the first arrival energy and various interfering noises. With increasing data density, this has become a very time-consuming and expensive process.

Machine learning is a fast-developing science that teaches computers to learn from data and human experiences. There are two potential applications of machine learning in automatic FB picking. One application of machine learning is automated editing of outlying picks by clustering (Tan et al, 2005). Another more advanced application is deep learning by training the networks with manually edited FB and classifying the first arrival energy waveforms as pre-FB and post-FB (Hu et al., 2019, Fernhout et al., 2020). With a catalogue of images of trained models, the deep neural works will be able to classify the first arrival energy waveforms of new datasets as pre-FB and post-FB as accurately as the trained technicians. In this report, we will review some of the automatic FB picking methods, clustering applications and one deep-learning application.

Automated first arrival picking

During a seismic experiment, ground motions are recorded after a seismic source is activated. A seismic record contains the refraction and reflection seismic signals caused by the seismic source, as well as surface-related seismic noises (Figure 1) caused by ambient noise, human and animal activities. First arrival energy is characterized by the relatively weak surface noises, followed by the stronger refraction seismic signal. We will review

and compare two automatic first arrival picking methods that use these amplitude characteristics. The signal to noise ratio (S2N) method defines the sum of peaks, $P(1:k-1)$, of a potential first arrival time pick k, as noise, and the peak at k, as the signal, and $S2N$ at k as:

$$S2N(k) = \frac{P(k)}{Sum(P(1:k-1))} \quad 7\text{-}1$$

Another method that uses the characteristics of the amplitude levels before and after the first arrival is the Akaike's information criterion (AIC) (Akaike, 1973; St-Onge, 2011). AIC is defined as:

$$AIC(k) = k * \log\left(var\big(y(1:k)\big)\right) + (nsamp - k - 1) * (\log\left(var(y(k+1:nsamp)\right)), \quad 7\text{-}2$$

where $var(y(i:j))$ is the variance for the time series y from sample i to k.

$$var\big(y(i:j)\big) = \frac{sum\left(v\left(y(i:j) - \overline{y(i:j)}\right)^2\right)}{j-i}_{ii=(i:j} \quad 7\text{-}3$$

When k is less than the first arrival time, $var\big(y(1:k)\big)$ is small, and $\left(var\big(y(k+1:nsamp)\big)\right)$ is large. Similarly, when k is greater than the first arrival time, $\left(var\big(y(1:k)\big)\right)$ becomes larger, and $var(k+1:nsamp)$ becomes smaller. When k is at the first arrival time, $AIC(k)$ is at its minimum.

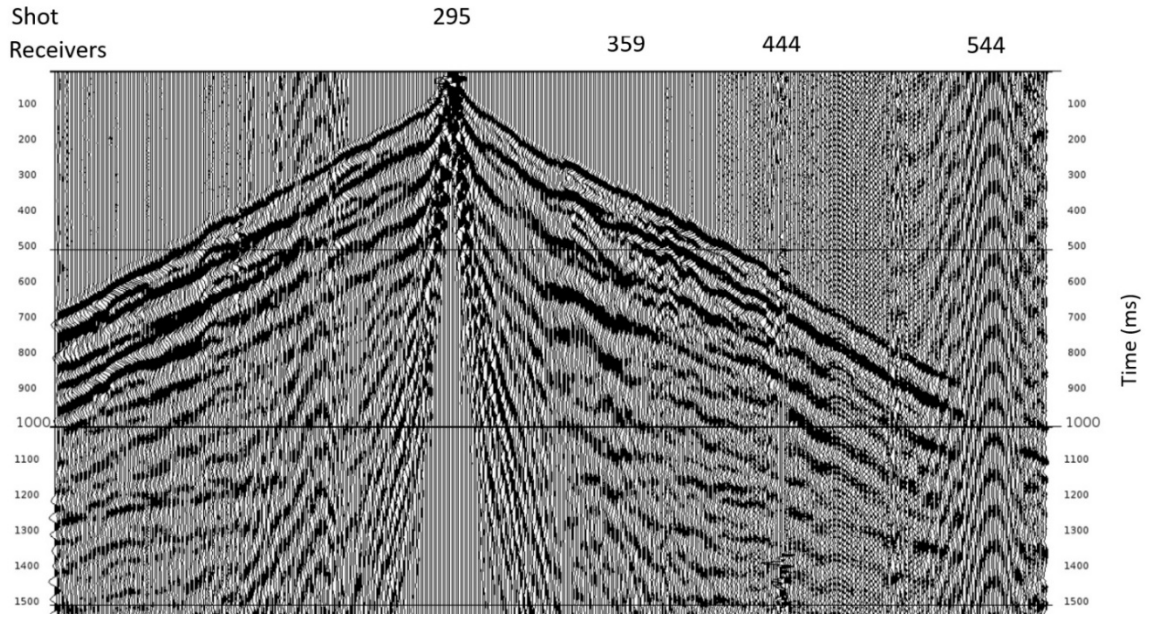


Figure 1. Seismic record with varying surface noise conditions. Receivers 359, 444 and 544 are identified with an increased noise level.

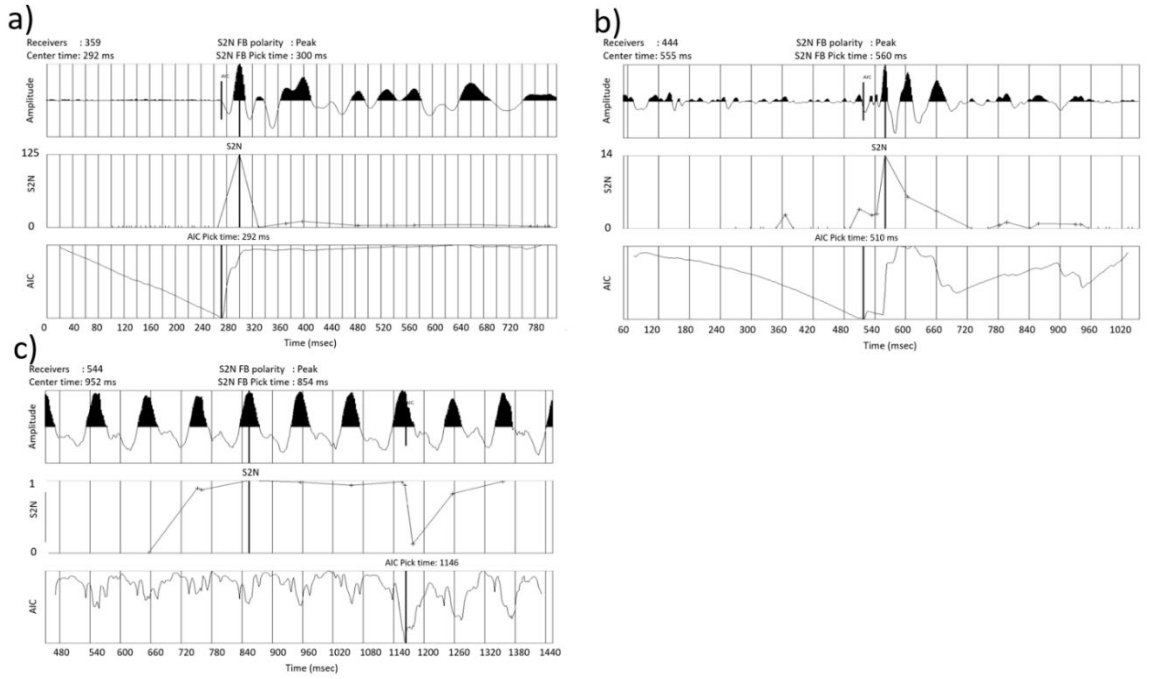


Figure 2. Results of S2N and AIC methods with increasing noise level (a to c). Top panel: first arrival energy. Middle panel: S2N analysis. Bottom panel: AIC analysis.

Figure 2a shows the first arrival energy with weak surface noise, followed by S2N and AIC analysis. Using positive FB polarity. The first arrival pick is chosen at a peak with the maximum $\boldsymbol{S2N(k)}$. The polarity of the S2N analysis method is peak; therefore, it picks a peak that meets the S2N criteria and arrives later than the zero-crossing picked by the AIC method. With a moderate increase in noise level (Figure 2b) both S2N and AIC methods are able to detect the consistent FB picks. However, with significantly stronger surface noises that overwhelm the first arrival energy, both S2N and AIC methods fail.

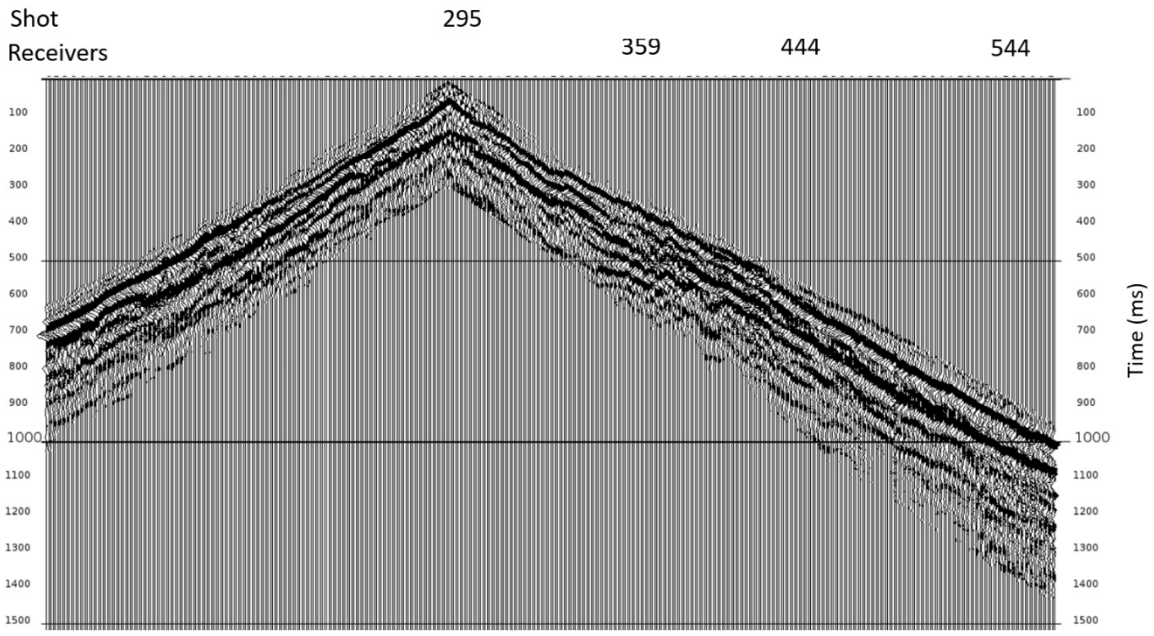


Figure 7. 3. Seismic shot 295 high-resolution linear Radon transform in windowed first arrival energy in receiver gather domain.

We sort the seismic records into common receiver gather domain and apply high-resolution Sparse Linear-Radon transform to remove the surface noise from windowed first arrival energy. Figure 3 shows the shot record 295 after being sorted back to the shot domain. This shows that high-resolution Sparse Linear-Radon transform in common receiver gather domain is an effective algorithm to remove surface noises that appear coherent within the shot gathers. Figure 4 shows both S2N and AIC methods can effectively determine the FB picks after the removal of the strong surface noise. However, this approach also creates some pre-first arrival artifacts that can negatively affect the FB picking accuracy. In general, AIC is superior to the S2N method; however, additional preconditioning processes may still be required to remove strong surface noises. Moreover, carefully editing of automatically picked FBs is often required by trained technicians with experience in identifying FB through complex first arrival waveforms and surface noises.

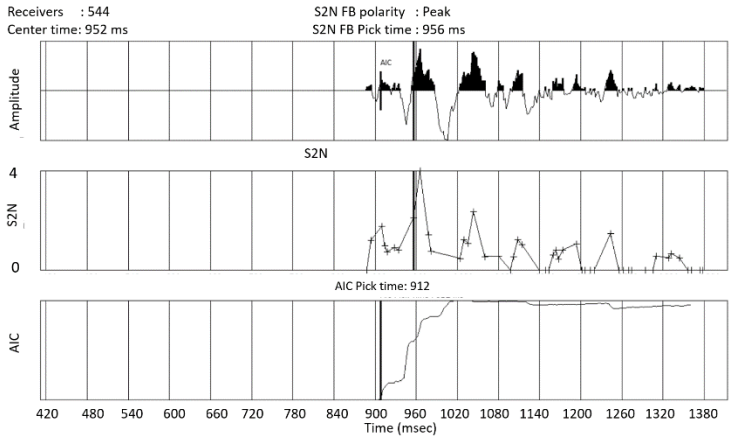


Figure 4. Results of S2N and AIC methods at shot 295 and receiver 544 after high-resolution linear Radon transform in windowed first arrival energy in common receiver gather domain.

Application of machine learning in automatic first arrival picking

Unlike the two trace-by-trace automated first arrival picking algorithms, machine learning algorithms explore the spatial relationship between data points. When technicians perform quality control of the computer-picked FBs, they reject FB picks that vary rapidly with respect to the neighbouring picks. Clustering is a machine learning technique that groups the data points according to their attributes (Tan et al, 2005). It has the potential of automating the human efforts in rejecting and modifying the FB picks that vary too rapidly (Smith et al, 2017). Three commonly used clustering algorithms are K-Means, Gaussian Mixture Models (GMM) and Density-Based Spatial Clustering of Applications with Noise (DBSCAN) (bin Waheed, 2019).

K-Means

K-Means assumes the data points distribution to be Euclidean or circular. The Euclidean distance square is defined as:

$$\Delta^2 = (x-\mu)^T(x-\mu), \qquad 7\text{-}4$$

where x is the dimensional matrix, μ is the mean matrix.

The following steps outline the K-Means algorithm:

1. Starts with randomly placing the centroids of N clusters.
2. For each data point, calculates the Euclidean distance between the data point and each of the centroids.
3. To find the clusters, assign the data point to the nearest centroid.
4. Recompute the coordinates of the centroids using the mean coordinates of the clusters.
5. Repeat steps 2, 3, and 4 until convergence.

Gaussian-Mixture-Models

GMM assumes the data points distribution to be Gaussian and is less restrictive than K-means.
The Gaussian distribution function is defined as:

$$f(x) = \frac{1}{2\pi |\sum|^{\frac{1}{2}}} \exp[-\frac{1}{2}(x-\mu)^T \sum^{-1}(x-\mu)] \quad 7\text{-}5$$

where x is the dimensional matrix, μ is the mean matrix and $\sum$ is the covariance matrix. For a two dimensions case and μ equals 0:

$$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \quad 7\text{-}6$$

$$\mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad 7\text{-}7$$

$$\sum = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix}, \quad 7\text{-}8$$

$$\sum^{-1} = \frac{1}{\sigma_1^2\sigma_2^2 - \sigma_{12}^2} \begin{bmatrix} \sigma_2^2 & -\sigma_{12} \\ -\sigma_{12} & \sigma_1^2 \end{bmatrix} = \begin{bmatrix} a & -b \\ -b & c \end{bmatrix}. \quad 7\text{-}9$$

The Mahalanobis or statistical distance is different from the Euclidean distance by the inclusion of the inverse covariance matrix:

$$\begin{aligned} \Delta &= (x-\mu)^T \sum^{-1}(x-\mu)] \\ &= [\, x_1 \quad x_2 \,] \begin{bmatrix} a & -b \\ -b & c \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \\ &= ax_1^2 - 2bx_1x_2 + cx_2^2 \end{aligned} \quad 7\text{-}10$$

For the special case of $a = c$, and $b = 0$, this is the same as Euclidean distance. For the case of $a > c$, and $b = 0$, this is a horizontal ellipse. For the case of $a < c$, and $b = 0$, this is a vertical ellipse. For the case of $a \neq c$, and $b \neq 0$, this is a rotated ellipse. Hence, Mahalanobis distance is more flexible and can handle elongated clusters better than Euclidean distance.

The following steps outline the algorithm that uses the Mahalanobis distance:

1. Starts with randomly placing the centroids of N clusters.

2. Compute the mean matrix μ and the covariance matrix $\sum$ of the clusters.

3. Compute the Mahalanobis distance between the data point and each of the centroids.

3. To find the clusters, assign the data point to the nearest centroid.

4. Repeat steps 2, 3, and 4 until convergence.

GMM available in Scikit-learn achieves the same result but is implemented differently.

The following steps outline the GMM algorithm:

1. Starts with randomly initializing the Gaussian distribution parameters μ and $\sum$ for each cluster.

2. GMM iterates until convergence using the Expectation-Maximization (EM) algorithm.

Density-Based Spatial Clustering of Application with Noise

DBSCAN is a density-based clustering algorithm that forms clusters of dense regions of data points and ignores the low-density areas by considering them as noise. Hence, DBSCAN has an advantage in handling clusters with irregular shapes and data points with noisy outliers.

DBSCAN uses two parameters eps, and min_$samples$. eps defines the maximum distance between two points for them to belong to the same cluster. min_$samples$ defines the minimum number of data points a cluster must-have. The following outlines the DBSCAN algorithm:

1. For each data point calculate its distance from all other points. If the distance is within eps, it is a neighbour of the corresponding data point. If the data point has a number of neighbours greater than or equal to min_$samples$, it is considered a core point.

2. For each core point that has not been assigned to a cluster create a new cluster. For this core point, find all its neighbouring points and assign them to the same cluster.

3. Continue step 2 until all the non-core points are covered.

I will demonstrate the effectiveness of the three clustering algorithms in recognizing trends and rejecting outliers using FB from shot records from the Hussar 2D (Margrave et.al 2012). The shot record is linear moveout (LMO) corrected with time correction of 0 ms at 0 m offset and 1600 ms at 4500 m offset, and bulk shifted by 500 ms (Figure 5). Automatic FB picking is done using the AIC algorithm. Three groups of mis-picks are introduced at around sequential receiver locations 120, 220 and 320. The FB picks are input to the K-Means algorithms, GMM algorithms and DBSCAN algorithms. The parameter for the K-Means and GMM algorithms is 20 clusters. The parameters for the DBSCAN algorithm are $eps = 0.08$ and min_samples=3. Figures 6a to 6c compare the results from the 3 algorithms. All three algorithms capture the trend of the FB picks; however, only DBSCAN can reject the mis-picks at around sequential receiver 120, 200 and 320. Figure 6c displays the input FB as black dots and the interpolated DBSCAN picks as a blue line. Figure 7a displays the input FB picks and interpolated DBSCAN in the same scale as the

LMO shot gather 203. Figure 7b overlays the interpolated DBSCAN picks on LMO shot gather 203. This shows DBSCAN is a good algorithm for rejecting outlying picks in FB picks.

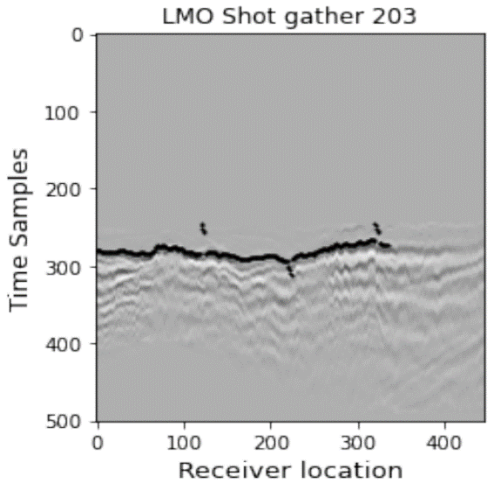


Figure 5. LMO and bulk shifted shot records with FB picks displayed as black dots.

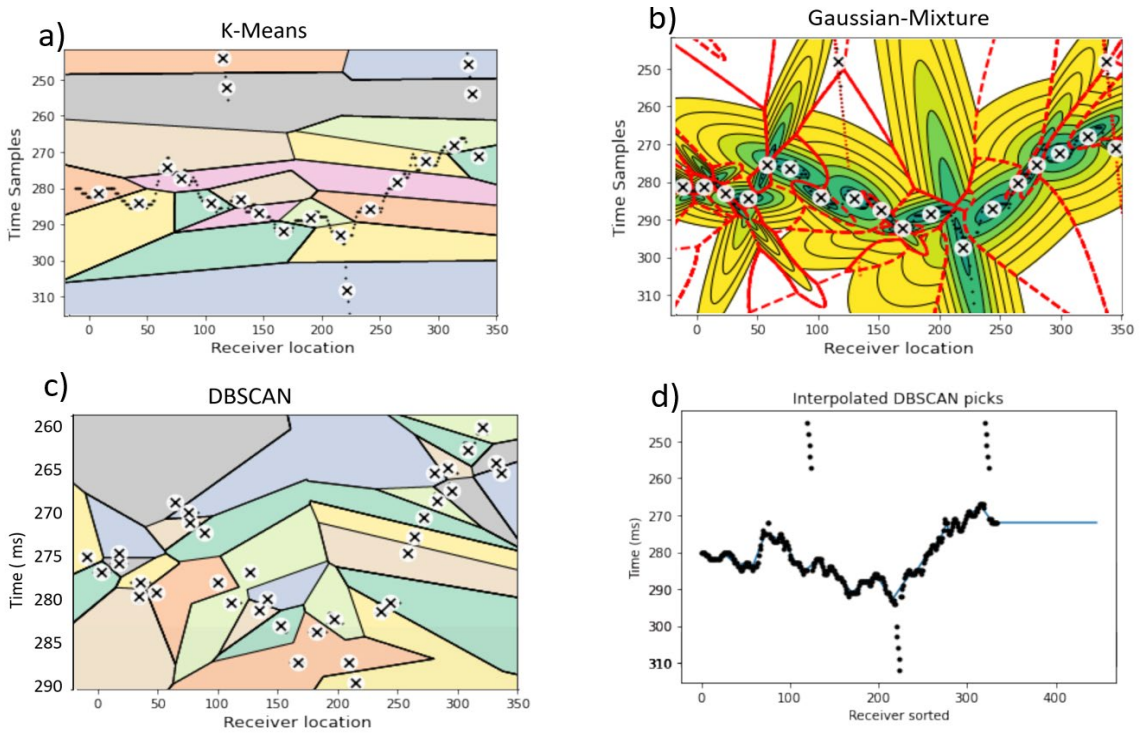


Figure 6. (a) K-Means cluster boundaries and centroid, b) GMM cluster boundaries and centroids, c) DBSCAN cluster boundaries and centroids, d) Interpolated DBSCAN centroids displayed as a blue line and input FB picks displayed as black dots.

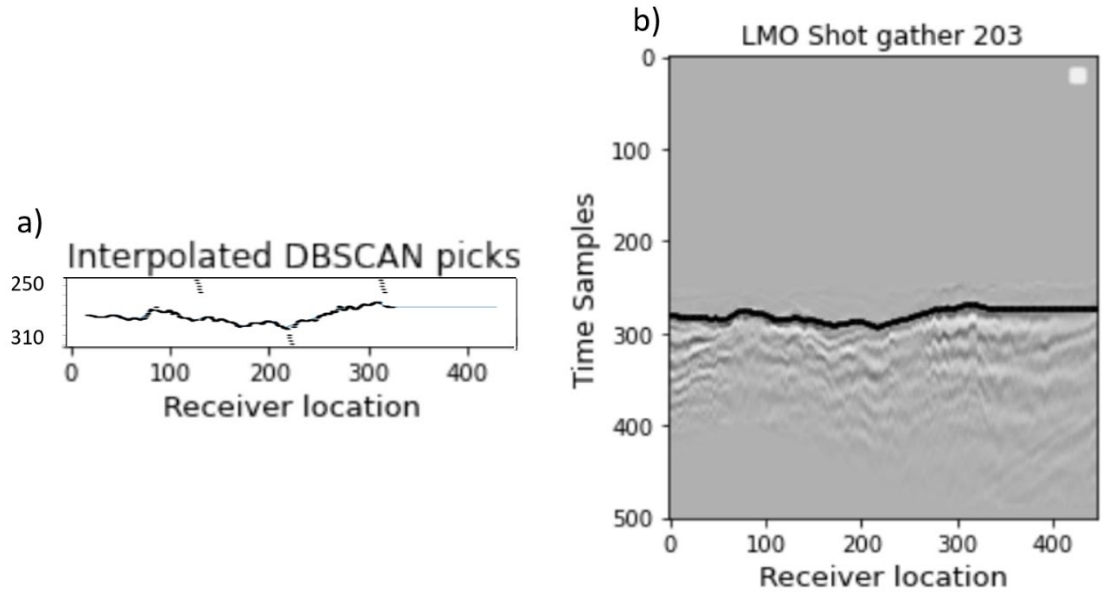


Figure 7. a) Interpolated DBSCAN centroids displayed as a blue line and input FB picks displayed as black dots in the same scale as the LMO shot gather, b) LMO shot gather 203 overlays with interpolated DBSCAN picks.

Deep learning with UNET

Another approach for FB picking is using image segmentation (Hu et al., 2019, Fernhout et al, 2020), which is a supervised deep learning technique. By training the networks with images of the first arrival energy waveform and pre-FB and post-FB masks created by automatically picked, and manually edited FB (Figure 8). With a catalogue of images of trained models from the regions of similar near-surface geology, the deep neural networks will be able to classify the first arrival energy waveforms of a new dataset as pre-FB and post-FB as accurately as the trained technicians. The ultimate goal of the network is to classify each pixel of an input image according to the class to which it belongs.

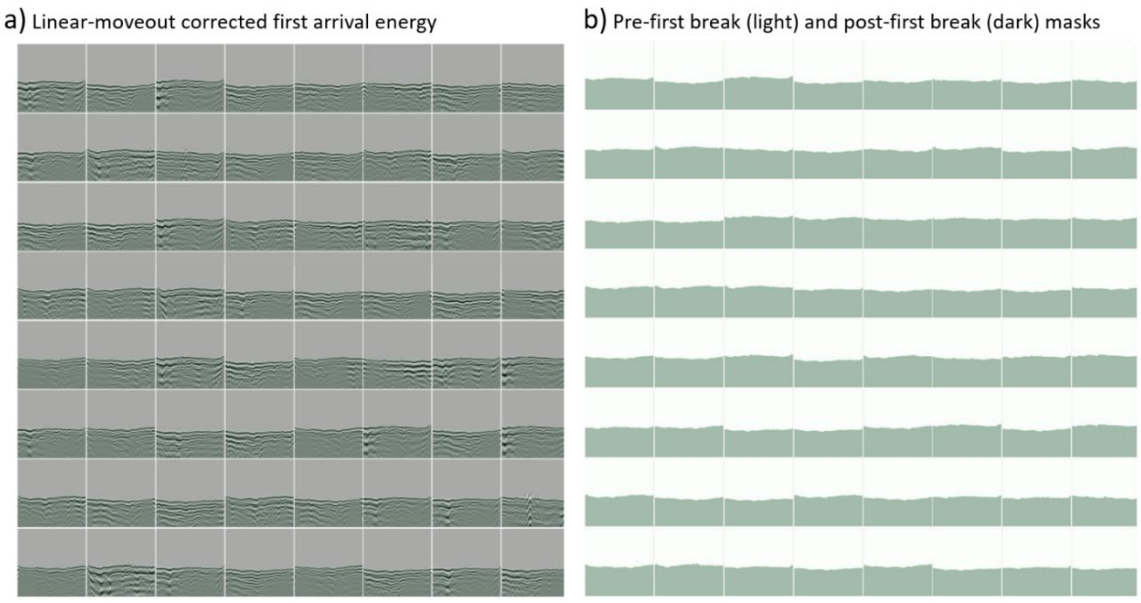


Figure 8. Linear-moveout corrected shot records, b) Corresponding pre-first break and post-first break masks.

This can be solved as an image segmentation problem using UNET (Ronneberger et al, 2015). One-half of the UNET increases the depth of the feature maps and downsizes the resolution of the image using convolutional filters during the encoding process, while the other side of UNET increases the resolution of the output using transpose convolution filters during the decoding process (Figure 9). The input image size of an LMO corrected shot record is 258 traces and 501 samples. This is resized to 256 traces by 256 samples for the UNET. The first break mask is either pre-FB or post-FB; therefore, has a depth of 1. Figure 10 describes the simple UNET used in the problem. The Conv2D filter has a 3x3 filter width and height and 2 feature maps. A 2x2 Maxpooling reduces the image size to 128x128. Another Conv2D filter with a kernel size of 3x3 increases the number of feature maps to 4. The other half of the UNET uses a transpose Conv2D filter to increase the image size back to 256 by 256. The subsequent Conv2D filter reduces the depth of the output segmentation map back to 1.

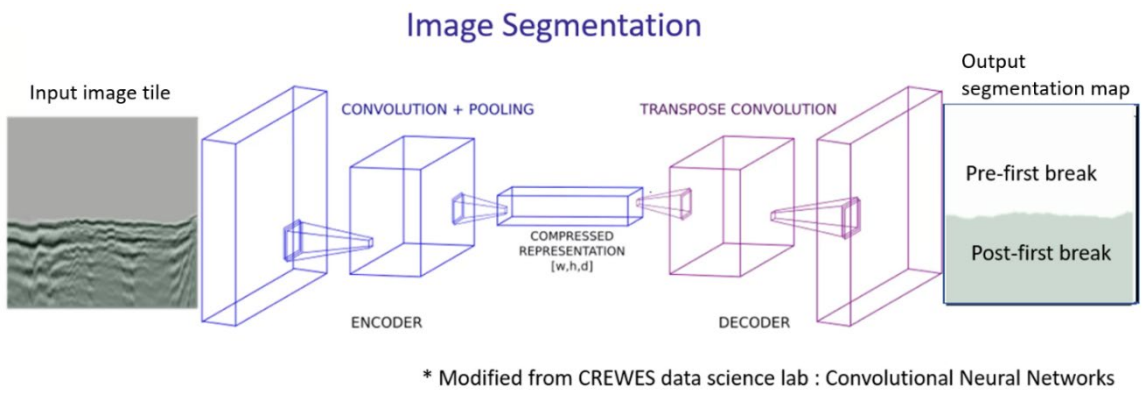
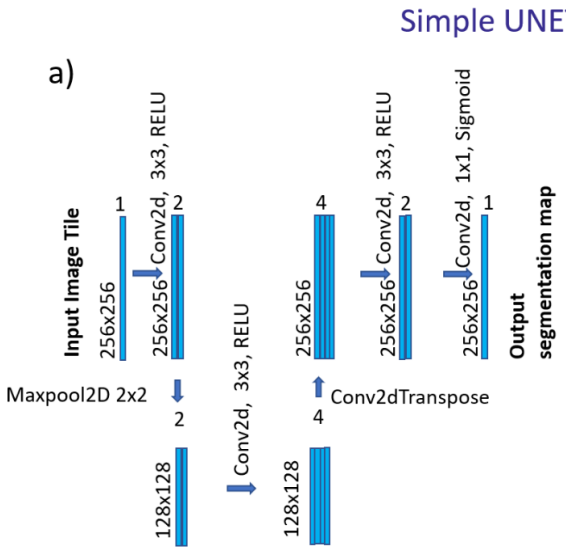


Figure 9. First, break picking as an image segmentation problem.



Layer	Output Shape	Param #
Input	(256,256,1)	0
Conv2D(2,3)	(256,256,2)	20
Maxpool2D(2,2)	(128,128,2)	0
Conv2D(4,3)	(128,128,4)	76
Conv2D_transpose(2,2,2)	(256,256,4)	34
Conv2D(2,3)	(256,256,2)	74
Conv2D(1,1)	(256,256,1)	3

Figure 10. Simple UNET for the first break image segmentation problem.

FIELD DATA EXAMPLE

The vertical component of the Hussar 2D multicomponent seismic survey is used for the UNET test. The first arrival energy is linear moveout corrected to reduce the data size, and the first arrival picking is done automatically using the AIC method, followed by manual editing. The edited FB times are used to create pre-FB and post-FB masks. Convolutional neural networks require input images to be of the same dimensions; therefore, we extract data from 0 to 2230 meters offset from each shot. The result is 258 positive and negative 2D spreads of 224 traces and 501 samples each. The input images are then resized to 256 traces by 256 samples. The 258 images are separated into 206 training images and 52 validation images. The number of training images is also expanded to 824 images by augmentation. A simple UNET is a setup as described in Figure 10. A validation test using 60 epochs is run. Figures 11a and 11b show the convergence history for the validation test. The network converges after 15 epochs with 96.8% accuracy. Figure

12a shows the superimposed classification predictions on the first arrival energy. Figure 12b shows the classification predictions. The prediction results are good but not perfect because a small percentage of the traces have predicted the earlier FB than what is expected according to the first arrival images. This may be the result of the lack of training images for the UNET. If more training images are available, the prediction results are expected to improve.

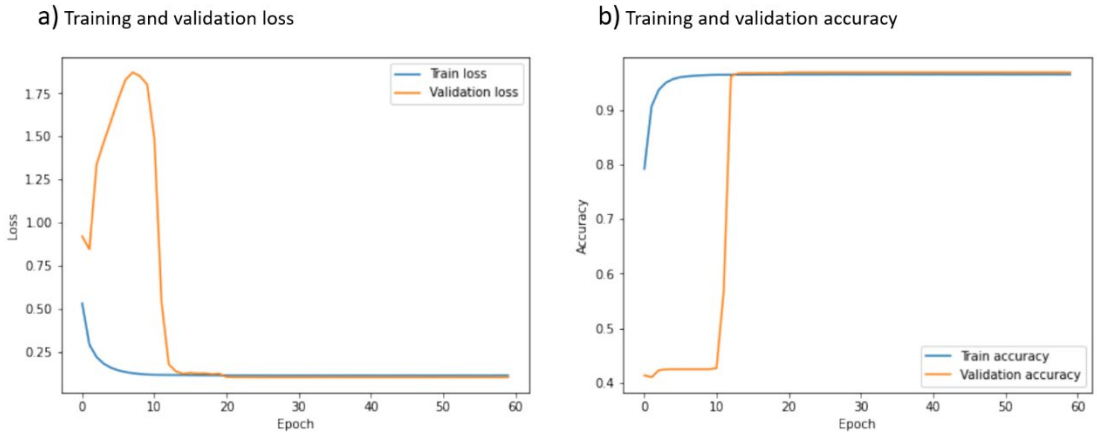


Figure 11. (a) Training and validation loss displayed in blue and green respectively, b) training and validation accuracy displayed in blue and green respectively.

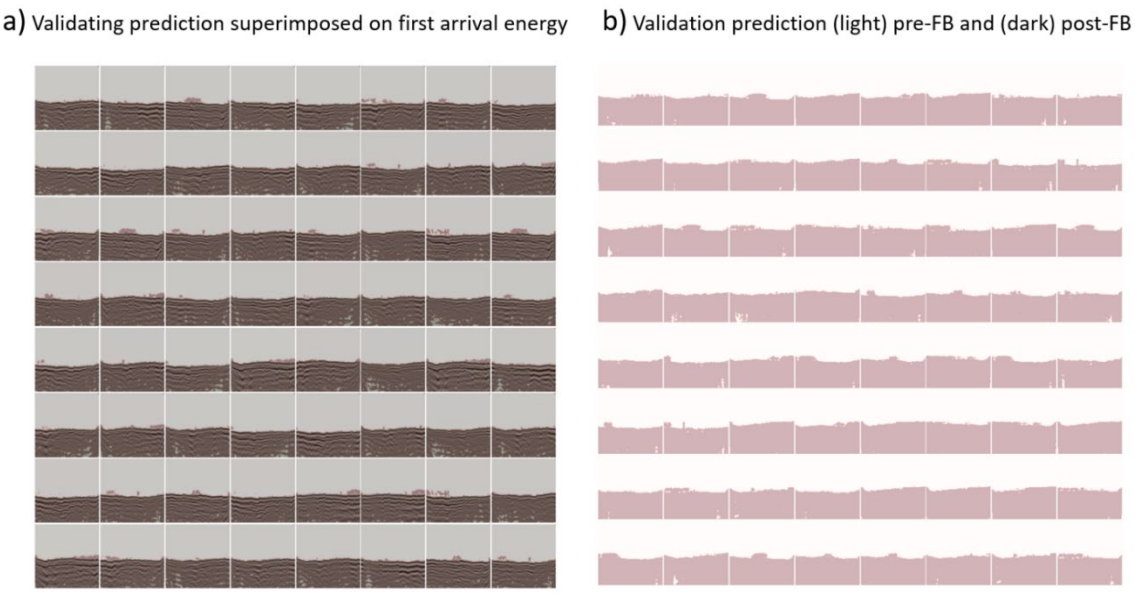


Figure 12. a) Validation prediction results and first arrival energy displayed together, b) Validation prediction results.

CONCLUSIONS

When the first arrival energy is contaminated with noise, experienced technicians are required to confirm, or modify FB picks based on their experience in the regional near-surface geology and in recognizing FB trends through the complex first arrival waveform mixed with noises. With increasing data density, this has become a very time-consuming and expensive process. We used the vertical component of the multi-components Hussar 2D survey to demonstrate the application of machine learning in the most important part of near-surface velocity model building: FB picking. The first application is to reject outlying FB picks using three clustering algorithms, K-Means, GMM and DBSCAN. Only DBSCAN can reject outlying FB picks introduced in the test. The more promising application of machine learning in FB picking is supervised deep learning. We define 2 classifications for the first arrival energy: pre-FB and post-FB. FB are picked automatically using the AIC method and edited interactively to create the training masks. With the Hussar dataset, we created 258 training images and masks of equal size. Because of not having enough samples for training, we were forced to use a very simple UNET to model the first arrival events. The network converges after 15 epochs with 96.8% accuracy. We only used 258 training images in this test, while in a production processing environment one will have more than thousands of training images. Therefore, a better match between the FB mask and first arrival energy can be achieved in a production processing environment with more training images.

ACKNOWLEDGEMENTS

We thank the sponsors of CREWES for their continued support. This work was funded by CREWES industrial sponsors and NSERC (Natural Science and Engineering Research Council of Canada) through the grant CRDPJ 543578-19.

REFERENCES

Akaike, H., 1973. Information theory and an extension of the maximum likelihood principle,[w:] proceedings of the 2nd international symposium on information, bn petrow, f. *Czaki, Akademiai Kiado, Budapest*.

bin Waheed, U., Al-Zahrani, S., Hanafy, S., 2019, Machine learning algorithms for automatic velocity picking: K-means vs. DBSCAN: SEG Technical Program Expanded Abstracts, 5110-5114.

Fernhout, C., Zwartjes, P. and Yoo, J., 2020, Automatic First Break picking with Deep Learning. IOSR Journal of Applied Geology and Geophysics (IOSR-JAGG) e-ISSN: 2321–0990, p-ISSN: 2321–0982.Volume 8, Issue 5 Ser. I (Sep. – Oct. 2020), PP 24-36 www.iosrjournals.org

Hu, L., Zheng, X., Duan, Y., Yan, X., Hu, Y. and Zhang, X., 2019. First-arrival picking with a U-net convolutional network. *Geophysics*, *84*(6), pp.U45-U57.

Ronneberger, O., Fischer, P. and Brox, T., 2015, October. U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical image computing and computer-assisted intervention (pp. 234-241). Springer, Cham.

Smith, K., 2017, Machine learning assisted velocity autopicking: SEG Technical Program Expanded Abstracts, 5686-5690.

St-Onge, A., 2011. Akaike information criterion applied to detecting first arrival times on microseismic data. In *SEG Technical Program Expanded Abstracts 2011* (pp. 1658-1662). Society of Exploration Geophysicists.

Tan, P.N., Steinbach, M., Kumar, V., 2005, Cluster Analysis: Basic Concepts and Algorithms. Introduction to Data Mining, first edition: Addison-Wesley, 487-568.

.