Unsupervised machine learning facies classification in the Delaware Basin and its comparison with supervised Bayesian facies classification

Satinder Chopra[+], Kurt J. Marfurt[†] and Ritesh Kumar Sharma[+]*
[+]TGS, Calgary; [†]The University of Oklahoma, Norman

Summary

An ongoing challenge to seismic interpreters is to identify and extract heterogeneous seismic facies on data volumes that are continually increasing in size. Geometric, geomechanical, and spectral attributes help to extract key features but add to the number of data volumes to be examined. Common analysis tools include interactive co-rendering, crossplotting, and 3D visualization where we examine more than one attribute at a time, data reduction, where we mathematically reduce the number of data volumes to a more manageable subset, clustering, where the goal is to identify voxels that have similar expressions, and supervised classification, where the computer attempts to mimic the skills of an experienced interpreter. In this study we compare several of the more well-established machine learning techniques: waveform classification, principal component analysis (PCA), *k*-means clustering, and supervised Bayesian classification to a seismic data volume from the Delaware Basin. We also examine some less common clustering techniques applied to seismic attributes including independent component analysis (ICA), self-organizing mapping and generative topographic mapping. We find that the machine learning methods hold promise as each of them exhibits more vertical and spatial resolution than the waveform classification, or the supervised Bayesian classification.

Introduction

The Delaware Basin is one of the most important resource plays in North America, with production being drawn from multiple zones. Depending on the operator and the acreage, the Bone Spring and Wolfcamp Formations are most prolific and thus serve as the most common development target, followed by the Barnett and the Mississippian (Figure 1). The Bone Spring Formation exhibits thicknesses varying from 700 m to 1000 m and comprises sequences of dark gray deep-marine shales interbedded with sands and black limestones. The Wolfcamp Formation consists of dark shale and limestone with silt and sand zones in the central parts of the basin but as carbonate buildups and banks on the shelf areas. Below the Wolfcamp lies the Late Mississippian age brittle siliceous, gray-to-dark gray, and organic rich Barnett Shale. The Early-to-Mid-Mississippian Limestone underlies the Barnett Shale and provides a distinct marker on the seismic data. Our goal is to map the vertical and lateral variations of the facies within each of these lithounits.

Seismic facies classification using machine learning techniques

Machine learning uses mathematical operations to learn from the similarities and differences in the provided data and make decisions or predictions. There are three broad families of machine learning algorithms. The first algorithm family includes dimensionality reduction algorithms such as PCA and ICA. When plotted against a 2D color bar, the interpreter may "see" clusters, but the algorithm output is a continuum of data in a lower dimensional space. The second, unsupervised classification algorithm family attempts to explicitly cluster the data into a finite number of groups that in some metric "best represent" the data provided. Before the analysis, there is no interpretation assigned to any given group; rather, "the data speak for themselves". However, the choice of input attributes biases the clustering to features of interpretation interest. Biasing the training data to favor geologic features of interest (e.g. by more heavily weighting a bright-spot anomaly) also provides interpreter control of the output. In this paper, we apply *self-organized mapping* (SOM) and *generative topographic mapping* (GTM) to the Delaware Basin data volume. The third, supervised classification algorithm family attempts to map each data point or voxel to a suite of features defined by the interpreter. In this paper, we illustrate supervised learning using the well-established Bayesian classification workflow, where the classes are defined by petrophysical analysis. Other supervised learning algorithms such as multilayer feed-forward neural networks, support vector machines, random forest decision trees, and convolutional neural networks require the interpreter to define facies of interest by picking voxels, drawing polygons, or extracting data about microseismic events and or image log anomalies.

Principal component analysis

Principal component analysis (PCA) is a useful dimensionality reduction tool and assumes that the input seismic attributes exhibit a Gaussian distribution.

As mentioned above, 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 crosscorrelation 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

© 2019 SEG
10.1190/segam2019-3214088.1

the plane of the first two principal components. 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.

In Figure 2a we show stratal slices extracted from PCA-1, PCA-2 and PCA-3 data volumes co-rendered together using RGB color scheme, at the top of the Mississippian marker of the Delaware 3D seismic survey. The 3D color bar shown alongside the image is multiplexed into a 1D color bar, which can be used for displaying the co-rendered data volumes from the three principal components. Overlaid in black are the fault/fracture lineaments from the most-positive curvature attribute using transparency.

Independent component analysis

Independent component analysis (ICA) is an elegant machine learning technique that separates multivariate data into independent components, without the requirement of a Gaussian distribution for data going into the analysis. The other differences between ICA and PCA are that the independent components are not orthogonal, and their order is not defined, in that the first, second and third ICAs are ordered by visual examination, and are not mathematically ordered in the process as in PCA (Lubo-Robles, 2018; Chopra et al., 2018).

Given a combination of different seismic attributes as input data, ICA attempts to find the 'mixer' that acts on a number of independent components, which is mathematically cast as a matrix equation, and solved using higher order statistics. We demonstrate its application to multiattribute seismic data, wherein the resultant independent components exhibit better resolution and separation of the geologic features.

In Figure 2b is shown an equivalent stratal display at the Mississippian level from the ICA-1, ICA-2 and ICA-3 RGB co-blended data volume. Notice the appearance of the clusters in different colors resemble the cluster patterns obtained from the PCA co-blended data display in Figure 2a.

k-means clustering

k-means clustering is one of the simplest clustering algorithms and is available in most seismic interpretation software. *k*-means organizes a given distribution of *N* data points, x_n, where $n = 1, 2, \ldots N$, into a desired number of *k* clusters. The clustering process begins by assigning at random *k* centroids which can serve as centers of the groups we wish to form, where each centroid defines one cluster. Next, the distance between each data point and the centroid of that cluster is calculated. A point may be within a cluster if it is closer to the centroid in that cluster than any other centroid. As some reorganization of the points in different clusters has taken place, the centroids are recalculated for each cluster. These two steps are carried out iteratively, until there is no more shifting of the centroids, and the process has converged. The calculation of distance between the centroid and the data points referred to above is the traditional Euclidean distance, which assumes there is no correlation between the classification variables. If this is the case, then the classification variables would exhibit a spherical shape of the clusters in crossplot space. In many cases, this is not found to be true, as the classification variables exhibit cluster that are elliptical in shape, and hence are correlated. In such cases, the traditional *k*-means clustering method might not achieve convergence and hence fail. To avoid this a different distance metric called Mahalanobis distance is used instead of the Euclidean distance. Thus, the *k*-means clustering method using the Mahalanobis distance metric correctly classifies nonspherical and nonhomogeneous clusters.

k-means can be computed along horizons or volumetrically. In Figure 2c we show a stratal slice at the Mississippian marker from the facies volume generated using *k*-means clustering method with five clusters. The input seismic attributes comprising P-impedance, S-impedance, instantaneous amplitude, weighted instantaneous frequency, GLCM-energy, GLCM-homogeneity and total energy. We see different colored patches on the display, which are a representation of the different facies in the data at that level.

Self-organizing maps

Like *k-means*, *self-organizing mapping* (SOM) is a technique that generates a seismic facies map from multiple seismic attributes, again in an unsupervised manner. In contrast to *k*-means, SOM defines its initial cluster centroids in an *N*-dimensional attribute data space, by least-squares fitting the data with a plane that best fits the data defined by the first two eigenvectors of the covariance matrix (Kohonen, 1982, 2001). This plane with centroids locked to it 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.

Figure 2d shows the equivalent stratal display at the Mississippian marker extracted from the SOM-1 and SOM-2 crossplot volume using a 2D color bar as shown alongside. Some of the clusters seen on this display are better defined than the ones shown earlier from PCA and ICA analysis in Figures 2a and b or the *k*-means clustering display in Figure 2c.

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, e.g. the first three principal components or the ICA components. Each of the grid points are then nonlinearly mapped onto a similar dimensional non-Euclidean curved surface as a corresponding vector ($\boldsymbol{m}_k$) embedded into different dimensional data space in GTM. Each data vector ($\boldsymbol{x}_k$) mapped into this space is modeled as a suite of Gaussian probability density functions centered on these reference vectors ($\boldsymbol{m}_k$). 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

10.1190/segam2019-3214088.1

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, ICA, 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. 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.

In Figure 2g we show a stratal slice at the Mississippian markers from the GTM crossplot volumes respectively. Comparing with the equivalent SOM display in Figure 2d, one may conclude that the GTM displays are crisper, and could lead to more accurate interpretations. The log strips by the side depict the facies obtained by machine learning technique application on log data comprising the V_P, GR, NPHI, DPHI log curves. As the input data for facies computation for the logs are different from the input data that goes into the facies computation from seismic data, the two facies classification results may be considered independent. As pointed out with the green arrows, the facies at the location of well W1 is different from the facies seen at the location of the well W2 and is corroborated with the colors on the display. Similarly, on the right-hand-side, the colors on the display at the two well locations (W3 and W4) pointed at with green arrows are similar, and so also are the facies as seen on the two facies strips.

Waveform classification

One of the earliest and more popular pattern recognition techniques is to define seismic facies along an interpreted horizon based on their seismic waveforms. Commonly known as "waveform classification", the actual clustering is generated using multiattribute self-organizing mapping (SOM) where the n^{th} sample of each trace is the n^{th} attribute. After the interpreter defines a hypothesized number of clusters, the SOM algorithm examines the data and determines which centroids in N-dimensional space best represent the data. Plotting a given centroid as attribute, followed by attribute 2, up to attribute N looks like a waveform, giving us the name waveform classification. Each windowed seismic trace is then compared to all of the cluster centroids (sometimes called neurons), where the result a color-coded map showing the nearest centroid. A variation of this workflow is to apply the analysis to stratal slices of Poisson's ratio rather than seismic amplitude, thereby classifying the geomechanical stacking pattern at each mapped location. The resulting map is essentially a facies map, or a similarity map of the actual traces to the centroids (waveforms) that best represent the variability in the data. The seismic facies so generated also can be overlaid on a vertical seismic section to study their lateral variation. Since this method does not require any input in the form of any well log or any guidance about where the character divisions should occur, this approach is referred to as *unsupervised* waveform classification. Figure 2e shows a waveform classification about top Mississippian horizon of the Delaware Basin 3D seismic survey, where six classes were generated.

Bayesian classification

As we are trying to characterize the different lithology units from seismic data between Bone Spring and Mississippian markers, it is possible that different models that we deduce have the same seismic response. Understandably, some of these models will be more probable than others, which we can term as being realistic. Consequently, we can follow an approach that accounts for the uncertainties associated with reservoir characterization in the different lithounits. This work follows the Bayesian classification approach (Grana, 2013) and provides a facies model reflecting the quality of the lithounits and a related uncertainty analysis.

When using Bayesian classification, the interpreter defines the different facies based on the cut off values of density, porosity and neutron porosity well curves. In this paper we identify eight facies for the broad zone from Bone Spring to Mississippian, knowing fully well that all these eight facies may not be seen at either the Bone Spring or the Mississippian levels, but would be distributed over the full zone. Projecting these predictions onto the log curves, augmented by from mud log curves, are then used to ascertain their validity. Sharma et al. (2019) shows that the probability distributions for each of these facies can be represented by Gaussian ellipses. The density porosity and neutron porosity attributes derived from the seismic data using a neural network approach, and the probability density functions for each facies generated from well log data analysis provides the information necessary to generate facies volumes based on Bayesian classification. Stratal displays from the facies volume were generated at different levels and a representative display at the Mississippian level is shown in Figure 2f. The display at the Bone Spring level did not have enough spatial detail due to the lower resolution of the seismic data, as well as the discrete number of facies it was organized into, hence is not shown.

Conclusions

We have shown a comparison of seismic facies classification using the traditional seismic waveform classification, the supervised Bayesian classification, as well as the machine learning methods such as *k*-means, PCA, ICA, SOM and GTM to a seismic volume from the Delaware Basin. Although supervised learning provides answer to questions we know how to ask, it does not answer questions that weren't asked. A common problem is to define classes based on well log data, say sand, shale, and carbonate. In this scenario, any anhydrite found in the seismic data volume would be guaranteed to be misclassified. In contrast, while unsupervised may identify the anhydrite as a distinct class, it provides no indication of what it means geologically. Finally, the selection of the input data is critical. If we wish to differentiate lateral changes in shale properties, geomechanical attributes are valuable input, while attribute such as coherence and curvature may be valuable in delineating lateral compartments.

In summary, we find that the machine methods hold promise as each of them exhibit more vertical and spatial resolution than the waveform classification, or the supervised Bayesian classification. Amongst the machine learning methods, the ICA furnishes more detail than the PCA. Both the SOM and GTM methods provide promising results, with the latter yielding more accurate definition as seen on the displays.

Acknowledgements

We wish to thank TGS for encouraging this work and for the permission to present and publish it. The well data used in this work was obtained from the TGS Well Data Library and is gratefully acknowledged.

10.1190/segam2019-3214088.1

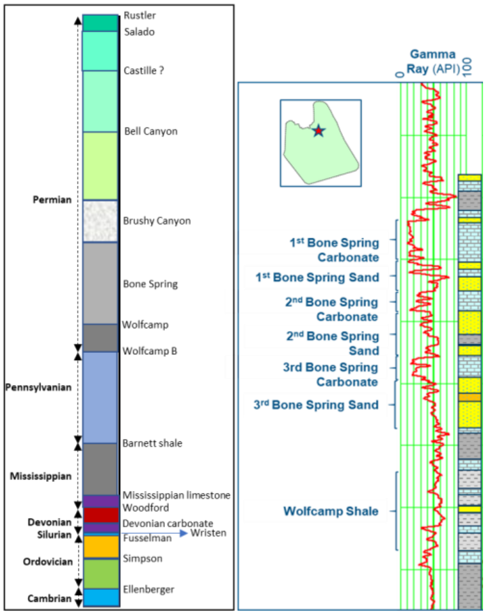


Figure 1 : The generalized stratigraphy of the Delaware Basin and the expanded litho-column for the Bone Spring and Wolfcamp intervals.

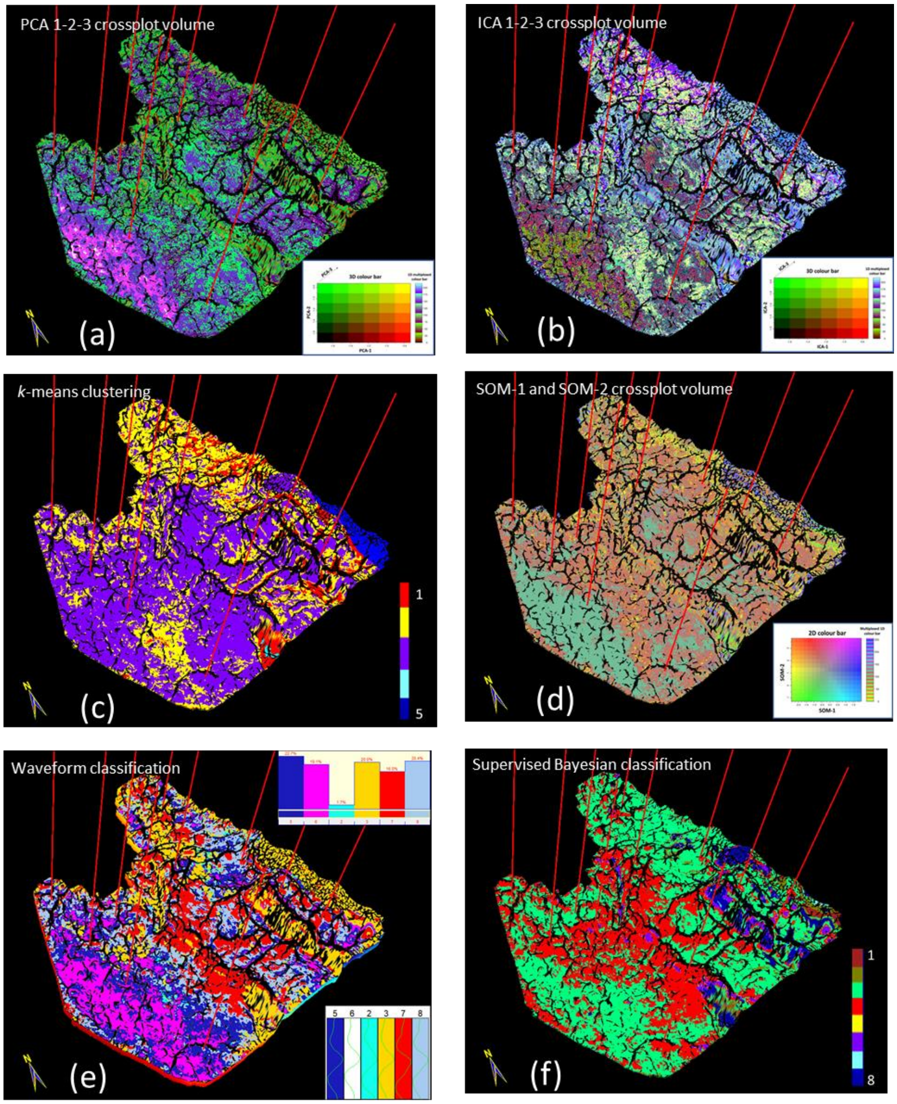
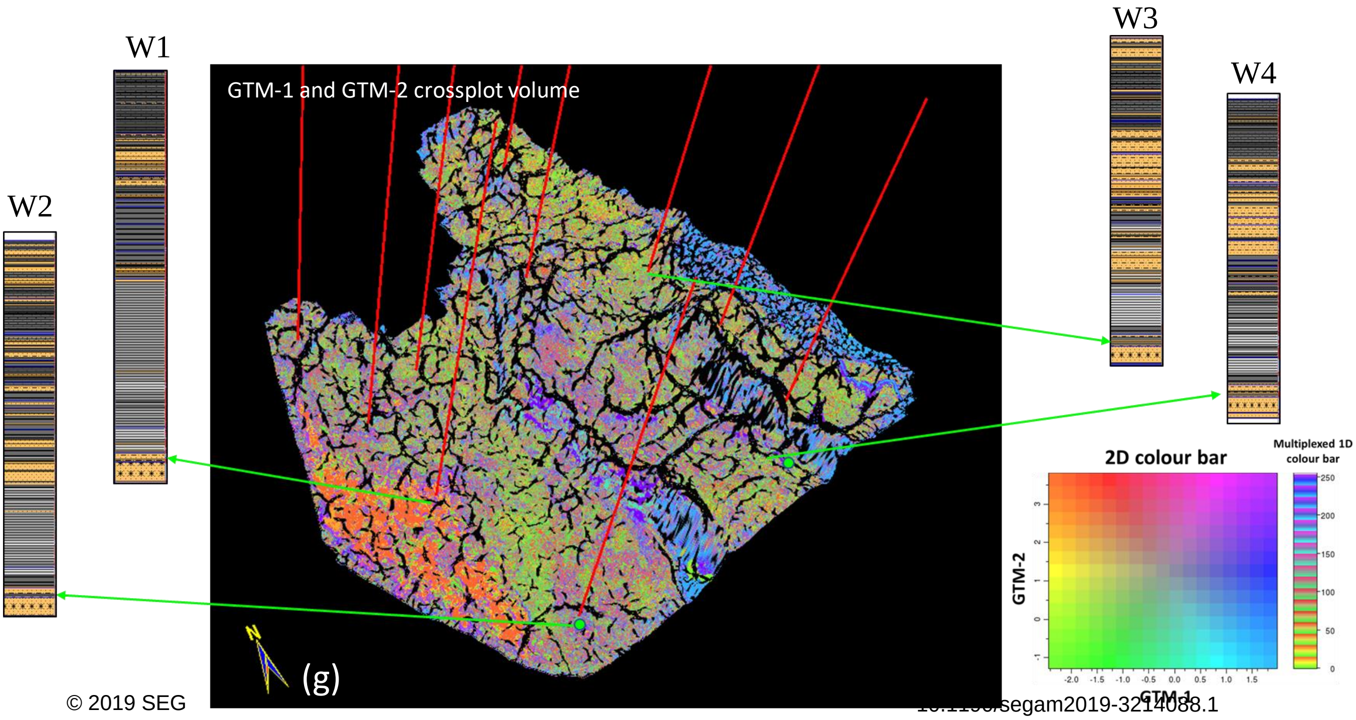

PCA 1-2-3 crossplot volume
(a)
ICA 1-2-3 crossplot volume
(b)
k-means clustering
1
5
(c)
SOM-1 and SOM-2 crossplot volume
2D colour bar
(d)
Waveform classification
5
6
2
3
7
8
(e)
Supervised Bayesian classification
1
8
(f)

Figure 2: Stratal slices at the Mississippian marker from different volumes (a to g) as indicated. The size of the survey is about 1050 km^2. Overlaid is the most-positive curvature attribute lineaments using transparency.



10.1190/segam2019-3214088.1

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**, 215–234.

Chopra, S., D. Lubo-Robles, and K. J. Marfurt, 2018, Some machine learning applications in seismic interpretation: AAPG Explorer, 22–24.

Grana, D., 2013, Bayesian inversion methods for seismic reservoir characterization and time-lapse studies: Ph.D. thesis, Stanford University.

Kohonen, T., 1982, Self-organized formation of topologically correct feature maps: Biological Cybernetics, **43**, 59–69.

Kohonen, T., 2001, Self-organizing Maps: Springer-Verlag.

Lubo-Robles, D., 2018, Development of independent component analysis for reservoir geomorphology and unsupervised seismic facies classification in the Taranaki Basin, New Zealand: M. Sc. thesis, University of Oklahoma.

Roy, A., 2013, Latent space classification of seismic facies: Ph.D. Dissertation, The University of Oklahoma.

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**, no. 1, SA31–SA47.

Sharma, R. K., S. Chopra, and L. R. Lines, 2019, Challenges and uncertainty in the seismic reservoir characterization of Bone Spring and Wolfcamp formations in the Delaware basin using rock physics: 89th Annual International Meeting, SEG, Expanded Abstracts.

10.1190/segam2019-3214088.1