Abstract accepted for presentation at the 2018 SEG Convention at Anaheim, California. Presentation to be made in session MLDA 3: Facies Classification and Reservoir Properties 2, on October 17, 2018 from 11:25 am to 11:50 am in Room 204B (Anaheim Convention Center)

Seismic facies classification using some unsupervised machine learning methods

Satinder Chopra[*†] and Kurt J. Marfurt[+]*

[†] *TGS, Calgary*; [+] *The University of Oklahoma, Norman*

Summary

The size of the individual seismic surveys has increased over the last decade, along with the generation of megamerge and even larger, what some operators call "gigamerge" surveys. The number of useful attribute volumes has also increased, such that interpreters may need to integrate terabytes of data. During the past several years, various machine learning methods including unsupervised, supervised and deep learning have been developed to better cope with such large amounts of information. In this study we apply several unsupervised machine learning methods to a seismic data volume from the Barents Sea, on which we had previously interpreted shallow high-amplitude anomalies using traditional interactive interpretation workflows. Specifically, we apply K-means, principal component analysis, self-organizing mapping and generative topographic mapping to a suite of attributes and compare them to previously generated P-impedance, porosity and V_{clay} displays, and find that self-organized mapping and the generative topographic mapping provide additional information of interpretation interest.

Introduction

In the late 1980s, seismic facies analysis was carried out on 2D seismic data by visually examining the seismic waveforms that can be characterized by their amplitude, frequency and phase expression. Such information would be posted on maps and contoured to generate facies maps. As seismic data volumes increased in size with the adoption of 3D seismic data in the early 1990s, interpreters found that 3D seismic attributes highlighted patterns that facilitated the human recognition of geologic features on time and horizon slices, thereby both accelerating and further quantifying the interpretation. More recently, computer-assisted seismic facies classification techniques have evolved. Such methods or workflows examine seismic data or their derived geometric, spectral, or geomechanical attributes and assign each voxel to one of a finite number of classes, each of which is assumed to represent seismic facies. Such seismic facies may or may not represent geologic facies or petrophysical rock types. In this workflow, well log data, completion data, or production data are then used to determine if a given seismic facies is unique and should be lumped (or "clustered") with other similar facies determined from attributes with similar attribute expression.

As 3D seismic data volumes and their generated attributes have increased in size, it has become increasingly difficult for the human interpreter to carefully analyze each seismic line and time slice. An increasing trend is to adapt methods such as pattern recognition and machine learning methods used in market research, voice recognition, security, and other fields to the problem of automated seismic facies analysis (Zhao et al., 2015).

Partially driven by the success of companies like Google and Amazon, the terms machine learning, deep data, and data analytics have become buzz words in geoscience. Machine learning refers to teaching computers via mathematical operations how to learn from the provided data and make decisions or predictions. The learning style adopted for a given problem to be solved with computers can be specific, depending not only on the geologic processes examined, but also on the quality of the seismic data used, where an example of the latter is the radically different appearance of the interior of salt domes on older, narrow azimuth time-migrated data and more modern wide-azimuth depth migrated data.

Unsupervised learning uses the attributes themselves as both training data and data to be analyzed. The simplest algorithm is K-means, wherein the interpreter defines the number of facies (clusters) to be found. The algorithm then finds means and standard deviations (more generally, covariance matrices) to determine the center and the extent of each cluster in multidimensional attribute space. Using Bayesian classification rules, the attribute vector at a given voxel is then assigned to the cluster to which it is nearest. Although the term neural networks and clustering are often used to describe self-organized mapping and generative mapping, they are implemented and perhaps better understood to be projection techniques, similar to principal component and independent component analysis, where the goal is to reduce the dimensionality of the data. In this case, the clustering is done after the projection, either through the use of a color map or by drawing polygons on the lower dimensional "latent space".

Supervised learning extracts seismic facies by using that part of the data for which we have already classified using well control, production data, or human experience, as training data, whereby the algorithm constructs a mathematical relation between the seismic attributes at a given voxel and the predefined classes. This relation is then applied to the remainder of the data with the goal of predicting the most likely class or rock property. Artificial neural networks and probabilistic neural networks commonly use seismic attributes as input and typically incorporate only one hidden layer (Hampson et al., 2001) to classify the data. Deep learning is a more recent innovation, whereby some developers propose eliminating the use of seismic attributes, and through the introduction of seven or more hidden layers, let the algorithm find the voxel-to-voxel amplitude relationships that define a given facies. All these methods come under the umbrella of machine learning.

In this study, we discuss some of the methods that are used under unsupervised learning and show their comparative performance on a seismic dataset from the Barents Sea.

Crossplotting

Interpreters have long used commercial software packages to crossplot two or three seismic attributes within a reservoir to define

lithofacies. To calibrate such crossplots, the interpreter also crossplots well log information, using gamma ray and resistivity responses not measured by surface seismic data, along with P-wave, S-wave, and density logs to construct probability density functions for each desired facies, thereby calibrating the seismic crossplot. By enclosing anomalous cluster points in crossplot space, and projecting them back onto the vertical or horizontal sections, the facies of interest are then plotted on vertical or horizontal slices. However, crossplotting more than three attributes becomes difficult to not only implement but also to comprehend.

k-means clustering

k-means clustering partitions a given distribution of unlabeled points into a desired fixed number of groups or clusters. If there are n data points x_i, where $I = 1,2, \ldots n$, then these need to be partitioned into k clusters, such that a cluster is assigned to each data point, and the points within each cluster have greater similarity with one another than the points in another group.

The clustering process begins by assigning at random, k centroids which can serve as centers of the group we wish to form. This way each centroid defines one cluster. The distance between each data point and the centroid of that cluster is now calculated. Traditionally, this distance has been Euclidean distance between two points. A point may now be considered to be within a particular cluster if it is closer to the centroid in that cluster than any other centroid. This way the points in each cluster get reorganized. The next step is to recalculate the centroids based on the reorganized points within each cluster. These two steps are carried out iteratively, until there is no more movement of the centroids, when convergence is achieved.

The Euclidean distance metric assumes that there is no correlation between the classification variables, and that is why it has been used traditionally. What this means is that when the classification variables exhibit a spherical shape of the clusters in crossplot space, a Euclidean distance approach will work. Many times, this requirement is not found to be met as the classification variables are correlated, and the clusters exhibit elliptical shape. In such cases, the traditional k-means cluster method will fail in that convergence may not be achieved. In those cases, the k-means clustering algorithm is modified in that it adopts a different distance metric called the Mahalanobis distance instead of the Euclidean distance.

Thus, by choosing the centroids as points that have high density of neighbouring points as well as using the Mahalanobis distance metric instead of the Euclidean distance in the k-means clustering algorithm, allows it to correctly classify nonspherical and nonhomogeneous clusters.

Principal component analysis

Principal component analysis (PCA) is a useful dimensionality reduction tool. Many of our attributes are coupled through the underlying geology, such that a fault may give rise to lateral changes in waveform, dip, peak frequency, and amplitude. Less desirably, many of our attributes are coupled mathematically, such as alternative measures of coherence (Barnes, 2007) or of a suite of closely spaced spectral components. The amount of attribute redundancy is measured by the covariance matrix. The first step in multiattribute analysis is to subtract the mean of each attribute from the corresponding attribute volume. If the attributes have radically different units of measure, such as frequency measured in Hz, envelope measured in mV, and coherence without dimension, a Z-score normalization is required. The element C_{mn} of an N by N covariance matrix is then simply the cross-correlation between the m^{th} and n^{th} scaled attribute over the volume of interest. Mathematically, the number of linearly independent attributes is defined by the value of eigenvalues and eigenvectors of the covariance matrix. The first eigenvector is a linear combination that represents the most variability in the scaled attributes. The corresponding first eigenvector represents the amount of variability represented. Commonly, each eigenvalue is normalized by the sum of all the eigenvalues, giving us a percentage of the variability represented.

By convention, the first step is to order the eigenvalues from the highest to the lowest. The eigenvector with the highest eigenvalue is the first principal component of the data set (PC1); it represents the vector representing the maximum variance in the data and thereby the bulk of the information that would be common in the attributes used. The eigenvector with the second-highest eigenvalue, called the second principal component, exhibits lower variance and is orthogonal to PC1. PC1 and PC2 will lie in the plane that represents the plane of the data points. Similarly, the third principal component (PC3) will lie in a plane orthogonal to the plane of the first two principal components. In the case of N truly random attributes, each eigenvalue would be identical, and equal to $1/N$. Since seismic attributes are correlated through the underlying geology and the band limitations of the source wavelet, the first two or three principal components will almost always represent the vast majority of the data variability.

Self-organizing maps

The self-organizing mapping (SOM) is a technique that generates a seismic facies map from multiple seismic attributes, again in an unsupervised manner. In a distribution of N attributes lying in an N-dimensional data space, the plane that best fits the data is defined by the first two eigenvectors of the covariance matrix. This plane is then iteratively deformed into a 2D surface called a manifold that better fits the data. After convergence, the N-dimensional data are projected onto this 2D surface, which in turn are mapped against a 2D plane or "latent" (hidden) space, onto which the interpreter either explicitly defines clusters by drawing polygons, or implicitly defines clusters by plotting the results against a 2D colorbar.

Generative Topographic Mapping

The Kohonen self-organizing map described above, while the most popular unsupervised clustering technique, being easy to implement and computationally inexpensive, has limitations. There is no theoretical basis for selecting the training radius, neighborhood function and learning rate as these parameters are data dependent (Bishop et al., 1998; Roy, 2013). No cost function is defined that could be iteratively minimized and would indicate the convergence of the iterations during the training process, and finally no probability density is defined that could yield a confidence measure in the final clustering results. Bishop et al. (1998) developed an alternative approach to the Kohonen self-organizing map approach that overcomes its limitations.

It is called a generative topographic mapping (GTM) algorithm, and is a nonlinear dimension reduction technique that provides a probabilistic representation of the data vectors in latent space.

The GTM method begins with an initial array of grid points arranged on a lower dimensional latent space. Each of the grid points are then nonlinearly mapped onto a similar dimensional non-Euclidean curved surface as a corresponding vector ($\mathbf{m_k}$) embedded into different dimensional data space in GTM. Each data vector ($\mathbf{x_k}$) mapped into this space is modeled as a suite of Gaussian probability density functions centered on these reference vectors ($\mathbf{m_k}$). The components of the Gaussian model are then iteratively made to move toward the data vector that it best represents. Roy (2013) and Roy et al. (2014) describe the details of the method and demonstrate its application for mapping of seismic facies to the Veracruz Basin, Mexico.

As it may have become apparent from the descriptions above, the PCA, SOM and GTM techniques project data from a higher dimensional space (8D when 8 attributes are used) to a lower dimensional space which may be a 2D plane or a 2D deformed surface.

Sometimes the above three techniques are collectively referred to as projection techniques. Once they are projected on a lower dimensional space, the data can be clustered in that space, or interactively clustered with the use of polygons. Though not shown here, this aspect will be demonstrated in the formal presentation

Case study

The case study we demonstrate the comparison of facies classification techniques on is a 3D seismic volume from the Hoop Fault Complex (HFC) area of the western Barents Sea. For this study, we focus on a 500 km^2 section of a larger 22,000 km^2 survey, where Chopra et al. (2017) found multiple bright seismic amplitude anomalies, some of which are associated with the presence of hydrocarbons and some that are not. This previous workflow we adopted consisted of the computation of a set of seismic attributes (spectral decomposition, prestack simultaneous impedance inversion, etc.), and quantification of the hydrocarbon bearing zones with the application of rock physics analysis, and an extended elastic impedance approach.

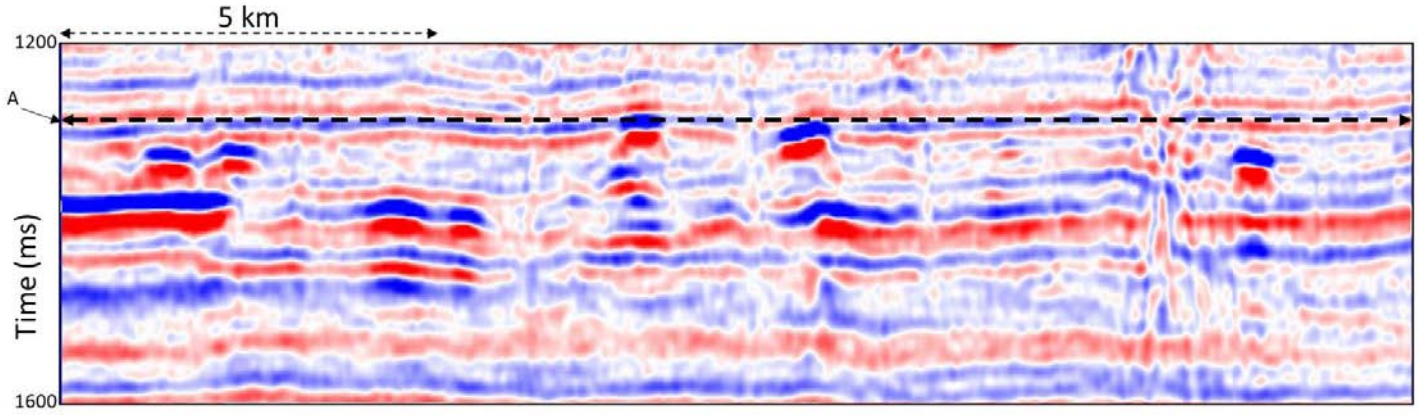

Figure 1: A segment of a section from a seismic volume from the Barents Sea. Time slices at 1280 ms indicated to the left have been extracted from the different attributes and are shown in Figures 2 and 3. Some of the high-amplitude anomalies seen on the section are associated with channel sands, which we attempt to characterize using some unsupervised machine learning techniques. *(Data courtesy: TGS, Asker)*

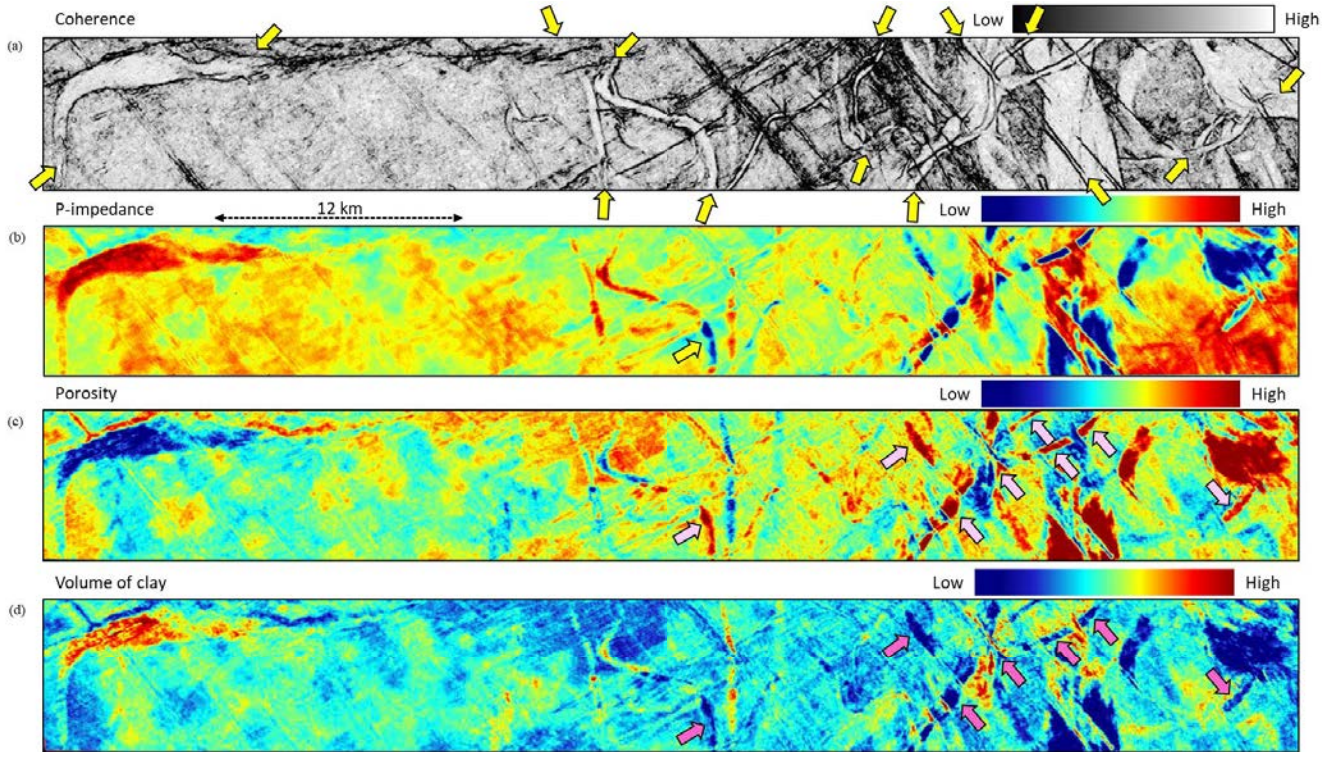

Figure 2: Time slices generated at 1280 ms from (a) energy-ratio coherence, (b) P-impedance, (c) porosity, and (d) V_{clay} volumes. The P-impedance volume was generated using prestack simultaneous impedance inversion, and the porosity and V_{clay} volumes generated using extended elastic impedance. The channels are defined well with the coherence attribute overlaid on the displays (b) to (d) using transparency. Some of the channels exhibiting low impedance, high porosity and low V_{clay} are indicated with pink arrows. *(Data courtesy: TGS, Asker)*

In Figure 1 we show a representative vertical slice through the data volume that exhibits many high amplitude anomalies, several of which are associated with the presence of hydrocarbons. Many of these bright spot anomalies appear to be associated with channels that can be identified on time or horizons slices. Figure 2a, a time slice through energy ratio coherence volume, shows northwest-southeast and northeast-southwest faults as well as a number of well-defined channels. Figures 2b-d show the equivalent time slices first through the P-impedance generated using prestack simultaneous impedance inversion, followed by porosity and V_{clay} volumes derived from the application of extended elastic impedance inversion reported earlier in Chopra et al. (2017). Note those channels associated with high porosity and low V_{clay} values at different levels, as indicated with pink and purple arrows respectively.

Next, we generated volumetric measures using k-means clustering, principal component analysis, SOM and GTM techniques, where the seven input attribute data volumes were P-impedance, S-impedance, GLCM-energy, GLCM-entropy, GLCM-homogeneity, total energy and instantaneous envelope volumes. Figure 3a-f show equivalent time slices at t=1280 ms through these attributes. With the exception of Figure 3a, all other figures are the result of plotting pairs of attributes using a 2D color bar, i.e. PCA1 and PCA2 in Figure 3b, SOM-1 and SOM-2 in Figure 3c, and GTM-1 and GTM-2 in Figure 3d. The energy-ratio coherence attribute shown in Figure 2a is overlaid on these displays using transparency, thereby clearly defining the faults and channel boundaries. We notice that these displays highlight the channel features very nicely in terms of the distribution of properties within their boundaries. The k-means clustering display in Figure 3a does a reasonably good job, and in fact better than the porosity display in Figure 2c. The first principal component (Figure 3b) displays higher values in those channels that are seen to have higher impedance, and are crisper. Between SOM and GTM, the GTM displays exhibit crisper channel features than the SOM displays.

Conclusions

We have shown a comparison of seismic facies classification using K-means clustering, principal component analysis, self-organizing mapping and generative topographic mapping unsupervised machine learning techniques as applied to a seismic volume from the Hoop Fault Complex area of the Barents Sea that had previously undergone careful human interpretation, thereby establishing "ground truth". Application of these techniques to the same data allowed us to assess their relative strengths as well as their suitability, considering the fact that simultaneous inversion and extended elastic impedance require more time and effort.

We found that principal component analysis provided more convincing results than K-means clustering. Both GTM and SOM show promising results, with GTM having an edge over SOM in terms of the detailed distribution of channel facies and distinct definition seen on the displays.

Acknowledgements:

We wish to thank TGS, for encouraging this work, data courtesy, and permission to present.

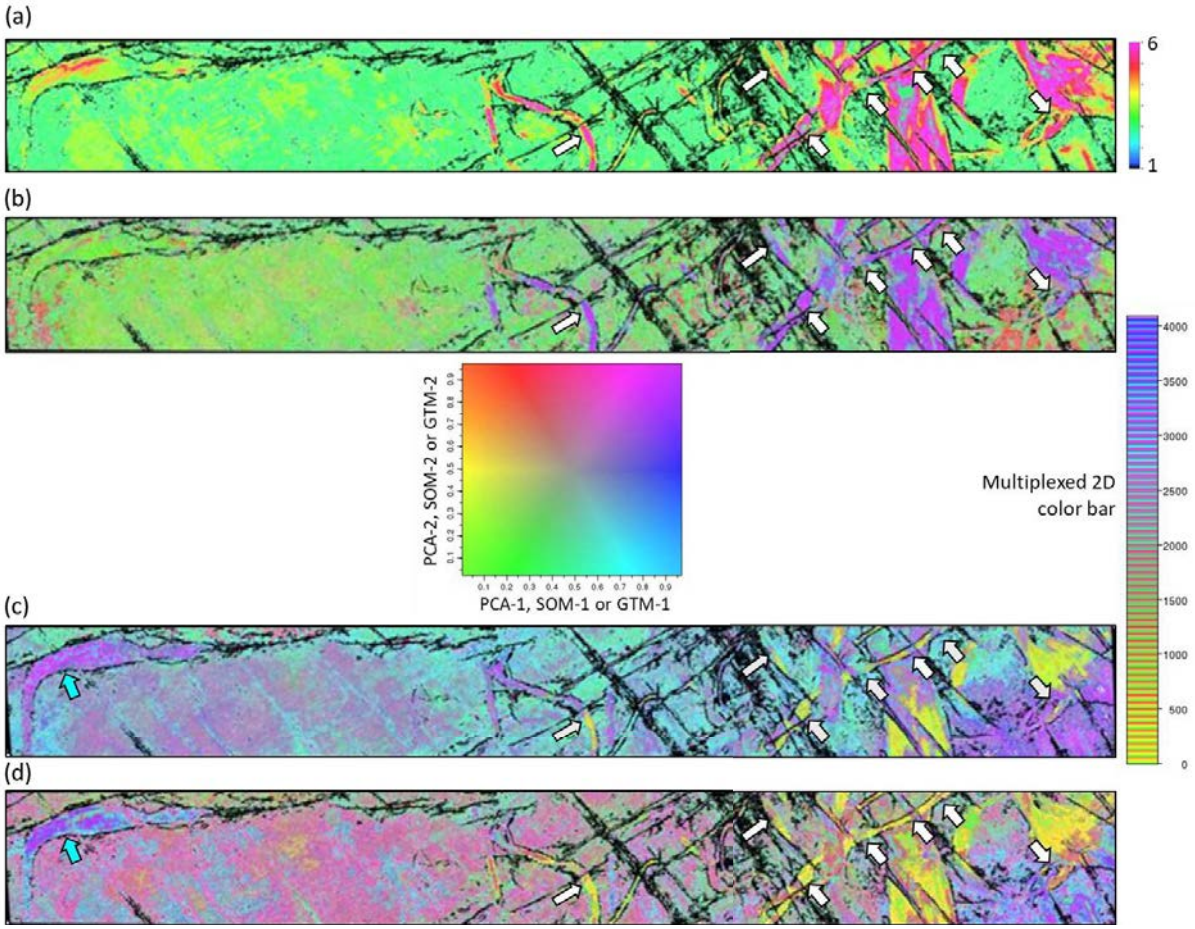

Figure 3: (a) Time slice generated at 1280 ms from (a) K-means clustering co-rendered with coherence using transparency. Equivalent time slice from the first two principal components plotted against a 2D color bar is shown in (b). The 2D color bar as well as the 1D multiplexed 2D color bars are shown in the middle. Similarly, the time slices at 1280 ms from SOM-1 and SOM-2 volume pair, and GTM-1 and GTM-2 volume pair, plotted against a 2D color bar are shown in (c) and (d) respectively. Notice, that while the K-means and principal component displays show the facies distribution within the channels, there is better distribution of the channel facies on the SOM display (channel to the left indicated with cyan arrow). The most detailed distribution of channel facies and their distinct definition is seen on the GTM display. *(Data courtesy: TGS, Asker)*

References

- Barnes, A. E., 2007, Redundant and useless seismic attributes, Geophysics, **72**, P33-P38.
- Bishop, C. M., M. Svensen, and C. K. I. Williams, 1998, The generative topographic mapping: Neural Computation, **10**, No. 1, 215-234.
- Chopra, S., R. K. Sharma, G. K. Grech, B. E. Kjølhamar, 2017, Characterization of shallow high-amplitude seismic anomalies in the Hoop Fault Complex, Barents Sea, Interpretation, **5**, T607-622.
- Hampson, D., Schuelke, J.S., and Quirein, J.A., 2001, Use of multi-attribute transforms to predict log properties from seismic data: Geophysics, **66**, 220-231.
- Roy, A., 2013, Latent space classification of seismic facies: Ph.D. Dissertation, The University of Oklahoma, 212 p.
- Roy, A., A. S. Romero-Pelaez, T. J. Kwiatkowski and K. Marfurt, 2014, Generative topographic mapping for seismic facies estimation of a carbonate wash, Veracruz Basin, southern Mexico, Interpretation, **2**, SA31-SA47.
- Zhao, T., V. Jayaram, A. Roy, and K. J. Marfurt, 2015, A comparison of classification techniques for seismic facies recognition, Interpretation, **3**, SAE29-SAE58.