Geologic pattern recognition from seismic attributes: Principal component analysis and self-organizing maps

Rocky Roden¹, Thomas Smith², and Deborah Sacrey³

Abstract

Interpretation of seismic reflection data routinely involves powerful multiple-central-processing-unit computers, advanced visualization techniques, and generation of numerous seismic data types and attributes. Even with these technologies at the disposal of interpreters, there are additional techniques to derive even more useful information from our data. Over the last few years, there have been efforts to distill numerous seismic attributes into volumes that are easily evaluated for their geologic significance and improved seismic interpretation. Seismic attributes are any measurable property of seismic data. Commonly used categories of seismic attributes include instantaneous, geometric, amplitude accentuating, amplitude-variation with offset, spectral decomposition, and inversion. Principal component analysis (PCA), a linear quantitative technique, has proven to be an excellent approach for use in understanding which seismic attributes or combination of seismic attributes has interpretive significance. The PCA reduces a large set of seismic attributes to indicate variations in the data, which often relate to geologic features of interest. PCA, as a tool used in an interpretation workflow, can help to determine meaningful seismic attributes. In turn, these attributes are input to self-organizing-map (SOM) training. The SOM, a form of unsupervised neural networks, has proven to take many of these seismic attributes and produce meaningful and easily interpretable results. SOM analysis reveals the natural clustering and patterns in data and has been beneficial in defining stratigraphy, seismic facies, direct hydrocarbon indicator features, and aspects of shale plays, such as fault/fracture trends and sweet spots. With modern visualization capabilities and the application of 2D color maps, SOM routinely identifies meaningful geologic patterns. Recent work using SOM and PCA has revealed geologic features that were not previously identified or easily interpreted from the seismic data. The ultimate goal in this multiattribute analysis is to enable the geoscientist to produce a more accurate interpretation and reduce exploration and development risk.

Introduction

The object of seismic interpretation is to extract all the geologic information possible from the data as it relates to structure, stratigraphy, rock properties, and perhaps reservoir fluid changes in space and time (Liner, 1999). Over the past two decades, the industry has seen significant advancements in interpretation capabilities, strongly driven by increased computer power and associated visualization technology. Advanced picking and tracking algorithms for horizons and faults, integration of prestack and poststack seismic data, detailed mapping capabilities, integration of well data, development of geologic models, seismic analysis and fluid modeling, and generation of seismic attributes are all part of the seismic interpreter's toolkit.

What is the next advancement in seismic interpretation?

A significant issue in today's interpretation environment is the enormous amount of data that is used and generated in and for our workstations. Seismic gathers, regional 3D surveys with numerous processing versions, large populations of wells and associated data, and dozens if not hundreds of seismic attributes, routinely produce quantities of data in terms of terabytes. The ability for the interpreter to make meaningful interpretations from these huge projects can be difficult and at times quite inefficient. Is the next step in the advancement of interpretation the ability to interpret large quantities of seismic data more effectively and potentially derive more meaningful information from the data?

¹Rocky Ridge Resources, Inc., Centerville, Houston, USA. E-mail: rodenr@wt.net.

²Geophysical Insights, Houston, Texas, USA. E-mail: tsmith@geoinsights.com.

³Auburn Energy, Houston, Texas, USA. E-mail: dsacrey@auburnenergy.com.

Manuscript received by the Editor 5 February 2015; revised manuscript received 15 April 2015; published online 14 August 2015. This paper appears in *Interpretation*, Vol. 3, No. 4 (November 2015); p. SAE59–SAE83, 18 FIGS., 4 TABLES.

<http://dx.doi.org/10.1190/INT-2015-0037.1>. © The Authors. Published by the Society of Exploration Geophysicists and the American Association of Petroleum Geologists. All article content, except where otherwise noted (including republished material), is licensed under a Creative Commons Attribution 4.0 Unported License (CC BY-NC-ND). See <http://creativecommons.org/licenses/by/4.0/>. Distribution or reproduction of this work in whole or in part requires full attribution of the original publication, including its digital object identifier (DOI). Commercial reuse and derivatives are not permitted.

For example, is there a more efficient methodology to analyze prestack data whether interpreting gathers, offset/angle stacks, or amplitude-variation with offset (AVO) attributes? Can the numerous volumes of data produced by spectral decomposition be efficiently analyzed to determine which frequencies contain the most meaningful information? Is it possible to derive more geologic information from the numerous seismic attributes generated by interpreters by evaluating numerous attributes all at once and not each one individually?

This paper describes the methodologies to analyze combinations of seismic attributes of any kind for meaningful patterns that correspond to geologic features. Principal component analysis (PCA) and self-organizing maps (SOMs) provide multiattribute analyses that have proven to be an excellent pattern recognition approach in the seismic interpretation workflow. A *seismic attribute* is any measurable property of seismic data, such as amplitude, dip, phase, frequency, and polarity that can be measured at one instant in time/depth over a time/depth window, on a single trace, on a set of traces, or on a surface interpreted from the seismic data (Schlumberger Oil Field Dictionary, 2015). Seismic attributes reveal features, relationships, and patterns in the seismic data that otherwise might not be noticed (Chopra and Marfurt, 2007). Therefore, it is only logical to deduce that a multiattribute approach with the proper input parameters can produce even more meaningful results and help to reduce risk in prospects and projects.

Evolution of seismic attributes

Balch (1971) and Anstey at Seiscom-Delta in the early 1970s are credited with producing some of the first generation of seismic attributes and stimulated the industry to rethink standard methodology when these results were presented in color. Further development was advanced with the publications by Taner and Sheriff

(1977) and Taner et al. (1979) who present complex trace attributes to display aspects of seismic data in color not seen before, at least in the interpretation community. The primary complex trace attributes including reflection strength/envelope, instantaneous phase, and instantaneous frequency inspired several generations of new seismic attributes that evolved as our visualization and computer power improved. Since the 1970s, there has been an explosion of seismic attributes to such an extent that there is not a standard approach to categorize these attributes. Brown (1996) categorizes seismic attributes by time, amplitude, frequency, and attenuation in prestack and poststack modes. Chen and Sidney (1997) categorize seismic attributes by wave kinematics/dynamics and by reservoir features. Taner (2003) further categorizes seismic attributes by prestack and by poststack, which is further divided into instantaneous, physical, geometric, wavelet, reflective, and transmissive. Table 1 is a composite list of seismic attributes and associated categories routinely used in seismic interpretation today. There are of course many more seismic attributes and combinations of seismic attributes than listed in Table 1, but as Barnes (2006) suggests, if you do not know what an attribute means or is used for, discard it. Barnes (2006) prefers attributes with geologic or geophysical significance and avoids attributes with purely mathematical meaning. In a similar vein, Kalkomey (1997) indicates that when correlating well control with seismic attributes, there is a high probability of spurious correlations if the well measurements are small or the number of independent seismic attributes is considered large. Therefore, the recommendation when the well correlation is small is that only those seismic attributes that have a physically justifiable relationship with the reservoir property be considered as candidates for predictors.

In an effort to improve the interpretation of seismic attributes, interpreters began to coblend two and three

Table 1. Seismic attribute categories and corresponding types and interpretive uses.

CATEGORY	TYPE	INTERPRETIVE USE
Instantaneous Attributes	Reflection Strength, Instantaneous Phase, Instantaneous Frequency, Quadrature, Instantaneous Q	Lithology Contrasts, Bedding Continuity, Porosity, DHIs, Stratigraphy, Thickness
Geometric Attributes	Semblance and Eigen-Based Coherency/Similarity, Curvature (Maximum, Minimum, Most Positive, Most Negative, Strike, Dip)	Faults, Fractures, Folds, Anisotropy, Regional Stress Fields
Amplitude Accentuating Attributes	RMS Amplitude, Relative Acoustic Impedance, Sweetness, Average Energy	Porosity, Stratigraphic and Lithologic Variations, DHIs
AVO Attributes	Intercept, Gradient, Intercept/Gradient Derivatives, Fluid Factor, Lambda-Mu-Rho, Far-Near, (Far-Near)Far	Pore fluid, Lithology, DHIs
Seismic Inversion Attributes	Colored inversion, Sparse Spike, Elastic Impedance, Extended Elastic Impedance, Prestack Simultaneous Inversion, Stochastic Inversion	Lithology, Porosity, Fluid Effects
Spectral Decomposition	Continuous Wavelet Transform, Matching Pursuit, Exponential Pursuit	Layer Thicknesses, Stratigraphic Variations

attributes together to better visualize features of interest. Even the generation of attributes on attributes has been used. Abele and Roden (2012) describe an example of this where dip of maximum similarity, a type of coherency, was generated for two spectral decomposition volumes high and low bands, which displayed high energy at certain frequencies in the Eagle Ford Shale interval of South Texas. The similarity results at the Eagle Ford from the high-frequency data showed more detail of fault and fracture trends than the similarity volume of the full-frequency data. Even the low-frequency similarity results displayed better regional trends than the original full-frequency data. From the evolution of ever more seismic attributes that multiply the information to interpret, we investigate PCA and self-organizing maps to derive more useful information from multiattribute data in the search for oil and gas.

Principal component analysis

PCA is a linear mathematical technique used to reduce a large set of seismic attributes to a small set that still contains most of the variation in the large set. In other words, PCA is a good approach to identify the combination of most meaningful seismic attributes generated from an original volume. The first principal component accounts for as much of the variability in the data as possible, and each succeeding component (orthogonal to

each preceding) accounts for as much of the remaining variability (Guo et al., 2009; Haykin, 2009). Given a set of seismic attributes generated from the same original volume, PCA can identify the attributes producing the largest variability in the data suggesting these combinations of attributes will better identify specific geologic features of interest. Even though the first principal component represents the largest linear attribute combinations that best represents the variability of the bulk of the data, it may not identify specific features of interest to the interpreter. The interpreter should also evaluate succeeding principal components because they may be associated with other important aspects of the data and geologic features not identified with the first principal component. In other words, PCA is a tool that, used in an interpretation workflow, can give direction to meaningful seismic attributes and improve interpretation results. It is logical, therefore, that a PCA evaluation may provide important information on appropriate seismic attributes to take into an SOM generation.

Natural clusters

Several challenges and potential benefits of multiple attributes for interpretation are illustrated in Figure 1. Geologic features are revealed through attributes as coherent energy. When there is more than one attribute, we call these centers of coherent energy *natural clusters*.

Figure 1. Natural clusters are illustrated in four situations with sets of 1000-sample Gaussian distributions shown in blue. From PCA, the principal components are drawn in red from an origin marked at the mean data values in x and y . The first principal components point in the direction of maximum variability. The first eigenvalues, as variance in these directions, are (a) 4.72, (b) 4.90, (c) 4.70, and (d) 0.47. The second principal components are perpendicular to the first and have respective eigenvalues of (a) 0.22, (b) 1.34, (c) 2.35, and (d) 0.26.

ters. Identification and isolation of natural clusters are important parts of interpretation when working with multiple attributes. In Figure 1, we illustrate natural clusters in blue with 1000 samples of a Gaussian distribution in 2D. Figure 1a shows two natural clusters that project to either attribute scale (horizontal or vertical), so they would both be clearly visible in the seismic data, given a sufficient signal-to-noise ratio. Figure 1b shows four natural clusters that project to both attribute axes, but in this case, the horizontal attribute would see three clusters and would separate the left and right natural clusters clearly. The six natural clusters would be difficult to isolate in Figure 1c, and even the two natural clusters in Figure 1d, while clearly separated, have a special challenge. In Figure 1d, the right natural cluster is clearly separated and details could be resolved by attribute value. However, the left natural cluster would be separated by the attribute on the vertical axis and resolution cannot be as accurate because the higher values are obscured by projection from the right natural cluster. Based on the different 2D cluster examples in Figure 1a–1d, PCA has determined the largest variation in each example labeled one on the red lines and represents the first principal component. The second principal component is labeled two on the red lines.

Overview of principal component analysis

We illustrate PCA in 2D with the natural clusters of Figure 1, although the concepts extend to any dimension. Each dimension counts as an attribute. We first consider the two natural clusters in Figure 1a and find the centroid (average x and y points). We draw an axis through the point at some angle and project all the attribute points to the axis. Different angles will result in a different distribution of points on the axis. For an angle, there is a variance for all the distances from the centroid. PCA finds the angle that maximizes that variance (labeled one on red line). This direction is the first principal component, and this is called the *first eigenvector*. The value of the variance is the first eigenvalue. Interestingly, there is an error for each data point that projects on the axis, and mathematically, the first principal component is also a least-squares fit to the data. If we subtract the least-squares fit, reducing the dimension by one, the second principal component (labeled two on red line) and second eigenvalue fit the residual.

In Figure 1a, the first principal component passes through the centers of the natural clusters and the projection distribution is spread along the axis (the eigenvalue is 4.72). The first principal component is nearly equal parts of attribute one (50%) and attribute two (50%) because it lies nearly along a 45° line. We judge that both attributes are important and worth consideration. The second principal component is perpendicular to the first, and the projections are restricted (eigenvalue is 0.22). Because the second principal component is so much smaller than the first (5%), we discard it. It is clear that the PCA alone reveals the importance of both attributes.

In Figure 1b, the first principal component eigenvector is nearly horizontal. The x -component is 0.978, and the y -component is 0.208, so the first principal eigenvector is composed of 82% attribute one (horizontal axis) and 18% attribute two (vertical axis). The second component is 27% smaller than the first and is significant. In Figure 1c and 1d, the first principal components are also nearly horizontal with component mixes of 81%, 12% and 86%, 14%, respectively. The second components were 50% and 55%, respectively. We demonstrate PCA on these natural cluster models to illustrate that it is a valuable tool to evaluate the relative importance of attributes, although our data typically have many more natural clusters than attributes, and we must resort to automatic tools, such as SOM to hunt for natural clusters after a suitable suite of attributes has been selected.

Survey and attribute spaces

For this discussion, seismic data are represented by a 3D seismic survey data volume regularly sampled in location x or y and in time t or in depth Z . A 2D seismic line is treated as a 3D survey of one line. Each survey is represented by several attributes, $f_1, f_2 \dots f_F$. For example, the attributes might include the amplitude, Hilbert transform, envelope, phase, frequency, etc. As such, an individual sample $\mathbf{x}$ is represented in bold as an attribute vector of F -dimensions in survey space:

$$\mathbf{x} \in \mathbf{X} = \{ \mathbf{x}_{c,d,e,f} \}, \quad (1)$$

where $\in$ reads “is a member of” and $\{ \dots \}$ is a set. Indices c, d, e , and f are indices of time or depth, trace, line number, and attribute, respectively. A sample drawn from the survey space with c, d , and e indices is a vector of attributes in a space $\mathbb{R}^F$. It is important to note for later use that $\mathbf{x}$ does not change position in attribute space. The samples in a 3D survey drawn from $\mathbf{X}$ may lie in a fixed time or depth interval, a fixed interval offset from a horizon, or a variable interval between a pair of horizons. Moreover, samples may be restricted to a specific geographic area.

Normalization and covariance matrix

PCA starts with computation of the covariance matrix, which estimates the statistical correlation between pairs of attributes. Because the number range of an attribute is unconstrained, the mean and standard deviation of each attribute are computed and corresponding attribute samples are normalized by these two constants. In statistical calculations, these normalized attribute values are known as *standard scores* or *Z-scores*. We note that the mean and standard deviation are often associated with a Gaussian distribution, but here we make no such assumption because it does not underlie all attributes. The phase attribute, for example, is often uniformly distributed across all angles, and envelopes are often lognormally distributed across amplitude values. However, standardization is a way to

assure that all attributes are treated equally in the analyses to follow. Letting T stand for transpose, a multiattribute sample on a single time or depth trace is represented as a column vector of F attributes:

$$\mathbf{x}_i^T = [x_{i,1} \quad \dots \quad x_{i,F}], \quad (2)$$

where each component is a standardized attribute value and where the selected samples in the PCA analysis range from $i = 1$ to I . The covariance matrix is estimated by summing the dot product of pairs of multiattribute samples over all I samples selected for the PCA analysis. That is, the covariance matrix

$$C = \frac{1}{I} \begin{bmatrix} \sum_{i=1}^I \mathbf{x}_{i,1} \cdot \mathbf{x}_{i,1} & \dots & \sum_{i=1}^I \mathbf{x}_{i,1} \cdot \mathbf{x}_{i,F} \\ \vdots & \ddots & \vdots \\ \sum_{i=1}^I \mathbf{x}_{i,1} \cdot \mathbf{x}_{i,F} & \dots & \sum_{i=1}^I \mathbf{x}_{i,F} \cdot \mathbf{x}_{i,F} \end{bmatrix}, \quad (3)$$

where the dot product is the matrix sum of the product $\mathbf{x} \cdot \mathbf{x} = \mathbf{x}^T \mathbf{x}$. The covariance matrix is symmetric, semipositive definite, and of dimension $F \times F$.

Eigenvalues and eigenvectors

The PCA proceeds by computing the set of eigenvalues and eigenvectors for the covariance matrix. That is, for the covariance matrix, there are a set of F eigenvalues λ and eigenvectors $\mathbf{v}$, which satisfy

$$C \mathbf{v} = \lambda \mathbf{v}. \quad (4)$$

This is a well-posed problem for which there are many stable numerical algorithms. For small F (≤ 10), the Jacobi method of diagonalization is convenient. For larger matrices, Householder transforms are used to reduce it to tridiagonal form, and then QR/QL deflation where Q, R, and L refer to parts of any matrix. Note that an eigenvector and eigenvalue are a matched pair. Eigenvectors are all orthogonal to each other and orthonormal when they are each of unit length. Mathematically, eigenvectors $\mathbf{v}_i \times \mathbf{v}_j = 0$ for $i \neq j$ and $\mathbf{v}_i \times \mathbf{v}_j = 1$ for $i = j$. The algorithms mentioned above compute orthonormal eigenvectors.

The list of eigenvalues is inspected to better understand how many attributes are important. The eigenvalue list that is sorted in decreasing order will be called the *eigenspectrum*. We adopt the notation that the pair of eigenvalue and eigenvectors with the largest eigenvalue is $\{\lambda_1 \mathbf{v}_1\}$, and that the pair with the smallest eigenvalue is $\{\lambda_F \mathbf{v}_F\}$. A plot of the eigenspectrum is drawn with a horizontal axis numbered one through F from left to right and a vertical axis that is increasing eigenvalue. For a multiattribute seismic survey, a plot of the corresponding eigenspectrum is often shaped like a decreasing exponential function. See Figure 3. The point where the eigenspectrum generally flattens is particularly important. To the right of this point, additional eigenvalues are insignificant. Inspection of the eigenspectrum constitutes the first and often the most important step in PCA (Figure 3b and 3c).

Unfortunately, eigenvalues reveal nothing about which attributes are important. On the other hand, simple identification of the number of attributes that are important is of considerable value. If L of F attributes are important, then $F-L$ attributes are unimportant. Now, in general, seismic samples lie in an attribute space $\mathbb{R}^F$, but the PCA indicates that the data actually occupy a smaller space $\mathbb{R}^L$. The space $\mathbb{R}^{F-L}$ is just noise.

The second step is to inspect eigenvectors. We proceed by picking the eigenvector corresponding to the largest eigenvalue $\{\lambda_1 \mathbf{v}_1\}$. This eigenvector, as a linear combination of attributes, points in the direction of maximum variance. The coefficients of the attribute components reveal the relative importance of the attributes. For example, suppose that there are four attributes of which two components are nearly zero and two are of equal value. We will conclude that for this eigenvector, we can identify two attributes that are important and two that are not. We find that a review of the eigenvectors for the first few eigenvalues of the eigenspectrum reveal those attributes that are important in understanding the data (Figure 3b and 3c). Often the attributes of importance in this second step match the number of significant attributes estimated in the first step.

Self-organizing maps

The self-organizing map (SOM) is a data visualization technique invented in 1982 by Kohonen (2001). This nonlinear approach reduces the dimensions of data through the use of unsupervised neural networks. SOM attempts to solve the issue that humans cannot visualize high-dimensional data. In other words, we cannot understand the relationship between numerous types of data all at once. SOM reduces dimensions by producing a 2D map that plots the similarities of the data by grouping similar data items together. Therefore, SOM analysis reduces dimensions and displays similarities. SOM approaches have been used in numerous fields, such as finance, industrial control, speech analysis, and astronomy (Fraser and Dickson, 2007). Roy et al. (2013) describe how neural networks have been used since the late 1990s in the industry to resolve various geoscience interpretation problems. In seismic interpretation, SOM is an ideal approach to understand how numerous seismic attributes relate and to classify various patterns in the data. Seismic data contain huge amounts of data samples, and they are highly continuous, greatly redundant, and significantly noisy (Coleou et al., 2003). The tremendous amount of samples from numerous seismic attributes exhibits significant organizational structure in the midst of noise. SOM analysis identifies these natural organizational structures in the form of clusters. These clusters reveal important information about the classification structure of natural groups that are difficult to view any other way. Dimensionality reduction properties of SOM are well known (Haykin, 2009). These natural groups and patterns in the data identified by the SOM

analysis routinely reveal geologic features important in the interpretation process.

Overview of self-organizing maps

In a time or depth seismic survey, samples are first organized into multiattribute samples, so all attributes are analyzed together as points. If there are F attributes, there are F numbers in each multiattribute sample. SOM is a nonlinear mathematical process that introduces several new, empty multiattribute samples called *neurons*. These SOM neurons will hunt for natural clusters of energy in the seismic data. The neurons discussed in this article form a 2D mesh that will be illuminated in the data with a 2D color map.

The SOM assigns initial values to the neurons, then for each multiattribute sample, it finds the neuron closest to that sample by the Euclidean distance and advances it toward the sample by a small amount. Other neurons nearby in the mesh are also advanced. This process is repeated for each sample in the training set, thus completing one epoch of SOM learning. The extent of neuron movement during an epoch is an indicator of the level of SOM learning during that epoch. If an additional epoch is worthwhile, adjustments are made to the learning parameters and the next epoch is undertaken. When learning becomes insignificant, the process is complete.

Figure 2 presents a portion of results of SOM learning on a 2D seismic line offshore of West Africa. For these results, a mesh of 8×8 neurons has six adjacent touching neurons. The 13 single-trace (instantaneous) attributes were selected for this analysis, so there was no communication between traces. These early results demonstrated that SOM learning was able to identify a great deal of geologic features. The 2D color map identifies different neurons with shades of green, blue, red, and yellow. The advantage of a 2D color map is that neurons that are adjacent to each other in the SOM analysis have similar shades of color. The figure reveals water-bottom reflections, shelf-edge peak and trough reflections, unconformities, onlaps/offlaps, and normal faults. These features are readily apparent on the SOM classification section, where amplitude is only one of 13 attributes used. Therefore, a SOM evaluation can incorporate several appropriate types of seismic attributes to

define geology not easily interpreted from conventional seismic amplitude displays alone.

Self-organizing map neurons

Mathematically, a SOM neuron (loosely following notation by Haykin, 2009) lies in attribute space alongside the normalized data samples, which together lie in $\mathbb{R}^F$. Therefore, a neuron is also an F -dimensional column vector, noted here as $\mathbf{w}$ in bold. Neurons learn or adapt to the attribute data, but they also learn from each other. A neuron $\mathbf{w}$ lies in a mesh, which may be 1D, 2D, or 3D, and the connections between neurons are also specified. The neuron mesh is a topology of neuron connections in a neuron space. At this point in the discussion, the topology is unspecified, so we use a single subscript j as a place marker for counting neurons just as we use a single subscript i to count selected multiattribute samples for SOM analysis.

$$\mathbf{w} \in \mathbf{W} = \{\mathbf{w}_j\}. \quad (5)$$

A neuron learns by adjusting its position within attribute space as it is drawn toward nearby data samples.

In general, the problem is to discover and identify an unknown number of natural clusters distributed in attribute space given the following: I data samples in survey space, F attributes in attribute space, and J neurons in neuron space. We are justified in searching for natural clusters because they are the multiattribute seismic expression of seismic reflections, seismic facies, faults, and other geobodies, which we recognize in our data as geologic features. For example, faults are often identified by the single attribute of coherency. Flat spots are identified because they are flat. The AVO anomalies are identified as classes 1, 2, 2p, 3, or 4 by classifying several AVO attributes.

The number of multiattribute samples is often in the millions, whereas the number of neurons is often in the dozens or hundreds. That is, the number of neurons $J \ll I$. Were it not so, detection of natural clusters in attribute space would be hopeless. The task of a neural network is to assist us in our understanding of the data.

Neuron topology was first inspired by observation of the brain and other neural tissue. We present here results based on a neuron topology $\mathbf{W}$, that is, 2D, so

Figure 2. Offshore West Africa 2D seismic line processed by SOM analysis. An 8×8 mesh of neurons trained on 13 instantaneous attributes with 100 epochs of unsupervised learning. A SOM of neurons resulted. In the figure insert, each neuron is shown as a unique color in the 2D color map. After training, each multiattribute seismic sample was classified by finding the neuron closest to the sample by Euclidean distance. The color of the neuron is assigned to the seismic sample in the display. A great deal of geologic detail is evident in classification by SOM neurons.

$\mathbf{W}$ lies in $\mathbb{R}^2$. Results shown in this paper have neuron meshes that are hexagonally connected (six adjacent points of contact).

Self-organizing-map learning

We now turn from a survey space perspective to an operational one to consider SOM learning as a process. SOM learning takes place during a series of time steps, but these time steps are not the familiar time steps of a seismic trace. Rather, these are time steps of learning. During SOM learning, neurons advance toward multiattribute data samples, thus reducing error and thereby advancing learning. A SOM neuron adjusts itself by the following recursion:

$$\mathbf{w}_j(n+1) = \mathbf{w}_j(n) + \eta(n) h_{j,k}(n)(\mathbf{x}_i - \mathbf{w}_j(n)), \quad (6)$$

where $\mathbf{w}_j(n)$ is the attribute position of neuron j at time step n . The recursion proceeds from time step n to $n+1$. The update is in the direction toward $\mathbf{x}_i$ along the “error” direction $\mathbf{x}_i - \mathbf{w}_j(n)$. This is the direction that pulls the winning neuron toward the data sample. The amount of displacement is controlled by learning controls, η and h , which will be discussed shortly.

Equation 6 depends on $\mathbf{w}$ and $\mathbf{x}$, so either select an $\mathbf{x}$, and then use some strategy to select $\mathbf{w}$, or vice versa. We elect to have all $\mathbf{x}$ participate in training, so we select $\mathbf{x}$ and use the following strategy to select $\mathbf{w}$. The neuron nearest $\mathbf{x}_i$ is the one for which the squared Euclidean distance,

$$d_j^2 = \|\mathbf{x}_i - \mathbf{w}_j\|^2 = \|\mathbf{x}_i - \mathbf{w}_j\|, \quad (7)$$

is the smallest of all $\mathbf{w}_j$. This neuron is called the *winning neuron*, and this selection rule is central to a competitive learning strategy, which will be discussed shortly. For data sample $\mathbf{x}_i$, the resulting winning neuron will have subscript j , which results from scanning over all neurons with free subscript s under the minimum condition noted as

$$j = \operatorname{argmin}_s \|\mathbf{x}_i - \mathbf{w}_s\|. \quad (8)$$

Now, the winning neuron for $\mathbf{x}_i$ is found from

$$\mathbf{w}_j(\mathbf{x}_i) = \{\mathbf{w}_j \mid j = \operatorname{argmin}_s \|\mathbf{x}_i - \mathbf{w}_s\| \forall \mathbf{w}_s \in \mathbf{W}\}, \quad (9)$$

where the bar $|$ reads “given” and the inverted $\forall$ reads “for all.” That is, the winning neuron for data sample $\mathbf{x}_i$ is $\mathbf{w}_j$. We observe that for every data sample, there is a winning neuron. One complete pass through all data samples $\{\mathbf{x}_i \mid i = 1 \text{ to } I\}$ is called one *epoch*. One epoch completes a single time step of learning. We typically exercise 60–100 epochs of training. It is noted that “epoch” as a unit of machine learning shares a sense of time with a geologic epoch, which is a division of time between periods and ages.

Returning to learning controls of equation 6, let the first term η change with time so as to be an adjustable learning rate control. We choose

$$\eta(n) = \eta_0 \exp(-n/\tau_2), \quad (10)$$

with η_0 as some initial learning rate and τ_2 as a learning decay factor. As time progresses, the learning control in equation 10 diminishes. This results in neurons that move large distances during early time steps move smaller distances in later time steps.

The second term h of equation 6 is a little more complicated and calls into action the neuron topology $\mathbf{W}$. Here, h is called the *neighborhood function* because it adjusts not only the winning neuron $\mathbf{w}_j$ but also other neurons in the neighborhood of $\mathbf{w}_j$.

Now, the neuron topology $\mathbf{W}$ is 2D, and the neighborhood function is given by

$$h_{j,k}(n) = \exp(-d_{j,k}^2 / 2 \sigma^2(n)), \quad (11)$$

where $d_{j,k}^2 = \|\mathbf{r}_j - \mathbf{r}_k\|^2$ for a neuron at $\mathbf{r}_j$ and the winning neuron at $\mathbf{r}_k$ in the neuron topology. Distance d in equation 11 represents the distance between a winning neuron and another nearby neuron in neuron topology $\mathbf{W}$. The neighborhood function in equation 11 depends on the distance between a neuron and the winning neuron and also time. The time-varying part of equation 11 is defined as

$$\sigma(n) = \sigma_0 \exp(-n/\tau_1), \quad (12)$$

where σ_0 is the initial neighborhood distance and τ_1 is the neighborhood decay factor. As σ increases with time, h decreases with time. We define an edge of the neighborhood as the distance beyond which the neuron weight is negligible and treated as zero. In 2D neuron topologies, the neighborhood edge is defined by a radius. Let this cut-off distance depend on a free constant ζ . In equation 11, we set $h = \zeta$ and solve for $d_{\max}$ as

$$d_{\max}(n) = [-2 \sigma^2(n) \ln \zeta]^{1/2}. \quad (13)$$

The neighborhood edge distance $d_{\max} \rightarrow 0$ as $\zeta \rightarrow 0$. As time marches on, the neighborhood edge shrinks to zero and continued processing steps of SOM are similar to K-means clustering (Bishop, 2007). Additional details of SOM learning are found in Appendix A on SOM analysis operation.

Harvesting

Rather than apply the SOM learning process to a large time or depth window spanning an entire 3D survey, we sample a subset of the full complement of multiattribute samples in a process called *harvesting*. This is first introduced by Taner et al. (2009) and is described in more detail by Smith and Taner (2010).

First, a representative set of harvest patches from the 3D survey is selected, and then on each of these patches, we conduct independent SOM training. Each harvest patch is one or more lines, and each SOM analysis yields a set of winning neurons. We then apply a har-

vest rule to select the set of winning neurons that best represent the full set of data samples of interest.

A variety of harvest rules has been investigated. We often choose a harvest rule based on best learning. Best learning selects the winning neuron set for the SOM training, in which there has been the largest proportional reduction of error between initial and final epochs on the data that were presented. The error is measured by summing distances as a measure of how near the winning neurons are to their respective data samples. The largest reduction in error is the indicator of best learning. Additional details on harvest sampling and harvest rules are found in Appendix A on SOM analysis operation.

Self-organizing-map classification and probabilities

Once natural clusters have been identified, it is a simple task to classify samples in survey space as members of a particular cluster. That is, once the learning process has completed, the winning neuron set is used to classify each selected multiattribute sample in the survey. Each neuron in the winning neuron set ($j = 1$ to J) is tested with equation 9 against each selected sample in the survey ($i = 1$ to I). Each selected sample then has assigned to it a neuron that is nearest to that sample in Euclidean distance. The winning neuron index is assigned to that sample in the survey.

Every sample in the survey has associated with it a winning neuron separated by a Euclidean distance that is the square root of equation 7. After classification, we study the Euclidean distances to see how well the neurons fit. Although there are perhaps many millions of survey samples, there are far fewer neurons, so for each neuron, we collect its distribution of survey sample distances. Some samples near the neuron are a good

fit, and some samples far from the neuron are a poor fit. We quantify the goodness-of-fit by distance variance as described in Appendix A. Certainly, the probability of a correct classification of a neuron to a data sample is higher when the distance is smaller than when it is larger. So, in addition to assigning a winning neuron index to a sample, we also assign a classification probability. The classification probability ranges from one to zero corresponding to distance separations of zero to infinity. Those areas in the survey where the classification probability is low correspond to areas where no neuron fits the data very well. In other words, anomalous regions in the survey are noted by low probability. Additional comments are found in Appendix A.

Case studies

Offshore Gulf of Mexico

This case study is located offshore Louisiana in the Gulf of Mexico in a water depth of 143 m (470 ft). This shallow field (approximately 1188 m [3900 ft]) has two producing wells that were drilled on the upthrown side of an east–west-trending normal fault and into an amplitude anomaly identified on the available 3D seismic data. The normally pressured reservoir is approximately 30 m (100 ft) thick and located in a typical “bright-spot” setting, i.e., a Class 3 AVO geologic setting (Rutherford and Williams, 1989). The goal of the multiattribute analysis is to more clearly identify possible DHI characteristics such as flat spots (hydrocarbon contacts) and attenuation effects to better understand the existing reservoir and provide important approaches to decrease risk for future exploration in the area.

Initially, 18 instantaneous seismic attributes were generated from the 3D data in this area (see Table 2). These seismic attributes were put into a PCA evaluation to determine which produced the largest variation in the data and the most meaningful attributes for SOM analysis. The PCA was computed in a window 20 ms above and 150 ms below the mapped top of the reservoir over the entire survey, which encompassed approximately 26 km² (10 mi²). Figure 3a displays a chart with each bar representing the highest eigenvalue on its associated inline over the displayed portion of the survey. The bars in red in Figure 3a specifically denote the inlines that cover the areal extent of the amplitude feature and the average of their eigenvalue results are displayed in Figure 3b and 3c. Figure 3b displays the principal components from the selected inlines over the anomalous feature with the highest eigenvalue (first principal component) indicating the percentage of seismic attributes contributing to this largest variation in the data. In this first principal component, the top seismic attributes include the envelope, envelope modulated phase, envelope second derivative, sweetness, and average energy, all of which account for more than 63% of the variance of all the instantaneous attributes in this PCA evaluation. Figure 3c displays the PCA results, but this time the second highest eigenvalue was se-

Table 2. Instantaneous seismic attributes used in the PCA evaluation for the Gulf of Mexico case study.

Gulf of Mexico Case Study Seismic Attributes Employed in PCA	
Acceleration of Phase	Instantaneous Frequency Envelope Weighted
Average Energy	Instantaneous Phase
Bandwidth	Instantaneous Q
Dominant Frequency	Normalized Amplitude
Envelope Modulated Phase	Real Part
Envelope Second Derivative	Relative Acoustic Impedance
Envelope Time Derivative	Sweetness
Imaginary Part	Thin Bed Indicator
Instantaneous Frequency	Trace Envelope

lected and produced a different set of seismic attributes. The top seismic attributes from the second principal component include instantaneous frequency, thin bed indicator, acceleration of phase, and dominant frequency, which total almost 70% of the variance of the 18 instantaneous seismic attributes analyzed. These results suggest that when applied to an SOM analysis, perhaps the two sets of seismic attributes for the first and second principal components will help to define two different types of anomalous features or different characteristics of the same feature.

The first SOM analysis (SOM A) incorporates the seismic attributes defined by the PCA with the highest variation in the data, i.e., the five highest percentage contributing attributes in Figure 3b. Several neuron counts for SOM analyses were run on the data with lower count matrices revealing broad, discrete features and the higher counts displaying more detail and less variation. The SOM results from a 5×5 matrix of neurons (25) were selected for this paper. The north-south line through the field in Figures 4 and 5 shows the origi-

nal stacked amplitude data and classification results from the SOM analyses. In Figure 4b, the color map associated with the SOM classification results indicates all 25 neurons are displayed, and Figure 4c shows results with four interpreted neurons highlighted. Based on the location of the hydrocarbons determined from well control, it is interpreted from the SOM results that attenuation in the reservoir is very pronounced with this evaluation. As Figure 4b and 4c reveal, there is apparent absorption banding in the reservoir above the known hydrocarbon contacts defined by the wells in the field. This makes sense because the seismic attributes used are sensitive to relatively low-frequency broad variations in the seismic signal often associated with attenuation effects. This combination of seismic attributes used in the SOM analysis generates a more pronounced and clearer picture of attenuation in the reservoir than any one of the seismic attributes or the original amplitude volume individually. Downdip of the field is another undrilled anomaly that also reveals apparent attenuation effects.

Figure 3. Results from PCA using 18 instantaneous seismic attributes: (a) bar chart with each bar denoting the highest eigenvalue for its associated inline over the displayed portion of the seismic 3D volume. The red bars specifically represent the highest eigenvalues on the inlines over the field, (b) average of eigenvalues over the field (red) with the first principal component in orange and associated seismic attribute contributions to the right, and (c) second principal component over the field with the seismic attribute contributions to the right. The top five attributes in panel (b) were run in SOM A, and the top four attributes in panel (c) were run in SOM B.

The second SOM evaluation (SOM B) includes the seismic attributes with the highest percentages from the second principal component based on the PCA (see Figure 3). It is important to note that these attributes are different than the attributes determined from the first principal component. With a 5×5 neuron matrix, Figure 5 shows the classification results from this SOM evaluation on the same north–south line as Figure 4, and it clearly identifies several hydrocarbon contacts in the form of flat spots. These hydrocarbon contacts in the field are confirmed by the well control. Figure 5b defines three apparent flat spots, which are further isolated in Figure 5c that displays these features with two neurons. The gas/oil contact in the field was very difficult to see on the original seismic data, but it is well defined and mappable from this SOM analysis. The oil/water contact in the field is represented by a flat spot that defines the overall base of the hydrocarbon reservoir. Hints of this oil/water contact were interpreted

from the original amplitude data, but the second SOM classification provides important information to clearly define the areal extent of reservoir. Downdip of the field is another apparent flat spot event that is undrilled and is similar to the flat spots identified in the field. Based on SOM evaluations A and B in the field that reveal similar known attenuation and flat spot results, respectively, there is a high probability this undrilled feature contains hydrocarbons.

Shallow Yegua trend in Gulf Coast of Texas

This case study is located in Lavaca County, Texas, and targets the Yegua Formation at approximately 1828 m (6000 ft). The initial well was drilled just downthrown on a small southwest–northeast regional fault, with a subsequent well being drilled on the upthrown side. There were small stacked data amplitude anomalies on the available 3D seismic data at both well locations. The Yegua in the wells is approximately 5 m (18 ft) thick and is composed of thinly laminated sands. Porosities range from 24% to 30% and are normally pressured. Because of the thin laminations and often lower porosities, these anomalies exhibit a class 2 AVO response, with near-zero amplitudes on the near offsets and an increase in negative amplitude with offset (Rutherford and Williams, 1989). The goal of the multiattribute analysis was to determine the full extent of the reservoir because both wells were performing much better than the size of the amplitude anomaly indicated from the stacked seismic data (Figure 6a and 6b). The first well drilled downthrown had a stacked data amplitude anomaly of approximately 70 acres, whereas the second well upthrown had an anomaly of about 34 acres.

The gathers that came with the seismic data had been conditioned and were used in creating very specific AVO volumes conducive to the identification of class 2 AVO anomalies in this geologic setting. In this case, the AVO attributes selected were based on the interpreter's experience in this geologic setting. Table 3 lists the AVO attributes and the attributes generated from the AVO attributes used in this SOM evaluation. The intercept and gradient volumes were created using the Shuey three-term approximation of the Zoepritz equation (Shuey, 1985). The near-offset volume was produced from the 0° – 15° offsets and the far-offset volume from the 31° – 45° offsets. The attenuation, envelope bands on the envelope breaks, and envelope bands on the

Figure 4. SOM A results on the north–south inline through the field: (a) original stacked amplitude, (b) SOM results with associated 5×5 color map displaying all 25 neurons, and (c) SOM results with four neurons selected that isolate interpreted attenuation effects.

Figure 5. SOM B results on the same inline as Figure 4: (a) original stacked amplitude, (b) SOM results with associated 5×5 color map, and (c) SOM results with color map showing two neurons that highlight flat spots in the data. The hydrocarbon contacts (flat spots) in the field were confirmed by well control.

Figure 6. Stacked data amplitude maps at the Yegua level denote: (a) interpreted outline of hydrocarbon distribution based on upthrown amplitude anomaly and (b) interpreted outline of hydrocarbons based on downthrown amplitude anomaly.

phase breaks seismic attributes were all calculated from the far-offset volume. For this SOM evaluation, an 8 × 8 matrix (64 neurons) was used.

Figure 7 displays the areal distribution of the SOM classification at the Yegua interval. The interpretation of this SOM classification is that the two areas outlined represent the hydrocarbon producing portion of the reservoir and all the connectivity of the sand feeding into the well bores. On the downthrown side of the fault, the drainage area has increased to approximately 280 acres, which supports the engineering and pressure data. The areal extent of the drainage area on the upthrown res-

ervoir has increased to approximately 95 acres, and again agreeing with the production data. It is apparent that the upthrown well is in the feeder channel, which deposited sand across the then-active fault and splays along the fault scarp.

In addition to the SOM classification, the anomalous character of these sands can be easily seen in the probability results from the SOM analysis (Figure 8). The probability is a measure of how far the neuron is from its identified cluster (see Appendix A). The low-probability zones denote the most anomalous areas determined from the SOM evaluation. The most anomalous areas typically will have the lowest probability, whereas the events that are present over most of the data, such as horizons, interfaces, etc., will have higher probabilities. Because the seismic attributes that went into this SOM analysis are AVO-related attributes that enhance DHI features, these low-probability zones are interpreted to be associated with the Yegua hydrocarbon-bearing sands.

Figure 9 displays the SOM classification results with a time slice located at the base of the upthrown reservoir and the upper portion of the downthrown reservoir. There is a slight dip component in the figure. Figure 9a reveals the total SOM classification results with all 64 neurons as indicated by the associated 2D color map. Figure 9b is the same time slice slightly rotated to the west with very specific neurons highlighted in the 2D color map defining the upthrown and downthrown fields. The advantage of SOM classification analyses is the ability to isolate specific neurons that highlight desired geologic features. In this case, the SOM classification of the AVO-related attributes was able to define the reservoirs drilled by each of the wells and provide a more accurate picture of their areal distributions than the stacked data amplitude information.

Eagle Ford Shale

This study is conducted using 3D seismic data from the Eagle Ford Shale resource play of south Texas. Understanding the existing fault and fracture patterns in the Eagle Ford Shale is critical to optimizing well locations, well plans, and fracture treatment design. To identify fracture trends, the industry routinely uses various seismic techniques, such as processing of seismic attributes, especially geometric attrib-

Table 3. The AVO seismic attributes computed and used for the SOM evaluation in the Yegua case study.

Yegua Case Study AVO Attributes Employed for SOM Evaluation	
Near	
Far	
Far-Near	Near = 0° – 15° Far = 31° – 45°
(Far-Near)Far	
Intercept	
Gradient	Shuey 3 Term Approximation $RC(\theta) = A + B \sin^2 \theta + C(\sin^2 \theta \tan^2 \theta)$
Intercept X Gradient	
1/2(Intercept + Gradient)	
Attenuation on Far Offsets	
Envelope Bands on Envelope Breaks from Far Offsets	
Envelope Bands on Phase Breaks from Far Offsets

Figure 7. The SOM classification at the Yegua level denoting a larger area around the wells associated with gas drainage than indicated from the stacked amplitude response as seen in Figure 6. Also shown is the location of the arbitrary line displayed in Figure 8. The 1D color bar has been designed to highlight neurons 1 through 9 interpreted to indicate those neuron patterns which represent sand/reservoir extents.

utes, to derive the maximum structural information from the data.

Geometric seismic attributes describe the spatial and temporal relationship of all other attributes (Taner,

2003). The two main categories of these multitrace attributes are coherency/similarity and curvature. The objective of coherency/similarity attributes is to enhance the visibility of the geometric characteristics of seismic

Figure 8. Arbitrary line (location in Figure 7) showing low probability in the Yegua at each well location, indicative of anomalous results from the SOM evaluation. The colors represent probabilities with the wiggle trace in the background from the original stacked amplitude data.

Figure 9. The SOM classification results at a time slice show the base of the upthrown reservoir and the upper portion of the dowthrown reservoir and denote: (a) full classification results defined by the associated 2D color map and (b) isolation of upthrown and dowthrown reservoirs by specific neurons represented by the associated 2D color map.

data such as dip, azimuth, and continuity. *Curvature* is a measure of how bent or deformed a surface is at a particular point with the more deformed the surface the

more the curvature. These characteristics measure the lateral relationships in the data and emphasize the continuity of events such as faults, fractures, and folds.

Figure 10. Three geometric attributes at the top of the Eagle Ford Shale computed from (left) the full-frequency data and (right) the 24.2-Hz spectral decomposition volume with results from the (a) dip of maximum similarity, (b) curvature most positive, and (c) curvature minimum. The 1D color bar is common for each pair of outputs.

The goal of this case study is to more accurately define the fault and fracture patterns (regional stress fields) than what had already been revealed in running geometric attributes over the existing stacked seismic data. Initially, 18 instantaneous attributes, 14 coherency/similarity attributes, 10 curvature attributes, and 40 frequency subband volumes of spectral decomposition were generated. In the evaluation of these seismic attributes, it was determined in the Eagle Ford interval that the highest energy resided in the 22- to 26-Hz range. Therefore, a comparison was made with geometric attributes computed from a spectral decomposition volume with a center frequency of 24.2 Hz with the same geometric attributes computed from the original full-frequency volume. At the Eagle Ford interval, Figure 10 compares three geometric attributes generated from the original seismic volume with the same geometric attributes generated from the 24.2-Hz spectral decomposition volume. It is evident from each of these geometric attributes that there is an improvement in the

Table 4. Geometric attributes used in the Eagle Ford Shale SOM analysis.

Eagle Ford Shale Case Study Attributes Employed	
Curvature Maximum	
Curvature Minimum	
Curvature Mean	
Curvature Dip Direction	
Curvature Strike Direction	
Curvature Most Negative	
Curvature Most Positive	
Dip of Maximum Similarity	
Curvature Maximum	

image delineation of fault/fracture trends with the spectral decomposition volumes. Based on the results of the geometric attributes produced from the 24.2-Hz volume and trends in the associated PCA interpretation, Table 4 lists the attributes used in the SOM analysis over the Eagle Ford interval. This SOM analysis incorporated an 8 × 8 matrix (64 neurons). Figure 11 displays the results at the top of the Eagle Ford Shale of the SOM analysis using the nine geometric attributes computed from the 24.2-Hz spectral decomposition volume. The associated 2D color map in Figure 11 provides the correlation of colors to neurons. There are very clear northeast–southwest trends of relatively large fault and fracture systems, which are typical for the Eagle Ford Shale (primarily in dark blue). What is also evident is an orthogonal set of events running southeast–northwest and east–west (red). To further evaluate the SOM results, individual clusters or patterns in the data are isolated with the highlighting of specific neurons in the 2D color map in Figure 12. Figure 12a indicates neuron 31 (blue) is defining the larger northeast–southwest fault/fracture trends in the Eagle Ford Shale. Figure 12b with neuron 14 (red) indicates orthogonal sets of events. Because the survey was acquired southeast–northwest, it could be interpreted that the similar trending events in Figure 12b are possible acquisition footprint effects, but there are very clear indications of east–west lineations also. These east–west lineations probably represent fault/fracture trends orthogonal to the major northeast–southwest trends in the region. Figure 12c displays the neurons from Figure 12a and 12b, as well as neuron 24 (dark gray). With these three neurons highlighted, it is easy to see the fault and fracture trends against a background, where neuron 24 displays a relatively smooth and nonfaulted region. The key issue in this evaluation is that the SOM analysis allows the breakout of the fault/fracture trends and allows the geoscientist to make better informed decisions in their interpretation.

Figure 11. SOM results from the top of the Eagle Ford Shale with associated 2D color map.

Figure 12. SOM results from the top of the Eagle Ford Shale with (a) only neuron 31 highlighted denoting northeast-southwest trends, (b) neuron 14 highlighted denoting east-west trends, and (c) three neurons highlighted with neuron 24 displaying smoothed nonfaulted background trend.

Conclusions

Seismic attributes, which are any measurable properties of seismic data, aid interpreters in identifying geologic features, which are not clearly understood in the original data. However, the enormous amount of information generated from seismic attributes and the difficulty in understanding how these attributes when combined define geology, requires another approach in the interpretation workflow. The application of PCA can help interpreters to identify seismic attributes that show the most variance in the data for a given geologic setting. The PCA works very well in geologic settings, where anomalous features stand out from the

background data, such as class 3 AVO settings that exhibit DHI characteristics. The PCA helps to determine, which attributes to use in a multiattribute analysis using SOMs.

Applying current computing technology, visualization techniques, and understanding of appropriate parameters for SOM, enables interpreters to take multiple seismic attributes and identify the natural organizational patterns in the data. Multiple-attribute analyses are beneficial when single attributes are indistinct. These natural patterns or clusters represent geologic information embedded in the data and can help to identify geologic features, geobodies, and aspects of geology that often cannot be interpreted by any other means. The SOM evaluations have proven to be beneficial in essentially all geologic settings including unconventional resource plays, moderately compacted onshore regions, and offshore unconsolidated sediments. An important observation in the three case studies is that the seismic attributes used in each SOM analysis were different. This indicates the appropriate seismic attributes to use in any SOM evaluation should be based on the interpretation problem to be solved and the associated geologic setting. The application of PCA and SOM can not only identify geologic patterns not seen previously in the seismic data, but it also can increase or decrease confidence in already interpreted features. In other words, this multiattribute approach provides a methodology to produce a more accurate risk assessment of a geoscientist's interpretation and may represent the next generation of advanced interpretation.

Acknowledgements

The authors would like to thank the staff of Geophysical Insights for the research and development of the applications used in this paper. We offer sincere thanks to T. Taner (deceased) and S. Treitel, who were early visionaries to recognize the breadth and depth of neural network benefits as they apply to our industry. They planted seeds. The seismic data offshore West Africa was generously provided by B. Bernth (deceased) of SCS Corporation. The seismic data in the offshore Gulf of Mexico case study is courtesy of Petroleum Geo-Services. Thanks to T. Englehart for insight into the Gulf of Mexico case study. Thanks also to P. Santogrossi and B. Taylor for review of the paper and for the thoughtful feedback.

Appendix A

SOM analysis operationals

We address SOM analysis in two parts — SOM learning and SOM classification. The numerical engine of SOM learning as embodied in equations 6–13 has only six learning controls:

- 1) η_0 — learning rate
- 2) τ_2 — learning decay factor
- 3) σ_0 — initial neighborhood distance
- 4) τ_1 — neighborhood decay factor
- 5) ζ — neighborhood edge factor
- 6) $n_{\max}$ — maximum number of epochs.

Typical values are as follows: $\eta_0 = 0.3$, $\tau_2 = 10$, $\sigma_0 = 7$, $\tau_1 = 4$, $\zeta = .1$, and $n_{\max} = 100$. We find the SOM algorithm to be robust and not particularly sensitive to parameter values. We first illustrate the amount of movement a winning neuron advances in successive time epochs. In the main paper, the product of equations 10 and 11 is applied to the projection of a neuron $\mathbf{w}_j$ in the direction of a data sample $\mathbf{x}_i$ as written in equation 6. When the neuron is the winning neuron as defined in equation 9, the distance between the neuron and winning neuron is zero, so the neighborhood function in equation 11 is one. Figure A-1 shows the total weight applied to the difference vector $\mathbf{x}_i - \mathbf{w}_j$ for the winning neuron.

Next, we develop the amount of movement of neurons in the neighborhood near the winning neuron. The time-varying part of the neighborhood function of equation 11 is shown in equation 12 and is illustrated in Figure A-2. Now, the distance from the winning neuron to its nearest neuron neighbor has a nominal distance of one. Figure A-3 presents the total weight that would be applied to the neuron adjacent to a winning neuron. Not only does the size of the neighborhood collapse as time marches on, as noted in equation 13, but also the shape changes. Figure A-4 illustrates the shape as a function of distance from the winning neuron on four different epochs. As the epochs increase, the neighborhood function contracts. To illustrate equation 13, Figure A-5 shows the distance from the winning neuron to the edge of the neighborhood across several epochs. Combining equations 10 and 11, Figure A-6 shows the weights that would be applied to the neuron adjacent to the winning neuron for each epoch. Notice

that they are always smaller than the weights applied to the winning neuron of Figure A-1. The winning neuron, as recipient of competitive learning, always moves the most.

Training epochs

Because SOM neurons adjust positions through an endless recursion, the choice of initial neuron positions and the number of time steps are important requirements. The initial neuron positions are specified by their vector components, which are often randomly selected, although Kohonen (2001) points out that much faster convergence may be realized if natural ordering of the data is used instead. Instead, we investigate three

Figure A-1. Winning neuron weight η at epoch n in equation 10 for $\eta_0 = 0.3$ and $\tau_2 = 10$.

Figure A-2. Neighbor weight factor σ at each epoch n in equation 12 for $\sigma_0 = 7$ and $\tau_1 = 10$.

neuron initial conditions: $\mathbf{w} = 0$, $\mathbf{w} = \text{rand}(-1 \text{ to } 1)$, and $\mathbf{w}_i = \mathbf{x}_i$. We make a great effort to ensure that the result does not depend on the initial conditions.

Haykin (2009) observes that SOM learning passes through two phases during a set of training time steps. First, there is an ordering phase when the winning neuron activity and neighborhood neuron activity (cooperative learning) are underway. This phase is followed by a convergence phase when winning neurons make refining adjustments to arrive at their final stable locations (competitive learning). Learning controls are important because there should be sufficient time steps for the neurons to converge on any clustering that is present in the

data. If learning controls are set too tight, the learning processes is quenched too soon.

Let us investigate the behavior of neighborhood edge $d_{\max}$ as epoch n increases. We observe in equation 13 that $d_{\max}$ monotonically decreases with increasing n . At some time $d_{\max}$ decreases to one, the neighborhood then has a radius so small that there are no adjacent neurons to the winning neuron. This marks the transition from cooperative learning to competitive learning:

$$d_{\max}(n_{\text{switch}}) = 1, \quad \text{learning transition,} \quad (\text{A-1})$$

$$n < n_{\text{switch}}, \quad \text{cooperative learning,} \quad (\text{A-2})$$

$$n > n_{\text{switch}}, \quad \text{competitive learning.} \quad (\text{A-3})$$

Figure A-3. Neighborhood weight function h in equation 11 for neuron nearest the winning neuron in network ($d = 1$).

Figure A-4. Neighborhood weight function h in equation 11 for epochs = (left) 20, 10, 5, and (right) 1.

Figure A-5 illustrates the transition from cooperative to competitive learning. The edge of the neighborhood collapses to one on approximately epoch 25. On epoch 11, the edge of the neighborhood is approximately five, which means that all neurons within a radius of five units will have at least some movement toward the data sample. Those neurons nearest the winning neuron move the most. On epochs after the 25th, the edge has shrunk to less than one unit, so only the winning neuron has movement.

Let us find the value of time step n for this transition. Starting with equations 12 and 13 and inserting condition in equation A-1, the transition n is

$$n_{\text{switch}} = \frac{\tau_1}{2} \ln(-2\sigma_0^2 \ln \zeta). \quad (\text{A-4})$$

Another way to set learning controls is to decide on the number of cooperative and competitive epochs before learning is started. Let the maximum number of epochs be $n_{\max}$ as before, but also let f be the fraction of $n_{\max}$ at which the transition from cooperative to competitive learning will occur; that is,

$$n_{\text{switch}} = f n_{\max} \mid 0 < f < 1. \quad (\text{A-5})$$

For example, $f = 0.25$ implies that the first 25% of the epochs are cooperative and the last 75% are competitive.

Combining equations A-4 with A-5, and then solving for τ_1 leads to

$$\tau_1 = \frac{2f}{\ln(-2\sigma_0^2 \ln \zeta)} n_{\max}. \quad (\text{A-6})$$

Substituting $f = 0.27$, $\sigma_0 = 7$, and $\zeta = 0.1$ into equation A-6 results in the learning control $\tau_1 = 10$, so the standard controls presented above have 27% cooperative and 73% competitive learning epochs.

Equation A-6 is useful for SOM learning controls where $n_{\max}$ is an important consideration. If so, then an alternate set of learning controls includes the following:

- 1) η_0 — learning rate
- 2) τ_2 — learning decay factor
- 3) σ_0 — initial neighborhood distance
- 4) τ_1 — neighborhood decay factor
- 5) ζ — neighborhood edge factor
- 6) f — fraction of $n_{\max}$ devoted to cooperative learning
- 7) $n_{\max}$ — maximum number of epochs.

The proportion of cooperative to competitive learning is important. We find that an f value in the range 0.2–0.4 yields satisfactory results and that $n_{\max}$ should be large enough for consistent results.

Convergence

Recall that all data samples in a survey are a set with I members. For seismic data, we point out that this amounts to many thousands or perhaps millions of data-point calculations. Because the winning neuron cannot be known until a data point is compared with all neurons, it is noted that the SOM computations are $O(I^2)$. Although the number of calculations for any one application of the SOM recursion is small, these computations increase linearly as the number of data points and/or the number of neurons increase. Calculations require many epochs to ensure that training results are reproducible and independent of initial conditions. Careful attention to numerical techniques, such as pipelining and multithreading are beneficial.

It is advantageous to select a representative training subset $\hat{\mathbf{X}}$ of $\mathbf{X}$ to reduce the number of samples from I to $\hat{I}$. We define an epoch as the amount of calculations required to process all samples in the reduced data. Mathematically, the training subset

$$\hat{\mathbf{X}} \subset \mathbf{X}, \quad (\text{A-7})$$

$$\hat{\mathbf{X}} = \{ \hat{\mathbf{x}}_1 \dots \hat{\mathbf{x}}_{\hat{I}} \}, \quad \text{training subset}, \quad (\text{A-8})$$

$$\mathbf{X} = \{ \mathbf{x}_1 \dots \mathbf{x}_I \}, \quad \text{full sample set}, \quad (\text{A-9})$$

where $\subset$ reads “is a subset of.”

Observe that a neuron is adjusted by an amount proportional to the difference vector $\hat{\mathbf{x}}_i - \mathbf{w}_j$. The error for this epoch is quantified by total distance movement of all neurons. From the total distance and total distance squared defined in equations A-10 and A-11 below, we compute the mean and standard deviation of neuron nearness distance in equations A-12 and A-13

Figure A-5. Neighborhood edge radius in equation 13 in nominal neuron distance units $\eta_0 = 0.3$, $\tau_2 = 10$, and $\zeta = 0.1$

Figure A-6. Weight $\eta(n)h_{|j-k|=1}(n)$ in equation 6 for the neuron nearest the winning neuron.

$$u(n) = \sum_i^{\hat{I}} \|\hat{\mathbf{x}}_i - \mathbf{w}_j(\hat{\mathbf{x}}_i)\|, \quad (\text{A-10})$$

$$v(n) = \sum_i^{\hat{I}} [[\hat{\mathbf{x}}_i - \mathbf{w}_j(\hat{\mathbf{x}}_i)]], \quad (\text{A-11})$$

$$\hat{d}(n) = \frac{1}{\hat{I}} u(n), \quad (\text{A-12})$$

$$\hat{\sigma}(n) = \sqrt{[v(n) - u^2(n)/\hat{I}]/(\hat{I} - 1)}. \quad (\text{A-13})$$

Now equations A-12 and A-13 are useful indicators of convergence because they decrease slowly as epochs advance in time. We often terminate training at an epoch, in which the absolute change in standard deviation of the total distance falls below a minimum level.

It is also useful to record how many samples switch the winning neuron index from one epoch to the next. These are counted by the sum

$$S(n) = \sum_i^{\hat{I}} \delta[\arg \mathbf{w}_j(\hat{\mathbf{x}}_i; n) - \arg \mathbf{w}_j(\hat{\mathbf{x}}_i; n - 1)], \quad (\text{A-14})$$

where δ is the Kronecker delta function and $\arg \mathbf{w}_j(\hat{\mathbf{x}}_i; n)$ is the winning neuron index number for a data sample at time step n . During early epochs, many data samples have winning neurons that switch from one to another. In later time steps, there are a diminishing number of changes. Early time steps are described by Haykin (2009) as an ordering phase. In later time steps as switching reduce, neurons enter a refinement phase, in which their motion decelerates toward their final resting positions.

Regardless of whether convergence is achieved by reaching $n_{\max}$ epochs or by terminating on some threshold of training inactivity, SOM training is completed and the set of neurons at the end of the last epoch is termed the *winning neuron set*.

We note that there is no known general proof of SOM convergence (Ervin et al., 1992), although slow convergence has not been an issue with multiattribute seismic data. SOM learning rates must not be so rapid that training is “quenched” before completion. To guard against this mistake, we have conducted extensive tests with slow learning rates and with several widely separated initial neuron coordinates, all arriving at the same results. Indeed, our experience with SOM learning yields satisfactory results when applied with learning parameters similar to those discussed here.

The field of mapping “big data” high-dimensionality information is of interest across several different fields, yet high dimensionality challenges all models with non-linear scaling and performance limitations among other issues (for instance, see the summary in Bishop, 2006,

pp. 595–599). For example, the generative topographic mapping, similar in some respects to SOM, offers a closed-form 2D analytic likelihood estimator, which bounds the computational effort, but it makes certain Gaussian distribution assumptions of the input space, which may be objectionable to some (Bishop et al., 1998). The properties of multiattribute seismic data, if not mixed with other types of data, are fairly smooth and that simplifies machine learning.

Harvest patches and rules

When the number of multiattribute samples is large, it is advantageous to select several harvest patches that span the geographic region of interest. For example, in reconnaissance of a large 3D survey with a wide time or depth window, we may sample the survey by selecting harvest patches of one line width and an increment of every fifth line across the survey.

In other situations, where the time or depth interval is small and bounded by a horizon, there may be only one harvest patch that spans the entire survey. The purpose of harvesting is to purposefully subsample the data of interest.

We have investigated several harvest rules:

- 1) select harvest line results with least error
- 2) select harvest line results with best learning
- 3) select harvest line with fewest poorly fitting neurons
- 4) select harvest line most like the average
- 5) select all harvest lines and make neuron “soup.”

Classification attribute

On completion of training on the training subset equation A-8 on the final epoch N , the set of J winning neurons is

$$\mathbf{W} = \{ \mathbf{w}_1(N) \dots \mathbf{w}_j(N) \dots \mathbf{w}_J(N) \}. \quad (\text{A-15})$$

Also at the end of the last epoch, each of the training subset samples has a winning neuron index. The training subset of data samples has a set of $\hat{I}$ indices:

$$\mathbf{K}_I = \{ \arg \mathbf{w}_k(\hat{\mathbf{x}}_1; N) \dots \arg \mathbf{w}_k(\hat{\mathbf{x}}_i; N) \dots \arg \mathbf{w}_k(\hat{\mathbf{x}}_I; N) \}. \quad (\text{A-16})$$

The set of I winning neuron indices for all data samples in the survey is

$$\mathbf{K} = \{ \arg \mathbf{w}_k(\mathbf{x}_1; N) \dots \arg \mathbf{w}_k(\mathbf{x}_i; N) \dots \arg \mathbf{w}_k(\mathbf{x}_I; N) \}, \quad (\text{A-17})$$

which we will simplify for later use as

$$\mathbf{K} = \{k_1 \dots k_I\}. \quad (\text{A-18})$$

These I samples of winning neuron indices are a 3D classification survey attribute as

$$\mathbf{K} = \{k_{c,d,e}\}, \quad (\text{A-19})$$

where c is the time or depth sample index, d is the trace number, and e is the line number.

The set of winning neuron indices $\mathbf{K}$ is a complete survey and the first of several SOM survey attributes.

Classification attribute subsets

Although $\mathbf{K}$ is the classification attribute, we turn now to the data samples associated with each winning neuron index. It has been noted previously that every data sample has a winning neuron. This implies that we can gather all the data samples that have a common winning neuron index into a classification subset. For example, the subset of data samples for the first winning neuron index is

$$\mathbf{X}_1 = \{\mathbf{x}_i \mid \arg \mathbf{w}_j(\mathbf{x}_i) = 1 \forall \mathbf{x}_i \in \mathbf{X}\} \subset \mathbf{X} \neq \hat{\mathbf{X}}. \quad (\text{A-20})$$

In total, when we gather all the J neuron subsets, we recover the original multiattribute survey as

$$\mathbf{X} = \sum_j^J \mathbf{X}_j. \quad (\text{A-21})$$

There may be some winning neurons with a large subset of data samples, whereas others have few. There may even be a few winning neurons with no data samples at all. This is to be expected because during the learning process, a neuron is abandoned when it no longer is the winning neuron of any data sample. More precisely, SOM training is nonlinear mapping, by which dimensionality reduction is not everywhere connected.

There are several interesting properties of the subset of $\mathbf{x}$ data samples associated with a winning neuron. Consider a subset of $\mathbf{X}$ for winning neuron index m , that is,

$$\mathbf{X}_m = \{\mathbf{x}_i \mid \arg \mathbf{w}_j(\mathbf{x}_i) = m\}, \quad (\text{A-22})$$

of which there are, say, n members

$$\mathbf{X}_m = \{\mathbf{x}_1 \dots \mathbf{x}_n\}. \quad (\text{A-23})$$

For subset m , the sum distance

$$u_m = \sum_i^n \|\mathbf{x}_i - \mathbf{w}_j(\mathbf{x}_i)\| \mid \mathbf{x}_i \in \mathbf{X}_m, \quad (\text{A-24})$$

and sum squared distance

$$v_m = \sum_i^n [|\mathbf{x}_i - \mathbf{w}_j(\mathbf{x}_i)|] \mid \mathbf{x}_i \in \mathbf{X}_m. \quad (\text{A-25})$$

The mean distance of the winning neurons in subset m is estimated as

$$\hat{d}_m = \frac{1}{n} u_m. \quad (\text{A-26})$$

Some data points lie close to the winning neuron, whereas others are far away as measured by Euclidean distance. For a particular winning neuron, the variance of distance for this set of n samples of $\mathbf{x}$ estimates the spread of data in the subset by

$$\hat{\sigma}_m^2 = [v_m - u_m^2/n]/(n-1). \quad (\text{A-27})$$

The variance of the distance distribution may be found for each of the J winning neurons. Moreover, some winning neurons will have a small variance, whereas others will be large. This characteristic tells us about the spread of the natural cluster near the winning neuron.

Classification probability attribute

Classification probability estimates the probable certainty that a winning neuron classification is successful. To assess a successful classification, we make several assumptions

- SOM training has completed successfully with winning neurons that truly reflect the natural clustering in the data.
- The number and distribution of neurons are adequate to describe the natural clustering.
- A good measure of successful classification is the distance between the data sample and its winning neuron.
- The number of data samples in a subset is adequate for reasonable statistical estimation.
- Although the distribution of distance between the data sample and winning neuron is not necessarily Gaussian, statistical measures of central tendency and spread are described adequately.

Certainly, if the winning neuron is far from a data sample, the probability of a correct classification is lower than one, where the distance is small. After the variance of the set has been found with equations A-24, A-25, and A-27, assigning probability to a classification is straightforward. Consider a particular subsample $\mathbf{x}_i \in \mathbf{X}_m$ in equation A-23, whose winning neuron has a Euclidean distance d_i . Then, the probability of fit for that winning neuron classification is

$$p(z) = 2(2\pi)^{-1/2} \int_0^z \exp(-v^2/2) dv = 1 - \operatorname{erf}\left(z/\sqrt{2}\right), \quad (\text{A-28})$$

where $z = |d_i - \hat{d}_m|/\sigma_m$ from equations A-26 and A-27.

In summary, we illustrate with equations A-20 and A-21 that the winning neuron of every data sample is a member of a winning neuron index subset and every data sample has a winning neuron. This means that every data sample has available to it the following:

- a winning neuron with an index that may be shared with other data samples, which together is a classification attribute subset

- a sum distance and sum square distance between the sample and its winning neuron computed with equations A-24 and A-25 for that subset
- and then also mean distance and variance computed with equations A-26 and A-27 for that subset.
- resulting in a probability of classification fit of the neuron to the data sample, which is found with equation A-28.

In the same fashion as with the set of winning neuron index k in equation A-17, the probability of a successful winning neuron classification for each of I samples of $\mathbf{x}$ with equation A-28 is shown as

$$\mathbf{P} = \{p_i(u_k, v_k, \hat{d}_k, \hat{\sigma}_k) | k = \arg \mathbf{w}_j(\mathbf{x}_i) \forall \mathbf{w}_j \in \mathbf{W} | \forall \mathbf{x}_i \in \mathbf{X}\}, \quad (\text{A-29})$$

which we simplify as members

$$\mathbf{P} = \{p_1 \dots p_I\}. \quad (\text{A-30})$$

Probability $\mathbf{P}$ is a complete survey, so it, too, is an SOM survey attribute.

These probability samples are a 3D probability survey attribute as

$$\mathbf{P} = \{p_{c,d,e}\}, \quad (\text{A-31})$$

where c is the time or depth sample index, d is the trace number, and e is the line number.

Anomalous geobody attributes

It is simple to investigate where in the survey winning neurons are successful and where they are not. We select a probability distance cut-off γ , say, $\gamma = 0.1$. Then, we nullify any winning neuron classification and its corresponding probability in their respective attribute surveys, which have a probability below the cut-off. Mathematically, from equation A-30, we have

$$\mathbf{P}_{\text{cut}} \subset \mathbf{P}, \quad (\text{A-32})$$

$$\mathbf{P}_{\text{cut}} = \{p_i = 0 | p_i < \gamma \forall p_i \in \mathbf{P}\}, \quad (\text{A-33})$$

where $\mathbf{P}$ divides into two parts, one part that falls less than the threshold γ and the other part that does not:

$$\mathbf{P}^*(\gamma) = \mathbf{P}_{\text{cut}} \cup \mathbf{P}_{\text{rest}}, \quad (\text{A-34})$$

where $\cup$ is the union of the two subsets and reads “and.” In terms of indices of a 3D survey, we have

$$\mathbf{P}^*(\gamma) = \{p_{c,d,e}^*\} = \{0 | p_{c,d,e} < \gamma\} \cup \{p_{c,d,e} | p_{c,d,e} \geq \gamma\}. \quad (\text{A-35})$$

We note that $\mathbf{P}^*(\gamma)$ is another 3D attribute survey with particularly interesting features apart from the classification and classification probability attributes previously discussed. We report in [Taner et al. \(2009\)](#) and confirm with SOM analysis of many later surveys that there are continuous regions in $\mathbf{P}^*(\gamma)$ of small probability values. These regions have been identified without human assistance, yet because of their extended volumetric extent cannot easily be explained as statistical coincidence. Rather, we recognize some of them as contiguous geobodies of geologic interest. These regions of low probability are an indication of poor fit and may be anomalies for further investigation.

The previous discussion computes an anomalous geobody attribute based on regions of low classification probability. We next turn to inspect the 3D survey for geobodies, where adjacent samples of low probability share the same neuron index. In an analogous fashion to equations A-32–A-34, we define the following:

$$\mathbf{K}_{\text{cut}} \subset \mathbf{K}, \quad (\text{A-36})$$

$$\mathbf{K}_{\text{cut}} = \{k_i = 0 | p_i < \gamma \forall p_i \in \mathbf{P} \cap k_i \in k_i^*\}, \quad (\text{A-37})$$

where k_i^* is a member of a connected region of the same classification index

$$\mathbf{K}^*(\gamma) = \mathbf{K}_{\text{cut}} \cup \mathbf{K}_{\text{rest}}. \quad (\text{A-38})$$

Attribute $\mathbf{K}^*(\gamma)$ is more restrictive than $\mathbf{P}^*(\gamma)$ because this one requires low probability and contiguous regions of single-neuron indices. The values $\mathbf{P}^*(\gamma)$ and $\mathbf{K}^*(\gamma)$ constitute additional attribute surveys for analysis and interpretation.

Three measures of goodness of fit

A good way to estimate the success of an SOM analysis is to measure the fraction of successful classifications for a certain choice of γ . The fraction of successful classifications for a particular choice of cut-off is the value

$$\check{M}_c(\mathbf{P}^* | \gamma) = \frac{1}{I} \sum_i^I \delta[p_i^*]. \quad (\text{A-39})$$

We use the letter M for goodness-of-fit measures. The check accent is used for counts and entropies (later).

An additional measure of fitness is the amount of entropy that remains in the probabilities of the classification above the cut-off threshold. Recall that entropy is a measure of degree of disorder ([Bishop, 2007](#)). First, we estimate a probability distribution function (PDF) for the probability survey by a histogram of $\mathbf{P}^*(\gamma)$. If the PDF is narrow, entropy is low, and if the PDF is broad, entropy is high. We view entropy as a measure of variety.

Let the number of bins be n bins, then the bin width

$$\Delta h = (1 - \gamma)/n\text{bins}. \quad (\text{A-40})$$

The count of probabilities in bin j of a probability histogram of $\mathbf{P}^*$ is

$$h_j(\mathbf{P}^*|\gamma) = \sum_i^I \delta[p_i^*|(j-1)\Delta h \leq p_i^* - \gamma < j\Delta h \forall p_i^* \in \mathbf{P}^*]. \quad (\text{A-41})$$

So the histogram of probability is the set

$$\mathbf{H}(\mathbf{P}^*|\gamma) = \{h_j(\mathbf{P}^*|\gamma) \forall j\}. \quad (\text{A-42})$$

The entropy of the probability PDF in equations A-40 through A-42 is

$$\check{M}_e[\mathbf{H}(\mathbf{P}^*|\gamma)] = -\frac{1}{n\text{bins}} \sum_i^{n\text{bins}} h_i(\mathbf{P}^*) \ln h_i(\mathbf{P}^*). \quad (\text{A-43})$$

We similarly find the PDF of classification indices and then the entropy of that PDF with histograms of counts of neuron above the cut-off of equation A-38 as

$$\check{M}_e[\mathbf{H}(\mathbf{K}^*|\gamma)] = -\frac{1}{n\text{bins}} \sum_i^{n\text{bins}} h_i(\mathbf{K}^*) \ln h_i(\mathbf{K}^*). \quad (\text{A-44})$$

When the probability distribution is concentrated in a narrow range, $\check{M}_e$ is small. When p is uniform, $\check{M}_e$ is large. We assume that a broad uniform distribution of probability indicates a neuron that fits a broad range of attribute values.

The entropy of a set of histograms is mean entropy. Consider first the PDF of distances for the subset of data samples associated with a winning neuron of a classification subset. The set of distances in subset m

$$\mathbf{D}_m = \{d_m\} = \{\|\mathbf{x}_i - \mathbf{w}_j(\mathbf{x}_i)\| \mid \mathbf{x}_i \in \mathbf{X}_m\}, \quad (\text{A-45})$$

where $\mathbf{X}_m$ is given in equation A-23. Then, the mean entropy of distance (total) is

$$\bar{M}(\mathbf{D}) = -\frac{1}{J n\text{bins}} \sum_j^J \sum_i^{n\text{bins}} h_i(\mathbf{D}_m) \ln h_i(\mathbf{D}_m). \quad (\text{A-46})$$

Notice that for ensemble averages, we use the bar accent. A set of broad distributions leads to a larger ensemble entropy average, and that is preferred.

It is important to have tools to compare several SOM analyses to assess which one is better. For example, several analyses may offer different learning parameters and/or different convergence criteria.

For comparing different SOM analyses, measurements $\hat{M}_e(\mathbf{P}^*|\gamma)$ and $\bar{M}_e(\mathbf{D})$ offer unequivocal tests of the best analysis.

The ensemble mean measure is also applied to attributes of winning neuron subsets. Recall that there are a total of F attributes, so we begin by mathematically isolating attributes in data samples. Note that there are a set of F attributes at a single location in a 3D survey at time or depth index c , trace index d , and line index e . The set of F attribute samples at fixed point i in the survey are

$$\{\mathbf{a}_{i,f}\} = \{\mathbf{x}_{c,d,e,f} \mid c, d, e, f \text{ fixed}\}, \quad (\text{A-47})$$

where the relation between reference index i and survey indices c , d , and e is given as $i = c + (d-1)D + (e-1)CD$.

We further manipulate the data samples by creating a set of attribute samples that are apart from the multiattribute data samples. The set of I samples of attribute k is

$$\{\mathbf{a}_{i,k}\} = \{\mathbf{x}_{c,d,e,f} \mid f = k\}. \quad (\text{A-48})$$

Reducing equation A-47 for attribute k from the full set to subset m , the set of attribute samples is

$$\mathbf{A}_{m,k} = \{\mathbf{a}_{i,k} \mid \arg \mathbf{w}_j(\mathbf{x}_i) = m \in \mathbf{x}_{c,d,e,f} \mid f = k\}. \quad (\text{A-49})$$

With the help of equation A-49, we have the ensemble mean entropy for attribute k as

$$\bar{M}_e(\mathbf{A} \mid \text{attribute } k) = \frac{1}{J} \sum_j^J \check{M}_e[\mathbf{H}(\mathbf{A}_{j,k})]. \quad (\text{A-50})$$

This allows us to evaluate and compare entropies of each attribute, which we find to be a useful tool to evaluate the learning performance of a particular SOM training.

We turn next to another aspect of goodness of fit as estimated from PDFs. Consider a PDF that does not have a peak anywhere in the distribution. If so, the PDF does not contain a “center” with mean and standard deviation. Certainly, such a distribution is of little statistical value.

Define a curvature indicator as a binary switch that is one if the histogram has at least one peak and zero if it has none. We define the curvature indicator of any histogram $\mathbf{H}$ as

$$c(\mathbf{H}) = \delta[\max h_i \mid h_{i-1} \langle h_i \rangle h_{i+1} \forall h_{1 < i < n\text{bins}} \in \mathbf{H}]. \quad (\text{A-51})$$

This leads to the definition of curvature measure as an indicator for a natural cluster suitable for higher dimensions. Here, the curvature measure is taken over a set of attribute histograms. Let the number of histograms be $n\text{histo}$, and then the mean curvature measure for the set of attribute PDFs

$$\bar{M}_{\text{cm}}(\mathbf{A} \mid \text{attribute } k) = \frac{1}{n\text{histo}} \sum_i^{n\text{histo}} c(\mathbf{H}_i(\mathbf{A}_{i,k})). \quad (\text{A-52})$$

The curvature measure is an indicator of attribute PDF robustness. If all the PDFs have at least one peak $\bar{M}_{\text{cm}} = 1$, indicating that all the PDFs have at least one peak. The case that $\bar{M}_{\text{cm}} < 0.8$ would be suspicious.

The curvature measure of all distance PDFs is found by forming a histogram of each PDF and then finding the curvature indicator. The curvature measure of distance for the entire set of classification distances from equation A-45 is summed over J subsets of D as

$$\overline{M}_{cm}(\mathbf{D}) = \frac{1}{J} \sum_j c[\mathbf{H}(\mathbf{D}_j)]. \quad (\text{A-53})$$

With the help of equation A-49, we write that the mean curvature measure for all PDFs of attribute k subset for the J winning neurons is

$$\overline{M}_{cm}(\mathbf{A}|\text{attribute } k) = \frac{1}{J} \sum_j c[\mathbf{H}(\mathbf{A}_{j,k})]. \quad (\text{A-54})$$

The curvature measure of an attribute is a valuable indicator of the validity of the attribute. The entropy of each attribute is similarly formulated. A mean curvature measure of at least 0.9 indicates a robust set of attributes.

In summary, we have presented in this section three measures of goodness of fit and suggest how they assist an assessment of SOM analysis:

- fraction of successful classifications, equation A-39
- entropy of classification distance, equations A-43 and A-44, mean entropy of distance equation A-46, and ensemble mean entropy, equation A-50
- curvature measures of distance, equation A-53 and attributes, equation A-54.

References

- Abele, S., and R. Roden, 2012, Fracture detection interpretation beyond conventional seismic approaches: Presented at AAPG Eastern Section Meeting, Poster AAPG-ICE.
- Balch, A. H., 1971, Color sonograms: A new dimension in seismic data interpretation: *Geophysics*, **36**, 1074–1098, doi: [10.1190/1.1440233](https://doi.org/10.1190/1.1440233).
- Barnes, A., 2006, Too many seismic attributes?: *CSEG Recorder*, **31**, 41–45.
- Bishop, C. M., 2006, Pattern recognition and machine learning: Springer, 561–565.
- Bishop, C. M., 2007, Neural networks for pattern recognition: Oxford, 240–245.
- Bishop, C. M., M. Svensen, and C. K. I. Williams, 1998, GTM: The generative topographic mapping: *Neural Computation*, **10**, 215–234, doi: [10.1162/089976698300017953](https://doi.org/10.1162/089976698300017953).
- Brown, A. R., 1996, Interpretation of three-dimensional seismic data, 3rd ed.: AAPG Memoir 42.
- Chen, Q., and S. Sidney, 1997, Seismic attribute technology for reservoir forecasting and monitoring: The Leading Edge, **16**, 445–448, doi: [10.1190/1.1437657](https://doi.org/10.1190/1.1437657).
- Chopra, S., and K. Marfurt, 2007, Seismic attributes for prospect identification and reservoir characterization, SEG, Geophysical Development Series.
- Coleou, T., M. Poupon, and A. Kostia, 2003, Unsupervised seismic facies classification: A review and comparison of techniques and implementation: The Leading Edge, **22**, 942–953, doi: [10.1190/1.1623635](https://doi.org/10.1190/1.1623635).
- Erwin, E., K. Obermayer, and K. Schulten, 1992, Self-organizing maps: Ordering, convergence properties and energy functions: *Biological Cybernetics*, **67**, 47–55, doi: [10.1007/BF00201801](https://doi.org/10.1007/BF00201801).
- Fraser, S. H., and B. L. Dickson, 2007, A new method of data integration and integrated data interpretation: Self-organizing maps, in B. Milkereit, ed., Proceedings of Exploration 07: Fifth Decennial International Conference on Mineral Exploration, 907–910.
- Guo, H., K. J. Marfurt, and J. Liu, 2009, Principal component spectral analysis: *Geophysics*, **74**, no. 4, P35–P43, doi: [10.1190/1.3119264](https://doi.org/10.1190/1.3119264).
- Haykin, S., 2009, Neural networks and learning machines, 3rd ed.: Pearson.
- Kalkomey, C. T., 1997, Potential risks when using seismic attributes as predictors of reservoir properties: The Leading Edge, **16**, 247–251, doi: [10.1190/1.1437610](https://doi.org/10.1190/1.1437610).
- Kohonen, T., 2001, Self organizing maps: Third extended addition, Springer, Series in Information Services.
- Liner, C., 1999, Elements of 3-D seismology: PennWell.
- Roy, A., B. L. Dowdell, and K. J. Marfurt, 2013, Characterizing a Mississippian tripolitic chert reservoir using 3D unsupervised and supervised multi-attribute seismic facies analysis: An example from Osage County, Oklahoma: *Interpretation*, **1**, no. 2, SB109–SB124, doi: [10.1190/INT-2013-0023.1](https://doi.org/10.1190/INT-2013-0023.1).
- Rutherford, S. R., and R. H. Williams, 1989, Amplitude-versus-offset variations in gas sands: *Geophysics*, **54**, 680–688, doi: [10.1190/1.1442696](https://doi.org/10.1190/1.1442696).
- Schlumberger Oilfield Glossary, 2015, online reference, <http://www.glossary.oilfield.slb.com/>.
- Shuey, R. T., 1985, A simplification of the Zoeppritz equations: *Geophysics*, **50**, 609–614, doi: [10.1190/1.1441936](https://doi.org/10.1190/1.1441936).
- Smith, T., and M. T. Taner, 2010, Natural clusters in multi-attribute seismics found with self-organizing maps: Source and signal processing section paper 5: Presented at Robinson-Treitel Spring Symposium by GSH/SEG, Extended Abstracts.
- Taner, M. T., 2003, Attributes revisited, http://www.rocksolidimages.com/pdf/attrib_revisited.htm, accessed 13 August 2013.
- Taner, M. T., F. Koehler, and R. E. Sheriff, 1979, Complex seismic trace analysis: *Geophysics*, **44**, 1041–1063, doi: [10.1190/1.1440994](https://doi.org/10.1190/1.1440994).
- Taner, M. T., and R. E. Sheriff, 1977, Application of amplitude, frequency, and other attributes, to stratigraphic and hydrocarbon determination, in C. E. Payton, ed., Applications to hydrocarbon exploration: AAPG Memoir 26, 301–327.
- Taner, M. T., S. Treitel, and T. Smith, 2009, Self-organizing maps of multi-attribute 3D seismic reflection surveys, Presented at the 79th International SEG Convention,

Rocky Roden received a B.S. in oceanographic technology-geology from Lamar University and an M.S. in geological and geophysical oceanography from Texas A&M University. He runs his own consulting company, Rocky Ridge Resources Inc., and works with numerous oil companies around the world on interpretation

technical issues, prospect generation, risk analysis evaluations, and reserve/resource calculations. He is a senior consulting geophysicist with Geophysical Insights helping to develop advanced geophysical technology for interpretation. He is a principal in the Rose and Associates DHI Risk Analysis Consortium, developing a seismic amplitude risk analysis program and worldwide prospect database. He has also worked with Seismic Micro-Technology and Rock Solid Images on the integration of advanced geophysical software applications. He has been involved in numerous oil and gas discoveries around the world and has extensive knowledge on modern geoscience technical approaches (past Chairman — *The Leading Edge* Editorial Board). As Chief Geophysicist and Director of Applied Technology for Repsol-YPF (retired 2001), his role comprised advising corporate officers, geoscientists, and managers on interpretation, strategy and technical analysis for exploration and development in offices in the United States, Argentina, Spain, Egypt, Bolivia, Ecuador, Peru, Brazil, Venezuela, Malaysia, and Indonesia. He has been involved in the technical and economic evaluation of Gulf of Mexico lease sales, farmouts worldwide, and bid rounds in South America, Europe, and the Far East. Previous work experience includes exploration and development at Maxus Energy, Pogo Producing, Decca Survey, and Texaco. Professional affiliations include SEG, GSH, AAPG, HGS, SIPES, and EAGE.

Thomas Smith received a B.S. (1968) in geology and an M.S. (1971) in geology from Iowa State University, and a Ph.D. (1981) in geophysics from the University of Houston. He founded and continues to lead Geophysical Insights, a company advancing interpretation practices through more effective data analysis. His research

interests include quantitative seismic interpretation, integrated reservation evaluation, machine learning of joint geophysical, and geologic data and imaging. He and his wife, Evonne, founded Seismic Micro-Technology (1984) to develop one of the first PC seismic interpretation packages. He received the SEG Enterprise Award in 2000. He is a founding sustaining trustee associate of the SEG Foundation (2014), served on the SEG Foundation Board (2010–2013) and was its Chair (2011–2013). He is an honorary member of the GSH (2010). He was as an exploration geophysicist at Chevron (1971–1981), consulted on exploration problems for several worldwide oil and gas companies (1981–1995) and taught public and private short courses in seismic acquisition, data processing, and interpretation through PetroSkills (then OGC 1981–1994). He is a member of SEG, GSH, AAPG, HGS, SIPES, EAGE, SSA, AGU, and Sigma Xi.

Deborah Sacrey is a geologist/geophysicist with 39 years of oil and gas exploration experience in the Texas and Louisiana Gulf Coast and Mid-Continent areas of the United States. She received a degree in geology from the University of Oklahoma in 1976 and immediately started working for Gulf Oil in their Oklahoma City

offices. She started her own company, Auburn Energy, in 1990 and built her first geophysical workstation using Kingdom software in 1995. She helped SMT/IHS for 20 years in developing and testing the Kingdom Software. She specializes in 2D and 3D interpretation for clients in the United States and internationally. For the past three years, she has been part of a team to study and bring the power of multiattribute neural analysis of seismic data to the geoscience public, guided by Tom Smith, founder of SMT.

She has been very active in the geological community. She is a past national president of Society of Independent Professional Earth Scientists (SIPES), past president of the Division of Professional Affairs of American Association of Petroleum Geologists (AAPG), past treasurer of AAPG and is now the president of the Houston Geological Society. She is also a DPA certified petroleum geologist #4014 and DPA certified petroleum geophysicist #2. She is a member of SEG, AAPG, PESA (Australia), SIPES, Houston Geological Society, and the Oklahoma City Geological Society.