Using multidimensional scaling and kernel principal component analysis to interpret seismic signatures of thin shaly-sand reservoirs

Piyapa Dejtrakulwong¹, Tapan Mukerji², and Gary Mavko¹
¹Department of Geophysics and ²Department of Energy Resources Engineering
Stanford University

Abstract

Characterizing thin shaly-sand reservoirs using seismic data can be challenging due to limited seismic resolution. This paper investigates seismic signatures of thin shaly-sand reservoirs with statistical attributes extracted using multidimensional scaling (MDS) and kernel principal component analysis (KPCA). We model sand-shale sequences as a first-order, discrete, 1-D Markov chain. Then we assign properties to the sand-shale layers using established rock-physics relations. From multiple realizations of synthetic seismograms, we extract statistical attributes using MDS and KPCA. Results show that these attributes can be used to differentiate sequences with different net-to-gross ratios or different water saturations. Thus a workflow similar to this synthetic study using MDS and KPCA can potentially be applied to real seismic data to characterize thin shaly-sand reservoirs.

Introduction

Estimating petrophysical properties from seismic response of thin shaly-sand reservoirs with layer thicknesses below seismic resolvability can be challenging. Previous studies (e.g., Khattri and Gir, 1976; Marion et al., 1994; Takahashi, 2000; Hart and Chen, 2004; Hart, 2008) have shown that both the scales of heterogeneity relative to the scale of seismic measurement and the difference in sediment stacking patterns affect seismic signatures and their relations with rock properties. This study aims to extract statistical attributes from seismograms and use them for quantitative seismic interpretation of thin shaly-sand reservoirs. The attributes considered in this study are obtained using multidimensional scaling (MDS) and kernel principal component analysis (KPCA).

In the next section, we present a workflow for generating thin shaly-sand sequences and their corresponding synthetic seismograms. Then we use MDS and KPCA to investigate seismic signatures of sequences with different net-to-gross ratios and different water saturations.
Forward modeling for seismic response and attributes

In this section, we briefly review methodology and tools used in forward modeling to generate seismic responses and their attributes for thin sand-shale sequences. Our forward-modeling process has three steps: using Markov chain models to generate sequences of sedimentary layers, using rock-physics models to assign physical properties to the layers and simulating seismic responses of the sequences, and extracting seismic attributes.

Markov chain models in stratigraphic sequences

We generate vertical arrangements of sand and shale layers in thin sand-shale sequences using Markov chain models. In stratigraphic analysis, a column of sediments can be described as a chain or spatial arrangement of a finite number of discrete states (i.e. lithology). Markov chains use conditional probabilities to describe the dependency of the current state on the previous states. If the transition from one state to the next depends only on the immediately preceding state, the chain is said to be first-order (Harbaugh and Bonham-Carter, 1970; Sinvhal and Khattri, 1983).

A Markov chain model is commonly represented by a transition matrix, whose element $p_{ij}$ (at the $i^{th}$ row and $j^{th}$ column) represents the probability of a transition from state $i$ to state $j$, or the probability of going to state $j$, given that $i$ is the current state. In a stratigraphic study, the transition matrix is usually obtained from real geological observations and is typically constructed in one of two ways either counting states using a fixed sampling interval, or counting states only when a transition occurs (an embedded form).

When counting states using a fixed sampling interval, the lithologic state is determined and considered only at discrete points equally spaced along a stratigraphic column. This allows successive points to have similar lithology, which implies that the diagonal element (i.e., probability that a state has a transition to itself) can be non-zero (Krumbein and Dacey, 1969). In practice, however, selecting a proper sampling interval for this method can be problematic. Choosing an interval that is too small relative to the overall average bed thickness can increase the counts of transitions of a state to itself. Consequently, the diagonal elements become very large, and probabilities of the state transiting into the others become unreasonably small. In contrast, using a sampling interval that is too large can miss very fine-layered characteristics of the sequences (Sinvhal and Sinvhal, 1992). A sequence simulated by this fixed-sampling type always yields lithologic states with thicknesses that are geometrically distributed (Krumbein and Dacey, 1969).
In contrast, in an embedded-form transition matrix, all diagonal elements are zero, since transitions are considered only when lithologic states change. In this case, the step size between two consecutive states is not a fixed interval, but the actual observed bed thicknesses (Parks et al., 2000). Thus, a sequence of states generated from an embedded transition matrix does not contain information about thicknesses of each layer. Thicknesses are simulated separately assuming that the layer thickness of each lithologic state is distributed according to some distributions (i.e., semi Markov process). A case when the layer thickness is exponentially distributed is called a continuous-time Markov chain model. Examples of transition matrices and sequences obtained from both the fixed-sampling and the embedded-form transition matrices are shown in Figure 1 and Figure 2.

The probability of transition from state $i$ to state $j$ in $n$ steps (denoted as $p^{(n)}_{ij}$) can be determined by raising the transition matrix to power $n$. If a limiting distribution exists, a successive multiplication leads to a row vector of fixed probabilities representing proportions of each state in the long-term behavior. In terms of stratigraphic application, this limiting distribution (or stationary distribution $\pi$) implies proportions of each lithology in the entire sequence provided that sufficient numbers of transition steps occur (Harbaugh and Bonham-Carter, 1970).

We characterize the lithology into four states: sand, shaly sand, sandy shale, and shale. Then using a set of 4x4 fixed-sampling transition matrices, we generate multiple realizations of lithology arrangements, keeping the layer thickness fixed at 0.5 m. Each sand-shale sequence has a total thickness of 100 m, and the sequence is embedded between two 200-m thick shale layers.
The lithologic states in the transition matrices are sand (s), shaly sand (sh-s), sandy shale (s-sh), and shale (sh). The off diagonal elements marked by arrows control the directionality of the sequences.

Figure 1: Three examples of transition matrices with fixed sampling intervals: retrogradational, progradational, and aggradational sequences. The lithologic states in the transition matrices are sand (s), shaly sand (sh-s), sandy shale (s-sh), and shale (sh). The off diagonal elements marked by arrows control the directionality of the sequences.

Figure 2: Examples of an embedded-form transition matrix with realizations of sequences. The lithologic states in the transition matrix are sand (s), shaly sand (sh-s), sandy shale (s-sh), and shale (sh). An example of thickness distributions used is shown in the lower left corner.
**Rock-physics models for sand-shale mixtures**

After obtaining multiple sequences we assign petrophysical properties to those sand-shale layers with rock-physics models. The four lithologic states are represented by mixtures of sand and clay with clay fraction values of 0.1, 0.3, 0.6 and 0.9 in the order of increasing shaliness. Then, porosity values corresponding to the mixtures of sand with the specified clay fractions are determined by using the Yin-Marion dispersed mixing model (Marion et al., 1992; Yin, 1992). This model describes the topology of bimodal mixtures and a V-shape relation between the volume fraction of clay and the mixture porosity. When a small volume of clay (i.e., less than sand porosity) is added to an original packing of sand, clay starts filling the sand pore space without disturbing the sand packing. Sand grains provide the load-bearing matrix of the mixture. At this stage, porosity decreases, because clay particles replace some portions of the original sand pore space. When the clay content is greater than the sand porosity, sand grains are displaced and disconnected. The mixture changes from grain-supported to clay-supported sediments. At this stage, porosity increases linearly with increasing clay content, because the solid sand grains are replaced by a porous chunk of clay. As a result, the plot of volume fractions of clay versus total porosity values of the mixture shows a V-shape pattern (Figure 3) (Marion et al., 1992; Yin, 1992). Then we obtain the corresponding P-wave velocities using the soft-sand model. This model uses the lower Hashin-Shtrikman bound to construct velocity-porosity trends for sand mixing with a specified clay volume (Avseth et al., 2005). Gassmann’s equation (Gassmann, 1951) is used to compute properties of rocks saturated with pore-fluid mixtures of water and oil. Density assigned to each layer is simply a weighted average of densities of all components in that particular layer. With the rock-physics models, velocity and density values are now assigned to each sedimentary layer in the sequences. We introduce uncertainties by assuming that each lithologic state has a distribution of velocities with a mean equal to the calculated velocity and a standard deviation of 0.1 km/s.

Layer properties (i.e., velocity, density and thickness) are then input into the Kennett algorithm (Kennett, 1983) to simulate full-waveform, normally-incident, reflected seismograms using a zero-phase Ricker wavelet with a central frequency of 30 Hz. The simulations are performed for all reverberations. A summary workflow from a transition matrix to multiple realizations of seismograms is shown in Figure 4.
Figure 3: Illustrations of sand-shale mixtures, with their porosity and velocity values related to clay content (Modified after Marion, 1990). Porosity versus clay content shows a V-shaped trend, where the two end points are the pure sand and pure shale porosity. Selected clay fractions corresponding to the four lithologic states are marked. The states are sand (s), shaly-sand (sh-s), sandy-shale (s-sh), and shale (sh).

Figure 4: Workflow for generating seismic responses of multiple realizations of thin sand-shale sequences. The four lithologic states in the transition matrix are sand (s), shaly-sand (sh-s), sandy-shale (s-sh), and shale (sh).
Seismic attributes

In our study of seismic signatures for thin sand-shale sequences, we extract attributes from seismograms and relate these attributes to net-to-gross ratios and water saturations. Even though the thin sand-shale layers are below seismic resolution, seismograms from sequences with similar underlying geology and properties potentially share more statistical similarities than those from sequences with different properties. Seismograms of length $n$ can be considered as $n$-dimensional vectors in the space $\mathbb{R}^n$. To compare the simulated seismograms, we obtain new representations of these seismograms in a lower dimensional space using multidimensional scaling (MDS) and kernel principal component analysis (KPCA).

Multidimensional scaling (MDS)

To compare $p$ seismograms of length $n$ using MDS, we first construct a $p$-by-$p$ dissimilarity matrix, whose element at the $i^{th}$ row and $j^{th}$ column ($\delta_{ij}$) is a dissimilarity measurement between the $i^{th}$ and $j^{th}$ seismograms. The dissimilarity values can be assigned from subjective judgment or objective measure (Cox and Cox, 2001). For seismogram comparison, dissimilarity need not always be the Euclidean distance between any two seismograms; we can define dissimilarity in many different ways. For example, we apply wavelet transform to seismograms and obtain wavelet coefficients for specified scales. Dissimilarity can be defined as pairwise Euclidean distance between their modulus of wavelet coefficients at selected scales (Figure 5).

![Figure 5: Modulus of wavelet coefficients (shown in the middle two boxes) of two seismogram segments. Dissimilarity between these seismograms is defined as distance between their modulus maps.](image)

To obtain attributes, this dissimilarity matrix is then input into an MDS algorithm which transforms the matrix into points in lower-dimensional Euclidean space. Points corresponding to their original objects (i.e., seismograms) are configured so that the inter-point Euclidean distances ($d_{ij}$) match the input dissimilarity values ($\delta_{ij}$) as much as possible. MDS is classified into various categories: classical (metric), metric, and non-metric MDS. While classical MDS solves an eigendecomposition problem for coordinates of points in lower-dimensional space,
both metric and non-metric MDS use iterative optimization techniques. Metric MDS minimizes difference between $d_{ij}$ and $f(\delta_{ij})$, where $f$ is a continuous monotonic function. Examples of such functions include identity transformation, logarithmic transformation, etc (Cox and Cox, 2001; Zhang, 2008). Non-metric MDS aims to find configuration of points so that the rank order of the dissimilarity matrix is preserved (Cox and Cox, 2001; Webb, 2002). The main result of MDS algorithms is the relative positions of points in lower-dimensional space; thus, absolute positions do not matter (Scheidt and Caers, 2009). We use these new coordinates obtained by MDS as seismic attributes for reservoir characterization.

**Kernel principal component analysis (KPCA)**

The other technique used for extracting seismic attributes is KPCA, which also provides new representations of seismograms in lower-dimensional space. We first briefly review the concept of linear principal component analysis (PCA). PCA is commonly used for dimensionality reduction and linear feature extraction. This method performs eigenvalue decomposition of a data covariance matrix to compute a set of orthogonal directions (principal components) which capture maximum variability of the data. The projection of each data point onto selected principal components creates a new representation of each point in a low-dimensional space. While PCA does not work well with non-linear features, kernel principal component analysis (KPCA), introduced by Schölkopf et al. (1997), provides an alternative nonlinear PCA method. The idea behind KPCA is to map input data nonlinearly into some high dimensional feature space, in which the data become linearly separable, and then linear PCA is performed (Scheidt and Caers, 2009). Instead of solving an eigenvalue problem of the covariance matrix, linear PCA in the feature space is done by solving an equivalent system involving dot products in the feature space. These dot products can be obtained using kernel functions without an explicit mapping from the original space to the feature space. Results from KPCA are the projections of input data onto the principal components.

To obtain attributes, we apply Gaussian, dynamic similarity, inverse multi-quadric, and polynomial kernels (Table 1) to the input seismograms and obtain kernel matrices. Then after centering these matrices, we compute eigenvalues and eigenvectors of the matrices, and obtain coordinates of points, each of which corresponds to a seismogram that is projected onto selected principal components. Hereafter we refer to KPCA with Gaussian, dynamic similarity, inverse multi-quadric, or polynomial kernels as Gaussian, dynamic similarity, inverse multi-quadric, or polynomial KPCA, respectively.
Table 1: Kernel functions used for extracting seismic attributes. Note that $x_i$ is the $i^{th}$ seismogram.

<table>
<thead>
<tr>
<th>Kernel function</th>
<th>Kernel matrix</th>
<th>Note</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gaussian kernel</td>
<td>$k(x_i, x_j) = e^{-\frac{|x_i - x_j|^2}{2\sigma^2}}$</td>
<td>$\sigma &gt; 0$, $|x_i - x_j|$ is the Euclidean distance between the $i^{th}$ and $j^{th}$ seismograms</td>
</tr>
<tr>
<td>Dynamic similarity kernel (Yan et al., 2006)</td>
<td>$k(x_i, x_j) = e^{-\frac{\text{DPF}(x_i, x_j)}{\sigma^2}}$</td>
<td>Dynamic partial function (DPF) by Li et al. (2003)</td>
</tr>
<tr>
<td>Inverse multi-quadric kernel</td>
<td>$k(x_i, x_j) = \frac{1}{\sqrt{c^2 + |x_i - x_j|^2}}$</td>
<td>$|x_i - x_j|$ is the Euclidean distance between the $i^{th}$ and $j^{th}$ seismograms</td>
</tr>
<tr>
<td>Polynomial kernel</td>
<td>$k(x_i, x_j) = (x_i \cdot x_j + c)^d$</td>
<td>$c \geq 0$ and $d$ is the degree of the polynomial ($d \in \mathbb{N}$)</td>
</tr>
</tbody>
</table>

Note: Performing KPCA with kernel functions that depend only on $\|x_i - x_j\|$ can be interpreted as solving a metric MDS as an eigenproblem (Williams, 2002).

The Euclidean distance ($\|x_i - x_j\|$) used in both the Gaussian kernel and the inverse multi-quadric kernel is a special case of a more general Minkowski metric, which is defined as

$$d(A, B) = \left(\sum_{k=1}^{n}|a_k - b_k|^r\right)^{\frac{1}{r}},$$ \hspace{1cm} (1)

where $d$ is the distance or similarity measure between objects $A$ and $B$, each of which is represented by a vector of length $n$ (e.g., $A = [a_1, a_2, a_3, ..., a_n]$). The Euclidean distance is the Minkowski metric with the parameter $r$ equal to 2. Using the Minkowski metric to measure similarity between two objects is to compare each and every element in the objects.

Based on the idea similar to human perceptual similarity in cognitive science, Li et al. (2003) defined a distance function called dynamic partial function (DPF) as

$$\text{DPF}(A, B) = \left(\sum_{k=1}^{m}|\delta_k \in \Delta_m \delta_k^r\right)^{\frac{1}{r}},$$ \hspace{1cm} (2)

where $\delta_k = |a_k - b_k|$ for $k = 1, ..., n$ and $\Delta_m$ is the set of the smallest $m$ $\delta$’s from $\{\delta_1, ..., \delta_n\}$. Thus parameter $m$ ranges from 1 to $n$. If $m = n$, Equation 2 is equivalent to the Minkowski metric.

Using DPF, Yan et al. (2006) proposed a new kernel for image classification and showed that this kernel yields higher classification accuracy than other kernels including Gaussian and polynomial kernels. As an extension of the Gaussian kernel, the new kernel is defined as

$$k(x_i, x_j) = e^{-\frac{\text{DPF}(x_i, x_j)}{\sigma^2}}.$$ \hspace{1cm} (3)
**Seismic signatures for 1-D Synthetic example**

In the previous section, we have described how we generate seismic responses and extract seismic attributes for thin sand-shale sequences. In this section, we illustrate 1-D synthetic examples exploring two scenarios to investigate effect of net-to-gross ratios and saturations on seismic signatures of thin sand-shale sequences. Note that we define the net-to-gross ratio as the proportion of the first lithologic state (sand) in the entire sequence.

**Scenario 1: Effect of net-to-gross ratios**

We study three transition matrices, which generate aggrading-type sequences (Table 2). Each transition matrix has a different limiting distribution ($\mathbf{\pi}$), a row vector of fixed probabilities representing proportions of each state in the long-term behavior. Thus in long sequences (i.e., large numbers of layers per sequence) the proportions of sand in the sequences should converge to a distinct value: $\mathbf{\pi}_{\text{sand}}$. In our simulations, all sequences consist of 200 sedimentary layers. We simulate 200 sequence realizations from each transition matrix and compute the true net-to-gross ratios of these sequences. Figure 6 shows that the computed net-to-gross values scatter around $\mathbf{\pi}_{\text{sand}}$ and that their distributions get narrower when the values of parameter $k$ in the transition matrices (Table 2) approach 1. Since the probabilities of going from sand (or shaly-sand) to shale and the probabilities of going from sandy-shale (or shale) to sand are equal to $k$, as $k$ gets larger the simulations will eventually turn into sequences with only sand and shale in alternating layers; as a result, the distributions of net-to-gross ratios become very narrow around the value of 0.5. Water saturation ($S_w$) is set to be the same in all simulations: $S_w=0.1$ for sand layers and $S_w=1$ for the other lithology states.

**Table 2:** Form of transition matrices for generating sequences used in investigating net-to-gross effects on seismic signatures. Values of parameter $k$ are 0.45, 0.75, and 0.95. Four lithologic states are sand (s), shaly-sand (sh-s), sandy-shale (s-sh), and shale (sh).

<table>
<thead>
<tr>
<th></th>
<th>Sand</th>
<th>Shaly-sand</th>
<th>Sandy-shale</th>
<th>Shale</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sand</td>
<td>(1-k)/3</td>
<td>(1-k)/3</td>
<td>(1-k)/3</td>
<td>k</td>
</tr>
<tr>
<td>Shaly-sand</td>
<td>(1-k)/3</td>
<td>(1-k)/3</td>
<td>(1-k)/3</td>
<td>k</td>
</tr>
<tr>
<td>Sandy-shale</td>
<td>k</td>
<td>(1-k)/3</td>
<td>(1-k)/3</td>
<td>(1-k)/3</td>
</tr>
<tr>
<td>Shale</td>
<td>k</td>
<td>(1-k)/3</td>
<td>(1-k)/3</td>
<td>(1-k)/3</td>
</tr>
</tbody>
</table>
Next, we run MDS and KPCA algorithms to extract seismic attributes, which simply are coordinates of points (i.e., seismograms) in the lower-dimensional space. To compare how well the attributes characterize net-to-gross ratios, we count the number of times the classification algorithm successfully sort points into their corresponding net-to-gross classes, which are defined to be less than 0.38, between 0.38 and 0.465, and greater than 0.465. The classification success rate is estimated using a stratified 10-fold cross validation. The validation process first partitions the attributes into 10 approximately equal-sized subsamples (i.e., folds) such that the proportions of each class in the subsamples are approximately equal to those in the whole sample (i.e., stratified). Then, a subsample is held as a validation set, while treating the rest as a training set in the classification step, where a success rate is obtained. This step is repeated 10 times so that each subsample is used once as a validation set. The overall success rate is simply an average of the 10 classification experiments. To improve the accuracy of the estimates, this stratified 10-fold cross validation is repeated 10 times. The 10 success rates are then averaged to obtain the final success rate (Witten et al., 2011).

Results from MDS and KPCA algorithms are shown in Figure 7 – Figure 13 as cross-plots between just two attributes, either the first two coordinates or the first two principal components. The classification results are summarized in Table 3. Note that three numbers are listed under the success rate column, representing the success rate when classifying samples into classes using only the first coordinate, the second coordinate, and both coordinates, respectively.

The metric and non-metric MDS yield better classification success rates (74% and 73% respectively) than the classical MDS (56%) when considering results in the polar coordinate system (i.e., theta and rho values; Figure 7 – Figure 9, lower right corner). Note that in the classical MDS, the first two coordinates correspond to the two largest eigenvalues. Among 2-
attribute results, there is not much difference in success rate between the metric and non-metric MDS. The success rate of each individual attribute clearly shows which attribute is more sensitive to net-to-gross changes. For example, the first principal components of the Gaussian, dynamic similarity, and inverse multi-quadric KPCA effectively capture variations in net-to-gross ratios. Using the first component alone, classification success rates of KPCA can be as high as 85%. When using only the first two principal components, the dynamic similarity KPCA best differentiates the three net-to-gross classes while the polynomial KPCA yields the poorest performance (Figure 10 – Figure 13).

The inter-point distances \( (d_{ij}) \) of MDS results do not reproduce the original dissimilarity \( (\delta_{ij}) \) of seismograms well if only two coordinates (i.e., 2-dimensional space) are considered. In this case, the correlation coefficients between the distances and dissimilarity for all three MDS algorithms are less than 0.9 (Figure 7 – Figure 9, lower left corner). However, the correlation increases as the number of coordinates increases. The classical MDS results have 10 large eigenvalues, which further indicates that more coordinates may need to be included in the results to better reproduce the dissimilarity matrix and improve classification performance.

The overall classification success rate increases as the number of coordinates included as attributes increases (Figure 14). However, using more coordinates involves higher-dimensional vectors; therefore, visualization of the multivariate results becomes more difficult. One way to visualize high-dimensional data is to use a parallel coordinate plot, where an \( n \)-dimensional vector is represented by a polyline connecting the value of each element in the vector. The horizontal and vertical axes of the plot are the non-negative integer line (e.g., from 1 to \( n \)) and the real number line, respectively. The polylines, each of which corresponds to one attribute vector (i.e., seismogram), can be color-coded by their net-to-gross classes. Instead of plotting individual polylines, the specified quantiles of each element can be plotted to better see the distributions of each class at a particular element. Figure 15 is an example of the parallel coordinate plots, using the first five components of the results from KPCA (Gaussian kernel). While the polylines do not provide much information on how each component is distributed in each class due to large overlaps, the plot using quantile values clearly shows that the three net-to-gross classes have quite distinct distributions of the first component. This observation supports the above mentioned KPCA result, which achieves a classification success rate of 85% by using only the first component. Note that 0.45 and 0.55- quantiles are selected only to get a better picture of how each component may contribute to the classification results. However, using 0.25- and 0.75- quantiles gives a better statistical summary of each component.
Figure 7: (Top left) Dissimilarity matrix showing pairwise Euclidean distances between any two seismograms of thin sand-shale sequences. (Bottom left) Correlation coefficient between dissimilarity and distance (between points in new coordinates resulting from classical MDS) versus numbers of included coordinates. (Top right) Classical MDS results with the first two coordinates. Each point represents one seismogram, color-coded by its true net-to-gross value from its corresponding sand-shale sequence. (Bottom right) Classical MDS results with the first two coordinates converted into polar coordinates.
Figure 8: (Top left) Dissimilarity matrix showing pairwise Euclidean distances between any two seismograms of thin sand-shale sequences. (Bottom left) Correlation coefficient between dissimilarity and distance (between points in new coordinates resulting from metric MDS) versus numbers of included coordinates. (Top right) Metric MDS results with the first two coordinates. Each point represents one seismogram, color-coded by its true net-to-gross value from its corresponding sand-shale sequence. (Bottom right) Metric MDS results with the first two coordinates converted into polar coordinates.
Figure 9: (Top left) Dissimilarity matrix showing pairwise Euclidean distances between any two seismograms of thin sand-shale sequences. (Bottom left) Correlation coefficient between dissimilarity and distance (between points in new coordinates resulting from non-metric MDS) versus numbers of included coordinates. (Top right) Non-metric MDS results with the first two coordinates. Each point represents one seismogram, color-coded by its true net-to-gross value from its corresponding sand-shale sequence. (Bottom right) Non-metric MDS results with the first two coordinates converted into polar coordinates.
Figure 10: (Left) Kernel matrix using a Gaussian kernel. Each element in the matrix \( K(x_i, x_j) \) corresponds to the Gaussian kernel function evaluated using a pair of seismograms \( (x_i, x_j) \). (Right) Projections of seismograms, which correspond to sequences with different net-to-gross ratios, onto the first two principal components from the Gaussian KPCA. Each point is color-coded by the net-to-gross ratio from its corresponding sand-shale sequence.

Figure 11: (Left) Kernel matrix using a dynamic similarity kernel. Each element in the matrix \( K(x_i, x_j) \) corresponds to the dynamic similarity kernel function evaluated using a pair of seismograms \( (x_i, x_j) \). (Right) Projections of seismograms, which correspond to sequences with different net-to-gross ratios, onto the first two principal components from the dynamic similarity KPCA. Each point is color-coded by the net-to-gross ratio from its corresponding sand-shale sequence.
Figure 12: (Left) Kernel matrix using an inverse multi-quadric kernel. Each element in the matrix \( K(x_i,x_j) \) corresponds to the inverse multi-quadric kernel function evaluated using a pair of seismograms \((x_i,x_j)\). (Right) Projections of seismograms, which correspond to sequences with different net-to-gross ratios, onto the first two principal components from the inverse multi-quadric KPCA. Each point is color-coded by the net-to-gross ratio from its corresponding sand-shale sequence.

Figure 13: (Left) Kernel matrix using a polynomial kernel. Each element in the matrix \( K(x_i,x_j) \) corresponds to the polynomial kernel function evaluated using a pair of seismograms \((x_i,x_j)\). (Right) Projections of seismograms, which correspond to sequences with different net-to-gross ratios, onto the first two principal components from the polynomial KPCA. Each point is color-coded by the net-to-gross ratio from its corresponding sand-shale sequence.
Table 3: Summary of the methods used to compute seismic attributes for net-to-gross estimation. Choices of parameters for each method are also included. Performance of each method is shown as a success rate in classifying a data point into three net-to-gross classes: <0.38, 0.38-0.465, and >0.465

<table>
<thead>
<tr>
<th>Method</th>
<th>Compared objects</th>
<th>Distance/ or Kernel</th>
<th>Classification success rate (using 1&lt;sup&gt;st&lt;/sup&gt;, 2&lt;sup&gt;nd&lt;/sup&gt;, both coordinates/components)</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>MDS (classical)</td>
<td>Seismograms</td>
<td>Euclidean</td>
<td>0.52, 0.49, 0.56 (theta/rho/both)</td>
<td>&gt; 10 large eigenvalues; the algorithm cannot reproduce dissimilarity well by using only the first two coordinates</td>
</tr>
<tr>
<td>MDS (metric)</td>
<td>Seismograms</td>
<td>Euclidean</td>
<td>0.47 0.69, 0.74 (theta/rho/both)</td>
<td>Correlation coefficient between distance and dissimilarity is 0.83 when using the first two coordinates; the value becomes greater than 0.95 when there are more than 4 coordinates.</td>
</tr>
<tr>
<td>MDS (non-metric)</td>
<td>Seismograms</td>
<td>Euclidean</td>
<td>0.48, 0.68, 0.73 (theta/rho/both)</td>
<td>Correlation coefficient between distance and dissimilarity is 0.88 when using the first two coordinates; the value becomes greater than 0.95 when there are more than 3 coordinates.</td>
</tr>
<tr>
<td>KPCA</td>
<td>Seismograms</td>
<td>Gaussian</td>
<td>0.81, 0.35, 0.81</td>
<td>( \sigma^2 = 0.5 )</td>
</tr>
<tr>
<td>KPCA</td>
<td>Seismograms</td>
<td>Dynamic similarity</td>
<td>0.85, 0.58, 0.90</td>
<td>( \sigma = 0.002, m = 80% ) of total seismogram length, ( r = 2 )</td>
</tr>
<tr>
<td>KPCA</td>
<td>Seismograms</td>
<td>Inverse multi-quadric</td>
<td>0.79, 0.35, 0.79</td>
<td>( c^* = 1 )</td>
</tr>
<tr>
<td>KPCA</td>
<td>Seismograms</td>
<td>Polynomial</td>
<td>0.50, 0.51, 0.59</td>
<td>( c = 1, d = 2 )</td>
</tr>
</tbody>
</table>
Scenario 2: Effect of saturations

We use sequences generated from a similar transition matrix, but with varying water saturation values (Sw). In the sand layers, three values of Sw are considered: 0.1, 0.5, and 1. All other lithologies always have Sw equal to 1. We consider three matrices (A, B, and C) which have the same limiting distribution: [0.45 0.05 0.05 0.45] (Figure 16). Long sequences generated using the three matrices should have their net-to-gross ratios within the sequence equal to 0.45;
however, in our simulation the values spread around 0.45. Therefore, to separate the net-to-gross effect from the saturation effect, we select only those sequences with net-to-gross ratios ranging from 0.4 – 0.5. Thus, the main difference among sequences generated from the three transition matrices is how the sand layers are distributed. For example, sand layers in sequences from matrix C are more clustered and blocky than those from the other two matrices. This blocky pattern is expected since matrix C has large probabilities of going to the sand and shale states ($p_{ss} = p_{sh,s} = p_{s,sh} = p_{shsh} = 0.85$).

Results from MDS and KPCA algorithms are color-coded by their corresponding saturation values (i.e., 0.1, 0.5, and 1) and shown in Figure 17 – Figure 23 as cross-plots between just two attributes, either the first two coordinates or the first two components. Results are then used in the same cross-validation scheme described in the previous section. The classification success rates are summarized in Table 4.

For the same transition matrix, plots of the first two coordinates from classical MDS and the two coordinates from metric and non-metric MDS in Figure 17 – Figure 19 show similar patterns, resulting in similar classification success rates with the rates of classical MDS being
slightly smaller than metric and non-metric MDS. In matrix A results, all MDS algorithms effectively separate out the sequences which have sand layers fully saturated with water (Sw = 1) from the other sequences relatively well. The other two saturation classes are almost inseparable because their distributions largely overlap. In matrix B results, the water-saturated sequences are still well separated by all MDS algorithms; however, in this case the saturation class of 0.1 starts to separate out and surround the saturation class of 0.5 with less overlaps than the results in matrix A. In matrix C, both the saturation classes of 0.1 and 1 surround the saturation class of 0.5. The overall classification success rates using two-coordinate MDS results are less than 70%.

Similar to the net-to-gross case, if only two coordinates are considered, the inter-point distances of MDS results do not reproduce the original dissimilarity of seismograms well, with the correlation coefficients between the distances and dissimilarity for metric and non-metric MDS algorithms being less than 0.9 (Figure 25). The correlation increases as the number of coordinates increases, with correlations of matrix C being the least for the same number of coordinates. For classical MDS results, there are more than two large eigenvalues, especially the eigenvalues in matrix C whose values are almost twice as large as those in matrix A and B (Figure 24). More coordinates can be included as attributes to better reproduce the dissimilarity matrix and improve classification performance. When the number of coordinates increases, the overall classification success rate also increases. For the same transition matrix and same number of coordinates, the success rates from classical, metric, and non-metric MDS turn out to be very similar (Figure 26).

By using the first two principal components, the dynamic similarity KPCA best differentiate the three saturation classes for all transition matrices (Figure 21). Its classification success rates are greater than 80% while the other kernels have relatively poorer performance. From Figure 20, the first principal component of the Gaussian KPCA for both the transition matrices A and B separate out the sequences which have sand layers fully saturated with water (Sw = 1) from the other sequences relatively well. Matrix C shows a similar result, except that the class which is effectively separated out by using the first component is that with Sw equal to 0.5. For all three transition matrices, the second component does not contribute much in differentiating the three saturation classes, and the overall classification success rates of the Gaussian KPCA are only in the 60-70 percent range.

Both the inverse multi-quadratic KPCA in Figure 22 and the polynomial KPCA in Figure 23 show patterns quite similar to the Gaussian KPCA, which the water-saturated sequences (Sw = 1) seem distinguishable from the other classes by just the first component for both matrices A and
B; however, in matrix C this saturation class (Sw = 1) is less separable from the rest. The classification success rates using the first two principal components of the inverse multi-quadratic KPCA are generally in the 60-70 percent range, which is slightly higher than the 50-65 percent range of the polynomial KPCA.

Two ways to improve the performance of KPCA attributes are to either increase the numbers of components included as attributes (e.g., from two to ten), or use parallel coordinate plots and select the components that can effectively distinguish different classes of saturation. By selecting only one to three components, visualization of the results becomes possible; moreover, with a small number of components the classification success rate can sometimes be very close to the success rate obtained when using a larger numbers of components. For example, the success rate when using the first and sixth components of Gaussian KPCA results for matrix B is 80%, while the success rate when using the first ten components is 82% (Figure 27). As another example, the success rate when using the first and sixth components of Inverse multi-quadratic KPCA results for matrix C is 73%, while the success rate when using the first ten components is 82% (Figure 28). Here our selection process is heuristic-based, selecting any two components that exhibit large separation among the three classes. However, this selection scheme does not necessarily work in all cases, especially when several components behave similarly. Therefore, a systematic comparison of success rates may be needed in order to select the best possible combination of components.

Figure 17: Projections of seismograms, which correspond to sequences with different water-saturation values, onto the first two coordinates from classical MDS. Each point is color-coded by the water-saturation value of the sand layers. From left to right, subplots correspond to transition matrices A, B, and C, respectively.
Figure 18: Projections of seismograms, which correspond to sequences with different water-saturation values, onto the two coordinates from metric MDS. Each point is color-coded by the water-saturation value of the sand layers. From left to right, subplots correspond to transition matrices A, B, and C, respectively.

Figure 19: Projections of seismograms, which correspond to sequences with different water-saturation values, onto the two coordinates from non-metric MDS. Each point is color-coded by the water-saturation value of the sand layers. From left to right, subplots correspond to transition matrices A, B, and C, respectively.

Figure 20: Projections of seismograms, which correspond to sequences with different water-saturation values, onto the first two principal components from the Gaussian KPCA. Each point is color-coded by the water-saturation value of the sand layers. From left to right, subplots correspond to transition matrices A, B, and C, respectively.
Figure 21: Projections of seismograms, which correspond to sequences with different water-saturation values, onto the first two principal components from the dynamic similarity KPCA. Each point is color-coded by the water-saturation value of the sand layers. From left to right, subplots correspond to transition matrices A, B, and C, respectively.

Figure 22: Projections of seismograms, which correspond to sequences with different water-saturation values, onto the first two principal components from the inverse multi-quadric KPCA. Each point is color-coded by the water-saturation value of the sand layers. From left to right, subplots correspond to transition matrices A, B, and C, respectively.

Figure 23: Projections of seismograms, which correspond to sequences with different water-saturation values, onto the first two principal components from the polynomial KPCA. Each point is color-coded by the water-saturation value of the sand layers. From left to right, subplots correspond to transition matrices A, B, and C, respectively.
Figure 24: First 21 eigenvalues from classical MDS, color-coded by the transition matrices used in generating sand-shale sequences for investigating the effect of saturations.

Figure 25: Correlation coefficient between dissimilarity and distance (between points in new coordinates resulting from an MDS algorithm) versus numbers of included coordinates from MDS results for investigating the effect of saturations. (Left) metric MDS and (right) non-metric MDS.

Figure 26: Change in classification success rate when the number of coordinates included as saturation attributes increases. Three MDS algorithms are compared. Results when using sequences generated by transition matrices A (left), B (middle) and C (right) are shown.
Figure 27: (Left) Parallel coordinate plot of the first ten principal components of the Gaussian KPCA for matrix B. The solid lines are the median (the 0.5 quantile). The dash lines surrounding the median are the 0.45 and 0.55 quantiles. (Right) Projections of seismograms onto the first and sixth principal components of the Gaussian KPCA.

Figure 28: (Left) Parallel coordinate plot of the first ten principal components of the inverse multi-quadric KPCA for matrix C. The solid lines are the median (the 0.5 quantile). The dash lines surrounding the median are the 0.45 and 0.55 quantiles. (Right) Projections of seismograms onto the first and sixth principal components of the inverse multi-quadric KPCA.
Table 4: Summary of the methods used to compute seismic attributes for water-saturation (in sand layers) effect. Choices of parameters for each method are similar to those listed in Table 3, unless otherwise specified. Performance of each method is shown as a success rate in classifying a data point into three water-saturation categories: Sw=0.1, 0.5, and 1. Results are shown in the columns corresponding to the transition matrices used in simulations. The actual transition matrices and sample sequences are shown in Figure 16.

<table>
<thead>
<tr>
<th>Method</th>
<th>Classification success rate (using $1^{st}$, $2^{nd}$, both coordinates/components)</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Transition matrix A (Aggrading)</td>
<td>Transition matrix B</td>
</tr>
<tr>
<td>MDS (classical)</td>
<td>0.66, 0.33, 0.65</td>
<td>0.66, 0.33, 0.66</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MDS (metric)</td>
<td>0.65, 0.36, 0.65</td>
<td>0.67, 0.31, 0.67</td>
</tr>
<tr>
<td>MDS (non-metric)</td>
<td>0.66, 0.35, 0.66</td>
<td>0.68, 0.33, 0.69</td>
</tr>
<tr>
<td>KPCA (Gaussian)</td>
<td>0.64, 0.32, 0.62 (1&amp;3 0.73 (1to10 0.80)</td>
<td>0.732, 0.35, 0.729</td>
</tr>
<tr>
<td></td>
<td>1to10 0.80)</td>
<td>(1&amp;6 0.80 (1to10 0.82)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KPCA (Dynamic similarity)</td>
<td>0.71, 0.76, 0.88 (1to10 0.89)</td>
<td>0.70, 0.73, 0.87 (1to10 0.89)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KPCA (Inverse multi-quadric)</td>
<td>0.65, 0.32, 0.64 (1&amp;3 0.71 (1to10 0.81)</td>
<td>0.68, 0.34, 0.67 (1&amp;7 0.78 (1to10 0.82)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KPCA (Polynomial)</td>
<td>0.68, 0.32, 0.65 (1&amp;7 0.79 (1to10 0.80)</td>
<td>0.66, 0.35, 0.65 (1&amp;8 0.76 (1to10 0.77)</td>
</tr>
</tbody>
</table>

Discussion and conclusion

The layer thickness in our simulations has the wavelength to thickness ratio of about 100. Seismic waves cannot distinguish the boundaries of these thin layers because of their band-limited characteristics. Hence, interpretations of these sub-resolution layers can only be based on statistical attributes of the seismograms. In this study, we extract attributes from seismograms of thin shaly-sand reservoirs using multidimensional scaling (MDS) and kernel principal component analysis (KPCA). Results show that these attributes can differentiate
sequences with different net-to-gross ratios or different water saturations. A workflow similar to our study using MDS and KPCA can be applied to real seismic data to characterize thin shaly-sand reservoirs. In real application, all parameters should be calibrated at the well locations. By assuming that the stratigraphy in the explored area demonstrates a lateral continuation within conformable sequences, inferred transition matrix from a calibration well could be used to explore statistically how the seismic attributes would change with varying net-to-gross ratios and saturations. These statistics of the attributes can then be applied to observations away from the well to characterize the area and quantify the uncertainties.

**Acknowledgements**

This work was supported by the Stanford Rock Physics and Borehole Geophysics project and the Stanford Center for Reservoir Forecasting.

**References**


Marion, D., Mukerji, T. and Mavko, G., 1994. Scale effects on velocity dispersion; from ray to effective medium theories in stratified media: Geophysics, 59, no. 10, 1613–1619


