UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

MACHINE LEARNING APPLICATIONS FOR SEISMIC PROCESSING & INTERPRETATION

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

DOCTOR OF PHILOSOPHY

By

YUJI KIM

Norman, Oklahoma

2020

MACHINE LEARNING APPLICATIONS
FOR SEISMIC PROCESSING & INTERPRETATION

A DISSERTATION APPROVED FOR THE
SCHOOL OF GEOSCIENCES

BY THE COMMITTEE CONSISTING OF

Dr. Kurt J. Marfurt, Chair

Dr. Deepak Devegowda

Dr. Heather Bedle

Dr. Brett Carpenter

Dr. Bradley Wallet

To my family and everyone who supported me in my studies

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my advisor Dr. Marfurt for his dedicated support, guidance and patience. He continuously provided encouragement and was always willing and enthusiastic to assist in any way he could throughout the research project.

I thank Dr. Bradley Wallet, Dr. Header Bedle, Dr. Brett Carpenter, and Dr. Deepak Devegowda who served on my PhD committee. Dr. Wallet gave me excellent statistical feedback on my research. Dr. Bedle held the SDA/AASPI seminar which gave me many critical insights to seismic interpretation. I also enjoyed Dr. Carpenter's Fractures, Faults, Earthquakes class where students from different academic backgrounds could share their views and thoughts on fracture mechanics. And Dr. Devegowda's comments on machine learning are greatly appreciated.

I appreciate all members and industrial sponsors of the Attribute Assisted Seismic Processing and Interpretation (AASPI) consortium for their continuous support and encouragement. Also, I want to thank to KIGAM for my early work on prestack least-squares time migration. I want to thank Larry, Shiguang and Thang for helping me when I started projects and research for the AASPI consortium. I also thank Dr. Nori Nakata for his discussions and feedback on research.

I would like to thank Rebecca Fay, Leah Moser, Ginger Leivas Ginny Guedez and Ashley Tullius in the geosciences department for their professional guidance and help.

Finally, I want to thank to my parents, my sister, my husband and my friends who supported and encouraged me to go through.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

During the past few years, exploration seismology has increasingly made use of machine learning algorithms in several areas including seismic data processing, attribute analysis, and computer aided interpretation. Since machine learning is a data-driven method for problem solving, it is important to adopt data which have good quality with minimal bias. Hidden variables and an appropriate objective function also need to be considered. In this dissertation, I focus my research on adapting machine learning algorithms that have been successfully applied to other scientific analysis problems to seismic interpretation and seismic data processing. Seismic data volumes can be extremely large, containing Gigabytes to Terrabytes of information. Add to these volumes the rich choice of seismic attributes, each of which has its own strengths in expressing geologic patterns, and the problem grows larger still.

Seismic interpretation involves picking faults and horizons and identifying geologic features by their geometry, morphology, and amplitude patterns seen on seismic data. For the seismic facies classification task, I tested multiple attributes as input and built an attribute subset that can best differentiate the salt, mass transport deposits (MTDs), and conformal reflector seismic patterns using a suite of attribute selection algorithms. The resulting attribute subset differentiates the three classes with high accuracy and has the benefit of reducing the dimensionality of the data. To maximize the use of unlabeled data as well as labeled data, I provide a workflow for facies classification based on a semi-supervised learning approach.

Compared to using only labeled data, I find that the addition of unlabeled data for learning results in higher performance of classification. In seismic processing, I propose a deep learning approach for random and coherent noise attenuation in the frequency – space domain. I find that the deep ResNet architecture speeds up the process of denoising and improves the accuracy, which efficiently separates the noise from signals. Finally, I review geophysical inversion and machine learning approaches in an aspect of solving inverse problems and show similarities and differences of these approaches in both mathematical formulation and numerical tests.

CHAPTER 1: INTRODUCTION

During the past few years, machine learning has gained popularity in exploration seismology not only in seismic data processing, but also in seismic attribute analysis and the interpretation of 3D data volumes. Machine learning is a computer algorithm that improves automatically through experience (Mitchell, 1997). The "automagical" process means that the learning algorithm can perform tasks without explicit programming. Instead of explicit programming, machine learning solves a problem based on statistical analyses using example data or past experience (Parsons, 2005).

In exploration seismology, seismic interpretation uses machine learning techniques extensively, with algorithms addressing automatic picking of seismic horizons and first break picking for refraction statics being used for 25 years or more. The task of an interpreter is to use 3D seismic data to not only map major features such as horizons and faults but also to infer the geology based on an understanding of both geologic processes and the limits of seismic data acquisition, processing, and inversion. Interpreters improve with age and experience; where a petroleum engineer may see nothing but wiggles, the interpreter sees a turbidite in the data. Because of the importance of experience in human interpretation, seismic interpretation is also amenable to n machine learning where the scientific discovery process is data-intensive and empirical. Machine learning tools such as self-organizing maps (SOM) (Kohonen, 1982), k-means (Forgy, 1965; Jancey, 1966), and artificial neural networks have been used for seismic facies classification since the 1990s (Poupon et al., 1999). The advance of convolutional neural networks (CNNs) has accelerated machine learning to complete more complicated tasks such as the interpretation

of fault systems (Xiong et al., 2018) and classifying sedimentary succession patterns (Dramsch and Lüthje, 2018).

Recent breakthroughs in machine learning and deep neural network encourage geophysicists to make use of machine learning algorithms to solve geophysical problems in seismic processing. Many seismic processing tasks such as velocity modeling, inversion and noise attenuation are based on physics-based models. The advantage of a data-driven model over a physics-based model is that a data-driven model does not require information on initial or boundary conditions, the constraints, or even the correct model of the physical phenomenon. Several machine learning applications have shown promising results in prestack seismic inversion (Yang and Ma, 2019; Li et al., 2019), full waveform inversion (Zhang and Alkhalifah, 2019), data interpolation (Jia and Ma, 2017; Wang et al., 2019), and signal to noise enhancement (Yu et al., 2019).

Since machine learning is a data-driven method where the model is decided by inferring relationships between training input and output data (supervised learning) or by measuring the distribution of input data (unsupervised learning), the model building does not require explicit programming. As the cost of computation has become lower, machine learning algorithms can process massive sizes of data. However, since model validity relies heavily on data, it is important to use training data that can accurately represent the problem. The investigator should pay attention to the quality, bias, and splitting of data. Data in explorational seismology are massive in size and contain significant amounts of information. Seismic data also suffer from noise due to the recording environment, malfunction of a geophone, limitations in processing, and most difficult, aliasing due to

undersampling. In addition to data quality, considering hidden variables and different objective functions is critical (Riley, 2019).

In this dissertation, I address multiple applications of machine learning for seismic interpretation and seismic data processing. For a facies classification task, I test multiple attributes to build a training dataset that can differentiate and describe specific geologic patterns and depositional environments. I propose a facies classification workflow to maximize the use of unlabeled data as well as labeled data based on the approach of semi-supervised learning. In seismic processing, I evaluate machine learning solutions for noise attenuation and geophysical inversion.

This dissertation is structured as follows:

In Chapter 2, I test multiple attributes to classify salt diapirs, mass transport deposits (MTDs), and the conformal reflector "background" for a 3D seismic marine survey acquired over the northern Gulf of Mexico shelf. I analyze attribute-to-attribute correlation and the correlation between the input attributes to the output classes to understand which attributes are relevant and which attributes are redundant and find that amplitude and texture attribute families are able to differentiate salt, MTDs, and conformal reflectors. I then apply the attribute selection workflow to a Barnett Shale play to differentiate limestone and shale facies. I rank multivariate analysis using filter, wrapper and embedded algorithms attributes by importance, facilitating the choice of the best attribute subset for classification. I show that using such attribute selection algorithms for supervised learning not only reduces computational cost but also enhances the performance of the classification.

In Chapter 3, to maximize the use of the unlabeled data, I build and test a conditional generative adversarial network (cGAN) framework to differentiate salt, mass-transport deposits (MTDs), and conformal reflectors based on a semi-supervised learning approach. I compare the classification results of the proposed cGAN model ed with other deep neural network architectures which are based on supervised learning, such as the U-net. I will show that the semi-supervised cGAN model results in higher validation accuracies when the number of labeled training samples is relatively small. I also compare single attribute input and multiple attribute input for the classification task. The multiple attribute input (seismic amplitude, chaos, total aberrancy magnitude, and instantaneous frequency) shows improved performance of classification compared with that of a single attribute (the original seismic amplitude volume). The predicted 3D facies volume facilitates the understanding of the depositional relation of the MTDs with the salt diapirs, as well as the geometry of salt diapirs themselves.

In Chapter 4, I apply a complex-valued residual convolutional neural network (ResNet) to attenuate random and coherent noise in the frequency – space (*f*-*x*) domain. Noise attenuation in the *f*-*x* domain has the advantage in that the spectra often exhibit a few dominant harmonics. I generalize the ResNet algorithm to treat complex-valued data with the goal of suppressing noise. When it is applied to random noise attenuation, the resulting time – space (*t*-*x*) displays of denoised signals show that the noise is significantly suppressed while preserving the amplitude of the signal. The denoising method is effective for removing coherent noise such as migration artifacts in poststack land data.

In Chapter 5, I compare geophysical inversion and machine learning approaches in an aspect of solving inverse problems and show similarities and differences of these approaches in a mathematical form and numerical tests. Both methods aid to solve ill-posed and non-linear problems and use similar optimization techniques. I take reflectivity inversion as an example of the inverse problem. I apply geophysical inversion based on the least-squares method and artificial neural network (ANN) as a machine learning approach to solve reflectivity inversion using 2D synthetic data sets and 3D field data sets. I find that a neural network with multiple hidden layers successfully generates the non-linear mapping function to predict reflectivity. For this inverse problem, I test different L1 regularizations for both approaches. L1 regularization alleviates some of the effects of noise in the seismic data and enhances sparsity for the least-squares method. Sparsity in the ANN method is introduced by using a sparse representation of the reflectivity data in the training data.

In Chapter 6, I evaluate a preconditioned least-squares migration workflow to address acquisition artifacts in marine seismic data. Marine seismic data often suffers from acquisition artifacts where high currents give rise to cable feathering and mispositioning. These artifacts negatively impact the data, providing inaccurate amplitudes and generating footprint artifacts that overprint the subsurface geology. I use a conjugate gradient method to minimize the objective function and a prestack structure-oriented filter as a preconditioner to reject aliased data and accelerate convergence. I apply this workflow to a 3D seismic survey acquired in three phases in the Jeju Basin of South Korea that suffered from severe cable feathering due to strong currents and adverse weather. The stacked images and time slices through attribute volume computed using my constrained least-

squared migration workflow shows significant attenuation of the artifacts arising from cable feathering resulting in irregular subsurface coverage.

References

Dramsch, J. S., and M. Lüthje, 2018, Deep-learning seismic facies on state-of-the-art CNN architectures: 88th Annual International Meeting, SEG Expanded Abstracts, 2036-2040.

Forgy, E. W., 1965, Cluster analysis of multivariate data: efficiency vs interpretability of classifications: Biometrics, **21**, 768-769.

Jancey, R. C., 1966, Multidimensional group analysis: Australian Journal of Botany, **14**, 127-130.

Jia, Y., and J. Ma, 2017, What can machine learning do for seismic data processing? An interpolation application: Geophysics, **82** - 3, V163-V177.

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

Li, S., B. Liu, Y. Ren, Y. Chen, S. Yang, Y. Wang, and P. Jiang, 2019, Deep learning inversion of seismic data: arXiv preprint arXiv:1901.07733.

Mitchell, T. M., 1997, Machine learning. 1997: Burr Ridge, IL: McGraw Hill, **45** - 37, 870-877.

Mitchell, T. M., 2006, The discipline of machine learning: Carnegie Mellon University, School of Computer Science, Machine Learning Department.

Parsons, S., 2005, Introduction to Machine Learning by Ethem Alpaydin: The Knowledge Engineering Review, **20** - 4, 432.

Riley, P., 2019, Three pitfalls to avoid in machine learning: Nature, **572**, 27–28

Wang, B., N. Zhang, W. Lu, and J. Wang, 2018, Deep-learning-based seismic data interpolation: A preliminary result: Geophysics, **84** - 1, V11-V20.

Xiong, W., X. Ji, Y. Ma, Y. Wang, N. M. BenHassan, M. N. Ali, and Y. Luo, 2018, Seismic fault detection with convolutional neural network: Geophysics, **83**, 1–28.

Yang, F., and J. Ma, 2019, Deep-learning inversion: a next generation seismic velocity-model building method: Geophysics, **84** - 4, 1-133.

Yu, S., J. Ma, and W. Wang, 2019, Deep learning for denoising. Geophysics, **84** - 6, V333-V350.

Zhang, Z. D., and T. Alkhalifah, 2019, Regularized elastic full-waveform inversion using deep learning. Geophysics, **84** - 5, R741-R751.

CHAPTER 2: ATTRIBUTE SELECTION IN SEISMIC FACIES CLASSIFICATION: APPLICATION TO A GULF OF MEXICO 3D SEISMIC SURVEY AND THE BARNETT SHALE

Introduction

In the exploration and production (E&P) industry, automated seismic facies classification is gradually being integrated into common workflows. Several machine learning algorithms, such as self-organizing maps (SOM) and K-means clustering, have been applied to automate seismic facies classification and are available in several commercial interpretation software packages. A great number of different seismic attributes can be used as inputs to machine learning algorithms for classification and pattern recognition. However, some attributes express geologic or depositional patterns more effectively than others. For instance, the envelope (reflection strength) is sensitive to changes in acoustic impedance and has long been correlated to changes in lithology and porosity (Chopra and Marfurt, 2005). In many cases, the instantaneous frequency typically enhances interpretation of vertical and lateral variations of layer thickness (Chopra and Marfurt, 2005). Coherence measures lateral changes in seismic waveforms which in turn can be correlated to lateral changes in structure and stratigraphy (Marfurt et al., 1998). Exploration seismic are large volumes of data and many attributes may be highly redundant. Adding to this problem, the original seismic amplitude data (and therefore subsequently derived attributes) may contain significant noise (Coléou. et al. 2003). Therefore, understanding the nature of seismic attributes is of crucial importance for providing the most reliable facies classifications.

According to the Hughes Phenomenon, adding attributes beyond a threshold value causes a classifier's performance to degrade (Hughes, 1968). A number of studies found dimensionality reduction in machine learning problems reduces computation time and storage space as well as having meaningful results for facies classification (Coléou. et al. 2003; Roy et al., 2010; Roden et al., 2015). Principal component analysis (PCA) is one of the most popular methods reducing a large multidimensional (multiattribute) data set into a lower dimensional data set spanned by composite (linear combinations of the original) attributes, while preserving variation. SOM also creates a lower-dimensional representation of high-dimensional data to aid interpretation. Both PCA and SOM are a type of unsupervised learning, where the goal is to discover the underlying structure of the input data.

Roden et al. (2015) used PCA to define a framework for multiattribute analysis to understand which seismic attributes are significant for unsupervised learning. In their study, the combination of attributes determined by PCA was used as input to SOM to identify geologic patterns and to define stratigraphy, seismic facies and direct hydrocarbon indicators. Zhao et al. (2018) built on these ideas and suggested a weight matrix computed from the skewness and kurtosis of attribute histograms to improve SOM learning.

In general, attribute selection in unsupervised learning relies on the data distribution of the input attributes and the correlation between input attributes. Supervised learning maps a relationship between input attributes and a desired output using an interpreter-defined training dataset. A number of supervised learning studies for attribute selection, also known as feature selection or variable selection, to reduce dimensionality (Jain and

Zongker, 1997; Chandrashekar and Sahin, 2014). We introduce multiple strategies to select appropriate attributes for seismic facies classification with a case study. Our goals are to provide a good classification model in terms of validation accuracy while avoiding overfitting and reduce computation and memory requirements needed for generating seismic attributes.

A desirable attribute subset might simply be built by detecting relevant attributes and discarding the irrelevant ones (Sánchez-Maroño et al., 2007). While relevant attributes are those which are highly correlated with the output classes, redundant attributes are highly correlated with each other. Barnes (2007) suggested that there are many redundant and useless attributes that breeds confusion in seismic interpretation, we argue that these attributes also pose problems in machine learning classification.

To avoid building an unnecessarily complex model, we evaluate several attribute selection algorithms to maximize relevance and minimize redundancy to build an efficient subset of attributes for supervised facies classification analysis. Attribute selection methods can be classified into three groups: 1) filter which uses a correlation or dependency measure, 2) wrapper which applies a predictive model to evaluate the performance of an attribute subset and 3) embedded which measures attribute importance during the training process. Since multiple attributes are analyzed simultaneously in the test, we consider our attribute selection algorithm to be a multivariate algorithm.

We compare the three types of attribute selection algorithms to build an efficient subset to differentiate seismic facies in a Gulf of Mexico survey. Twenty attributes from five attribute categories are generated. The attribute categories consist of amplitude,

instantaneous, geometric, texture and spectral. The aim of the case study is to classify specific facies based on patterns from a labeled training dataset. We define the target classes of training data as being the facies corresponding to salt diapirs, MTDs, and conformal reflectors which are created from manual geological and stratigraphical interpretation. Correlations between attributes and correlations between attributes and output classes are analyzed using different measures to investigate the relevance and redundancy of each seismic attribute. The selected attributes are tested using a Random Forest (RF) algorithm and classification results are discussed. We also apply our workflow to the Barnett Shale play in Fort Worth Basin to differentiate shale and limestone facies using inverted physical properties as input attributes. Output class is labeled based on stratigraphical interpretation aided by adjacent wireline logs. The classification results using different attribute subsets are discussed.

Correlation measures to maximize relevance, minimize redundancy

Finding an optimal subset of attributes can be achieved by maximizing the relevance between attributes and output classes, while minimizing redundancy among attributes (Yu and Liu, 2004; Peng et al., 2005). To maximize relevance, attributes that are highly correlated with output classes are selected. On the other hand, redundancy is caused by attributes that are highly correlated to each other. Thus, measuring and analyzing the correlation between attributes and classes, or correlations between attributes are prioritized to evaluate the performance of attribute subset. A number of correlation measures can be

used in feature selection; We examine Pearson correlation (Pearson, 1895), rank correlation (Spearman, 1904), mutual information (MI) (Shannon and Weaver, 1949; Cover and Thomas, 1991), and distance correlation **(**Székely et al. 2007**)**. Refer to Appendix A for a mathematical description. Pearson correlation (Pearson, 1895) is the most common measure, and detects only linear relationship between two random variables. Spearman's rank correlation (Spearman, 1904) measures the tendency of a positive or negative relation, without requiring the increase or decrease to be explained by a linear relationship. Figure 2.1 illustrates different types of relationships between variables *X* and *Y*. Four types of correlation measure are able to detect linear relationship (Figure 2.1a and 2.1d). In Figure 2.1b and 2.1e the rank correlation has a higher coefficient value than Pearson's correlation, because rank correlation is able to detect non-linear positive relationships, while Pearson's correlation is not. Also dependence among attributes are not always linear. Mutual information (Shannon and Weaver, 1949; Cover and Thomas, 1991) and distance correlation ***(****Székely et al. 2007****)***, detect non-linear and non-monotonic relationship. Figure 2.1c and 2.1f, the Pearson and rank correlation coefficients are approximately zero which indicates that these two correlation measures do not detect non-linear and non-monotonic relationship.

In terms of dependence between an input attribute and an output class, it is also important to identify each predictive model's ability to map non-linear relationship between an attribute and a class. Even though the attribute is a powerful variable which can have a high correlation with class, some predictive models may degrade prediction accuracy if the model cannot properly map the relationship. Figure 2.1 describes four types

of predictive models: linear Bayesian, neural network (NN), random forest (RF), and support vector machine (SVM) and their ability to map input and desired output using regression methods. Using data points described in each plot, 5-fold cross-validation is applied. The hyper parameters for each predictive model were selected based on grid-searches that give the best validation score. Linear-Bayesian models are not able to map non-linear relationships which gives accuracy of 0.84 in the monotonic case (Figure 2.1b) and -0.04 in the non-monotonic case (Figure 2.1c). Except for linear-Bayesian, the other three of models map the input and output with high accuracy (1.0) when noise is not added (Figure 2.1b and 2.1c). We select our test predictive model to be NN, RF, and SVM for our case study since they are able to map the nonlinear relationships appropriately. The noise in the signal can affect the correlation, since it gives more uncertainty in predicting the output class. The sensitivity to noise also differs with correlation measure. MI and distance correlation coefficients decrease more than the other when 10 percent of Gaussian noise is added to variable *Y* as shown in Figure 2.1d, 2.1e and 2.1f.

The correlation measures are affected not only by non-linearity but also by the covariance of two variables or noise. Figure 2.2 shows the scatter plots of two attributes which have relatively high correlations. RMS amplitude and total energy in Figure 2.2a have a positive, monotonic relationship since the energy attribute is equivalent to the square of amplitude. These two attributes exhibit high rank correlation coefficient (0.97). The relationship between peak magnitude and instantaneous envelope (Figure 2.2b) is linear and has a higher Pearson coefficient than the other two scatter plots. Figure 2.2c describes correlation between entropy computed from grey-level co-occurrence matrix (GLCM) and

variance. MI and distance correlation can detect non-linear relationships but their values for GLCM entropy and variance are lower in Figure 2.2c because their entropy is high. Additionally, we also test analysis of variance (ANOVA) to determine which attributes are significant to differentiate output classes. ANOVA is an analysis tool that splits the variability found in a data set into systematic factors and random factors. If the variation can be explained from systematic factors, then the variable is significant in distinguishing classes.

Attribute selection algorithms: filters, wrappers and embedded methods

The goal of attribute selection is to differentiate seismic facies effectively with an optimal combination of different attributes. To choose an optimal subset, the relationship between attributes as well as their relevance to the output should be analyzed in a multivariate manner. To measure redundancy is simple when attributes are perfectly correlated. If two attributes are perfectly correlated, then adding them does not provide additional information. Guyon and Elisseeff (2003), however, suggest that if two variables are highly correlated, then they have a possibility to complement each other. Also, two variables which are not relevant by themselves can be useful when they are used together. Selecting attributes when considering relevance and redundancy together can be a complicated problem, but if the attribute selection algorithms are developed in a multivariate manner, they can be applied for attribute selection as well.

In supervised classification, there are three major approaches to select attributes in a multivariate manner: filters, wrappers and embedded methods. Figure 2.3 describes the mechanisms of these methods. Filter methods use a suitable measure or ranking criterion, such as correlation or mutual information to select attributes. Relief (Kira and Rendell, 1992) is a distance-based filter algorithm which evaluates attributes according to feature value differences between nearest neighbor instance pairs. ReliefF (Kononenko, 1994) which is an updated Relief algorithm can deal with multi-class problems and is more robust to incomplete and noisy data. Correlation-based Feature Selection (CFS) (Hall, 1999) is an algorithm based on a heuristic evaluation function which is calculated from correlations between attribute-class and attribute-attribute. Fast correlation-based filter (FCBF) algorithm (Yu and Liu, 2003) is also a correlation-based measure but designed for high-dimensional data. Filter methods are computationally less expensive than the wrapper algorithm which requires computation of classification model.

Wrapper methods use a classification model to select the attribute subset. Wrapper methods require greater computational resources but provide better performance in that they maximize classification accuracy. The sequential forward selection (SFS) algorithm (Kittler, 1978), for example, starts with an empty subset and adds an attribute to the subset sequentially to yield the highest increase in score. Sequential backward selection (SBS), on the other hand, subtracts an attribute from a full subset sequentially whose elimination gives the lowest decrease in score.

Embedded methods implement attribute selection as a part of the training process of classification. In addition to having a low computational cost, embedded methods do not

require a separate process for attribute selection. For instance, a random forest classifier calculates the variable importance (Breiman, 2001b; Liaw and Wiener, 2002) during training. Another embedded technique is to compute the weights of each attribute in the SVM classifier (Guyone et al.,2002) and logistic regression (Ma and Huang, 2005).

Application 1: Gulf of Mexico survey - attribute selection to differentiate salt, mass transport deposits (MTDs) and conformal reflectors

Seismic expression of salt and MTDs

Salt diapirs inherently have poor internal reflectivity and are easily overprinted by crossing coherent migration artifacts (Jones and Davison, 2014) in part due to their geometry and their higher P-wave velocities compared to surrounding strata. In general, the mismigrated noise gives rise to a relatively low amplitude, chaotic and discontinuous seismic patterns that result in low coherence and high GLCM entropy inside the salt body. Therefore, texture attributes, such as GLCM (entropy, homogeneity, energy) are used to differentiate salt diapirs (Berthelot et al., 2013; Qi et al., 2016). Mass transport deposits (MTDs) are slumps, slides, and debris flows generated by gravity-controlled processes (Nelson et al., 2011). MTDs often show chaotic or highly disrupted seismic patterns with great internal complexity (Frey-Martinez, 2010). In general, the resulting attribute anomalies are high RMS amplitudes and low coherence (Brown, 2011; Omosanya and Alves, 2013). The conformal reflectors around salt diapirs and MTDs show a relatively

continuous seismic pattern, which leads to high coherence and low to moderate values of GLCM entropy.

Methodology

The 3D marine seismic data was acquired in offshore Louisiana over an area of 3089 mi^2 (Qi et al., 2016). The poststack seismic volume includes 4367 in-lines, 1594 cross-lines and 475 time samples with a sampling interval of 4 ms. Twenty seismic attributes in five categories were calculated from the seismic volume. The five attribute categories consist of amplitude, geometric, instantaneous, texture and spectral attributes (Table 2.1). For supervised learning, we use a voxel type training dataset which is rendered from geological and stratigraphical interpretation: salt, MTDs and conformal reflectors are interpreted inside the red box (751 in-lines × 551 cross-lines) as shown in Figure 2.4, and cropped as a 3D seismic volume using a polygon. From the cropped volume, 10,000 voxels were randomly selected for each facies and labeled for training output (e.g. conformal reflectors: 0, salt diapirs: 1, and MTDs: 2). The training input data were then extracted from the twenty attribute volumes at the same voxel locations. Figure 2.5 summarizes the attribute selection workflow. First, we look into attribute - attribute correlations using four measures of correlation: Pearson, rank, MI, and distance correlation. These measures are valid to analyze relationships between two continuous variables. To investigate the relationship between attributes and an output class, we used ANOVA and MI. Both metrics can provide information on how well a single attribute can differentiate classes and is relevant to each output class individually. Even if the correlation measures are not used to

build the attribute subsets, correlation measures help to explain and evaluate the results from the attribute selection methods. We apply multivariate algorithms using three approaches: filter, wrapper and embedded methods. Among several filter methods, ReliefF and FCBF are tested. For the wrapper method, we applied SFS and SBS with three classifiers: NN, RF, SVM. For the embedded method, the RF classifier is adopted which produces a ranking of variables during the training process. For each method, we test the performance and evaluate the error rates of the attribute subset using the NN, RF and SVM classifiers. We predict 3D facies using an RF classifier for the best attribute subset to test the validity of the model.

Results and discussion

Table 2.2 shows attribute - attribute correlation using Pearson, rank, MI, and distance correlation measures. We note that correlation measures have different susceptibilities to non-linearity, presence of noise and outliers, and also whether or not the attributes are normally distributed. For instance, Pearson correlation coefficient changes substantially compared to rank correlation coefficient when an outlier is included (Mukaka, 2012). In case of mutual information, the response of one variable is due to both stimuli and noise (Shannon and Weaver, 1949). Figure 2.1 d, e and f show that the presence of noise decreases MI and the distance correlation coefficient significantly. The GLCM homogeneity and entropy are perfectly anti-correlated when measured with Pearson, rank and distance correlation, suggesting we can select only one of them for the subset to avoid redundancy. Amplitude attributes such as RMS amplitude and total energy are highly

correlated (corr. coeff. > 0.9). Also, both amplitude attributes are highly correlated with instantaneous envelope and peak magnitude. Because MI is more sensitive to noise or the distribution of data points, the MI coefficients are lower than those of other measures.

The attribute-to-class relationship is analyzed using ANOVA and MI (Figure 2.6). Both methods show that the amplitude family of attributes (e.g. RMS amplitude, total energy, instantaneous envelope and peak magnitude) are relevant to output classes. Also, texture attributes (GLCM and chaos) are strongly related to training classes. The difference to highlight between ANOVA and MI is that ANOVA is a linear model while MI is a non-linear model. MI shows that the output classes have high dependence with total energy and RMS amplitude. On the other hand, in the ANOVA model indicated RMS amplitude exhibits a high F-value, which indicates that RMS amplitude explains a significant amount of variation in facies. Although ANOVA and MI tell us which attributes can better differentiate the facies of interest, attributes which are selected by both methods can have redundancy. For instance, both GLCM entropy and GLCM homogeneity are highly ranked in both ANOVA and MI (Figure 2.6) which shows they are powerful variables for classification. However, we can select one of two attributes for subset. Because two attributes are perfectly anti-correlated, which indicates using both is redundant.

To take into account relevance and redundancy, we test several attribute selection algorithms to build the attribute subsets. The ten highest ranked attributes obtained from each attribute selection algorithm are shown in Table 2.3. Algorithms belonging the same categories (e.g. ReliefF and FCBF algorithms of the filter method, six algorithms of the wrapper method) show similar attribute rankings. However, the filter and the wrapper

algorithms yield quite different attribute subsets. Wrapper methods select relevant attributes, according to input-to-output dependence. At the same time, wrapper methods more efficiently reject redundant attributes. For instance, when total energy is chosen in the subset, the wrapper algorithm rejects RMS amplitude and vice-versa. Random forest variable selection is an example of embedded method which tends to choose important attributes, but they also include redundant attributes. For instance, the subset has total energy and peak magnitude close together, while GLCM homogeneity and entropy are ranked close as well. Figure 2.7 shows the error rate of the attribute subsets selected using the filter, wrapper and embedded methods. A 5-fold cross-validation is implemented to compute accuracy score and error rate when each attribute subset is applied. Input attributes were split into 5 groups randomly, then selected 1 group and remaining 4 groups were used as test dataset and training datasets respectively. The cross-validation process is repeated 5 times and the average value of accuracy is used to compute error rate. Wrapper methods reduce error rate with a small number of attributes compared to other methods, because the methods are based on the performance of the predictive model.

To understand how noise in the dataset affects classification performance, Gaussian noise with different signal to noise ratios (S/N) (noise-free, 10, 5 and 0 dB) were added to attributes of training dataset (Figure 2.8). We measure the S/N as a ratio of signal power compared to noise power in decibels (dB). The higher level of noise in data generally degrades classification accuracy. One key point of observation is using larger number of attributes significantly reduces the error rate in the case of noisy data. Especially RF and SVM wrapper models in Figure 2.8e and 2.8f show that the error rate of 0 dB data decreases

substantially as the number of attributes increases. This implies that if the data is contaminated with noise, using other attributes together can improve classification.

We tested subsets of attributes using a RF classifier with differing numbers of attributes to differentiate salt, MTDs and conformal reflectors in the same seismic volume we used for the training set (Figures 2.9, 2.10 and 2.11). The error rate with respect to the number of attributes in each subset is computed from training data and is shown in Figure 2.9b. The subset consisting of just the first highest-ranked attribute does not differentiate MTDs and conformal reflectors (Figure 2.9c). The three highest-ranked attributes distinguish MTDs and conformal reflectors better than the one attribute subset (Figure 2.9d). However, some parts of the conformal reflectors are misclassified into salt and MTD. The top seven highest-ranked attributes differentiate the three facies as effectively as twenty, the full set of attributes. The subset with the top seven highest-ranked attributes include relevant attributes that map different geology while also avoiding the redundancy. Figure 2.11 shows the predicted facies within the entire seismic volume. The salt domes which show high coherence values in Figure 2.11b are correctly predicted as salt facies in Figure 2.11c. A limitation of this classification is that some of the MTD facies are misclassified as salt because both facies are highly discontinuous and have low coherence. Also, other discontinuous geologic features, such as faults, are misclassified as MTDs.

Application 2: Barnett Shale play in Fort Worth Basin - attribute selection to differentiate limestone and shale facies

We test our workflow on the Barnett Shale to classify limestone and shale facies which are dominant in the play. In the survey area, the Barnett Shale is separated into upper and lower shale units by a thin Forestburg Limestone. The Barnett Shale, which is relatively brittle and acts as the reservoir, lies between the Marble Falls and the Viola Limestones which are more ductile (Perez and Marfurt, 2014; Li et al., 2016; Verma et al., 2016). In this example, defining the output classes of the training data is aided by wireline logs and stratigraphical interpretation. A vertical slice through a seismic line along six wells is described in Figure 2.12. Facies were estimated based on each log set including gamma ray, P-wave sonic, and bulk density. We defined the facies mainly based on gamma ray, with values of limestone ranging from 10 to 40 API, while those of shale range from 60 to 150 API. From a stratigraphical interpretation that was aided by well logs, we labeled data points adjacent to each well log as limestone and shale, which is equivalent to the training output (Figure 2.13). The input attributes are comprised of twelve physical properties calculated from prestack seismic inversion: P and S impedances, P and S velocities, V_P/V_S, Density, MuRho, Young's modulus, Poisson's ratio, Mu, Lambda and LambdaRho. The general workflow is the same as that of the first case study: attribute – attribute correlations and attribute – class correlations are analyzed, and attribute subsets are built using filter and wrapper algorithms. Among the four correlation measures, we opted for MI for the second case study since it is able to assess non-linear relationships.

The attributes that describe physical properties inverted from seismic amplitude are highly correlated with each other because many of these physical properties can be calculated from other physical properties in the attribute set (Table 2.4). Elastic properties can especially be determined from thr two elastic moduli in the case of homogeneous isotropic media. In terms of attribute – class correlations, MI coefficients from 7 attributes (P impedance, Young's modulus, P velocity, Mu, S velocity, MuRho and S impedance) are higher than 0.5 (Figure 2.14b). From the correlation analysis we found attributes are highly correlated with each other and many of the attributes are also highly correlated to the corresponding target class. Young's modulus and P impedance are highly ranked in the sequential forward selection of the wrapper methods (Table 2.5), and decrease error rate efficiently (Figure 2.15 d, e and f). The score in Figure 2.15 does not increase significantly after three components in each method, since the other attributes are highly correlated to these three attributes. Like the first case study, we tested subsets with differing numbers of attributes to differentiate shale and limestone in the same seismic volume we used for the training set (Figure 2.16). Two thin lime layers are intervening in the Barnett shale, which is interpreted in well logs section in Figure 2.13: a limestone layer in the Upper Barnett and the Forestburg limestone between the Upper and Lower Barnett Shale. The subsets with the four highest-ranked attributes (Young's modulus, P impedance, density, and Lambda) differentiate thin limestone layers between the Barnett shale as effectively as twelve attributes (Figure 2.16).

Conclusions

Analyzing attribute-to-attribute dependence and attribute-to-class relationships helps to understand which attributes are redundant and which are relevant. However, a high correlation between attributes does not always imply that attributes are redundant. We need to analyze all attributes together using a framework which can quantitatively rank the attributes to build a subset. The multivariate attribute selection algorithms result in the subsets which have smaller number of attributes but show good performance in differentiating salt and MTD facies from conformal reflectors. From a geological point of view, it is challenging to define the depositional environments in the survey area into only three discrete classes. Turbidites, faults, overpressured shale, and seismic noise will be misclassified into one of the target classes. However, understanding each seismic attribute's characteristic is crucial to implement automated facies classification and to aid rendering of a seismic volume where the interpreters target. Even though the case study is focused on mapping different facies in geology, the attribute selection algorithms can be applied to other supervised classification problems. For instance, the workflow can be applied to select physical properties and seismic attributes to yield reservoir properties such as porosity, permeability and brittleness from input quantitative interpretation attributes.

Acknowledgements

The authors thank the sponsors of OU Attribute Assisted Seismic Processing and Interpretation (AASPI) consortium for their encouragement and financial support.

Appendix A

Correlation measures

Pearson's product-moment correlation (Pearson, 1895)

A correlation measure which is most widely used is Pearson's product-moment coefficient. The correlation between two random variables *X* and *Y* is defined as

$$\boldsymbol{corr(X,Y) = \frac{cov(X,Y)}{\sigma_x \sigma_y}} \quad (1)$$

where cov(*X, Y*) is the covariance between *X* and *Y*, σ_x and σ_y are the standard deviation of *X* and *Y*, respectively. The Pearson's coefficient describes linear dependence between two variables. Among scatter plots in Figure 2.1a, 2.1b and 2.1c, only 1a is perfectly correlated or anti-correlated in terms of Pearson's correlation.

Spearman's rank correlation (Spearman, 1904)

Spearman correlation coefficient is defined as Pearson correlation between the ranked variables. A positive Spearman correlation corresponds to an increasing monotonic trend, while a negative one corresponds to a decreasing monotonic trend between two random variables. The correlation assesses positive or negative relationships whether they are linear or not. According to the Spearman's definitions of correlation, two variables X and Y in Figure 2.1 b are highly correlated even if the relationship is nonlinear.

Mutual information (Shannon and Weaver, 1949; Cover and Thomas, 1991).

In information theory, the uncertainty involved in the value of a random variable is quantified as entropy. The Shannon entropy which is a measure of the uncertainty of a random variable is defined as

$$\boldsymbol{H} = -\sum_i \boldsymbol{p_i log(p_i)} \tag{2}$$

where $\mathbf{p_i}$ is the probability of occurrence of the i-th possible value of the source symbol. Mutual information measures the gain of information about one random variable by observing another. Mutual information of two discrete random variables **X** and **Y** are denoted by

$$\begin{aligned} \boldsymbol{I(X;Y)} &= \boldsymbol{H(X) - H(X|Y)} \\ &= \boldsymbol{H(Y) - H(Y|X)} \end{aligned} \tag{3}$$

Where H(**X**), H(**Y**) are marginal entropies and H(X|Y), H(Y|X) are conditional entropies. Substituting equation (2) to equation (3) gives

$$\boldsymbol{I(X;Y)} = \sum_{y \in Y} \sum_{x \in X} \boldsymbol{p(x,y) log} \left(\frac{\boldsymbol{p(x,y)}}{\boldsymbol{p(x)p(y)}} \right) \tag{4}$$

where p(x), p(y) are marginal probability functions and p(x, y) is the joint probability function.

Distance correlation (Székely et al. 2007)

The distance correlation of two random variables is defined as distance covariance divided by distance standard deviation. Distance correlation is denoted as

$$\boldsymbol{dCor(X,Y)} = \frac{\boldsymbol{dCov(X,Y)}}{\sqrt{\boldsymbol{dVar(X) dVar(Y)}}} \tag{5}$$

where dCov(X, Y) is distance covariance, dVar(X) and dVar(Y) are distance variance of X and Y, respectively. In contrast to Pearson's covariance which is defined as inner product of two centered vectors, distance covariance is defined as product of centered Euclidean distances $D(x_i, x_j)$ and $D(y_i, y_j)$ in arbitrary dimensions.

$$\boldsymbol{dCov(X,Y) = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} D(x_i, x_j) \cdot D(y_i, y_j)} \quad (6)$$

where $x \in X$, $y \in X$ and n is number of samples of X and Y. Distance correlation can detect non-linear relationships and its values are non-negative.

Appendix B

Attribute selection methods

1) Filter

Relief and ReliefF

Relief algorithm *(Kira and Rendell, 1992)* estimates attributes according to how well their values differentiate among the instances near to each other. Relief searches for its two nearest neighbors: one from the same class which is called nearest hit and the other from a different class called nearest miss. At each iteration, Relief estimates weight vector W of a given attribute,

$$W_i = W_i + (x_i - nearest\ miss(x)_i)^2 - (x_i - nearest\ hit(x)_i)^2 \quad (7)$$

where x is an instance randomly selected in training data. Attributes are selected if their average weight is greater than a threshold τ. ReliefF *(Kononenko, 1994,* 1997*)* improved

Relief by estimating probabilities more reliably and extended the algorithm to handle noisy, incomplete and multi-class data sets.

Correlation based feature selection (CFS) and Fast correlation-based filter (FCBF)

CFS's feature subset evaluation function (Hall, 1999) is

$$M_s = \frac{k\overline{r_{cf}}}{\sqrt{k+k(k-1)\overline{r_{ff}}}} \tag{8}$$

where M_s is the heuristic merit of a feature subset containing k attributes, $\overline{r_{cf}}$ is the mean attribute – class correlation and $\overline{r_{ff}}$ is the average attribute – attribute correlation. FCBF (Yu and Liu, 2003) starts with full set of features, uses symmetrical uncertainty to calculate dependences of features and finds best subset using backward selection for high dimensional data.

2) Wrapper

Sequential selection algorithms

Sequential forward selection (SFS) algorithm (Kittler, 1978) starts from an empty set and sequentially adds the attributes that maximize classification accuracy. The process is repeated until the required number of features are added. Sequential backward selection (SBS) algorithm starts from the full set and sequentially removes the attribute that its removal gives the lowest decrease in classification performance. Sequential floating forward selection (SFFS) and sequential floating backward selection (SFBS) (Pudil et al.,

1994) introduce an additional backtracking step, which is more flexible than the simple SFS algorithm (Chandrashekar and Sahin, 2014).

3) Embedded

Attribute importance in tree-based methods (Random forest)

In the classification and regression tree (CART) algorithm (Breiman et al., 2001, 2002), the best split is made using Gini impurity at each internal node for prediction. Gini importance can be computed as byproduct during training process of tree-based predicted model which is given by

$$Gini\ Importance\ i_G(\theta) = \sum_T \sum_\tau \Delta i_\theta(\tau, \mathrm{T}), \tag{9}$$

where, $\Delta i(\tau)$ is node purity gain which is denoted

$$\Delta i(\tau) = i(\tau) - p_l i(\tau_l) - p_r i(\tau_r). \tag{10}$$

Decrease in Gini impurity, $\Delta i(\tau)$ results from splitting the samples to left and right sub-nodes.

References

Abdi, H., and L. J. Williams, 2010, Principal component analysis: Wiley Interdisciplinary Reviews: Computational Statistics, **2**, 433-459.

Bahorich, M. S., and S. L. Farmer, 1995, 3-D seismic coherency for faults and stratigraphic features: The Leading Edge, 1053-1058.

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

Berthelot, A., A. H. Solberg, and L. J. Gelius, 2013, Texture attributes for detection of salt: Journal of Applied Geophysics, **88**, 52-69.

Breiman, L., 2001. Random forests: Mach. Learning, **45**, 5–32.

Brown, A.R., 2011, Interpretation of Three-dimensional Seismic Data, Society of Exploration Geophysicists and American Association of Petroleum Geologists.

Chandrashekar, G., and F. Sahin, 2014, A survey on feature selection methods: Computers & Electrical Engineering, **40**, 16-28.

Chopra, S., and K. J. Marfurt, 2005, Seismic attributes—A historical perspective: Geophysics, **70**, 3SO-28SO.

Coléou, T., M., Poupon and K. Azbel, 2003, Unsupervised seismic facies classification: A review and comparison of techniques and implementation: The Leading Edge, **22**, 942-953.

Cover, T. M. and J. A. Thomas, 1991, Entropy, relative entropy and mutual information: Elements of Information Theory, **2**, 1-55.

Frey-Martinez, J., 2010, 3D Seismic interpretation of mass transport deposits: Implications for basin analysis and geohazard evaluation: In Submarine Mass Movements and Their Consequences, Springer, Dordrecht, 553-568.

Guyon, I., and A. Elisseeff, 2003, An introduction to variable and feature selection: Journal of Machine Learning Research, **3**, 1157-1182.

Hall, M. A., 1999, Correlation-based feature selection for machine learning.

Hall, M. A., 2000, Correlation-based feature selection of discrete and numeric class machine learning: Proc. 17th Int'l Conf. Machine Learning, 359-366.

Hughes, G., 1968, On the mean accuracy of statistical pattern recognizers: IEEE transactions on information theory, **14**-1, 55-63.

Jain, A., and D. Zongker, 1997, Feature selection: Evaluation, application, and small sample performance: IEEE Transactions on Pattern Analysis and Machine Intelligence, **19**, 153-158.

Jones, I. F., and I. Davison, 2014, Seismic imaging in and around salt bodies: Interpretation, **2**, SL1-SL20.

Kira, K., and L. A. Rendell, 1992, A practical approach to feature selection: In International Conference on Machine Learning, 368–377.

Kittler, J., 1978, Feature set search algorithms: In Pattern Recognition and Signal Processing, 41–60.

Kononenko, I., 1994, Estimating attributes: analysis and extensions of RELIEF: In European conference on machine learning, Springer, Berlin, Heidelberg.

Kononenko, I., E. Šimec, and M. Robnik-Šikonja, 1997, Overcoming the myopia of inductive learning algorithms with RELIEFF: Applied Intelligence, **7** - 1, 39-55.

Li, F., S. Verma, H. Zhou, T. Zhao, and K. J. Marfurt, 2016, Seismic attenuation attributes with applications on conventional and unconventional reservoirs: Interpretation, **4**-1, SB63-SB77.

Liaw, A. and M. Wiener, 2002, Classification and regression by random Forest: R News, **2** - 3, 18–22.

Marfurt, K. J., R. L. Kirlin, S. L. Farmer, and M. S. Bahorich, 1998, 3-D seismic attributes using a semblance-based coherency algorithm: Geophysics, **63**, 1150-1165.

Mukaka, M. M., 2012, A guide to appropriate use of correlation coefficient in medical research: Malawi Medical Journal, **24**, 69-71.

Nelson, C. H., C. A. Escutia, J. E. Damuth, and D. C. Twichell, 2011, Interplay of mass-transport and turbidite-system deposits in different active tectonic and passive continental margin settings: External and local controlling factors: Sediment. Geol, **96**, 39-66.

Omosanya, K. O., and T. M. Alves, 2013, A 3-dimensional seismic method to assess the provenance of Mass-Transport Deposits (MTDs) on salt-rich continental slopes (Espírito Santo Basin, SE Brazil): Marine and Petroleum Geology, **44**, 223-239.

Pearson, K., 1894, Contributions to the mathematical theory of evolution: Philosophical Transactions A, **185**, 71-110.

Peng, H., F. Long, and C. Ding, 2005, Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy: IEEE Transactions on Pattern Analysis and Machine Intelligence, **27**, 1226-1238.

Pudil, P., J. Novovičová, and J. Kittler, 1994, Floating search methods in feature selection. Pattern recognition letters, **15**, 1119-1125.

Perez, R., and K. J. Marfurt, 2014, Mineralogy-based brittleness prediction from surface seismic data: Application to the Barnett Shale: Interpretation, 2, no. 4, T255–T271

Qi, J., T. Lin, T. Zhao, F. Li, and K. J. Marfurt, 2016, Semisupervised multiattribute seismic facies analysis: Interpretation, **4**, SB91-SB106.

Randen, T., S. Pedersen, and L. Sonneland, 2001, Automatic extraction of fault surfaces from three-dimensional seismic data: 71st Annual International Meeting, SEG, Expanded Abstracts, 551–554.

Robnik-Šikonja, M., and I. Kononenko, 2003, Theoretical and empirical analysis of ReliefF and RReliefF: Machine learning, **53**(1-2), 23-69.

Roden, R., T., Smith, and D. Sacrey, 2015, Geologic pattern recognition from seismic attributes:Principal component analysis and self-organizing maps: Interpretation, **3**, SAE59-SAE83.

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

Roy, A., M. Matos, and K. J. Marfurt, 2010, Automatic seismic facies classification with kohonen self organizing maps-a tutorial: Geohorizons Journal of Society of Petroleum Geophysicists, 6-14.

Sánchez-Maroño, N., A. Alonso-Betanzos, and M. Tombilla-Sanromán, 2007, Filter methods for feature selection–a comparative study: In International Conference on

Intelligent Data Engineering and Automated Learning, Springer, Berlin, Heidelberg, 178-187

Shannon, C., and W. Weaver, 1949, The mathematical theory of communication: University of Illinois Press.

Spearman, C., 1904, The proof and measurement of association between two things: American journal of Psychology, **15**-1, 72-101.

Székely, G. J., M. L. Rizzo, and N. K. Bakirov, 2007, Measuring and testing dependence by correlation of distances: The annals of statistics, **35**-6, 2769-2794.

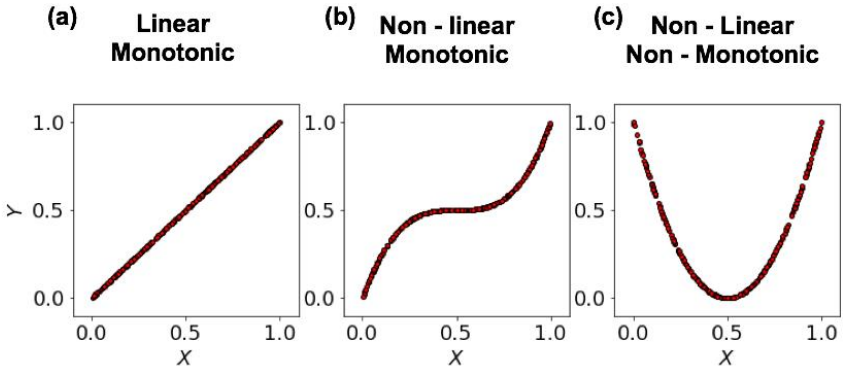
Verma, S., T. Zhao, K. J. Marfurt, and D. Devegowda, 2016, Estimation of total organic carbon and brittleness volume: Interpretation, **4**-3, T373-T385.

Yu, L., and H. Liu, 2003, Feature selection for high-dimensional data: A fast correlation-based filter solution: Proceedings of the 20th international conference on machine learning (ICML-03) 856-863.

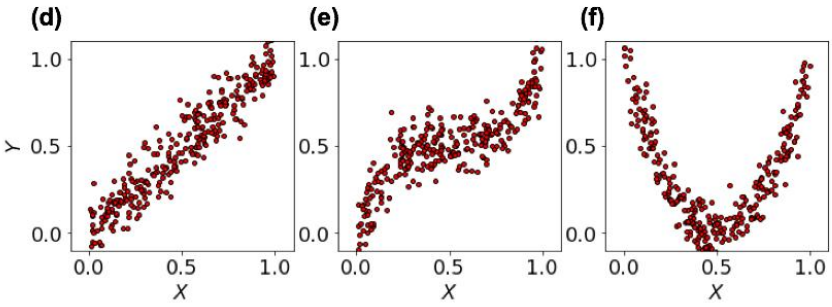
Yu, L., and H. Liu, 2004, Efficient feature selection via analysis of relevance and redundancy: Journal of Machine Learning Research, **5**, 1205-1224.

Zhao, T., F. Li, and K. J. Marfurt, 2018, Seismic attribute selection for unsupervised seismic facies analysis using user-guided data-adaptive weights: Geophysics, **83**, O31-O44.

Figures



Correlation between X and Y	Pearson corr.=1.00 Rank corr.=1.00 MI=1.00 Distance corr.=1.00	Pearson corr.=0.92 Rank corr.=1.00 MI=1.00 Distance corr.=0.84	Pearson corr.=0.05 Rank corr.=-0.01 MI=1.00 Distance corr.=0.26
Regression score between X and Y	Linear Bayesian = 1.00 DNN = 1.00 RF = 1.00 SVM = 1.00	Linear Bayesian = 0.84 DNN = 1.00 RF = 1.00 SVM = 1.00	Linear Bayesian = -0.04 DNN = 1.00 RF = 1.00 SVM = 1.00



Correlation between X and Y	Pearson corr.=0.95 Rank corr.=0.95 MI=0.89 Distance corr.=0.88	Pearson corr.=0.84 Rank corr.=0.83 MI=0.62 Distance corr.=0.65	Pearson corr.=0.02 Rank corr.=-0.01 MI=0.87 Distance corr.=0.23
Regression score between X and Y	Linear Bayesian = 0.89 DNN = 0.89 RF = 0.84 SVM = 0.89	Linear Bayesian = 0.68 DNN = 0.81 RF = 0.74 SVM = 0.81	Linear Bayesian = -0.02 DNN = 0.90 RF = 0.86 SVM = 0.90

Figure 2.1. Different types of relationships between variables X and Y and their correlation coefficients and regression score. Each scatter plot describes a different relationship between X and Y: (a) and (c) linear and monotonic relationship, (b) and (e) non-linear,

monotonic relationship, (c) and (f) non-linear, non-monotonic relationships. Gaussian noise of 10 percent has been added to variable Y in (d), (e) and (f). Coefficients are computed using Pearson, rank, mutual information, and distance correlation methods. Regression score are computed with linear Bayesian, NN, RF, SVM repressor predictive algorithms. The best hyper parameters for each model were obtained using a grid-search algorithm.

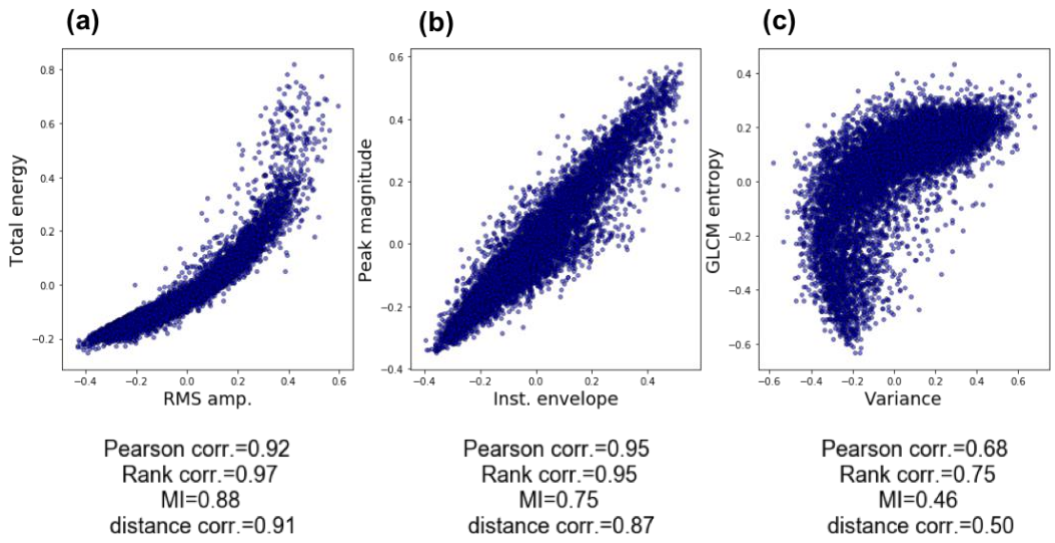


Figure 2.2. Relations between different attribute pairs (a) total energy vs. RMS amplitude, (b) peak magnitude vs. instantaneous envelope, and (c) GLCM entropy vs. variance. Correlation coefficients are computed using Pearson, rank, mutual information and distance measures.

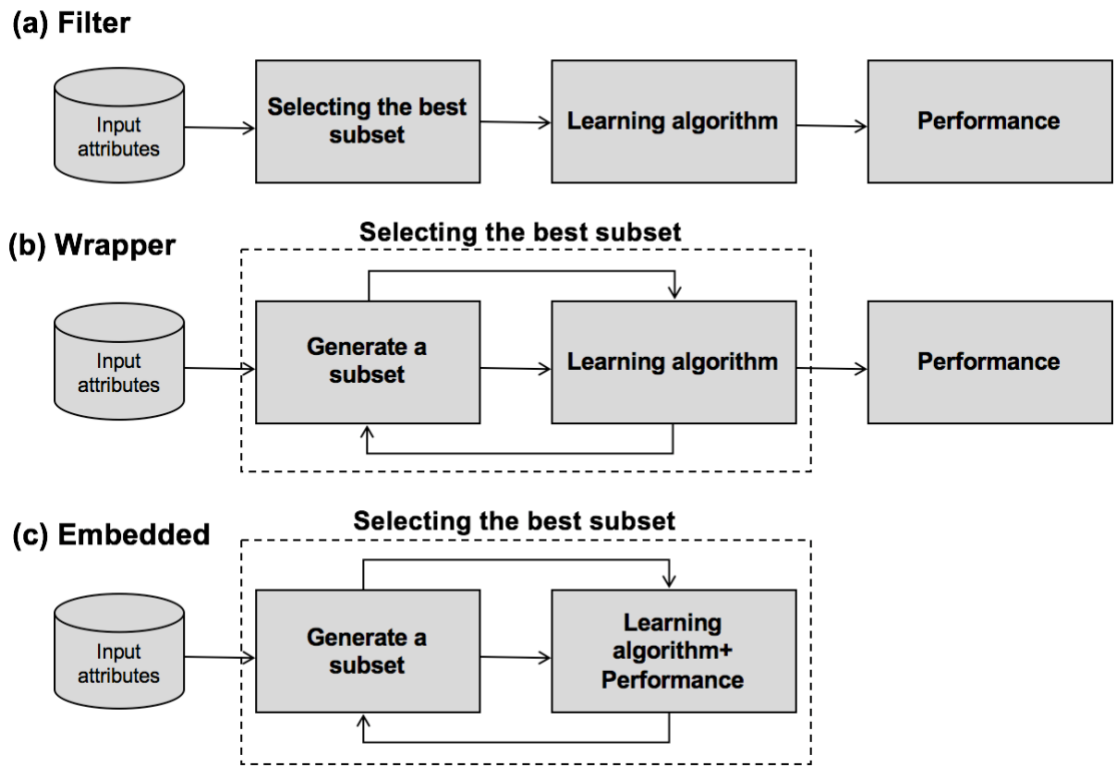


Figure 2.3. Schematic diagram summarizing the the steps from the (a) filter, (b) wrapper and (c) embedded attribute subset selection workflows. Note that there is no feedback in the filter workflow. The examples of each method and their mathematical description is given in the Appendix.

Categories of seismic attributes evaluated in facies classification

Amplitude attributes	Instantaneous attributes	Geometric attributes	Texture attributes	Spectral attributes
RMS amplitude	Instantaneous envelope	Variance	Chaos	Peak magnitude
Total energy	Instantaneous frequency	Dip magnitude	GLCM entropy	Peak frequency
Relative acoustic impedance	Instantaneous phase	Dip azimuth	GLCM homogeneity	Peak phase
		Most-positive curvature		
		Most-negative curvature		
		Curvedness		
		Aberrancy magnitude		
		Aberrancy azimuth		

Table 2.1. The seismic attribute families and the 20 specific attributes used to classify conformal sediments, salt and mass transport deposits.

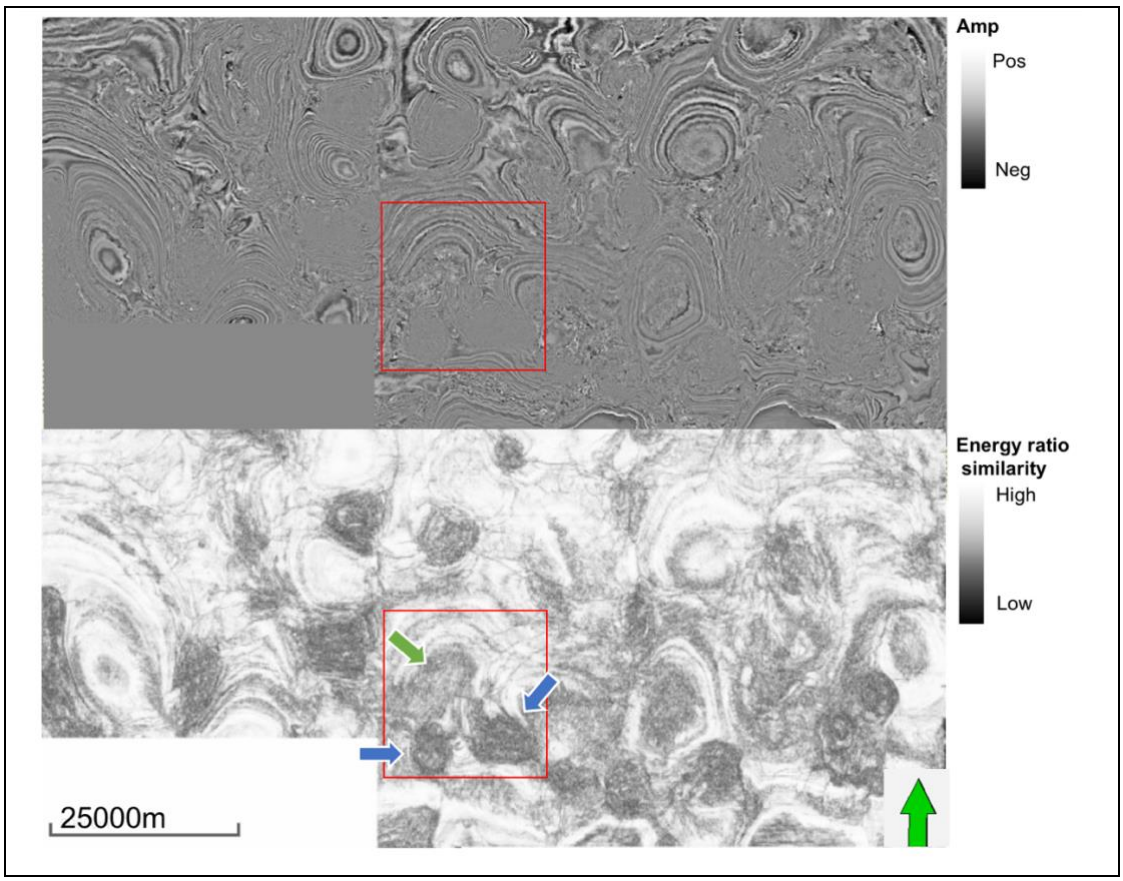


Figure 2.4. Time slice through seismic amplitude (top) and energy ratio similarity attribute (bottom). The red box indicates the volume where the training data are sampled. Green arrow indicates a mass transport deposits (MTDs) while blue arrows indicate salt diapirs, both of which exhibit low value of similarity.

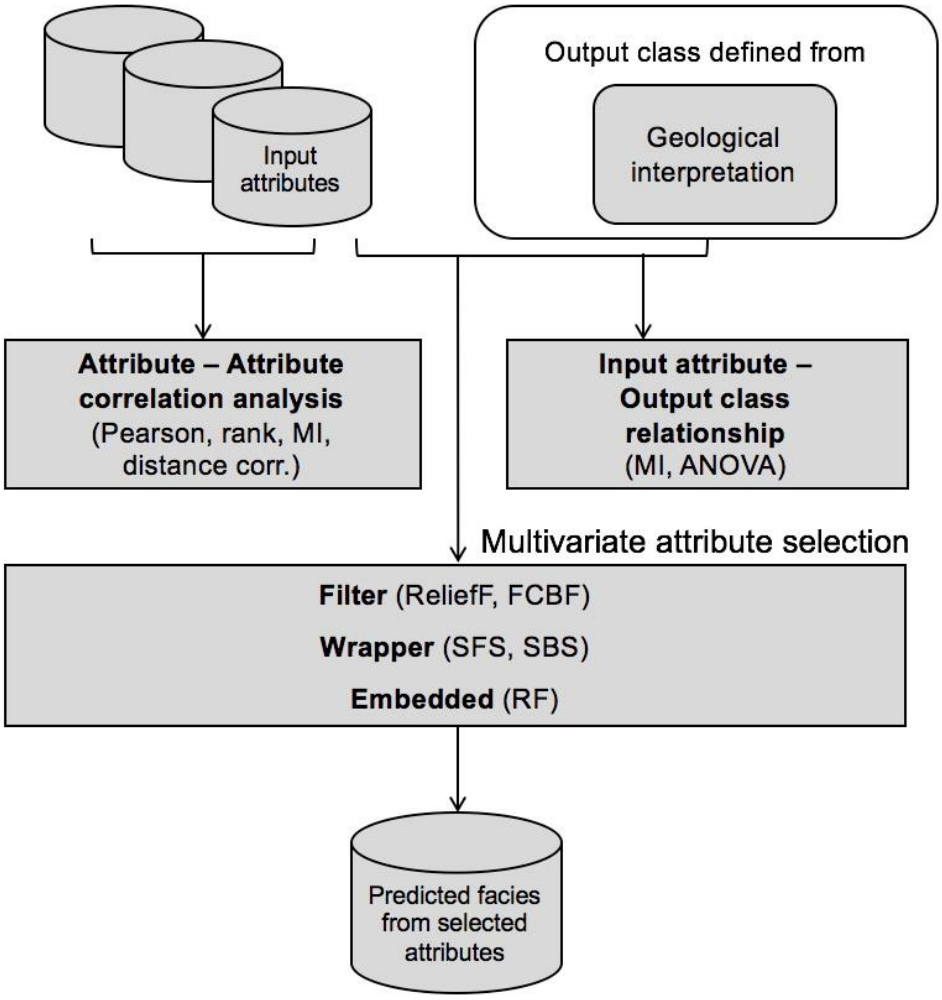


Figure 2.5. The workflow to select the best subset of attributes based on geologic relevance as well as attribute-to-attribute redundancy using three types of multivariate approaches.

Attribute – attribute correlation analysis

Correlation measures	Attributes highly correlated with the other attributes (corr. coeff. > 0.6)
Pearson correlation	GLCM entropy – GLCM homogeneity (-1.0)
	Instantaneous envelope – Peak magnitude (0.96)
	RMS amplitude - Instantaneous envelope (0.93)
	RMS amplitude – Peak magnitude (0.90)
	RMS amplitude – Total energy (0.90)
	Total energy – Instantaneous envelope (0.84)
	Total energy – Peak magnitude (0.83)
	Instantaneous phase – Relative acoustic impedance (0.73)
	GLCM entropy – Chaos (0.71)
	GLCM entropy – Variance (0.70)
	Instantaneous frequency – Peak frequency (0.62)
Rank correlation	GLCM entropy – GLCM homogeneity (-1.0)
	RMS amplitude – Total energy (0.99)
	Instantaneous envelope – Peak magnitude (0.95)
	RMS amplitude – Instantaneous envelope (0.94)
	RMS amplitude – Total energy (0.93)
	RMS amplitude – Peak magnitude (0.92)
	Total energy – Peak magnitude (0.82)
	Instantaneous phase – Relative acoustic impedance (0.81)
	GLCM entropy – Variance (0.80)
	GLCM entropy – Chaos (0.71)
	Instantaneous frequency – Peak frequency (0.64)
	Instantaneous envelope – GLCM homogeneity (0.62)
Mutual information	RMS amplitude – Total energy (0.9)
	GLCM entropy – GLCM homogeneity (0.85)
	Instantaneous envelope - Peak magnitude (0.74)
	RMS amplitude – Instantaneous envelope (0.72)
	Total energy – Instantaneous envelope (0.71)
	RMS amplitude – Peak magnitude (0.69)
	Total energy – Peak magnitude (0.68)
Distance correlation	GLCM entropy – GLCM homogeneity (0.97)
	RMS amplitude – Total energy (0.92)
	Instantaneous envelope – Peak magnitude (0.87)
	RMS amplitude – Instantaneous envelope (0.83)
	RMS amplitude – Peak magnitude (0.78)
	Total energy – Instantaneous envelope (0.78)
	Total energy – Peak magnitude (0.74)

Table 2.2. Attribute-to-attribute attribute correlation analysis using Pearson, rank, MI, and distance correlations. Attribute pairs exhibiting high correlation (correlation coefficient > 0.6) are ranked in descending order. Pairs with yellow background are the amplitude attributes or the attributes highly correlated with amplitude attributes. Pairs with green background are texture attributes or the attributes highly correlated with texture attributes.

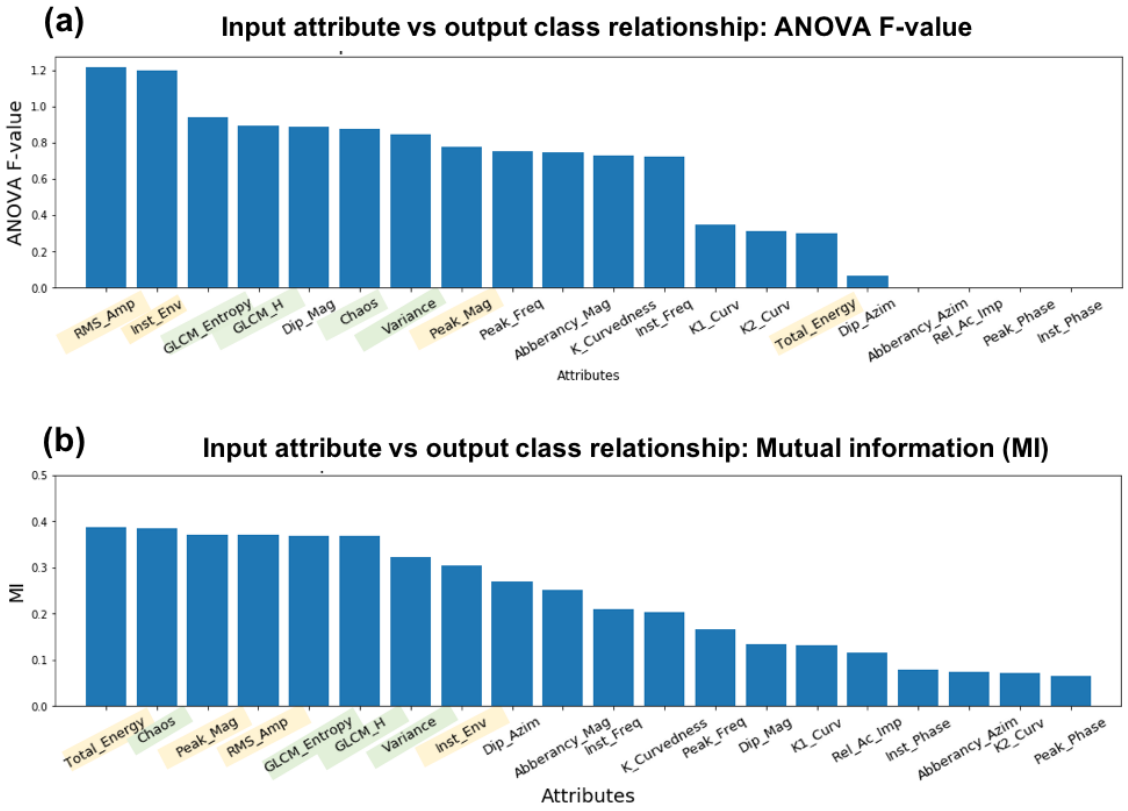


Figure 2.6. Relationships between a single input attribute and the desired output classes using (a) analysis of variance (ANOVA) F-value, and (b) mutual information. Both analyses show amplitude and texture type attributes are important variables for classification. (Yellow background indicates the amplitude attributes or the attributes highly correlated with amplitude attributes. Green background indicates texture attributes or the attributes highly correlated with texture attributes.) Unless the prediction is restricted to a specific horizon, phase varies between 180° and +180° with increasing time and is poorly correlated to output class. Examining Figure 2.4, it is clear that the azimuth of reflector dip, faults, and flexures for this data set also varies between 180° and +180° and is not correlated to any one facies.

Attribute selection algorithms			Ranking of attributes (Ten highest ranked attributes)				
Filter	ReliefF		Peak phase	Peak frequency	Total energy	Relative acoustic impedance	Instantaneous Envelope
			Instantaneous frequency	Instantaneous phase	Variance	Dip magnitude	Dip azimuth
	FCBF		Peak freqency	Peak phase	RMS amplitude	Total energy	Relative acoustic impedance
			Instantaneous Envelope	Instantaneous frequency	Instantaneous phase	Variance	Dip magnitude
Wrapper	SFS	NN	Total energy	Chaos	Aberrancy magnitude	Instantaneous frequency	Variance
			Dip azimuth	Dip magnitude	Aberrancy azimuth	Peak frequency	Most-positive curvature
		RF	Total energy	Chaos	Aberrancy magnitude	Instantaneous frequency	Dip azimuth
			Variance	Dip magnitude	Curvedness	Peak frequency	Aberrancy azimuth
		SVM	RMS amplitude	Chaos	Aberrancy magnitude	Instantaneous frequency	Variance
			Dip Magnitude	Peak frequency	Curvedness	Dip azimuth	Aberrancy azimuth
	SBS	NN	Total energy	Variance	Dip magnitude	Instantaneous frequency	Aberrancy magnitude
			Dip azimuth	Aberrancy azimuth	Chaos	Peak frequency	Most-positive curvature
		RF	Total energy	Chaos	Aberrancy magnitude	Peak frequency	Variance
			Dip azimuth	Dip magnitude	Most-positive curvature	Instantaneous frequency	Most-negative curvature
		SVM	RMS amplitude	Chaos	Aberrancy magnitude	Instantaneous frequency	Variance
			Dip magnitude	Peak frequency	Most-negative curvature	Dip azimuth	Aberrancy azimuth
Embedded	RF		Total energy	Peak magnitude	Chaos	RMS amplitude	Aberrancy magnitude
			GLCM homogeneity	GLCM entropy	Instantaneous Envelope	Variance	Instantaneous frequency

Table 2.3. Selected attribute subsets using filter (RelifF, FCBF), wrapper (NN, RF, SVM) and embedded (RF) methods. Each subset includes the ten best attributes ranked in descending order (from left to right).

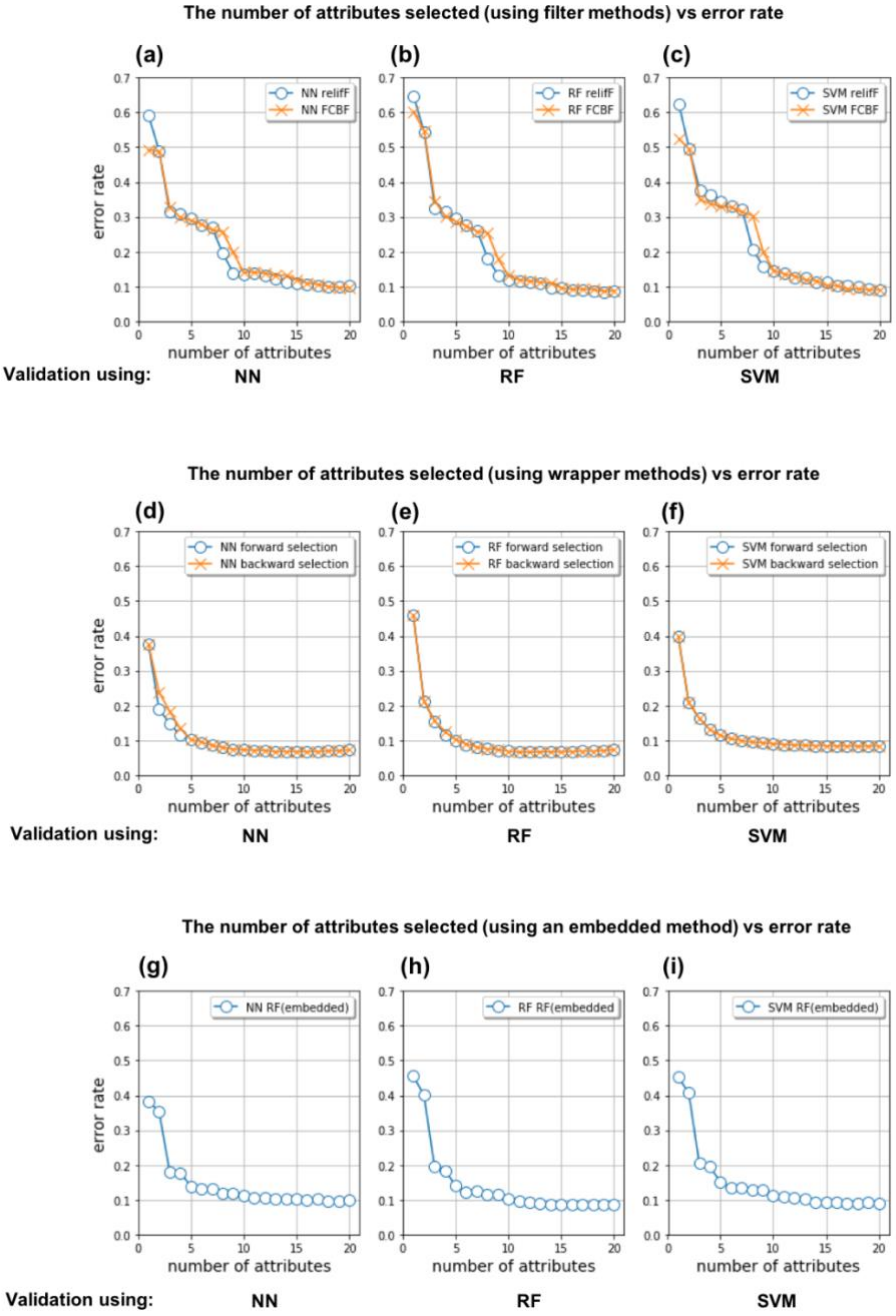
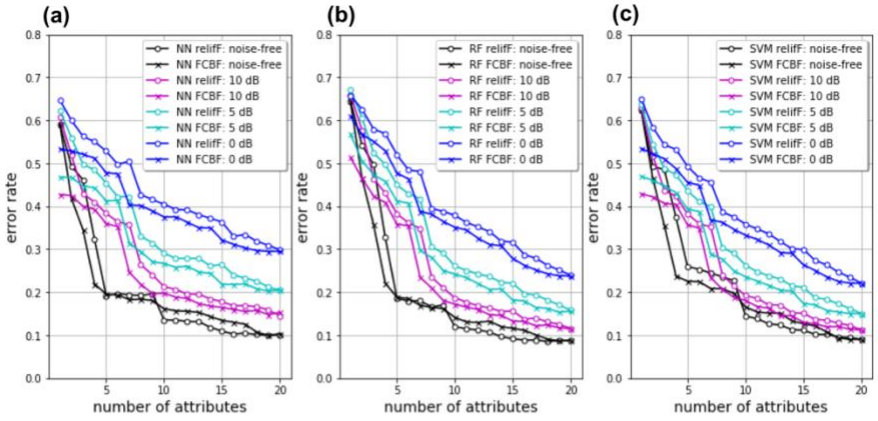


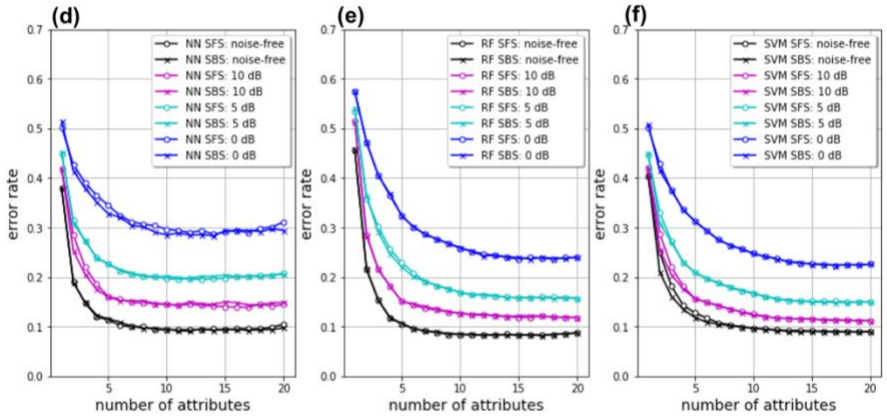
Figure 2.7. The number of attributes included in the attribute subset vs error rate. For (a), (b), and (c) Attributes in the subset were selected using filter methods (RelifF, FCBF). For (d), (e) ,and (f) attributes included in the subset were selected using wrapper methods (SFS, SBS). For (g), (h), and (i) attributes included in the subset were selected using an embedded method (RF). Each attribute subset was validated using NN, RF and SVM classifiers.

The number of attributes selected (using filter methods) vs error rate



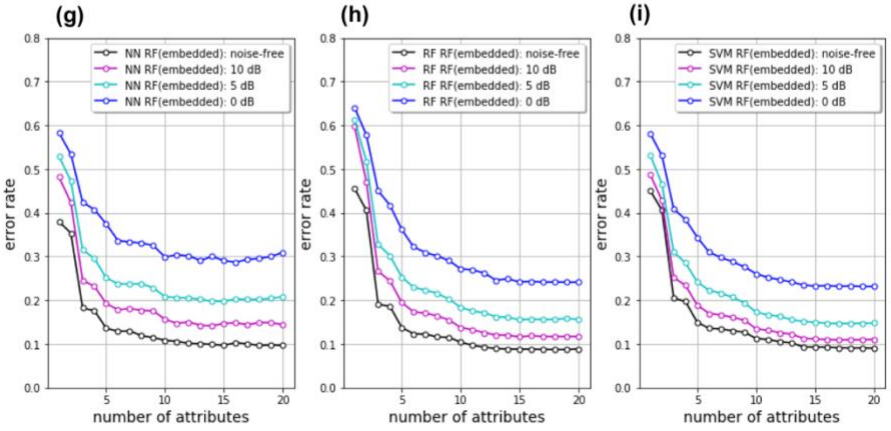
Validation using: **NN** **RF** **SVM**

The number of attributes selected (using wrapper methods) vs error rate



Validation using: **NN** **RF** **SVM**

The number of attributes selected (using an embedded method) vs error rate



Validation using: **NN** **RF** **SVM**

Figure 2.8. The number of attributes included in the attribute subset vs error rate when Gaussian noise with different S/N is added (noise-free, 10, 5 and 0 dB) to attributes. For (a), (b), and (c) a attributes in the subset were selected using filter methods (RelifF, FCBF). For (d), (e), and (f) attributes included in the subset were selected using wrapper methods (SFS, SBS). For (g), (h), and (i) attributes included in the subset were selected using an embedded method (RF). Each attribute subset was validated using NN, RF and SVM classifiers.

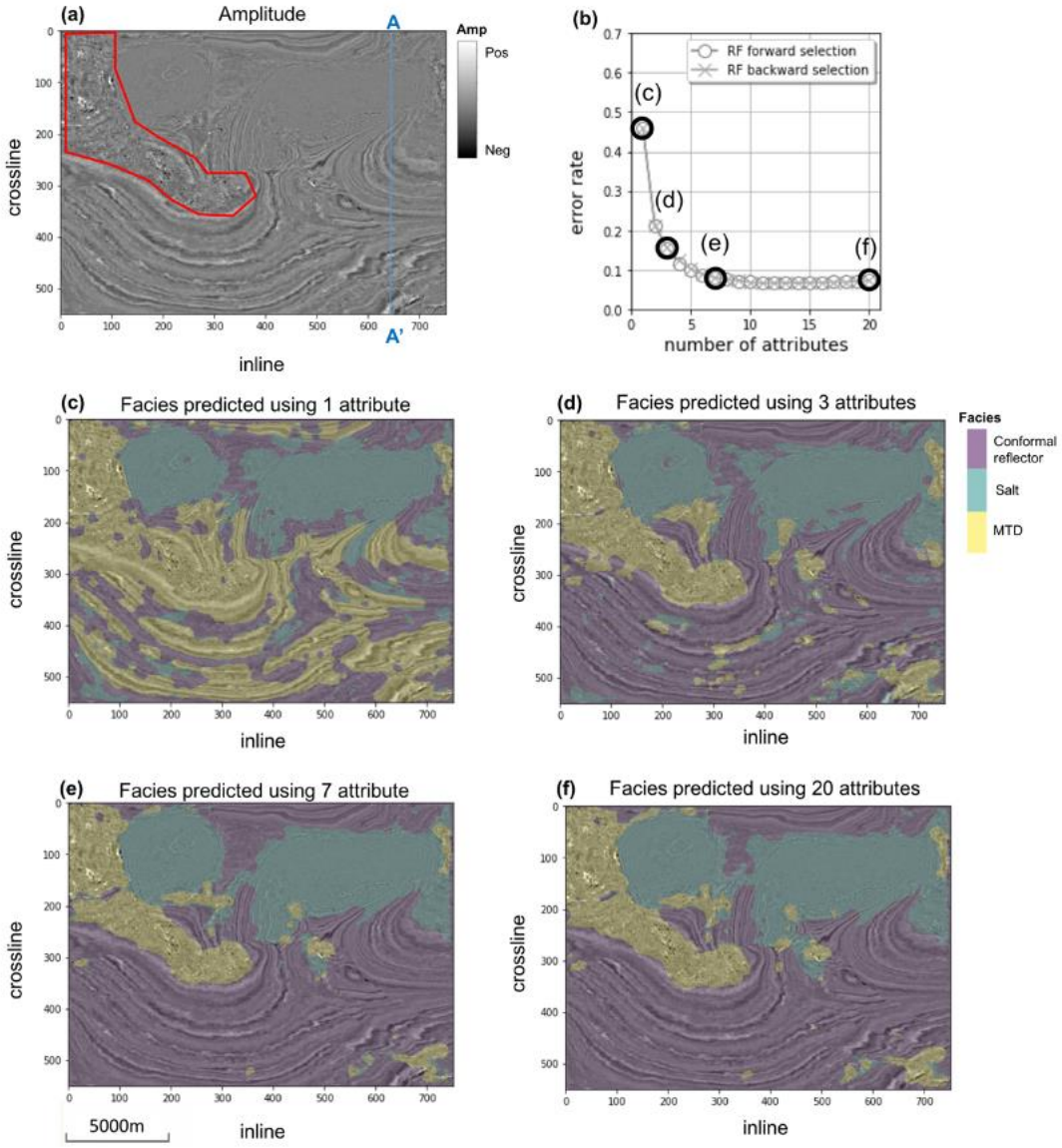


Figure 2.9. (a) A representative time slice at t = 1.1 s through amplitude, (b) error rate with respect to the number of attributes in the subset selected by the wrapper method (RF) using training data. Facies predicted using (c) the highest-ranked attribute, (d) top three highest-ranked attributes, (e) top seven highest-ranked attributes selected by the wrapper method (RF), and (f) all twenty attributes. The red polygon in (a) is human-delineated a MTD.

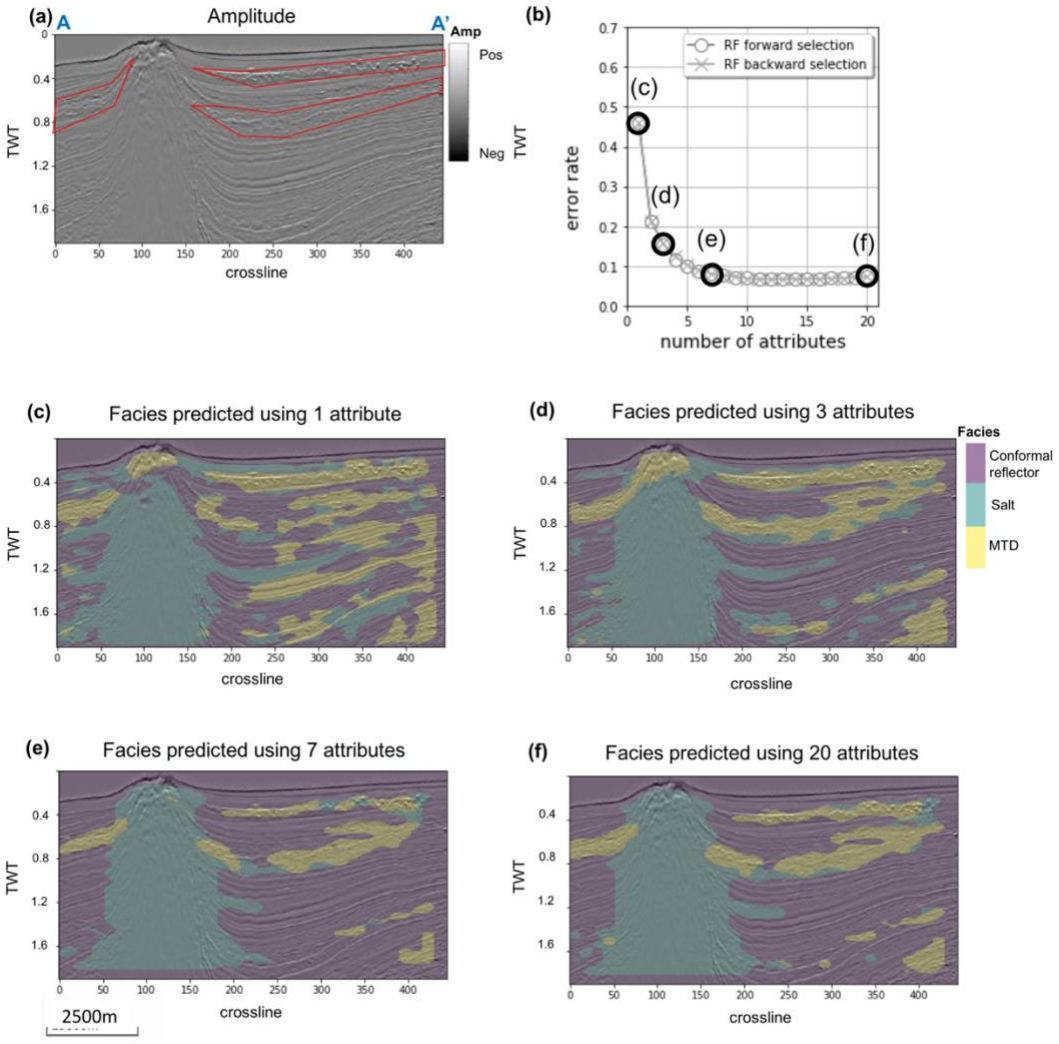


Figure 2.10. (a) A representative vertical slice along line AA' through amplitude, (b) error rate with respect to the number of attributes in the subset which is selected by the wrapper method (RF) using training data. Facies predicted using (c) the highest-ranked attribute, (d) top three highest-ranked attributes, (e) top seven highest-ranked attributes selected by the wrapper method (RF), and (f) all twenty attributes. The three red polygons in (a) are human-interpreted MTDs.

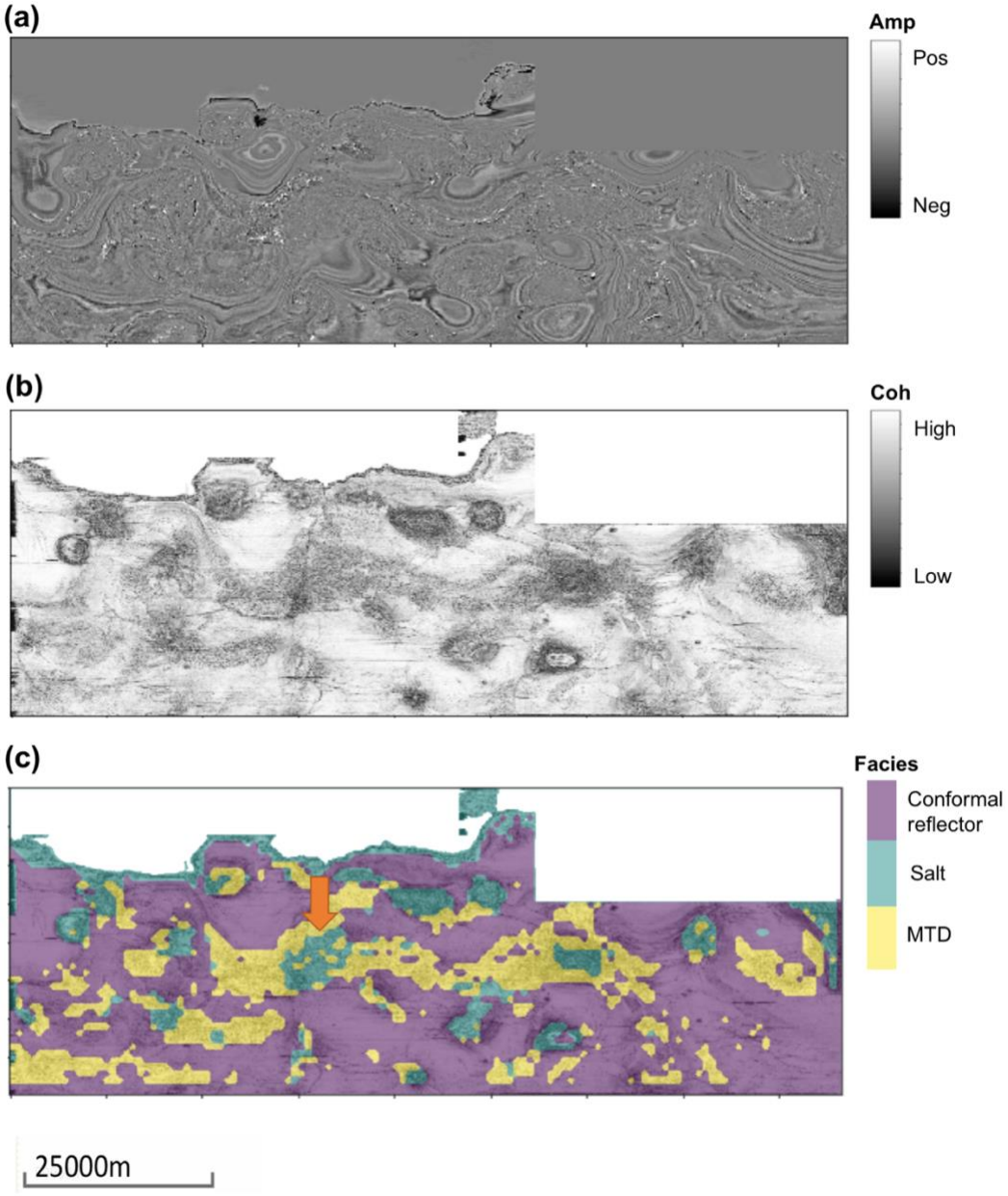


Figure 2.11. A time slice at t = 0.612 s through (a) amplitude, (b) coherence, and (c) facies predicted using 7 high-ranked attributes using the wrapper method (RF). The arrow in (c) indicates the area where MTDs are misclassified as salt.

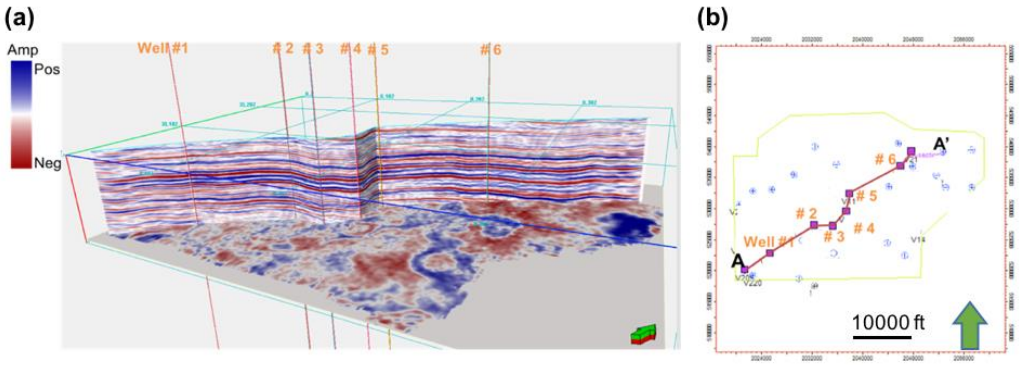


Figure 2.12. (a) The vertival slice along AA' and representative time slice through seismic amplitude, and (b) map of seismic survey and location of wells used for data training.

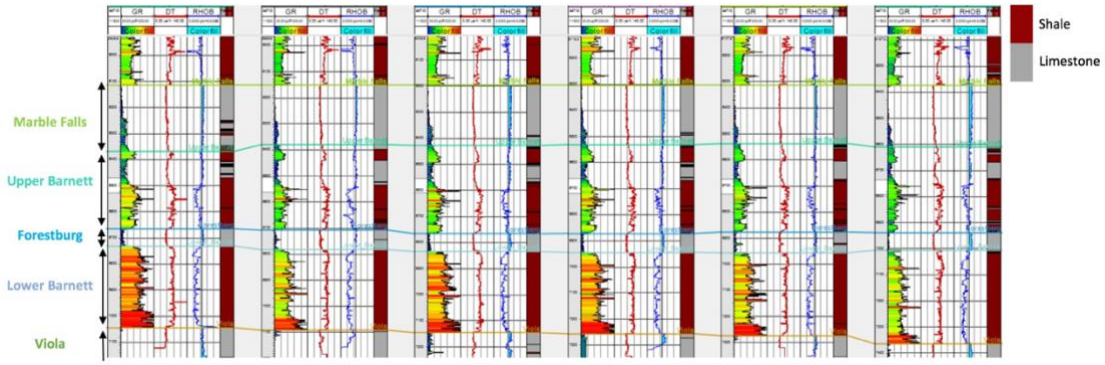


Figure 2.13. Well logs through the Barnett Shale showing the relevant section (Marble Falls - Upper Barnett – Forestburg - Lower Barnett - Viola). The section is flattened based on Marble falls. Wireline log data includes gamma ray, P-sonic, bulk density. Facies were estimated based on each log set. Gray color represents limestone and brown is shale.

Attribute – attribute correlation analysis

Correlation measures	Attributes highly correlated with the other attributes (corr. coeff. > 0.6)
Mutual information	Lambda - LambdaRho (0.96) MuRho - Mu (0.93) Is - Mu (0.93) Young - Mu (0.89) Vs - Mu (0.88) Vs – Young's modulus (0.87) MuRho - Young's modulus (0.83) Is - Young's modulus (0.83) Ip - Young's modulus (0.79) Ip - Vp (0.79) Is - Vs (0.75) MuRho - Vs (0.75) Vp - LambdaRho (0.74) Vp - Lambda (0.72) Vp/Vs - Lambda (0.69) Poisson's ratio - Lambda (0.69) Vp - Young's modulus (0.65) Ip - Mu (0.64) Ip - Vs (0.63) Ip - LambdaRho (0.63) Ip - Is (0.62) Ip - MuRho (0.62)

Table 2.4. Attribute-to-attribute attribute correlation analysis using MI. In the Barnett Shale survey, the attributes are physical properties calculated from prestack seismic inversion. Attribute pairs exhibiting high correlation (correlation coefficient > 0.6) are ranked in descending order.

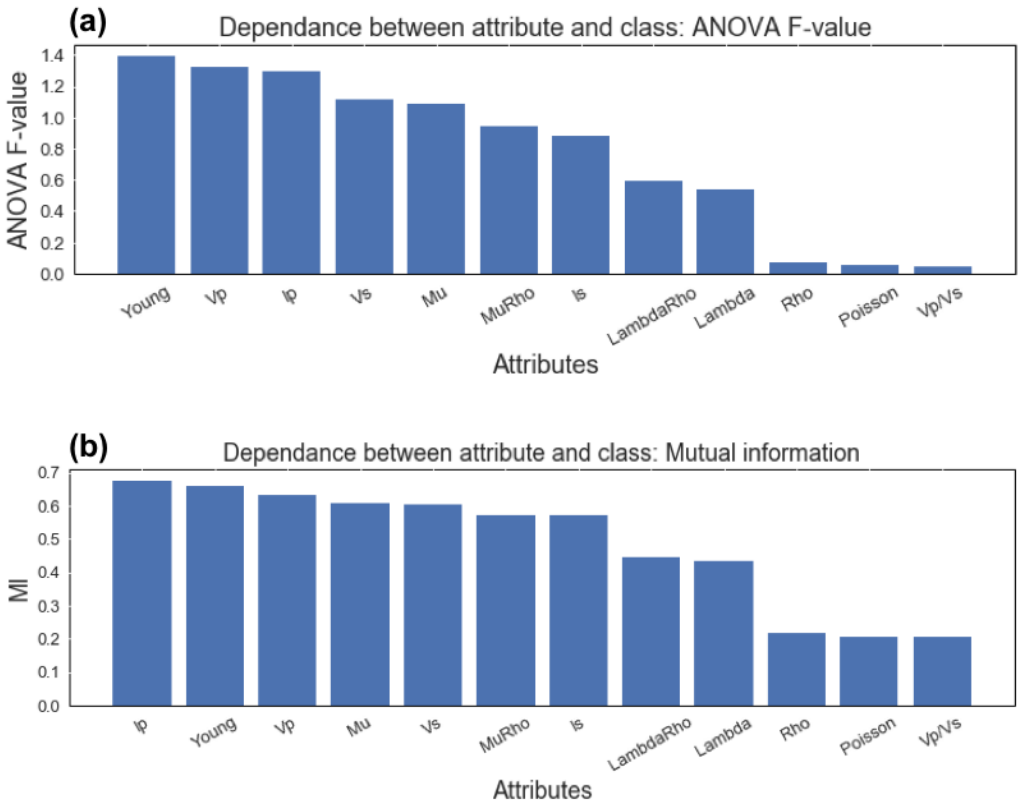


Figure 2.14. Relationships between a single input attribute and the desired output classes using (a) analysis of variance (ANOVA) F-value, and (b) mutual information.

Attribute selection algorithms			Ranking of attributes (Ten highest ranked attributes)				
Filter	ReliefF		LambdaRho	Lambda	Mu	Poisson	Young
			Rho	Vp/Vs	Vs	Vp	MuRho
	FCBF		Vs	Mu	Is	Vp	Ip
			MuRho	Young	Lambda	LambdaRho	Rho
Wrapper	SFS	NN	Young	Ip	Rho	Vp/Vs	LambdaRho
			MuRho	Vs	Vp	Poisson	Is
		RF	Young	Ip	Rho	Lambda	LambdaRho
			Is	Vp/Vs	Vp	Poisson	Vs
		SVM	Young	Ip	Mu	Vs	Is
			Vp/Vs	Rho	MuRho	Poisson	LambdaRho
	SBS	NN	Is	Poisson	Vs	LambdaRho	MuRho
			Vp/Vs	Vp	Young	Lambda	Mu
		RF	Vp	Is	Rho	Lambda	Vs
			Ip	MuRho	LambdaRho	Mu	Young
		SVM	Ip	Is	Vs	MuRho	Poisson
			Vp	Young	Vp/Vs	Lambda	Mu
Embedded	RF		Young	Ip	Mu	Vs	Vp
			MuRho	Is	LambdaRho	Lambda	Rho

Table 2.5. Selected attribute subsets using filter (RelifF, FCBF), wrapper (NN, RF, SVM), and embedded (RF) methods. Each subset includes the ten best attributes ranked in descending order (from left to right).

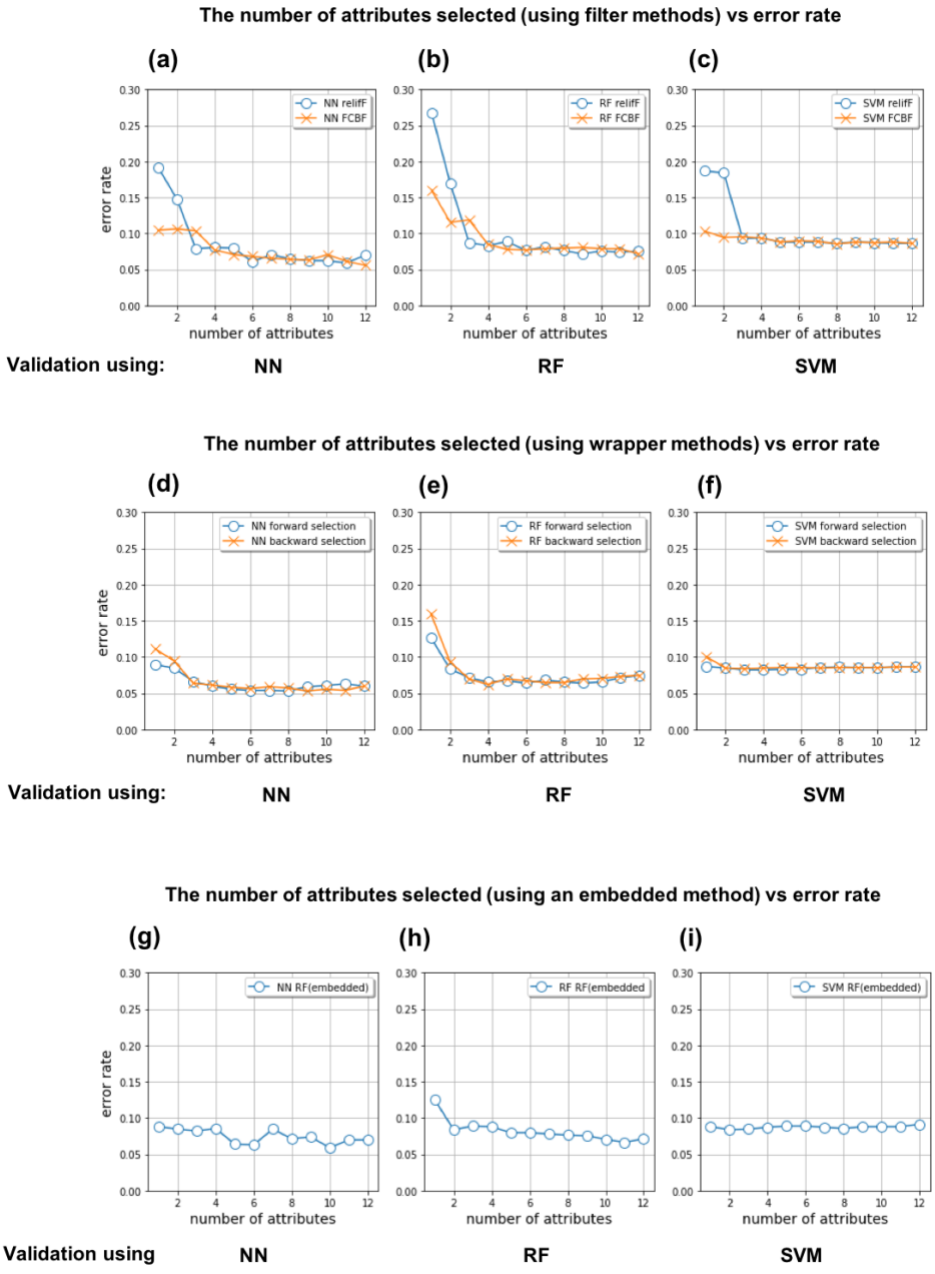


Figure 2.15. The number of attributes included in the attribute subset vs error rate. For (a), (b) and (c) Attributes in the subset were selected using filter methods (RelifF, FCBF), For (d) (e) and (f) attributes included in the subset were selected using wrapper methods (SFS, SBS). For (g), (h), and (i) attributes included in the subset were selected using an embedded method (RF). Each attribute subset was validated using NN, RF and SVM classifiers.

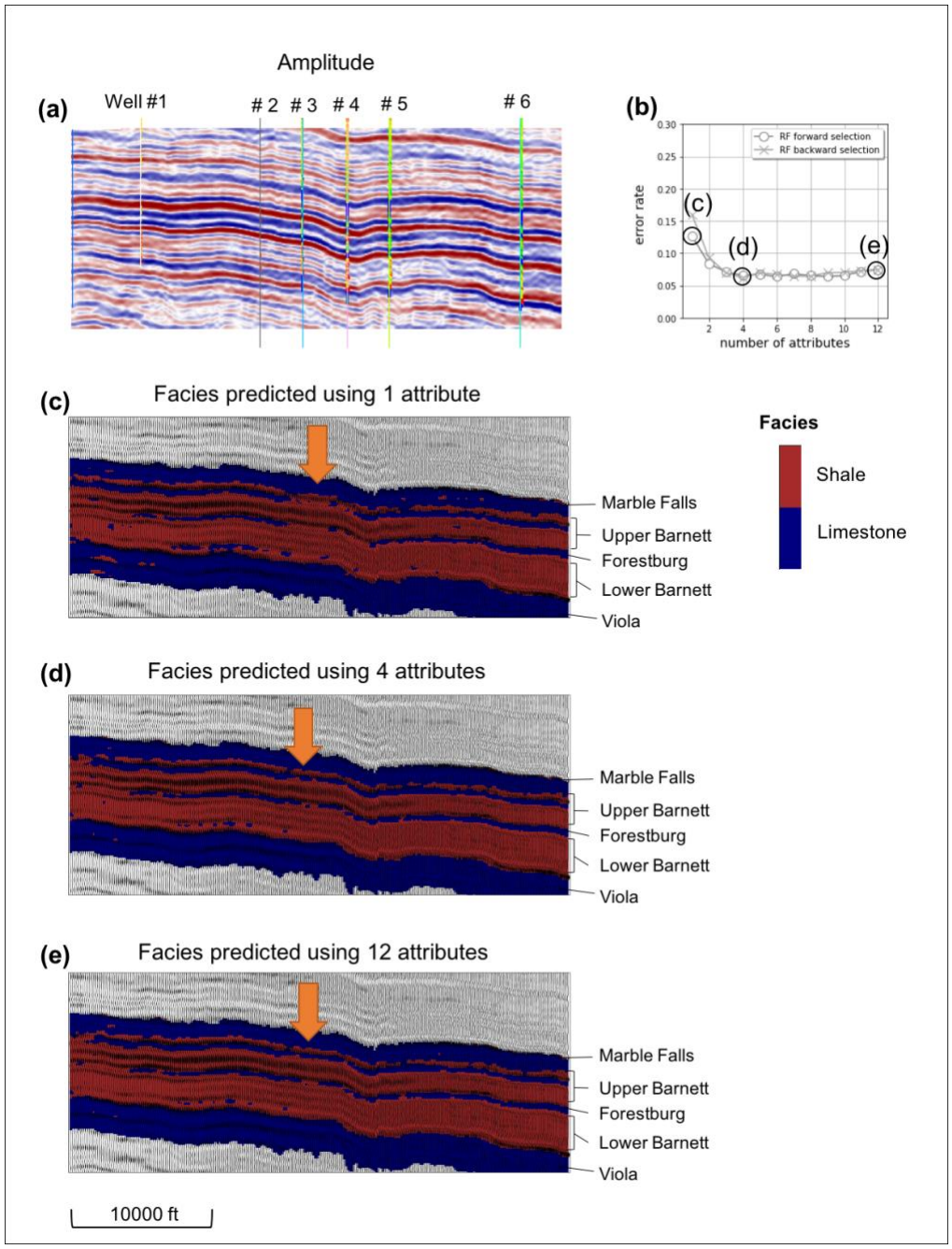


Figure 2.16. (a) A representative vertical slice through amplitude and well logs, (b) error rate with respect to the number of attributes in each subset which is selected by the wrapper method (RF) using training data. Facies predicted using (c) first highest-ranked attribute, (d) top four highest-ranked attributes selected by the wrapper method (RF), and (e) all 12 attributes. Subsets with the top four highest-ranked attributes differentiate the thin limestone layers as effectively as all 12 attributes (orange arrows).

CHAPTER 3: SEMI-SUPERVISED SEISMIC FACIES CLASSIFICATION USING CONDITIONAL GENERATIVE ADVERSARIAL NETWORK WITH MULTIPLE ATTRIBUTES AS INPUT

Introduction

Beginning with the early work of Poupon et al. (1999), Strecker and Uden (2002), and Coléou et al. (2003) almost two decades ago, automated machine learning for seismic facies analysis has been successfully applied to stratigraphic plays). Since that time advances of machine learning techniques and computing power have accelerated the development of seismic classification and interpretation algorithms. By mapping higher dimensional data to a lower dimensional 1D or 2D manifold, self-organizing maps (SOM), an unsupervised learning algorithm, generates clusters that can be color-coded by their similarity (Kohonen, 1982). SOM has been widely used for facies classification and has been applied to a target horizon or 3D volume with multiple seismic attributes at each voxel as input (Roy et al., 2013; Roden et al., 2015). The advantage of unsupervised learning is that the algorithm attempts to discover the underlying structure of the input data and does not suffer from the potential bias of explicit interpreter labeling. However, because SOM classes are constructed in a mathematical versus geological space, such that correlating a resulting SOM class to a specific geologic facies may be difficult (Wrona et al., 2018).

In contrast, supervised learning maps a relationship between input attributes and interpreter-defined output classes. Neural networks have had wide applications to mapping reservoir properties or facies from training data that are calibrated with well logs (Meldahl et al., 2001; West et al., 2002). Rather than using the seismic amplitude data volume itself,

depositional patterns can be classified using multiple attributes that better differentiate the facies of interest as the input to supervised learning. Wrona et al. (2018) tested multiple attributes as input, as well as different classification algorithms to distinguish the facies in terms of their continuity and structural orientation. They found that a neural network with deep convolutional layers (CNN) enabled them to build a model that captured complex geologic features and patterns such as horizons, faults, salt bodies, and channels on 2D or 3D data, where their input consisted of interpreter-defined patches and masks (labels) of the target geologic features. The resulting facies are predicted using a semantic segmentation algorithm. Wang et al. (2018) and Shi et al. (2019) adopted a CNN U-net architecture for segmentation of a salt body. Dramsch and Lüthje (2018) found that deep CNN such as VGG16 and ResNet50 provide good results in extracting the seismic textural response. Even if deep CNN can capture complex patterns, CNN requires a large number of labeled samples for training. The number of unlabeled data, however, is often larger than that of labeled data, since constructing a segmentation mask requires considerable time.

To maximize the use of unlabeled as well as of labeled data, we adopt a semi-supervised approach to CNN using a generative adversarial network (GAN) (Odena, 2016). GAN consists of a generator that creates *fake* samples and a discriminator that distinguishes *true data* samples from the *fake* samples. Salimans et al. (2016) reports that the GAN framework improves semi-supervised learning performance and reduces the need for an excessive number of labels. Semantic segmentation of images has been improved using GAN, especially when labeled training samples are insufficient (Souly et al., 2017; Hung et al., 2018). In seismic interpretation, Liu et al. (2019) applied GAN to classify channel,

levee, and floodplain facies with training data calibrated from a limited number of well logs.

We begin with a review of multiattribute facies classification, followed by a summary of key details of the cGAN network. We then apply our cGAN facies classification algorithm to a 3D marine seismic data volume acquired in the Gulf of Mexico (GOM) to differentiate salt, mass transport deposits (MTDs) and conformal reflector facies (Figure 3.1). Next, we compare our classification with other segmentation methods such as the U-net. We conclude with a summary of the benefits and limitations of this classification workflow and of machine learning based classification in general.

Facies classification with multiple attributes as input

Seismic attribute extract and quantify spatial and temporal patterns in the seismic data, where a "useful" seismic attribute for a given project extracts patterns similar to those found to be meaningful by interactive human interpreters. Depending on whether the objective is to define stratigraphy, seismic facies, or the presence of hydrocarbons, different combinations of attributes are used as input to learning tools such as the self-organizing map (SOM) (Roden et al., 2015). For instance, texture attributes such as gray-level co-occurrence matrix (GLCM) entropy and homogeneity can help differentiate salt diapirs (Berthelot et al., 2013; Qi et al., 2016) from the more continuous sedimentary basin fill. Mud fill in channel systems can be highlighted with frequency, magnitude, shape index, and coherent energy (Zhao et al., 2016). Combining several families of attributes helps when there are multiple target classes. Kim et al. (2019) found that the simultaneous use

of the amplitude family of attributes (e.g. RMS amplitude and total energy) and texture family of attributes (chaos, GLCM, coherence) efficiently differentiates salt and MTDs from conformal reflectors.

The four attributes shown in Figure 3.2 (amplitude, chaos, aberrancy total magnitude, and instantaneous frequency) partially differentiate salt and MTD facies from the conformal sedimentary background. Chaos is a measure of how well a reflector can be represented by a constant amplitude plane. Aberrancy, also known as flexure, measures the lateral changes in the curvature of a seismic reflector, thereby making it a good indicator of the intensity of faults and fractures, localized deformation (Gao, 2013; Qi and Marfurt 2018), and in our case rapid variations in dip associated with seismic noise internal to salt domes. For an isolated event, the instantaneous frequency (Taner et al., 1979) is the mean frequency of the wavelet. In our implementation, we follow Taner et al. (1979) and weight the frequency by the envelope in a small window to improve stability. Halpert and Clapp (2008) observed a rapid change in instantaneous frequency at salt boundaries. Amplitude provides detailed features of depositional patterns in salt, MTDs and conformal reflectors. Since salt and MTDs exhibit both discontinuous and chaotic textures, the chaos attribute helps to differentiate salt and MTDs from the more continuous conformal sedimentary reflectors that fill the basin. Aberrancy and instantaneous frequency differentiate salt diapirs from the other two facies.

In principle, a highly trained CNN can predict features of interest from the seismic amplitude data alone. In practice, constructing the 3D labeled data to train the CNN to work only on amplitude data can be extremely time consuming. Because attributes such as chaos

and aberrancy are inherently 3D, we hypothesize that the far easier task of creating 2D labeled data from a suite of vertical slices might by sufficient. In terms of classification and segmentation, multiple feature learning can process several types of features simultaneously (Gao et al., 2018). In this work, we examine whether using multiple attributes together can improve the classification task when compared using a single attribute, amplitude, as input to the deep neural network.

Methodology

In the deep learning community, the term "sample" represents an image, rather than the value at a point in time on a seismic trace commonly used by the geophysics processing community. In the examples that follow later in this paper our samples will be 2D patches of the seismic amplitude and seismic attribute data that measure 64 by 64 voxels in size. For clarity to our geoscience readers, we will denote such an array of values with a bold **x**. Generative adversarial network (GAN) consists of a generator and a discriminator. The generator generates "fake" samples and attempts to fool the discriminator into thinking they are the input data samples, **x**. The role of the discriminator is to distinguish true data samples from fake samples. For the seismic facies classification task, we implemented a conditional generative adversarial network (cGAN) based on work by Mirza and Osindero (2014) whereby the generator generates samples with specific conditions, instead of generating samples from an unknown noise distribution. In the cGAN algorithms, the generator G tries to minimize the objective function against a discriminator, D, that tries to maximize it. The resulting binary cross entropy log loss function L is then:

$$\min_{G}\max_{D} L_{cGAN}(D,G) = E_{x,y}[logD(\mathbf{x},\mathbf{y})] + E_{x,z}[\log\,(1 - D(\mathbf{x},G(\mathbf{x},\mathbf{z})))], \quad (1)$$

where

E represents the expectation,

x is an observed input data sample,

y is an output data sample, and

z is an input noise sample.

The generator, *G*, outputs a fake sample, *G(***x***,* **z***),* from the input data sample, **x***,* and a random noise sample **z**. The discriminator, *D*, is used to distinguish the image generated by *G* from the corresponding ground truth, **y**. Referring to the flow diagram in Figure 3.3a, the generator generates a fake facies labels, *G(***x***,* **z***)*, from the seismic attribute input and noise. Instead of using Gaussian noise, we use noise in the form of dropout, which is applied on several layers of the generator (Isola et al., 2017). The generator shares weights with the segmentation network to speed up the training process. The discriminator distinguishes fake samples from ground truth using two kinds of inputs: 1) input attribute samples, **x**, and desired output (labeled) samples, **y**, which the discriminator needs to find to be true, and 2) input attribute samples, **x**, and generated fake facies labels, $\mathbf{y}_{\text{fake}}=G(\mathbf{x},\mathbf{z})$, which the discriminator needs to find to be false. If a well-trained discriminator can't tell the difference between a human and a machine interpretation, then we believe that we have a good generator (interpretation) system. As the cGAN network is trained, the generator output gradually changes as shown in Figure 3.3b. The machine interpretation becomes difficult to distinguish from the human interpretation as the cGAN model is well-fitted.

We will use a U-Net architecture for the generator and segmentation networks. An advantage of the U-net is that a great amount of feature detail is shared between the input and output, such that the process of training is accelerated. Detailed specification for the generator and discriminator can be found in Isola et al. (2017) (Appendix A).

Training and validation process

We applied the semi-supervised learning approach to a 3D seismic marine survey acquired on the northern Gulf of Mexico shelf. Examining Figure 3.1, note that several salt diapirs as well as mass transport deposits (MTDs) are seen in the seismic amplitude volume. While MTDs are common on the shelf edge, in this data volume away from the shelf edge, most of the gravity-controlled processes of the MTDs are induced by salt diapirism and salt withdrawal forming minibasins and rotating the seafloor. In contrast to the relatively conformal sediments that fill the bulk of the basin, both salt and MTDs exhibit a more chaotic behavior. With the exception of salt welds and an occasional block of sediments incorporated in the diapir, the interior of the salt should be relatively reflector free. In older, narrow azimuth, prestack time-migrated data like that shown in Figure 3.1, the interior of the salt is filled with potentially moderate amplitude coherent multiples and improperly imaged converted waves, as well as lower amplitude operator aliasing artifacts and other noise that appears random. These latter events appear chaotic with inconsistent waveforms, lower frequency content, and random dips. Careful examination of the MTDs reveal a different seismic expression. Towards the toe of the MTD the data may be chaotic, with random, but fairly gentle dips, lower frequencies, and moderate amplitude. Updip, MTDs are characterized by rotated fault blocks. Internal to each fault block, the amplitude,

frequency spectrum, and continuity of the MTD looks much like that of the surrounding conformal sedimentary matrix. However, there are rapid breaks in continuity and dip between adjacent fault blocks and between the top and base of the MTD from the conformal sediments in which it is incased. The goal of our cGAN is therefore to differentiate between these two "chaotic" seismic facies as well as from the conformal sedimentary matrix.

Three seismic attributes (chaos, total magnitude of aberrancy, and instantaneous frequency) are computed from seismic amplitude. For training and validation input, 75 vertical slices of 512 crosslines by 384 time samples are cropped out of each volume of the 3 attribute and the seismic amplitude volumes. The input data are normalized to range between -1 to +1 using a linear transformation. The training and validation output are labeled by a skilled interpreter drawing polygons around the features of interest. The training input and output are fed as patches with a size of 128 × 128 voxels. A total of 900 patches are generated from the 75 vertical slices. Each voxel in a given sample is labeled as 0 for conformal sediments, 1 for salt, and 2 for MTDs. We then randomly selected 70% of the samples for training, using the remaining 30% samples for validation.

We run two main tests to verify 1) the efficiency of the cGAN algorithm as a semi-supervised learning technique, and 2) the advantage of using multiple attributes as input data over a single input (amplitude only). To test whether the semi-supervised cGAN architecture is efficient, the segmentation results are compared with those of a U-net model based on supervised learning. For the U-net model 5, 30, and 100 labeled samples are used as the training data in each test. The cGAN model uses 5, 30, and 100 labeled samples and 700 samples were fed into generator as unlabeled training data. The output of the generator

was, then fed into the discriminator. The detailed training process of cGAN is described Appendix B.

To test how the multiple attribute input affects the classification performance, two cases of input (amplitude and three attributes versus amplitude only) are fed into the cGAN and U-net models. In order to input four attributes into the deep-learning architecture, each attribute is treated as one channel of a 4-channel input stream.

Results and Discussion

We use the accuracy of predicting the three seismic facies correctly as an evaluation metric. Specifically, we define the accuracy to be the percentage of samples that have all their labels classified correctly. The classification results of the proposed semi-supervised cGAN model show higher validation accuracies than those of the U-net, especially when the number of labeled training samples are smaller (5, 30, and, 100) than the entire 900 training samples available (Figure 3.4 and Table 3.1). Also, the U-net results have a higher variance of accuracies, which indicates that the performance of the model is susceptible to the specific set of labeled training samples. When the entire training data are trained as labeled samples, the U-net model has slightly higher validation accuracy. The results imply cGAN enhances segmentation with unlabeled samples based on the semi-supervised learning approach, avoiding overfitting of training data. Comparing the classification performance of the single attribute input and multiple attribute input, the multiple attribute input enhances validation accuracies in most cases, especially when not all of the training samples are labeled.

Using the proposed cGAN model with the multiple attribute input, we made a prediction for the cropped 3D seismic volume (750 lines × 512 crosslines × 384 time samples). 75 vertical slices out of 750 lines were labeled training data and fed into discriminator with their labels. 750 lines were fed into the generator as unlabeled training data, then fed into the discriminator with generated samples (Figure 3.5-3.8). A 3D median filter with a size of 15 × 15 × 15 voxels is applied to the predicted volume to smooth each facies. The cropped volume used for testing consists of two salt diapirs as well as several MTDs surrounding the salts (Figure 3.5). The segmented 3D volume effectively shows the depositional structure of MTDs around two salt diapirs. As anticipated, we are able to see that the MTDs are sliding away from the salt dome into the evacuated mini basins.

Limitations

The training and application in this paper were based on 2D samples. Clearly, picking seismic facies on a suite of seismic lines builds on standard interpretation workflows. Although salt and MTDs are 3D objects, the chaotic nature of salt and the toes of MTDs, and for the piecewise chaotic nature of the rotated fault blocks of the MTDs provides a similar appearance on inlines, crosslines, and arbitrary lines. Such is not the case for faults where deep learning works better on 3D samples (Qi et al., 2020). Faults appear as a fault stick on a vertical slice perpendicular to the fault surface and as a (hard to pick) U-shaped feature on a vertical slice parallel to the fault surface. We do not expect our 2D samples to work as well on fluvial systems, where the appearance of a channel or of a fan is quite different when seen parallel or perpendicular to the depositional axis.

The use of seismic attributes partially ameliorates these shortcomings. Although the spectral response is computed trace by trace, chaos and aberrancy use 3D windows. In this work, chaos was computed in a 3×3 trace by 20 ms analysis window while aberrancy was computed in a 10×10 trace by 100 ms analysis window, thereby capturing more of the 3D geologic seismic response.

Limitations

In this work, our sample size was fixed to be 2D patches of 128 × 128 voxels. Therefore, the facies are classified based on local seismic and attribute patterns and do not "see" patterns that range beyond this limit. As with any machine learning technique, the data quality needs to be consistent within the target where facies are to be predicted. If the training data were limited to the shallower, easier to label, section of the seismic survey, the algorithm has no knowledge of the loss of frequency and lateral resolution with depth. Likewise, if a part of the survey contaminated by noise is not used in the training, the quality of the prediction is uncertain. Unlabeled geologic features such as canyons, channels, and erosional unconformities may be lumped in with one of the three target facies, or alternatively, have a low probability of being any of the facies. The method is based on deep learning with multiple convolutional layers that can capture complicated features and seismic patterns. However, the pitfall of machine assisted interpretation is that the algorithm makes decisions without an understanding of geologic processes and the limits of seismic acquisition, processing, and imaging, routinely used by skilled human interpreters. Specifically, for this case, a human interpreter would not accept a mass

transport complex internal to a salt dome, or a salt block inclusion within a mass transport complex.

Conclusions

We present a workflow of semi-supervised seismic facies classification with the goal of obtaining an accurate segmentation particularly with a minimum amount of labels. The cGAN architecture of deep learning enhances the accuracy of classification of seismic facies and avoids overfitting to training data. By sharing weights with the generator network, the segmentation network speeds up the reducing loss of the generator, which results in speeding up of the overall training process as well. Use of multiple attributes, which contain 3D characteristics of each facies (texture, geometry, frequency components), improves the classification performance especially when the training data is only partially labeled. The predicted 3D classification volume helps us to understand the geometry and depositional structure of MTDs and salts.

Acknowledgements

The authors thank the sponsors of OU Attribute Assisted Seismic Processing and Interpretation (AASPI) consortium for their encouragement and financial support.

Appendix A

Network architectures of generator and discriminator

The generator, *G*, consists of encoder and decoder blocks. Each encoder block includes a 2D convolutional (in TensorFlow, *conv2D*) layer, a batch normalization to recenter and rescale the data, and a leaky rectified linear unit (*ReLU*) activation function, where

$$. \qquad ReLU(u) = \begin{cases} 0 \ \ if \ u < 0 \\ u \ if \ \ u \geq 0 \end{cases} \qquad . \tag{A1}$$

The decoder blocks include a transposed 2D convolution layer for upsampling, followed by a *LeakyReLU*, and batch normalization. The generator output is connected to a 'tanh' activation function, Finally, the model uses binary cross-entropy as the loss function.

Each block in the discriminator consists of a *conv2D* layer, *BatchNormalization*, and *LeakyReLU*. A sigmoidal activation function is connected at the end of the last block after which we use a binary cross-entropy log loss function.

Appendix B

The training of labeled and unlabeled samples

We test cGAN model using 5, 30, and 100 labeled samples and 700 samples as unlabeled data. Figure 3.3a describes the overall procedure of training. For instance, five samples of input attributes are fed into the segmentation network with desired output **y** which is label of input attributes. The weights in the network are adjusted to reduce the loss function of supervised learning. The weights are then shared with the generator. The unlabeled input attributes $\mathbf{x}_{unsup}$ are fed into the generator after which the generator constructs a fake sample G($\mathbf{x}_{unsup}$, **z**) which has the same form as a labeled facies map. Next, the discriminator differentiates the fake labeled samples from the desired output

using two kinds of input 1) true labeled input attribute samples, $\mathbf{x}_{\mathbf{sup}}$, and desired output labeled data **y**, which the discriminator needs to find to be true, and 2) unlabeled input attribute samples, $\mathbf{x}_{\text{unsup}}$, and generated fake labeled samples, $\mathbf{y}_{\text{fake}}=G(\mathbf{x}_{\mathbf{unsup}}, \mathbf{z})$, which the discriminator needs to find to be false.

References

Coléou, T., M., Poupon and K. Azbel, 2003, Unsupervised seismic facies classification: A review and comparison of techniques and implementation: The Leading Edge, **22**, 942-953.

Dramsch, J. S., and M. Lüthje, 2018, Deep-learning seismic facies on state-of-the-art CNN architectures: 88th Annual International Meeting, SEG Expanded Abstracts, 2036-2040.

Gao, D., 2013, Integrating 3D seismic curvature and curvature gradient attributes for fracture characterization: Methodologies and interpretational implications: Geophysics, **78**, O21-O31.

Halpert, A., and R. G. Clapp, 2008, Salt body segmentation with dip and frequency attributes: SEP-Report, **136**, 113-124.

Hung, W. C., Y. H. Tsai, Y. T. Liou, Y. Y. Lin, and M. H. Yang, 2018, Adversarial learning for semi-supervised semantic segmentation: arXiv preprint arXiv:1802.07934.

Isola, P., J. Y. Zhu, T. Zhou, and A. A. Efros, 2017, Image-to-image translation with conditional adversarial networks: Proceedings of the IEEE conference on computer vision and pattern recognition, 1125-1134.

Kim, Y., R. Hardisty, and K. J. Marfurt, 2019, Attribute selection in seismic facies classification: Application to a Gulf of Mexico 3D seismic survey and the Barnett Shale: Interpretation, **7**, SE281-SE297.

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

Liu, M., W. Li, M. Jervis, and P. Nivlet, 2019, 3D seismic facies classification using convolutional neural network and semi-supervised generative adversarial network: 89th Annual International Meeting, SEG Expanded Abstracts, 4995-4999.

Meldahl, P., R., Heggland, B. Bril, and P. de Groot, 2001, Identifying faults and gas chimneys using multiattributes and neural networks. The leading edge, **20**, 474-482.

Mirza, M., and S. Osindero, 2014, Conditional generative adversarial nets: arXiv preprint arXiv:1411.1784.

Mondal, A. K., J. Dolz, and C. Desrosiers, 2018, Few-shot 3d multi-modal medical image segmentation using generative adversarial learning: arXiv preprint arXiv:1810.12241.

Odena, A., 2016, Semi-supervised learning with generative adversarial networks. *arXiv preprint arXiv:1606.01583*.

Qi, J., B. Zhang, B. Lyu, and K. Marfurt, 2020, Seismic attribute selection for machine-learning-based facies analysis. Geophysics, **85**, O17-O35.

Qi, X., and K. J. Marfurt, 2018, Volumetric aberrancy to map subtle faults and flexures: Interpretation, **6**, T349-T365.

Roden, R., T., Smith, and D. Sacrey, 2015, Geologic pattern recognition from seismic attributes: Principal component analysis and self-organizing maps: Interpretation, **3**, SAE59-SAE83.

Roy, A., 2013, Latent space classification of seismic facies. University of Oklahoma.

Salimans, T., I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen, 2016, Improved techniques for training gans: In Advances in neural information processing systems, 2234-2242.

Souly, N., C. Spampinato, and M. Shah, 2017, Semi supervised semantic segmentation using generative adversarial network: Proceedings of the IEEE International Conference on Computer Vision, 5688-5696.

Shi, Y., X. Wu, and Fomel, S., 2019, SaltSeg: Automatic 3D salt segmentation using a deep convolutional neural network: Interpretation, **7**, no. 3, SE113-SE122.

Taner, M. T., F. Koehler, and R. E. Sheriff, 1979, Complex seismic trace analysis: Geophysics: 44, 1041–1063.

Wang, Z., H. Di, M. A. Shafiq, Y. Alaudah, and G. AlRegib, 2018, Successful leveraging of image processing and machine learning in seismic structural interpretation: A review: The Leading Edge, **37**, 451-461.

West, B. P., S. R. May, J. E. Eastwood, and C. Rossen, 2002, Interactive seismic facies classification using textural attributes and neural networks. The Leading Edge, **21**, 1042-1049.

Wrona, T., I. Pan, R. L. Gawthorpe, and H. Fossen, 2018, Seismic facies analysis using machine learning: Geophysics, **83**, O83-O95.

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-SB89.

Figures

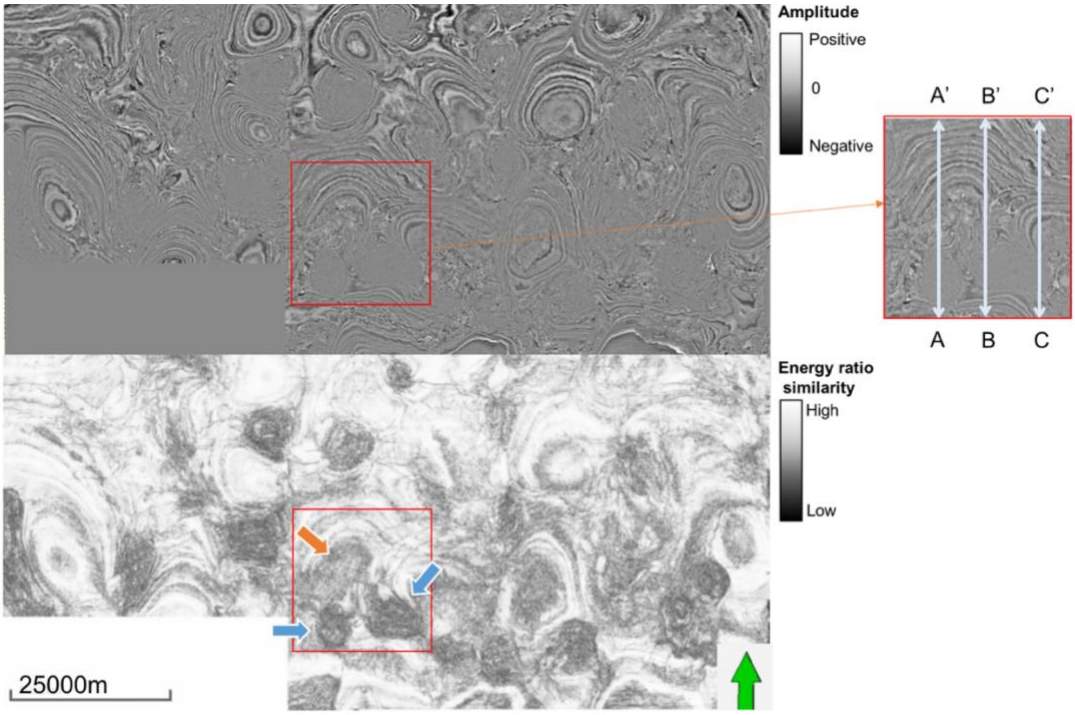


Figure 3.1. Time slices at t = 1.1 s through the (a) seismic amplitude and (b) energy ratio similarity volumes. The red box indicates the volume where the training data are sampled. The orange arrow in (b) indicate MTD and the cyan arrows indicate salt diapirs, both of which exhibit a low value of similarity. (c) Lines AA', BB', and CC' are used in training.

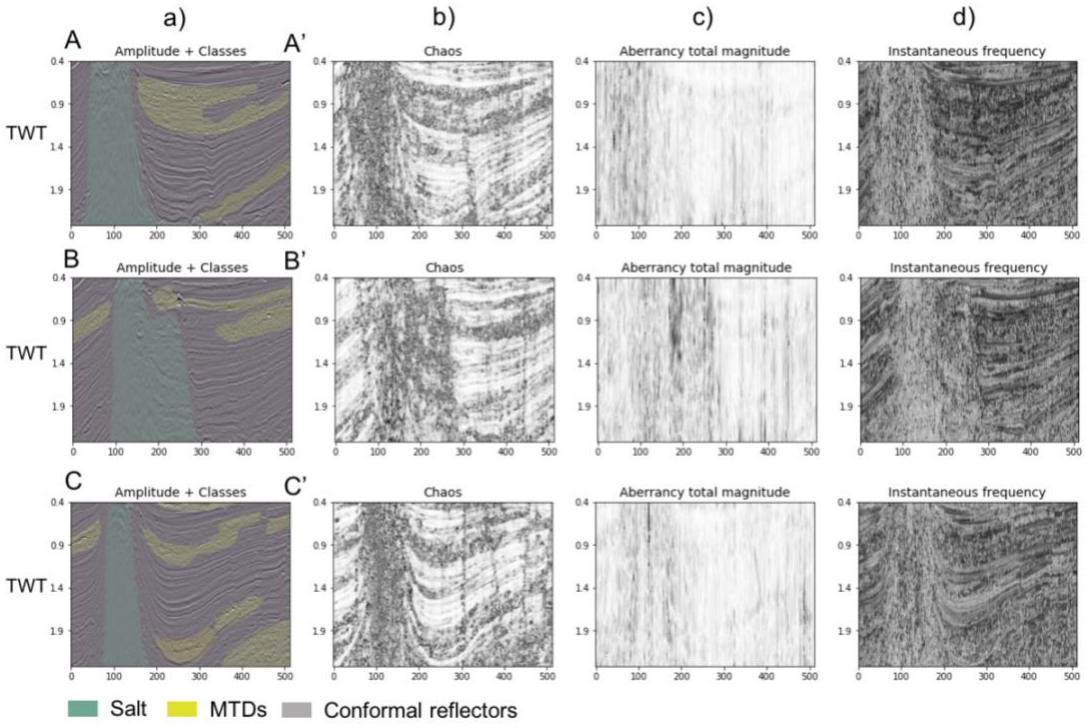


Figure 3.2. Vertical slices AA', BB', and CC' through (a) amplitude co-rendered with the three labeled seismic facies, (b) chaos, (b) total magnitude of aberrancy, and (d) instantaneous frequency. Attributes in (b), (c), and (d) differentiate the salt and MTDs from each other and from the background conformal seismic reflectors. The salt exhibits low amplitude, high chaos, high aberrancy, and relatively low frequency. The MTDs are highly chaotic.

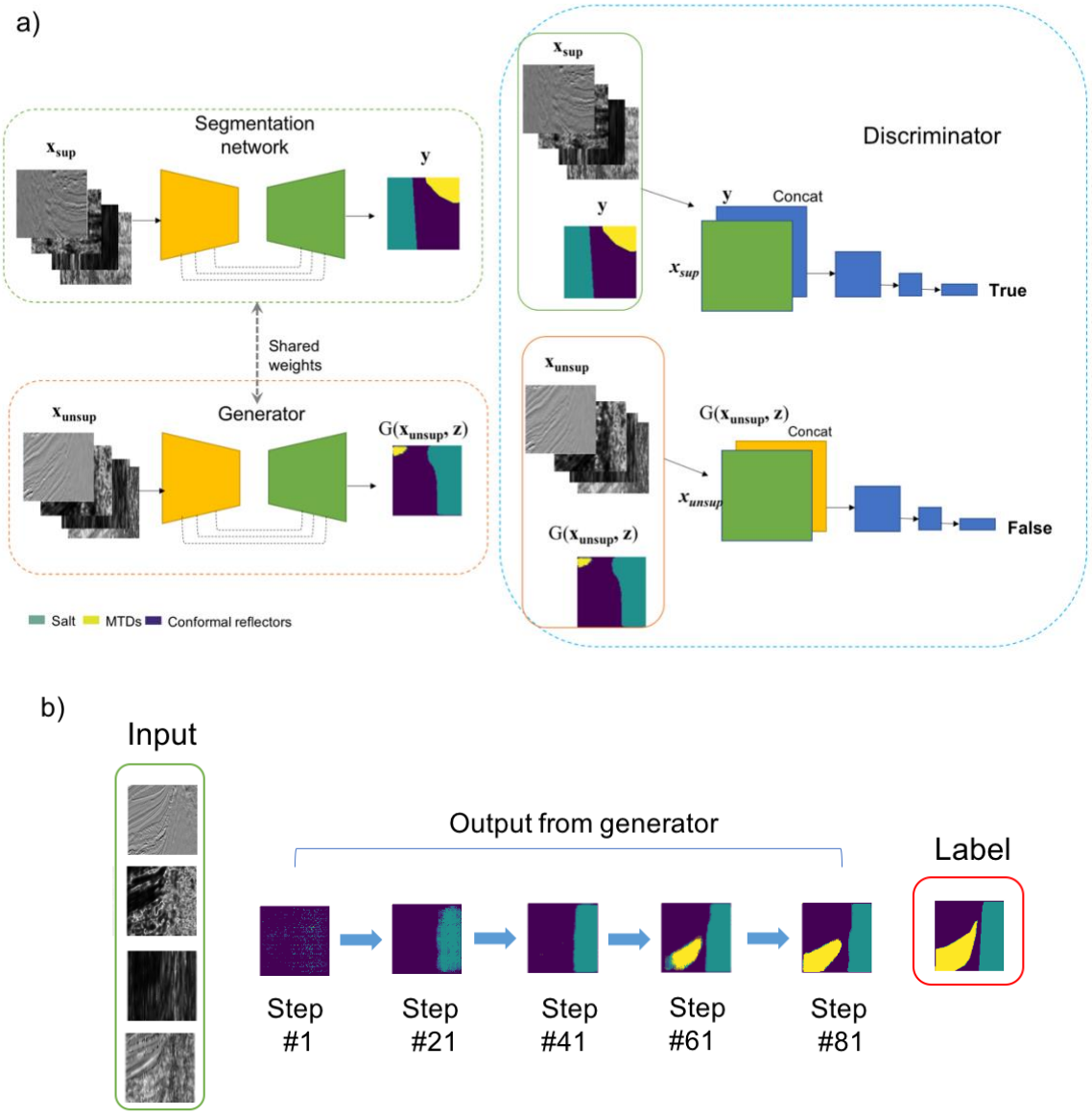


Figure 3.3. (a) Overview of conditional Generative Adversarial Network architecture. The model contains a segmentation network, a generator, and a discriminator. To begin, the generator constructs a fake facies map G(**x**$_{unsup}$, **z**) out of the unlabeled input attribute **x**$_{unsup}$ with noise. Next, the discriminator differentiates the ground truth **y** which is the label of input attributes and fake output G(**x**$_{unsup}$, **z**). If a well-trained discriminator can't tell the difference between a human and a machine interpretation, then we believe that we have a good generator (interpretation) system. The segmentation network which is trained with labeled input attribute **x**$_{sup}$ and ground truth **y**, shares weights with the generator (After Mirza and Osindero, 2014; Isola et al., 2017). (b) The gradual change of the generator

output as the training process develops. The machine interpretation becomes difficult to distinguish with human interpretation as the cGAN model is well-fitted.

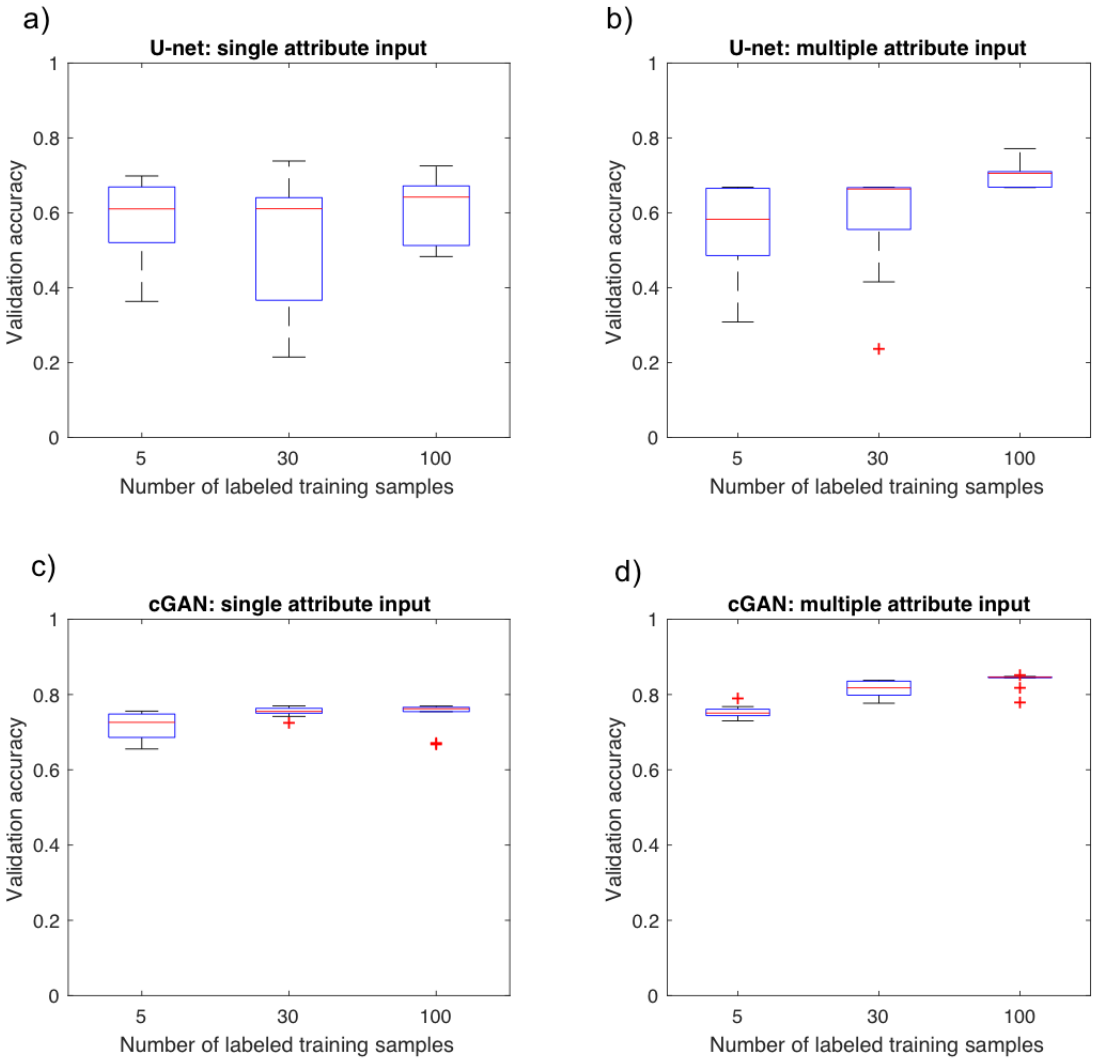


Figure 3.4. Comparison of validation accuracy versus the number of labeled training samples between different models and datasets that have iterated 10 times. (a) U-net model with single attribute input. (b) U-net model with multiple attribute input. (c) Proposed cGAN model with single attribute input. (d) Proposed cGAN model with multiple attribute input. Note the validation accuracy increases when we use multiple rather than a single attribute as input for both the U-net and cGAN algorithms. The median accuracy of the cGAN exceeds the median accuracy of the U-net for all six tests run.

Number of Labeled training samples	5 samples	30 samples	100 samples	Entire training set
U-net Single attribute	0.59	0.53	0.61	0.91
U-net Multi-attribute	0.56	0.58	0.71	0.91
cGAN Single attribute	0.71	0.75	0.74	0.89
cGAN Multi-attribute	0.75	0.81	0.84	0.90

Table 3.1. Comparison of validation accuracy versus the number of labeled training samples between different models and datasets that have been iterated 10 times and averaged. With one small exception, the accuracy increases with increasing the amount of training data.

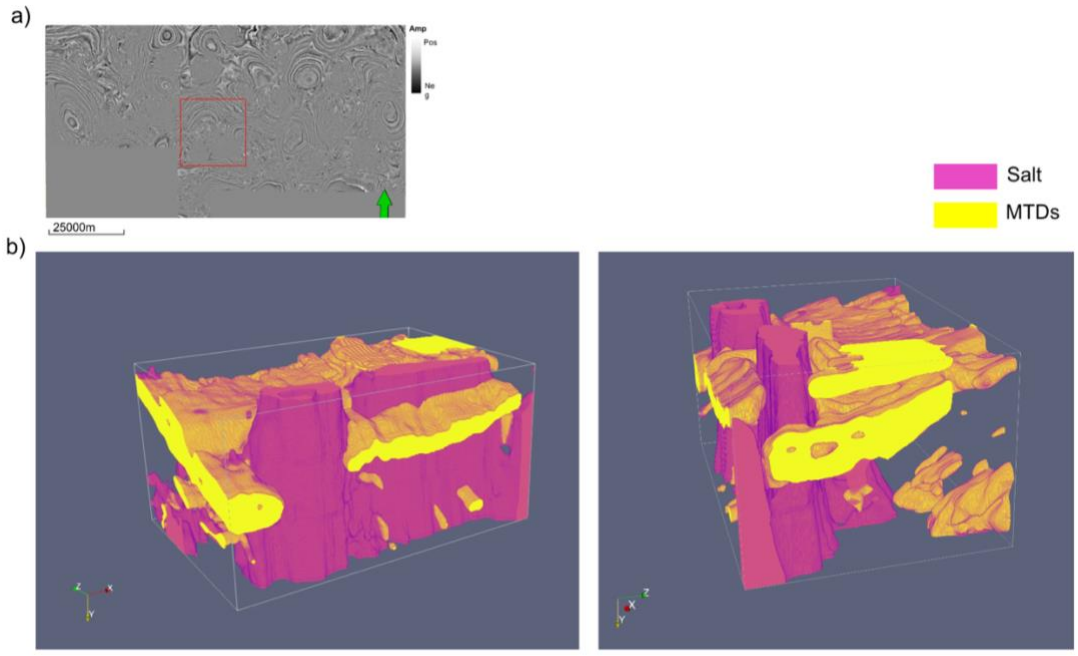


Figure 3.5. (a) Time slice at t = 1.1 s through the seismic amplitude volume. The red box indicates the 3D test volume where the facies prediction is made. (b) Predicted facies in the 3D volume. Using seismic amplitude and attributes as input, two salt diapirs and two MTDs are mapped in the area of interest. The predicted facies volume implies the hydrodynamic behavior and transport direction of the MTDs in the area is induced by the growth of the salt diapirs.

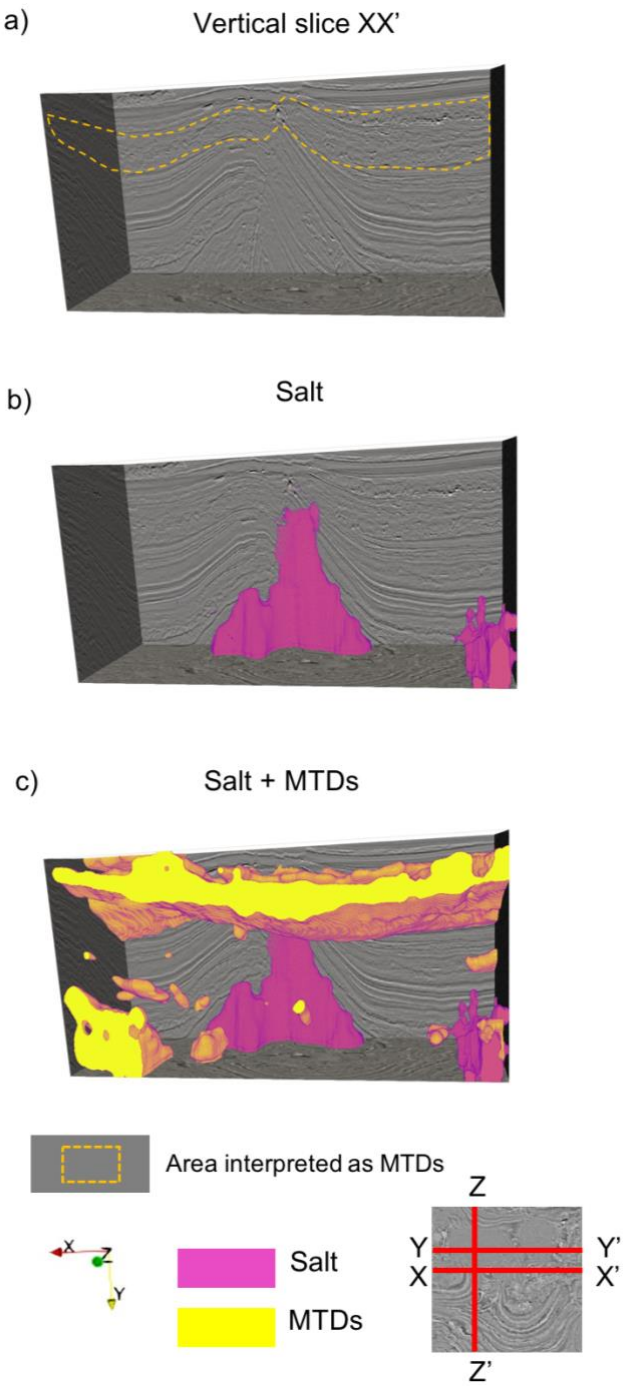


Figure 3.6. (a) EW vertical slice XX' through the seismic amplitude volume. The dashed line indicates MTDs not use in the training. 3D views of the (b) predicted salt facies and (c) the predicted salt facies and MTD facies.

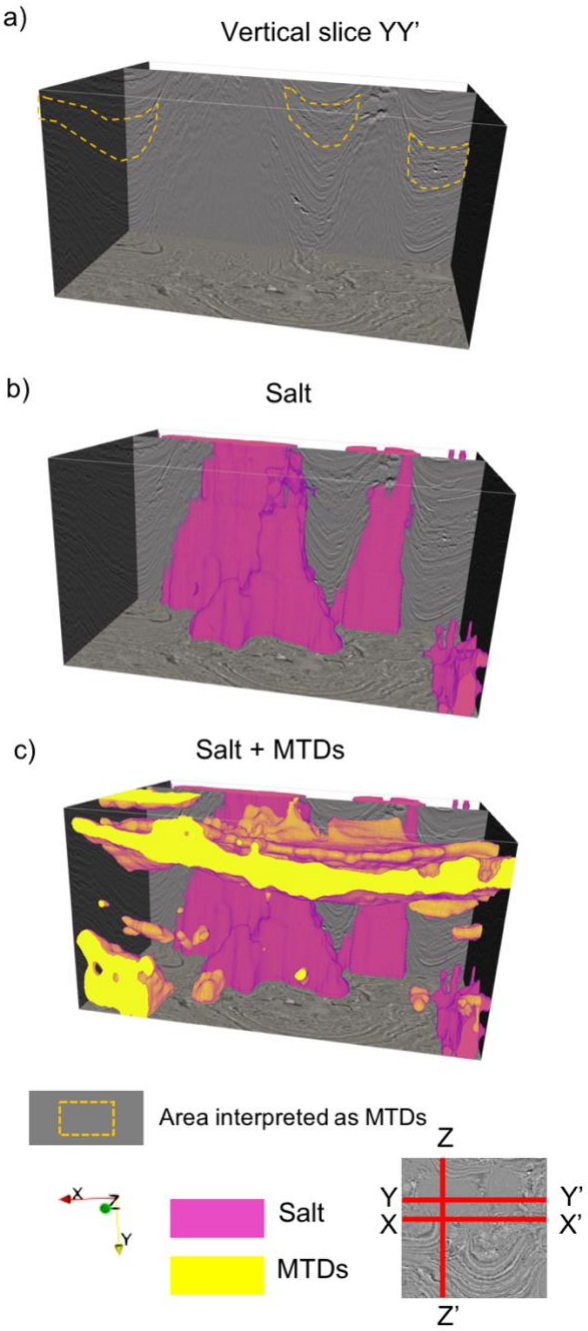


Figure 3.7. (a) EW vertical slice YY' through the seismic amplitude volume. The dashed line indicates MTDs not use in the training. 3D views of the (b) predicted salt facies and (c) the predicted salt facies and MTD facies.

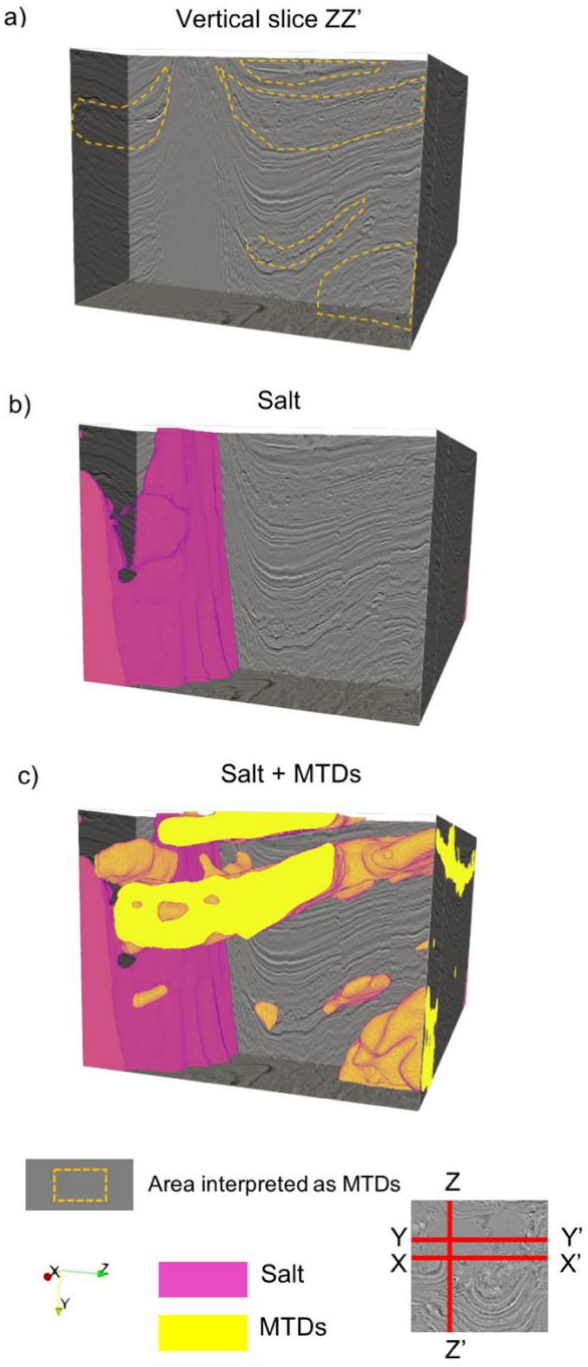


Figure 3.8. (a) NS vertical slice ZZ' through the seismic amplitude volume. The dashed line indicates MTDs not use in the training. 3D views of the (b) predicted salt facies and (c) the predicted salt facies and MTD facies.

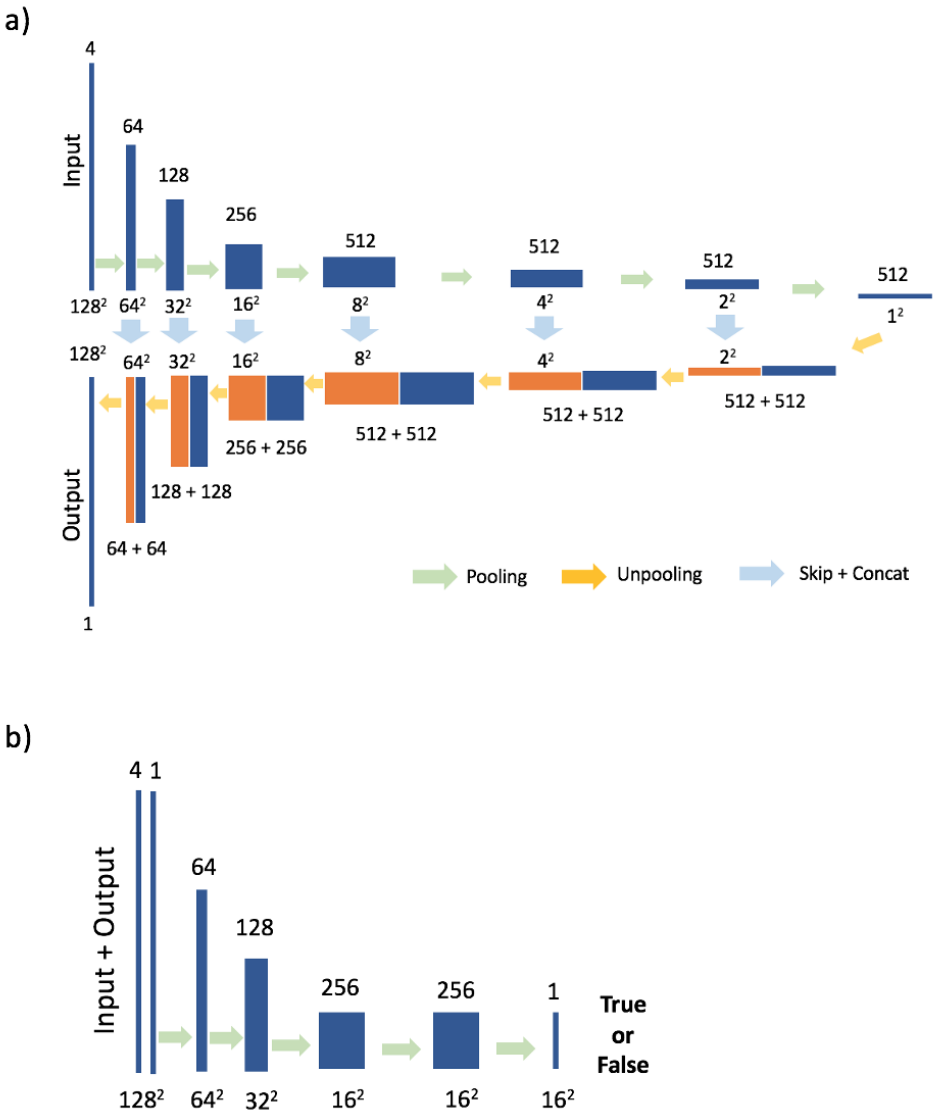


Figure 3.A-1. (a) Specification of generator. (b) Discriminator architecture. Each blue box is multi-channel feature map. The number of channels is indicated on top of box. The size of feature map is denoted below box. Orange boxes are copied feature maps.

CHAPTER 4: SEISMIC NOISE ATTENUATION IN THE F-X DOMAIN USING A COMPLEX-VALUED RESIDUAL CONVOLUTIONAL NEURAL NETWORK

Introduction

Due to limits in recording, processing, and imaging as well as to the overlying overburden, seismic data are inevitably contaminated with noise. For this reason, noise reduction that enhances the reflectors and diffractors of interest are a key part of signal analysis and critical to accurate seismic interpretation. In terms of random noise reduction in post-stack seismic data, there are multiple approaches that can be implemented including predictive filtering, wavelet-transforms and other decomposition techniques as well as rank-reduction (Zhou et al., 2017).

Rank-reduction based approaches, which can be applied in the *f-x* domain, retrieve low-rank data of coherent signals after removing the random noise which cause the data to be more highly-ranked. For example, Oropeza and Sacchi (2009) describe the singular spectrum analysis rank-reduction procedure, which requires first constructing a Hankel matrix in the *f-x* domain and then applying singular value decomposition (SVD) to reduce its rank. Noise reduction methods in the *f-x* domain have a number of advantages over those in the time-space domain. The most important advantage is the assumption of a stationary signal. Fourier-based denoising improves the signal-to-noise ratio by finding and then preserving the more dominant harmonics (Naghizadeh and Sacchi, 2010, 2012).

Recently, Jain and Seung (2009), Zhang et al. (2017) and others have shown that a convolutional neural networks (CNN), a class of deep neural networks, are effective for image denoising. Si and Yuan (2018) and Li et al. (2018) apply a CNN-based workflow to

random noise attenuation and ground roll attenuation. Ye et al. (2018) suggests that a deep convolutional framelet is closely related to the theory of annihilating filter-based low-rank Hankel matrix approaches.

We introduce a seismic denoising workflow using a residual convolutional neural network (ResNet) in the *f-x* domain. With residual learning, deep CNNs are easier to optimize, improve the accuracy of image classification, and better detect objects (He et al., 2016). The general workflow consists of a Fourier transform of the signal, constructing a ResNet training model, denoising in the *f-x* domain using a CNN, and inverting the Fourier transform of the desnoised data. To implement the feed-forward neural network in the *f-x* domain, we adopt a complex-valued ResNet as the denoising model. Denoising complex signals in the frequency domain using a deep neural network is well established in speech recognition and processing areas. Because the signals in the frequency domain are complex valued, a number of denoising approaches can be used. For example, time-frequency masking (Yilmaz and Rickard, 2004) first analyzes the signal as a magnitude spectrum, after which a time-frequency mask is applied for the separation of signals. In seismic processing, this technique has been used to suppressed aliased air waves on seismic shot gathers for more than 20 years. Weiqiang et al. (2018) suggested a denoising workflow of seismic signals in the time-frequency domain, which splits the complex signal into real and imaginary values and then feeds them into a two-channel deep neural network. In this study, we implement a complex-valued neural network (ComplexNet) with input and output being complex-valued signals comprised of complex building blocks. As seen in Figure 4.1, the signals are less coherent and less sparse when they are split into real and imaginary

parts. In a ComplexNet our objective is to enhance components of the spectral magnitude and to maintain a simple denoising process.

We begin our paper with a brief review of the theory of the ResNet. We then construct a suite of noisy and noise-free synthetics to both train and test our network on both random and coherent noise. After training and testing, we apply our ResNet to prestack-time migrated, stacked data from a marine data volume acquired in the Jeju Basin, Korea, and to a land data volume acquired in Texas. We conclude with a summary of the limitations and advantages of the ResNet to the seismic denoising problem.

Theory

Fourier transforming a noisy observation, $a_{\text{noisy}}(t, x)$ using

$$A_{\text{noisy}}(f,x) = \int_{-\infty}^{\infty} a_{\text{noisy}}(t,x) e^{-j2\pi f t}\, dt, \tag{1}$$

provides a noisy signal, were $A_{\text{noisy}}(f,x)$ in the f-x domain. This noisy signal can be written as the sum of desired clean signal, $A_{\text{clean}}(f,x)$, and noise $W(f,x)$:

$$A_{\text{noisy}}(f,x) = A_{\text{clean}}(f,x) + W(f,x). \tag{2}$$

In order to denoise an image in the space-space domain (x-y) or t-x domain, the loss function to minimize is the mean squared error between the desired residual and estimated noise from the noisy input:

$$Loss(\theta) = \frac{1}{2N} \sum_{n=1}^{N} \left\| R\left(a_{\text{noisy } n}; \theta\right) - \left(a_{\text{noisy } n} - a_{\text{clean } n}\right) \right\|_F^2, \tag{3}$$

where θ contains parameters for the ResNet model, N is the number of noisy-clean training pairs, and F indicates the Frobenius Norm (Zhang et al., 2017).

In the frequency domain, the signal is complex-valued. We therefore adopt the following loss function:

$$Loss_{freq}\ (\theta) = \frac{1}{2N}\sum_{n=1}^{N}(C_n\overline{C_n})_F \tag{4}$$

where $Cn = R\left(A_{noisy\ n};\theta\right) - (A_{noisy\ n} - A_{clean\ n})$ and $\overline{C_n}$ is the complex conjugate of Cn.

Attenuation of random noise

To begin, we generate a suite of synthetic data sets with different types and levels of noise. By knowing both the noise and the noise-free signal, we can train a ResNet model. Then we test the ResNet model using noise-contaminated data not used in the training step (Figure 4.2). In our first tests, we add random noise to the synthetic data. To test our denoising workflow on synthetic data, we modeled 100 synthetic seismic records of multiple reflectors with different amplitudes, polarities, and dip structures. The seismic records were computed by convolving impedance models with a Ricker wavelet having a peak frequency of 40 Hz. Each seismic record has 200 time samples and 100 traces, with a sample interval of 2 ms and a trace spacing of 100 ft. Gaussian noise is added with different signal-to-noise ratios (S/N) with values ranging from 0 dB to 10 dB at 1 dB intervals, where we measure the S/N as the ratio of signal power of the true reflectivity

model compared to the noise power. The result is a total of 1100 instances of seismic records (100 synthetic records × 11 noise levels) for training and testing. We applied a 1D fast Fourier transform (FFT) to both noisy and clean input signals in the *t-x* domain to generate noisy – clean pairs which serve as training input and output, respectively. We randomly selected 70% of the data for training and 30% of the data for testing. For training, we generated patches from noisy and clean data with a patch size of 40 × 40 samples, where the amplitudes are normalized to the interval of [0,1]. To build the complex-valued ResNet training model, we modify an architecture used by Zhang et al. (2017) shown in Figure 4.3. The ResNet model includes: 1) a convolutional layer containing 64 filters with a size of 3 × 3 × 2 (width × height × channels (real and imaginary)), a stride of 1, and a rectified linear unit, or ReLU, activation function, 2) five blocks each containing a single convolutional layer, batch normalization, and a ReLU activation function, and 3) a final convolutional layer.

We implement our denoising model with building blocks consisting of complex convolutions, complex-valued activations, and complex batch normalizations (Trabelsi et al., 2017). The training input and output samples are complex valued as well. The batch size of the neural network is 128 and the number of epochs or iterations is 60.

Applying our denoising method to seismic records with linear dipping reflectors reduces a significant amount of noise while maintaining the amplitude of the true data (Figure 4.4, 4.5 and Table 4.1). The 2 dB Gaussian random noise in the *t-x* domain appears as a randomized pattern after the noisy data are transformed into the *f-x* domain, while the reflectors are coherent in the *f-x* domain (Figure 4.4b). This randomized pattern causes the

signal to be high-rank, which can be reduced through use of deep convolutional layers. After applying the denoising process using the complex-valued ResNet, the random noise is significantly alleviated (Figure 4.4c). Moreover, the loss function tends to minimize the difference of magnitude between the true and noisy data. The effect of denoising is evident in the magnitude as well as in the real and imaginary parts. Applying the same workflow to more curvilinear reflectors shows the same improvements after denoising (Figure 4.5). The linear and curvilinear reflector models both have overlapping events, which is challenging to denoise in the *t-x* domain because denoising possibly damages the amplitude of the reflectors. However, by implementing noise attenuation in the *f-x* domain, the overlapping reflectors preserve their geometry and amplitude after denoising (Figure 4.4c and 4.5c). In this synthetic example, the denoising method recovers a thinning bed which is challenging to see in the noisy data (Figure 4.6). Also, a low-amplitude horizon becomes more recognizable after random noise attenuation (Figure 4.7).

Next, we apply the random noise attenuation method to a data volume that has been prestack time migrated and stacked with a bin spacing of 25×25 m and time increment of 2 ms.. For this test, we cropped the 3D seismic to include 100 inlines, 1600 crosslines and 800 time samples (1600 ms). We used the same training data and procedure as we used for synthetic example.

Significant amounts of random noise are removed after applying the noise attenuation method (Figure 4.8c), whereas the underlying signal becomes more coherent (Figure 4.8b). The coherent signals consist of migration artifacts as well as horizons, which

indicates the ResNet removes the random noise only, since the difference of training input and output is random noise.

Attenuation of coherent noise

Our noise attenuation method is tested 1) to remove the linear noise from synthetically generated data and 2) to alleviate migration operator aliasing artifacts in field data. To test the denoising method on synthetic data, we modeled 500 synthetic seismic records of multiple parabolas with linear noise with different amplitudes, polarities and dips. The seismic records have 200 time samples and 128 traces, with a sample interval of 2 ms and a trace spacing of 100 ft. Like our previous random noise attenuation exercise, the data are split into training and test subsets with ratio of 70%: 30%, then Fourier transformed in the *f-x* domain. The neural network architecture is the same as the one used for random noise attenuation except that the patch size is now 64 × 64. The neural network architecture successfully distinguishes the aliased linear noise from the unaliased parabolic events in the frequency – space domain (Figure 4.9). However, if the linear noise patterns are stronger that the underlying reflectors, the noise attenuation fails, as indicated by the red arrow in Figure 4.10 c.

We apply our denoising method to land seismic data which is acquired in the Fort Worth Basin in Texas to verify whether the denoising method is effective for removing coherent noise in real data. we crop a 3D post-stack land seismic to 198 inlines × 200 crosslines × 1001 time samples. We sampled 198 vertical slices with size of 1001 time samples and 200 crosslines. As shown in Figure 4.11 a, the data suffer from a lot of U-

shaped migration artifacts. A structure-oriented filter (SOF) is applied to the data after stacking to generate denoised training output. SOF attenuates both rand noise and coherent noise that cut across the reflector dip. Discontinuities are preserved in SOF using coherence information. After splitting the data into training and test subsets with a ratio of 70%:30%, we iterated the denoising neural network using the same architecture as we used for denoising of the synthetic data. The difference of noisy data and denoised data shows a great amount of migration artifacts are eliminated (Figure 4.11 c).

Conclusions

We introduce a deep learning approach of random noise attenuation in the *f-x* domain using a complex-valued ResNet where we find that the deep convolutional layers effectively exploit the coherent characteristics of signals. ResNet speeds up the process of denoising and improves the accuracy, which efficiently separates the noise from signals. Above all, the biggest merit of this approach is using complex-valued inputs, outputs, and building blocks which enables us to implement the noise attenuation process in *f-x* domain and successively recover the denoised coherent signals. Subtle feature such as thin beds and reflectors with low amplitude masked by noise are recovered by applying random noise attenuation. This neural network architecture can also be used to attenuate coherent noise such as aliased linear noise or migration artifacts. One limitation is that noise can remain when the amplitude of the noise is significantly higher than signal. Therefore, the ResNet may be less effective in eliminating noise such as high amplitude ground-roll on common shot records.

Acknowledgements

The authors thank the sponsors of OU Attribute Assisted Seismic Processing and Interpretation (AASPI) consortium for their encouragement and financial support.

References

He, K., X. Zhang, S. Ren, and J. Sun, 2016, Deep residual learning for image recognition: In Proceedings of the IEEE conference on computer vision and pattern recognition, 770-778.

Jain, V., and S. Seung, 2009, Natural image denoising with convolutional networks: In Advances in neural information processing systems 769-776.

Li, H., W. Yang, and X. Yong, 2018, Deep learning for ground-roll noise attenuation: 88th Annual International Meeting, SEG, Expanded Abstracts, 1981-1985.

Naghizadeh, M., and M. D. Sacchi, 2010, On sampling functions and Fourier reconstruction methods: Geophysics, **75**, no. 6, WB137–WB151.

Naghizadeh, M., and M. D. Sacchi, 2012, Multicomponent *f-x* seismic random noise attenuation via vector autoregressive operators. Geophysics, **77**, no. 2, V91-V99.

Oropeza, V. E., and M. D. Sacchi, 2009, Multifrequency singular spectrum analysis: 79th Annual International Meeting, SEG, Expanded Abstracts, 3193–3197.

Si, X., and Y. Yuan, 2018, Random noise attenuation based on residual learning of deep convolutional neural network: 88th Annual International Meeting, SEG, Expanded Abstracts, 1986-1990.

Trabelsi, C., O. Bilaniuk, Y. Zhang, D. Serdyuk, S. Subramanian, J. F.Santos, S. Mehri, N. Rostamzadeh, Y. Bengio, C. J. Pal, 2017, Deep complex networks. arXiv preprint arXiv:1705.09792.

Ye, J. C., Y. Han, and E. Cha, 2018, Deep convolutional framelets: A general deep learning framework for inverse problems. SIAM Journal on Imaging Sciences, 11(2), 991-1048.

Yilmaz, O., and S. Rickard, 2004, Blind separation of speech mixtures via time-frequency masking. IEEE Transactions on signal processing, **52**-7, 1830-1847.

Yin, R., T. Gao, Y. M. Lu, and I. Daubechies, 2017, A tale of two bases: Local-nonlocal regularization on image patches with convolution framelets: SIAM Journal on Imaging Sciences, **10**-2, 711-750.

Zhang, K., W. Zuo, Y. Chen, D. Meng, and L. Zhang, 2017, Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising: IEEE Transactions on Image Processing, **26**-7, 3142-3155.

Zhou, Y., S. Li, D. Zhang, and Y. Chen, 2017, Seismic noise attenuation using an online subspace tracking algorithm: Geophysical Journal International, **212**-2, 1072-1097.

Figures

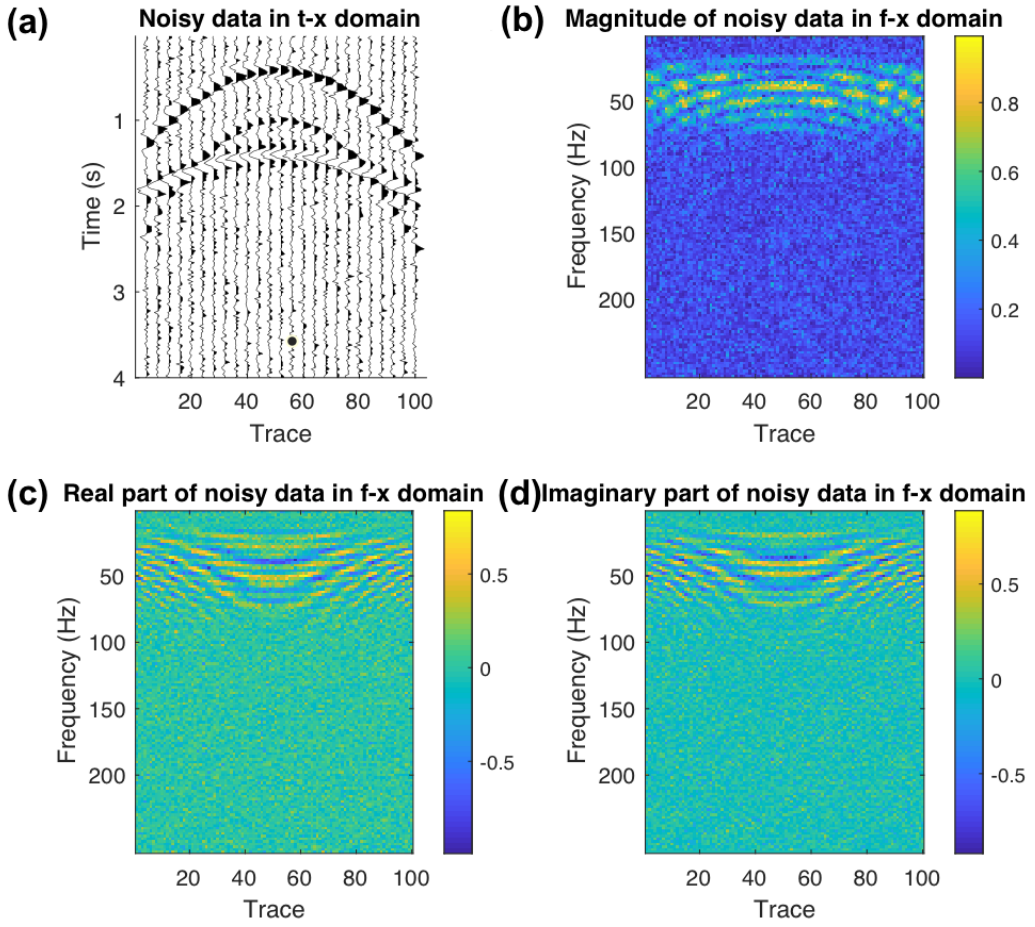


Figure 4.1. (a) An unmigrated CMP gather in the *t-x* domain. The signal-to-noise ratio is 2dB. (b) Spectral magnitude, (c) real part, and (d) imaginary part of (a) in the *f-x* domain. Note that the signals are less coherent and less sparse in the real and imaginary parts of the *f-x* domain than in the *t-x* domain.

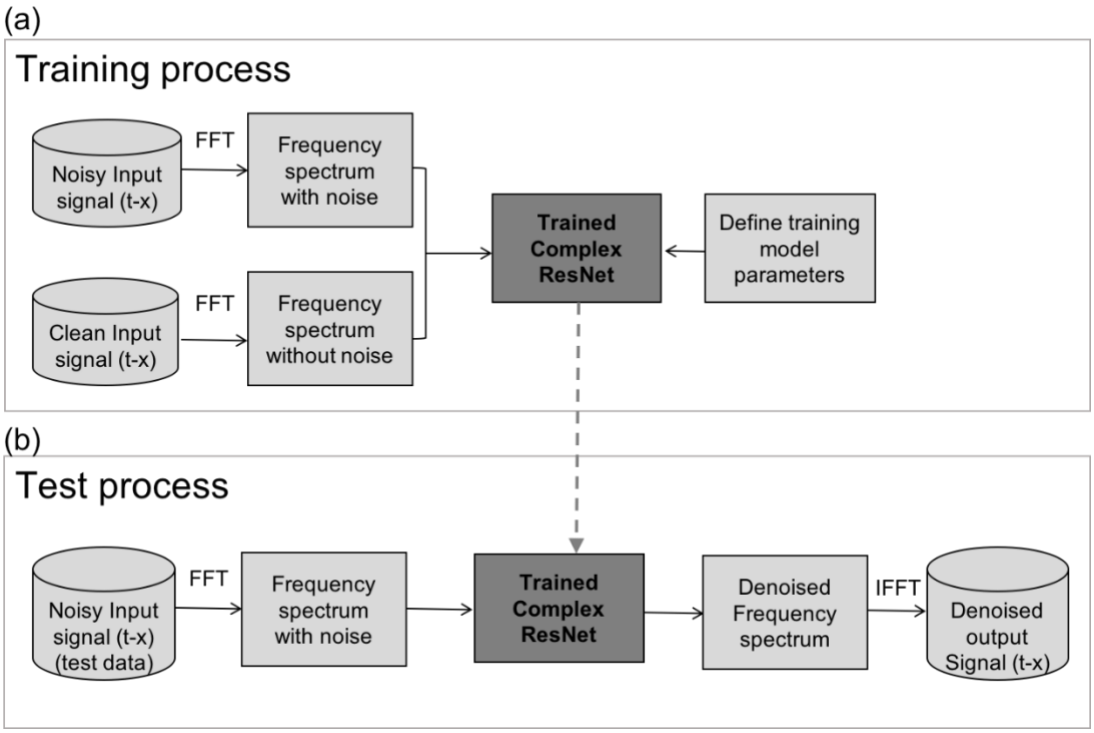


Figure 4.2. The workflow for denoising. (a) Training and validation process to build a denoising model. (b)Test process where actual denoising is applied using the trained model.

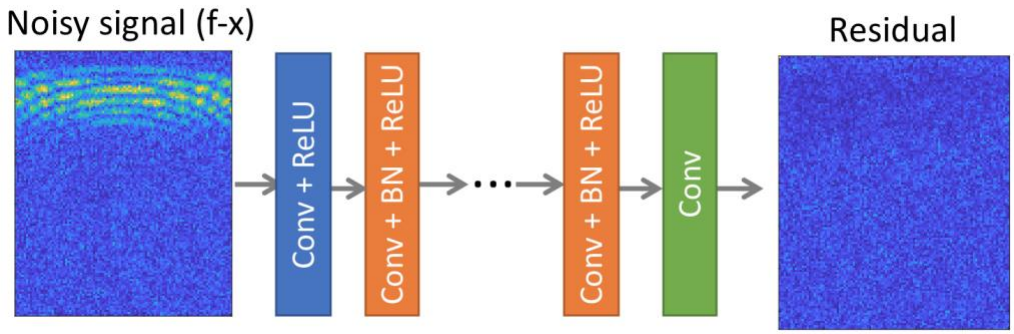


Figure 4.3. Architecture of the complex-valued ResNet consisting of 7 Convolutional layers.

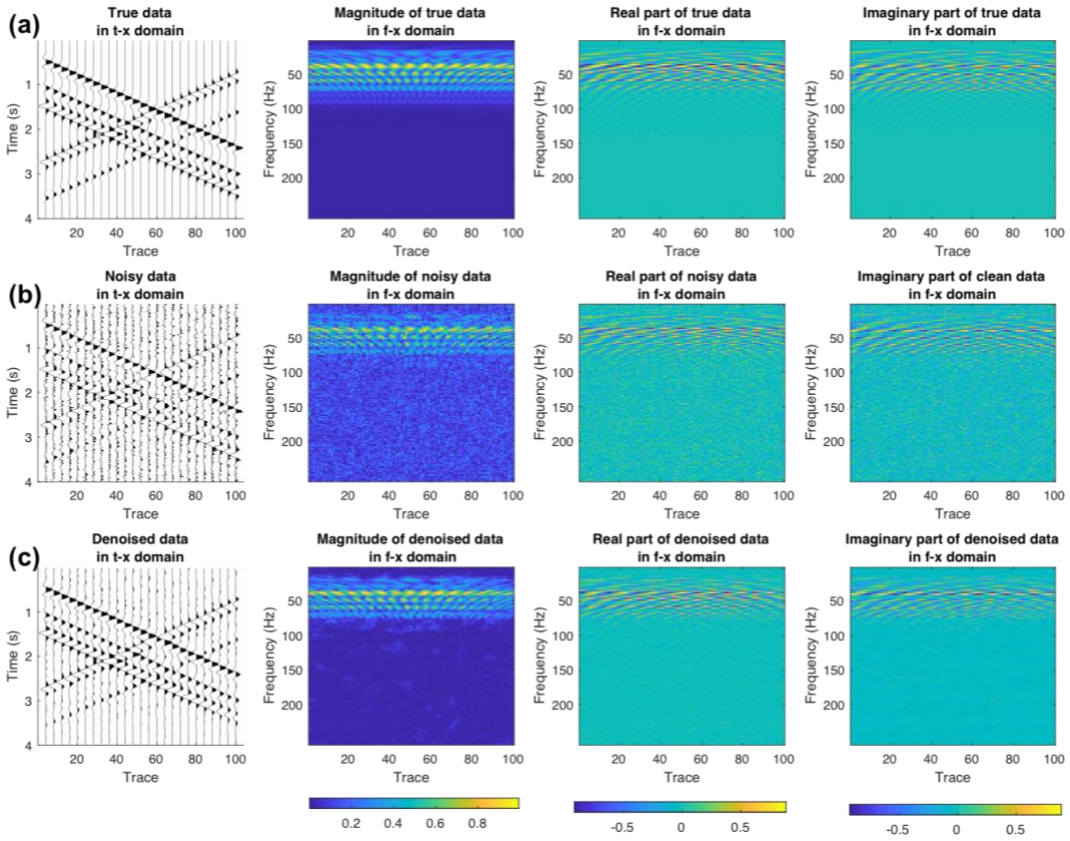


Figure 4.4. Noise attenuation of synthetic seismic records with linear dipping reflectors as seen in common shot gathers. Amplitude of the data displayed in the *t-x* domain, magnitude, real part and imaginary part of data in *f-x* domain for the (a) desired noiseless signal (b) noisy data (S/N: 2 dB) and (c) denoised data.

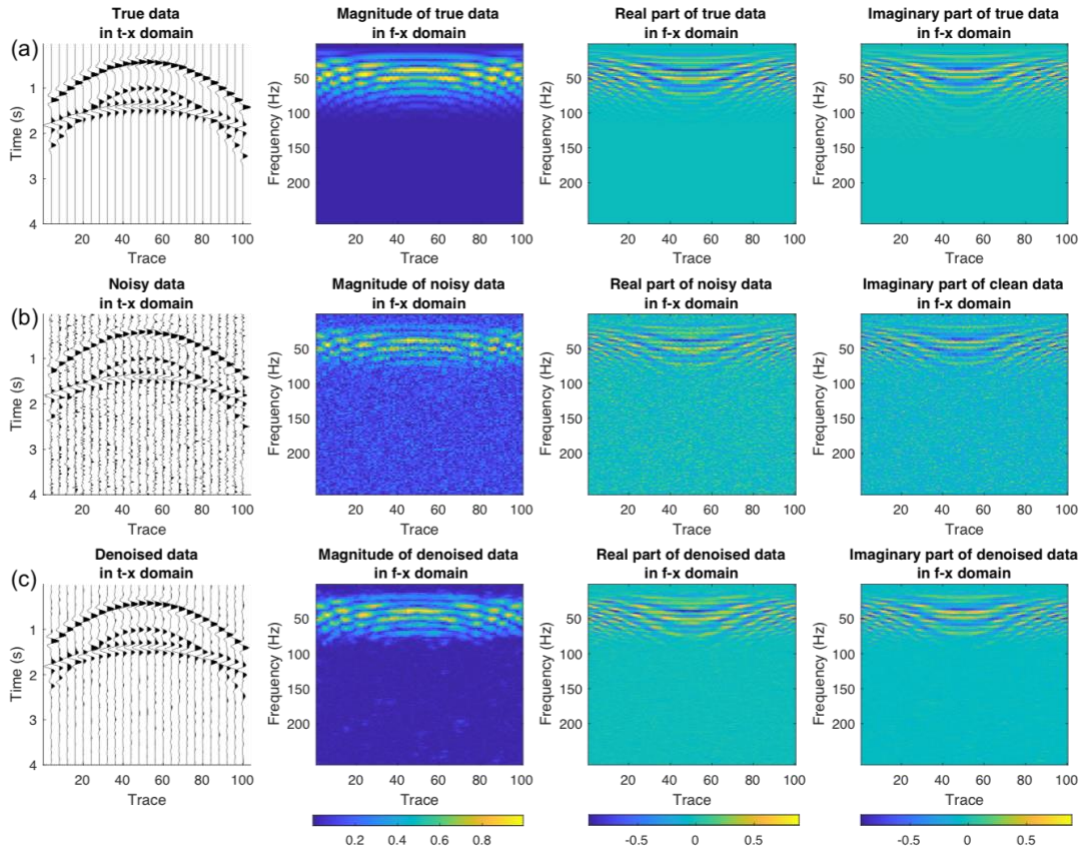


Figure 4.5. Noise attenuation of synthetic seismic records with hyperbolic reflectors as seen in CMP gathers. Amplitude of the data displayed in the *t-x* domain, magnitude, real part and imaginary part of data in the *f-x* domain for the (a) desired noiseless signal (b) noisy data (S/N: 2 dB) and (c) denoised data.

Linear model	Input SNR (dB)	2.00	4.00	6.00	8.00	10.00
	Output SNR (dB)	10.66	11.96	13.25	14.26	15.00
Curvilinear model	Input SNR (dB)	2.00	4.00	6.00	8.00	10.00
	Output SNR (dB)	10.90	12.19	13.44	14.60	15.25

Table 4.1. Comparison of SNRs for different models and input noise levels.

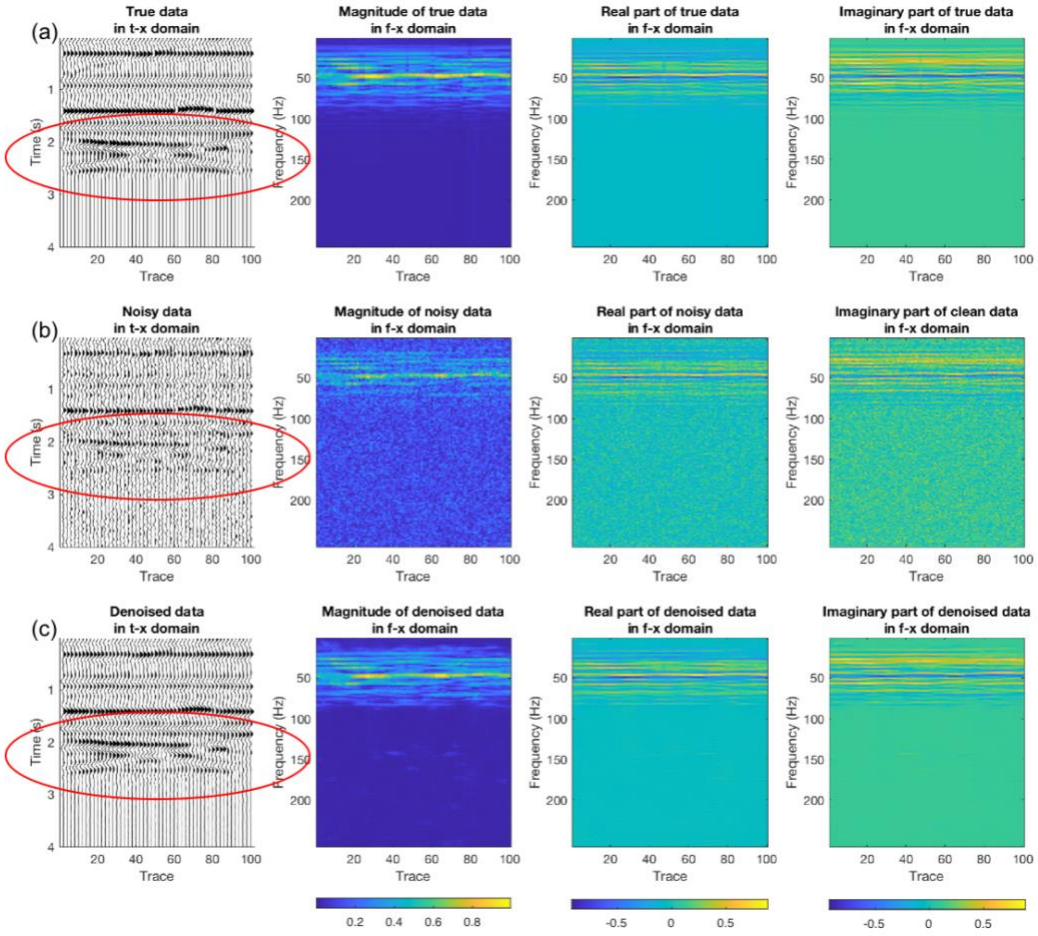


Figure 4.6. Noise attenuation of synthetic seismic records representing a line through a migrated data volume with stratigraphic features where thin bed tuning is indicated by the red ellipses. Amplitude of the data displayed in the *t-x* domain, magnitude, real part and imaginary part of data in the *f-x* domain for the (a) desired noiseless signal (b) noisy data (S/N: 0 dB) and (c) denoised data. Note that the thin bed tuning becomes clear after random noise attenuation.

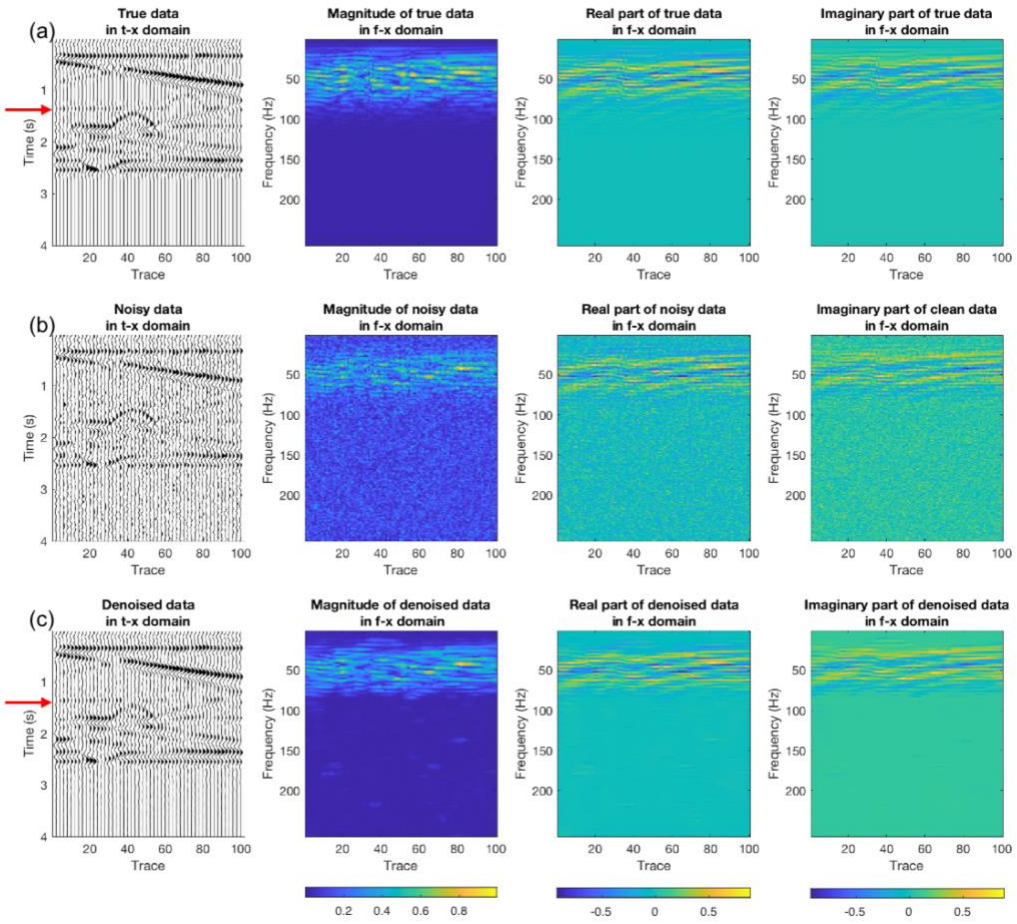


Figure 4.7. Noise attenuation of synthetic seismic records representing a line through a migrated data volume with structurally complex. Amplitude of the data displayed in the *t-x* domain, magnitude, real part and imaginary part of data in the *f-x* domain for the (a) desired noiseless signal (b) noisy data (S/N: 0 dB) and (c) denoised data. The red arrows indicate low amplitude horizon which becomes distinguishable after random noise attenuation.

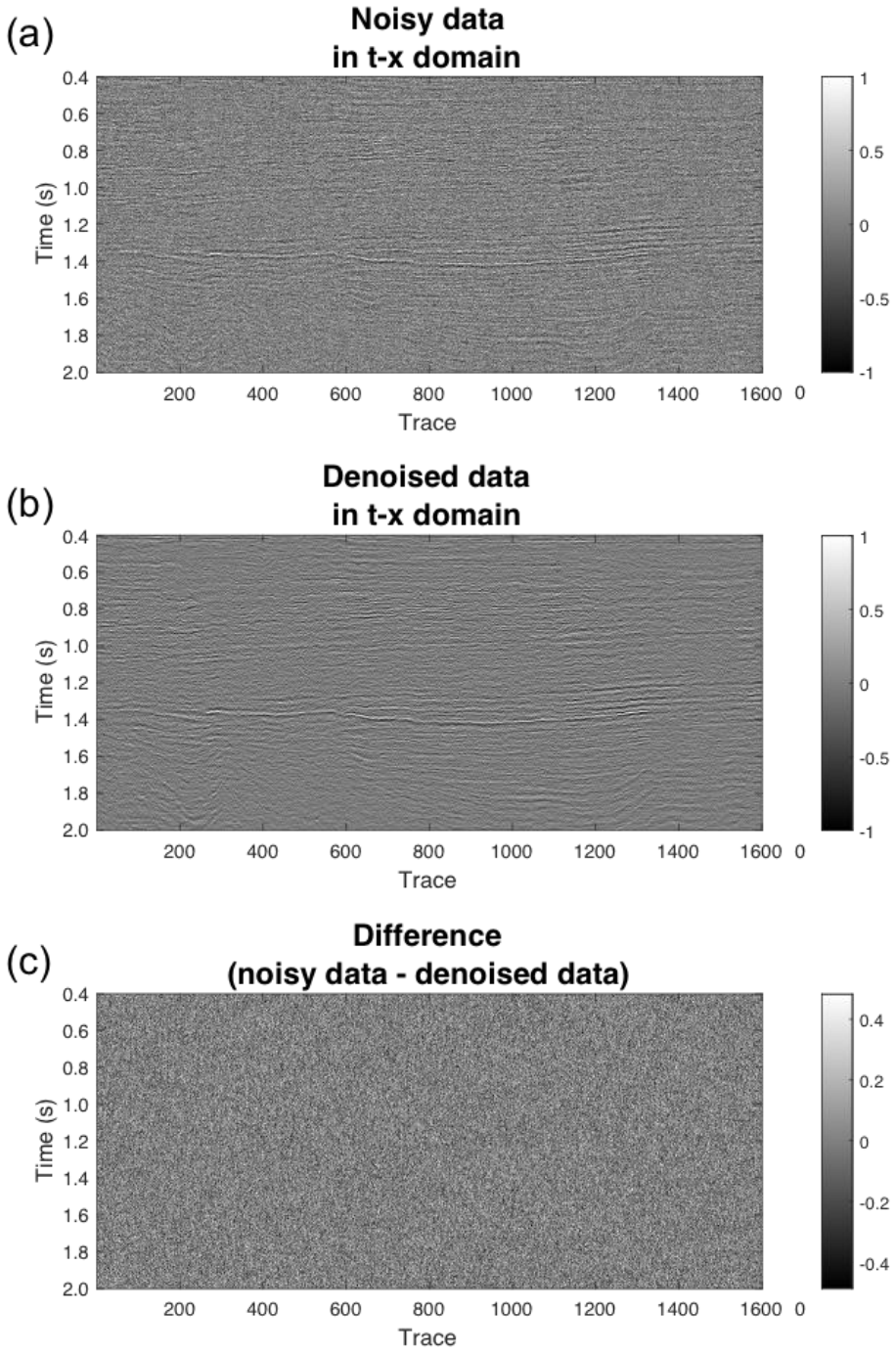


Figure 4.8. Random noise attenuation of a marine 3D prestack time migrated seismic records from the Jeju Basin, Korea, using the training constructed on the previous synthetics. Amplitude of (a) noisy data, (b) denoised data, and (c) difference of noisy data

and denoised data. The coherent signal becomes more distinguishable after applying the noise attenuation method. (Data courtesy of KIGAM).

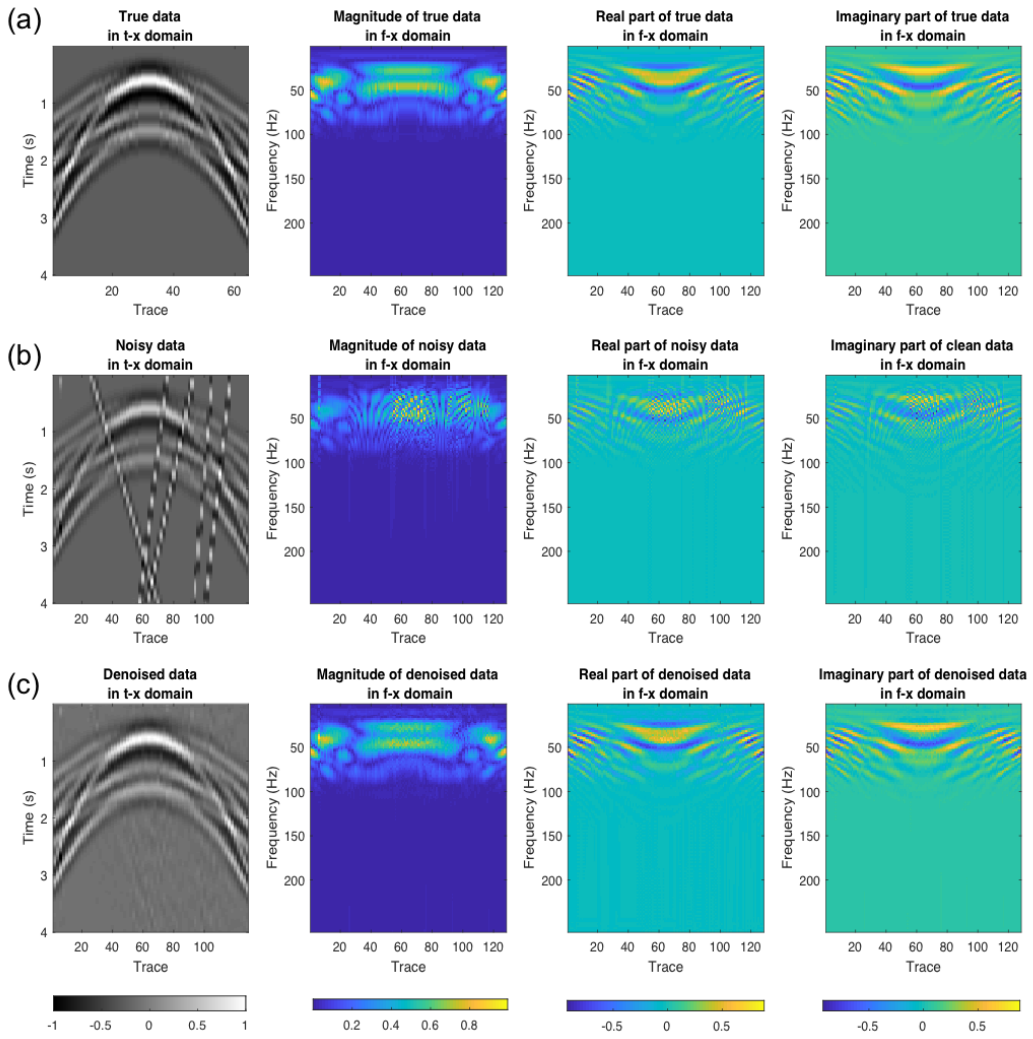


Figure 4.9. Noise attenuation of a synthetic CMP gather with hyperbolic reflectors contaminated by steeply dipping, aliased, linear noise Amplitude of the data displayed in the *t-x* domain, magnitude, real part and imaginary part of data in the *f-x* domain for the (a) desired noiseless signal (b) data with linear noise and (c) denoised data. Unlike conventional seismic processing workflows, the CNN algorithm is able to identify and then suppress the aliased noise events, much like a human interpreter can.

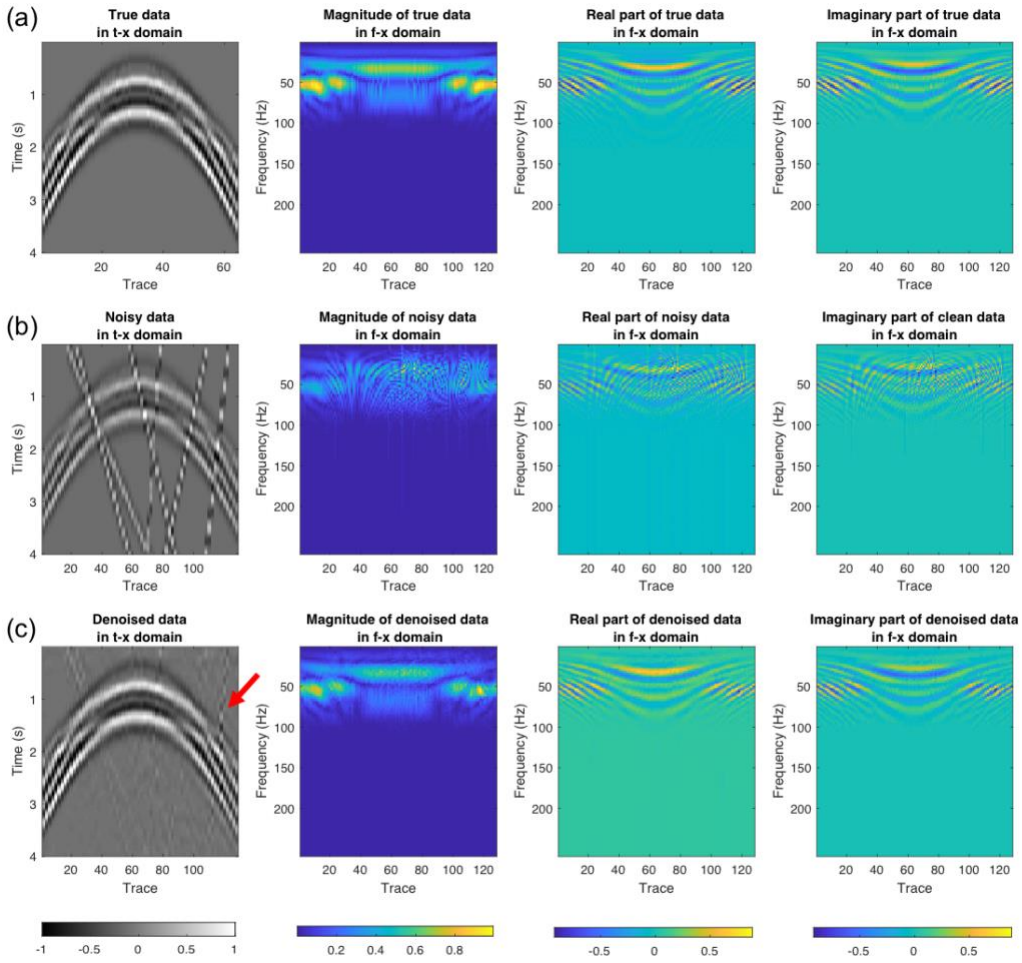


Figure 4.10. Noise attenuation of synthetic CMP gather with hyperbolic reflectors contaminated by steeply dipping, aliased, linear noise. Amplitude of the data displayed in the *t-x* domain, magnitude, real part and imaginary part of data in the *f-x* domain for the (a) desired noiseless signal (b) data with linear noise and (c) denoised data. Part of the linear noise which has high amplitude remains after noise attenuation as indicated by the red arrow.

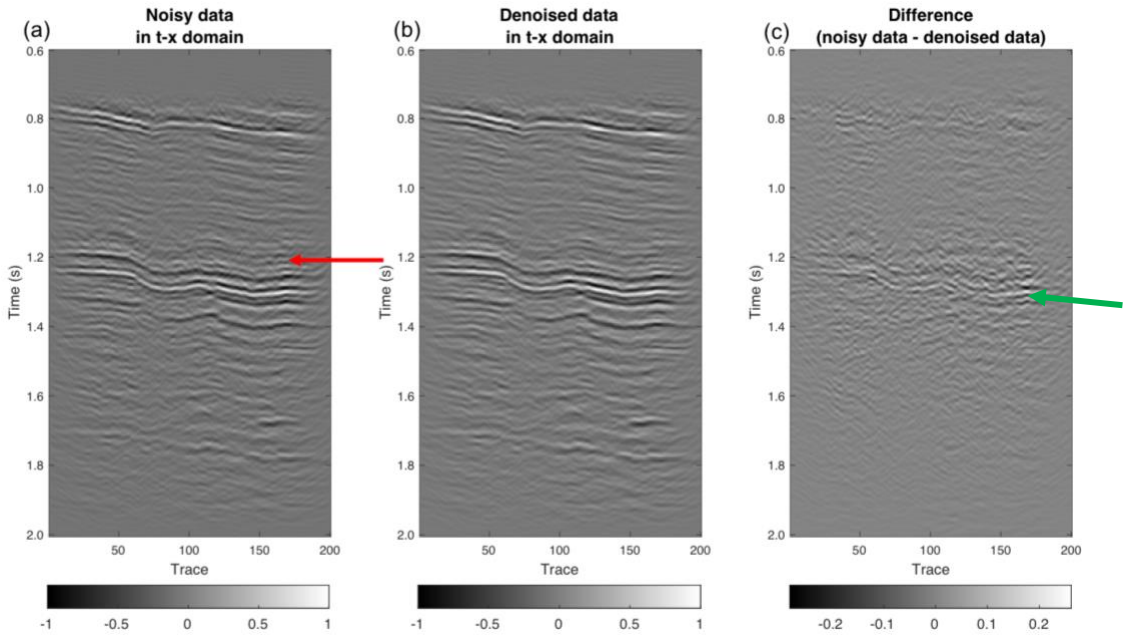


Figure 4.11. Coherent noise attenuation of land prestack time-migrated data volume. Amplitude of (a) noisy data, (b) denoised data, and (c) difference between noisy data and denoised data. The seismic record suffers significant migration artifacts which is indicated by red arrow. The difference of noisy data and denoised data shows a great amount of migration operator aliasing artifacts that are eliminated. However, it appears that some of the reflectors of interest have also been rejected (green arrow). (Data courtesy of Devon Energy).

CHAPTER 5: A COMPARISON OF GEOPHYSICAL INVERSION AND MACHINE LEARNING IN INVERSE PROBLEMS

Introduction

The objective of an inverse problem is to predict causal factors or underlying parameters from measured data, given a partial description of a physical system. Tarantola (2005) defines an inverse problem as a reverse process that predicts observations using a model of the system. In solid earth and exploration geophysics, inverse problems aim to retrieve subsurface physical properties from measured geophysical data such as surface seismic, magnetotelluric. and controlled source electromagnetic data. Full waveform inversion (FWI), simultaneous inversion and amplitude versus offset (AVO) inversion are common inversion methods used to recover physical properties such as P- and S wave velocities or P- and S-impedances using prestack seismic data.

Geophysical inverse problems are often ill-posed, non-unique, and non-linear. Deterministic approaches to solve inverse problems usually require minimizing an objective function where iterative linearized inversion algorithms such as the Gauss-Newton, steepest descent. or conjugate gradient algorithms are widely used. Geophysical inversion is usually based on an understanding of the physics of the recorded data and may include wave equations and scattering theory, as well as sampling theory. In contrast to deterministic approaches, machine learning inversion is data-driven and statistical and in principal requires only minimal understanding of the underlying physics, petrophysics, or geology. The use of machine learning has rapidly increased during the past decade and is widely used in computer science and many other disciplines including geophysics. One of

the major drivers has been ever-increasing volumes of data providing the motivation for "big data analysis". At the same time, computational power for the same budget has increased dramatically through the use of multicore processors and graphical processing units.

Machine learning solves a problem by optimizing a performance criterion based on statistical analyses using example data or past experience (Parsons, 2005). Machine learning falls into two broad categories – called supervised and unsupervised learning. Artificial neural network (ANN), Naïve Bayes and support vector machine (SVM) are popular supervised learning methods for classification and regression in geophysical analysis. Self-organizing maps (Coléou et al., 2003) and generative topographical mapping (Roy et al., 2013) are two unsupervised learning algorithms.

Kim and Nakata (2018) provided some preliminary analysis of the benefits of classical inversion vs. machine learning. They found that both classical physics-based inversion and machine learning methods are useful for solving inverse problems. In this study, we compare geophysical inversion based on a least-squares method with a neural network based supervised machine learning method to improve the vertical resolution of seismic amplitude data. Least-squares inversion minimizes the sum of squared error between the modeled and measured data. Least-squares estimates of reflectivity to improve seismic resolution go back 50 years or more with the original work by Robinson and Treitel (1969), and many others with the application of Wiener filtering to seismic deconvolution. Since then, many other algorithms have been developed ranging from maximum entropy

deconvolution (give reference) to reflectivity inversion using sparse spike constraints (Zhang and Castagna, 2011).

Classical inversion for reflectivity is a highly nonlinear problem. For this reason, we wish to evaluate the nonlinear capabilities of neural networks as an inversion method. Unlike other problems in geophysical analysis such as the facies classification problems discussed in Chapters 2 and 3 of this dissertation, the training for the neural network inversion is readily available – we simply use the same physics-based forward modeling algorithms used in classical inversion to generate the training data.

We begin our paper with review of the theory for both classical inversion and for neural network inversion. Next, we show how regularization to compensate for noise in the data operates in two algorithms. Next, we test the sensitivity of both algorithms to noise in the data. Finally, we apply these algorithms to a 3D data volume from a Texas Barnett Shale survey and compare their results.

Theory

Both geophysical inversion and machine learning involve processes which converts input data from data space to model space. In the case of reflectivity inversion, the seismic trace $s(t)$ can be represented as a convolution of the seismic wavelet and Earth's reflectivity denoted by

$$s(t) = w(t) * r(t) + n(t), \tag{1}$$

where *w(t)* is seismic wavelet, *r(t)* is reflectivity and *n(t)* is seismic noise. The objective of the inverse problem is to obtain an unknown reflectivity model *r(t)*. In the context of geophysical inversion, *w(t)* and *s(t)* are presumably known, and *n(t)* and *r(t)* are unknown to investigators. Ideally, the wavelet *w(t)* is constructed by comparing sonic and density logs to the surface seismic data, In the absence of quality well control we can assume properties about Earth's subsurface and compute a statistical wavelet.

The objective function (also known as cost function or loss function) of inverse problem using a constrained least-squares is

$$L = \frac{1}{2}\left\|s(t) - w(t) * \widehat{r(t)}\right\|_2^2 + \lambda\left\|\widehat{r(t)}\right\|, \tag{2}$$

where $\widehat{r(t)}$ is the modeled reflectivity. $\lambda\left\|\widehat{r(t)}\right\|$ is an L1 norm regularization term and λ is non-negative regularization parameter. The L1-norm regularizer recovers a sparse solution and is called basis pursuit (Chen et al., 2001; Zhang and Castagna, 2011) in signal processing and Lasso in statistics.

In the machine learning approach, the forward wavelet operator *w(t)* can be unknown. Instead, some portion of input dataset $r_n(t)$ and output dataset $s_n(t)$ are provided to a supervised learning algorithm as a training dataset. The machine learning approach to inverse problems can be denoted as

$$L_{learn} = \frac{1}{2}\left\|r_n(t) - W_\theta^\dagger s_n(t)\right\|_2^2 + \lambda\|\theta\|, \tag{3}$$

where θ is a parameter set optimized during the learning process and $W_{\theta}^{\dagger}$ is a pseudo-inverse operator or mapping function that is given by θ (Adler and Öktem, 2017). For instance, neural networks approximate the inverse mapping from the data space into the model space using non-linear basis functions with weights and biases. In this case, weights and biases which are determined during the learning process are equivalent to the parameter set, θ. This set of weights and biases in the neuron layers defines the pseudo-inverse operator $W_{\Theta}^{\dagger}$. The L1 regularizer term $\lambda\|\theta\|$ is applied to favor sparsity of reflectivity and to prevent overfitting.

Methodology

In this study, we test and compare geophysical inversion and machine learning methods in the aspects of: 1) inversion sensitivity to Gaussian noise in the data, and 2) the effect of the regularization terms when inverting noisy data. To build a deep neural network (DNN) model as an example of the machine learning approach, we synthetically generate the training datasets by convolving 25 Hz Ricker wavelet and reflectivity sequences. The total number of training samples is equivalent to the multiplication of the 30 reflectivity models used and the 266,200 number of observations in each reflectivity model (266,200 × 30 = 7986000 observations).

The seismic traces modeled from each reflectivity model includes 30 time samples with a 2 ms interval. A ”feature” in DNN corresponds to windowed part of the seismic

trace ranging for -40 to 40 ms about the analysis point. A reflectivity value corresponding to the modeled seismic trace then becomes the desired training output for the inverse problem. In each reflectivity model, three reflectors vary in location and magnitude. The location changes at 2 ms increments, whereas the magnitude ranges from -1.0 to 1.0 with an increment of 0.2 thereby generating 266,200 synthetic datasets (Figure 5.1). For field datasets, reflectivity magnitudes ranges from -1.0 to 1.0 with 0.1 increment. When the source wavelet is not close to 25 Hz, we can squeeze or stretch the signals and can still use the mapping function for different datasets.

We construct our DNN model with an input layer, three hidden layers and an output layer. The three hidden layers have 200, 100, and 50 neuron units, respectively. In a neural network, an activation function decides whether to convert an input signal of a node to an output signal or not. We find the rectified linear unit (ReLU) to be efficient for training DNN with large datasets and use it as the activation function. To minimize the objective function given by equation 3, we adopt an adaptive moment estimation (ADAM) gradient-descent optimization algorithm. During the training process, 10% of the neuron units are dropped to prevent overfitting of the DNN model.

For the field data example, we apply the two algorithms to evaluate the resolution of the inverted reflectivity. The seismic data are acquired over the Barnett Shale play in the Fort Worth Basin, Texas. We crop a vertical slice with size 367 x 271 traces from seismic volume. Traces are sampled up to 2.5 s with the sample interval of 2 ms. We estimate the source wavelet, $w(t)$, using a statistical method (Figure 5.2). For the real data, we use this statistical source wavelet to generate the training dataset for neural network. The same

source wavelet is used internally as part of forward modeling operator used in least-squares inversion. The L1 regularization parameter λ for the neural network and least-squares methods are 0 and 1.5E-5, respectively, which provide the minimum error between the measured seismic traces and seismic traces modeled from the inverted reflectivity for each method.

Results and discussion

We examine the sensitivity of two methods for noisy data (Figures 5.3 and 5.4). White Gaussian noise is added to seismic traces with signal-to-noise ratio (S/N) of 0 dB, 2 dB and 4 dB. Here we measure the S/N as the ratio of signal power of the true reflectivity model compared to noise power. To test the sensitivity of the noise for the two methods and evaluate an effect of L1 regularization with respect to the noise, we calculate the correlation coefficient between the true reflectivity and the inverted reflectivity for different regularization coefficients (Figure 5.4). We use ten different models (one realization is shown in Figure 5.3a) to estimate statistically reliable correlation coefficients. For noise-free data, the neural network model yields high accuracy without a regularization term (Figures 5.3c and 5.4b). Adding regularization results in under-fitting, which reduces the accuracy of the results. In the presence of noise, however, a proper value of the regularization term enhances the accuracy and suppresses the noise (Figures 5.3d-e and 5.4a-b). The under-fitted model means that the mapping function is smoother, and thus the model can recover a good solution in the wide range in noisy datasets. For the least-squares model, adding the L1-regularization term gives better results whether or not there is any

noise. For the least-squares model the inverse operator has the effect of a bandwidth extension filter (Zhang and Castagna, 2011), such that without regularization the noise may be amplified during the inversion process. The regularization term filters out these undesired components when their eigenvalues are small.

We also add the different levels of noise to the training data and apply inversion to a noisy test set to examine the effect of noise in neural network training. Noise in the training data appears to alleviate the effect of noise in the test data and enhances the accuracy of predictions (Figure 5.5) since it prevents overfitting in the training process. Bishop (1995) reports that training with noise is equivalent to adding a regularization factor. In the case of noise-free test data, however, the use of noisy training data reduces the accuracy of the inversion, which has the same effect as adding regularization.

In the case of the Barnett Shale example, the inversion result using a neural network shows higher resolution compared to the inputseismic traces without the ringing effect experienced by aggressive spectral balancing (Figure 5.6 and 5.7). The least-squares method also resolves thin beds; however, the reflectivity is less definite, and the result shows a ringing effect or noise around traces 220-240. The total computation time of the two methods are compared in Table 5.1. Even if the NN method requires more computing time for training, the time elapsed for the inversion process in NN method is comparatively shorter than that of the LS method.

Conclusions

We examine conventional geophysical inversion and compare to machine learning as a methodology to solve an inverse problem. We show the similarities and differences of such methods based on mathematical expression and take reflectivity inversion as an example of the more general inverse problem. The appropriate value of the regularization term in the least-squares method helps to recover true reflectivity in both cases of noise-free and noise contaminated data. In both least-squares and neural network methods, adding a regularization term yields a smooth solution which has advantages in inverting noisy data. The regularization term in least-squares alleviates the ringing effect when data contains noise, however, regularization also suppresses the magnitude of impedance.

Although there are no explicit geological or petrophysical constraints in the DNN solution, there are implicit constraints. By construction, all of my models can be considered to be sparse spike. If I had generated model data from real well logs, resulting in a more continuous reflectivity series, the results would be quite different.

Acknowledgements

The authors thank the sponsors of OU Attribute Assisted Seismic Processing and Interpretation (AASPI) consortium for their encouragement and financial support.

References

Adler, J. and O. Öktem, 2017, Solving ill-posed inverse problems using iterative deep neural networks: Inverse Problems, **33**, no. 12, 124007.

Alpaydin, E., 2009, Introduction to machine learning: MIT press.

Baddari, K., N. Djarfour, T. Aïfa and J. Ferahtia, 2010, Acoustic impedance inversion by feedback artificial neural network: Journal of Petroleum Science and Engineering, **71,** 106-111.

Bishop, C. M., 1995, Training with noise is equivalent to Tikhonov regularization: Neural computation, **7**-1, 108-116.

Calderon-Macias, C., M. K. Sen and P. L. Stoffa, 2000, Artificial neural networks for parameter estimation in geophysics: Geophysical Prospecting, **48**, no. 1, 21-47.

Chen, S. S., D. L. Donoho, and M. A. Saunders, 2001, Atomic decomposition by basis pursuit: Society for Industrial and Applied Mathematics Review, **43**, no. 1, 129–159.

Chung, H., and D. C. Lawton, 1995, Frequency characteristics of seismic reflections from thin beds: Canadian Journal of Exploration Geophysics, **31**, 32–37.

Constable, S., A. Orange, and K. Key, 2015, And the geophysicist replied: "Which model do you want?": Geophysics, **80**, no.3, E197-E212.

Ji S.H., Xue Y. and Carin L. 2008. Bayesian compressive sensing. IEEE Transactions Signal Processing, **56**, 2346–2356.

Kappler, K., H. A. Kuzma, and J. W. Rector, 2005, A comparison of standard inversion, neural networks and support vector machines: SEG Technical Program Expanded Abstracts 2005, 1725-1727.

Kuzma, H. A., and J. W. Rector, 2004, Non-linear AVO inversion using Support Vector Machines. SEG Technical Program Expanded Abstracts 2004, 203-206.

McCann, M. T., K. H. Jin and M. Unser, 2017, A review of convolutional neural networks for inverse problems in imaging. arXiv preprint arXiv:1710.04011.

Ray, A., K. Key, T. Bodin, D. Myer and S. Constable, 2014, Bayesian inversion of marine CSEM data from the Scarborough gas field using a transdimensional 2-D parametrization: Geophysical Journal International, **199**-3, 1847-1860.

Reading, A. M., M. J. Cracknell, D. J. Bombardieri and T. Chalke, 2015, Combining machine learning and geophysical inversion for applied geophysics: ASEG Extended Abstracts 2015, **1**, 1-5.

Robinson, E. A., and S. Treitel, 2000, Geophysical signal analysis. Society of Exploration Geophysicists.

Rojo-Álvarez, J., M., Martínez-Ramón, J. Muñoz-Marí, G. Camps-Valls, C. M. Cruz and A. R. Figueiras-Vidal, 2008, Sparse deconvolution using support vector machines: EURASIP Journal on Advances in Signal Processing, **2008**, 1-13

Röth, G., and A. Tarantola, 1994, Neural networks and inversion of seismic data: Journal of Geophysical Research, **99**, 6753-6768.

Roy, A., 2013, Latent space classification of seismic facies. University of Oklahoma.

Schmidhuber, J., 2015, Deep learning in neural networks: An overview: Neural networks, **61**, 85-117.

Snieder, R., 1998, The role of nonlinearity in inverse problems. Inverse Problems, **14**-3, 387.

Tarantola, A., 2005, Inverse problem theory and methods for model parameter estimation V. 89: siam.

Van der Baan, M. and C. Jutten, 2000, Neural networks in geophysical applications. Geophysics, **65**, no. 4, 1032-1047.

Wipf, D. P., and B. D. Rao, 2004, Sparse Bayesian learning for basis selection: IEEE Transactions on Signal processing, **52,** no. 8, 2153-2164.

Yuan, S., and S. Wang, 2013, Spectral sparse Bayesian learning reflectivity inversion: Geophysical Prospecting, **61**, no. 4, 735-746.

Zhang, R., and J. Castagna, 2011, Seismic sparse-layer reflectivity inversion using basis pursuit decomposition: Geophysics, **76,** no. 6, R147-R158.

Figures

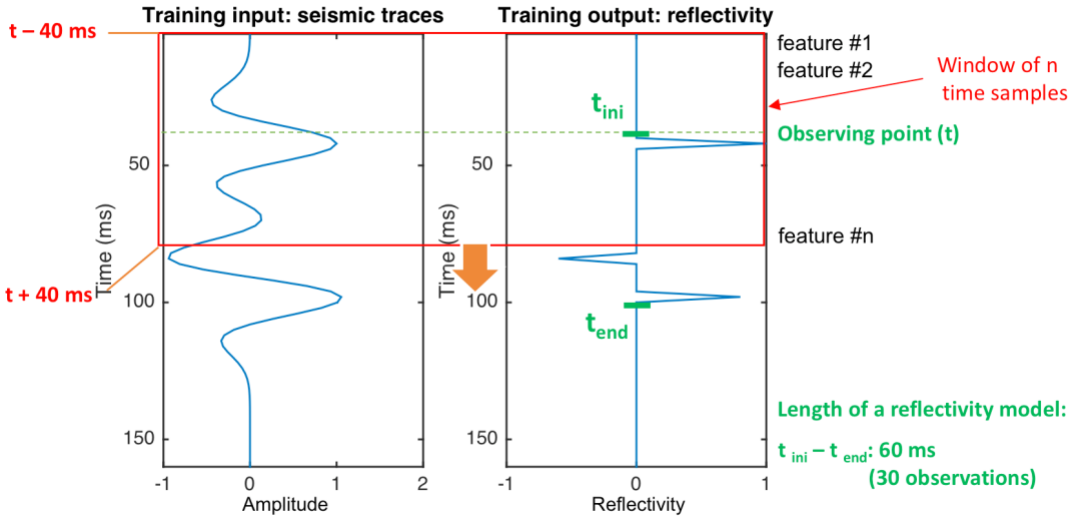


Figure 5.1. Description of training input and output data for reflectivity inversion using neural network methods. Each of the 61 reflectivity time samples is a neural network "feature".

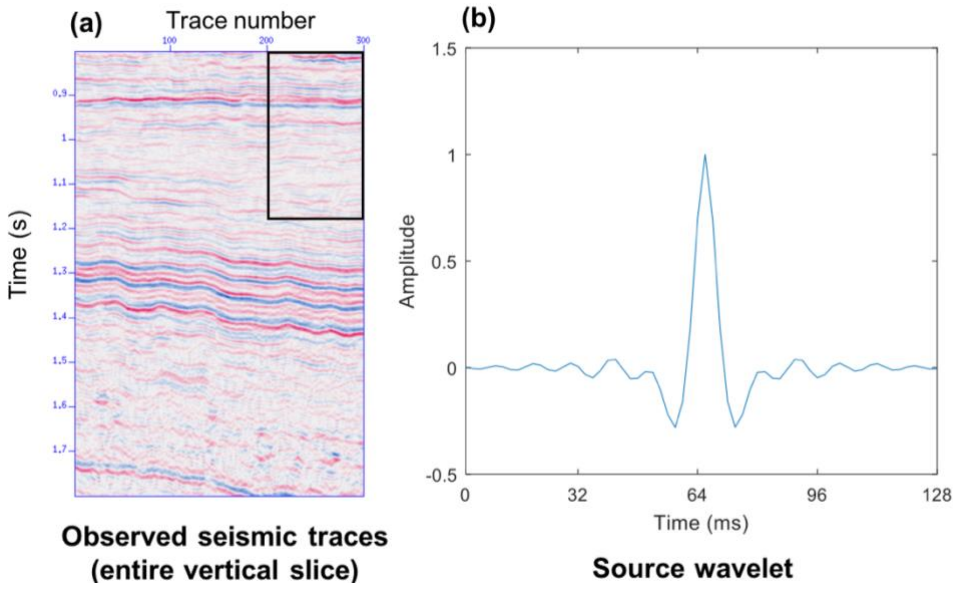


Figure 5.2. (a) A representative vertical slice from a 3D post-stack seismic data from Texas Barnett Shale survey used to evaluate least-squared method vs. neural network method of reflectivity inversion. The black box indicates a zoomed area shown in Figure 5.5. (b) Seismic source wavelet extracted using a statistical method used for reflectivity inversion. (Data courtesy of Devon Energy).

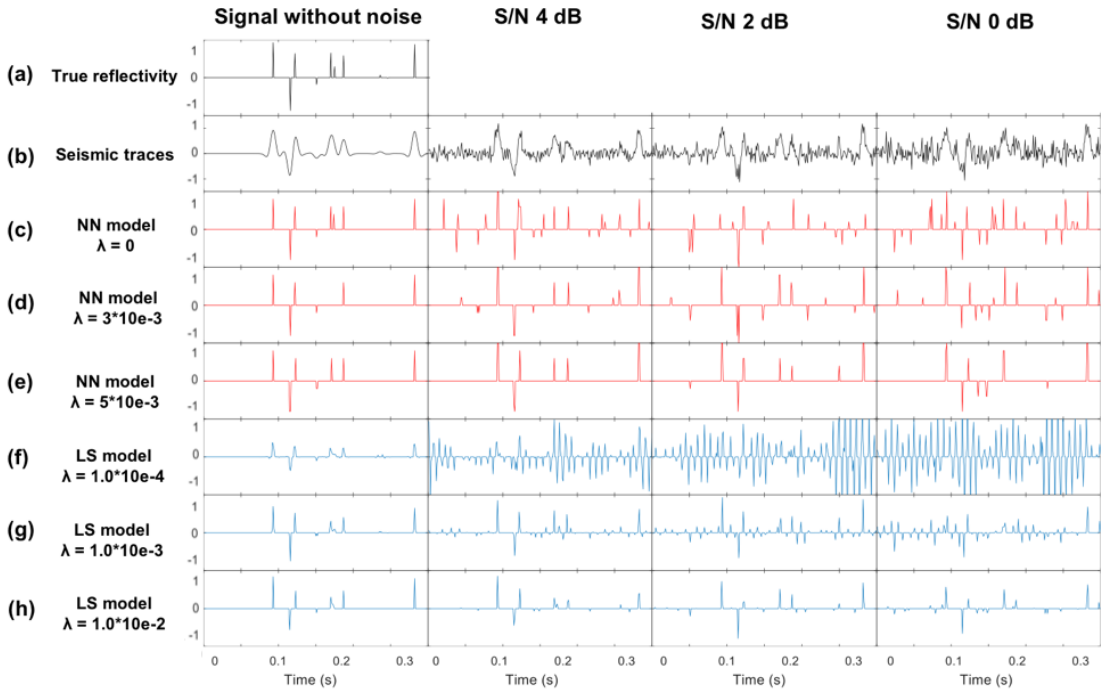


Figure 5.3. (a) True reflectivity, (b) Synthetic seismic traces computed from the true reflectivity using a simple convolutional model. White Gaussian noise is added with different S/N levels: noise-free, 4, 2, and 0 dB. Recovered reflectivity using neural network method with the L1 regularization parameter given by equation 5 of (c) $\lambda = 0$, (d) $\lambda = 3 \cdot 10e\text{-}3$, (e) $\lambda = 5 \cdot 10e\text{-}3$. (f-h) Recovered reflectivity using least-squared method with the L1 regularization parameter (equation 7): (f) $\lambda = 10e\text{-}4$, (g) $\lambda = 10e\text{-}3$, (h) $\lambda = 10e\text{-}2$.

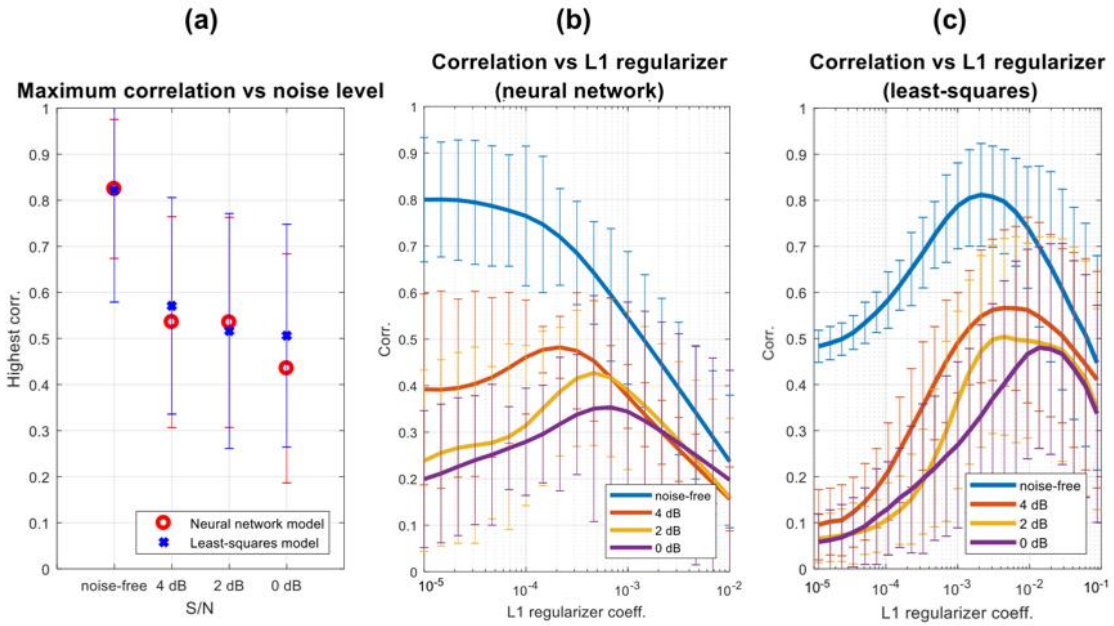


Figure 5.4. Sensitivity of the two algorithms as a function of the noise component of the data. (a) Correlation coefficients between the true reflectivity model and the inverted model at different noise levels. White Gaussian noise is added with S/N level: noise-free, 4 dB, 2 dB and 0 dB. Correlation coefficients for different L1 regularization coefficients at each noise level using the (b) neural network and (c) least-squares methods. We test 10 different models to compute the correlation coefficients where the error bars represent the standard deviation of different models.

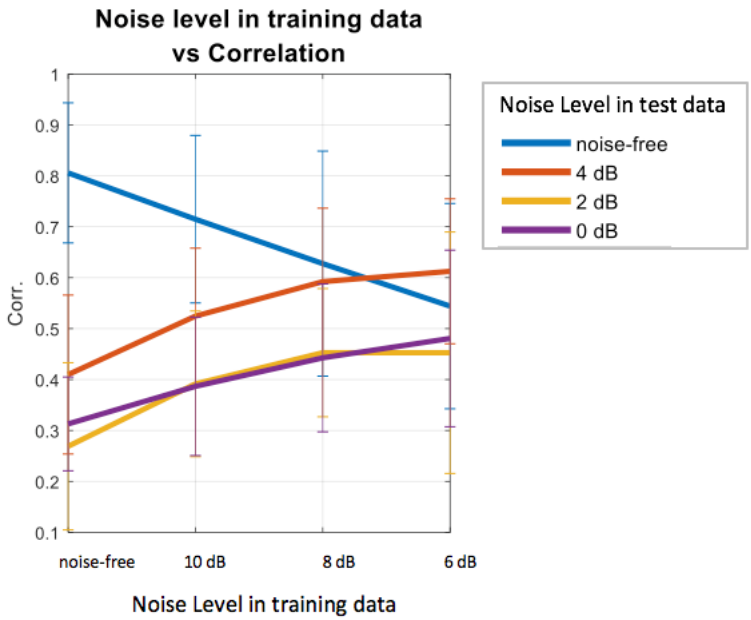


Figure 5.5. Correlation coefficients for different levels of noise in the training data for the NN method. Noise in the training data can alleviate the effect of noise in the test data. In the case of noise-free test data, however, noisy training data reduces the accuracy of the inversion.

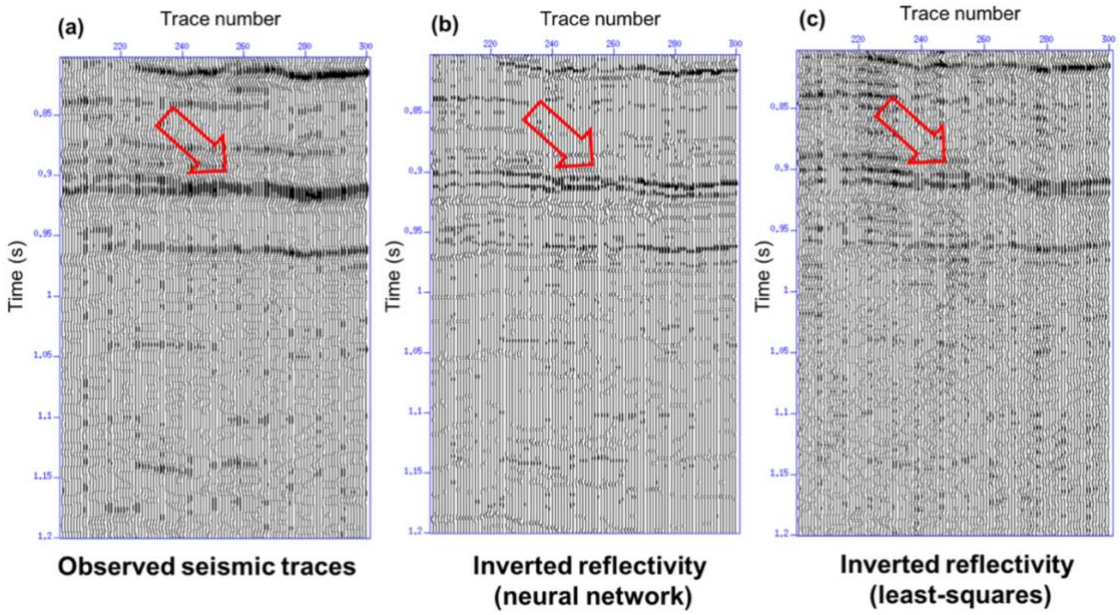


Figure 5.6. (a) Observed seismic traces (input data). Inverted reflectivity model estimated by the (b)neural network and (c) least-squares method. Inverted reflectivity in panels (b) and (c) show higher resolution than that of the seismic traces in (a). Note that the neural network method used in (b) better recovers is the thin bed indicated by the red arrows.

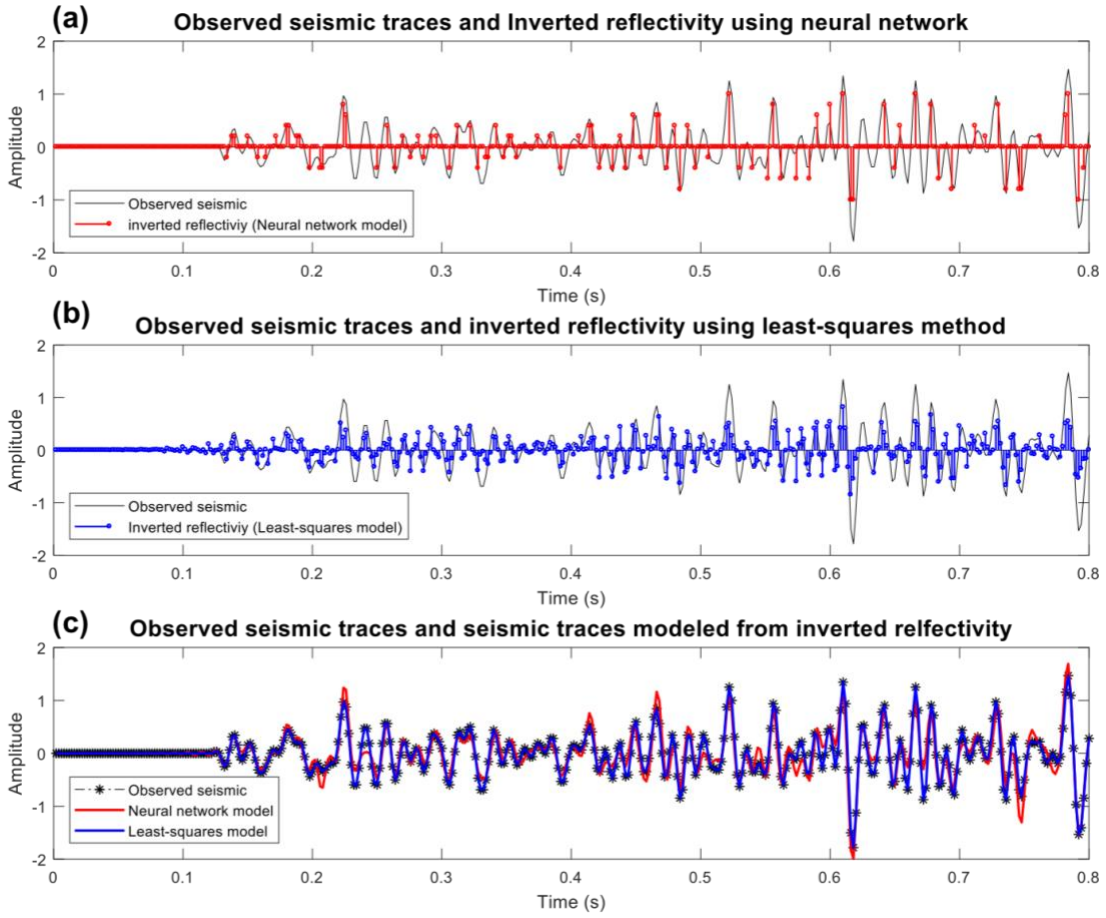


Figure 5.7. Comparison of observed seismic traces and inverted reflectivity using the (a) neural network and (b) least-squares methods. (c) Comparison of the observed seismic traces (indicated by asterisks) using neural networks (in red) and least-squares (in blue).

Inversion methods		Number of processes	Elapsed time (in seconds)
Neural network	Training	1	837 s
	Inversion	1	0.344 s / traces × (367 × 271) traces = 34213 s
Least-squares		1	1.89 s / traces × (367 × 271) traces = 187973 s

Table 5.1. Computation time of reflectivity inversion neural network and least-squares methods.

CHAPTER 6: PRECONDITIONED LEAST-SQUARES MIGRATION TO ADDRESS ARTIFACTS DUE TO CABLE FEATHERING

Introduction

Acquiring marine seismic data is unfavorably affected by weather conditions, currents, or tides which give rise to irregular sail line. This artifact known as cable feathering results in irregular acquisition geometry and fold coverage. Severe cable feathering resulting in high-fold or low-fold areas can give rise to inaccurate amplitude information (Ronen and Liner, 2000). Such amplitude artifacts negatively impact further analysis including not only AVO and impedance inversion, but also simpler geometric attributes such as coherence and curvature. Anomalously high-fold coverage can be often alleviated simply by removing (editing out) the irregular acquisition lines to make the coverage more uniform. Low fold coverage is often addressed by copying or averaging traces of neighboring area in processes called flex binning or interpolation. A more rigorous approach used by some large commercial processing shops is to use each common offset gather to define Delauney triangles. Each trace is then given a weight that is proportional to 1/3 of area of the triangles it defines. Although these methods are commonly used for marine seismic data, such heavy trace editing and weighting takes considerable skill and sophisticated software, with flex binning and interpolation becoming quite challenging if the fold coverage is highly irregular.

In some of the earliest work on least-squares migration (LSM), Nemeth et al. (1999) and Schuster (1997) showed its power in imaging incomplete data. The simplest way to understand this leverage is to think of oversampled data. If we measured three traces rather

than one trace, a naïve migration would result in using amplitudes that are three times too large. Careful weighting would divide each trace by 1/3, thereby correcting this error. In contrast, least-squares migration poses the question "when forward modeled, which subsurface reflectivity best fits the data measured at the earth's surface?". In the case of redundant measurements, we now constrain the misfit three times stronger at this location. Trad et al. (2015) finds that correct amplitudes are obtained by a few iterations of data-fitting. The only disadvantage comes from the higher computational cost.

Preconditioning is key to reducing the computational effort, reducing the number of iterations needed to suppress artifacts caused by missing or oversampled data. Wei and Schuster (2009) used a deblurring filter to generate skeletonized reference model for migration for phase-shift plus interpolation migration. Aoki and Schuster (2009) and Wang et al. (2009) followed up with *f-x* domain prediction filters as a structure-preserving constraint. Applying Kirchhoff migration to irregularly sampled land data, Guo et al. (2014) used a prestack structure-oriented filter as a preconditioner. Since the geology of the east China Sea including the Jeju Basin area is characterized by steeply-dipping igneous complexes and intrusive, we follow Guo et al. (2014) and use an edge-preserving prestack structure oriented filter (SOF) that preserves geology but rejects cross-cutting aliased noise (Zhang et al., 2016).

We apply PLSM to 3D marine seismic data acquired in the Jeju Basin of South Korea, which is severely contaminated by feathering and migration artifacts arising from irregular acquisition geometry. To demonstrate the effectiveness and efficiency of PLSM

to address acquisition footprint, we compare PLSM images with those of conventional Kirchhoff migration using amplitude and geometric attributes.

Geologic Setting

Figure 6.1 shows regional map of the East China Sea and its structural and tectonic elements.where the Jeju Basin forms part of the northern East China shelf basin (Kwon and Boggs., 2002; Xu and Le, 1989; Kirillova, 1993). Since a significant amount of petroleum resources were developed within extensional basins of eastern China (Ren et al., 2002; Cukur et al., 2011), the Jeju Basin has been seen to hold economic potential for exploration and oil productions for many years. Although several studies were carried out in the region, most of previous studies have been carried out with 2D seismic data. The East China shelf basin is a tectonically convergent zone. Basement faults trend mostly NE-SW, parallel to the structural trend of the basin (Cukur et al., 2011). The basin underwent rifting during the Late Cretaceous, resulting in grabens filled by fans and fluviolacustrine deposits. This rapid rifting terminated in Late Eocene with the Yuquan movement and gave rise to an angular unconformity. In the early Miocene postrift stage, regional subsidence occurred and was interrupted by the Longjing movement which resulted in a thrust-fold belt in the eastern part of the area. Afterwards, regional subsidence reoccurred with the area becoming a continental shelf since the late Miocene (Lee et al., 2006; Cukur et al., 2011).

During the Cenozoic, volcanic or igneous activity was pervasive through eastern China (Cukur et al., 2010). Various igneous features characterized by high amplitudes are observed in a vertical slice through the seismic amplitude volume (Figure 6.2a). Seismic

mapping of igneous rock is significant since igneous rocks and low-permeability sills can act as seals and affect fluid migration pathways as well as providing potential reservoirs if the igneous rocks are fractured (Cukur et al., 2010). Unfortunately, due to the irregular sampling, the igneous complexes may be either brighter or dimmer than anticipated in the amplitude volume. Complex channel features are well imaged in the megasequence deposited through early Miocene and late Miocene (Figure 6.2b) where small channel features are best interpreted on the horizon and/or stratal slices through the coherence attribute volume. Unfortunately, linear coherence artifacts due to the irregular acquisition overprint the meandering channels, hampering their interpretation. For this reason, we will apply least-squares migration to recover accurate amplitude information on subsurface volcanic features and the footprint seen on the coherence images.

Data Description

The 3D marine seismic datasets were acquired in three campaigns by the Korean Institute of Geoscience and Mineral Resources (KIGAM). In years 2012 and 2013, the first two surveys were acquired covering 160 km^2 of the Jeju Basin. The third survey was completed in 2014 covering an additional 120 km2 with sail lines perpendicular to those of surveys 1 and 2 (Figure 6.3). The data were acquired using two 2.4 km streamers 100 m apart each with 192 channels at 12.5 m spacing. Two air gun sources were 50 m apart, fired every 50 m in flip-flop mode. Record length is 5 s with sampling increment of 1 ms. The survey parameters summarizing the shot and receiver geometry are described in Table 6.1. The data exhibit a natural bin size of 6.25 × 25 m, coarser in crossline direction. We migrate

the data using a square 12.5 × 12.5 m bin size. Figure 6.4 shows fold of each survey exhibiting the actual geometry. Due to adverse weather conditions and high currents, seismic data acquired during phase 1 suffered from significant feathering. Although average fold number is 24, some acquisition lines overlap, resulting in higher fold while significant gaps occur in other areas. As anticipated, the initial Kirchhoff migration exhibits corresponding amplitude and energy anomalies that correspond to the fold (Figure 6.5).

Preconditioning

The data have broad bandwidth and a good signal-to-noise ratio. Figure 6.6 describes the preconditioning workflow before applying LSM. To eliminate swell noise, a bandpass filter was applied with corner frequencies 8 – 15 – 100 – 120. Then initial velocities were picked on a 1 km by 1 km grid prior to SRME (surface related multiple elimination) to suppress long period multiples over all offsets. Finally, spiking deconvolution was applied to compress the air bubble source wavelet. After preconditioning, the data were imaged with pre-stack Kirchhoff time migration. We migrated the common midpoint gathers at 12.5 m ×12.5 m bin-size using velocity picked on a 500 m by 500 m grid. These initial velocities were updated through reverse moveout and residual velocity analysis.

Implementation of Preconditioned Least-Squares Migration

Since the data suffer from irregular sampling resulting from feathering, and not having access to state-of-the-art migration trace weighting software, we applied a simpler, but computationally more expensive preconditioned least-squares migration (PLSM) to alleviate acquisition footprint. A conjugate gradient method was used to minimize the objective function which is simply the difference between observed data and estimated data from the current (forward modeled, or demigrated) image.

Forward modeling of earth's reflectivity is represented as:

$$\mathbf{d} = \mathbf{Lm},$$

where **d** is the acquired data, **m** is reflectivity model to be determined and **L** is the forward modeling operator. To approximate the earth's true reflectivity **m** in a least-squares manner, we form the objective function

$$\mathbf{P(m)} = \|\mathbf{Lm} - \mathbf{d}\|^2.$$

The model **m** is then estimated using the normal equations:

$$\mathbf{m} = (\mathbf{L}^T\mathbf{L})^{-1}\mathbf{L}^T\mathbf{d}.$$

A preconditioning term can be added to the objective function to improve the convergence rate:

$$\mathbf{P(m)} = \|\mathbf{Lm} - \mathbf{d}\|^2 + \|\mathbf{Cm}\|^2.$$

where **C** is the constraint matrix (Guo, 2012). Conjugate directions are recursively derived using the gradient of the objective function (Lewis et al., 2006). The search direction and step size are recursively determined from the migrated residual and demigrated conjugated gradient of each iteration. We use prestack lower-upper-middle (LUM). structure-oriented

filtering (SOF) as a preconditioner to accelerate the speed of convergence in conjugate method The LUM filter involves two parameters that contribute to smoothing and sharpening of reflection gathers. Our goal that is to remove high amplitude noise (typically aliased artifacts) that cut across reflectors while preserving edges that may have geological meaning. –Three iterations of the conjugate gradient method were applied to achieve convergence, implying that the PLSM costs six times more than our conventional migration. Figure 6.7 shows representative common mid point (CMP) gathers before after PLSM. Note that the demigrated (modeled) reflectos indicated by the red arrow in demigrated data exhibits an enhanced signal-to-noise ratio. As described by Nemeth et al. (1999) the migration-demigration least-squares-migration pair acts much like a Radon transform pair used in more conventional filtering, with reflection events supported by the wave equation passing through the filter and short wavenumber, unaliased events being suppressed. In our implementation, SOF suppresses aliased signal inconsistent with the dominant pattern of the reflectivity.

Data Analysis and Interpretation

Figure 6.8 compares the resulting stacks of the LSM results with simple Kirchhoff time migration. The Kirchhoff migration (and corresponding adjoint modeling or demigration) internal to LSM is the same software used to generate Figure 6.8a. Since the sampling in the cross-line direction is irregular and sparse, the migrated stack in Figure 6.8a exhibits ellipse-shaped migration artifacts (indicated by red arrows). These artifacts repeat on parallel lines, giving rise to the acquisition footprint seen on coherence and

curvature slices. These migration artifacts were reduced after three iterations of PLSM (Figure 6.8b).

Figure 6.9 shows time slices through both amplitude and coherence volumes. The time slice through the coherence volume better exhibits the benefit of PLSM. Coherence measures continuity between two or more windowed seismic traces (Gersztenkorn and Marfurt, 1999). Ideally, the attribute illuminates geological discontinuities such as structural and stratigraphic features; however, the linear anomalies parallel to the sail lines are due to unattenuated migration ellipses, overprints the underlying stratigraphy, hampering further interpretation (red arrows in the left time slice through coherence). After applying LSM, the acquisition footprint in coherence image is alleviated while discontinuities from the channel edges are preserved.

High amplitude anomalies arising from feathering are seen in the amplitude time slice in Figure 6.10. These amplitude artifacts correspond to the high and low areas seen on the fold map. The feathering lines which indicated by yellow arrow are reduced after three iterations of PLSM.

Conclusions

Cable feathering due to strong currents and adverse weather results in severe cable feathering and irregular fold that is exhibited as amplitude and discontinuity anomalies. By asking the question "which reflectivity when forward modeled best represents the measured surface data?" preconditioned least-squares migration reduces these migration artifacts using redundant data as locations that are more strongly constrained and missing data as areas that are only weakly constrained. The application of prestack structure-

oriented filtering suppresses coherent, aliased artifacts that cut across the reflectors of interest. At each iteration, the residuals corresponding to these aliases has an opportunity to be mapped to the correct location, thereby reducing their impact with each iteration. The result is an amplitude section that preserves small scale geological discontinuities such as faults and meandering channel edges and provides amplitudes that honor the measured data. While simpler to write than complex data weighting algorithms commonly used in prestack Kirchhoff migration, the main disadvantage of PLSM is cost, with each iteration of migration and demigration taking twice the time of migration alone. We find that acquisition footprint suppression occurs with three iterations, resulting in an increase in cost by a factor of six.

Acknowledgments

The authors thank the Korea Institute of Geoscience and Mineral Resources (KIGAM) for data, encouragement, guidance, and financial support. Most of the software was developed as part of the OU Attribute Assisted Seismic Processing and Interpretation (AASPI) consortium.

References

Aoki, N., and G. T. Schuster, 2009, Fast least-squares migration with a deblurring filter: Geophysics, 74, WCA83-WCA93.

Cabrales-Vargas, A., and K. J. Marfurt, 2013, Amplitude-preserving imaging of aliased data using preconditioned Kirchhoff least-squares depth migration: 83th Annual International Meeting, SEG, Expanded Abstracts, 3726 – 3730.

Cukur, D., S. Horozal, D. C. Kim and H. C. Han, 2011, Seismic stratigraphy and structural analysis of the northern East China Sea Shelf Basin interpreted from multi-channel seismic reflection data and cross-section restoration: Marine and Petroleum Geology, 28, 1003-1022.

Cukur, D., S. Horozal, G. H. Lee, D. C. Kim, and H. C. Han, 2012, Timing of trap formation and petroleum generation in the northern East China Sea Shelf Basin: Marine and Petroleum Geology, 36, 154-163.

Cukur, D., S. Horozal, G. H. Lee, D. C. Kim, H. C. Han and M. H. Kang, 2010, The distribution and characteristics of the igneous complexes in the northern East China Sea Shelf Basin and their implications for hydrocarbon potential. Marine Geophysical Researches, 31, 299-313.

Dai, W., and J. Schuster, 2009, Least-squares migration of simultaneous sources data with a deblurring filter: 79th Annual International Meeting, SEG, Expanded Abstracts, 2990-2994

Gersztenkorn, A. and K. J. Marfurt, 1999, Eigenstructure-based coherence computations as an aid to 3-D structural and stratigraphic mapping. Geophysics, 64, 1468-1479.

Guo, S., 2014, Seismic solutions for unconventional resource plays: Ph.D. thesis, University of Oklahoma.

Guo, S., B. Zhang, K. J. Marfurt, and A. Cabrales-Vargas, 2012, Noise suppression using preconditioned least-squares prestack time migration: Application to the Mississippian Limestone: 82th Annual International Meeting, SEG, Expanded Abstracts, 1-6.

Kirillova, G.L., 1993, Types of Cenozoic sedimentary basins of the East Asia and Pacific Ocean junction area: Palaeogeogr., Palaeoclimatol., Palaeoecol. 105, 17–32.

Kwon, Y. I. and S. Boggs, 2002, Provenance interpretation of Tertiary sandstones from the Cheju Basin (NE East China Sea): a comparison of conventional petrographic and scanning cathodoluminescence techniques, Sedimentary Geology 152, 29-43.

Lee, G. H., B. Kim, K. S. Shin and D. Sunwoo, 2006, Geologic evolution and aspects of the petroleum geology of the northern East China Sea shelf basin: AAPG Bulletin, 90, 237-260.

Lewis, J. M., S. Lakshmivarahan, and S. Dhall, 2006, Dynamic data assimilation: a least squares approach: Cambridge University Press.

Nemeth, T., Sun, H., and G. T. Schuster, 1997, Separation of signal and coherent noise by migration filtering: Geophysics, 65, 574–583.

Nemeth, T., C. Wu, and G. T. Schuster, 1999, Least-squares migration of incomplete reflection data: Geophysics, 64, 208–221.

Wang, J., and M. D. Sacchi, 2009, Structure constrained least-squares migration. 79th Annual International Meeting, SEG, Expanded Abstracts, 2763-2767.

Ronen, S., and C. L. Liner, 2000, Least-squares DMO and migration: Geophysics, 65, 1364-1371.

Schuster, G. T.,1997, Acquisition footprint removal by least-squares migration. 1997 Annual UTAM Report, 73-99.

Trad, D. 2015, Least-squares Kirchhoff depth migration: implementation, challenges, and opportunities. 85th Annual International Meeting, SEG, Expanded Abstracts, 4238-4242.

Xu, W., Le, J., 1989, Structural history of the East China Sea: China Earth Sci. 1, 59– 73.

Zhang, B., T, Lin, S. Guo, O. E. Davogustto, and K. J. Marfurt, 2016, Noise suppression of time-migrated gathers using prestack structure-oriented filtering: Interpretation, 4, SG19-SG-29.

Figures

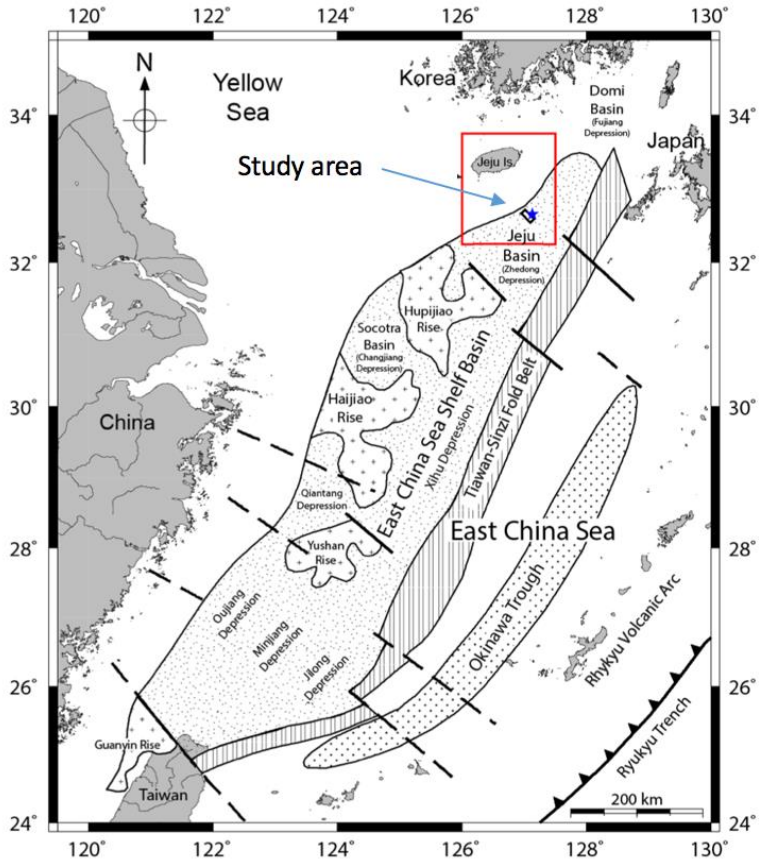


Figure 6.1. Regional map of the East China Sea showing major structural and tectonic elements (Modified after Zhou et al., 1989; Lee et al., 2006, Curkur et al., 2001).

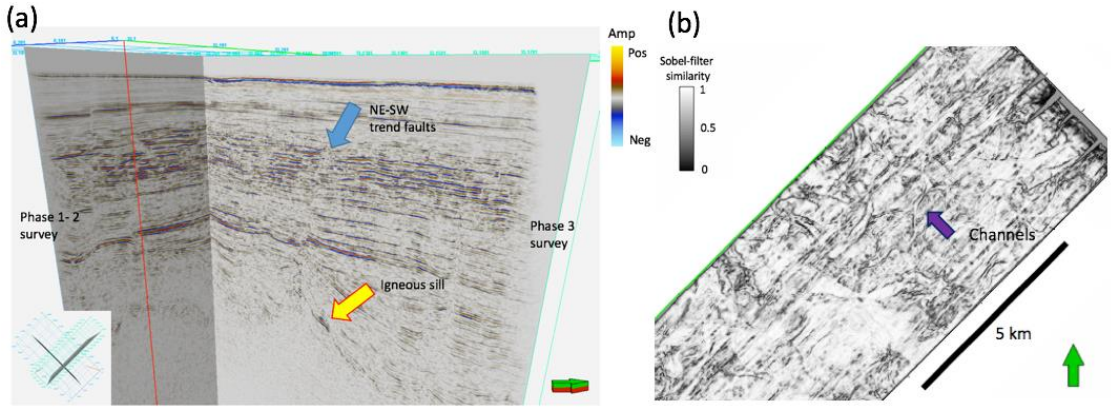


Figure 6.2. (a) Representative image through the prestack time migrated, stacked data volume and (b) a time slice through the corresponding coherence attribute at *t*=710 ms exhibiting strong inline (NE-SW) artifacts due to cable feathering superposed on the underlying geology.

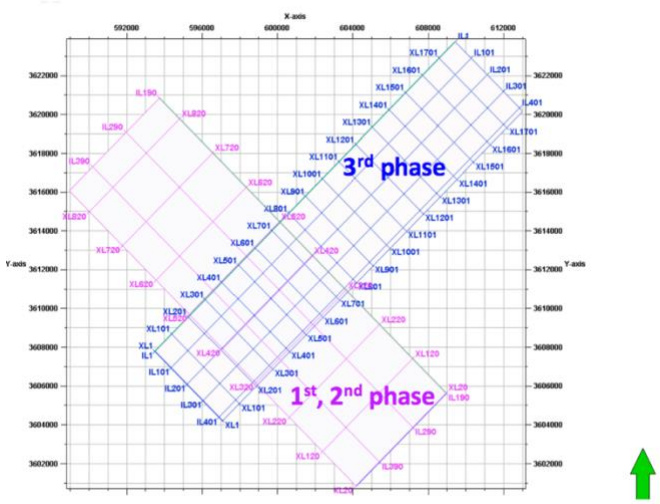


Figure 6.3. Map showing the three phases of the seismic acquisition program.

Acquisition Parameter	
Record length	5 s
Sample interval	1.0 ms
Shots	25 m interval 5 m below sea surface
Receivers	2 streamer cables, 2400 m - length 12.5 m interval, 384 channels 7 m below sea surface
Natural bin size	6.25 m (in-line) × 25 m (cross-line)
Sail direction	135° / 315° (Phase 1, 2) 45° / 225° (Phase 3)

Table 6.1. Summary of marine acquisition parameters.

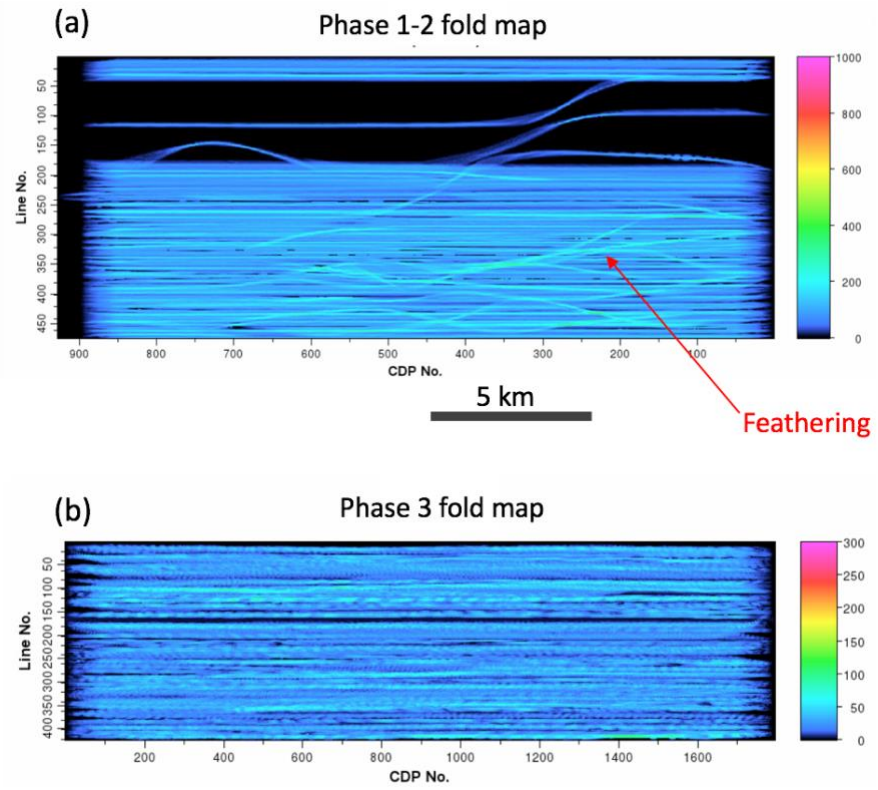


Figure 6.4. Fold maps of (a) phases 1 and 2 and (b) phase 3 mapped to the natural 6.25 by 25 m bins. The red arrow indicates anomalously high fold due to feathering arising from adverse weather condition during acquisition. Black areas represent zero fold. The irregular fold gives rise to acquisition footprint.

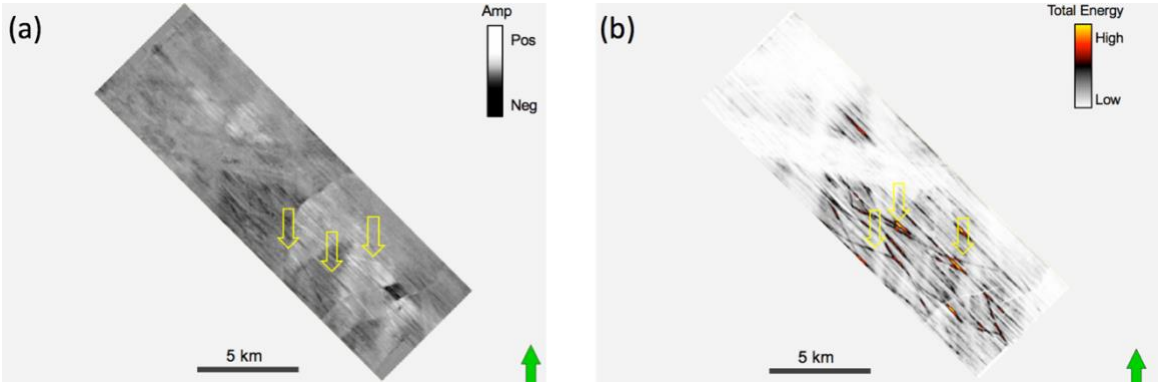


Figure 6.5. Time slices at *t*=400 ms through (a) migrated amplitude and (b) total energy computed in a 3-trace by 3-trace by 20 ms window. Yellow arrows indicate high amplitude and energy arising from overlapped acquisition due to feathering. Low and high amplitudes in the inline (NW-SE) direction correlate to less erratic low and high fold areas.

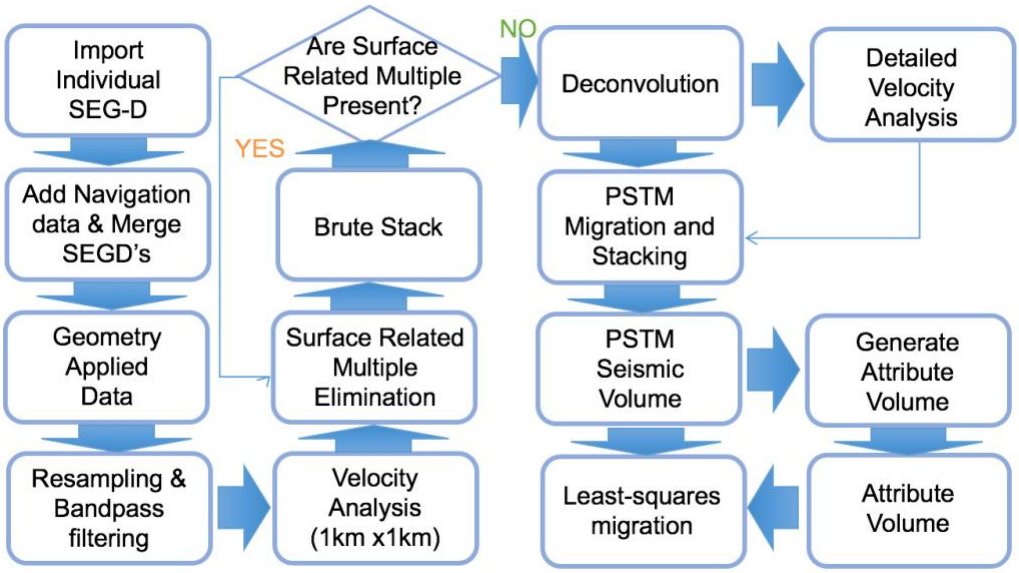


Figure 6.6. Processing workflow applied to the 3D marine seismic data.

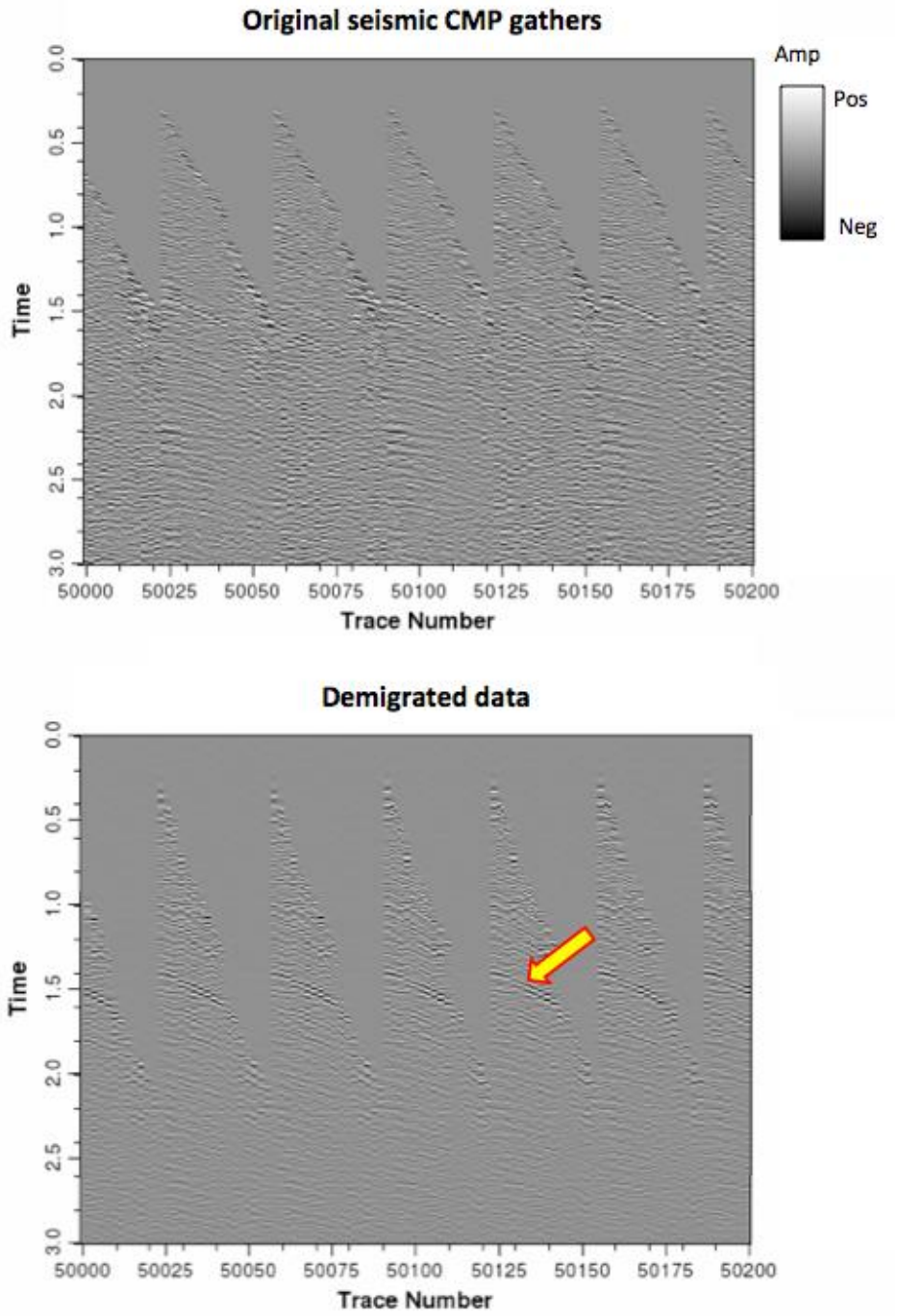


Figure 6.7. CMP gathers of (a) original seismic data and (b) demigrated data. Red arrow indicates reflectors exhibiting higher signal to noise ratio after applying demigration.

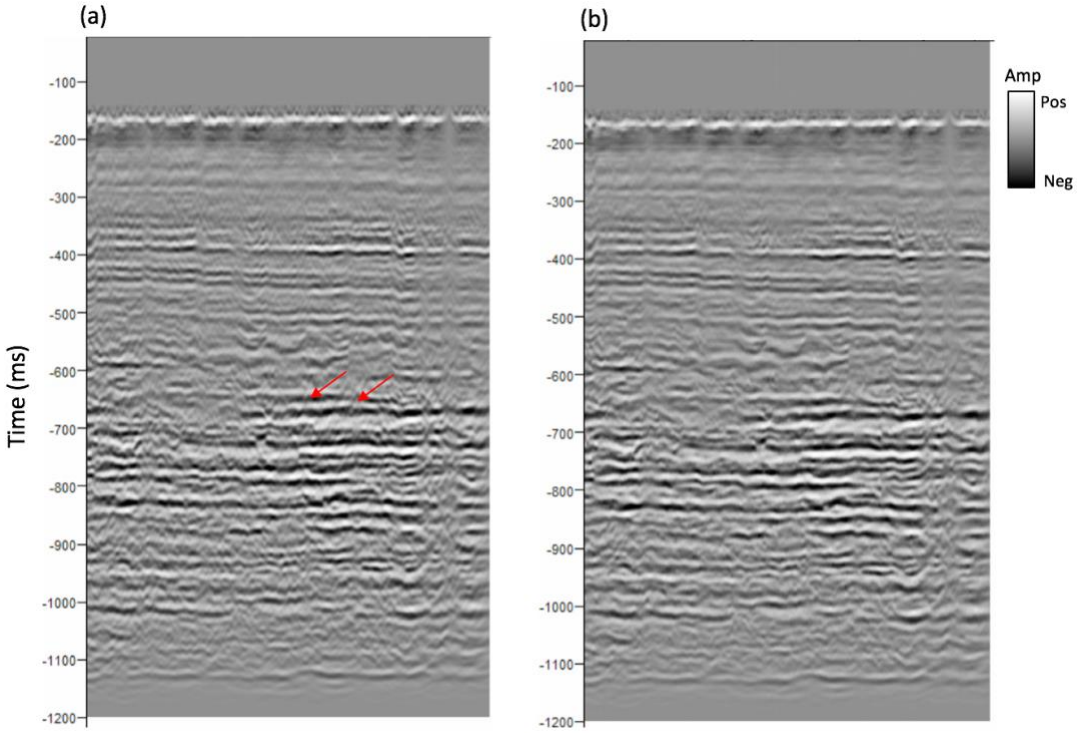


Figure 6.8. Migration stack (cross-line direction) after (a) conventional Kirchhoff migration and (b) PLSM. After three iterations PLSM, many (but not all) of the ellipse-shaped migration artifacts are suppressed.

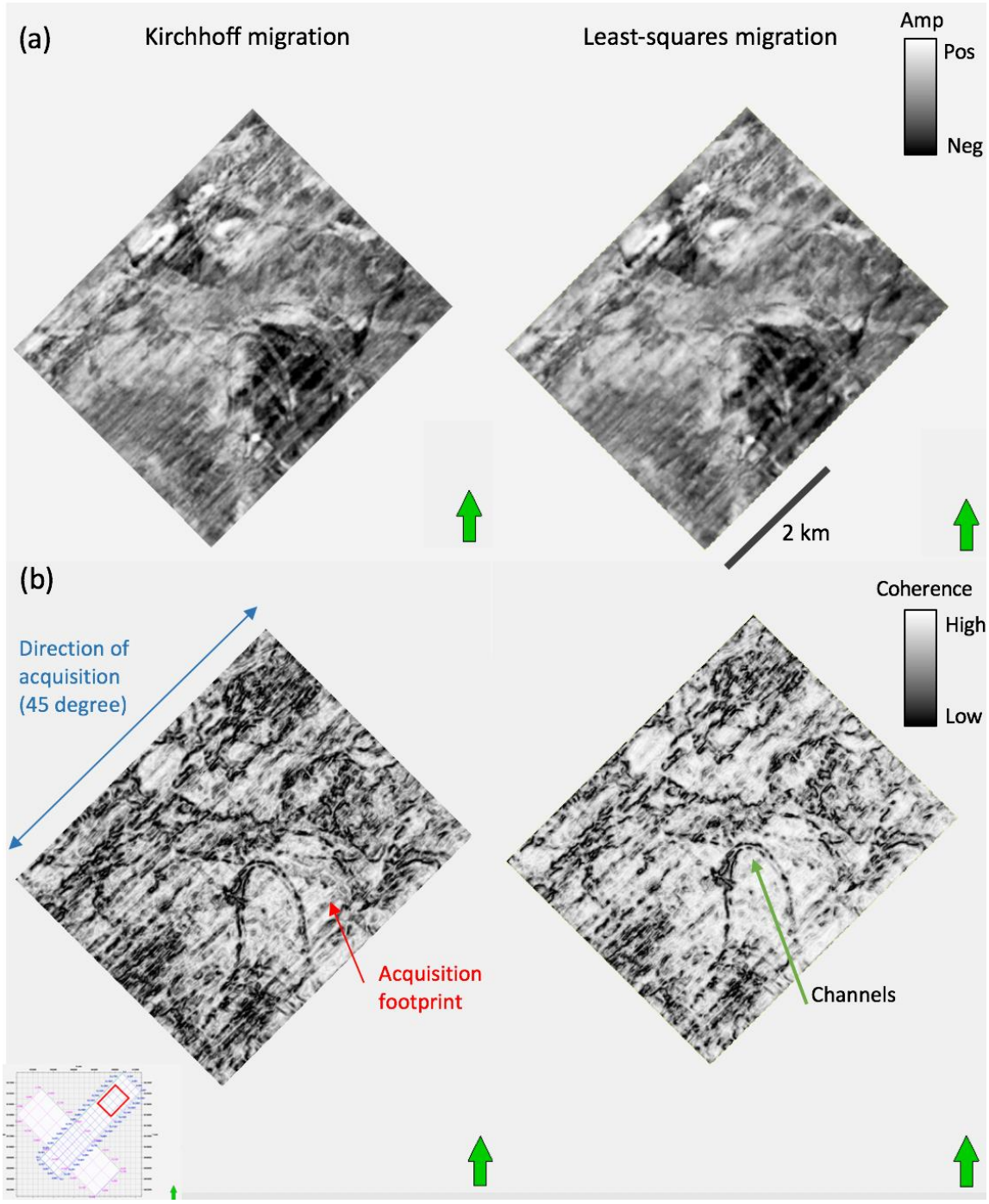


Figure 6.9. Time slices at *t*=740 ms through (a) amplitude and (b) coherence, (left) after conventional Kirchhoff migration, and (right) after applying PLSM. The acquisition footprint indicated by the red arrow is reduced after PLSM while the channel features indicated by green arrow is preserved.

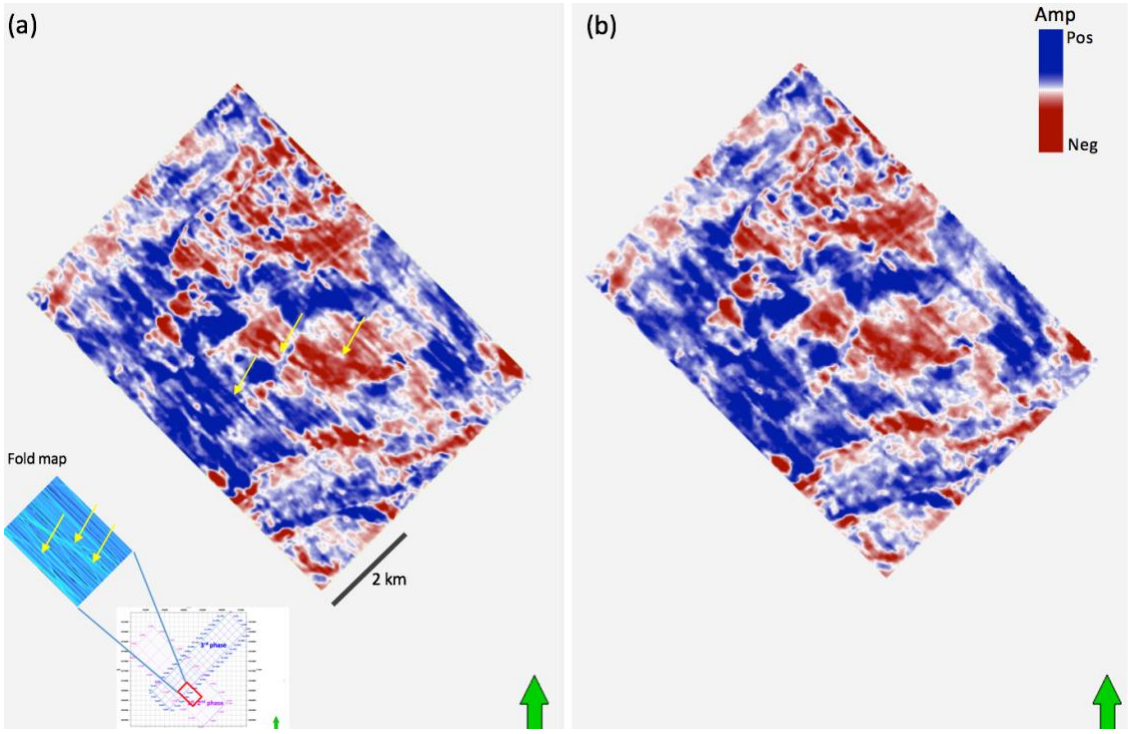


Figure 6.10. Time slices at *t*=720 ms through amplitude (a) after conventional Kirchhoff migration, and (b) after applying PLSM. Yellow arrows indicate high amplitudes arising from overlapped acquisition due to feathering.

CHAPTER 7: CONCLUSIONS

In this dissertation I explored machine learning approaches to solve problems in seismic interpretation and seismic data processing. Even though machine learning algorithms have been developed to complete quite complicated tasks ranging from self-driving cars to stock market training, there are many limitations including insufficient, excessive, and unanticipated data, as well as the misinterpretation of results. To build a reasonable learning model, both the geophysical and geological processes as well as machine learning algorithms should be understood.

For the seismic facies classification task, in Chapter 2 I analyzed all attributes together using a multivariate framework which can quantitatively rank the attributes to build an optimum subset. The resulting attribute subset maintains high accuracy to differentiate salt and MTD facies from conformal reflectors while also reducing the dimensionality of the data. A limitation of the classification is that some of the MTD facies are misclassified as salt because both facies are highly discontinuous and have low coherence. To address this shortcoming, in Chapter 3, I found that facies classification based on a semi-supervised learning approach enhances the accuracy of the classification of seismic facies especially when the number of labeled training data is insufficient. Using a CGAN architecture of a deep learning method, overfitting to training data is avoided.

In seismic data processing, I reviewed deep learning-based noise attenuation in and reflectivity inversion. In Chapter 4, I found that the architecture of the deep residual neural network is effective in attenuating random as well as coherent noise such as aliased migration artifacts. A limitation is that the noise is not eliminated entirely when the

amplitude of the noise is significantly higher than the signal in the frequency – space domain. In Chapter 5, I examined conventional geophysical inversion and compared the results to machine learning as a methodology to solve an inverse problem. The appropriate value of the regularization term in the least-squares method helps to recover true reflectivity for both noise-free and noise contaminated data. In least-squares and neural network methods, the choice of the regularization term yields a sparse spike solution which has advantages in inverting noisy data.

The recent breakthrough of machine learning technology enabled me to make use machine learning tools to address a wide variety of applications for seismic processing and interpretation. However, machine learning has limitations and is far from replacing the human interpreter or model-based problem solver. In the seismic processing area, the training dataset should be large enough to apply the training data to test problems. Consequently, the training process requires a great amount of time and computational power when the training dataset is large, even if the test process is not computationally intensive. For seismic interpretation, machine learning does not presently have access to the knowledge of geologic processes used by skilled interpreters, or the laws of physics and principals of signal analysis used by skilled seismic processors. To make the most use of machine learning as a tool, the capabilities and limitations of machine learning models, data, and applications need to be clearly understood by the geoscientists who use them and management that base their decisions on their results. I hope that this dissertation moves our community further towards that understanding.