A knowledge-integration framework for interpreting seismic facies

Iván D. Marroquín

Abstract

In recent years, the size of seismic data volumes and the number of seismic attributes available have increased. As a result, the task of recognizing seismic anomalies for the prediction of stratigraphic features or reservoir properties can be overwhelming. One way to evaluate a large amount of data and understand potential geologic trends is to automate seismic facies classification. However, the interpretation of seismic facies remains an elusive issue. Interpreters are confronted with the selection of the clustering technique and the optimal number of seismic facies that best uncover the spatial distribution of seismic facies. An interpretation framework combining data visualization with the results from various clustering techniques was evaluated. The framework allows interpreters to be directly involved in the seismic facies classification process. Because of the active participation, interpreters (1) gain insight into the detected seismic facies, (2) verify hypotheses with respect to the spatial distribution of seismic facies, (3) compare different seismic facies classification, and (4) gain more confidence with the seismic facies interpretation.

Introduction

The task of seismic interpreters for exploring hydrocarbon accumulations becomes more complex as large volumes of seismic data and their derivatives (e.g., prestack, instantaneous, geometric attributes, etc.) are made available. The introduction of automated seismic facies classification has provided an efficient means to analyze huge amounts of these data. The output of the classification is a single map (or volume) of seismic facies that can be interpreted to represent changes in lithology, rock properties, or fluid content of the strata.

Interpreters often encounter challenging issues and constraints when conducting seismic facies classification. There are challenges inherent to the characteristics of the seismic data (e.g., acquisition, processing, lateral continuity, coherent noise, redundancy, etc.) that may prevent well-separated clusters from forming or even obscure the cluster structure in the data (Coléou et al., 2003; Marroquín et al., 2009a). Additionally, the subjectivity surrounding the choice of an appropriate clustering technique and the optimal number of seismic facies can influence the quality of the facies classification (Grira et al., 2004; Jain, 2010). A visual-based interpretation framework is proposed to diminish the subjectivity involved in undertaking seismic facies. The framework promotes the integration of the interpreter’s expertise, such that the relationships among the found clusters of seismic facies are examined in a geologic context. To demonstrate the application of the proposed framework, channelized deposits imaged in a 3D seismic data volume were analyzed.

Methodology

Unsupervised clustering analysis is a process involving unsupervised cluster analysis for discovering and identifying meaningful clusters (e.g., seismic facies), without a priori information concerning the membership of a seismic response to a given cluster (Halkidi et al., 2001; Xu and Wunsch, 2005). The seismic responses belonging to a cluster share the same characteristics and are distinct from those seismic responses in other clusters (John et al., 2008).

Interpreters use automated seismic facies classification to identify important stratigraphic or reservoir characteristics from the seismic data, usually without well data to guide the classification. Examples are the mapping of thin clastic reservoirs interbedded between coal and shale layers (Chandra et al., 2003), the delineation of channel systems (Poupon et al., 2004; Cao et al., 2005), the prediction of thin-bed reservoirs (Xie et al., 2004), and the identification of lithofacies geometry within carbonate buildups (Farzadi, 2006). More recent studies include the interpretation of palaeokarst geobodies and sedimentary patterns of a carbonate turbidite (Farzadi and Hesthammer, 2007), the mapping of lithologic changes in a tidal channel and pinnacle reefs (Marroquín et al., 2009a), and the identification of hydraulic fracturing on a shale formation (Roy and Marfurt, 2011).

Automated seismic facies classification is a process involving unsupervised cluster analysis for discovering and identifying meaningful clusters (e.g., seismic facies), without a priori information concerning the membership of a seismic response to a given cluster (Halkidi et al., 2001; Xu and Wunsch, 2005). The seismic responses belonging to a cluster share the same characteristics and are distinct from those seismic responses in other clusters (John et al., 2008).

Interpreters often encounter challenging issues and constraints when conducting seismic facies classification. There are challenges inherent to the characteristics of the seismic data (e.g., acquisition, processing, lateral continuity, coherent noise, redundancy, etc.) that may prevent well-separated clusters from forming or even obscure the cluster structure in the data (Coléou et al., 2003; Marroquín et al., 2009a). Additionally, the subjectivity surrounding the choice of an appropriate clustering technique and the optimal number of seismic facies can influence the quality of the facies classification (Grira et al., 2004; Jain, 2010). A visual-based interpretation framework is proposed to diminish the subjectivity involved in undertaking seismic facies. The framework promotes the integration of the interpreter’s expertise, such that the relationships among the found clusters of seismic facies are examined in a geologic context. To demonstrate the application of the proposed framework, channelized deposits imaged in a 3D seismic data volume were analyzed.

Methodology

Unsupervised clustering analysis is a learning technique to discover the natural grouping of patterns in
a data set. In this mode of learning, there is no training set or a prior knowledge of the number of clusters. On the contrary, the analysis is exploratory by nature (Jain, 2010). Numerous clustering techniques from diverse application domains have been developed (Kotsiantis and Pintelas, 2004; Boryczka, 2008). However, the successful application of clustering techniques remains a challenging task (Griya et al., 2004). First, a clustering technique can give excellent results with one data set type, but may produce poor results — or fail — with other data sets. Second, the number of clusters influences the quality of the partitioning (Xu and Wunsch, 2005; Jain, 2010).

Because each clustering technique has its own approach to impose a clustering structure in the data set (Halkidi and Vazirgiannis, 2001), it is important to assess the results from various clustering techniques in a systematic and comparative manner. The following series of steps (Figure 1) is proposed to standardize the seismic facies classification process (Halkidi et al., 2001; Simpson et al., 2010):

1) Data selection: The data are selected based on seismic features, well data, or background geologic knowledge. Even though this choice is crucial to the effectiveness of clustering techniques, data are rarely selected in this way. In fact, faulty assumptions for data selection may lead to improper seismic facies and incorrect interpretation of the resulting classification.

2) Clustering technique selection: Interpreters should not have a preference for any particular clustering technique. Instead, they need to test several techniques to determine a partitioning that best reveals the spatial distribution of seismic facies.

3) Facies classification validation: The use of appropriate analysis tools is required to confirm the validity and quality of the seismic facies. The assessment should address questions such as, How many seismic facies are needed to partition the data? or, Why choose one clustering technique over another? In this analysis phase, interpreters explore the clusters of seismic facies using the coordinated visual examination of graphic display components: (1) facies map (Figure 2a) and (2) neuron viewer (Figure 2b). The facies map (Figure 2a) shows the seismic trace colored based on the color of the seismic facies with which the trace has maximum correlation. The neuron viewer display has two panels (Figure 2b). The seismic facies shape and its color are shown in the upper part, and the correlation curve in the lower part illustrates the cumulative difference from one seismic facies to the next. The following interpretation guidelines for the correlation curve are proposed:

- A straight line sloping upward is the desired curve (e.g., each facies is equally different from its neighbors).
- A flat segment could indicate redundant seismic facies.
- A sharp bend could suggest too few seismic facies or the presence of an abrupt geologic boundary.
- A curvilinear trend could denote too few seismic facies or not enough trace shape variability in the seismic data.

4) Results interpretation: The interpretation of the seismic facies provides interpreters with meaningful insights into the seismic features of the data. Although the previous step sheds light onto the reliability of the seismic facies, the results of the interpretation step involve the discovery of a natural clustering structure underlying the seismic data. So then, the seismic facies can be related to stratigraphic features or reservoir properties.

With the objective of evaluating the visual-based framework, three commercially available clustering techniques...
were tested. A review of the characteristics of these techniques is discussed below:

1) The self-organizing maps technique: Kohonen’s (1995) algorithm is implemented in this technique. Nodes are organized as a sequence of reference vectors in a 1D grid. The technique projects the input data onto these nodes. As a result, data samples near each other in the input space are mapped to nearby nodes. The clustering starts by first specifying arbitrary nodes. Then, a sample input is chosen and mapped to the closest node (Figure 3a). The winner node and its adjacent neighbors are adjusted toward the sample input. The above step is repeated over several iterations, in which other winner nodes — along with their closest neighbors — are updated to become more like the selected input sample (Figure 3b). The aspect of competitive and cooperative learning of the algorithm results in a gradational pattern of the seismic facies. For this technique, the number of nodes corresponds to the requested number of seismic facies.

2) The hierarchical technique: The hierarchical technique uses an agglomerative algorithm (Kauffman and Rousseeuw, 1990) to organize the input data into a cluster hierarchy. The resulting cluster scheme is depicted by a dendogram (Figure 4). The height of the joining nodes indicates the distance between two pair of clusters (Schonlau, 2002). The algorithm begins with each sample input as a separate cluster, and then it successively combines the closest pair of clusters based on a dissimilarity matrix computed between cluster centers. The algorithm iterates until the number of grouped clusters of the dendogram in construction is equal to the requested number of seismic facies. The hierarchical technique produces a hard partitioning of the input data, in which the patterns of the seismic facies vary abruptly from one facies to the next.

3) The hybrid technique: The hybrid technique is based on a three-stage clustering approach. It combines the best characteristics of the partition-based K-means (MacQueen, 1967) and agglomerative clustering (Kauffman and Rousseeuw, 1990) algorithms to optimize the clustering output. The three stages implemented in the hybrid technique correspond to (Klein and Peloso, 2006):

- Preprocessing: creates an initial partition of the input data set using the K-means algorithm (Figure 5a).
- Clustering process: merges the initial clusters to create many new clusters equal to the requested number of seismic facies using the agglomerative algorithm (Figure 5b).
- Postprocessing: fine tunes the new cluster centers to improve the inter-class inertia of the clusters (e.g., creates tightly bound clusters) by iteratively reallocating the cluster members using the K-means algorithm (Figure 5c).

The hybrid technique, similar to the hierarchical technique, produces a hard partitioning of the input data. The result is a seismic facies pattern that varies sharply from one facies to the next.
Visual-based framework

Determining the most suitable clustering technique and finding the optimal number of seismic facies (K) are the key components that benefit interpreters when assessing the validity and quality of the seismic facies. Several approaches have been introduced to find an optimal K. De Matos et al. (2007) suggest combining 2D self-organizing maps (SOMs) with the K-means algorithm to estimate K; Marroquín et al. (2009a, 2009b) show a visual data-mining computational methodology for finding K. Xiong et al. (2010) propose an index criterion to determine K.

Xiong et al. (2010) argue that the interpreter’s workload and the level of uncertainty of the reservoir properties prediction increase because the selection of the number of seismic facies is performed, essentially, by trial and error. To address this area of concern, Marroquín et al. (2009a, 2009b) discuss the importance of data visualization to guide the interpreter’s analysis and understanding of the seismic facies grouping and distribution, and then relate these facies to stratigraphic features or reservoir properties.

For the seismic facies classification process, exploratory data analysis (EDA) (Tukey, 1977) is an effective approach to judge the presence and nature of a cluster structure. EDA performs an analysis using visual techniques for establishing the objectives to evaluate the results and potentially reformulates the objectives according to the analysis of the results (Behrens, 1997). An essential step in visual data exploration is the requirement that interpreters formulate hypotheses:

1) How many seismic facies probably exist based on the inferred geologic setting?
2) What information does a particular seismic facies contain with regard to the geologic setting?
3) How does a single seismic facies relate to the other facies in terms of continuity and spatial relationship?
4) How does a result from one clustering technique compare to the results from other techniques?

Thomas and Cook (2005) argue that data visualization plays an important role in supporting the task of EDA because of the following reasons:

1) Increase cognitive resources: Visual techniques help to expand geologic knowledge and experiences.
2) Reduce search for patterns: Large amounts of data are summarized using graphical components.
3) Enhance the recognition of patterns: Multiple graphic components help to reveal subtle information.
4) Support the perceptual inference of relationships: The combination of various graphic components helps to identify relationships.

The value of combining data visualization with the results from various clustering techniques led to the development of the visual-based interpretation framework. Note that the framework is an intuitive and iterative process divided into three main steps (Figure 6).

In each step, interpreters accomplish well-defined tasks. The first step is to choose a clustering technique. Then, the classification process is run with three different numbers of seismic facies. The values should be adjusted to represent cases from few (K_LOW) to a large (K_HIGH) number of seismic facies through an intermediate case (K_MIDDLE). Thereafter, interpreters select one classification scheme to be tested. The second step is to define the optimal number of seismic facies (K_OPT) by running another classification with K_NEW and compare their results with the selected classification scheme from the previous step. After choosing the optimal number of seismic facies (K_OPT), the next step takes place. In the third step, the classification scheme of the previous step is compared with the output from other clustering techniques. In this step, interpreters select the facies partitioning that best captures the features of interest in the strata.

**Results**

The seismic data volume comes from a survey grid in the Western Canadian Sedimentary Basin in western Saskatchewan. The survey covers an area of approximately 16 km² (6.2 mi²) with a bin size of 20 × 30 m (66 × 98 ft). The seismic facies was conducted over an interval consisting of 20 seismic amplitude samples. The interval contains a series of blocky, discontinuous reflectors interpreted to represent a multichannel

![Figure 6. Flowchart of the proposed framework for interpreting seismic facies.](http://library.seg.org/)
A representative seismic section is shown in Figure 7. No wells or core are available from the survey area, so the interpretation of channelized deposits is inferred from the facies maps and the geologic facies that would be expected in this type of depositional environment (e.g., Miall, 1992).

First step: The analysis started with the SOMs technique. Three seismic facies classifications were run with $K_{LOW} = 6$ (Figure 8a), $K_{MIDDLE} = 9$ (Figure 8b), and $K_{HIGH} = 15$ (Figure 8c). From the facies maps, the overall topological ordering of the seismic facies does not change significantly when going from $K_{LOW}$ (Figure 9a) to $K_{HIGH}$ (Figure 9c) through $K_{MIDDLE}$ (Figure 9b). All facies maps effectively distinguish what are interpreted to be fluvial areas from the surrounding rocks. In general, the fluvial areas are represented by facies in green, blue, and violet. The character of channel-fill deposits is expected to be highly variable. Channels migrated laterally, resulting in a mixed of sand bodies, sand interbedded with mud layers, or flood plains truncated by other channels. The stratigraphic succession in the interfluvial areas is characterized by facies in brown, red, and yellow. Note that the presence of an oxbow lake is indicated by patterns of facies in brown. The facies map using $K_{LOW}$ (Figure 8a) suggests a patchy facies distribution within the channels. With an increased number of seismic facies (i.e., $K_{MIDDLE}$ in Figure 8b and $K_{HIGH}$ in Figure 8c), there is a suggestion of even finer lithologic variations inside the channels.

The neuron viewer displays using $K_{LOW}$, $K_{MIDDLE}$, and $K_{HIGH}$ are shown in Figure 9a–9c respectively. The correlation curve using $K_{LOW}$ (Figure 9a) exhibits a curvilinear trend indicating too few seismic facies. There is a bend in the slope at facies 3 (light green), supporting the separation of channel deposits (facies 3 [light green] to 6 [violet]) from the surrounding rocks (facies 1 [brown] and 2 [red]). The correlation curve using $K_{HIGH}$ (Figure 9c) shows two relatively flat segments, and this suggests too many seismic facies. The first segment is defined by clusters of facies 1 (brown) to 6 (yellow) representing the surrounding rocks. In the second segment, clusters of facies 9

**Figure 7.** Cross section A-A’ view through the interval of analysis showing channelized features. The locations of the cross section are shown in Figures 8a–8c, 10a, and 11a–11b.

**Figure 8.** Facies map generated using the SOMs technique with (a) $K_{LOW} = 6$, (b) $K_{MIDDLE} = 9$, and (c) $K_{HIGH} = 15$ seismic facies. The channel margins are indicated by white polygons. The cross section A-A’ is shown in Figure 7. Note the presence of an oxbow lake indicated by facies in brown.
(turquoise) to 15 (violet) correspond to channel deposits. Between the two flat segments, there is an upward-sloping segment defined by clusters of facies 7 (light green) and 8 (green), and they are interpreted as a transitional lithologic zone. The correlation curve using $K_{\text{MIDDLE}}$ (Figure 9b) shows an upward-sloping trend. The trend is a good indication that the seismic facies are different from one another, capturing most of the trace shape variability within the interval of analysis. Two bends in the correlation curve are also observed. The first bend occurs on facies 4 (light green) and discriminates the surrounding rocks (facies 1 [brown] to 3 [orange]) from the channel deposits (facies 5 [green] to 9 [violet]). Similar to the correlation curve using $K_{\text{HIGH}}$ (Figure 9c), the second bend on facies 7 (light blue) corresponds to the passage from the transitional lithologic zone (facies 5 [green] and 6 [turquoise]) into the channel-fill deposits (facies 7 [light blue] to 9 [violet]).

Second step: Based on the tests conducted, the seismic facies classification using $K_{\text{MIDDLE}} = 9$ is an excellent choice for the optimal $K$. To refine the clustering procedure’s output another classification using $K_{\text{NEW}} = 11$ was conducted (Figure 10a and 10b). The facies map using $K_{\text{NEW}}$ (Figure 10a) is similar to the one using $K_{\text{MIDDLE}}$ (Figure 8b). However, the visual comparison of the correlation curves using $K_{\text{NEW}}$ (Figure 10b) versus $K_{\text{MIDDLE}}$ (Figure 9b) shows that there are significant differences. The correlation curve for $K_{\text{NEW}}$ (Figure 10b) has the same behavior as the one observed for $K_{\text{HIGH}}$ (Figure 9c). The classification using $K_{\text{NEW}}$ was interpreted to have too many seismic facies. Based on these deductions, the optimal $K$ is nine seismic facies.

Third step: The classification output of the previous step is compared with the results from the hierarchical and hybrid techniques. Major stratigraphic features defined in the facies map shown in Figure 8b are also observed in the facies maps for hierarchical (Figure 11a).
and hybrid (Figure 11b) techniques. However, the visual appearance of the facies map for the hierarchical technique (Figure 11a) shows important differences. In this facies map, the surrounding rocks are represented by a single facies 4 (light green). Additionally, the rest of the eight facies define the channel features. As a result, the geometry of the channel bodies and their associated deposits are better imaged. Although the correlation curve for the hierarchical (Figure 12a) and hybrid (Figure 12b) techniques show a rapidly increasing upward trend, similar to that observed for the SOMs technique (Figure 9b). The correlation curve for the hierarchical technique (Figure 12a) exhibits a trend that is closer to the ideal curve, confirming this classification to be more suitable to capture the subtle variations in the seismic trace shape. In light of these interpretations, the seismic facies with nine facies using the hierarchical technique is selected to provide the best classification of the multichannel fluvial system.

Conclusions
A visual-based framework to assist the interpretation of seismic facies is proposed. The overall goal of the framework is to reduce the subjectivity surrounding the selection of the clustering technique and the optimal number of seismic facies that best reveals significant geologic trends in the seismic data. To address these difficulties, the framework combines data visualization with the results from various clustering techniques. It does this in a manner that interpreters can (1) easily incorporate their geologic knowledge in the seismic facies, (2) come up with new hypotheses regarding the distribution of seismic facies, (3) compare different clustering techniques and their results, and (4) gain confidence in the interpretation of the seismic facies.

However, the framework has limitations. First, the quality of the seismic facies is contingent on the characteristics of the input data. It is possible that clustering techniques will fail to reveal the natural clustering structure in the data. Therefore, the interpretation of the seismic facies could be flawed. Because of this, one must compare the outcome of several clustering techniques. The choice of appropriate graphic display components is a second limitation. Additional investigation of these limitations is necessary to ensure the accuracy of the interpretation.
has to be done to evaluate other visual representations that could improve the presentation of the seismic facies distribution. Although the framework needs more in-depth testing to determine its effectiveness and general applicability, the concept of gradually refining the classification process offers an intuitive mechanism to identify complex patterns, discriminate between relevant and irrelevant features, and validate relationships among patterns.

Acknowledgments
The author thanks Y. Sun, A. Barnes, and S. Al-Dossary for their constructive comments that have strengthened the content within this manuscript. The author also thanks G. Fischer at Paradigm and N. Rose and B. Payne at Chevron for their valuable input and suggestions. Also, the author thanks Paradigm for permission to publish this work.

References


Tukey, J. W., 1977, Exploratory data analysis: Addison-Wesley.


Iván Dimitri Marroquín received a Ph.D. in geophysics from McGill University, and he has been a senior geophysicist with Paradigm since 2006. He leads the company’s efforts in interpreting seismic facies classifications by means of clustering techniques. His research interests include the use of tools for visual data mining as an approach to discern patterns and trends in seismic data sets.