PaperID **AU478**

Author **Kalyan Saikia , Halliburton , India**

Co-Authors **KALYAN SAIKIA**

Enhancement in Seismic Facies Delineation and Better Reservoir Characterization Through Improved Waveform Classification Method—A Case Study

Rohit Arora and Kalyan Saikia, Halliburton

Abstract

Supervised and unsupervised methods are commonly used for seismic waveform classification (SWC). However, these techniques are not appropriate for correlation and classification of time series data, such as seismic volumes, because the Euclidian distance algorithm used to measure distances between time series samples is not able to compensate for the low correlation caused by the scaling effect on the waveforms. This paper presents a new approach that is capable of capturing the scaling effect on the waveforms while classifying the data into groups. This is achieved by merging waveforms of various size but the same shape. This approach improves the process of facies identification using seismic data and helps reduce uncertainty associated with using the waveform classification method for seismic data.

Introduction

SWC is a popular method among seismic interpreters for grouping seismic responses and identifying various rock facies within a defined time window in the dataset. Waveform classification is achieved by choosing a horizon—generally, the reservoir top or bottom—and a zone of interest in which the comparison of traces will occur to identify different waveform groups. Such groups help identify certain hidden properties within the dataset, such as an indication of hydrocarbon, the presence of sand, the location of sweet spots, etc. SWC has been used for several decades for exploration as well as in development assets to overcome challenges, such as heterogeneity around a well, identifying the extent of overpressure around a well, determining where to drill around a producing well, creating a facies map for a reservoir model, etc. In most cases, it is a reinforcement technique that helps provide certainty of a subsurface phenomenon if the SWC reinforces previous knowledge (Zhao et al., 2015). The concept of waveform classification is based on the idea that similar vertical stacking of lithologies (and fluid content) will produce a similar seismic response (Hill et al., 2006).

Currently, the oil and gas industry uses two methods for waveform classification: supervised and unsupervised. The supervised technique begins with a predefined number of groups based on reference locations—generally, well locations—and classifies the entire data on a horizon into similar waveforms. In the unsupervised method, the number of groups is provided but no reference waveforms exist. However, while applying these techniques, the interpreter needs to be aware of certain limitations associated with each. Most importantly, the supervised (k-means) algorithm is a valuable tool for clustering points but is not suitable for time series data because the correlation between two time series is heavily dependent on scaling associated with the waveforms. To overcome this challenge, a new approach is proposed in which waveform classes are merged to compensate for the scaling effect and re-group the facies classes.

Theoretical Background: Seismic Signal and Waveform Generation

While acquiring seismic data, a seismic wavelet (short pulse, amplitude) is sent into the earth. This wavelet interacts with each lithological interface below the surface and then returns to the recording device. Each interface can scale or change the polarity of the pulse (Figure 1). For example, if an unit amplitude interacts with an interface, it can become 0.7 or -0.3, but it is not possible to generate exactly same unit amplitude [all energy in zero time cannot be generated, hence it is only possible to send a short pulse (i.e., the wavelet, generally represented by a Ricker wavelet)].

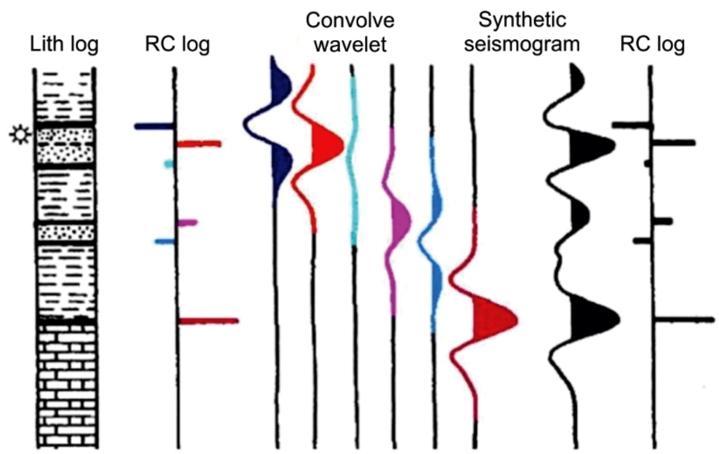

Figure 1: The input wavelet interacts with each interface (dark blue, red, blue, purple, light blue, and dark red interfaces), and each of these can scale it or flip its polarity. When these individual contributions are combined, it produces the theoretical model of the seismic trace called the **synthetic** seismogram, and this process is referred to as convolution.

The natural process that produces seismic amplitude traces is referred to as convolution. Hence, with a known lithology (density and sonic velocity), it is possible to create synthetic seismograms. The litho-column is important while recording the seismic traces or while creating synthetic seismograms. Thus, similar vertical stacking of lithologies should give rise to similar seismic waveforms. This concept is used in the SWC procedure.

Supervise Classification (k-means) Algorithm

The k-means clustering method is used to partition any dataset into k clusters. The important concept in this algorithm is that *a priori* knowledge exists of the number of clusters (k). The first step of the process specifies the number of clusters. Each of these clusters has its own "centroid." Any data point (x, y) can be chosen as the centroid of a cluster. If two clusters are selected (k = 2), two centroids will exist. Any chosen point (x, y) could be taken as the initial centroid because they continuously update with the clustering process. In fact, this process can be completely random. If there are six points to cluster and it is necessary to assign all points to two clusters, then one can throw a dice two times and choose the two centroids. In the following example (Table 1), a dice was thrown twice, which resulted in 1 and 2; hence, the first and second points were selected as centroids. The distance of all the other points in the dataset is calculated from each of the centroids per the Euclidean distance formula **$d = \surd\ (x_2\text{-}x_1)^2 + (y_2\text{-}y_1)^2$**.

X	Y	Distance from C1	Distance from C2	Assignment
195	82			**C1**
180	66			**C2**
178	70	20.808	4.472	Belongs to Cluster 2
189	78	7.21	15	Belongs to Cluster 1
192	82	3	20	Belongs to Cluster 1
198	87	5.83	27.65	Belongs to Cluster 1

Table 1: Centroid assignment.

The distance of point (178, 70) from C1 is determined as follows:

$Dc1 = \surd\ (195 - 178)^2 + (82 - 70)^2 = (17)^2 + (12)^2 = \surd 433$

The distance of the same point (178, 70) from the other centroid (of Cluster2) is determined as follows:

$Dc1 = \surd(180 - 178)^2 + (66 - 70)^2 = 2^2 + 4^2 = 4 + 16 = \surd 20$

The new centroid for each group is the average coordinate of all of its members. Hence, new centroid C1:

C1* = (195 + 189 + 192 + 198)/4, (82 + 78 + 82 + 87)/4 = (193.5, 82.25)

C2* = (180 + 178)/2, (66 + 70)/2 = (179, 68)

Now, in the second iteration, the distance of all points is calculated again from these new centroids, and this process is repeated until the objects stop moving to new clusters. At this point, it can be considered that k-means has reached its stability and no more iteration is required.

K-means for Time Series vs. Dynamic Time Warping

Euclidean distance is used in both supervised and unsupervised waveform classification. However, Euclidian distance cannot compare two waveforms at different scales. The concept is similar to someone not being able to recognize a voice speaking very fast vs. very slow; it is essentially the same voice, but Euclidean distance cannot identify the similarity between the two audio time series.

Conversely, dynamic time warping (DTW) is an algorithm that measures similarities between two time series that are not aligned in the time axis and might be of different durations. In the previous example, the i^{th} sample of one time series is not aligned with the i^{th} sample of the other series; hence, Euclidean distance will conclude that the time series are dissimilar (Figure 2).

Figure 2: Comparison of Euclidean and DTW methods.

Proposed SWC Method

Similarity between time series using correlation is heavily affected by random noise present in the data. It is surprising that even after applying a number of processing and post-processing steps that random noise somehow is present in the final seismic volume. Thus, it is suggested to apply noise-reduction filters, such as a structural filter, or to calculate attributes, such as fault likelihood, before using the dataset for waveform classification (Figure 3). This helps ensure achieving better classification with cleaner seismic data.

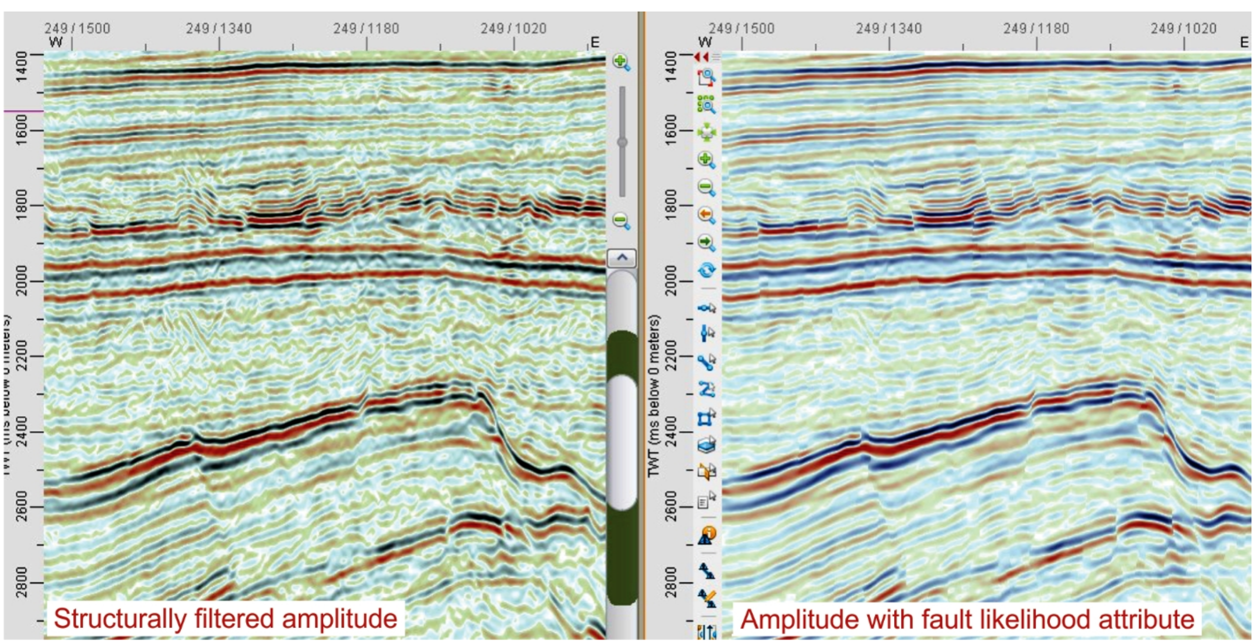

Figure 3: Cleaned seismic section after applying noise-reduction filters.

After cleaning data, the reservoir top (or reference horizon) is interpreted and a window is chosen. Next, the standard k-means classification method with Euclidean distance is used to classify the data. At this stage, it should be noted that many similar waveforms will be incorrectly classified and considered in separate groups. This shortcoming is a result of using the standard classification process through k-means and Euclidean distance, as discussed previously. To compensate for this negative effect on classification, the following workflow was adopted.

In the current study, nine wells as reference waveforms are used to arrive at a data grouping within a predefined window. Figure 4 shows the location of the nine wells. The color code represents the groups (i.e., facies) as obtained from standard k-means classification. The next step was to investigate well-to-well correlation. One method is to calculate the synthetic seismogram at these well locations and compare them. Figure 5 shows the correlation of all nine wells with computed synthetic seismograms.

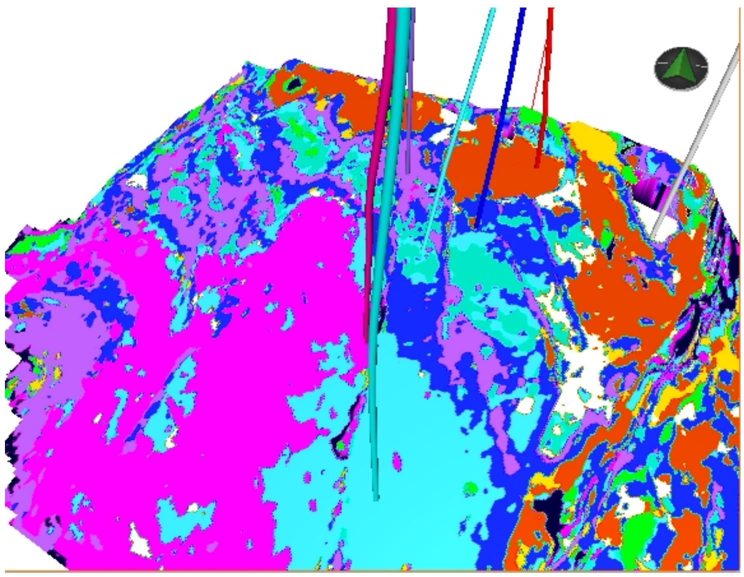

Figure 4: Classified seismic data with location.

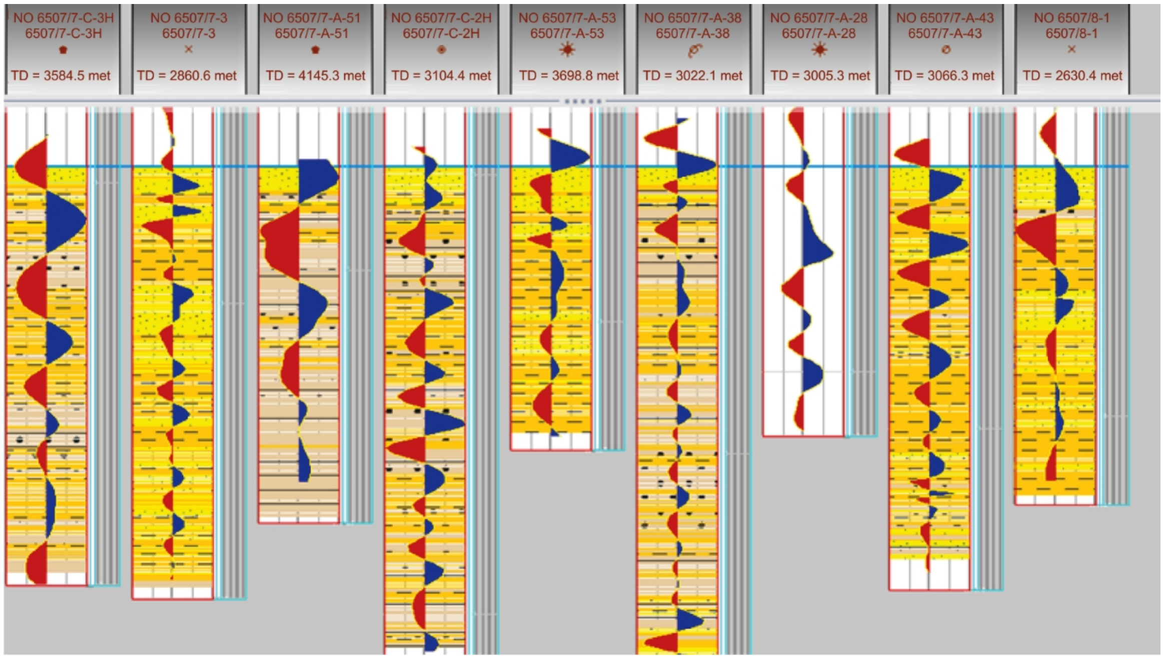

Figure 5: Correlation of synthetic seismograms.

A closer observation of all nine wells revealed that these synthetic seismograms are different from one another. Further investigation into the entire litho-column of each well showed that the overburden height on the top of the reservoir varied considerably among the wells. This resulted in stretching and squeezing of reservoir rock at these well locations with respect to varying overburden load, leading to varying waveforms sizes (shorter waveforms for squeezed and longer waveforms for stretched zones). Then, it was attempted to stretch and squeeze the synthetic seismograms in the correlation panel to determine whether these waveforms were actually identical to each other. After several attempts of stretching and squeezing, matches among the synthetic seismograms of the nine wells were obtained, resulting in four groups (Figure 6).

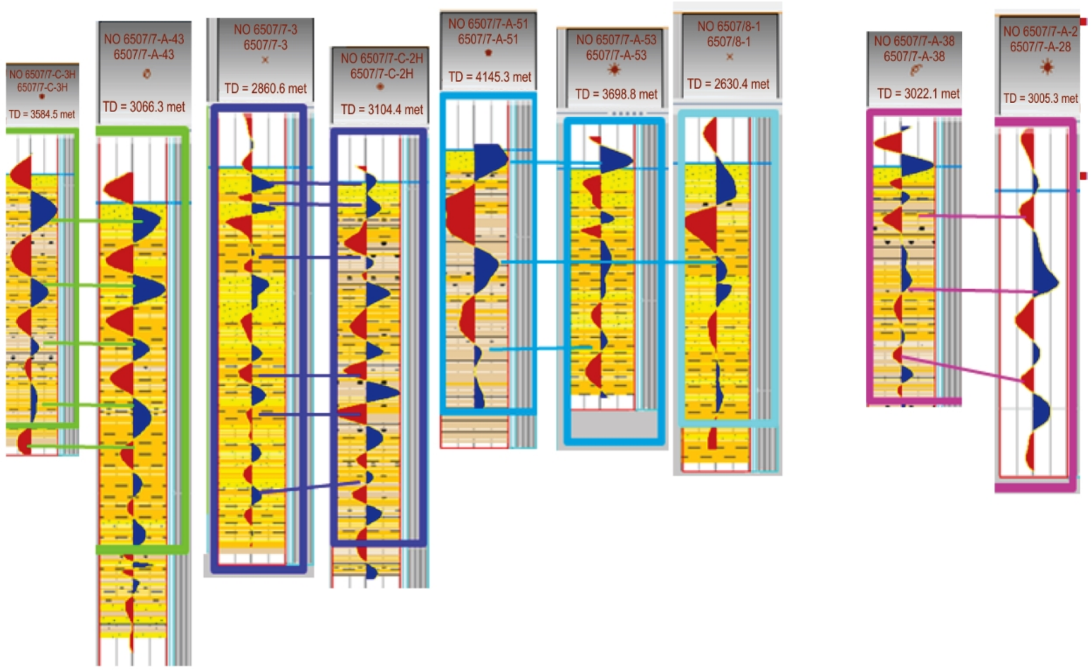

Figure 6: Correlated synthetic seismograms resulting in four groups.

The stretching and squeezing operation of synthetic seismograms, as explained previously, revealed underlying similarities in terms of lesser groups among different waveforms, which was used to perform the supervised classification. Thus, the original classification could be regrouped into four facies instead of nine classes. Some of the facies groups were scaled versions of the other nine facies groups originally identified. The last step of the process was to generate the final classification map after regrouping the original nine classes into four groups (Figure 7). This provided a better waveform classification and was a more accurate representation of the subsurface lithological distribution.

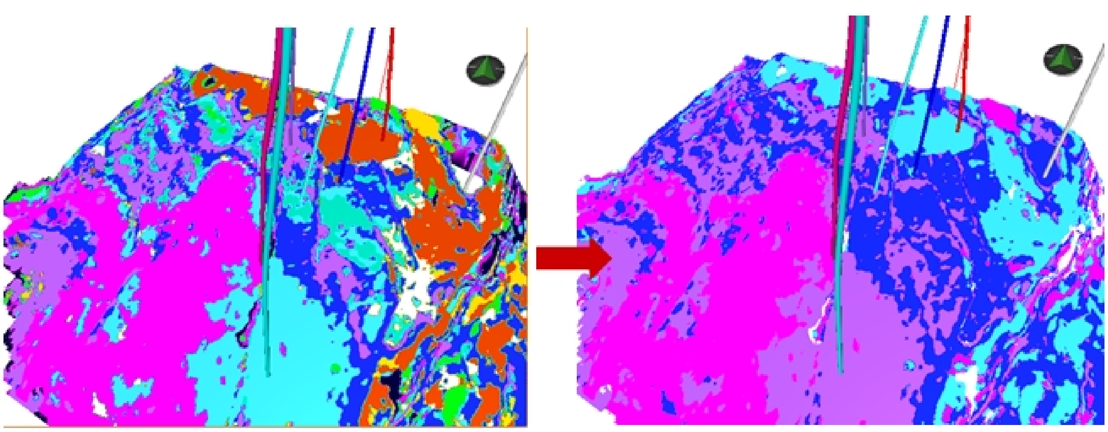

Figure 7: Result of the merged waveform classes.

Conclusion

Using the Euclidean distance method in the k-means clustering technique for comparing two scaled waveforms to classify seismic data is not accurate. To improve the result of such classification, a method is proposed in which similar waveforms are merged through stretching and squeezing. This process produces a more realistic group of rock facies. While applying this approach, using filtered seismic data, particularly those obtained after the fault likelihood process, is highly recommended.

A limitation of this approach is that the litho-columns are not exactly identical. However, it should be noted that sand and shale are relative terms and the litho-columns were created "per well." Additionally, it might not be correct to assume the percentage of sand is driving the shape waveform. The waveform is a result of the location of the interfaces and not the percentage of sand in a column. The same percentage of sand present at different interfaces can result in different waveforms.

References

Hill, S. J., Marfurt, K., and Chopra, S., 2006, Searching for similarity in a slab of seismic data. The Leading Edge, Vol. 25, pp. 168-177.

Zhao, T., Jayaram, V., Roy, A. and Marfurt, J. K., 2015, A comparison of classification techniques for seismic facies recognition. Interpretation, Vol. 3(4), pp. SAE29-SAE58.