SPE-194889-MS

Deep Learning Method for Latent Space Analysis

Bradley Wallet, Aramco Services Company: Aramco Research Center - Houston; Thang Ha, University of Oklahoma

This paper was prepared for presentation at the SPE Middle East Oil and Gas Show and Conference held in Manama, Bahrain, 18-21 March 2019.

Abstract

Seismic attributes are a well-established method for highlighting subtle features buried in seismic data in order to improve interpretability and suitability for quantitative analysis. Seismic attributes are a critical enabling technology in such areas thin bed analysis, 3D geobody extraction, and seismic geomorphology. When it comes to seismic attributes, we often suffer from an "abundance of riches" as the high dimensionality of seismic attributes may cause great difficulty in accomplishing even simple tasks. Spectral decomposition, for instance, typically produces 10's and sometimes 100's of attributes. However, when it comes to visualization, for instance, we are limited to visualizing three or at most four attributes simultaneously.

My co-authors and I first proposed the use of latent space analysis to reduce the dimensionality of seismic attributes in 2009. At the time, we focused upon the use of non-linear methods such as self-organizing maps (SOM) and generative topological maps (GTM). Since then, many other researchers have significantly expanded the list of unsupervised methods as well as supervised learning. Additionally, latent space methods have been adopted in a number of commercial interpretation and visualization software packages.

In this paper, we introduce a novel deep learning-based approach to latent space analysis. This method is superior in that it is able to remove redundant information and focus upon capturing essential information rather than just focusing upon probability density functions or clusters in a high dimensional space. Furthermore, our method provides a quantitative way to assess the fit of the latent space to the original data.

We apply our method to a seismic data set from the Canterbury Basin, New Zealand. We examine the goodness of fit of our model by comparing the input data to what can be reproduced from the reduced dimensional data. We provide an interpretation based upon our method.

Introduction

Seismic attributes offer a powerful method for improving the interpretation process of seismic data. Seismic attributes enhance subtle features that are not readily apparent in the data. Seismic attributes extract information that is 3D in nature and not visible in seismic slices, emphasis coherent information, and allow for the extraction of spectral information. Barnes (2007) argued for selecting only attributes that were not correlated with one another. We would tend to disagree, believing that sometimes that difference variations in correlated attributes are meaningful. However, even if we restrict ourselves to minimally correlated, meaningful (task dependent) attributes, we will still have many attributes to choose from.

A common workflow would be to look at a number of different attributes in succession, interpreting those that made sense or were preferred by the interpreter. More advanced workflows could involve the multi-attribute visualization using color blending (Marfurt, 2015). Proper use of color blending is especially challenging, and expert level skills in both seismic attributes and visualization is often required to properly visual multiple attributes.

Beyond the issues with visualization, a large number of seismic attributes pose considerable mathematical complications. For instance, in the case of six attributes, mathematically modeling the original data set requires constructing models in a 6D vector space. Bellman's Curse of Dimensionality (Bellman, 1957; Wallet, 2013) dictates that all points in high dimensional space become outliers. Therefore, beyond just concerns around visualization, reducing the dimension of the attribute space is necessary for such tasks as pattern recognition and other machine learning applications.

Dimensionality reduction and latent space learning

A number of methods for doing dimensionality reduction of multiple attributes exists. The most basic and most common is attribute selection. As suggested above, the interpreter hand selects the attributes believed to be most useful. Hand selection has the advantage in that it forces the interpreter to give careful consideration to the choice of attributes. However, this method ignores the fact that the excluded attributes likely have useful information which could contribute to the task at hand.

Wallet (2013) proposed an interactive method of constructing linear combinations of attributes. This method has the advantage that it allowed the interpreter to consider spatial information when combining attributes. However, this method is labor intensive and relies heavily upon the skill and perception of the interpreter.

Most current work in reducing the dimensionality of attributes spaces focused upon latent space learning. The concept behind latent spaces is that while the attribute space is itself high dimensional, most of the associated probability mass lies in a lower dimensional, hidden or latent space. The challenge is then to learn a (possibly nonlinear) projection from the higher dimensional space to a lower dimensional space that preserves the embedded information.

Guo et al. (2006) used principal component analysis (PCA) to combine multiple attributes. This method is simple, often effective, and widely available in a large number of software platforms. Additionally, the reliance upon eigenvalue decomposition allows for a quantitative assessment of how much information was retained by the process. Furthermore, under the assumption of normality (Gaussian distributions), the retained information is statistically independent. However, PCA is limited by its linear nature, the definition of variance as information, and the failure of the normality assumption.

Self-organizing maps (SOM) (Coleou et al., 200x; Roy et al., 2013) is a commonly used method of dimensionality reduction that has been implemented in a number of commercially available seismic interpretation platforms. Conventionally considered a clustering algorithm, Wallet et al. (2009) view it as a latent space learning method where a mesh of nodes are deformed to form a lower dimensional surface approximating the input data in the higher dimensional space. This has proven to be a powerful method for dealing with multiple attributes. However, SOM lacks a measure of goodness-of-fit or other quantitative method for assessing the results. Additionally, computational limitations tend to restrict the projected space to no more than 2D.

There exist a number of other methods for doing latent space learning. Generative topological maps (GTM) (Wallet et al., 2009) model the latent space as a mesh of Gaussian terms and use an expectation-maximization (EM) algorithm (Dempster et al., 1977) to optimize the fit. EM results in a probability (generative) model. However, we are unaware of any work that exploits this fact, and GTM is not currently available in any commercial software. Diffusion maps (Wallet et al., 2014) performs eigen-analysis on a matrix of inter-point similarities. Unfortunately, this method is computationally intractable with modern computers even when dealing with moderate size seismic data volumes.

In this paper, we propose a new method of latent space learning for visualization of seismic data based upon a deep learning technique, autoencoders. Autoencoders remove redundant information while retaining information that is designed to reproduce the original, high dimensional data. In this way, the method removes information that is repetitive due to correlation while retaining independent information to the degree possible as dictated by the input data, the retained number of dimensions, and the topology of the underlying neural network.

Autoencoders

An autoencoder is a type of neural network algorithm aimed to compress (encode) data in an unsupervised manner (Liou et al., 2014). It consists of two steps: (1) compressing the input data domain into a smaller, lower-dimension encoded data domain, and (2) decompressing the encoded data domain back to the original data domain while trying to minimize the difference between the decompressed data and the input data. It has been widely used for data denoising (Vincent et al., 2010) and image recognition (Makhzani and Fray, 2013).

Key to the functioning of an autoencoder is a choke point, often called the code. The code consists of a layer with a limited number of nodes. If the output of these nodes allows for the reproduction of the input data then the information contained in the input data is adequately captured in this code. The network output at the code is thus a learned latent space for the input data, containing a lower dimensional representation of the higher dimensional data set.

Commonly, the code is used as an input into a follow-on neural network tasked with pattern recognition or other similar task. Though not commonly acknowledged in the literature, autoencoders produce a learned latent space, and the code could be used as input to other methods including clustering, probability density function (PDF) estimation, and visualization. Our workflow involves using an autoencoder with a 3D code. Each node of the code is associated with a primary color, red, green, and blue (Figure 1). In this way, the code can be visualized in image format as an RGB color image.

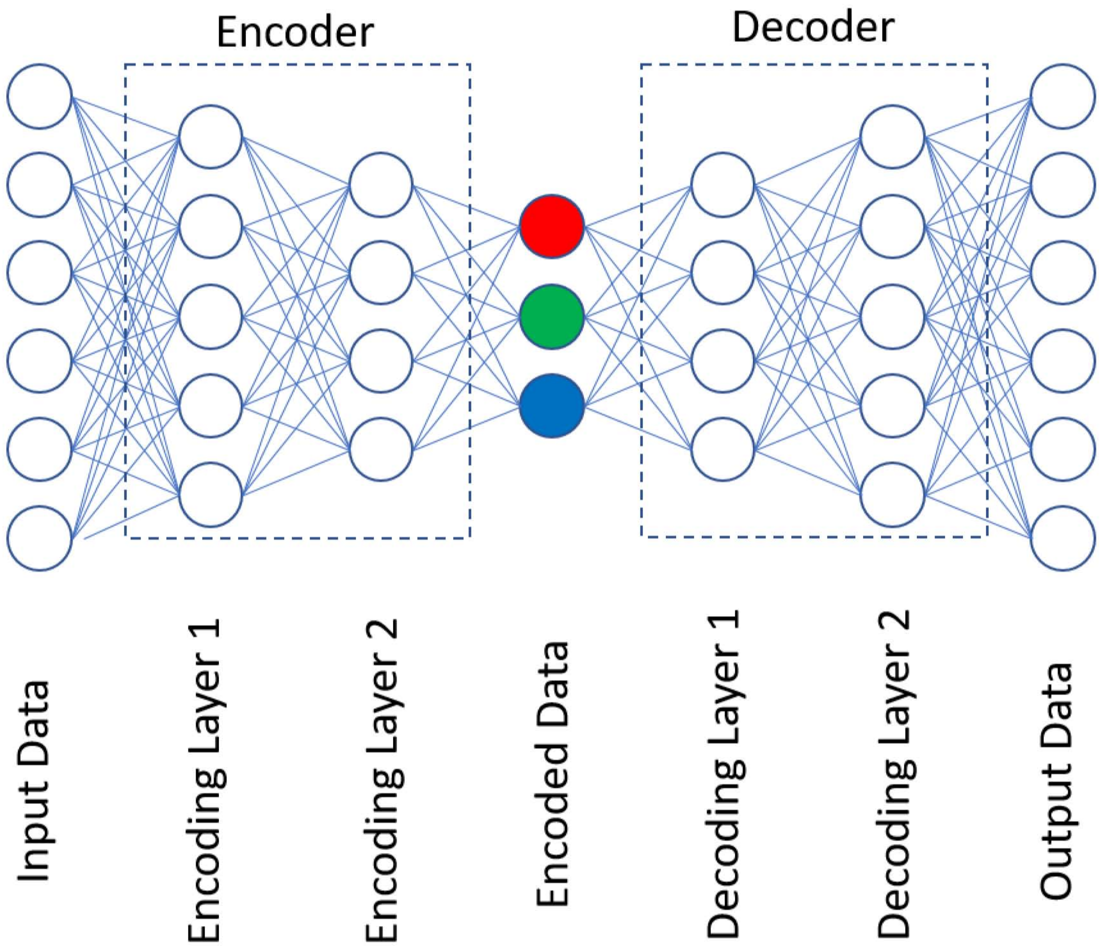

Figure 1—An example autoencoder with a three node code. Each node of the code is associated with a primary color, red, green, and blue.

Application and Discussion

To demonstrate the utility of our approach, we applied it to a seismic survey located on the Canterbury Basin, offshore New Zealand (Figure 2). The Canterbury Basin underwent multiple stages of rifting, passive subsidence, and minor uplift since the mid-Cretaceous (Sutherland and Browne, 2003). Sediments in the basin were deposited in a major transgressive-regressive cycle driven by tectonics (Zhao et al., 2016). During the middle to late Cretaceous, rifting and subsidence created the major structure of the Canterbury Basin, allowing a thick layer of clastic, coaly sediment to be deposited (Sutherland and Browne, 2003). From the late Cretaceous to the mid-Tertiary, the basin entered a transgression and was gradually filled with fluvial deposits, marine sandstone, and massive mudstones. Toward the mid-Tertiary, organic-rich black shales were deposited, followed by widespread limestones. The Canterbury Basin contains more than 6000ft of late Cretaceous to mid-Tertiary deposits (Cozens, 2011). From the late Tertiary to recent time, due to uplift and minor inversion in the NW, the basin entered a regression. Coarse clastic and shallow marine sediments were deposited in the northern and western parts of the basin, while mudstone was continued to be deposited in the eastern part of the basin (Sutherland and Browne, 2003). The seismic survey images the transition zone of the continental rise and continental slope and contains many paleo-canyons and turbidite deposits (Zhao et al., 2016).

In term of petrology, primary source rocks are late Cretaceous coaly sediments and mid-Tertiary organic-rich black shales. Reservoir rocks include late Cretaceous fluvial and marine sandstones, early Tertiary sandstones, and late Tertiary limestones. Prominent seal rocks include late Cretaceous fluvial mudstones and Tertiary regional marine mudstones (Sutherland and Browne, 2003).

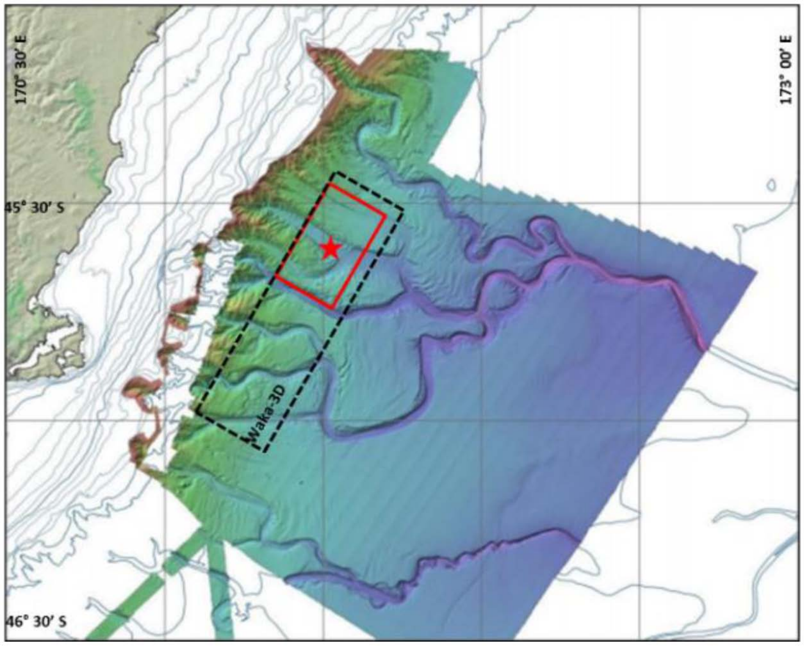

Figure 2—This map shows the turbidite channel system for a portion of the Canterbury Basin, New Zealand. The black rectangle shows the location of the Waka 3D while the red rectangle designates the portion of the survey used in this work. The colors represent the depth for a portion of the basin. (Zhao et al., 2016).

A Miocene turbidite system was interpreted using a phantom horizon slice tied to a picked continuous reflector below the turbidite system by Zhao et al., (2016). Figures 3-5 show the seismic data, the picked horizon, and seismic amplitudes extracted along a phantom horizon 200 ms deeper.

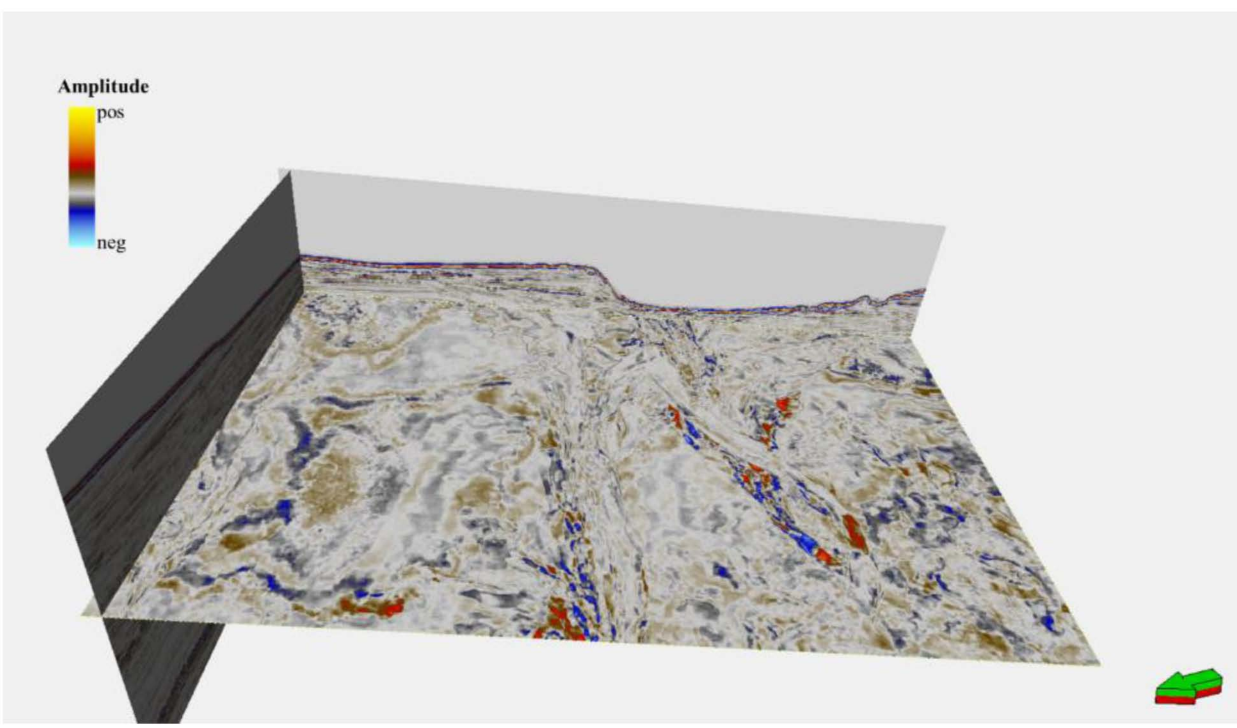

Figure 3—Time slice through the amplitude volume at t=1.88s looking downdip.

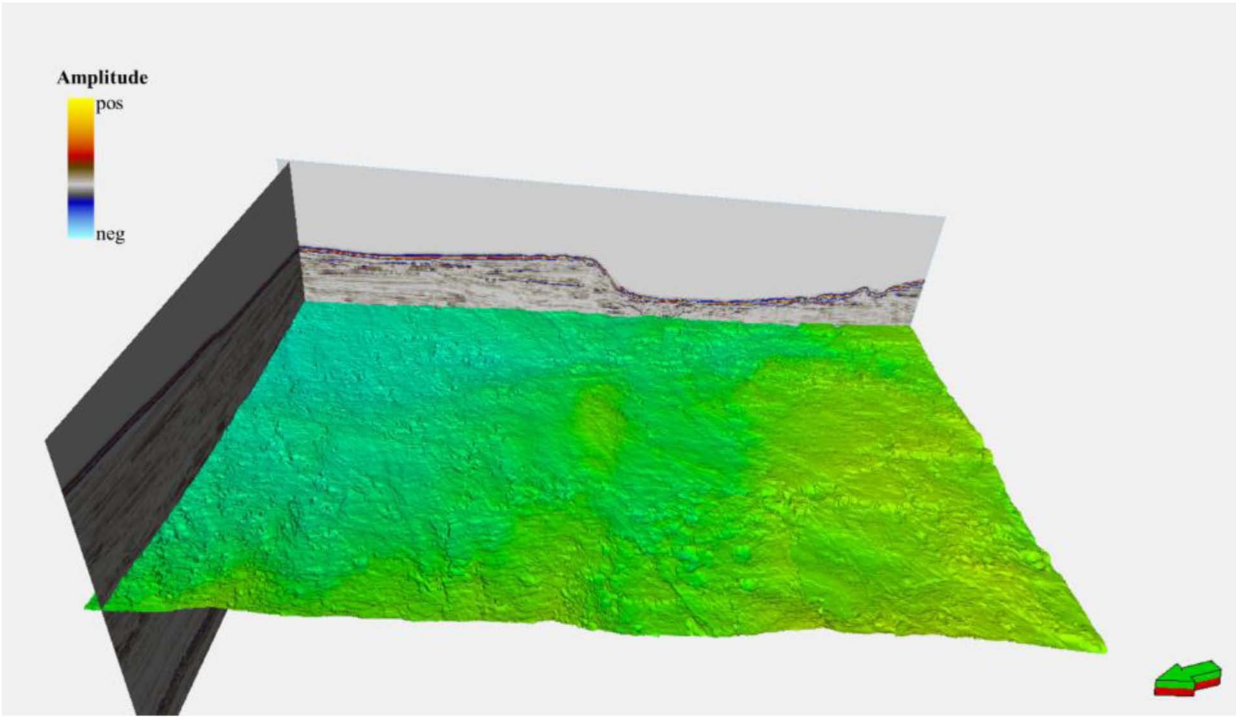

Figure 4—The picked horizon looking downdip.

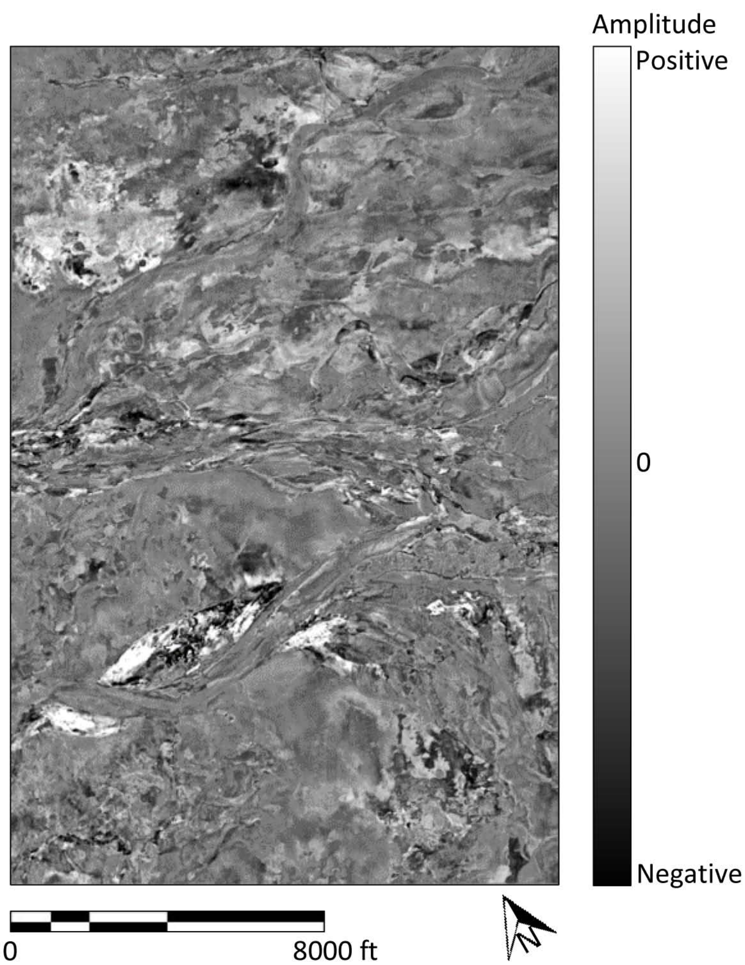

Figure 5—Phantom horizon slice 200 ms below the phantom horizon through the seismic amplitude volume.

Six seismic attributes, coherent energy, GLCM entropy, GLCM homogeneity, curvedness, peak frequency, and peak magnitude, were calculated and extracted for each point in the interpreted horizon. These were chosen for their known utility in mapping architectural elements in turbidite systems. In particular:

– Coherent energy is the energy of a 3-trace by 3-trace structure-oriented window of amplitude data, resulting in high values for strong, coherent reflectors.
– Gray Level Co-occurrence Matrix (GLCM) homogeneity is computed in a 5-trace by 5-trace structure-oriented window and measures the amplitude smoothness along structure. In contrast, GLCM entropy measures the randomness is lateral amplitude variation along structure.
– Curvedness is the magnitude of structural curvature, sensitive to both anticlinal and synclinal features, and is thus sensitive to both negative channel axes and positive channel levees.
– Peak frequency and peak magnitude are two spectral decomposition attributes that are sensitive to layer thickness.

For each voxel on the horizon, a 6D attribute vector was constructed, and an autoencoder was trained to reproduce the input data. The underlying neural network consisted of five hidden layers including a code of three nodes. The original attributes as well as the reconstructed data can be found in Figure 6 Figure 11. Examining these figures shows that the attributes were accurately recreated albeit with somewhat reduced contrast. This reduced contrast is a scaling issue, and it in no way impacts the contained information. We could have chosen to rescale the images, but we did not for the purposes of comparison.

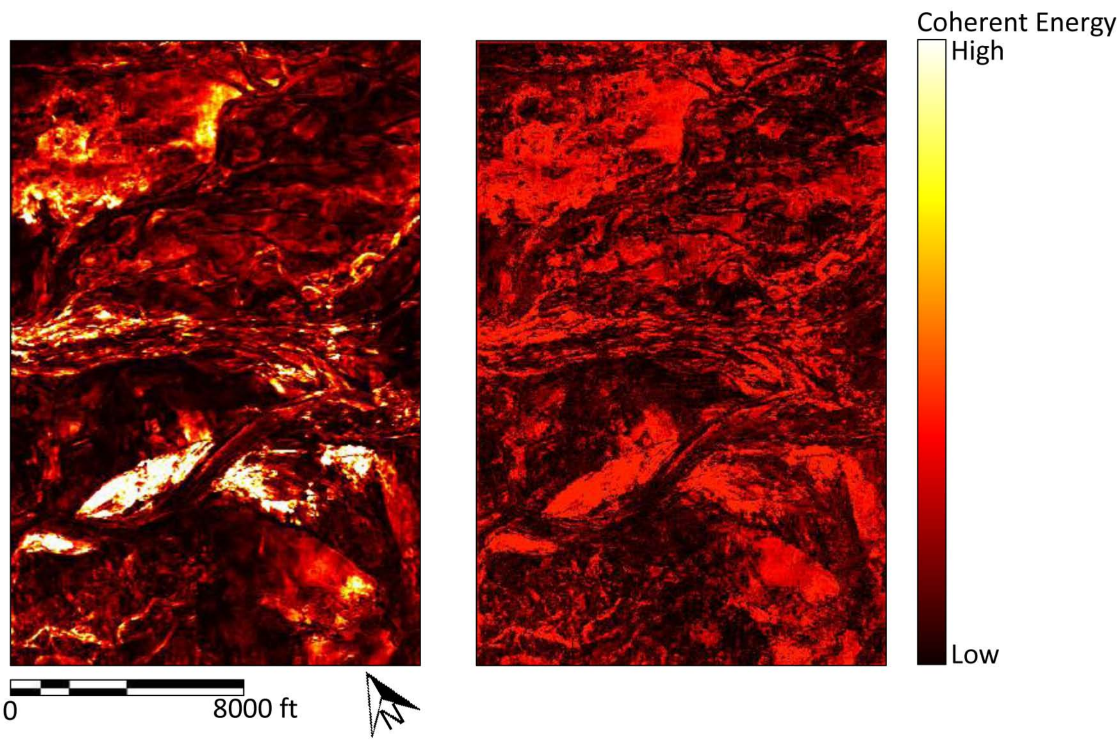

Figure 6—Coherent energy attribute extracted along the picked horizon. The left image shows the original attribute and the right image shows the reconstructed coherent energy attribute.

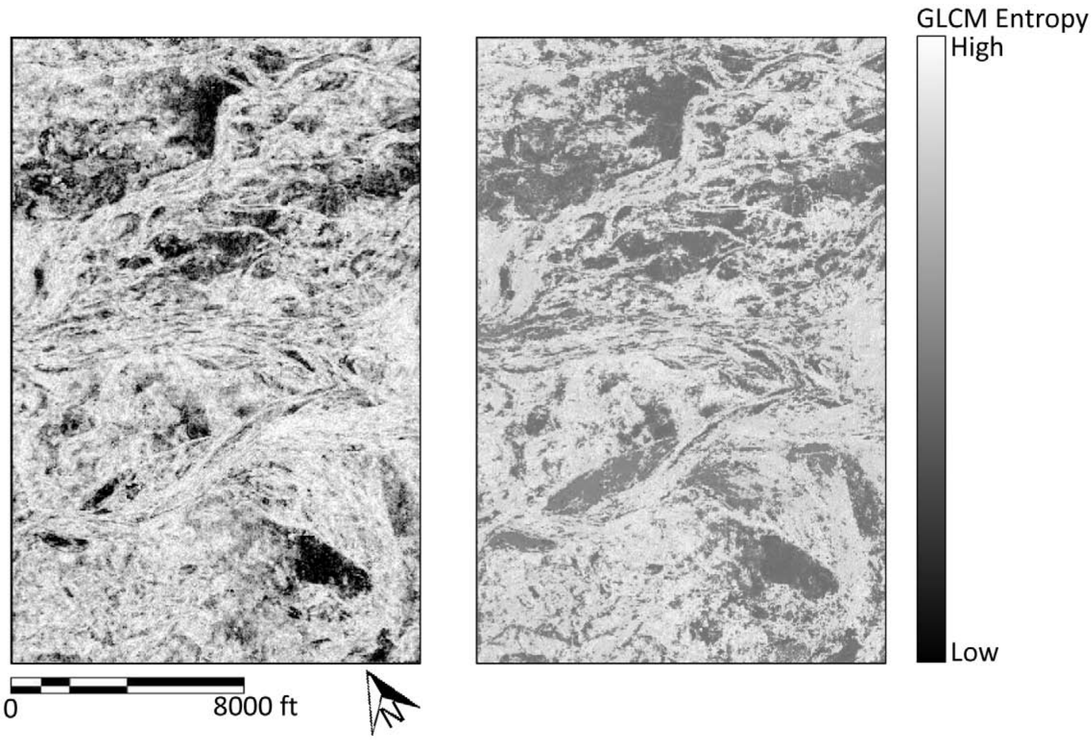

Figure 7—GLCM (entropy) attribute extracted along the picked horizon. The left image shows the original attribute and the right image shows the reconstructed GLCM (entropy) attribute.

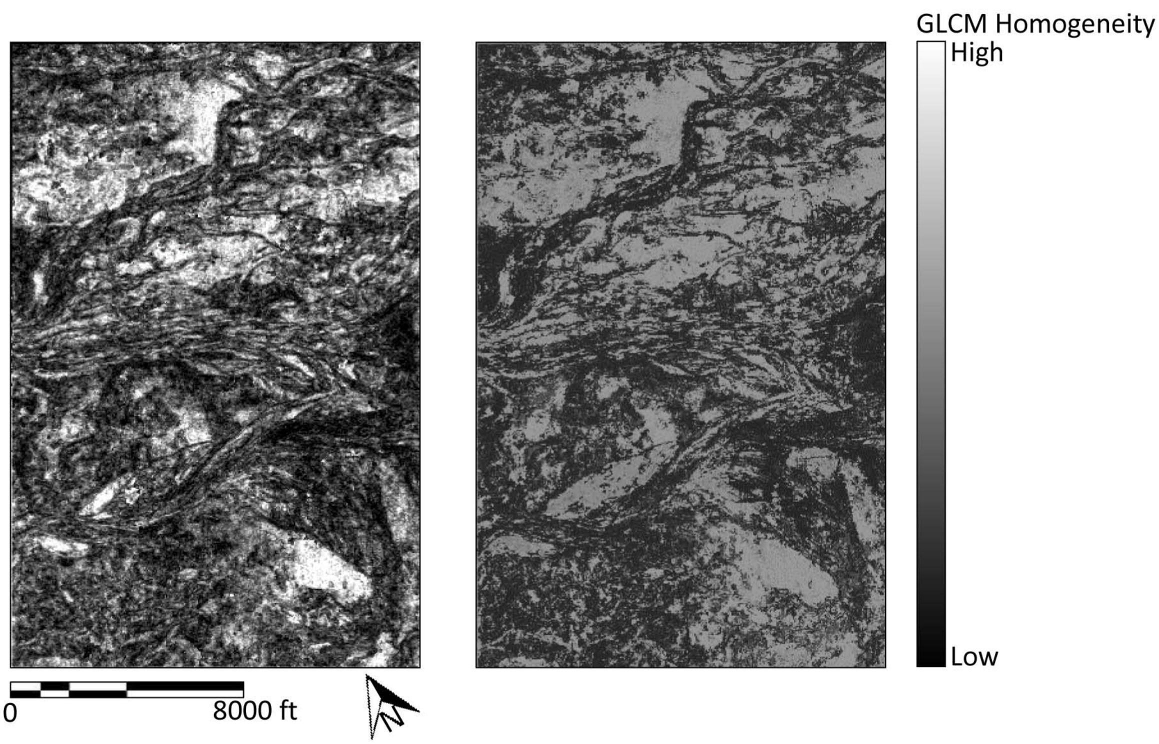

Figure 8—GLCM (homogeneity) attribute extracted along the picked horizon. The left image shows the original attribute and the right image shows the reconstructed GLCM (homogeneity) attribute.

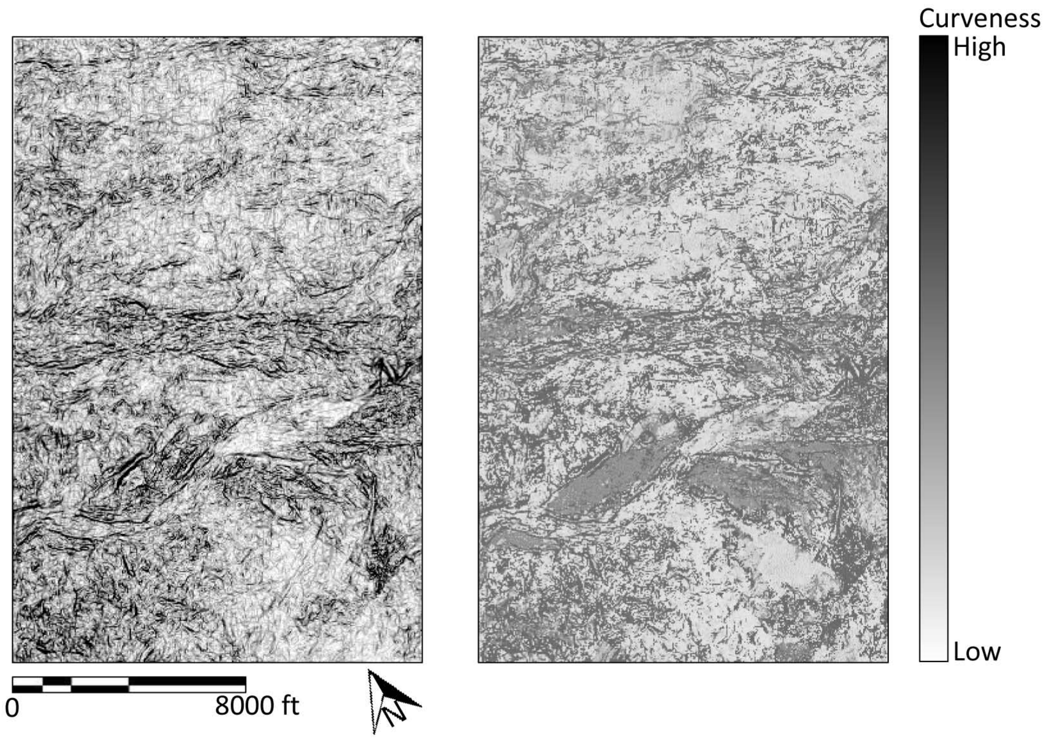

Figure 9—Curvedness attribute extracted along the picked horizon. The left image shows the original attribute and the right image shows the reconstructed curvedness attribute.

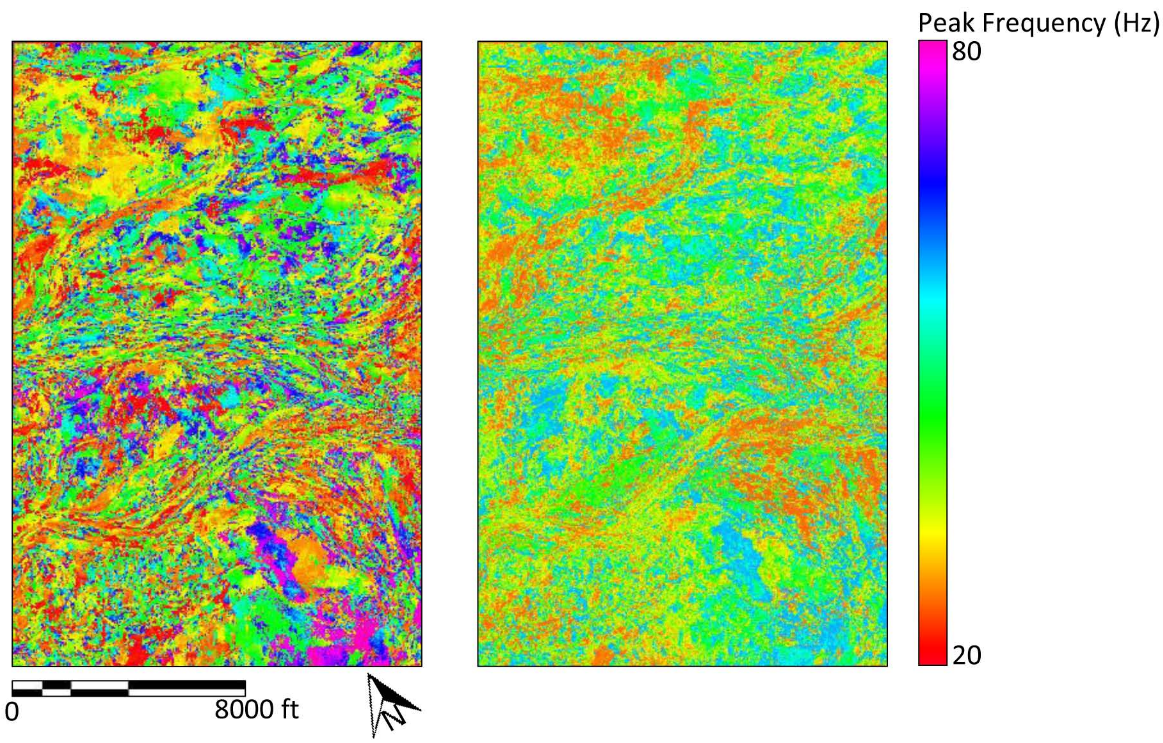

Figure 10—Peak frequency attribute extracted along the picked horizon. The left image shows the original attribute and the right image shows the reconstructed peak frequency attribute.

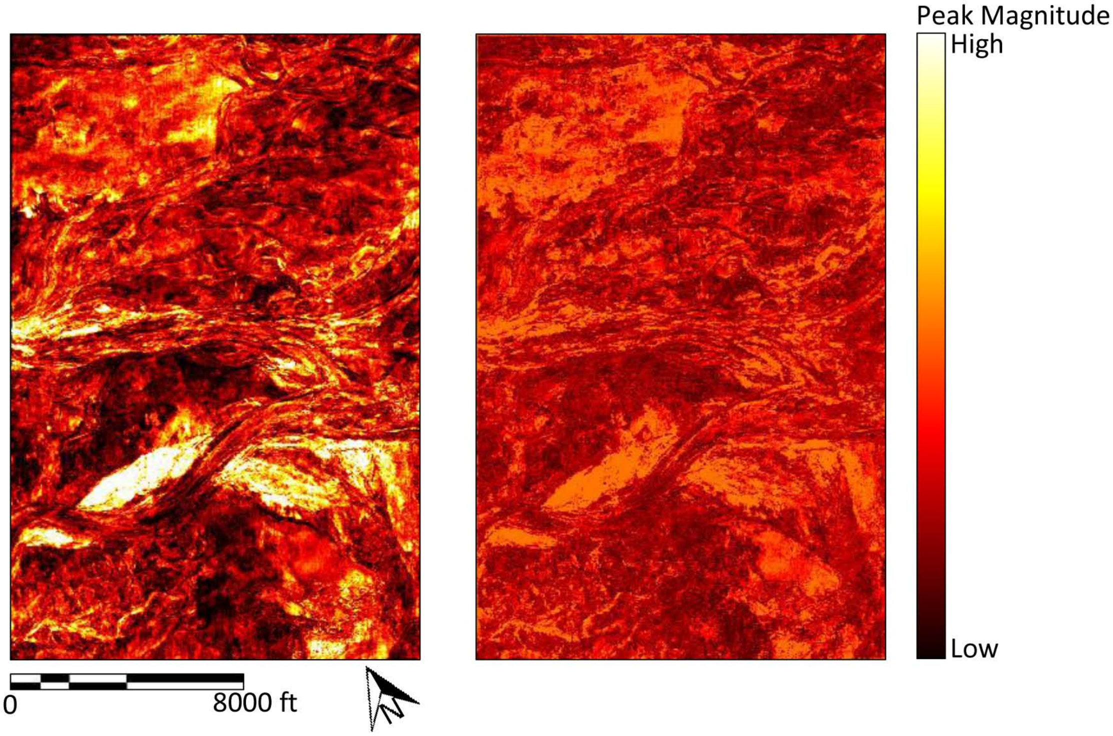

Figure 11—Peak magnitude attribute extracted along the picked horizon. The left image shows the original attribute and the right image shows the reconstructed peak magnitude attribute.

After examining these images, we concluded that the three modeled latent variables have done a good job of encoding the input attributes, with the output of the network closely matching the input. Because the coding retains the information inherent in the data set, the code is a reasonable latent space representation of the data. The three latent space images based upon the outputs at the code nodes are shown in Figure 12.

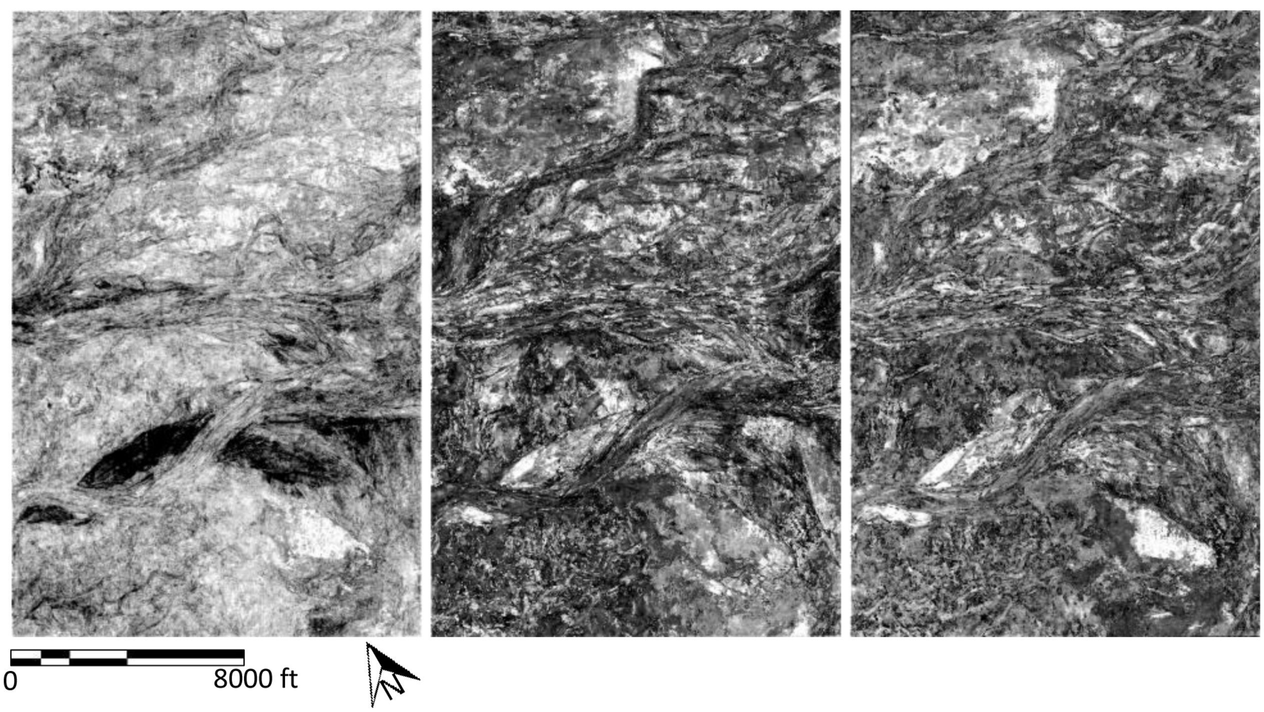

Figure 12—Encoded latent attributes. Ordering and definition of black or white being a low value is arbitrary.

A 3D latent space was chosen to facilitate visualization using an RGB color map. For the purposes of this visualization, the ordering of the attributes is arbitrary. Also, the output at the code layer ranges in value from zero to one. Each of these will be used to represent intensity of a color from black to red, green, or blue respectively. There is nothing inherent about the orientation of these coded attributes, and their values may just as properly be reversed in the visualization, i.e. black areas in the images could have been displayed in white and white areas could have been displayed in black with no change in meaning.

Given three images to view as a RGB image, there are 3×2×1 or six different ways to choose the ordering. Within each ordering, there are 2^3 or eight different ways to choose the low to high orientation of the images. Therefore, there are 48 different ways to visualize the output of a three attribute code in image format. Figure 13 shows eight different orientation sets based upon the ordering in Figure 12 with the left to right images being display as red, green, and blue respectively.

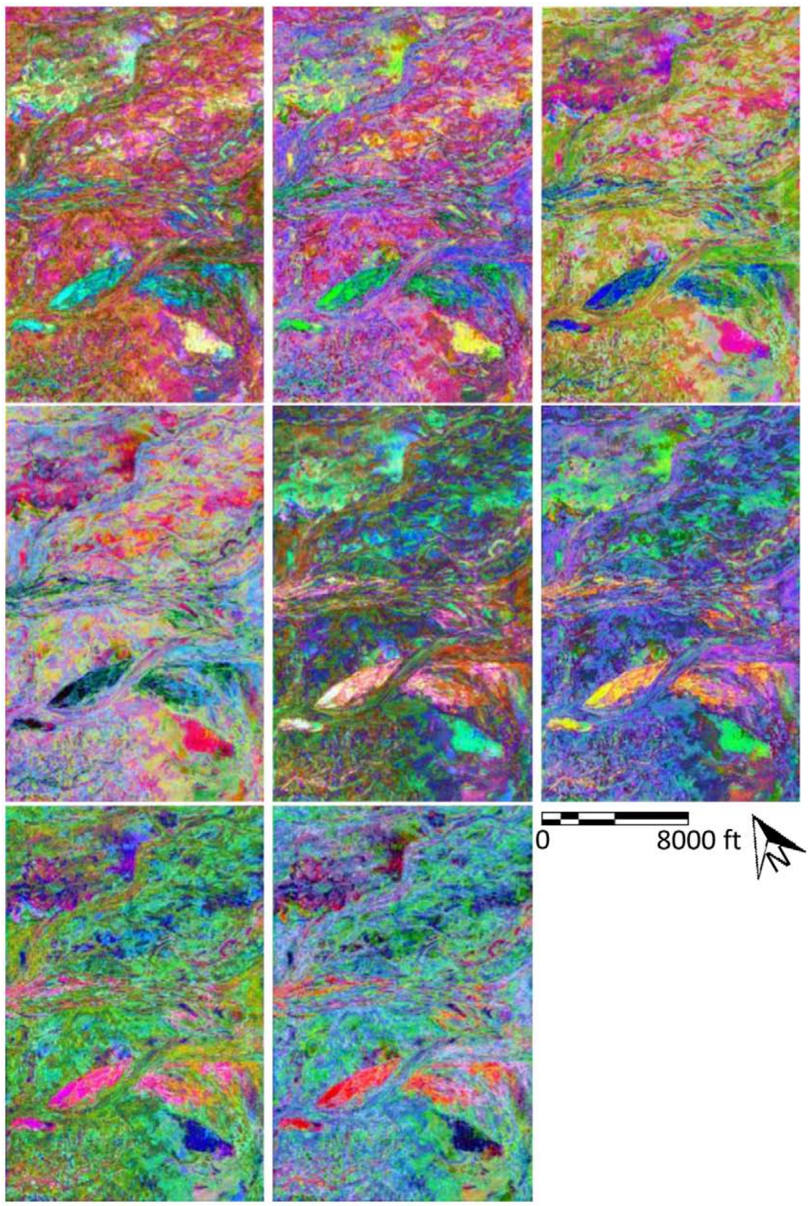

Figure 13—Eight different images that can be formed from one ordering of the code attributes. Each one uses a different color encoding as to whether a high or low value is represented as black.

In making the decision on which ordering and orientation to use, it is legitimate to consider aesthetics in these choices, i.e. optimizing according to the preference of the interpreter. Even within this decision, different configurations may tend to emphasis different architectural elements and geological features, such that it may be useful to use multiple images sequentially. However, our recommendation is to animate through all choices and to make a decision as to which is best suited according to the preferences of the interpreter.

Figure 14 shows one of the images from Figure 13 with an interpretation of various architectural elements shown using different color arrows. The shown image has the red channels color bar reversed such that high values are black and low values are red. The blue and the green channels are shown such that low and high values give black and full colors respectively. The interpretation is based upon the authors' knowledge of seismic geomorphology as well as a careful examination of the 3D profiles of the interpreted architectural elements. This image shows a great deal of detail and internal structure to the interpreted features. A full interpretation of this image would integrate well logs that would be used to calibrate to interpretation.

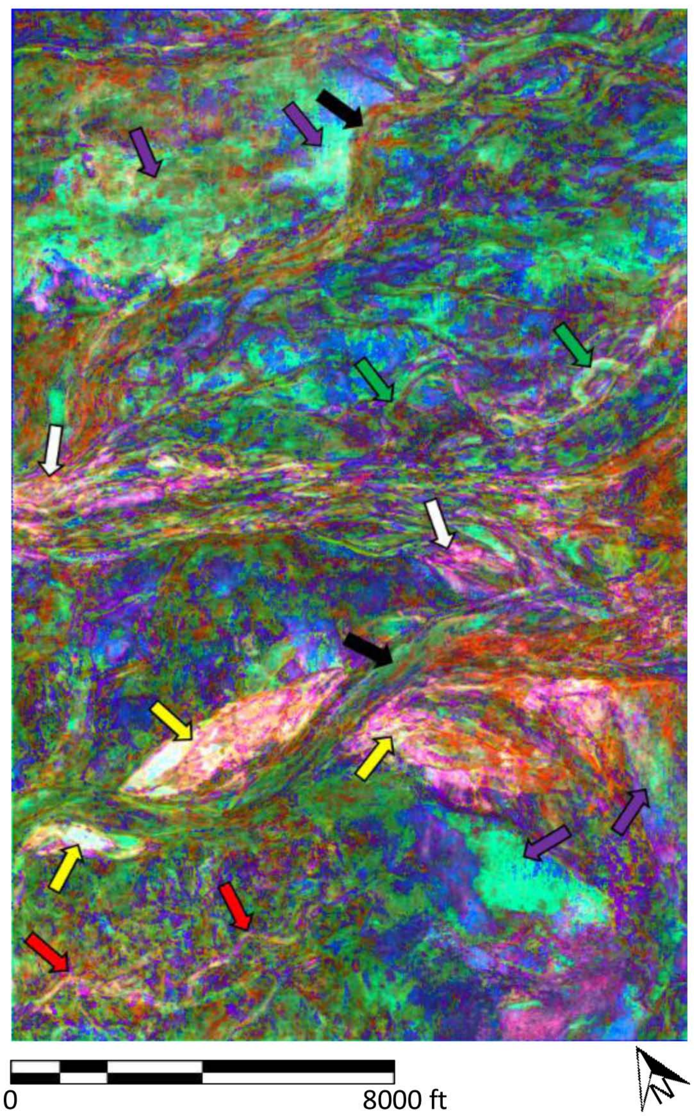

Figure 14—An RGB composite image with a possible interpretation. Red arrows denote a sinuous channel complex. Yellow arrows denote a possible sand filled channel. Black arrows denote possible mud filled channels. White arrows denote possibly sand filled lateral accretion packages. Purple arrows indicate likely fans and lobe deposits. Green arrows denotes a likely levee overbank.

Conclusions

We have presented a powerful new method for doing latent space visualization of seismic attributes using a deep learning technique, autoencoders. This method removes redundant information by forcing a concise encoding of the data that allows for the reproduction of the original information. Encoding allows for focusing upon preserving information rather than removing supposed noise. This focus upon preservation is important since the definition of noise is relative to the task for which the data is intended. Unlike SOM, our method allows for an easy way to assess the goodness of information preservation by comparing the reproduced data with the input data.

We have demonstrated that our method is effective and produces images with considerable interpretational value on horizon slices. Future work is necessary to assess its power in 3D applications including thin bed mapping and 3D modeling of architectural elements. Additionally, the incorporation of information from well logs would allow for calibration of the seismic data to facies, allowing for a more complete and definitive interpretation.

In the future, we will compare the results of our workflow to other methods such as PCA and SOM. The literature currently lacks work comparing different methods of latent space analysis of seismic attributes. While we can assess strengths and limitations of various methods based upon their assumptions and implementations, there is no practical knowledge of what methods work best in which situations and for which tasks.

Finally, we will continue to look for ways to incorporate spatial information into the process. Seismic data are inherently spatial in nature, and the definition of information should include a component of spatial variations. Our method, as well as all methods of latent space learning that we are aware of, do not incorporate spatial information in the learning process.

Acknowledgements

Thang Ha would like to thank the sponsors of the Attribute-Assisted Seismic Processing and Interpretation Consortium for their financial support. We would also like to thank the New Zealand Petroleum and Minerals for providing public access to their seismic data.

People we would also like to thank are Tao Zhao for his previous efforts related to this research. A special thanks to Prof. Roger Slatt for providing valuable insight and discussions about turbidite systems. Finally, we would like to express our gratitude to Prof. Kurt Marfurt for his ongoing support and advice.

References

Barnes, A., 2007, Redundant and useless seismic attributes: *Geophysics*, **72**, P33-P38.

Bellman, R. E., 1957, Dynamic Programming: Princeton University Press.

Cozens, N., 2011, A study of unconventional gas accumulation in Dannevirke Series (Paleogene) rocks, Canterbury Basin, New Zealand: Master's thesis, Victoria University of Wellington.

Dempster, A. P., N. M. Laird, and D. B. Rubin, 1977, Maximum likelihood from incomplete data via the EM algorithm: *Journal of the Royal Statistical Society Series B*, **39**, 1-38.

Guo, H., K. J. Marfurt, J. Liu, and Q. Duo, 2006, Principal component analysis of spectral components: 76th Annual International Meeting, SEG, Expanded Abstracts, 988–992.

Liou, C., W. Cheng, J. Liou, and D. Liou, 2014, Autoencoder for words: *Neurocomputing*, **139**, 84-96.

Makhzani, A. and B. Frey, 2013, k-Sparse Autoencoders. arXiv: 1312.5663

Marfurt, K. J., 2015, Techniques and best practices in multiattribute display: *Interpretation*, **3**, B1-B23.

Roy, A., B. L. Dowdell, and K. J. Marfurt, 2013, Characterizing a Mississippian tripolitic chert reservoir using 3d unsupervised and supervised multiattribute seismic facies analysis: An example from Osage County, Oklahoma: *Interpretation*, **1**, SB109-SB124.

Sutherland, R., and G. Browne, 2003, Canterbury basin offers potential on South Island, New Zealand: *Oil & Gas Journal*, **101**, 45–49.

Vincent, P., H. Larochelle, I. Lajoie, Y. Bengio, and P. Manzagol, 2010, Stacked denoising autoencoders: Learning useful representations in a Deep Network with a Local Denoising Criterion: *The Journal of Machine Learning Research*, **11**, 3371–3408.

Wallet, B. C., 2013, Using the image grand tour to visualize fluvial deltaic architectural elements in south Texas, USA: *Interpretation*, **1**, SA117-SA129.

Wallet, B. C., M. C. de Matos, J. T. Kwiatkowski, and Y. Suarez, 2009, Latent space modeling of seismic data: An overview: *The Leading Edge*, **28**, 1454-1459.

Wallet, B. C., R. P. Altimar, and R. M. Slatt, 2014, Unsupervised classification of $\lambda\rho$ and $\mu\rho$ attributes derived from well log data in the Barnett Shale: 84th Annual International Meeting of the SEG, Expanded Abstracts, 1594-1598.

Zhao, T., J. Zhang, F. Li, and K. J. Marfurt, 2016, Characterizing a turbidite system in Canterbury Basin, New Zealand, using seismic attributes and distance-preserving self-organizing maps: *Interpretation*, **4**, SB79-89.