^{GC}Some Machine Learning Applications in Seismic Interpretation*

Satinder Chopra[1], David Lubo-Robles[2], and Kurt J. Marfurt[2]

[1]Arcis Seismic Solutions, TGS, Calgary, Canada (schopra@arcis.com)
[2]University of Oklahoma, Norman, Oklahoma, USA

General Statement

"Big data" and "data analytics" are the buzzwords these days. The oil and gas industry has always had large volumes of data to acquire, process and interpret, and since the introduction of 3-D and 4-D seismic data acquisition, the handling of large quantities of data has only become more challenging. As our industry moved from large mainframe computers coupled with array processors to scalable multiprocessors for crunching large volumes of data, seismic software, data storage and visualization capabilities have been able to keep pace.

But in the last decade, our industry grappled not only with ever-larger volumes of data, but also with increased data heterogeneity. Fortunately, advancements in handling such large, heterogeneous, "big data" volumes have come along. Recent developments in data analytic capabilities applied to other industries hold significant promise for those of working in the hydrocarbon exploration and development. "Data analytics" refers to a special class of analytical tools or methods that are used to study complex systems, many of which are not amenable to traditional analysis, such as multivariant statistics.

Deductive versus Inductive

To better understand traditional interpretation versus data analytic workflows, we need to distinguish two terms: "deductive" and "inductive" reasoning. Using logic or reason to form a conclusion or opinion about something is deductive, whereas using examples to reach a general conclusion about something is inductive.

Interpreters routinely use deductive reasoning, analyzing the data using principles of geology, physics and petrophysics. Examples might be as simple as constructing synthetics to tie a well log to seismic data, or as complex as defining the environment of deposition using pattern recognition and modern analogues. There are two limitations to this approach. The first is that, try as we will, we may not be able to understand

the physical reasons why one area of a survey is more productive, or alternatively completes better, than another. The second is that we may simply not have enough time to carefully correlate multiple attribute volumes using principles of physics and geology.

In contrast, data analytics uses inductive reasoning to find patterns between multidimensional data volumes. Petrophysical analysis tells us that there is a theoretical basis for porosity to correlate to P-impedance and is an example of deductive theory-based reasoning. In contrast, if there is a statistically significant correlation between TOC and P-impedance for multiple wells in a specific play, and if we can successfully validate this correlation on new wells, we have an example of transductive (good for a limited number of data sets only) or inductive (good for most data sets) databased reasoning. Often, we do not know the reason behind a good correlation but given significant validation, we can use it as a statistically valid prediction tool. In other cases, the correlations identify a feature that can be explained by an already established theory. In still other cases, the correlations allow us to formulate a hypothesis based on physics or geology that, with further validation, can lead to a new theory.

Supervised and Unsupervised Learning

Machine learning algorithms can be broken into supervised and unsupervised learning subsets. Supervised learning is perhaps the easier subset to understand. Here, the interpreter provides training data, or "labels," to the algorithm in addition to multiple seismic attribute volumes. Common labels include the names of seismic facies described by interpreter-constructed polygons or of assignment of voxels along a well bore to measured lithology, geomechanical behavior, or fracture intensity. Key to supervised learning is selecting attributes that differentiate the feature of interest from the background geology.

As in cross plotting electric log properties, a shortcoming of supervised learning is that it will only search for explicitly defined features, such as carbonate versus dolomite versus shale. If there is also anhydrite in the system, it may be misclassified into one of the defined classes. Common machine learning techniques include:

- Decision trees

- Multilinear feedforward neural networks

- Probabilistic neural networks

- Support vector machines

Unsupervised learning is slightly more difficult to understand. Here, the training data are a random set of voxels drawn from multiple attribute volumes themselves. The objective is to find patterns that in some measure represent the bulk of the data. A point of confusion is that most interpreters think of patterns as a reflectivity pattern seen on vertical, horizontal or horizon slices. These latter structural and spectral patterns are measured by seismic attributes. Rather, in unsupervised learning, the "patterns" are measured across multiple attribute volumes at a given voxel. For example, a salt dome might be represented by the four-dimensional attribute pattern of low-coherence, high entropy, low envelope

and low reflector parallelism, while conformal sand/shale reflectors might be represented by high coherence, low entropy, moderate to high envelope and high reflector parallelism.

Not all the tools in the data analytics component analysis, self-organizing mapping, fuzzy logic, support vector machines, neural networks, etc. have all been used in interpretation for some 20 years, but in relatively focused applications, such as multi-attribute analysis along a picked horizon, or multi-log analysis along a small set of wells. The major limitation has been the "big" part of our modern data. The advent of multicore desktops machines, graphical processor units and interpreter access to supercomputers previously limited to seismic imaging and flow simulation, along with advances in software development now allow the analysis of large data volumes.

Humans Still Needed

A common misconception is that machine learning will replace human interpreters. The most common use of decision tree-based machine learning is a horizon autopickers. Autopickers have been in use for 20 years – yet each horizon needs to be examined and usually modified by a human interpreter.

First-break pickers for statics corrections have used neural networks for at least 10 years. Here, the human processor needs to quality control the results and add additional control (or corrections) where needed. The role of the interpreter will change from mundane picking to evaluating alternative hypotheses and evaluating the results. It is a pity that, these days, we see and hear about expert knowledge getting phased out by way of chosen or forced retirements due to the economic downturn from which we are still recovering. Can we somehow capture this expertise as part of a rule-based machine learning application? If so, data analytics applications on big data, where the machine learns from the human quality control, and where the interpreter poses new hypotheses, is the future for your industry.

There are several ways of combining multiple attributes, with visualization in red-green-blue (RGB) color space, coupled with transparency as one of the more powerful means. Unfortunately, such a color display is limited to three, and with transparency, four attributes. One of the methods commonly used for this purpose is principal component analysis, and a more recent one is independent component analysis. Both these methods 'churn' the different attributes and yield one, two or three volumes that represent the maximum variation in the input attributes. Such analysis reduces the redundancy in the input attributes. We present the results of our investigation into the application of both these methods on a seismic data volume from central Alberta, Canada.

Machine Learning Tools

Principal component analysis is a useful statistical technique that has found many applications, including image compression and pattern recognition in data of high dimensionality. We are familiar with the usual statistical measures like mean, standard deviation and variance, which are essentially one-dimensional. Such measures are calculated one attribute at a time with the assumption that each attribute is independent of the others. In reality, many of our attributes are coupled through the underlying geology, such that a fault may give rise to lateral changes in waveform, dip, peak frequency and amplitude. Less desirably, many of our attributes are coupled mathematically, such as alternative measures of coherence or of a suite of closely spaced spectral components. The amount of attribute redundancy is measured by the

covariance matrix. The first step in multiattribute analysis is to subtract the mean of each attribute from the corresponding attribute volume. If the attributes have radically different units of measure, such as frequency measured in Hertz, envelope measured in millivolts, and coherence without dimension, a Z-score normalization is required. Mathematically, the number of linearly uncorrelated attributes is defined by the value of eigenvalues and eigenvectors of the covariance matrix. The first eigenvector is a linear combination that represents the most variability in the scaled attributes. The corresponding first eigenvector represents the amount of variability represented. Commonly, each eigenvalue is normalized by the sum of all the eigenvalues, giving a percentage of the variability represented.

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

PCA is based on the statistical assumption that the input multivariate data exhibit a Gaussian distribution. Independent component analysis is an elegant machine learning technique that separates multivariate data into independent components, assuming that the data going into the analysis have non-Gaussian distribution. The other differences between ICA and PCA are that the independent components are not orthogonal, and their order is not defined, in that the first, second and third ICs are ordered by visual examination, and are not mathematically ordered in the process as in PCA.

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

Some of the seismic attributes commonly used for multi-attribute analysis are as follow:

● Discontinuity attributes: Coherence (see Multispectral Coherence Attribute Applications, Search and Discovery Article #42258.)

and curvature attributes are commonly used for interpreting faults, fractures, reef edges, channel edges, etc. Coherence, most-positive curvature (longwavelength), most-negative curvature (long-wavelength), most-positive curvature (short-wavelength), most-negative curvature (short-wavelength), are the commonly used discontinuity attributes.

● GLCM texture attributes (energy, entropy and homogeneity): GLCM (grey-level co-occurrence matrix) texture attributes are useful for the determination of seismic facies analysis. GLCM energy is a measure of textural uniformity in an image, GLCM entropy is a measure of disorder or complexity of the image, and GLCM homogeneity is a measure of the overall smoothness of the image. More information on these

attributes can be found in [Adding Texture Attributes to the 3-D Mix, Search and Discovery Article #41244](#) and [Extracting Information from Texture Attributes, Search and Discovery Article #42258](#).

● Spectral decomposition frequency attributes (spectral magnitude components, peak frequency and peak magnitude): Spectral decomposition refers to the transformation of seismic data into individual frequency components within the seismic bandwidth. The derived frequency data have found application for the interpretation of bed thickness, discontinuities and distinguishing fluids in the reservoirs. Spectral decomposition has been described extensively in [Spectral Decomposition's Analytical Value, Search and Discovery Article #41260](#), [Spectral Decomposition Helps Define Channel Morphology Search and Discovery Article #41272](#), [Effective Ways to Eliminate Side-Lobe Effects, Search and Discovery Article #41273](#), [Extracting Information from Texture Attributes, Search and Discovery Article #41330](#), [Spectral Decomposition for 3-D Geomodeling, Search and Discovery Article #41408](#), [Spectral Balancing Aids Seeing Faults More Clearly, Search and Discovery Article #41594](#), and [Spectral Decompostion for a More Accurate Image, Search and Discovery Article #41807](#).

Applications

The dataset chosen for this exercise is from central Alberta, Canada. We focus on the Mannville channels that are filled with interbedded units of shale and sandstone. On the 3-D seismic volume, these channels show up at the level indicated with a yellow arrow in [Figure 1](#).

The input attributes used for the principal component and independent component multivariate analysis are the multispectral coherence, GLCM-energy, GLCM-entropy, GLCM-homogeneity, spectral magnitudes at 30, 40 and 50 Hertz, and coherent energy. The stratal slices at the level of the yellow arrow in [Figure 1](#) are shown from the PCA and ICA in [Figure 2](#) and [Figure 3](#). The first, second and third components from both the methods are depicted as well as their co-rendered displays using RGB. Notice the second, third and co-rendered displays show crisper definition of the paleo channels for the independent components than on the principal components.

Conclusion

Thus, in conclusion, we state that while the data reduction in principal component and independent component analysis is powerful, the latter has an edge over the former. These smaller number of components can then be used in more sophisticated machine learning tools such as self-organizing mapping and generative topographic mapping. We will discuss the applications of these machine learning tools in another article sometime.
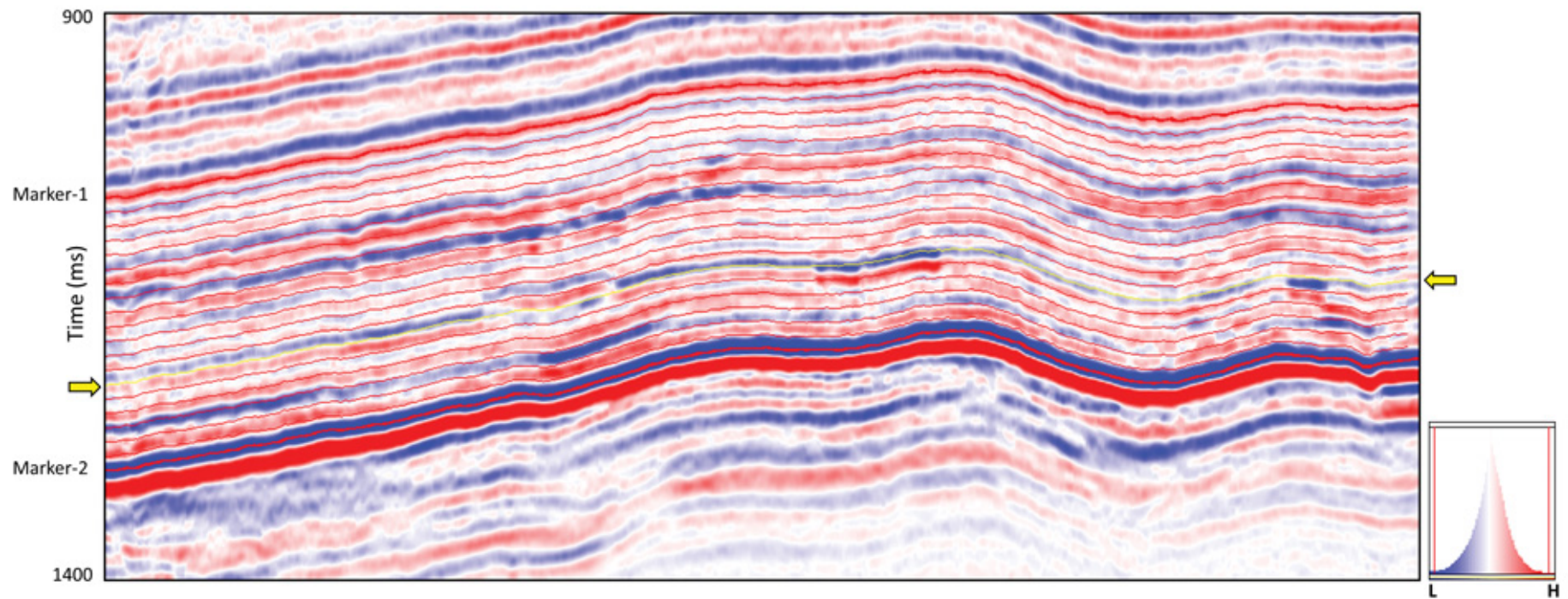
Figure 1. Segment of a section from a 3-D seismic volume from southcentral Alberta, Canada, which has two markers tracked on it. The strata grid shown overlaid was generated within these two markers and divided into 20 proportional slices. The stratal displays shown in Figure 2 and Figure 3 are along slice 14 indicated with the yellow arrows.
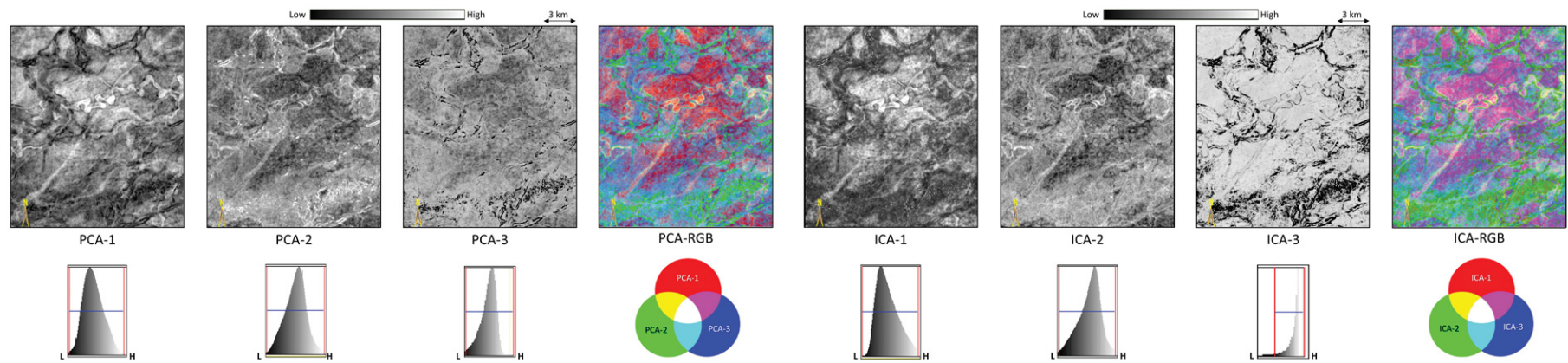
Figure 2 (left). Stratal slices from the first three principal components, and their co-rendered display using red (R), green (G) and blue (B) colors. Paleo channels of different dispositions and sizes are seen on the displays. Each of the first three displays exhibit somewhat different information. The RGB display puts it all together. Figure 3 (right). Stratal slices from the first three independent components, and their co-rendered display using red (R), green (G) and blue (B) colors. The different paleo channels seen in figure 2 are seen here again, but with somewhat different definition, which is interestingly seen more pronounced in the ICA-3 display. The RGB display puts all the information together, and the overall channel definition seems to be better in the ICA-RGB display.