The Dissertation Committee for Son Dang Thai Phan certifies
that this is the approved version of the following dissertation:

Machine Learning Algorithms for Solving Some Seismic Inversion Challenges

Committee:

Mrinal Sen, Supervisor

Kyle Spikes

Sergey Fomel

Charles Jackson

Douglas Foster

Machine Learning Algorithms
For Solving
Some Seismic Inversion Challenges

by

Son Dang Thai Phan

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

The University of Texas at Austin

May 2021

Dedicated to my wife, ***Thuy Tinh Tu Ngo****. Her hours of work in caring for our child,* ***Nhi****, enabled the hours of research, brainstorming, and writing to complete this study. She is my invaluable treasure, worth more than any jewels.*

Acknowledgements

I would like to sincerely thank my supervisor, Dr. Mrinal K. Sen for his guidance, encouragement and discussions throughout my studies. I would like to thank Dr. Kyle Spikes, Dr. Sergey Fomel, Dr. Charles Jackson, and Dr. Douglas Foster for serving as examining and dissertation committee members, and for their suggestions to improve the studies and this dissertation.

I would like to thank Dr. Brian Horton, the Graduate Adviser, for his inputs to all my inquiries about the program.

I would like to extend my gratitude to my colleagues at the Institute for Geophysics at The University of Texas at Austin: Nam Pham, Debanjan Datta, Badr Alulaiw, Ram Tuvi, Xiaohui Cai, Janaki Vamaraju, Reetam Biswas, Hala Alqatari, Arnab Dhara, Xin Liu, Ricardo de Braganca, Dimitri Voytan, and Jackson Tomski.

I also thank Philip Guerrero for his magnificent supports in all aspect of student life, Mark Wiederspahn for his wonderful technical support, Susanne Morrison, Rosaline Gamble, Lulu Curry for their assistances in all paperwork process, and Jennifer Jordan for her guidance to my internship experiences.

I would like to acknowledge the financial supports from the EDGER forum for my graduate studies.

This dissertation would not have been completed without the love and support of my parents and family.

Abstract

Machine Learning Algorithms for Solving Some Seismic Inversion Challenges

Son Dang Thai Phan, Ph.D.

The University of Texas at Austin, 2021

Supervisor: Mrinal K. Sen

Seismic inversion is a popular quantitative approach to extract some of the subsurface properties from seismic amplitudes by utilizing the physics of the wave propagation through earth layers, in forms of the empirical formulations that simulate the energy distribution phenomenon. Typical inversions are performed with limited offset angles, and in a small time-window, within which the wavelet is assumed to be stationary, and the property contrasts are small across the boundaries. These exposes the process to several major problems: (1) the limited resolution due to wavelet effects, (2) the dependence on some rock physics models when inverting for petrophysical properties, and (3) the resolution discrepancy between the time domain seismic signal and well logs in depth domain. The primary goal of this research is to use machine learning to solve these challenges, by designing and applying proper neural network structures and suitable training schematics. Firstly, a single-layer Boltzmann machine is implemented as an unsupervised learning algorithm to predict the elastic properties at higher resolution than that can be achieved by conventional approaches, while still retaining the physical relationship between the seismic amplitudes and reflectivity series. The high-resolution results are produced from the accurate post-inversion reflectivity series, which is not bounded by the wavelet effects, and novel model update schemes. Secondly, a new multimodal Cross-shape deep Boltzmann machine is designed to

simultaneously capture six possible relationships between four different input training data to invert for petrophysical properties from the pre-stack seismic amplitudes in datasets with limited well coverages. This algorithm has a significant advantage in avoiding the uncertainties associated with the data fitting algorithms to create the rock physics models to guide the solution. Last but not the least, a novel multimodal deep learning network is applied to predict the posterior distribution of the subsurface elastic properties from a seismic gather, to resolve the resolution discrepancy challenge, by a smart preparation of the training label in the form of time dependent probability distributions. The biggest advantage of this algorithm is the avoidance of the heuristic calculation of the partition function, which is required to calculate the posterior distributions of common neural network outputs. While the first algorithm requires an input wavelet to constraint the results, the other two algorithms do not, which make them appropriate for inversion in depth domain, or with nonstationary signals.

Table of Contents

List of Figures

Chapter 1 : Introduction

1.1. Motivation

Recorded seismograms are important sources of information to understand subsurface properties. On a seismic section with normal polarization, the positive amplitudes indicate the transition boundaries from softer to stiffer materials, and vice versa for the negative amplitudes. Seismic inversion (e.g Sen, 2006) is a popular quantitative approach to extract the subsurface elastic properties from seismic amplitudes by utilizing the physics of the wave propagation through earth layers. Various studies have been carried out to improve the overall quality of inversion results.

Typical seismic inversions, assuming small property contrasts across the boundaries, are performed on a small time-window, within which the wavelet is assumed to be stationary, and limited offset angles up to 35 degrees (Ikelle and Amundsen, 2018) These work best in fields where hydrocarbon signatures are recognizable on seismic sections. However, they may not be valid when the background geology becomes complicated with significant property contrasts, or in fields in which reservoir zones occupy large time intervals such that waveform changes are unavoidable due to energy dissipation and absorption. To improve the quality of inversion results in such conditions, non-stationary wavelets and large incident angle reflections are required. Despite countless efforts (e.g., Lines and Treitel, 1985; Fomel, 2007; Dai et al., 2016, Wang and Morozov, 2020), the estimation of source wavelets remains a difficult task. An alternative approach that does not directly require input wavelets is more desirable.

Artificial neural networks are a relatively new technology that has been successfully developed and applied in many areas of science and engineering, with noticeable applications ranging from the classical speech and text recognitions to the current live data prediction and optimization in many engineering and financial problems.

In seismic data analysis, the learning and adaptive capabilities of the ANN are valuable assets in performing automation tasks such as horizon picking, velocity analysis etc. (e.g., Murat et al., 1992, McCormack et al., 1993, Fish et al 1994, Biswas et al., 2018), or inversion of seismic data (Roth et al. 1994, Calderon-Marcias et al., 1993, Phan and Sen, 2019, Biswas and Sen, 2019). Recent applications in the advanced interpretation process include pattern recognition for structural features (Saggaf et al 1998, Diersen et al 2011, Brekovic et al 2017).

A neural network is built on a large system of interconnected units called the neurons. The interaction between neural nodes, once information is passed through the system, decides the learning pattern from the input information. The training process can be either (1) supervised, where known input data and corresponding labels are used to guide the network structure toward a desired relationship, or (2) unsupervised where the algorithm works on its own to extract hidden information from the unlabeled input data. Depending on the complexity of the objective feature, a network may have no or multiple hidden layers. A deep learning network system includes multiple layers of neurons to capture hidden features or properties from the data (e.g, Goodfellow et al., 2016). The extracted information is contained within the weight terms of network structures such as perceptron, multi-layer feedforward network, or recent deep learning methods with different structures involving the convolution neural networks (CNN-Lecun et al., 1995); or the responses of the neurons, such as Hopfield Network (Hopfield et al, 1985), the self-organizing maps (Kohonen, 1988), or the generative adversarial network (GAN-Goodfellow et al., 2016).

Although machine learning has been evolving as a promising technical advancement to seismology, a full understanding of all available algorithms becomes an impossible task. One may find it difficult to choose from a large number of different

methods that could do the same task, with almost similar outcomes. The question of which network structure to use for a particular problem has been a challenging topic for any researcher interested in applying machine learning to their studies.

The primary goal of my research is to use machine learning to solve some outstanding problems in seismic inversion: (1) the wavelet effects limiting the resolution of recorded seismic data, (2) the source signal changing shape and amplitude when traveling through the earth, (3) the dependence on data fitting algorithms to create some rock physics models to guide the solutions, and (4) the resolution discrepancy between seismic signals and well log data when combined for modeling. I hypothesize that with proper network structures and suitable training schematics, the machine learning algorithms would be able to mitigate these challenges, to a great extent.

In this chapter, I will review the common seismic inversion approaches and their limitations, the general neural network structures, and how to determine the efficiency of network training. I then outline my research objectives, and describe the organization of the rest of my dissertation chapters.

1.2. Common Seismic Inversion Approaches and Their Limitations

The inversion problem for quantitative seismic interpretation comprises (1) a forward problem (Equation 1.1), which assumes a convolution model[1] between the source signal **w** (also represented by the Toeplitz matrix $\boldsymbol{W}$) and a reflectivity series ($\boldsymbol{r}$) that represents the changes in rock properties, (2) an objective function that compares the difference between the recorded and synthetic signals, and (3) a model update scheme. The source wavelet is often difficult to measure for its non-stationary property: its shape and intensity change due to the high frequency attenuation when traveling through the

[1] A convolution model is a highly simplified approximation of wave propagation – full waveform inversion uses a more rigorous numerical solver of the wave equation.

subsurface (Yilmaz, 2001). Meanwhile, the reflectivity series is typically calculated using a full Zoeppritz equation (Zoeppritz, 1919), or its linearized version such as the Aki-Richards (1980) approximations. Such calculations are validated for reflected signals whose incident angles are up to the critical threshold (Wang, 1999).

$$\boldsymbol{s} = \boldsymbol{w}(*)\boldsymbol{r} = \boldsymbol{W}\boldsymbol{r}, \tag{1.1}$$

For a conventional pre-stack inversion approach, low frequency starting models are required at the beginning of the inversion process. This requires building property models from the horizons interpreted on seismic sections, which come with large uncertainties, especially in areas with complex geological settings. Common seismic inversion methods also suffer from the limited bandwidth of the seismic data, where the low frequency (0-10 Hz) and the high frequency (above 50 Hz) contents are missing. The lack of high frequency content limits the resolution of fine features, while the lack of low frequency content results in unreliable trends in the results. The methods based on statistical approaches such as that described by Srivastava and Sen (2010) have been applied successfully to broaden the high frequency bandwidth to improve the seismic resolution. Meanwhile, the common approach to recover the low frequency content is to use low frequency starting models to guide the solution (e.g., Sirgue, 2006; Innanen et al., 2011). The models are created by populating well log values throughout a 3D volume using some interpolation and extrapolation techniques. This process requires a geological framework built on the interpreted surfaces from a given seismic dataset. In areas with complex geological settings, the interpretation may have large uncertainties, which transform into uncertainties in the inversion results. To reduce the dependence on the guiding model, better approaches that are independent of human inputs are preferred.

1.3. Neural Networks

A neural network is a system of interconnected units that simulates the neurons within human brains. The neurons are connected via communication links (weights), through which an activation signal passes from one to the other once an incoming signal is received. Each neuron can only send one signal at a time. The visible neuron layer receives data, while hidden layer(s) serve(s) to extract hidden properties within input data. Depending on the complexity of the problem, a network may have one or multiple hidden layers.

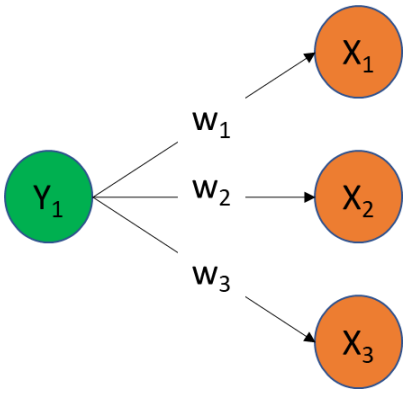

Figure 1.1: A simple neural network structure with 3 neurons X_1, X_2 and X_3 receiving information from a neuron Y_1 via communication links W_1, W_2 and W_3.

A neural network can be characterized by its architecture (the number of neurons and how they are connected to each other), the method of determining the connection weights (via training or optimization), and the output signal via activation function (Fausett, 2002). The training process can be either (1) supervised, where known input data and corresponding labels are used to guide the network structure toward a desired relationship; or (2) unsupervised where the algorithm works on its own to extract hidden information from the unlabeled input data. For an example, in Figure 1.1, a supervised training will have labeled data fed into both neurons **Y** and **X**'s so that the extracted weight

terms w's represent possible relationship between **Y** and **X**. Meanwhile, an unsupervised training will have data fed into only neuron **Y**, so that the internal mechanism rearranges and optimizes the outputs of neurons **X**'s and weights w's to represent some features of input signal.

Let $\boldsymbol{X} = [X_1, X_2 \dots X_N]^T$ and $\boldsymbol{Y} = [Y_1, Y_2 \dots Y_M]^T$ be the neuron values of layers **X** and **Y**, and $\boldsymbol{W} = \{W_{ij}\}$ be a weighting matrix, where w_{ij} is the weighting term connecting X_i to Y_j, and $\boldsymbol{b} = [b_1, b_2 \dots b_M]^T$ be the threshold term, which determines how sensitivity the neuron is to the incoming signal. The data is fed into the system at layer **Y**. The learning process is to find the best parameters that minimize an energy function at Equation 1.2. The incoming signal is translated at each individual neuron in layer **X** via connection weight, with an amplitude in Equation 1.3. If the signal amplitude surpasses the threshold terms, the neurons emit feedbacks whose amplitudes are determined by an activation function θ in Equation 1.4. The activation function varies from the simplest form of a dipole response with either 0 or 1, or -1 or 1; to more complicated forms such as the logistic or step functions, depending on the desired output. Returning energy toward layer **Y** are used to update the weighting terms, and in some cases, the threshold term, if not preset, following a backpropagation of energy differentiation in Equation 1.5. The learning terms $\alpha_{\boldsymbol{W}}, \alpha_{\boldsymbol{b}}$ are used to control the updating process. The traveling back and forth of data between neuron layers stops when the minimum energy stage is reached, at which the network is trained and could be used for further analyses.

$$E(\boldsymbol{W}, \boldsymbol{b}, \boldsymbol{X}, \boldsymbol{Y}) = -\boldsymbol{Y}^{\boldsymbol{T}}\boldsymbol{W}\boldsymbol{X} + \boldsymbol{b}^{\boldsymbol{T}}\boldsymbol{X}, \tag{1.2}$$

$$\varphi(\boldsymbol{Y}, \boldsymbol{W}, \boldsymbol{b}) = -\frac{\partial E}{\partial X} = \boldsymbol{W}^{\boldsymbol{T}}\boldsymbol{Y} - \boldsymbol{b}, \tag{1.3}$$

$$\boldsymbol{X} = \theta(\boldsymbol{\varphi}), \tag{1.4}$$

$$\Delta \boldsymbol{W} = -\alpha_{\boldsymbol{W}} \frac{\partial \boldsymbol{E}}{\partial \boldsymbol{W}};\ \Delta \boldsymbol{b} = -\alpha_{\boldsymbol{b}} \frac{\partial \boldsymbol{E}}{\partial \boldsymbol{b}}. \tag{1.5}$$

1.4. Neural Network Training: Determination of Efficiency

The network accuracy is measured by how close the energy function (or loss function) is toward the global minimum value, while its generalization is determined by how well it is able to predict the independent dataset that is not used in the training process. A properly trained neural network should be able to generalize the expected property and accurately predict results. A network is said to be underfitted if it is trapped at a local minimum during the training process. A network which has high prediction capability, but poor generalization is considered an overfitted one.

While an underfitted network can be optimized by using different optimization algorithms, or providing more data, an overfitted network requires careful analysis. A common approach to deal with such a challenge is to split the data into two parts used for two purposes: training and testing (Reitermanova, 2010). The system is first trained with training dataset, and then cross-validated (Picard et al., 1984) with the testing dataset. The matured network will ensure that the energy functions of both training and testing datasets reach stable solutions when reaching a global minimum stage (Figure 1.2). An overfitted network may have a stable loss curve of the training data, and an overturned test loss curve.

Underfitting may be either due to getting trapped at a local minimum solution, or not reaching the globally optimal solution, given the number of epochs and data. The solution is to provide more training data to the network, or using a better optimization technique (e.g, Sen and Stoffa, 2013). Meanwhile, for an overfitting case, starting of a gradual increase in the test data's loss curve indicates that the network has begun to lose its generalization. In such a situation, either providing more data, or modifying the hyperparameters may largely influence the performance.

Overfitting is a common problem in deep learning. Several data manipulation techniques have been developed to deal with such a challenge. Pooling is a technique to create feature maps that summarize all features in the input data. The common ones are MaxPooling (getting the maximum value), or AveragePooling (taking the average value). Dropout (Srivastava et al., 2014) is a technique to prevent overfitting when training a neural network by introducing noise to the learning process. At each epoch, a population of features are randomly chosen and dropped from the network training process. This has been proven to significantly reduce overfitting and improve the network performance.

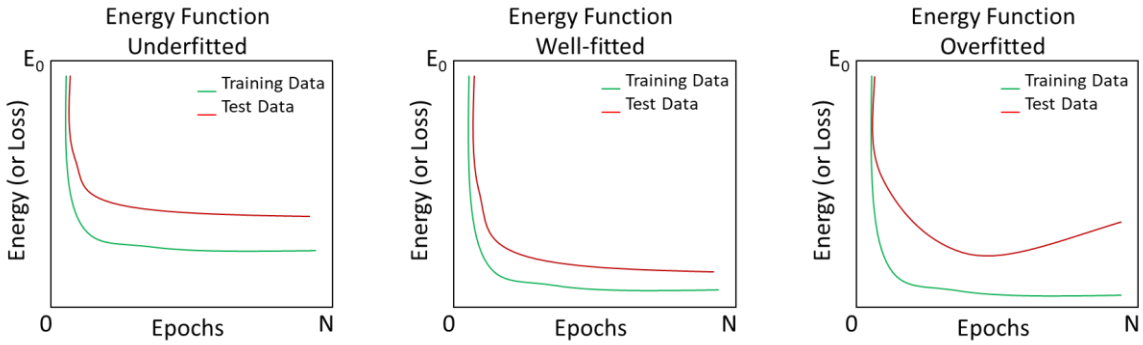

Figure 1.2: The determination of network quality by looking at the energy (loss) functions of the training and testing datasets.

1.5. Objectives and Organization

The primary goal of my research is to examine if machine learning algorithms can be used to solve three outstanding challenges in seismic inversions: (1) the limited resolution due to wavelet effects, (2) dependence on some rock physics models when inverting for petrophysical properties, and (3) quantification of the posterior distribution of the prediction results to resolve the resolution discrepancy between the time domain seismic signal and well logs in depth domain.

In the second chapter, I show how a single-layer network namely Boltzmann machine can be implemented as an unsupervised learning algorithm to predict elastic properties (impedances and density) at higher resolution than conventional approaches, while still retaining the physical relationship between the seismic amplitudes and reflectivity series. In the third chapter, I introduce a new multimodal network structure called the Cross-shape Deep Boltzmann machine that is able to simultaneously capture six possible relationships between four different input training data and use those to invert for petrophysical properties from pre-stack seismic amplitudes in a dataset with limited well coverages. In the fourth chapter, I design a multimodal deep learning network to construct the posterior distributions of the subsurface elastic properties from seismic gathers and demonstrate with a field data application. Chapter five summarizes the research and highlights some of the future research areas.

Chapter 2 : A Boltzmann Machine for High Resolution Pre-stack Seismic Inversion

Publication: Interpretation 2019, Vol 7: SE215-SE224, ISSN 2324-8858

https://doi.org/10.1190/INT-2018-0234.1

2.1. Introduction

Seismic inversion aims at predicting some indicative rock properties such as elastic impedances from the recorded dataset (e.g., Sen, 2006; Singha et al., 2014; Chatterjee et al., 2016) by utilizing the physics of wave propagation within the subsurface. Typically, it uses the Zoeppritz equation (Zoeppritz 1919) or its linearized versions such as the Aki-Richards (1980) approximation for computing reflection amplitudes. This inversion problem comprises of (a) a forward model, which is the convolution between a source wavelet, which represents seismic signal sent to the subsurface, and a reflectivity series, which represents the changes in subsurface elastic properties, (b) an objective function that compares the difference between the recorded and synthetic signals, and (c) a model update.

Common seismic inversion methods suffer from the limited bandwidth of the seismic data, where the low frequency (0-10 Hz) and the high frequency (above 50 Hz) contents are missing. The lack of high frequency content limits the resolution of fine features, while the lack of low frequency content results in unreliable trends in the results. The methods based on statistical approaches such as those from Srivastava and Sen (2010) have been applied successfully to broaden the high frequency bandwidth to improve the seismic resolution. Meanwhile, the common approach to recover the low frequency content is to use low frequency starting models to guide the solution (e.g., Sirgue, 2006; Innanen et al., 2011). The models are created by populating well log values throughout a 3D volume using some interpolation and extrapolation techniques. This process requires a geological

framework built on the interpreted surfaces from a given seismic dataset. In areas with complex geological settings, the interpretation comes with large uncertainties, which transform into uncertainties of the inversion results. In order to reduce the dependence on the guiding model, better approaches that are independent of human inputs are preferred.

Tikhonov (1963) introduced the first and second derivative operators as regularization terms to improve the stability of the underdetermined inverse problems. Many scholars used those to stabilize the solution of the seismic inverse problems (e.g., Menke, 1989). Studies demonstrated the effectiveness of these regularization terms in stabilizing the solutions (Bijwaard et al., 1998; Abubakar et al., 2009; Habashy et al., 2011). Zhang et al. (2013) successfully employed this approach to improve the lateral continuity of their results from basis pursuit inversion.

A Boltzmann machine is a special neural network which comprises of a Hopfield Neural Network (HNN; Hopfield et al.,1985) and the Mean Field Annealing (MFA) (Bilbro et al., 1989). The HNN is known to converge quickly toward an optimal solution once an initial state of the neurons is provided and an update rule is applied. Wang and Mendel (1992) employed HNN in their reflectivity solver and wavelet extractor. However, the HNN update rule only guarantees convergence to a local minimum (Rojas 1996). To ensure a global minimum solution, I integrate a stochastic model update from the MFA. This annealing algorithm reaches the equilibrium state 1-2 orders of magnitude faster than simulated annealing while still retaining a close relationship with the Hopfield network (Bilbro et al., 1989). Reported successful applications of HNN and MFA include Ansari et al. (1993), Dixon et al. (1995) in communication network design. In seismology, Calderon-Marcias et al. (1997) improved Wang and Mendel (1992) solution of the deconvolution problem by integrating HNN with MFA and applied it to multiple attenuation problem as well. Kahoo et al. (2006) later replicated this method and found that

the HNN was not sensitive to noise, the reflectivity estimator was not sensitive to frequency bandwidth of the wavelet, and the method improved the temporal resolution of a seismic section.

In this study, I employ the Boltzmann machine framework to perform pre-stack inversion with the goal of obtaining high resolution elastic properties. In my inverse problem, the forward modeling is a convolution model, the objective function is the Hopfield objective function, and the model update is guided by the stochastic MFA algorithm for a global minimum solution. The stochastic model update schematic of the Boltzmann machine helps broaden the search radius for all possible solutions to reduce the non-uniqueness of the final solution that common iterative least square solvers suffer from. In addition, the incorporated Hopfield network brings a huge advantage to the computational cost for its great capability of parallel calculations when dealing with large datasets.

The introduced algorithm includes two main steps: the first performs deconvolution to estimate the offset dependent reflectivity terms using a set of stationary wavelets calculated from the seismic data, which is similar to Calderon-Marcias et al. (1993) approach; the second inverts for Elastic P-, S-Impedances and density from the output reflectivity of the former step. To improve the lateral continuity of the results, I integrate the Tikhonov regularization into simultaneous inversion at multiple common depth point (CDP) locations. Common AVA inversions are performed on CDP-by-CDP location, which may result in a lack of lateral continuity in the final solutions as the algorithm does not take into account the relationship between neighboring traces. By reorganizing the order of input trace samples and introducing the first or second order Tikhonov regularization terms, the inversion preserves more geological events than the conventional trace-by-trace inversion approaches (Zhang et al., 2013).

For a conventional pre-stack inversion approach, low frequency starting models are required at the beginning of the inversion process. This requires building property models from the horizons interpreted on stack sections, which come with large uncertainties, especially in areas with complex geological settings. On the other hand, my approach directly inverts for reflectivity terms, bypassing the requirements of a starting model to guide the solution. Upon completion of the individual property reflectivity inversion step, I start to pick horizons on the reflectivity sections, build property models from those horizons and use those to predict the absolute target log values.

I show the results of this approach on a real 2D pre-stack seismic dataset that contains multiple vertical well paths with sonic, shear and density logs. The results confirm that this method is capable of producing indicative sections with realistic features supported by well results and suggest some possible targets within the vicinity of the seismic line.

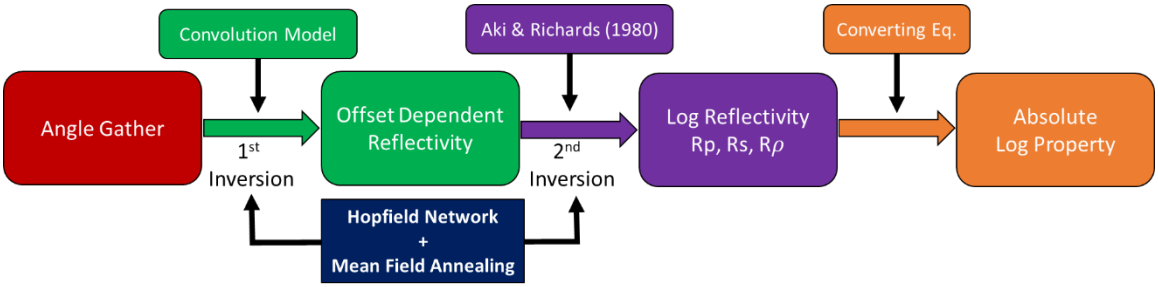

Figure 2.1: Inversion flow to obtain final log properties with Boltzmann machine.

2.2. Methods

The general workflow of this study is summarized in Figure 1, which involves two inversion steps: (1) inverting for an offset dependent reflectivity series from an input angle gather by assuming known stationary wavelets extracted from the seismic gathers, and (2) inverting for individual log reflectivity series from the output of (1) using a linearized

Zoeppritz equation by Aki and Richards (1980). Both inversion steps use the Hopfield network along with Mean Field Annealing as the main optimization algorithm. To improve the lateral continuity of the results, I incorporate a Tikhonov spatial regularization term into the first inversion step. The final step involves converting from reflectivity to absolute log values using a simple conversion formula.

2.2.1. Hopfield Neural Network

The Hopfield Neural Network (HNN) introduced by Hopfield et al. (1985) is an auto-associative network that requires a fixed input pattern to obtain the outputs. It is well-known for its fast convergence toward a stable solution once an initial stage of the neurons is provided and an update rule is applied. This neural system has been applied widely in engineering problems such as in communication network design. In exploration seismology, Wang and Mendel (1992) was the first to apply HNN in their automatic deconvolution and wavelet extraction problems.

A Hopfield network is a fully interconnected neural network in which each unit is connected to every other unit. The net has symmetric weights with no self-connections, as described in the schematic plot in Figure 2.2. Consider an inverse problem $\mathbf{y} = \mathbf{G}\mathbf{x}$ where the recorded data are fed into **y,** an operator **G**, and the target is to find the values of **x**. In case **G** is a square invertible matrix; the solution is a simple $\mathbf{x} = \mathbf{G}^{-1}\mathbf{y}$. Unfortunately, it is rarely the case as **G** could be a singular matrix. One could attempt to solve for this equation by using a direct inversion with some regularization term such as the damped least squares method. This is viable for a small size and simple problem. However, for large problems with strong heterogeneity in reservoir property, this method shows its weakness in computation cost in finding the inverse matrix and the requirement to update the damping parameter.

One way is to iteratively search for the values of **x** that minimizes the L_2 norm error function $\boldsymbol{E(x) = \frac{1}{2}\|y - Gx\|_2}$, which is expanded into the general form in Equation (2.1):

$$\boldsymbol{E(x) = \frac{1}{2}y^T y - \left(-\frac{1}{2}x^T G^T G x + y^T G x\right).} \quad (2.1)$$

The Hopfield network can be used to solve this optimization problem. Detailed description of HNN can be found in Hopfield et al. (1985), Wang and Mendel (1992), and Sen and Stoffa (2013). The energy or the cost function, and update rules of the network are given by,

$$\boldsymbol{E(x_i) = -\frac{1}{2}x_i^T W x_i + \phi^T x_i,} \quad (2.2)$$

$$\boldsymbol{x_{i+1}} = \begin{cases} 1 \; if(-\boldsymbol{W^T x_i + \phi}) > 0 \\ 0 \; else \end{cases}. \quad (2.3)$$

where **W** represents the symmetric weighting matrix **W**, $\mathbf{x}_i$ represents the status of the i^{th} neuron, $\boldsymbol{\phi}$ represents the threshold by which the input energy must surpass to have the neuron change its stage. Equation (2.2) is the energy function of the network, which is nothing but a reorganized form of Equation (2.1) without the constant term $\boldsymbol{\frac{1}{2}y^T y}$; $\boldsymbol{W = G^T G}$**-diagonal(**$\boldsymbol{G^T G}$**)** is a symmetric matrix; and the threshold term $\boldsymbol{\phi = G^T y +}$ **diagonal(**$\boldsymbol{G^T G}$**)**$\boldsymbol{x}$ is dynamically dependent on the unknown $\boldsymbol{x}$. Equation (2.3) is the activation function that represents the update rule of the network: a neuron changes its status when the incoming external force, which is $\boldsymbol{W^T x_i}$, surpasses the threshold $\boldsymbol{\phi}$.

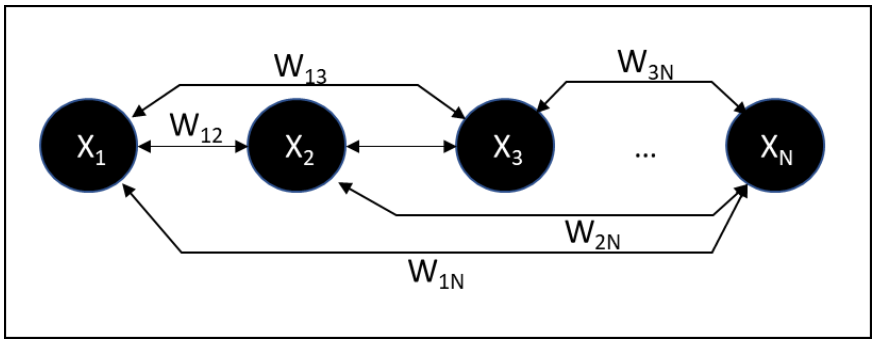

Figure 2.2: The schematic of a discrete Hopfield network.

2.2.2. Mean Field Theory

For conventional networks, a neuron can only feedback with a discrete value of either {0,1} or {-1,1}, depending on the format of the results. Meanwhile, the reflectivity series can take any value between the interval [-1, 1]. In order to capture these values, Wang and Mendel (1992) decomposed the reflectivity term using the geometric coding of Young (1971). However, such additional step increases the size of the inverse system, which consequently reduces the computational advantage of the network.

A better choice is to use a stochastic activation function, like a Mean Field Annealing (Bilbro et al., 1989). This uses a sigmoid function, such as in Equation (2.4), that allows the network to capture all possible solutions in the interval [-1, 1] without sacrificing the calculation cost. In this equation, a temperature term T represents the cooling schedule of the system, which is borrowed from the simulated annealing algorithm (Kirkpatrick et al., 1983). The formula of this annealing schedule is shown in Equation (2.5), where the temperature at the k^{th} iteration (T_k) is controlled by an initial temperature T_0 that decays at a rate β. Details on MFA can be found in Bilbro et al (1989), Calderon-Macias et al. (1997) and Sen and Stoffa (2013).

$$\boldsymbol{x}_{i+1} = \frac{1}{1 + exp\left(-\frac{\boldsymbol{W}^T\boldsymbol{x}_i - \boldsymbol{\phi}}{T_k}\right)}; \tag{2.4}$$

$$T_k = T_0 exp\left(-\beta\sqrt{k-1}\right). \tag{2.5}$$

2.2.3. Forward Modeling Formulation

Convolution Model

The convolution model is used to connect the reflectivity term with the seismic trace. Ignoring the external noise, an input seismic trace $\boldsymbol{y} = [y_1, y_2 \ldots y_N]^T$ can be written using a stationary wavelet **s** and reflectivity series $\boldsymbol{x} = [x_1, x_2 \ldots x_N]^T$ as:

$$y_k = \sum_{i=1}^{N} s_{k-i} x_i, \tag{2.6}$$

which can be rewritten in the matrix form $\boldsymbol{y} = \boldsymbol{Sx}$, where $\boldsymbol{y}$ represents the input seismic trace; $\boldsymbol{S}$ represents the wavelet kernel which is a Toeplitz matrix created from the wavelet term $\boldsymbol{s}$; and $\boldsymbol{x}$ represents the reflectivity term. The HNN algorithm can be employed in this case by considering unknown $\boldsymbol{x}$ to be the status of the neurons in Equations (2.2) and (2.3).

Linearized Zoeppritz Equation from Aki and Richards (1980)

Estimating the reflectivity series is the prerequisite for the estimation of rock properties, which in this case the Vp, Vs and Density. The Zoeppritz equation (Zoeppritz, 1919) gives exact reflection coefficients. Several linearized approximations of Zoeppritz equation by Aki and Richards (1980), Fatti et al. (1994) are widely used. In this study, I employ the Aki-Richards (1980) equation to calculate the PP reflection coefficient at a time sample as a function of offset angle θ as:

$$R_{pp}(\theta) = C_1(\theta)\frac{\Delta V_p}{V_{p0}} + C_2(\theta)\frac{\Delta V_s}{V_{s0}} + C_3(\theta)\frac{\Delta \rho}{\rho_0}, \tag{2.7}$$

where $$C_1(\theta) = \frac{1}{2}(1 + tan\theta^2),\ \ C_2(\theta) = -4sin\theta^2\frac{V_{s0}^2}{V_{p0}^2},$$

$$C_3(\theta) = \frac{1}{2}\left(1 - 4sin\theta^2\frac{V_{s0}^2}{V_{p0}^2}\right).$$

In Equation (2.7), V_{p0}, V_{s0}, ρ_0 are the average Vp, Vs and Density across the boundary. Extending Equation (13) to full angle offset range θ_i for i=1…M, full data range j=1…N, and using the following rewritten terms $R_p = {\Delta V_p}/{V_{p0}}, R_s = {\Delta V_s}/{V_{s0}}, R_\rho = {\Delta\rho}/{\rho_0}$, I obtain a series of equation, which can be written in the matrix form as:

$$\begin{bmatrix} R_{pp}(t,\theta_1) \\ R_{pp}(t,\theta_2) \\ \vdots \\ R_{pp}(t,\theta_M) \end{bmatrix} = \begin{bmatrix} C_1(t,\theta_1) & C_2(t,\theta_1) & C_3(t,\theta_1) \\ C_1(t,\theta_2) & C_2(t,\theta_2) & C_3(t,\theta_2) \\ \vdots & \vdots & \vdots \\ C_1(t,\theta_M) & C_2(t,\theta_M) & C_3(t,\theta_M) \end{bmatrix} \begin{bmatrix} R_p(t) \\ R_s(t) \\ R_\rho(t) \end{bmatrix}, \tag{2.8}$$

which can be represented in the form $\boldsymbol{y} = \boldsymbol{Gx}$ with unknown $\boldsymbol{x} = [R_p(t) \quad R_s(t) \quad R_\rho(t)]^T$. The HNN algorithm can be used in this case to quickly find the solution of this inverse problem.

$$\begin{bmatrix} d_{1,1} \\ d_{1,2} \\ \vdots \\ d_{1,N} \\ d_{2,1} \\ d_{2,2} \\ \vdots \\ d_{2,N} \end{bmatrix} = \begin{bmatrix} G_{1,1} & G_{1,2} & \dots & G_{1,N} & 0 & 0 & \dots & 0 \\ G_{2,1} & G_{2,2} & \dots & G_{2,N} & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ G_{N,1} & G_{N,2} & \dots & G_{N,N} & 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & G_{1,1} & G_{1,2} & \dots & G_{1,N} \\ 0 & 0 & \dots & 0 & G_{2,1} & G_{2,2} & \dots & G_{2,N} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \dots & \ddots & \vdots \\ 0 & 0 & \dots & 0 & G_{N,1} & G_{N,2} & \dots & G_{N,N} \end{bmatrix} \begin{bmatrix} m_{1,1} \\ m_{1,2} \\ \vdots \\ m_{1,N} \\ m_{2,1} \\ m_{2,2} \\ \vdots \\ m_{2,N} \end{bmatrix} \tag{2.9a}$$

$$\begin{bmatrix} d_{1,1} \\ d_{2,1} \\ d_{1,1} \\ d_{2,2} \\ \vdots \\ d_{1,N-1} \\ d_{2,N-1} \\ d_{1,N} \\ d_{2,N} \end{bmatrix} = \begin{bmatrix} G_{1,1} & 0 & G_{1,2} & 0 & \dots & G_{1,N} & 0 \\ 0 & G_{1,1} & 0 & G_{1,2} & \dots & 0 & G_{1,N} \\ G_{2,1} & 0 & G_{2,2} & 0 & \dots & G_{2,N} & 0 \\ 0 & G_{2,1} & 0 & G_{2,2} & \dots & 0 & G_{2,N} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ G_{N,1} & 0 & G_{N,2} & 0 & \dots & G_{N,N} & 0 \\ 0 & G_{N,1} & 0 & G_{N,2} & \dots & 0 & G_{N,N} \end{bmatrix} \begin{bmatrix} m_{1,1} \\ m_{2,1} \\ m_{1,1} \\ m_{2,2} \\ \vdots \\ m_{1,N-1} \\ m_{2,N-1} \\ m_{1,N} \\ m_{2,N} \end{bmatrix} \quad (2.9b)$$

Spatial Regularization

The 1st and 2nd order Tikhonov regularization matrices (Tikhonov, 1963) have been used widely to constrain the solution of some linear inverse problems. In applied seismology, they have been used in improving the solutions to some problems such as: (a) constraining the underdetermined seismic tomography problem (Bijwaard, 1998), (b) stabilizing the solutions of the full Waveform Inversion at some sharp boundaries (Abubakar et al., 2009), or (c) improving the lateral continuity of the inverted results (Zhang et al., 2013).

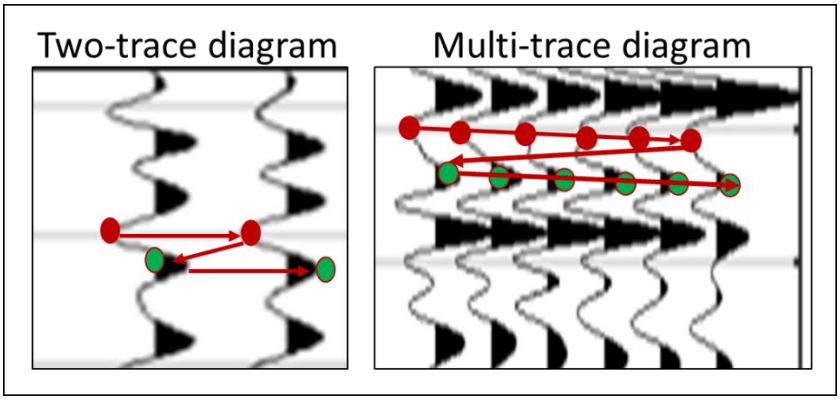

Figure 2.3: The "Z" shape trace sample sorting method for multi-trace case. Left panel shows the order of sampling for two traces, while right panel shows the order of sampling for multiple traces.

In this study, I use the "Z" shape trace sampling method to re-organize the samples as described in Figure 2.3. For details, see Zhang et al (2013). As an example, for an inverse

problem $\boldsymbol{d} = \boldsymbol{Gm}$ with two traces of length N, the conventional trace by trace sorting is expressed as in Equation (2.9a), and the "Z" shape trace sampling is expressed in Equation (2.9b).

Conversion from Reflectivity to Absolute Log Values

Once the individual reflectivity terms at Equation (8) are obtained, I convert those into the absolute Vp, Vs and Density terms. One condition for Equation (7) is the small change in property across the boundary. With this assumption, I use the following conversion equation from Zhang et al (2013), where the subscripts O represent the low frequency models:

$$\begin{aligned} Vp(t) &= Vp_0(t)exp\left(\int Rp(t)dt\right) \\ Vs(t) &= Vs_0(t)exp\left(\int Rs(t)dt\right). \\ \rho(t) &= \rho_0(t)exp\left(\int R\rho(t)dt\right) \end{aligned} \tag{2.10}$$

2.3. Results

I applied this algorithm to a 2D field dataset with several wells located along the line. The available logs include sonic, shear and density with target to be a hot shale interval with low density and P-wave velocity lying between two stiffer intervals characterized by high Vp and density values, as shown in Figure 2.4. The pre-stack seismic data was converted into angle gathers, ranging from 0-35 degrees at 5 degrees interval. Correspondingly, I extracted 7 stationary statistical wavelets (except the zero offset) that best tie well logs with the seismic data to feed into the inversion algorithm.

The provided pre-stack seismic data comes with four already interpreted horizons, where the purple horizon is represented on the reservoir top, as shown in Figure 2.5. The quality of the purple and green horizons is supported by strong seismic events well

correlated with well log data. The objective is to better delineate the geobody distributions in two areas: between the blue and purple horizons, which is highlighted by yellow arrows.

From the well log property analysis in Figure 2.6, I observe that the P- and S-impedance can highlight the hot shale from the over- and under-lying stiffer materials. The P-impedance versus density cross-plot with color coding of gamma ray index (GR) strongly outlines a hot shale interval to be located toward the low P-impedance values. Moreover, the histogram plots of the P- and S-impedance values show clear separations between reservoir and non-reservoir values. Therefore, I performed pre-stack AVA inversion to obtain the elastic P- and S-impedance and the density section to assist further mapping process.

I first tested the algorithm at the well locations for quality control before inverting the data for the entire 3D line. I performed the multi-trace inversion with a pair of neighborhood CDP locations and used the first order Tikhonov regularization term. I was able to obtain the best-fit results that match very well with the controlling angle gather, and the final inverted P-reflectivity matches very well with the actual P-reflectivity calculated from well logs, as shown with three example wells in Figure 2.7.

The inverted P- and S-reflectivity sections are of higher resolutions with sharper events with clear terminations, as shown in the highlighted zone in Figure 2.8. This allows me to perform horizon tracing with higher confident level to build up a more detailed structural framework for the P-, S and Density models used for converting reflectivity sections into absolute log values. I used a low frequency model of 10-15 Hz of P-, S-impedances and density to convert the log reflectivity into absolute log values using Equations 10. With this workflow, I expect to obtain results with more details. The final predicted logs match very well with the well logs, as shown in Figure 2.9. With the inversion results, I was able to map the low density, Vp and Vs reservoir intervals. The

predicted logs in the shallow section of Well 1 follows the log trends in Well 2 and 3 of the same intervals, therefore I believe these to be of reliable quality. However, I also observe some minor mismatches in areas where very strong property contrasts are predicted within thin sections. These random features are below seismic resolution, therefore may not be captured by this dataset.

With satisfactory results at well locations, I applied the inversion results to the 2D line. Figures 2.10-2.12 show the comparisons between the estimated P-, S- impedances and Density sections using Boltzmann machine and from a regular model-based deterministic inversion. The inversion algorithm is capable of outlining more features than from the latter method. I observe multiple geobodies with low P-impedances in the shallower sections and below the base of the reservoir. Comparing those with the inverted S-impedance and density sections, and incorporate with the well analysis result, I predict that there are possible targets distributed in the deeper section toward the left portion of the 2D seismic lines highlighted by outstandingly low P-impedances, S-impedances and average density values. The termination of the bright amplitude anomaly along with the possible structural closure from dipping event are two strong indicators of possible change in fluid content at this location.

2.4. Summary

I present a new application of a machine learning algorithm for pre-stack inversion to produce high resolution indicative sections to assist in the interpretation process. I employed a Boltzmann machine, which comprises of a Hopfield neural network and a stochastic model update from Mean Field Annealing algorithm to automatically invert for the elastic impedances and density. I first inverted for offset dependent reflectivity series and use those to solve for the reservoir properties. To improve the quality of the inverted

sections, I applied a "Z" shape trace sorting schematic and the 1st order Tikhonov as a spatial regularization term to improve the lateral continuity of the inverted sections.

One significant aspect of this method is that the inversion process does not require a starting model. Therefore, I was able to avoid any uncertainties or biases introduced from the horizon mapping process to create a model. I introduced a low frequency model only after the deconvolution by inversion was performed and observed that the resulting reflectivity sections are of high resolution.

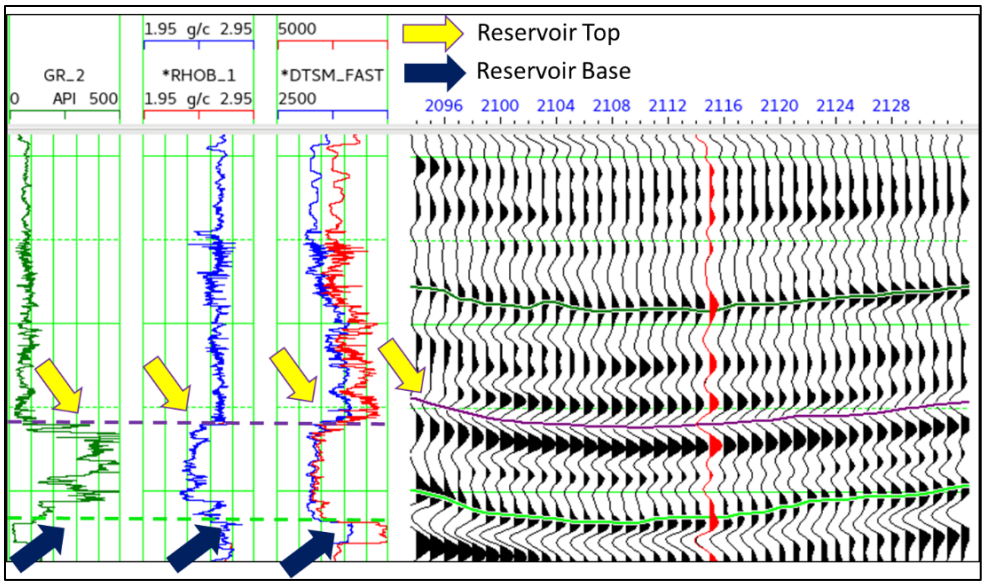

Figure 2.4: The well log responses at one well and the corresponding horizons on seismic section. In this case, the top of target zone is the purple horizon, while the base is the green event.

I demonstrated the algorithm on a 2D pre-stack dataset with several control wells. The primary objectives were to estimate high resolution P- and S-impedance sections to delineate possible features in the shallow and deeper sections from the main reservoir interval. The input parameters calibrated at well locations show excellent matches with seismic data and original well logs. The algorithm is capable of capturing multiple high

frequency features of the well logs and populates those away from the wells with the guidance of seismic data. The satisfactory results when applying the methods to 2D lines suggest that there might be some isolated geobodies in the deeper left portion of the 2D seismic line.

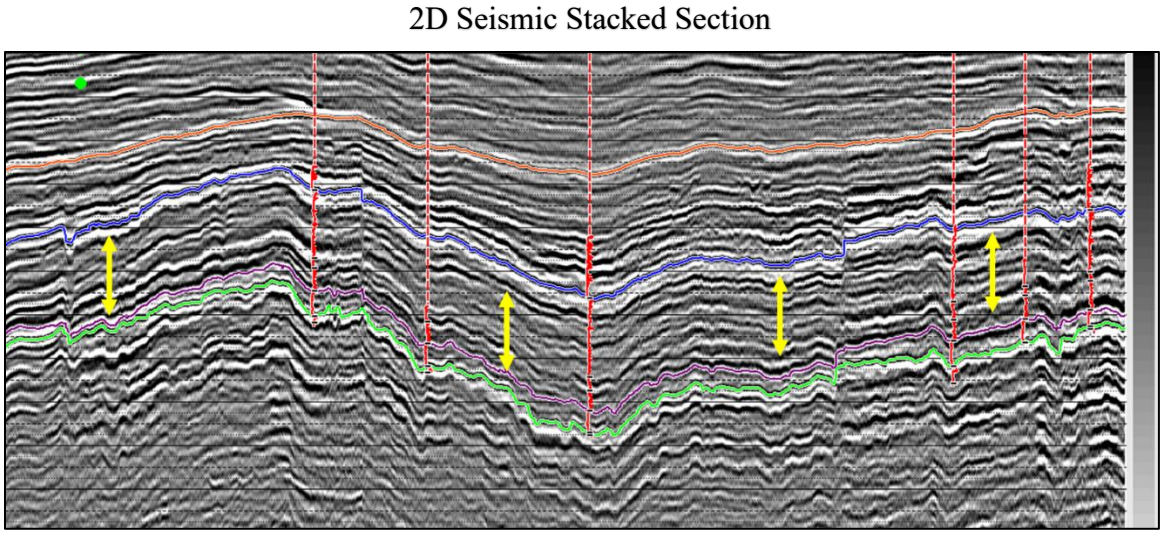

Figure 2.5: The seismic section showing interpreted horizons. The reservoir interval is between purple and green horizons. The objective is to better delineate geobody distributions in two areas: between the blue and purple horizons, which is highlighted by yellow arrows.

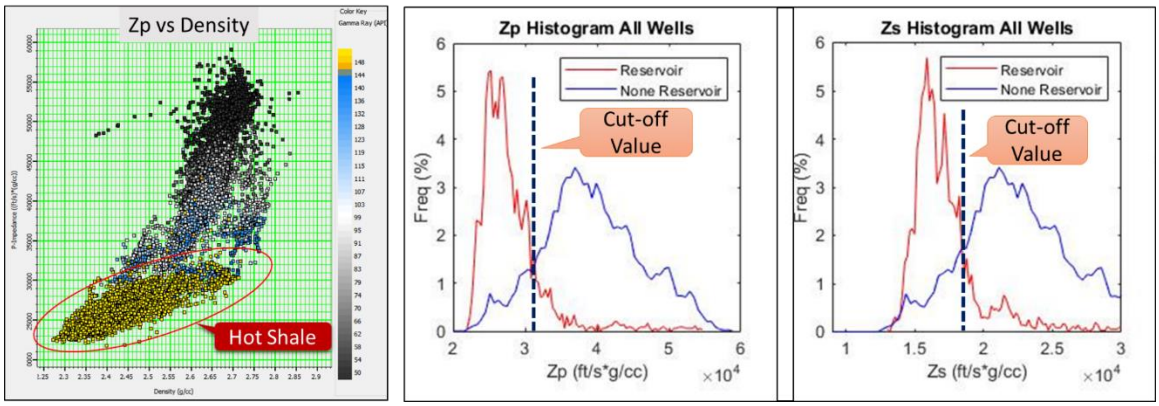

Figure 2.6: The well log analyses with cut-off value for reservoir and non-reservoir zones using P-Impedance (Zp) and S-impedance (Zs). The histogram of P-impedance and

S-impedances for reservoir and non-reservoir intervals suggests clear separation and the cut-off values are useful for adjusting color-bar for displaying purposes.

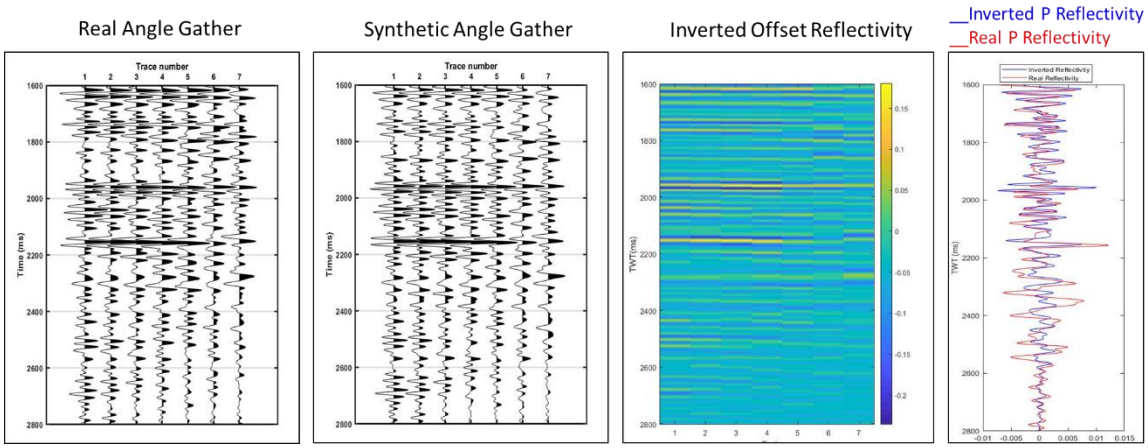

Figure 2.7: Inverted Offset Reflectivity is supported by excellent match between the synthetic and the real angle gathers. The final inverted P reflectivity matches very well with actual log calculated from well log.

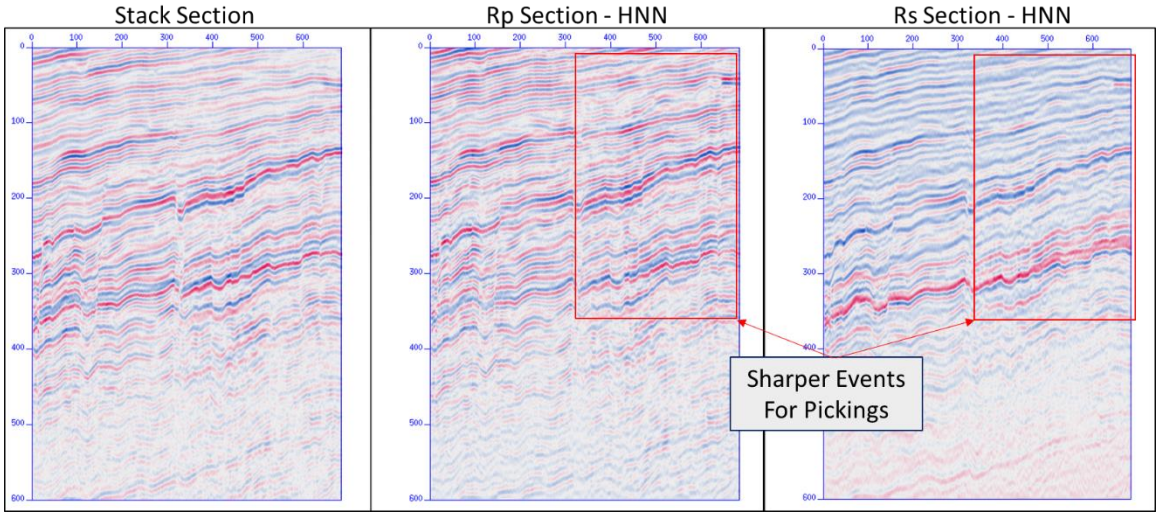

Figure 2.8: Inverted Rp and Rs sections (right two panels) are of higher resolution than the original stacked section. Sharper events for picking are in highlighted area.

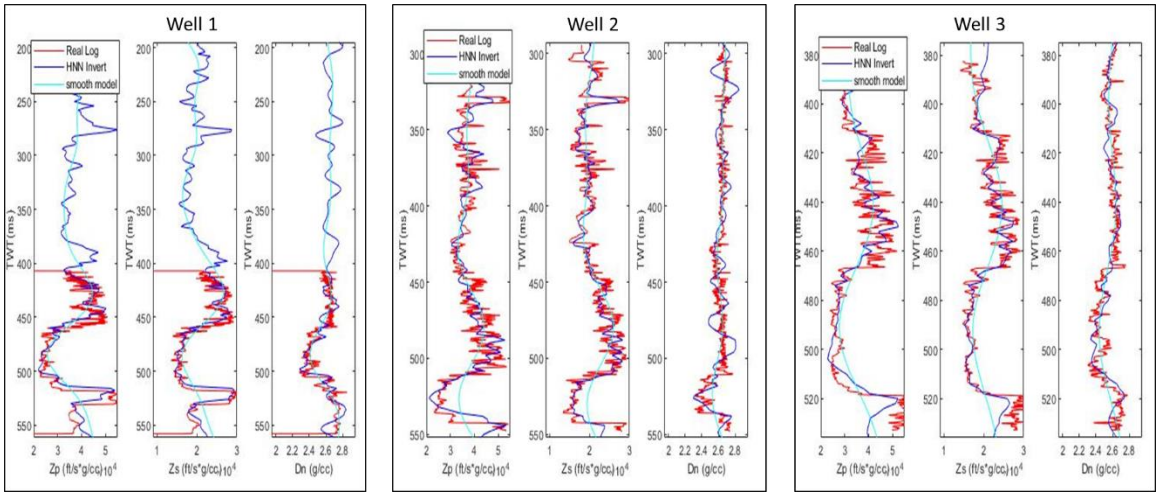

Figure 2.9: Inverted results (blue curves) plotted against the real logs (red curves) and the smooth model used to convert from reflectivity to absolute log values (cyan curves) at some well locations. Note the time scale is for reference only.

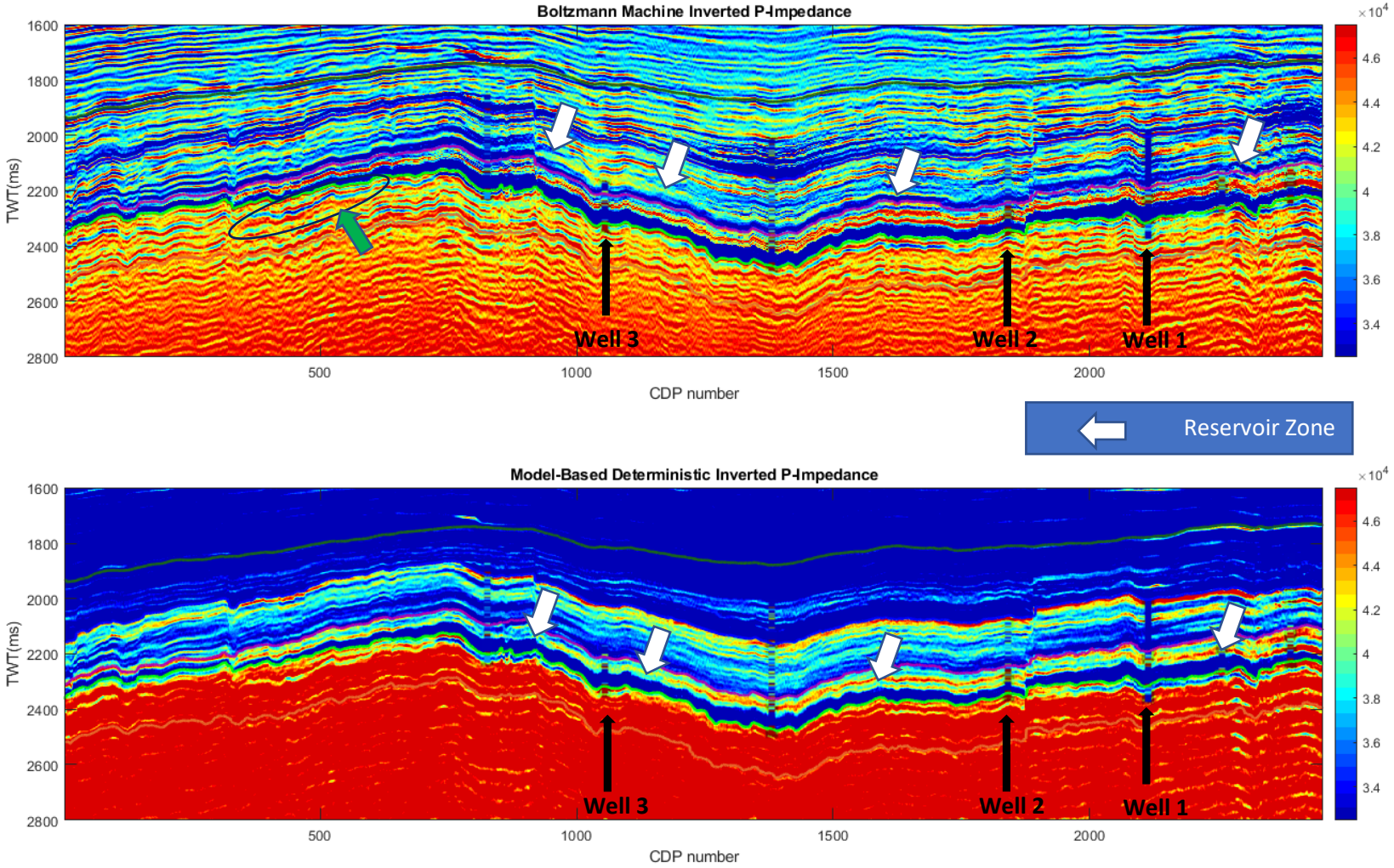

Figure 2.10: Inverted P-Impedance shows excellent matches at well locations. Notice the low impedance events in the reservoir interval as highlighted by white arrows. A prospective zone is highlighted with black ellipse and green arrow.

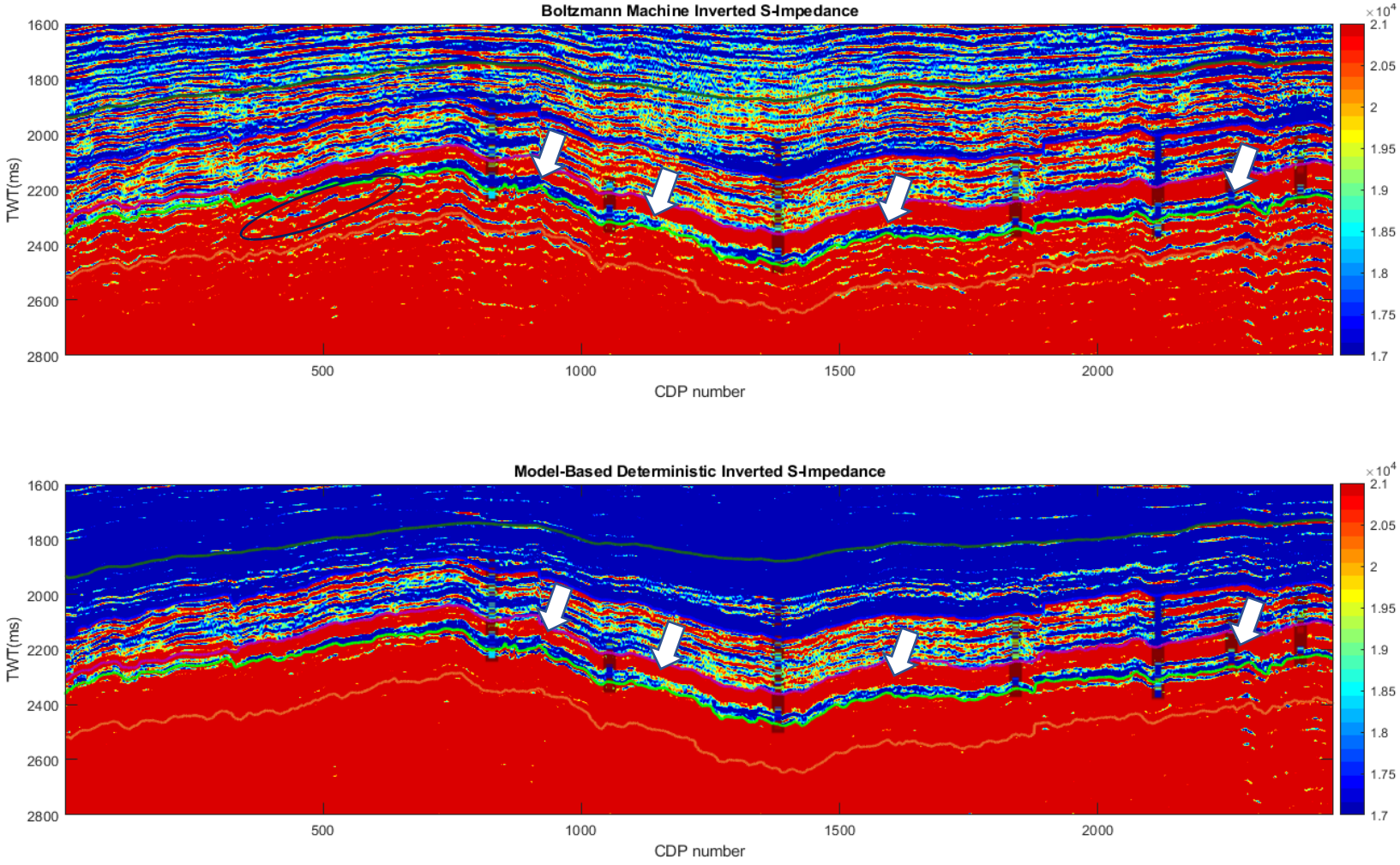

Figure 2.11: Inverted S-Impedance shows excellent matches at well locations. The reservoir zone is highlighted by low shear impedances highlighted by white arrows. The additional prospective zone is in white ellipse with low S-Impedance.

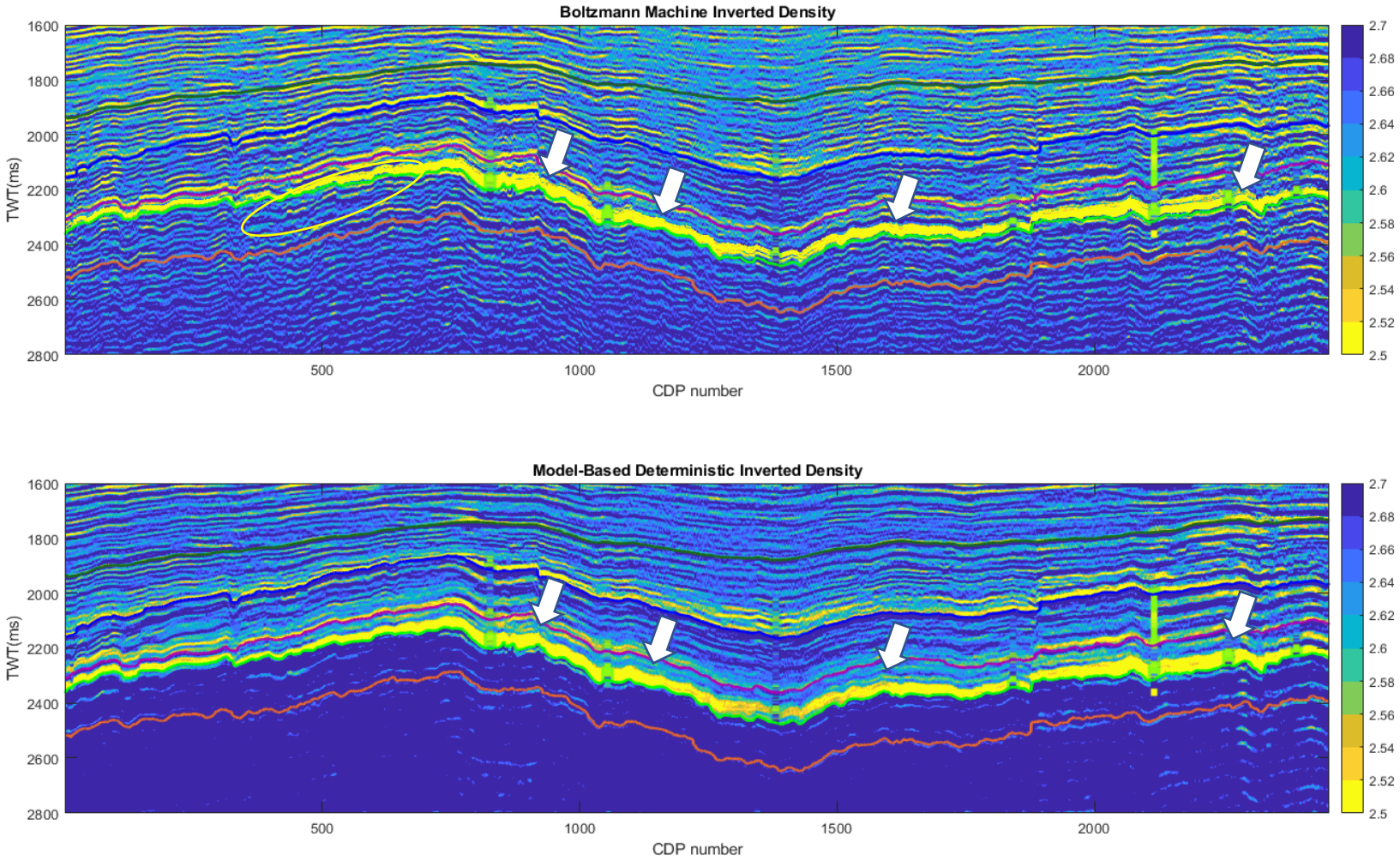

Figure 2.12: Inverted Density section shows excellent matches at well locations. Notice the reservoir is well defined by the low-density interval in the middle of the section, following the white arrows. The prospective zone is in white ellipse with low density.

Chapter 3 : A Cross-shape Deep Boltzmann Machine for Petrophysical Seismic Inversion

Manuscript Submitted to the AAPG Bulletin

3.1. Introduction

Typical petrophysical logs include porosity (ϕ), water saturation (Sw) and shale volumetric (Vshale). The lithological units in sand-shale environments are determined by the shale volume contrasts, and water saturation indicates hydrocarbon bearing zones. Several approaches (e.g Bachrach, 2006, Spikes et al., 2007) have been developed and applied to predict these properties from recorded seismic data via direct inversion using some rock physics models that represent statistical relationships between target logs and elastic properties.

Seismic inversion techniques (e.g Sen, 2006) utilize the physics of the wave propagation within the subsurface to predict rock elastic properties such as P-impedance, S-impedance, and density. The inversion process comprises of (1) a forward problem, which assumes a convolution model between the source signal and a reflectivity series that represent the changes in rock properties, (2) an objective function that compares the difference between the recorded and synthetic signals, and (3) a model update scheme. The source wavelet is often difficult to measure for its non-stationary property; its shape and intensity change due to the high frequency attenuation when traveling through the subsurface (Yilmaz, 2001). Meanwhile, the reflectivity series is typically calculated using a full Zoeppritz equation (Zoeppritz, 1919), or its linearized version such as Aki-Richards (1980) approximations. Such calculations are validated for reflected signals whose incident angles are up to the critical threshold (Wang, 1999).

Commonly, inversions are performed within a small time-window, within which the wavelet is assumed to be stationary with minimum changes in amplitudes and

frequency content. The three-term Aki-Richards or Shuey (1985) approximation is often used to calculate the reflectivity series for incident angles up to 35 degrees (Ikelle and Amundsen, 2018), assuming the property contrasts across the boundaries to be small. These work best in fields where hydrocarbon signatures are recognizable on seismic sections. However, they might not be valid when the background geology becomes complicated with significant property contrasts or in fields in which reservoir zones occupy large time intervals such that waveform changes are unavoidable. To improve the quality of inversion results in such conditions, non-stationary wavelet and varying maximum incident angles are required. However, the estimation of source wavelets still remains a difficult task (e.g., Lines and Treitel, 1985; Fomel, 2007; Dai et al., 2016, Wang and Morozov, 2020). An alternative approach that does not directly require input wavelets is more desirable.

The step after seismic inversion is to relate elastic properties to petrophysical properties at the well locations using the relationship established at the well locations, which is then used to infer the petrophysical properties away from the wells from seismic inversion results. Artificial neural networks (ANN) that have been applied successfully in many areas of science and engineering fields, have also been used for this purpose. A neural network is a system of interconnected units that simulate the neurons within human brains. The network learning capability is trained by the interaction between neural nodes when information is passed through the system. A properly trained network is capable of recognizing important features of a new dataset for predicting target outputs. The training process could be either (1) supervised, where known input data and corresponding labels are used to guide the network structure toward a desired relationship, or (2) unsupervised, where the algorithm works on its own to extract hidden information from the unlabeled input data.

The learning capabilities of ANN have been valuable assets in performing automatic computations. Unsupervised networks such as the Hopfield neural networks have been used for velocity analysis (Calderon-Marcias et al., 1993), or pre-stack inversion of seismic data (Phan and Sen, 2019) by rearranging the network structure to resemble the physical framework of the inverse problem. Supervised deep learning systems such as convolution neural networks (CNN-Lecun et al., 1995) or autoencoder-decoders (Hinton et al., 2011) are widely used to simulate the inversion operator from labeled data to compute elastic impedances (e.g., Li et al, 2020, Zheng et al., 2019, Biswas et al., 2019). The supervised networks for seismic inversion may be designed such that they do not require any prior knowledge of the wavelets. However, they do require a large amount of training data coverage to be able to capture all possible scenarios.

In common network designs for applications in seismic inversion, the training data are fed into visible layers located at two ends of the network structures. For pre-stack inversion problems that involve solving for multiple rock property terms, the training models are first merged and reorganized into a single neuron layer before they are input into the network. During the learning process, model parameters are updated to minimize an objective function. These update schemes do not clearly provide information on how individual model is manipulated.

A Boltzmann machine is a probabilistic deep learning network that is composed of two layers of neuron units which connects with each other via bidirectional weights. A cross-shape deep Boltzmann machine (CDBM) is a multimodal deep Boltzmann machine (Srivastava et al., 2014) created by connecting four different Boltzmann machines at the vertices via a hidden neuron layer located at the center of the cross. This arrangement allows the center units to distribute changes of input data at any visible vertex throughout the network toward remaining vertices. In an inversion problem for petrophysical

properties, the CDBM network structure simulates simultaneous interaction between the rock properties and the seismic amplitudes: any changes in any output terms are transmitted and results instantly in a change in seismic amplitude, and vice versa. Once the network is trained, one could deduce any geological information from the established mathematical relationships between various input terms.

A CDBM is designed here to carry out pre-stack seismic inversion that can estimate petrophysical properties directly. Unlike the conventional seismic inversion, this approach is not restricted to stationary wavelets, and does not require any rock physics model to direct the inversion algorithm. Four different data types, including seismic gathers, and three target logs of porosity, water saturation and shale volumetric, are used to train the network. Six different nonlinear relationships between 4 data types are simultaneously captured, among which the three connections from seismic amplitudes to the petrophysical logs are expected for property predictions. Unlike any conventional deep learning system, which expects large data for all what-if scenarios into training, the CDBM only requires the anticipated upper and lower bounds of the data and labels, which could be obtained from available well data. A field dataset with 2D seismic angle gather sections and 7 wells along the line is used for demonstration purposes. A total of 140 intervals ranging from the minimum to maximum property values are used for each data type fed to the vertices, which make up of 140^3 combinations of possible scenarios of these three parameters. This makes the algorithm a perfect candidate for problems with limited dataset, but powerful and robust enough to account for uncertainties while still retaining the generalization outside of the training information.

3.2. Method

3.2.1. Neural Network and Deep Learning

A neural network is a system of interconnected units that simulate the neurons within human brains. The neurons are connected via communication links (weights), through which activation signal is passed from one to the other once an incoming signal is received (Figure 3.1). The neuron layer where data is fed into the system is called a visible layer. The other layer that serves to extract information from data is called hidden layer. Depending on the complexity of the data property, a network may have no or multiple hidden layers. A deep learning network system includes multiple layers of hidden neurons to extract hidden features or properties from the data (e.g, Goodfellow et al., 2016). The extracted relationship between input and output is contained within the weight terms, or the responses of the hidden neurons, depending on the nature of the data and the expectations of the algorithms.

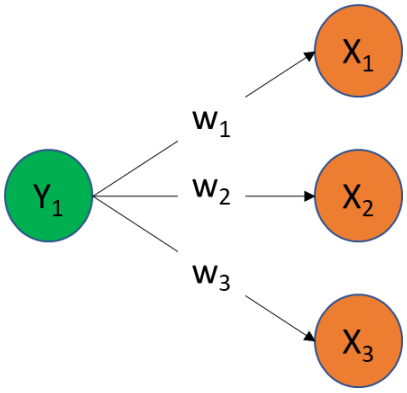

Figure 3.1: A simple neural network structure with 3 neurons $\mathbf{X}_1$, $\mathbf{X}_2$ and $\mathbf{X}_3$ receiving information from a neuron $\mathbf{Y}_1$ via communication links $\mathbf{w}_1$, $\mathbf{w}_2$ and $\mathbf{w}_3$. A supervised training will have labeled data fed into both neurons **Y** and **X**'s so that the extracted weight terms **w**'s represent possible relationship between **Y** and **X**. Meanwhile, an unsupervised training will have data fed into only neuron **Y**, so that the internal mechanism rearranges and optimizes output of neurons **X**'s and weights **w**'s to represent some features of input signal.

A neural network can be characterized by its architectures (the number of neurons and how they are connected to each other), the method of determining the connection weights (via training or optimization), and the output signal via activation function (Fausett, 2002). The training process could be either (1) supervised, where known input data and corresponding labels are used to guide the network structure toward a desired relationship; or (2) unsupervised where the algorithm works on its own to extract hidden information from the unlabeled input data. For an example, in Figure 3.1, a supervised training will have labeled data fed into both neurons ***Y*** and ***X****'s* so that the extracted weight terms ***w****'s* represent possible relationship between ***Y*** and ***X***. Meanwhile, an unsupervised training will have data fed into only neuron ***Y***, so that the internal mechanism rearranges and optimizes output of neurons ***X****'s* and weights ***w****'s* to represent some features of input signal.

3.2.2. Restricted Boltzmann Machine (RBM)

A Boltzmann machine is an energy based probabilistic deep learning network that is composed of two layers of neuron units which connects with each other via bidirectional weights. A restricted Boltzmann machine (RBM) by Smolensky (1986) is a variant of the single layer Boltzmann machine without the connections between neurons of the same group (Figure 3.2). The removal of the interconnection weights allows more efficient training such as the Contrastive Divergence algorithm of Hinton (2002). The learning process does not involve any back propagation, instead, requires random initializations of Markov chains to reach an equilibrium stage, at which the network is able to reconstruct the input data probabilistically.

$$E(\boldsymbol{v}, \boldsymbol{h}) = -\boldsymbol{v}^{\boldsymbol{T}}\boldsymbol{W}\boldsymbol{h} - \boldsymbol{\phi}_{\boldsymbol{v}}^{\boldsymbol{T}}\boldsymbol{v} - \boldsymbol{\phi}_{\boldsymbol{h}}^{\boldsymbol{T}}\boldsymbol{h}, \tag{3.1}$$

$$P(\boldsymbol{h}=\mathbf{1}|\boldsymbol{v})=\boldsymbol{\sigma}\left(\frac{\partial E}{\partial h}\right);\ P(\boldsymbol{v}=\mathbf{1}|\boldsymbol{h})=\boldsymbol{\sigma}\left(\frac{\partial E}{\partial v}\right), \tag{3.2}$$

$$\boldsymbol{\sigma}(x)=1/(1+\exp(-x)),$$

$$\Delta\boldsymbol{W}=\alpha(H_{data}[\boldsymbol{v}\boldsymbol{h}^{\boldsymbol{T}}]-H_{model}[\boldsymbol{v}\boldsymbol{h}^{\boldsymbol{T}}]). \tag{3.3}$$

For a binary network with M visible neurons $\boldsymbol{v}$ and N hidden neurons $\boldsymbol{h}$ connected by a weighting matrix $\boldsymbol{W}$, the energy function of the system is shown in Equation (3.1), where $\boldsymbol{\phi}_{\mathrm{v}}$ and $\boldsymbol{\phi}_{\mathrm{h}}$ are the threshold terms of the visible and hidden layers. The probability of the neuron to change its status are shown in Equation (3.2), where $\boldsymbol{\sigma}$ represents the logistic activation function. The update of the weighting matrix $\boldsymbol{W}$ is not a back propagation process, but is governed by the difference between the data-dependent expectation ($\boldsymbol{H}_{data}$) and the model expectation ($\boldsymbol{H}_{model}$) scaled with a learning constant α (Equation 3.3). The expectation values are calculated using the Markov chain random initializations to find the best fit responses to the input data.

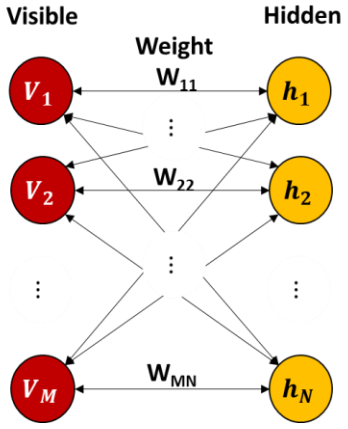

Figure 3.2: A schematic diagram of a restricted Boltzmann machine. The data is fed into visible layer with red neurons $\mathbf{V_i}$, fully interconnected with a hidden layer of orange neurons $\mathbf{h_i}$ via a weighting matrix $\mathbf{W}$.

3.2.3. Cross-shape Deep Boltzmann Machine

A deep Boltzmann machine (DBM – Salakhutdinov and Hinton 2009) is a multi-layer probabilistic network consisting of multiple layers of RBM for capturing hidden patterns within an unlabeled dataset. The layers are linked together with bidirectional connections, which allows the DBM to be more robust when incorporating uncertainty about ambiguous inputs.

A cross-shape deep Boltzmann machine (CDBM) is created by connecting four different RBMs at the vertices via a hidden neuron layer located at the center of the cross (Figure 3.3). This arrangement allows the center units to distribute changes of input data at any visible vertex throughout the network toward remaining vertices. In a pre-stack inversion problem for petrophysical properties, the CDBM network structure simulates simultaneous interactions between the rock properties and the seismic amplitudes: any changes in any output term are transmitted and results instantly in a change in seismic amplitude.

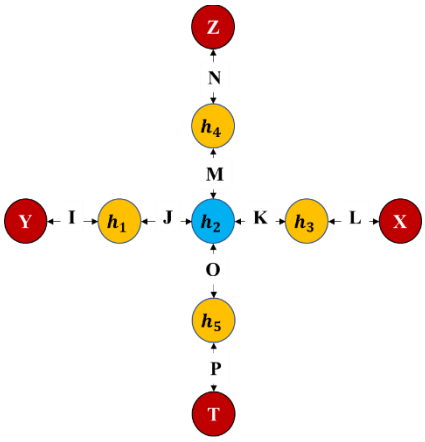

Figure 3.3: The schematic diagram of the cross-shape deep Boltzmann machine with four red visible neuron layers {**X**, **Y**, **Z**, **T**} in the outer area; and 5 hidden layers {**h_1, h_2, h_3, h_4, h_5**} toward the center. The neuron layers are connected with bidirectional weighting matrices **I, J, K, L, M, N, O** and **P**. Hidden neuron **h_2** (in blue) reflects with simultaneous interactions between visible input neurons.

The four RBM includes four visible layers {***X, Y, Z, T***}, and five hidden layers {$\boldsymbol{h_1, h_2, h_3, h_4, h_5}$} located in the inward direction, as shown in Figure 3.2. The hidden layer $\boldsymbol{h_2}$ in the center of the cross instantly transmits changes of any input into other visible layers via the bidirectional weight terms ***I, J, K, L, M, N, O, P***. When data is fed into visible neuron layers, the network transmits information inward to the center via $\boldsymbol{h_1}$, $\boldsymbol{h_3}$, $\boldsymbol{h_4}$ and $\boldsymbol{h_5}$. When incoming signals arrive at layer $\boldsymbol{h_2}$, these neurons handle and send back response signal outward to the vertices in the opposite direction. This back-and-forth data traveling are repeated until an equilibrium stage is achieved.

$$E = -\boldsymbol{Y^T I h_1} - \boldsymbol{h_1^T J h_2} - \boldsymbol{h_3^T K h_2} - \boldsymbol{X^T L h_3} - \boldsymbol{h_4^T M h_2} - \boldsymbol{Z^T N h_4} - \boldsymbol{h_5^T O h_2} - \boldsymbol{T^T P h_5} - \boldsymbol{\phi_Y^T Y} - \boldsymbol{\phi_X^T X} - \boldsymbol{\phi_Z^T Z} - \boldsymbol{\phi_T^T T}. \quad (3.4)$$

The general form of the energy or loss function for this design is shown in Equation (3.4), where $\boldsymbol{\phi_i}$ represents the threshold of the layer noted in subscription. The learning process of this network is described by Equations (3.5) – (3.14). The training data is first input into the visible layers, after which the hidden neurons start to interact to extract the patterns from dataset. Major steps in the algorithm include:

- Calculating the data-dependent expectation by fixing the visible neurons with input data and estimating hidden neuron responses until equilibrium.
- Estimating the model expectation by running multiple Gibbs sampling on all neuron responses to capture the uncertainties in model parameters.

$$P(\boldsymbol{h_i} = \boldsymbol{1}|\boldsymbol{v}) = \sigma\left(\frac{\partial E}{\partial \boldsymbol{h_i}}\right); \text{i} = [1,5] \quad (3.5)$$

$$P\left(\boldsymbol{v_j} = \boldsymbol{1}\middle|\boldsymbol{h}\right) = \sigma\left(\frac{\partial E}{\partial \boldsymbol{v_j}}\right); j = [1,4] \quad (3.6)$$

$$\Delta \boldsymbol{I} = \alpha\left(H_{data}\left[\boldsymbol{Y h_1^T}\right] - H_{model}\left[\boldsymbol{Y h_1^T}\right]\right) \quad (3.7)$$

$$\Delta \boldsymbol{J} = \alpha\left(H_{data}\left[\boldsymbol{h_1 h_2^T}\right] - H_{model}\left[\boldsymbol{h_1 h_2^T}\right]\right) \tag{3.8}$$

$$\Delta \boldsymbol{K} = \alpha\left(H_{data}\left[\boldsymbol{h_3 h_2^T}\right] - H_{model}\left[\boldsymbol{h_3 h_2^T}\right]\right) \tag{3.9}$$

$$\Delta \boldsymbol{L} = \alpha\left(H_{data}\left[\boldsymbol{X h_3^T}\right] - H_{model}\left[\boldsymbol{X h_3^T}\right]\right) \tag{3.10}$$

$$\Delta \boldsymbol{M} = \alpha\left(H_{data}\left[\boldsymbol{h_4 h_2^T}\right] - H_{model}\left[\boldsymbol{h_4 h_2^T}\right]\right) \tag{3.11}$$

$$\Delta \boldsymbol{N} = \alpha\left(H_{data}\left[\boldsymbol{Z h_4^T}\right] - H_{model}\left[\boldsymbol{Z h_4^T}\right]\right) \tag{3.12}$$

$$\Delta \boldsymbol{O} = \alpha\left(H_{data}\left[\boldsymbol{h_5 h_2^T}\right] - H_{model}\left[\boldsymbol{h_5 h_2^T}\right]\right) \tag{3.13}$$

$$\Delta \boldsymbol{P} = \alpha\left(H_{data}\left[\boldsymbol{T h_5^T}\right] - H_{model}\left[\boldsymbol{T h_5^T}\right]\right) \tag{3.14}$$

Reorganizing the visible layers as $\boldsymbol{v} = \{\boldsymbol{X, Y, Z, T}\}$, the hidden layers as $\boldsymbol{h} = \{\boldsymbol{h_1, h_2, h_3, h_4, h_5}\}$ and the weighting matrices as $\boldsymbol{W} = \{\boldsymbol{I, J, K, L, M, N, O, P}\}$. The hidden neuron responses are calculated by following the order of data transmission within the network, which is:

1. Update hidden neurons $\boldsymbol{h_1}$, $\boldsymbol{h_3}$, $\boldsymbol{h_4}$, $\boldsymbol{h_5}$ using Equation (3.5),
2. Update hidden neurons $\boldsymbol{h_2}$ using Equation (3.5),
3. Update visible neuron layers using Equation (3.7),
4. Update weighting terms using Equations (3.7) - (3.14),
5. Update hidden neurons and weighting terms until equilibrium is achieved.

One significant advantage of the CDBM network, besides the structural framework, is its capacity in accounting for what-if scenarios the training data. Unlike other supervised deep learning algorithm, the CDBM only requires the upper and lower bounds of input labeled data so that the internal mechanism automatically generate all possible combinations of the data fed to the vertices.

3.2.4. CDBM Implementation in Inversion for Petrophysical Properties

The CDBM is implemented into an inversion problem to predict the shale volumetric, porosity and water saturation from pre-stack angle gathers. This supervised

training is set-up to capture the nonlinear relationship between seismic data and rock property terms. Once the network is trained, it can predict properties away from the wells.

Let the seismic gather be input $\boldsymbol{Y}$, while Vshale, $\boldsymbol{\phi}$ and Sw are fed to vertices $\boldsymbol{X}$, $\boldsymbol{Z}$ and $\boldsymbol{T}$. The training efficiency is quantified by a least square misfit function of all inputs, as shown in Equation (3.15). Expanding this equation, and replacing the terms $\boldsymbol{h_1}$, $\boldsymbol{h_3}$, $\boldsymbol{h_4}$ and $\boldsymbol{h_5}$ with corresponding function from $\boldsymbol{h_2}$ (Equation 3.16), and the intermediate terms such as $\bar{\boldsymbol{J}}$, $\bar{\boldsymbol{K}}, \bar{\boldsymbol{O}}, \boldsymbol{\phi_X} = \frac{\mathbf{1}}{\mathbf{2}}\boldsymbol{X}; \boldsymbol{\phi_Y} = \frac{\mathbf{1}}{\mathbf{2}}\boldsymbol{Y}$, $\boldsymbol{\phi_Z} = \frac{\mathbf{1}}{\mathbf{2}}\boldsymbol{Z}, \boldsymbol{\phi_T} = \frac{\mathbf{1}}{\mathbf{2}}\boldsymbol{T}$, the final misfit function is obtained, as in Equation 3.17, and shares the same form with the general misfit function in Equation 3.6.

$$E = \frac{1}{2}\|\boldsymbol{Y} - \boldsymbol{I}\boldsymbol{h_1}\|_2 + \frac{1}{2}\|\boldsymbol{X} - \boldsymbol{L}\boldsymbol{h_3}\|_2 + \frac{1}{2}\|\boldsymbol{Z} - \boldsymbol{N}\boldsymbol{h_4}\|_2 + \frac{1}{2}\|\boldsymbol{T} - \boldsymbol{P}\boldsymbol{h_5}\|_2 \tag{3.15}$$

$$\text{Where } \boldsymbol{h_1} = \boldsymbol{J}\boldsymbol{h_2}; \boldsymbol{h_3} = \boldsymbol{K}\boldsymbol{h_2}; \boldsymbol{h_4} = \boldsymbol{M}\boldsymbol{h_2}; \boldsymbol{h_5} = \boldsymbol{O}\boldsymbol{h_2}, \tag{3.16}$$

$$\begin{aligned} E = -\boldsymbol{Y}^T\boldsymbol{I}\boldsymbol{h_1} \pm \boldsymbol{h_1^T}\bar{\boldsymbol{J}}\boldsymbol{h_2} - \boldsymbol{h_3^T}\bar{\boldsymbol{K}}\boldsymbol{h_2} - \boldsymbol{X}^T\boldsymbol{L}\boldsymbol{h_3} - \boldsymbol{h_4^T}\bar{\boldsymbol{M}}\boldsymbol{h_2} - \boldsymbol{Z}^T\boldsymbol{N}\boldsymbol{h_4} - \\ \boldsymbol{h_5^T}\bar{\boldsymbol{O}}\boldsymbol{h_2} - \boldsymbol{T}^T\boldsymbol{P}\boldsymbol{h_5} + \boldsymbol{\phi_Y}^T\boldsymbol{Y} + \boldsymbol{\phi_X}^T\boldsymbol{X} + \boldsymbol{\phi_Z}^T\boldsymbol{Z} + \boldsymbol{\phi_T}^T\boldsymbol{T}, \end{aligned} \tag{3.17}$$

where

$$\bar{\boldsymbol{J}} = -\frac{1}{2}\boldsymbol{J};\ \bar{\boldsymbol{K}} = -\frac{1}{2}\boldsymbol{K}; \bar{\boldsymbol{M}} = -\frac{1}{2}\boldsymbol{M}; \bar{\boldsymbol{O}} = -\frac{1}{2}\boldsymbol{O},$$

$$\Delta\ \boldsymbol{\phi_X} = \frac{\mathbf{1}}{\mathbf{2}}\boldsymbol{X}; \boldsymbol{\phi_Y} = \frac{\mathbf{1}}{\mathbf{2}}\boldsymbol{Y}, \boldsymbol{\phi_Z} = \frac{\mathbf{1}}{\mathbf{2}}\boldsymbol{Z}, \boldsymbol{\phi_T} = \frac{\mathbf{1}}{\mathbf{2}}\boldsymbol{T}.$$

3.3. Examples

The designed network is applied to a 2D field dataset (Figure 3.4) for demonstration purposes. Seven vertical wells are located along the line. All wells come with logs fully covering most of the target intervals shown with low impedances between stiffer intervals of high impedance values, and outstandingly high gamma ray reading, while leaving out the underlying sections in some wells. The pre-stack gathers were converted into angle gathers, ranging from 0 to 28 degrees at 5 degrees increment to ensure the highest correlation between well logs and seismic data. The petrophysical logs are calculated using

common approaches: total porosity is calculated from the density log with a sandstone base line; Vshale is extracted by normalizing the gamma ray index, and water saturation is calculated using the Wyllie time average rock physics model. The target zone is highlighted by a hot shale interval. To examine the generality of the network, recorded seismic data and logs at Well 2 are used for blind test. This well is located in the middle of the line, with full data coverage of the target intervals.

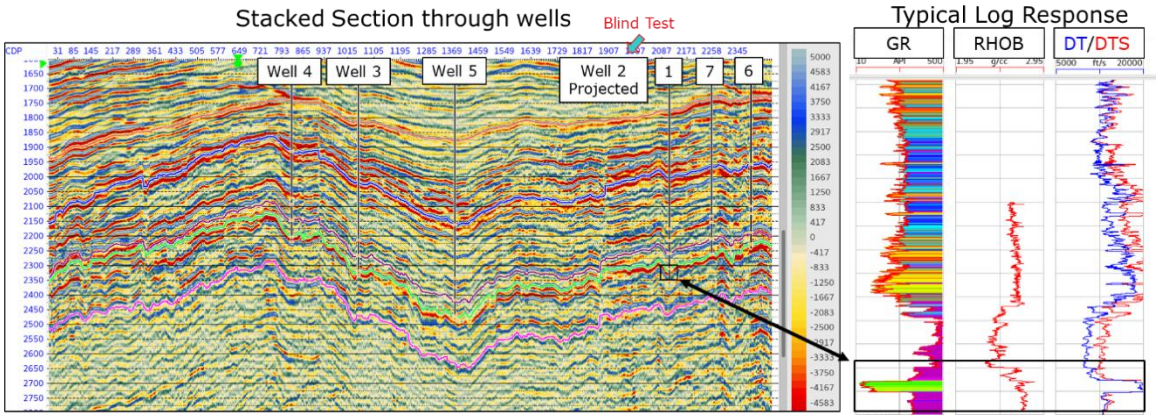

Figure 3.4: The seismic stack section through all wells within the 2D field data line. The available logs include the gamma ray (GR), density (RHOB), sonic (DT) and shear (DTS) velocity, and resistivity. The target is a hot shale interval with outstandingly high GR values (above 175 API) of softer material (smaller density and sonic readings than overlaid stiffer materials. Correspondingly, the top of hot shale interval is a strong trough event on the seismic stacked section.

The CDBM is ideal for this dataset, due to the limited number of wells available, which restricted the amount of data that could be used for training any conventional deep learning system. A total of 140 intervals ranging from the minimum to maximum property values are used for each data type fed to the vertices, which comprise 140^3 combinations of possible scenarios of these three parameters.

Following the same implementation procedure, the data is fed into the deep learning system at visible vertices. All hidden layers are set with the same number of units. A small

learning rate of 0.001 is used to accommodate the updating of all weight matrices. The network reaches an equilibrium stage when the overall error value is flattened after 10 epochs (Figure 3.5). At this stage, the network successfully reconstructed all peaks and troughs of the input data (Figure 3.6-3.9). In detail, reconstructed angle gathers, porosity and water saturation well overlay the real values at all visible events, with minor differences in amplitudes at Well 2 and Well 3 locations. Meanwhile, the machine underpredicts the hot shale interval in most of the wells. This is reasonable considering the seismic signature of this zone is similar to the signature of a clean sandstone interval: strong trough overlaying peak events. Also, the inconsistent data coverages of the environment below the target interval strongly affect the algorithm's decision in predicting the features there.

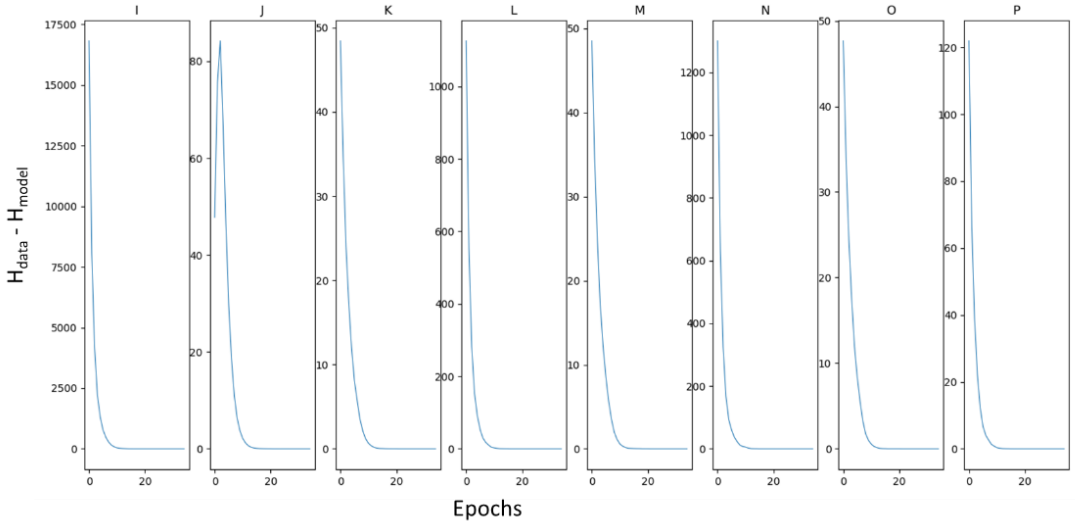

Figure 3.5: The plot of update values for the internal weighting **I, J, K, L, M, N, O, P** illustrate the maturation of this CDBM network toward stability after 20 epochs.

The trained network is applied to the entire 2D line to predict petrophysical sections (Figure 3.10-3.12). The resulting porosity section shows likely prospective intervals of high values at the bottom of the target. Meanwhile, the predicted Vshale and water

saturation sections suffer lateral resolution where multiple vertical stripes strongly occupy the display along with the geological sound events. These are due to insufficient data coverage to represent features outside of the target zones.

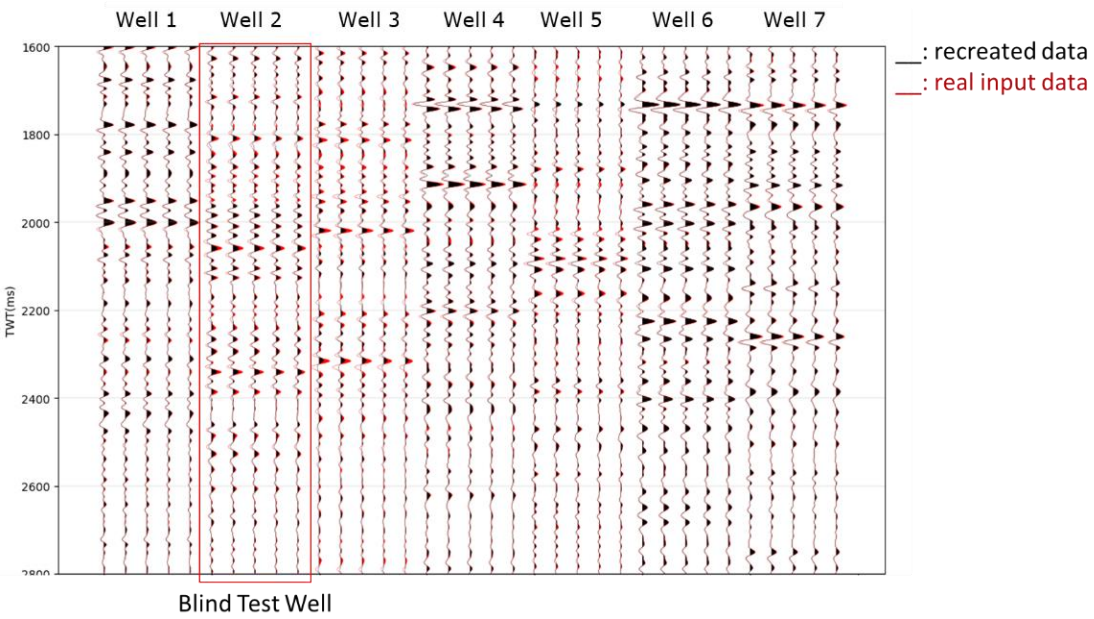

Figure 3.6: The comparison between reconstructed (black) and real seismic gathers that were used to train the network. Notice that Well 2 was not used for training, and the trained network was able to recreate the data at this location.

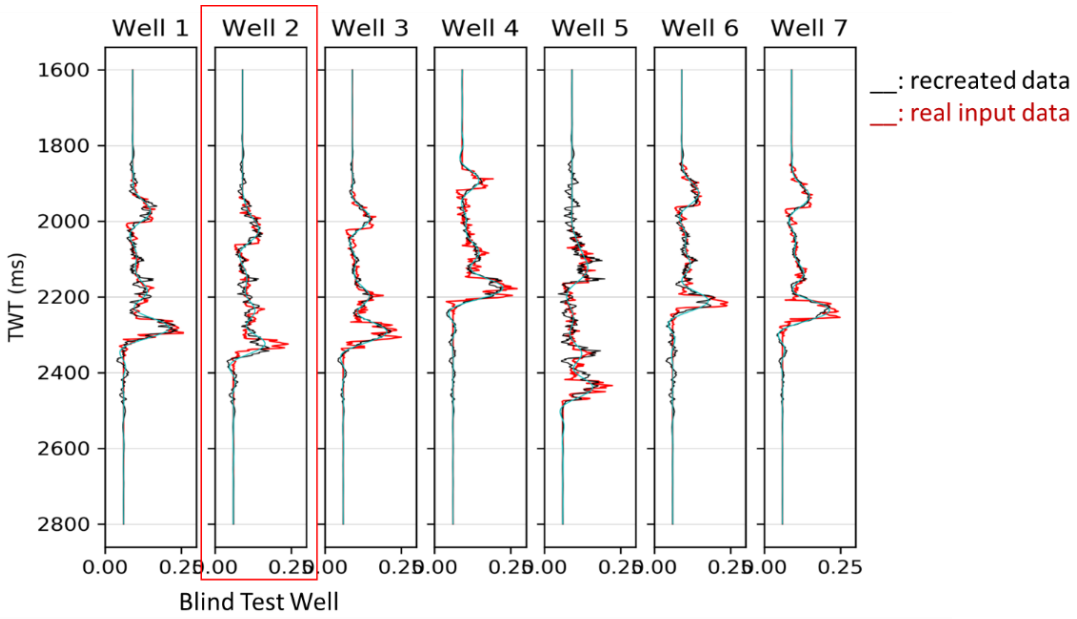

Figure 3.7: The comparison between reconstructed (black) and real total porosity values that were used to train the network. Notice that Well 2 was not used for training, and the trained network was able to recreate the data at this location.

3.4. Summary

This study introduces a newly designed cross-shape deep Boltzmann machine to capture the non-linear relationship between seismic amplitude and rock properties in the pre-stack inversion problem. The network includes four separated vertices of visible neuron layers following four hidden layers, all of which are connected by a hidden neuron at the center of the cross-shape.

For the prestack inversion for petrophysical properties, this deep learning algorithm requires four different input data types, including a seismic amplitude, and three petrophysical logs such as the porosity, water saturation and shale volumetric for training. Inherently from the deep Boltzmann family, the CDBM, unlike any conventional deep learning system, which expects large data for all what-if scenarios into training, only requires the anticipated upper and lower bounds of the data and labels, which could be used by the internal mechanism to account for any possible value combinations.

In the demonstration, 2D dataset with 7 well controls with limited data coverage around the low impedance hydrocarbon bearing shale interval overlaid by stiffer sealing shales is used. From the property bounds, the CDBM generates a total of 140 intervals ranging from the minimum to maximum property values for each data type to feed to the vertices, which make up of 140^3 combinations of possible scenarios of these three parameters. The trained network was able to recognize and replicate the majority of features in the original logs, except for the shale volumetric readings of the hydrocarbon shale interval. This is due to the similarity in seismic signatures of the hot shale and sandstone overlaid by similar stiffer materials, and the complication of labeling the shale volumetric values to better distinguish between the sealing and prospective shale.

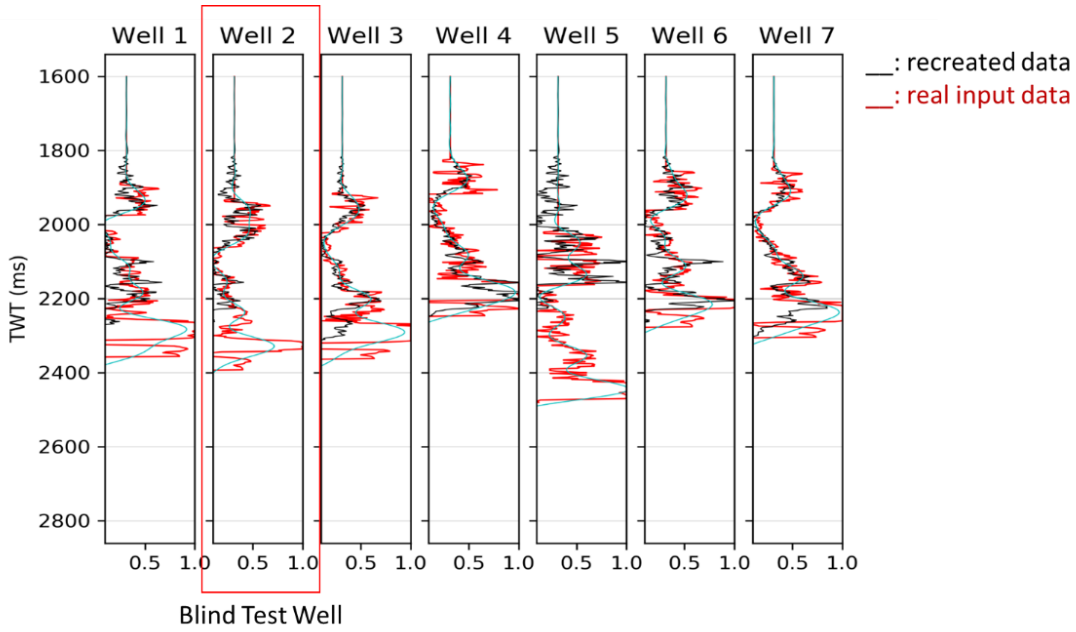

Figure 3.8: The comparison between reconstructed (black) and real shale volumetric (Vshale) values that were used to train the network. The network underestimates the Vshale readings of hot shale interval, which is expected as the hot shale is having similar seismic characterization as porous sandstones of other intervals.

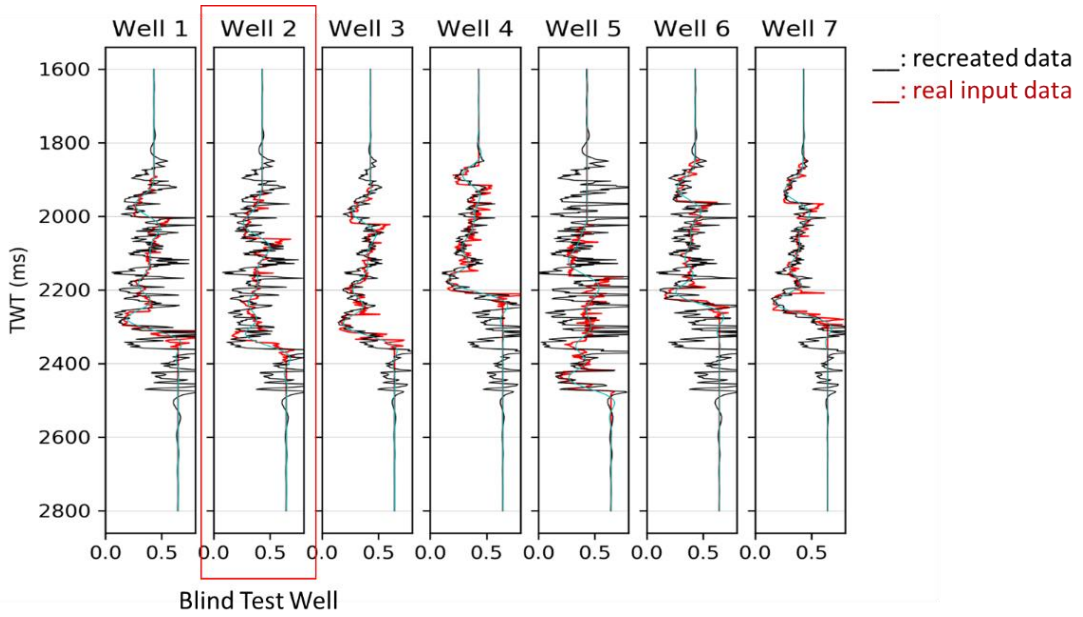

Figure 3.9: The comparison between reconstructed (black) and real water saturation values that were used to train the network. The network over-estimates the water saturation readings in some wells, such as Well 5.

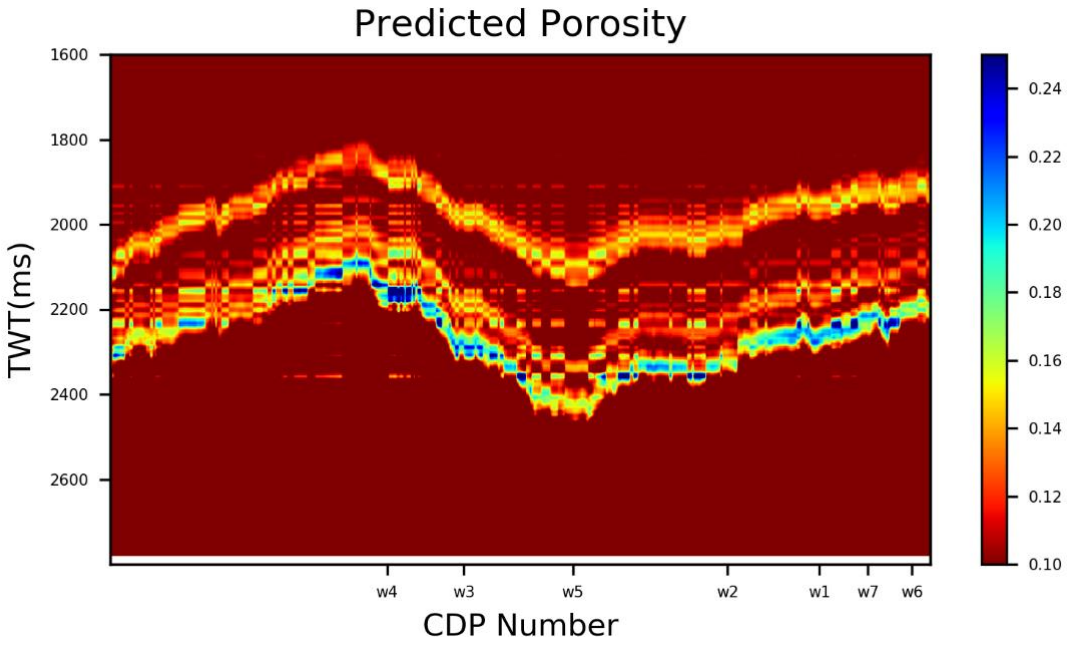

Figure 3.10: The predicted porosity section from trained CDBM network.

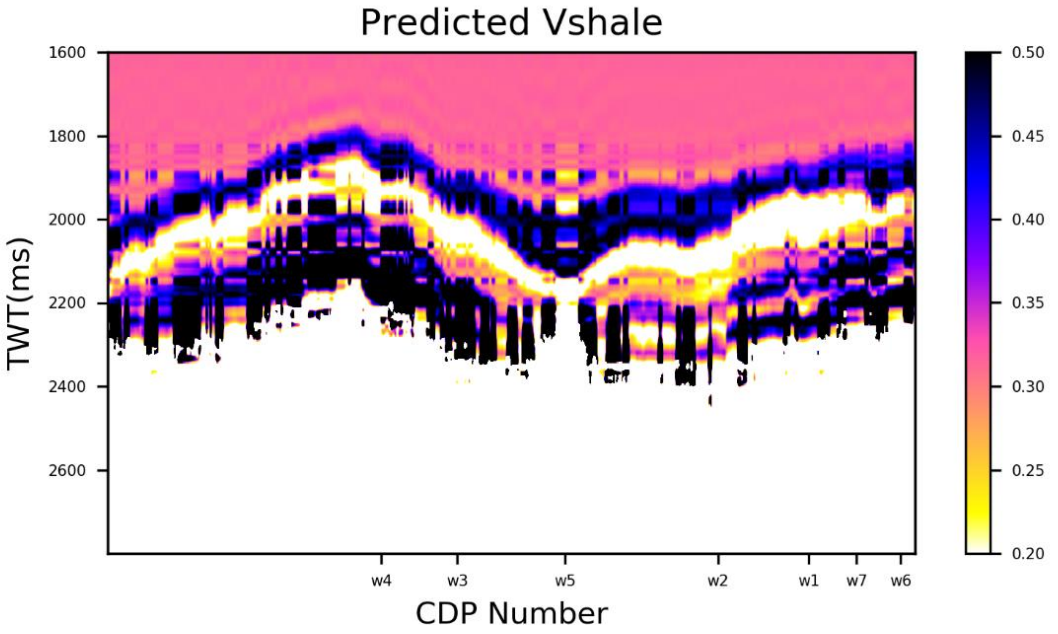

Figure 3.11: The predicted shale volumetric section from trained CDBM network.

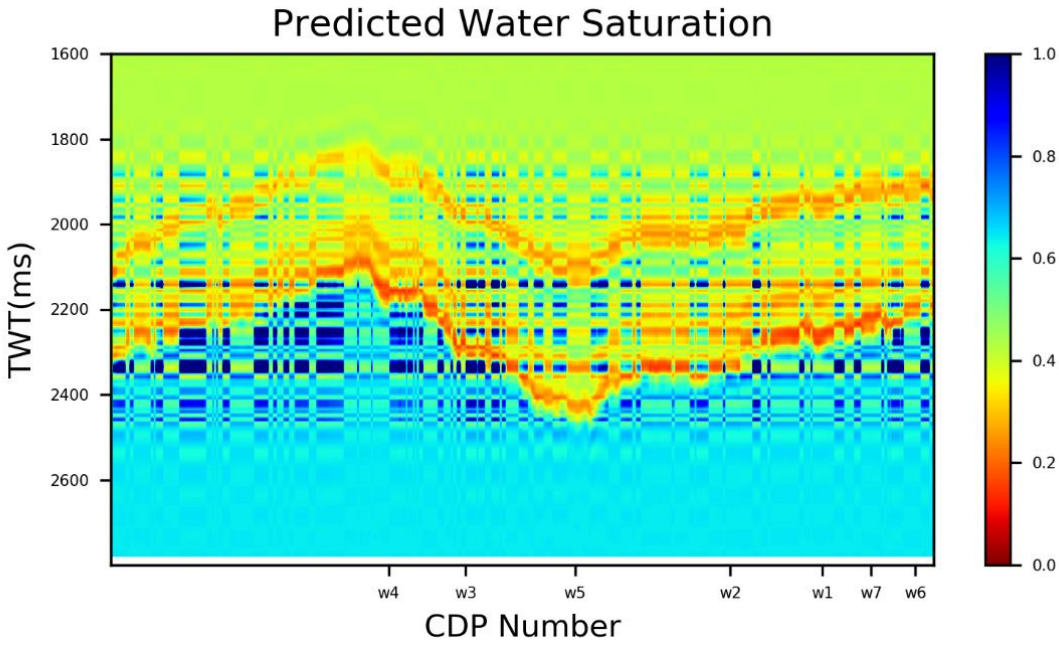

Figure 3.12: The predicted water saturation section from trained CDBM network. Notice the patchy results along the line, which insufficient training data.

Chapter 4 : Quantifying Uncertainty in AVA Inversion Using Deep Learning

Manuscript submitted to Geophysics.

4.1. Introduction

Recorded seismograms are an important source of information to understand subsurface rock properties. Pre-stack inversion is a popular seismic reservoir characterization method to extract the subsurface properties from the recorded seismograms. On a seismic section with normal polarization, the positive amplitudes are at the transition boundaries from softer to stiffer materials, and vice versa for the negative amplitudes. However, the inverse problem usually suffers from non-uniqueness due to the fact that multiple combinations of elastic properties may be able to match the observed data, the noise and measurement errors (Brown, 1998).

A set of recorded logs in depth domain from drilling locations is used to guide the inversion process and reduce uncertainties. They are first converted to time domain via seismic well-tie process (Walden and White, 1984; White and Shimm, 2003) before being used for further analyses. There lies a huge discrepancy in resolution between the seismic signals and well log events: a 2 milli-second two-way-time interval of 3,000 meters per second acoustic velocity is equal to 30 meters of depth, which could correlate within an interval of 200 recorded log values sampled at 15 centimeters spacing. As a result, matching a single log value to a single amplitude is not enough, unless the interval is completely homogenous. A probabilistic relation expressing all possible depictions of such a situation is more reasonable. A good prediction of this posterior distribution provides better understanding of the uncertainty in the inversion results associated with the seismic amplitudes.

Many studies have been carried out to quantify the uncertainty of predicted results using some statistical methods, such as the Bayesian approach (Sen and Stoffa, 2013; Mukerji et al., 2001; Buland et al. 2003; Tarantola, 2005), which combines the prior information on the target model with the likelihood function that links the model parameters to the observed data. The prior information is estimated from the conceptual models of existing well data analyses or inferred from regional studies. The likelihood function represents the physical relationship between the model parameters and recorded seismic signal, such as the full-wave equation, or the Zoeppritz equation and its linearized approximation (Aki and Richards, 1980). The resulting posterior distribution calculated from the seismic sections allow quantifying the uncertainty of the inverse problem.

The prediction of elastic properties and their uncertainties are challenging due to the complexity of the geological background, and the choice of an algorithm itself. For a full-wave inversion algorithm (Virieux et al., 2017), the complete physics of the wave propagation is integrated into the forward modeling, and the solution is obtained via an iterative gradient-based model update schematic to minimize the misfit between synthetic and observed data. The result will be accurate if the starting model is within the vicinity of the true model. However, this method is computationally expensive, especially for large datasets. Meanwhile, the amplitude variation with offset (AVO) inversion, which takes advantage of the linearized Zoeppritz equation for reflectivity calculation and a convolution model to convert reflectivity to seismic amplitudes, is generally faster, but limited to small incident angles and the non-converted signals (reflection and transmission).

Artificial neural networks have become powerful computational tools with successful applications in many fields of studies. A network structure is built on a system of neurons, which transmits signals via connection units once information is passed into

the system. A network structure is capable of capturing hidden information within the data, provided with a proper structure and suitable training schematic to best exploit training information (Fausett, 2002). Deep learning is one artificial neural network algorithm that employs the computing powers of processing units to extract hidden information from the input dataset. By properly manipulating and organizing seismic data and the elastic property logs into an input dataset, the deep learning can be taught to establish a mathematical sound relationship that could later be used to predict the distribution of elastic properties from a seismic section.

To map the posterior distribution of any multi-label output, a regular deep learning system requires the calculation a partition function. This process involves a heuristic backpropagation process to estimate all possible labels from the input data, and use those to calculate the probability of individual label within the output layer. However, the exhaustive computation associated with the backpropagation can be shown to be susceptible to false minima (Specht, 1990). For a single layer network, such as Hopfield neural network, the partition function can be approximated by using a sigmoid activation function (e.g, Bilbro et. al, 1990). Meanwhile, for a multi-layer deep learning system such as the probabilistic neural network (Specht,1990), the probability map of the outputs can be made using an exponential function that can compute nonlinear decision boundaries which approach the Bayes optimal.

This study examines a complete alternative approach to predict the posterior distribution of elastic properties from seismic data, without the need of directly calculating the partition function or using modified activation function to approximate it. Instead, the expected outputs in form of time varying probability maps of the acoustic and shear impedances, and density terms are used as training labels and the deep learning system is designed to establish the nonlinear function that maps these labels with the training data,

which is prestack seismic amplitude. We assume that the recorded well log and seismic data are of reliable quality. Two major steps are involved in this study: (1) generating training data and labels, and (2) designing a deep learning system to extract the desired relationship. The training data, which includes the seismic angle gathers, are transformed via wavelet transform operation to generate extra information that leads to better extraction of hidden features than the regular amplitude data from raw angle stacks. Meanwhile, the training labels, which are the measurements of the acoustic impedance (Zp), shear impedance (Zs) and density (RHOB) are resampled from depth to time domain, and mapped to probability distribution to accommodate the deep learning process. Of note, the multimodal network design allows multiple output labels to be input into the system, without the need of complicated merging and relocating different label values as needed in a common deep learning process with single output. The algorithm is demonstrated with a field dataset that includes 2D seismic gathers and multiple wells along the line.

4.2. Methodology

4.2.1. Deep Learning Structure with Convolution Network

Convolution neural network (CNN - Lecun et al.,1995) is a popular neural network structure to analyze hidden features from input data. Unlike a regular neural network, which applies a linear transformation to the input layer, a CNN performs the convolution operation on a pre-set window size that slides through the extent of the input signal.

A common deep learning structure with convolution network implementation includes one or multiple packages of CNN-MaxPooling-Dropout sequence, as depicted in Figure 4.1. The CNN layer extracts hidden features within input data, while MaxPooling and Dropout improve the performance of the network. The extracted features are flattened into a column (or row) format that is then fed into a system of interconnected layers to simulate the values of the input labels.

The network accuracy is measured by how close the energy function (or loss function) is to the global minimum value, while its generalization is determined by how well it could predict the independent dataset that is not involved in the training process. A properly trained neural network should be able to generalize the expected property and make accurate predictions. A network is called underfitted if it is trapped at a local minimum during the training process. A network which has high prediction capability, but poor generalization is considered an overfitted one.

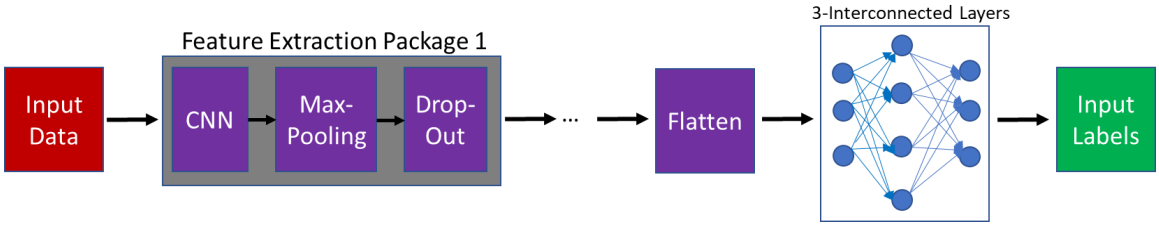

Figure 4.1: A common deep learning structure with convolution network implementation.

While an underfitted network can be improved by choosing a different type of optimization algorithm, or providing more data, an overfitted network is more complicated. A common approach to deal with such challenge is to split the data into two parts: training and testing (Reitermanova, 2010). The system is first trained with the training dataset, and then cross-validated (Picard et al., 1984) with testing dataset. The matured network will ensure the energy functions of both training and testing datasets reach stable solution when reaching a global minimum stage (Figure 4.2). An overfitted network could have a stable loss curve of the training data, and an overturned test loss curve.

Overfitting is a common problem in deep learning. Several data manipulation techniques have been developed to deal with this challenge. Pooling is a technique to create feature maps that summarize all features in the input data. The common ones are MaxPooling (getting the maximum value), or AveragePooling (taking the average value).

Dropout (Srivastava et al., 2014) is a technique to prevent overfitting when training a neural network by introducing noise to the learning process. At each epoch, a population of features is randomly chosen and dropped from the network training process. This has been proven to significantly reduce overfitting and improve the network performance.

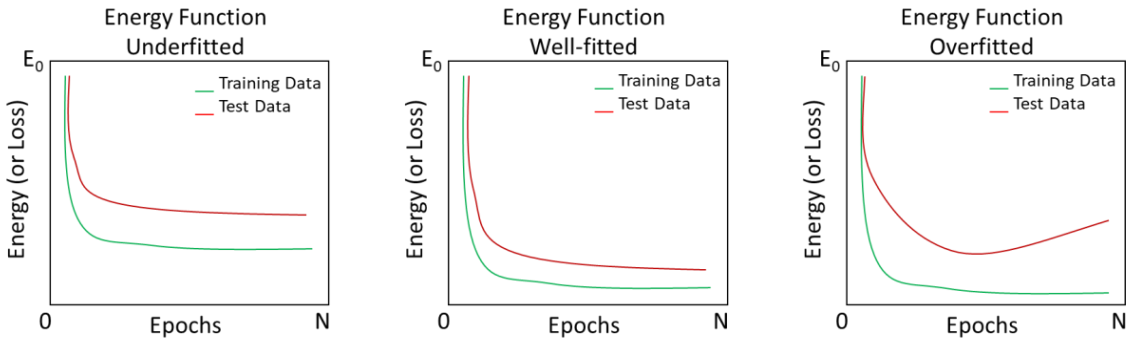

Figure 4.2: The determination of network quality by looking at the energy (loss) functions of the training and testing dataset. A matured network will ensure the energy functions of both training and testing datasets reach stable solution when reaching a global minimum stage.

4.2.2. Multi-modal Design for Pre-stack Seismic Inversion

The multimodal deep learning design aims at separating different data label types by locating them in different output modals (e.g, Srivastava et al, 2014, Phan and Sen, 2019). In the pre-stack seismic inversion problem, the data labels include P-impedance, S-impedance and density, which consequently requires three output branches connecting to the main feature extraction flow at a flattening node, as shown in Figure 4.3.

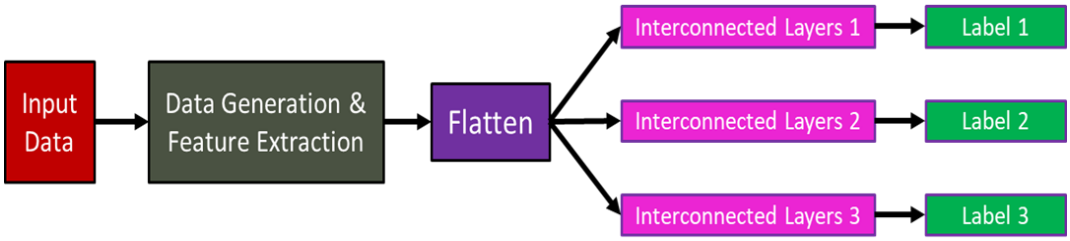

Figure 4.3: The general schematic diagram of the multimodal deep learning network to predict the posterior distribution (e.g probability) of the N different categorical data labels.

Unlike the common deep learning flow described in the previous section, the input labels do not need be merged into one layer. This allows flexible manipulations of the label data, where users can combine their interpretation (e.g., prior knowledge) into training the network. Consider the case where unknown data $\boldsymbol{X = (X_1, X_2, X_3)}$ are three target logs from a measured data $\boldsymbol{Y}$, and the weight terms $\boldsymbol{W}$ represent the connections between data and labels. The loss (energy) function of the common deep learning design measures the difference between the labels and network prediction results (Equation 4.1). Meanwhile, the loss function of the multimodal design deep learning is a weighted sum of all individual parameters (Equation 4.2).

$$E(\boldsymbol{W}, \boldsymbol{X}|\boldsymbol{Y}) \sim \left\|\boldsymbol{X} - \boldsymbol{X}^{\boldsymbol{predict}}\right\|, \tag{4.1}$$

$$E(\boldsymbol{W}, \boldsymbol{X}|\boldsymbol{Y}) \sim \alpha_1\left\|\boldsymbol{X_1} - \boldsymbol{X_1^{predict}}\right\| + \alpha_2\left\|\boldsymbol{X_2} - \boldsymbol{X_2^{predict}}\right\| + \alpha_3\left\|\boldsymbol{X_3} - \boldsymbol{X_3^{predict}}\right\|, \tag{4.2}$$

where α_i is the weighted term of the unknown data

4.2.3. Data preparation

The training data is fed into the input layer, while labels are fed into the output layer. The preparation of these data is essential to the overall performance of the network. This section briefly introduces the additional processing steps to generate training data

from angle gather seismograms, as we create labels in form of the prior distributions of elastic properties.

Training Data Generation with Continuous Wavelet Transform

The continuous wavelet transform (CWT) is an invertible mathematical operation to examine the similarity of an input signal to a model function. Popular applications include image compressions, edge and corner detection, solving partial differential equations, texture analyses (Mallat, 1999). In reflection seismology, considering varying the scale factor, the CWT is applied to decompose the seismograms into different corresponding contents, which consequently increase the amount of data available for deep learning purposes.

Let *f(t)* be the time varying seismic signal, and *ψ(t)* be the model function, which is usually chosen as a wavelet signal. The CWT transformation is expressed in Equation (4.3), where $b \in \mathcal{R}$ is the translational term, which relates to the phase of the model function, and $a \in \mathcal{R}$ is the scale factor, which determines the width of the model function. The model function is called a mother wavelet when $b = 0$ and $a = 1$.

$$X(a, b, f, \Psi) = \frac{1}{a^{1/2}} \int_{-\infty}^{\infty} f(t) \Psi \left(\frac{t-b}{a} \right) dt. \tag{4.3}$$

Theoretically, the translational and scaler terms can take any value. In reality, these can be determined from the content of input seismic data. The translational term is directly related to the phase of the statistical wavelet extracted on the target interval, while the scaler term corresponds to the frequency content of the wavelet: the higher the frequency, the narrower the model function, which reflects smaller scaler value; and vice versa. Figure 4.4 shows the CWT section (with b = 0 and $a \in [5,25]$) of a single seismogram trace.

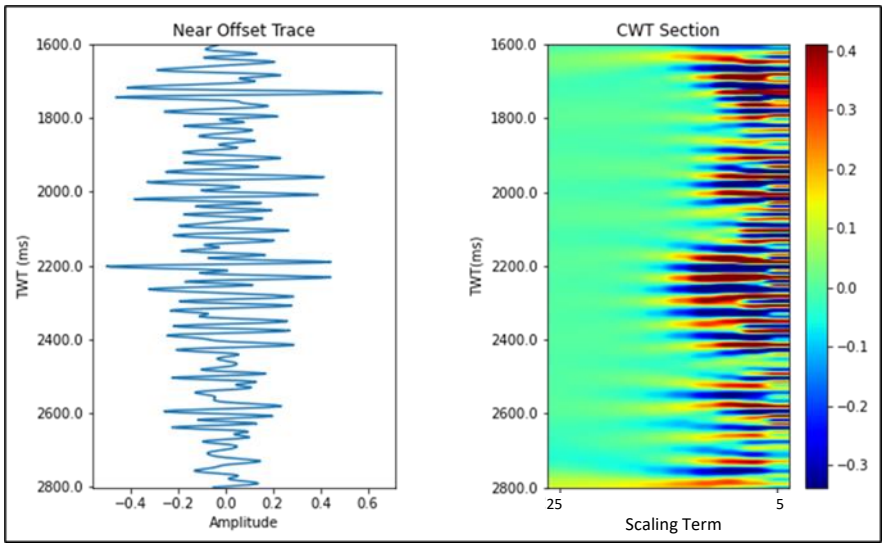

Figure 4.4: The continuous wavelet transform of a single seismic trace with a Ricker wavelet and zero transition and a scaling term ranging from 5 to 25.

Creating Probability Distribution as Training Label

The training labels are required to be in forms of probability distribution. The conversion from a single value to a distribution is described in Figure 4.5, where the log is first normalized to fall within a desired range that can best describe the data, and then categorized into binary distribution by setting a value 0 or 1 at the proper bin of the original value.

$$Zp_{norm} = \frac{Zp - Zp_{min}}{\Delta Zp}, \tag{4.4}$$

$$Zs_{norm} = \frac{Zs - Zs_{min}}{\Delta Zs}. \tag{4.5}$$

For this particular study where P-, S- impedances and density are label data, the desired range is determined to cover the minimum and maximum values of the corresponding types. The normalization process is performed by subtracting the impedance log with their expected minimum value (Equation 4.4-4.5) to reduce the range of the data

coverage. Then, the categorization process converts the normalized log readings into a distribution by putting binary values (0 or 1) at the corresponding bins.

Example: Value = 0.65

Bin number	0	...	65	...	100
Binary Value	0	0	1	0	0

Figure 4.5: An illustration of the categorization process to convert a regular log into a spiky distribution. A log value of 0.65 is decomposed into a 101-bin probability distribution with a spike value (1) at the sixty-sixth bin and flat 0 value at the remaining locations.

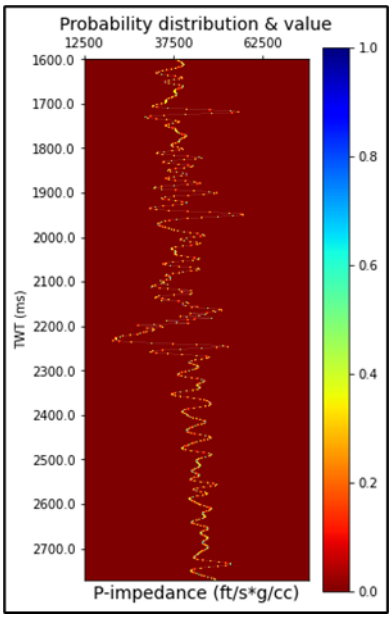

Figure 4.6: Spiky distribution from synthetic model due to one-to-one correspondence between the model and time sample.

4.3. Results

The network design is examined in three different tests for three different scenarios. The objective is to predict the posterior distributions of the acoustic and shear impedances and density from angle gathers. There are seven wells available along the line. The incident angles from the gather range from 0 to 28 degrees at 5-degree increment. The seismograms

after wavelet transformation are treated as training data, while the impedances and density logs are used as label data to the corresponding ends of the deep learning system. Well 2, which is located in the middle of the line, with full data coverage of the target intervals, is used for blind test purpose.

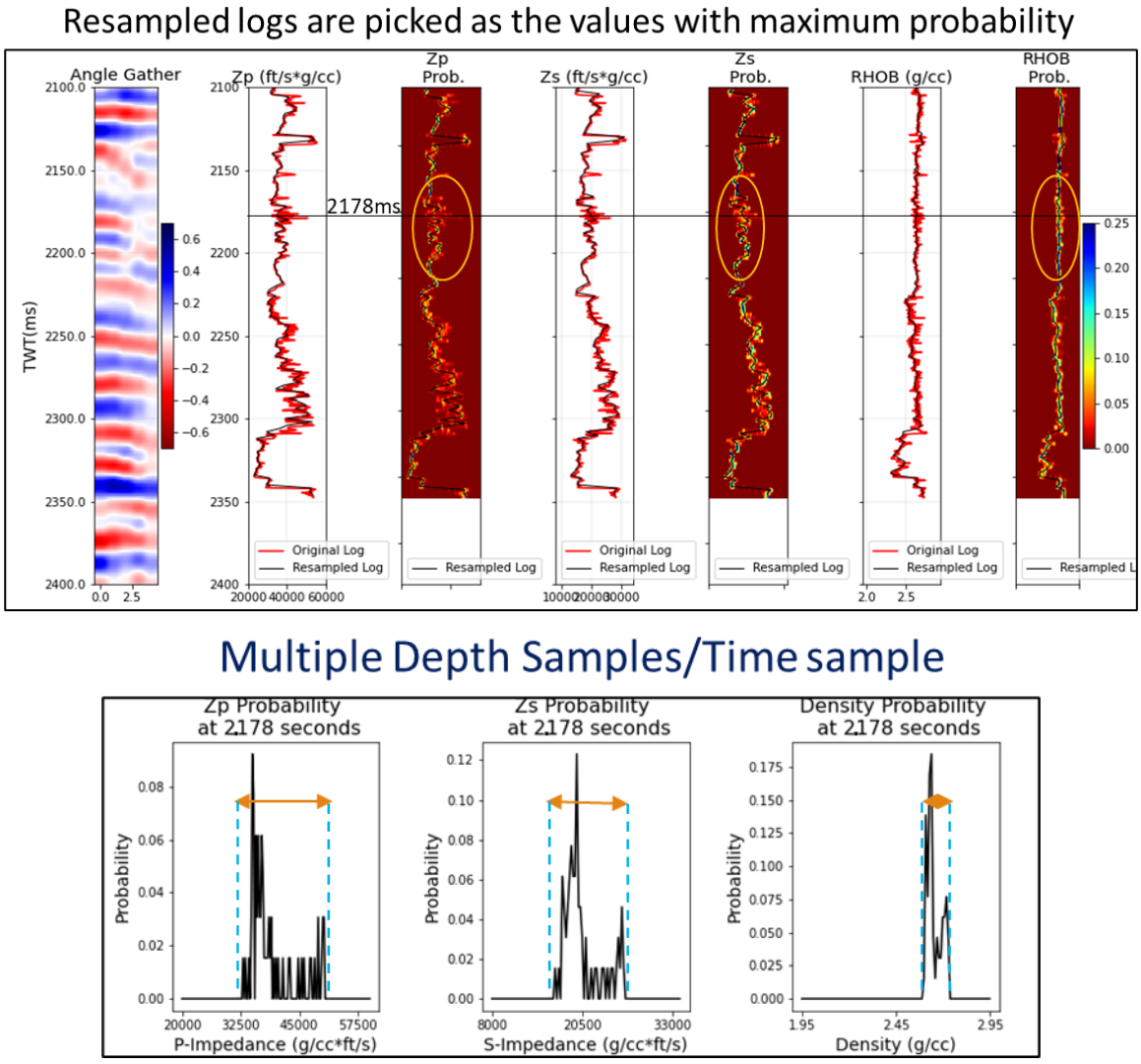

Figure 4.7: An illustration of probability distribution generated from real log data after depth-to-time conversion extracted at the 2178 millisecond time sample (lower) and the corresponding log curve and probability distribution maps (upper). In the comparison plots, the resampled curves are extracted from the values with maximum probability values. Notice the slight mismatch between the resampled log curves (black) and the true models (reds) are due to the truncated values of other possible solutions.

The first scenario requires examining the accuracy with a set of synthetic seismograms generated from an anticipated impedance and density volume that covers a large interval of 1,200 milliseconds. This study will also determine whether the training data and labels could be of enough coverage such that the network is able to predict away from the well locations.

In the second scenario, the model is trained with real seismic gathers and applied to a real 2D section ranging from 1,600ms to 2,800ms. In this case, the wells are with limited log coverages in the same reservoir interval, ranging from 2,100ms to 2,400ms. On log responses, those are indicated with outstandingly high gamma ray reading and low acoustic impedances located between stiffer intervals of high impedance readings (Figure 4.8). The result of this scenario demonstrates the algorithm performances with limited training data.

In the third scenario, the pre-trained network with synthetic data is applied to a real seismic dataset to predict the elastic properties. This is to verify the effectiveness of cross-model training, and examine how insufficient input data affects the algorithm performance.

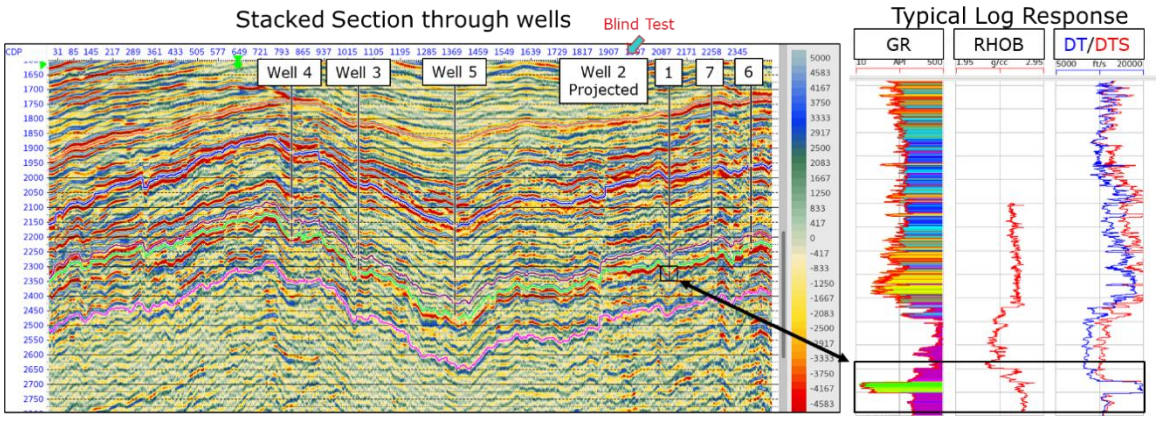

Figure 4.8: The seismic data and typical well log responses of the coverage interval.

4.3.1. Synthetic Data Application

The synthetic seismograms are generated for the whole 2D section using synthetic models of P-, S-impedances and density. The amplitudes are calculated with the 3-term Aki-Richards (1980) approximations of the Zoeppritz equation for angle offsets ranging from 5 to 25 degrees in 5 degrees increment, and a 10 percent of random noise is added to each seismogram. A sample synthetic seismogram of corresponding Zp, Zs and density is shown in Figure 4.9.

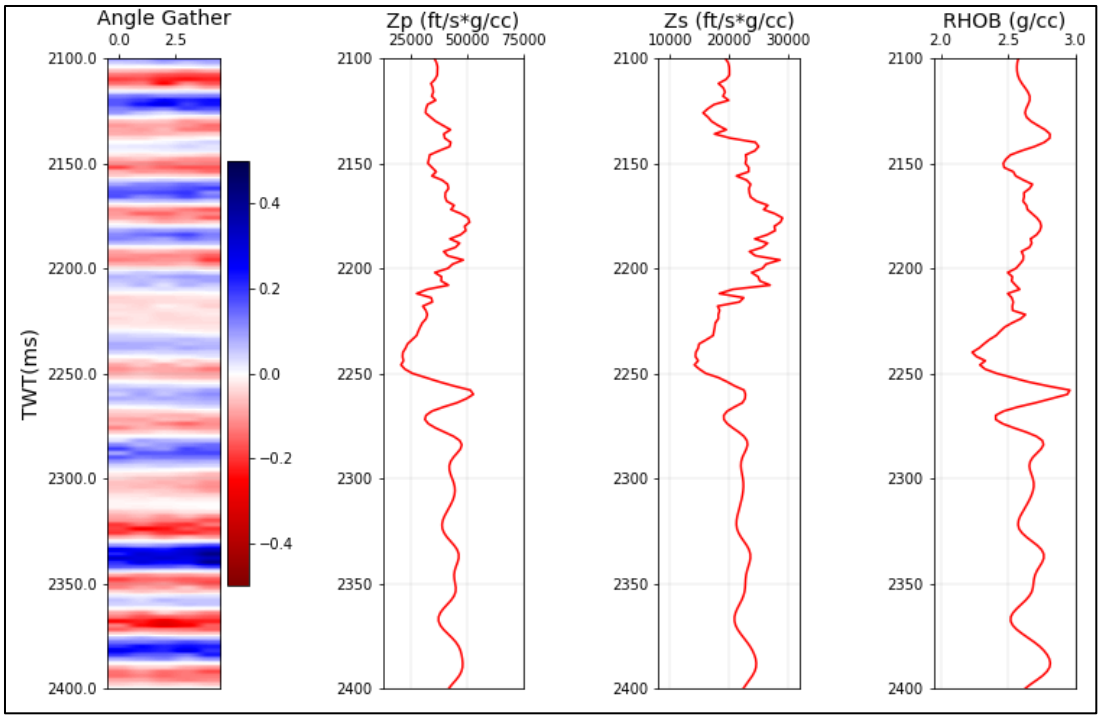

Figure 4.9: A synthetic angle gather generated from the impedances (Zp and Zs) and density log using the 3-term Aki-Richards approximation.

The models used for synthetic generation are recorded at the same time sampling interval with seismic trace, which results in spiky training label distributions. The network is trained with a small learning rate (10^{-7}), on a batch size of 32 using the adaptive moment

(ADAM - Kingma and Ba, 2014) optimization algorithm. The log values corresponding to the maximum predicted probability distribution are used for comparison with the original well logs. The learning stops when the validation error (blind test well) starts to increase from the global minimum value, which happens around the 2,800th epoch (Figure 4.10). At the blind test well location, the network reproduces all features in the original impedance and density models. The posterior distribution maps of properties (Figure 4.11) fully cover the associated true models, which suggests the network structure is well trained, and can provide reliable predictions away from the training wells. Meanwhile, the broad variation in the likelihood in the shear impedance prediction suggests non-uniqueness in the model predictions, where multiple possible solutions exist.

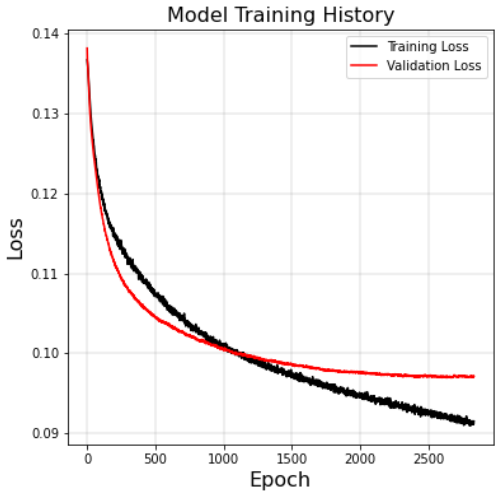

Figure 4.10: Loss curves of training and validation data for QC purposes. The training stops after 2800 epochs when validation loss starts to increase.

The same network structure is applied to the raw training data, which is not wavelet-transformed. For comparison purposes, the network structure is kept the same, where the only change is the input shape of the training data to better reflects the dimension of this different dataset. As displayed in Figure 4.12, at the blind-test location, after well trained,

the network is able to partially reproduce the training labels for P-impedance and density logs, where most of the predicted posterior distribution maps fully cover the real models. Meanwhile, it encounters difficulties reproducing the S-impedance, since the resulting distribution map does not accurately cover the true model curve. The scattering patterns in the S-impedance suggests multiple solutions that could satisfy the training data, which suggest current input dataset is not enough for the algorithm to clearly identify the hidden patterns. This suggests that the application of the CWT before training the network does significantly improve the overall performance of the deep learning system.

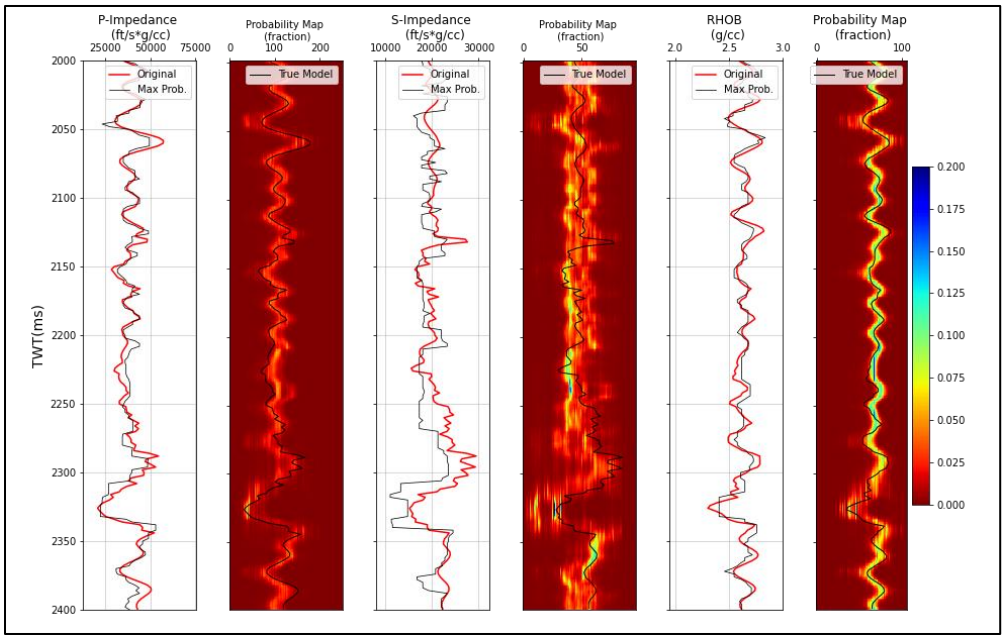

Figure 4.11: The network predictions at blind test well in comparison with real log when trained with data undergone wavelet transform. The predicted posterior distribution map covers the true model, which suggests that the network well performs in this dataset.

With satisfactory results at the well locations, the trained network with wavelet transform is applied to the whole 2D synthetic section. Similar to the blind test results, the

predicted P-impedance (Figure 4.13) and density (Figure 4.15) highly resemble the true models, which suggest that the training data is sufficient enough to represent the whole 2D dataset. Also, several variations in the shallow zones in the P-impedance suggests the network to have difficulty recognizing the corresponding patterns for correct solution. The predicted S-impedance section (Figure 4.14) also observes the variations from the true model in the shallower zone.

In summary, the structure of such a neural network seems to be complicated enough to capture the features, and the data from 7 wells are sufficient to represent the whole 2D section.

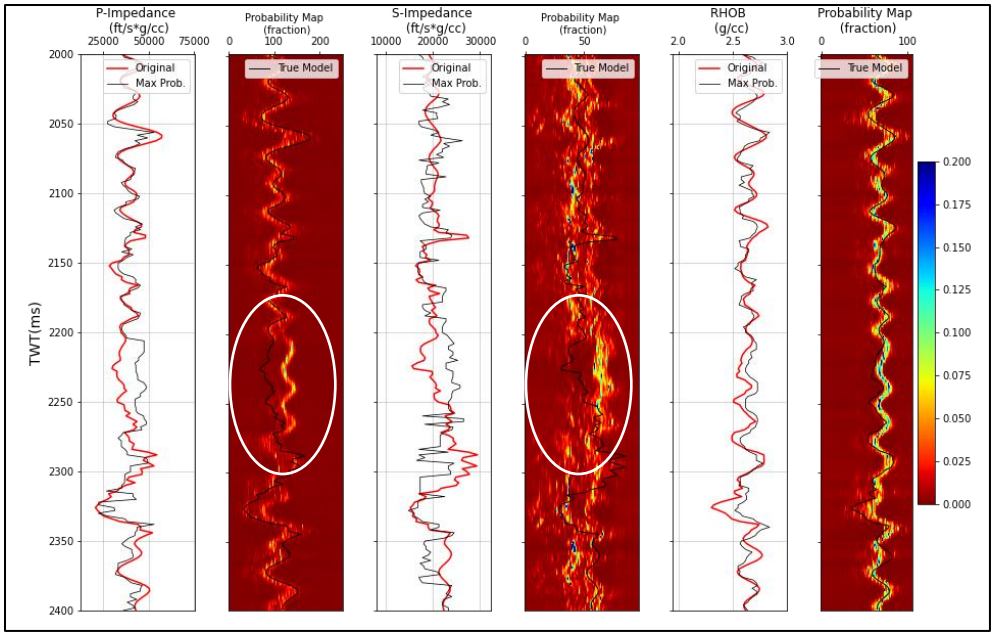

Figure 4.12: The network predictions at blind test well when trained with raw seismic angle gather using the same network structure and hyperparameters, except for different training data structure. The predicted posterior distribution maps partly cover the true models for the P-impedance and density logs, while mis-predicting the distribution of most of the S-impedance log and part of the P-impedance (highlighted in white oval). The scattering prediction of the S-impedance indicates the existence of multiple solutions that could satisfy the training data, which in this case is the seismic amplitudes.

4.3.2. Real Data Application

After being tested and proven effective with synthetic dataset, the same network structure is applied to the real seismic dataset. The hyperparameters are set the same as those used in the synthetic dataset, except for the maximum epoch to be 6,000. With the validation test data (or blind test well), the network reproduces most of the features in in the true models. The dominant events on the predicted posterior probability maps cover most of the original data, except for two zones: the shallow moderate impedance zone between 2,100-2,150ms, and the stiffer zone (2,275-2,300) overlaying the low impedance hydrocarbon bearing interval (Figure 4.16). Such differentiation suggests the training data does not provide sufficient information to help the model avoid mispredictions in such zones.

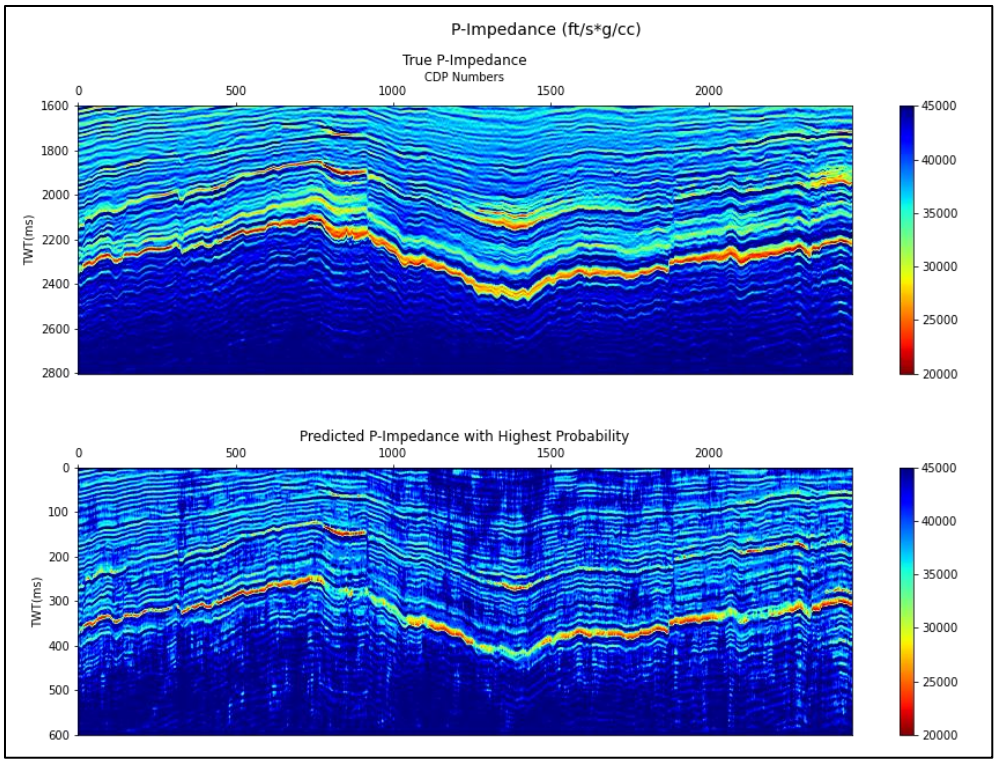

Figure 4.13: The comparison between the predicted and real P-impedance. The matured network is able to reproduce all features of the original section.

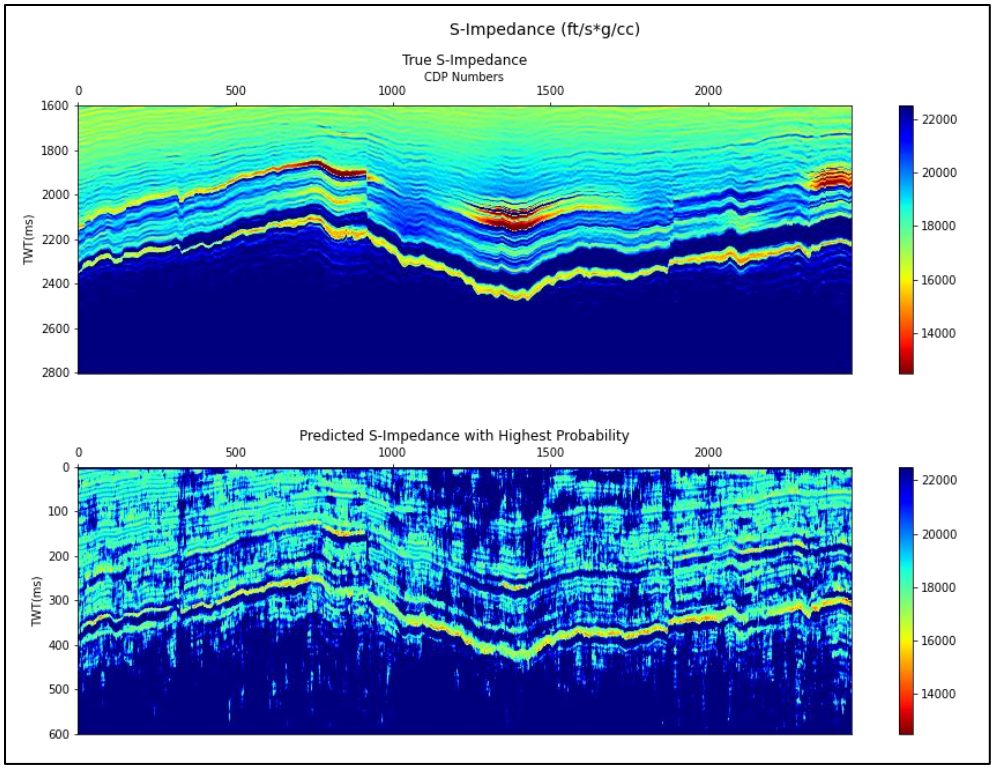

Figure 4.14: The comparison between the predicted and real S-impedance. The matured network is able to reproduce all features of the original section.

The trained network is applied to the 2D data to predict the attribute sections (Figure 4.17). Due to the limited well data coverages, the model is only capable of predicting reliable results around the reservoir interval. Outside of the target zone, the uncertainty increases, which lead to patchy predictions, especially on the S-impedance and density.

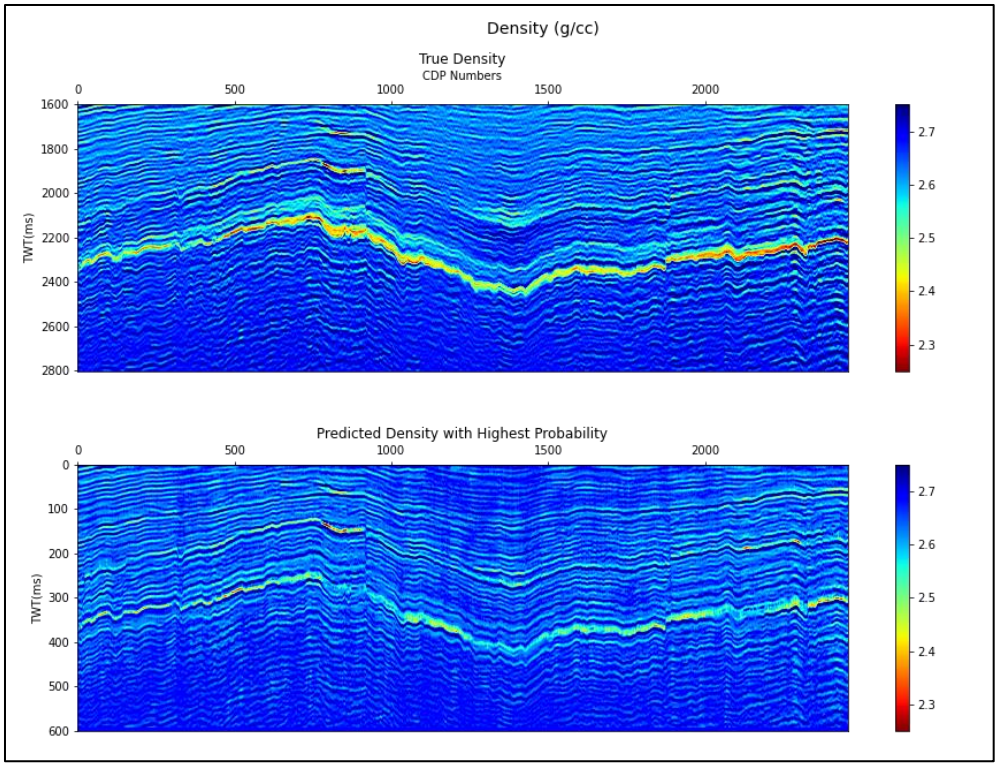

Figure 4.15: The comparison between the predicted and real density. The matured network is able to reproduce all features of the original section.

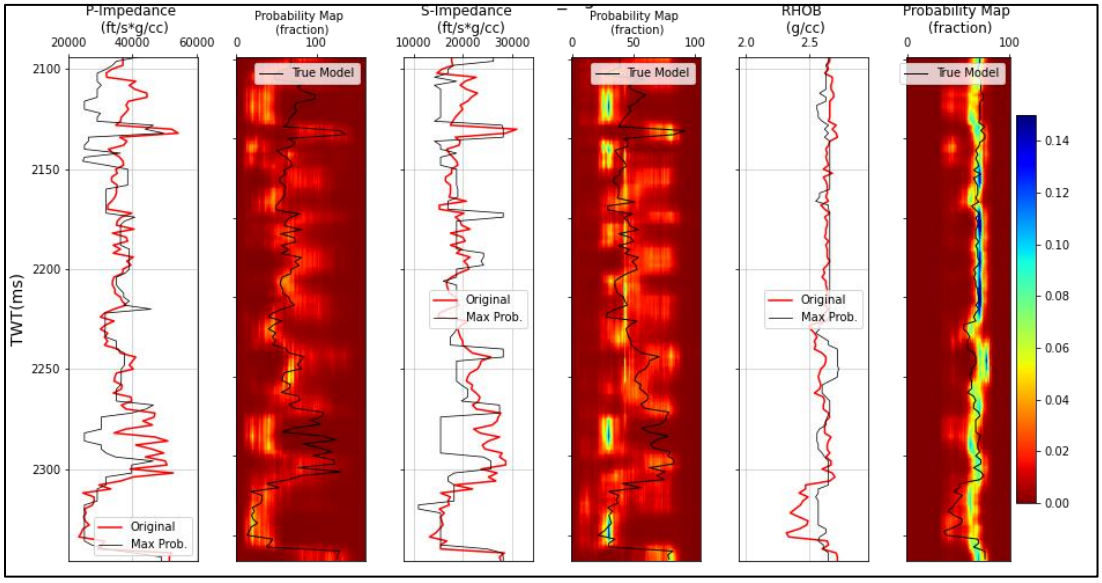

Figure 4.16: The network predictions at blind test well location.

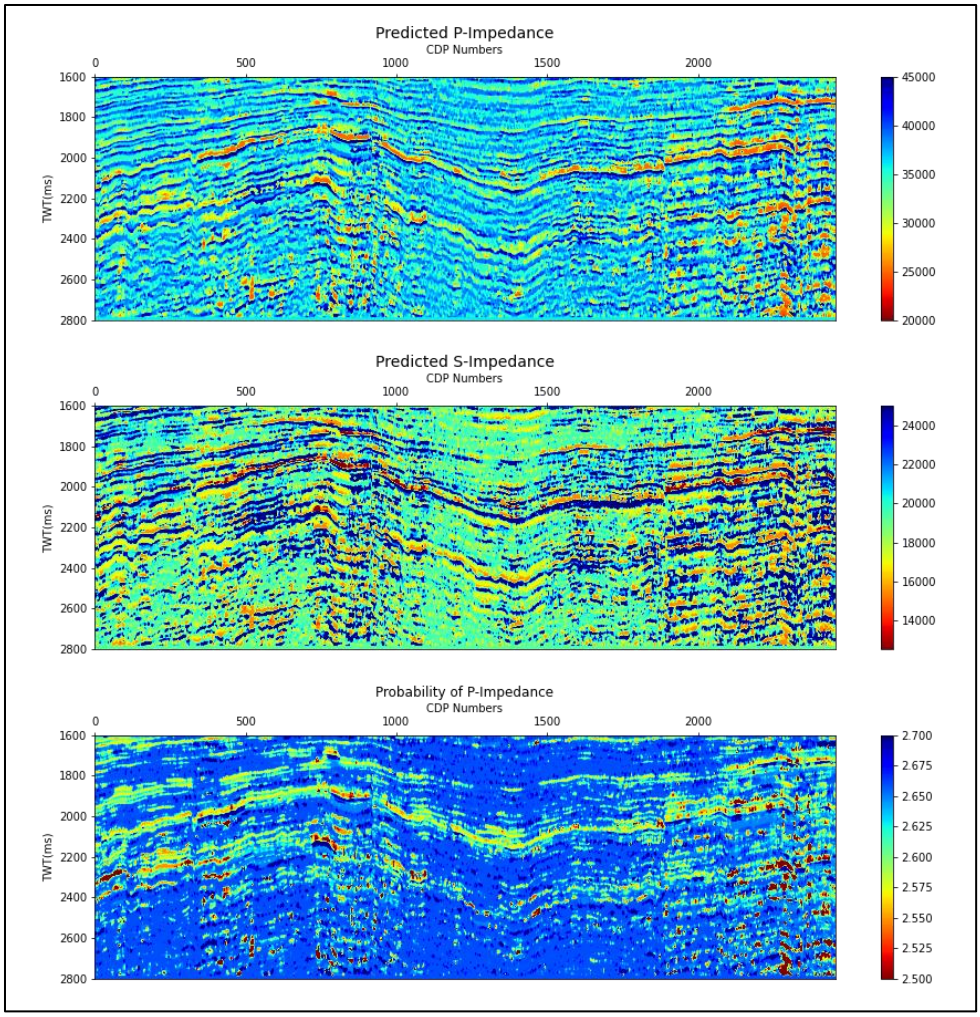

Figure 4.17: The predicted property sections corresponding to maximum probability values from trained network with real data.

4.3.3. Cross Model Application

The network trained with the synthetic dataset is used to predict the elastic impedances and density probabilities using the real seismic section. The predictions are of good quality, with the posterior distributions well covering the real models at well locations, as shown in Figure 4.18. There are some underpredicted zones due to the difference in dynamic ranges of the training data and models from the synthetic dataset in comparison with the real dataset. However, using the Pearson correlation to measure the

similarity between the predicted property curves (either corresponding with the maximum probability, or mean values) and the true models, the predictions with cross-model applications are better (Figure 4.19). This proves that the additional information provided from the shallow and deeper sections of the synthetic dataset is the key contribution to the advancement of the cross-model application.

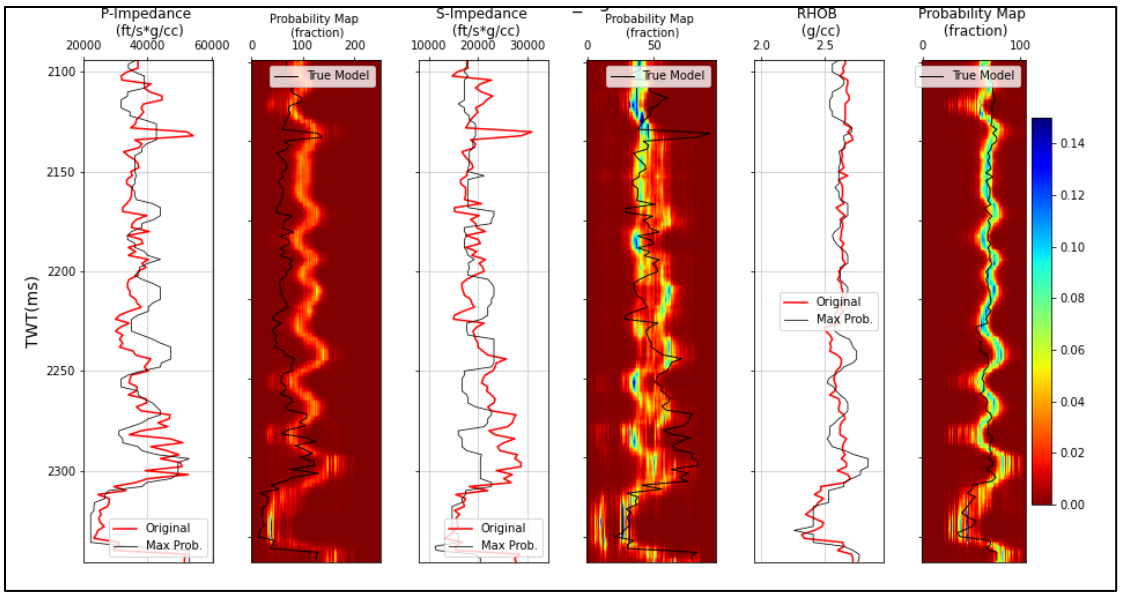

Figure 4.18: The predicted probability maps by applying the model trained with synthetic data on real seismograms.

On the predicted 2D sections, as shown in Figure 4.20, the properties are more continuous and well conforming with the seismic events. Especially in the P-impedance section, the outstandingly low impedances are clearly highlighted, which suggests potential lateral continuity of the reservoir bodies away from controlling wells.

4.4. Summary

This study examines a multimodal deep learning structure to estimate the posterior distributions of elastic properties from seismic gathers, which provide insights into possible

what-if scenarios from the recorded field data. Two major sources of uncertainties are accounted for in this study: (1) the discrepancy in domain resolution in mapping between seismic (two-way traveling time in seconds) and well log data (depth in meters); and (2) the different combinations of elastic properties resulting on similar peak or trough amplitudes. By converting the training labels from regular time varying property values to 2-dimensional time varying probability distributions, the network can bypass the heuristic computation of the partition function and directly map the posterior distributions of the P- and S- impedances, and density.

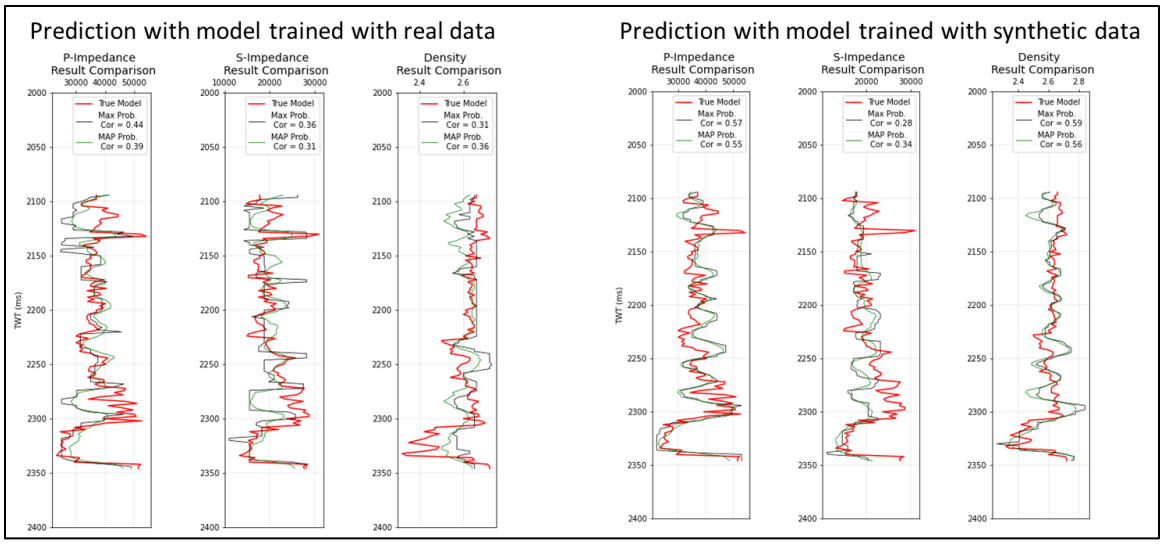

Figure 4.19: The comparison between the property predictions using model trained with real dataset, and using model trained with synthetic data. Using the Pearson correlation as a benchmarking term, the predictions using model trained with synthetic data are matching better.

The multimodal design allows more flexible training label arrangement, due to the introduced weighting terms that could be input by physical expectations of the desired outcomes. The input seismic gather is decomposed with the continuous wavelet transform using a Ricker mother wavelet to further generate more useful data for the deep learning

process. The addition of wavelet transform is shown to significantly improve the quality of the prediction. The properly trained network structure is demonstrated to accurately predict the likelihood maps of the outcomes, as the most likely distributions fully cover the true models. Different training scenarios, from synthetic to real dataset, indicate the importance of sufficient data coverage for better deep learning performance, as the more data coverage allows the algorithm to better capture the broader dynamic range of the desired properties.

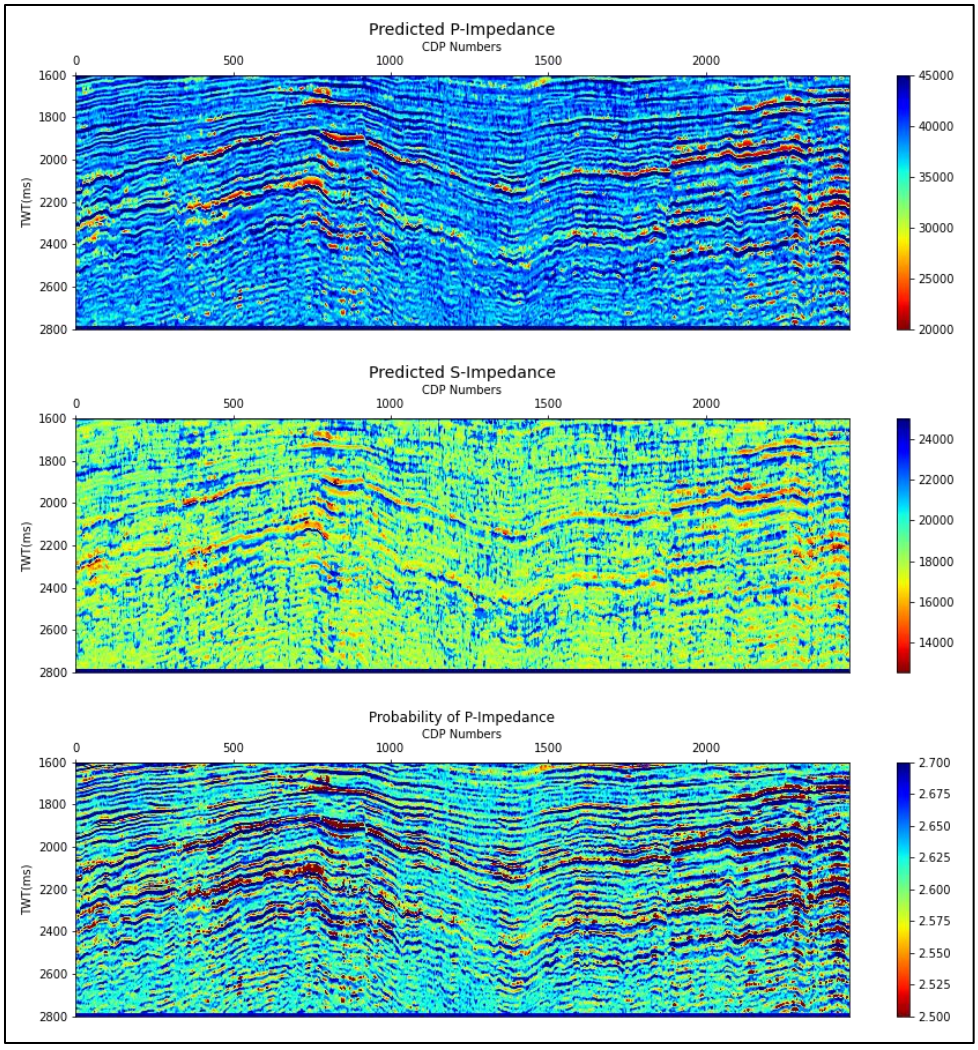

Figure 4.20: The predicted property sections corresponding to the maximum probability values from cross model application.

Chapter 5 : Summary and Future Work

5.1. Summary

Recorded seismograms contain useful information that are used to infer subsurface elastic properties. The seismic inversion is a powerful tool to extract the hidden information from seismic data, provided reasonable assumptions are made to linearly approximate the physical theory behind. Machine learning, on the other hand, is a more flexible and adaptive alternative approach to deal with the nonlinearity of the inverse problem and has been evolving as a promising advancement to seismology. A properly configured neural network structure built up on the system of interconnected units is capable of capturing hidden patterns from the input information, which can be integrated into the inversion workflow to infer the elastic properties from seismic amplitudes. I hypothesize that with proper network structures and suitable training schematics, the machine learning algorithms could resolve some outstanding challenges in seismic inversion: (1) the limited resolution due to wavelet effects; (2) the dependence on some rock physics models when inverting for petrophysical properties; (3) the quantification of the posterior distribution of the prediction results to resolve the resolution discrepancy between the time domain seismic signal and well logs in depth domain.

In this dissertation, I discussed the development and applications of three different machine learning algorithms for seismic reservoir characterization: (1) the single layer Boltzmann machine to predict high resolution elastic properties (Chapter 2); (2) the cross-shape deep Boltzmann machine to estimate petrophysical properties from seismic gathers (Chapter 3); (3) the multimodal deep learning network to map the posterior distributions of the elastic properties from seismic amplitudes (Chapter 4). Also, in Appendix 1, I introduced a quantitative workflow to correct for potential anisotropy and fluid effects on recorded well logs into hydrocarbon-bearing sandstone intervals with heavy sand-shale

laminations, which could be effectively used to condition well log data before input to train any machine learning algorithms.

In Chapter 2, the Boltzmann machine is designed to automatically extract the elastic impedances and density from input angle gather seismograms. This single layer neural network comprises of a Hopfield network structure, which is known to quickly converge toward an optimal solution once an initial state of neurons is provided and an update rule is applied; and the Mean Field Annealing optimization process that ensures a global minimum solution that reaches equilibrium state 1-2 orders of magnitude faster than simulated annealing. The optimization process evolves by minimizing the energy function, which shares the same characteristic of the linearized Zoeppritz equation and convolution model that relate the elastic impedances and density to the recorded seismic amplitudes. The high-resolution results are produced from the post-inversion reflectivity series, which is not bounded by the wavelet effects. Meanwhile, the accuracy of the overall prediction is achieved from the more geologically sound property models built from horizons interpreted on the reflectivity sections that contain multiple features that are not available in the original seismic stack sections commonly used in the interpretation process.

In Chapter 3, a cross-shape deep Boltzmann machine algorithm is designed to predict multiple petrophysical properties from seismic amplitudes. The algorithm has the potential to overcome the typical challenges in common petrophysical inversion approaches, which are dependent on data fitting algorithms to create the rock physics models during well analyses, which is prone to uncertainties in input data. The cross-shape deep Boltzmann machine is designed by arranging four different restricted Boltzmann machines located at the vertices interconnected via a hidden neuron layer at the center. Four different input types, including the seismic amplitude, and three petrophysical logs: the porosity, water saturation and shale volumetric, are fed into visible layers located at the

vertices to training the network. Ultimately, six different nonlinear relationships between the inputs are simultaneously captured during the learning process, among which three connections between target petrophysical logs, and seismic amplitudes are used to predict property values. Unlike any conventional deep learning system, which expects large data for what-if scenarios into training, the CDBM does not involve any back propagation, but only requires the anticipated upper and lower bounds of the data and labels for random initializations of Markov chains to reach an equilibrium state, at which the network is able to reconstruct the input data probabilistically. This makes the algorithm ideal for problems with limited dataset, and for being powerful and robust enough to account for uncertainties while still retaining the generalization outside of the training information. A 2D field dataset with limited well coverage is used to demonstrate the capability of this algorithm with accurate reconstruction of geologically plausible petrophysical property sections.

The deep learning algorithm in Chapter 4 is to predict the posterior distributions of the subsurface elastic properties from seismic gathers. It is designed to resolve the limited resolution of the time domain seismic signals, which are not enough to capture all features present in the recorded well logs in depth domain, by converting the resolution discrepancy to probability distributions to represent all what-if scenarios from the well log responses, and use them to constrain the inversion process. The greatest advantage of this algorithm is the avoidance of the heuristic calculation of the partition function, which is required to calculate the posterior distributions of common neural network outputs. In addition, the algorithm bypasses the assumption of the forward modelling steps, which usually include the convolution model and linearized Zoeppritz equation, to connect the target models to the observed data. Thus, it has the potential of inverting depth domain seismic gathers. Besides, unlike the common deep learning structure for prestack inversion, the new design does not require combination and reorganization of the impedances and density logs into

one single layer, but allows them to be input to the system at separate branches. To further improve on the features in capturing the hidden relationships, the seismograms are processed with a continuous wavelet transform to generate training data. The algorithm is applied to two datasets: the synthetic angle gathers generated from some elastic models using the three-term approximation to examine the network stability, and the field angle gathers with limited well coverage to demonstrate its application in real noisy data.

The successful results from these algorithms agree with the original hypothesis that the network structures and training schemes can be designed and fine-tuned to solve different challenges in seismic inversion, while still able to simulate the underlying physical process that relate seismic amplitudes to the subsurface properties.

5.2. Future Work

The Boltzmann machine is a powerful probabilistic neural network that has not been fully investigated for seismological applications. Its deep learning process does not require any back propagation to reach the stable solutions, which is a significant advantage comparing to other deep learning algorithms. Therefore, the Boltzmann machine is an excellent candidate to integrate into computationally intensive inversion algorithms such as the full waveform inversion, which heavily depends on solving the wave equation and error back propagation to update the model.

On the other hand, most of the machine learning algorithms are using gradient based optimizations to minimize the energy or loss functions. For simple network structure, this approach produces reasonable results. Meanwhile, for complex deep learning structures with multilayer designs, the model update involves the heuristic backpropagation process, which can be shown to get trapped to local minima. Recent developments include trying to incorporate the stochastic algorithms such as the very fast simulated annealing (VFSA), stochastic gradient descent etc. into updating the model

properties. More extensive studies on how to best benefit from such stochastic processes are required for better understanding of the effectiveness, and the associated pitfalls.

There are multiple sources of uncertainties affecting the performance and overall results of the deep learning algorithms. The algorithm introduced in this dissertation to map the posterior distribution from seismic amplitudes mostly deals with the time-depth resolution discrepancy between recorded well logs and seismic data. The additional data preparation to convert original log values into probabilistic distributions is a novel data arrangement process to avoid the complications of a true probabilistic neural network. It is capable of incorporating multiple sources of uncertainties: from training data and labels, from hyperparameters from the network systems, and all what-if scenarios. A deep investigation on how to implement the full probabilistic network is essential to fully resolve all uncertainty sources in seismic reservoir characterization using machine learning algorithms.

Appendices

A. Isotropy Correction and Fluid Substitution for Deviated Wells

Paper presented at the SEG 2020 Annual Meeting

A.1. Introduction

Seismic reservoir characterization aims to extract further useful information from the elastic properties, which are represented by the velocity terms (P-wave and S-wave). These are calibrated with sonic logs measured at borehole locations through the synthetic generations and tying process, after which any further examinations are performed with the guidance of the well log analyses. A set of good quality sonic logs are essential prerequisites to any characterization processes. Therefore, it is important to carefully examine the input dataset and perform any necessary calibrations to improve the reliability of sequential results.

Sonic logs are extracted from the refraction acoustic signals along the borehole surface (Sheriff et al., 1999). Bypassing the influences of the processing steps, the quality of recorded data is strongly dependent on the elastic properties of the host environment through which the logs are recorded. For clastic reservoirs, the complexity of the environment depends on the burial history, the diagenetic process and the lithology and fluid contents.

As shale makes up a large percentage of the sedimentary basin and exists within the reservoir structures, they contribute greatly to the anisotropy of the subsurface features (Hornby, 2003). Wave propagation through these environments suffers great uncertainties which are reflected on the recorded log values. Also, for an anisotropic environment, measured acoustic logs in deviated well-paths are not true representative of the environment within the well vicinity. Furre et al. (1998) reported differences in log

measurements between vertical and deviated wells through anisotropic environment, which lead to problematic usage of synthetic seismograms to tie well logs with seismic data. Hornby (2003) introduced methods to correct for anisotropy effects for deviated-well sonic logs in shale intervals. The approach assumes weak anisotropic environment to take advantage of the Thomsen parameters (Thomsen, 1986) to simplify the mathematical formulations and better accommodate the calculation process.

Fluid content within a reservoir also affects the recorded sonic logs. Commonly, the compressional signal slows down when passing through fluid-filled porous media and speeds up when entering stiffer materials surrounding the pore space. Wells penetrating similar reservoir intervals with different fluid contents will report different sonic log responses. In order to compare the log quality between wells of similar reservoir intervals, the logs need to be calibrated toward similar fluid content.

For an anisotropic environment, the common isotropic Gassmann fluid substitution model (Gassmann, 1951) over- or under-estimates the log responses. Meanwhile, the anisotropic Gassmann equation is governed by the stiffness tensors that are not fully characterized due to the lack of necessary measurements. Mavko and Bandyopadhyay (2009) proposed an approximate fluid substitution for vertical velocities in weakly anisotropic VTI rocks, which employs the Thomsen parameters.

This study examines the variation in recorded sonic logs at three deviational wells targeting similar clastic units. The reported in situ laminated shale layers is thought to introduce anisotropy to recorded logs along the well bore. In order to evaluate this hypothesis, a quantitative approach is proposed, which involves two main steps. Firstly, anisotropy effect is removed by converting recorded sonic log to vertical velocity along the symmetry axis of the anisotropic environment using a model-based inversion algorithm.

Then, the fluid effect is eliminated by performing anisotropic fluid substitution to restore similar reservoir intervals at different well locations toward the same fluid content.

Successful results from this study suggests a novel approach to perform calibrations to sonic logs measured along well bores penetrating through anisotropic environments before inputting these logs into further analyses for reservoir characterizations.

A.2. Theory

A.2.1. Weak Anisotropy Correction for VTI Media

Thomsen (1986) proposed the formulation of the velocities along a direction inclined an angle θ with the symmetry axis with the assumptions of weak anisotropic medium. These are the phase velocities (P- and S-wave) that represent the traveling speed of the wave-front. The P-wave phase velocities are described in Equations A.1 with two Thomsen parameters δ, ε, which are the combinations of the stiffness tensors as in Equations A.2-A.3, and a phase angle θ between the wave direction and the symmetry axis.

$$Vp(\theta) = Vp_0[1 + \delta sin^2\theta cos^2\theta + \varepsilon sin^4\theta] \quad \text{(A.1)}$$

$$\varepsilon = \frac{C_{11} - C_{33}}{2C_{33}} \quad \text{(A.2)}$$

$$\delta = \frac{(C_{13} + C_{44})^2 - (C_{33} - C_{44})^2}{2C_{33}(C_{33} - C_{44})} \quad \text{(A.3)}$$

The measured sonic logs were processed with semblance method. Wang et al (2012) concluded that the resulting velocity from the semblance method is actually the phase velocity. Therefore, in this study, it is safe to use above equations with the measured sonic logs.

In Equation A.1, the term on the left is the measured sonic log along the well bore. The angle θ is equivalent to the angle between the well bore and the bed symmetry axis.

This information can be obtained from simple geometric calculations using the bed dipping angle, the well inclinations and azimuth.

Equation (A.1) can be rewritten in form of $\boldsymbol{d} = \boldsymbol{Gm}$, where $\boldsymbol{d} = \{Vp_i(\theta)\}_{i=1}^{N}$ is the recorded P-wave log. At each recorded sample, the matrix $\boldsymbol{G} = [1 \quad sin^2\theta cos^2\theta \quad sin^4\theta]$ and the unknown $\boldsymbol{m} = [Vp_o \quad \delta Vp_o \quad \varepsilon Vp_o]^T$. By inverting for unknowns in this linear problem, the symmetry axis velocity Vp_o, and the Thomsen parameters (δ, ε) can be obtained.

A.2.2. Anisotropic Fluid Substitution

Gassmann (1951) published the fluid substitution to simulate the changes in elastic properties with the changes in pore fluid contents in both isotropic and isotropic media. The isotropic fluid substitution assumes low frequency isotropic environment with homogeneous mineral property. Meanwhile, the anisotropic fluid substitution model deals with variations in stiffness tensors due to changes in fluid content. The process requires the bulk moduli of saturated and dry rock framework. These could be calculated using isotropic substitution, as shown in Equations A.4-A.5 where ϕ is the total porosity, K_{fl}, Km, K_{dry} and K_{sat} are the bulk moduli of the fluid content, mineral, dry framework, and the saturated rock, respectively. The fluid dependent stiffness tensors, whose indices ***i, j, k, l*** are the directions 1, 2 or 3, are calculated using Equation (A.6).

$$K_{dry} = \frac{K_{sat}\left(\frac{\phi K_m}{K_{fl}} + 1 - \phi\right) - K_m}{\frac{\phi K_m}{K_{fl}} + \frac{K_{sat}}{K_m} - 1 - \phi} \tag{A.4}$$

$$K_{sat} = K_{dry} + \frac{\left(1 - \frac{K_{dry}}{K_m}\right)^2}{\frac{\phi}{K_{fl}} + \frac{1-\phi}{K_m} - \frac{K_{dry}}{K_m^2}} \tag{A.5}$$

$$c_{ijkl}^{sat} = c_{ijkl}^{dry} + \frac{(K_m\delta_{ij} - c_{ij\alpha\alpha}^{dry}/3)(K_m\delta_{kj} - c_{\beta\beta kl}^{dry}/3)}{(K_m/K_{fl})\phi(K_m - K_{fl}) + (K_m - c_{ppqq}^{dry}/9)} \quad \text{(A.6)}$$

$$\text{Where } \delta_{ij} = \begin{cases} 1, for\ i = j \\ 0, for\ i \neq j \end{cases} \quad \text{(A.7)}$$

$$c_{3333}^{dry} \approx c_{3333}^{sat} - \frac{\left(\frac{K_{fl}}{K_m}\right)\left[K_m - K_{iso}^{sat} - \frac{2}{3}c_{3333}^{sat}\delta\right]^2}{\phi(K_m - K_{fl}) - \left(\frac{K_{fl}}{K_m}\right)(K_m - K_{iso}^{sat})} \quad \text{(A.8)}$$

$$c_{3333}^{sat} \approx c_{3333}^{dry} - \frac{\left(\frac{K_{fl}}{K_m}\right)\left[K_m - K_{iso}^{dry} - \frac{2}{3}c_{3333}^{dry}\delta\right]^2}{\phi(K_m - K_{fl}) - \left(\frac{K_{fl}}{K_m}\right)(K_m - K_{iso}^{dry})} \quad \text{(A.9)}$$

Since it is not possible to measure all stiffness components to fully describe the tensor, it is difficult to use Equation (A.6) to perform the fluid substitution. By assuming low frequency and mono-mineralic rock matrix, Mavko and Bandyopadhyay (2009) provided the approximation of this equation in case of VTI media, which is summarized in Equations (A.8) and (A.9). The order of calculations is similar to that of an isotropic case, where the stiffness tensor of the dry matrix is first calculated using Equation (A.8) and then fed into Equation (A.9) to predict the stiffness tensor of desired fluid saturated case. The symmetry P-wave velocity is related to the stiffness tensor c_{3333} (i.e c_{33} in Voigt notation - Voigt, 1910) as $c_{3333} = c_{33} = \rho V p_0^2$.

A.3. Results

The workflow is applied to a set of logs from three deviated wells at a producing field. The first vertical well was drilled first with sonic log (only P-wave) measured by log-while-drilling (LWD) and wireline log and the logs were confirmed to be of reliable quality. Other two wells were drilled later with directional paths to reach the targets and the sonic logs (P- and S-wave) were recorded with LWD method. The inclination of the second well was of 30 degree and the third well was of 54 degrees. The third well reached

targets which are within the vicinity of well 1, while the second well targeted an area far away from the previous two. From log responses, as described in Figure A.1, calculated petrophysical properties and production results, Well 1 encountered hydrocarbon intervals in both 2 reservoirs; Well 6 encountered hydrocarbon bearing interval in reservoir 1 and a mixture of hydrocarbon and water in reservoir 2; while Well 1 only recorded hydrocarbon in the reservoir and did not capture the existence of reservoir 2. The measured sonic log (*Vp*) at the well locations varies largely, spanning from 2750 m/s to 4000m/s, with large separation between well 3 and well 1 and 2, as shown in the histogram of the log values in Figure A.2.

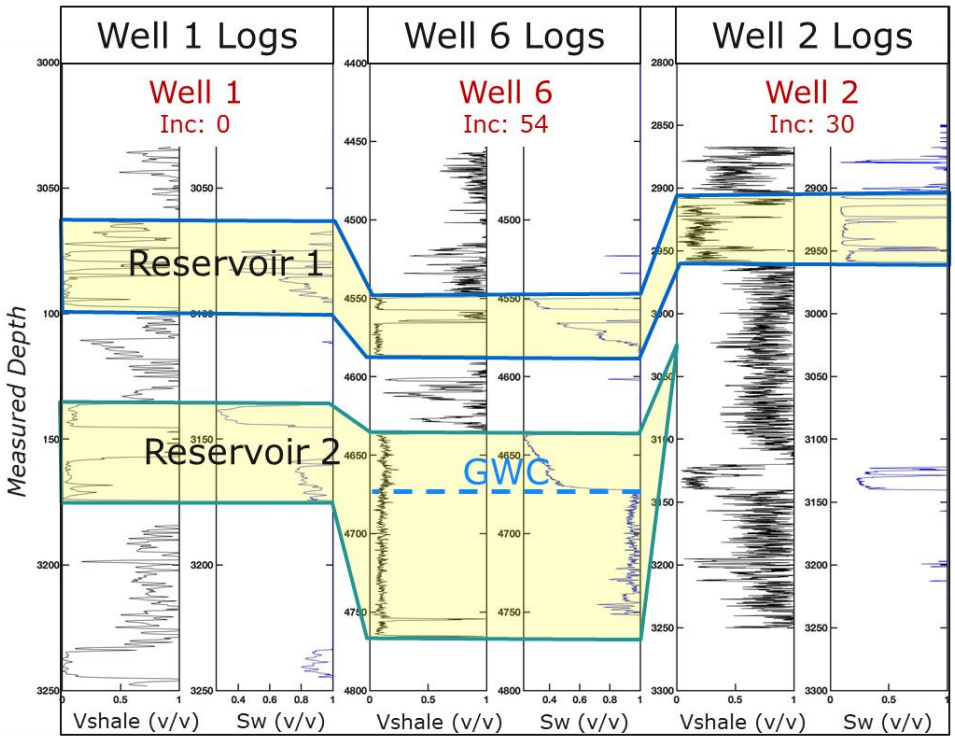

Figure A.1: Summary of the locations, the available logs with fluid contents of 3 wells used in the study.

The reliability of these histograms is questionable, due to: a) data being captured with different fluid contents from all wells, therefore suffers from the fluid effects, b) anisotropy effects at different well locations, especially at Well 6 and Well 2.

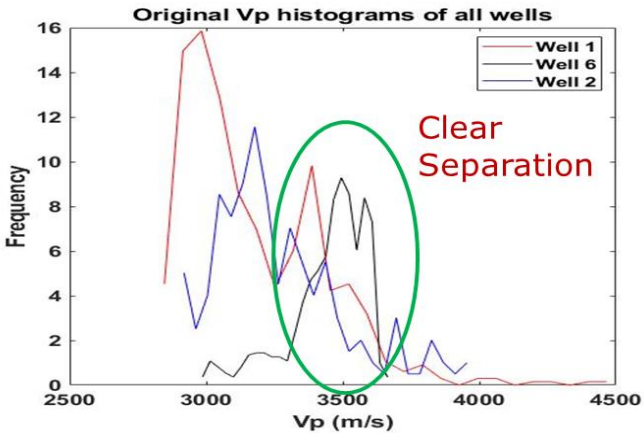

Figure A.2: Initial histogram of the P-wave velocities from reservoir 1 and 2 intervals at 3 wells showing clear separations of values.

The anisotropy effect from measured P-wave velocities at all wells is removed by inverting for symmetry axis P-wave velocity using an assumption of a weak VTI medium. Since the reservoir bodies are flat within the area of Well 1 and Well 6, the corresponding well inclinations are equal to the phase angle of the measured P-wave. Meanwhile, at Well 2 location, simple geometrical calculations resulted in an equivalent phase angle of 27 degree for this well path. The resulting axis P-wave velocities at three well locations are shown in Figure A.3. Well 1 is aligned with the symmetry axis, therefore the measured phase velocity is equivalent to the symmetry P-wave velocity. At Well 2, the small phase angle resulted in small drop from measured to symmetry P-wave. At Well 6, the deviation or the phase angle was high enough to result in significant drop in P-wave from measured log to symmetry axis values. Variation in the Thomsen parameters δ and ε observed at well locations reflects the minor differences in reservoir conditions at three well locations,

especially on gamma ray readings and the reservoir thicknesses. The resulting histograms of symmetry direction P-wave for each reservoir at all wells after VTI corrections in Figure A.4 show shifting of histograms toward the lower range of Well 1 for both reservoirs. For reservoir 2, there is no histogram plot from Well 2 as this reservoir does not reach out to this well.

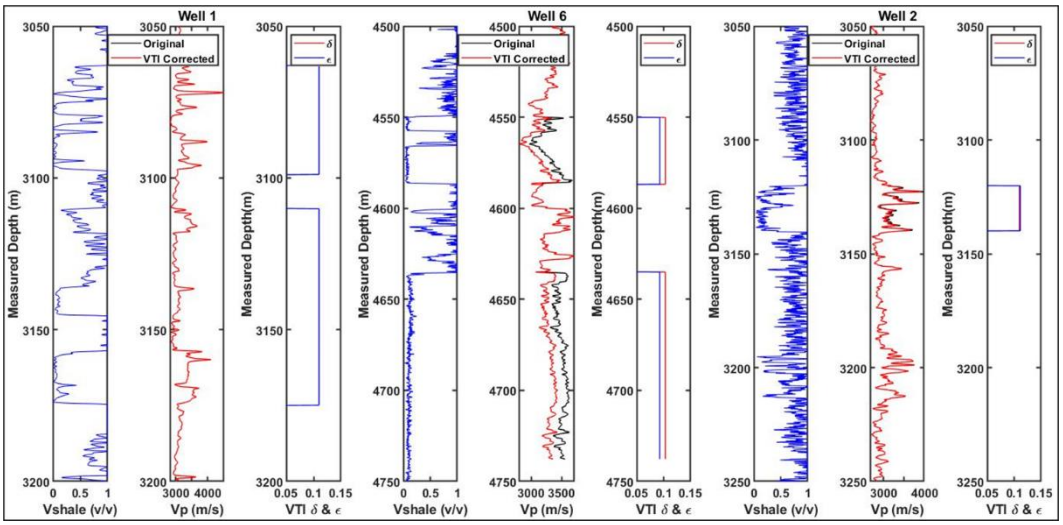

Figure A.3: Comparisons between the measured velocities (red curves) and the calculated symmetry velocities (blue) using weak VTI from Thomsen (1986).

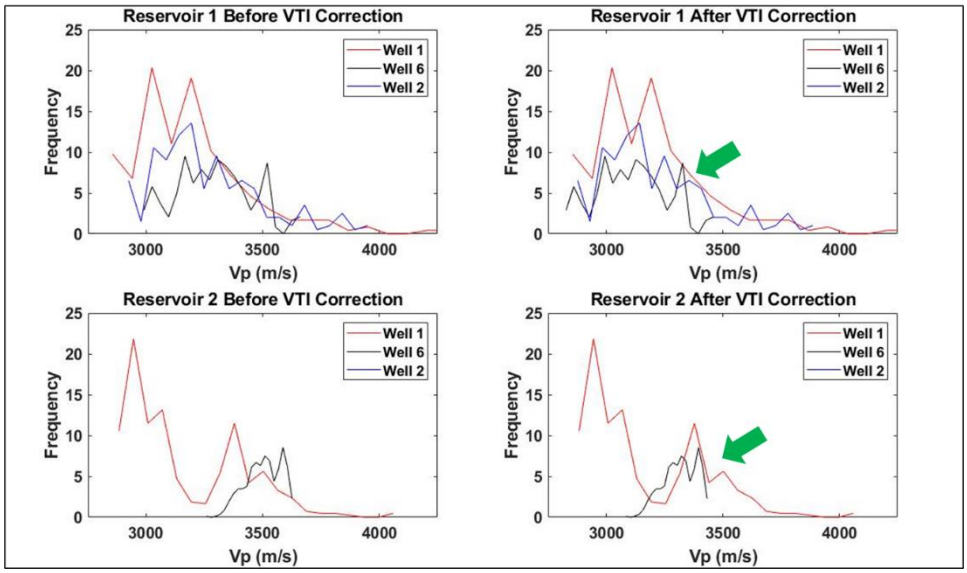

Figure A.4: Comparisons of histograms of symmetry direction P-wave for each reservoir at all wells after VTI corrections.

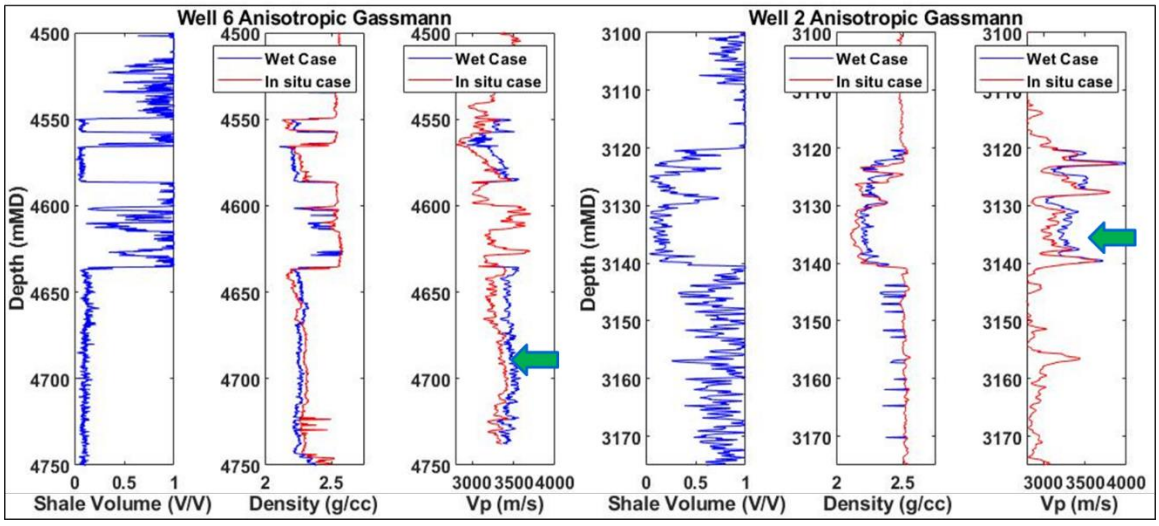

Figure A.5: The anisotropic fluid substitutions performed at Well 6 and Well 2 from in situ case to 100% water saturation show jumps in Vp velocity and density as the stiffer material (water) replaces the softer material (gas) in the pore space.

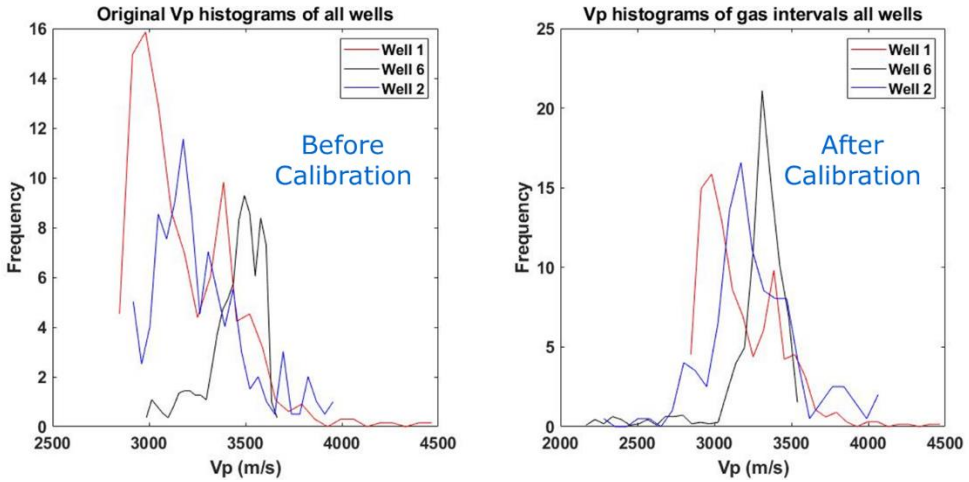

Figure A.6: The comparison between the histogram of original P-wave from all wells and the histogram of calibrated P-wave with similar fluid content from all wells.

The anisotropic fluid substitution is then performed to remove the fluid effects and restore reservoirs to be of the same fluid types before comparing the velocities. When all intervals are at 100 percent water saturation, the P-wave velocities and densities increase in all reservoir intervals, as water is a stiffer and heavier material than the in-situ gas. Figure A.5 shows the expected results in Well 6 and Well 2. At 25 percent water saturation, which is about the same fluid content of reservoirs in Well 1, the obtained histograms of the P-wave velocities of all reservoirs concentrate around the velocity ranges from 2,750 to 3,600 m/s, and no significant separations in histogram shapes were observed, as shown in Figure A.6. This confirms that the deviation in reading values of measured P-wave logs along the well bores are due to anisotropy and fluid effects.

A.4. Summary

This study proposed a novel approach to calibrate anisotropy and fluid effects on deviational wells penetrating sand-shale laminated clastic reservoirs. A set of recorded logs from two deviational and one vertical well path was used to demonstrate the process. The

anisotropy removal is performed by inverting for symmetry axis P-wave velocities at well locations using Thomsen's weak VTI phase velocity equation. Then, the anisotropic fluid substitution is applied to restore the reservoir intervals in two wells toward similar fluid content of the vertical reference well. After the calibration process, all velocity values concentrated toward a reasonable distribution, with no significant discrepancy. This suggests that the measured P-wave logs of deviational wells into anisotropic environment needs to be calibrated for better usage in further investigations.

References

Abubakar A; Hu W; Habashy T M; Van Den Berg P M, 2009, '*Application of the finite-difference contrast-source inversion algorithm to seismic full-waveform data'*, Geophysics, **74**, WCC47–58.

Aki, K; Richards, PG, 1980, '*Quantitative seismology: Theory and methods*', v.1: W.H. Freeman and Co.

Ansari, N; Li, K, 1993, '*Landmark-based shape recognition by a modified Hopfield neural network'*, Pattern Recognition, **26**(4), 531-542.

Bachrach, R, 2006, '*Joint estimation of porosity and saturation using stochastic rock-physics modeling*', Geophysics, **71**(5), O53-O63.

Bijwaard H; Spakman W; Engdahl E R, 1998, '*Closing the gap between regional and global travel time tomography'*, J. Geophys. Res. **103**, 30055–30078.

Bilbro, G; Mann, R; Miller, T; Snyder, W; Van den Bout, D, White, M, 1989, '*Optimization by Mean Field Annealing'*, Advances in Neural Network Information Processing System, Morgan-Kaufmann, San Mateo.

Biswas, R; Sen, M; Das, V; Mukerji, T, 2019, '*Prestack and poststack inversion using a physics-guided convolutional neural network*', Interpretation, **7**(3), SE161-SE174.

Biswas, R; Vassiliou, A; Stromberg, R; Sen, MK, 2018, '*Stacking velocity estimation using recurrent neural network'*, SEG technical Program Expanded Abstracts 2018, 2241-2245.

Brekovic, A; Kovacevic, M; Cvetkovic, M; Mocilac, I; Rukavina, D; Saftic, B, 2017, '*Application of artificial neural networks for lithofacies determination based on limited well data'*, Central European Geology, **60**(3), 299-315.

Brown, 1998, '*Common techniques for quantitative seismic interpretation'*, found in Avseth, P; Mukerji, T; Mavko, G, 2005, '*Quantitative seismic interpretation: Applying Rock Physics Tools to Reduce Interpretation Risk'*, Cambridge University Press, 168-257

Buland, A; Kolbjornsen, O; Hauge, R; Skjaeveland, O; Duffaut, K, 2008, '*Bayesian lithology and fluid prediction from seismic prestack data'*, Geophysics, **73**(2), C13-C21.

Calderon-Macias, C; Sen, M, 1993, '*Geophysical interpretation by artificial neural systems: a feasibility study'*, 63rd SEG Annual International Meeting Expanded Abstracts, 254-256.

Calderon-Macias, C; Sen, M; Stoffa, P, 1997, '*Hopfield neural networks, and mean field annealing for seismic deconvolution and multiple attenuation'*, Geophysics, **62**(3), 992-1002

Chatterjee, R; Singha, D; Ojha, M; Sen, MK; Sain, K, 2016, '*Porosity estimation from pre-stack seismic data in gas-hydrate bearing sediments, Krishna-Godavari basin, India*', Journal of Natural Gas Science and Engineering, **33**, 562-572.

Dai, Y; Wang, R; Li, Z; Zhang, P; Tan, Y, 2016, '*A time-varying wavelet extraction using local similarity'*, Geophysics, **81**(1), V55-V68.

Diersen, S; Lee, E; Spears, D; Chen, O; Wang, L, 2011, '*Classification of seismic windows using artificial neural networks'*, Procedia Computer Science, **4**, 1572-1581.

Dixon, M; Cole, G; Bellgard, M; 1995, '*Using the Hopfield neural network with mean field annealing to solve the shortest path problem in a communication network'*, Neural Networks, Proceedings., IEEE International Conference.

Fatti, J; Smith, G; Vail, P; Strauss, P and Levitt, P; 1994, '*Detection of gas in sandstone reservoirs using AVO analysis: a 3-D seismic case history using the Geostack Technique'*, Geophysics, **59**(9), 1362-1376.

Fausett, L, 2002, '*Fundamentals of Neural Networks: Architectures, Algorithms and Applications*', Pearson, 2nd Edition

Fomel, S, 2007, '*Local seismic attributes'*, Geophysics, **72**(3), A29-A33.

Gassmann, F, 1951, '*Uber die elastizitat poroser medien*', Veirteljahrsschrift der Naturforschenden Gesellschaft, **96**, 1–23.

Furre, A; Brevik, I, 1998, '*Characterization of angle dependency in sonic logs*', 68th SEG Annual International Meeting Expanded Abstract.

Goodfellow, I; Bengio, Y; Courville, A, 2016, '*Deep Learning'*, MIT Press.

Habashy, TM; Abubakar, A; Pan, G; Belani, A, 2011, '*Source–receiver compression scheme for full-waveform seismic inversion*', Geophysics, **76,** R95–108.

Hinton, G, 2002, *'Training Product of Experts by Minimizing Contrastive Divergence',* Neural Computation, **14**(8), 1771–1800.

Hornby, B; Howie, J; Ince, D, 2003, '*Anisotropy correction for deviated-well sonic logs: application to seismic well tie*', Geophysics, **68**(2), 464-471.

Hopfield, J; Tank, D, 1985, '*"Neural" computation of decisions in optimization problems'*, Biological Cybernetics, **53**(3), 141-152.

Ikelle, L; Amundsen, L, 2018, '*Introduction to petroleum seismology'*, Society of Exploration Geophysicists, 2nd Ed.

Innanen, K; Margrave, G, 2011, *'Seismic inversion and importance of low frequencies'*, CREWES research report, **23**(2011), 1-13.

Kahoo, A.R; Javaherian, A, Araabi, B.N, 2006, '*Seismic deconvolution using Hopfield neural network*', Proceedings of 8th SEGJ International Symposium, Kyoto, Japan, 26-28 Nov, 1-6.

Kingma, D; Ba, J; 2014, '*Adam: A Method for Stochastic Optimization'*, International Conference for Learning Representations, arxiv: 1412:6980

Kirkpatrick, S; Gelatt Jr, C.D; Vecchi, M.P, 1983, '*Optimization by Simulated Annealing*', Science, **220**(4598), 671-680

Kohonen, T, 1988, *'Self-Organization and Associative Memory'*, SpringerLink, 2nd Edition.

Lecun, Y; Yoshua, B, 1995, *'Convolutional networks for images, speech, and time series'*, The handbook of brain theory and neural networks, 2nd Edition, MIT press, 276-278.

Li, S; Liu, B; Ren, Y; Chen, Y; Yang, S; Wang, Y, 2020, *'Deep-Learning Inversion of Seismic Data*', IEEE Transaction on Geoscience and Remote Sensing, **58**(3), 2135-2149.

Lines, L, Treitel, S, 1985, '*Wavelets, well logs and Wiener filters'*, First Break, **3**, 9-14.

Mallat, S, 1999, *'A wavelet tour of signal processing'*, Academic Press, 2nd Edition.

Menke, W, 1989, '*Geophysical Data Analysis: discrete inverse theory'*, 2nd edition (Sandiego, CA: Academic).

Mavko, G; Bandyopadhyay, K, 2009, '*Approximate fluid substitution for vertical velocities in anisotropic VTI rocks*', Geophysics, **74**(1), D1-D6.

McCormack, M; Zaucha, D; Dushek, D, 1993, '*First-break refraction event picking and seismic data trace editing using neural networks'*, Geophysics, **58**(1),67-78.

Mukerji, T; Jorstad, A; Avseth, P; Mavko, G; Granli, J, 2001, '*Mapping lithofacies and pore-fluid probabilities in a North Sea reservoir: Seismic inversions and statistical rock physics'*, Geophysics, **66**, 988-1001.

Murat, M; Rudman, A, 1992, '*Automated first arrival picking: a neural network approach'*, Geophysical Prospecting, **40**(6), 587-604.

Phan, S; Sen, MK, 2019, '*A Boltzmann machine for high-resolution prestack seismic inversion'*, Interpretation, **7**(3), SE215-SE224

Reitermanova, Z, 2010, *'Data Splitting'*, WDS'10 Proceedings of Contributed Papers, Part 1, 31-36.

Rojas, R, 1996, '*Neural networks: a systematic introduction'*, Springer Verlag Berlin Heidelberg, 342-343.

Roth, G; Tarrantola, A, 1994, '*Neural networks and inversion of seismic data'*; Journal of Geophysical Research, **99**(B4), 6753-6868.

Saggaf, M; Nebrija, E, 1998, '*Estimation of lithologies and depositional facies from wireline logs'*, 68th SEG Annual International Meeting Expanded Abstracts, 288-291.

Salakhutdinov, R; Hinton, G, 2009, *'Deep Boltzmann Machines',* Proceedings of the 12th International Conference on Artificial Intelligence and Statistics, Florida, **5**, 448-455.

Sheriff, R; Geldart, L, 1999, '*Exploration Seismology*', Cambridge University Press, 2nd Edition.

Shuey, R, 1985, *'A simplification of the Zoeppritz equations'*, Geophysics, **50**(9), 609-614.

Singha, DK; Chatterjee, R; Sen, MK; Sain, K, 2014, '*Pore pressure prediction in gas-hydrate bearing sediments of Krishna-Godavari Basin in, India*', Marine Geology, **357**, 1-11.

Sirgue, L, 2006, '*The importance of low frequency and large offset in waveform inversion'*, EAGE 68th Conference and Exhibition incorporating SPE EUROPEC 2006.

Sen, M, 2006, '*Seismic Inversion'*, Society of Petroleum Engineers.

Sen, MK; Stoffa, PL, 2013, *'Global Optimization Methods in Geophysical Inversion*', Cambridge University Press, 2nd Edition.

Smolensky, P, 1986, *"Chapter 6: Information Processing in Dynamical Systems: Foundations of Harmony Theory*", In Rumelhart, David E.; McLelland, James L*," Parallel Distributed Processing: Explorations in the Microstructure of Cognition",* Volume 1: Foundations, MIT Press, 194–281.

Specht, D, 1990, *'Probabilistic Neural Networks'*, Neural Networks, Vol 3, pg 109-118

Spikes, K, Mukerji, T, Dvorkin, J, Mavko, G, 2007, '*Probabilistic seismic inversion based on rock-physics models*', Geophysics, **72**(5), R87-R97.

Srivastava, R; Sen, M, 2010, '*Stochastic inversion of prestack seismic data using fractal-based initial models*', Geophysics, **75**(3), R47-R59.

Srivastava, N; Hinton, G; Krizhevsky, A; Sutskever, I; Salakhutdinov, R, 2014, *'Dropout: A Simple Way to Prevent Neural Networks from Overfitting'*, Journal of Machine Learning Research, **15**, 1929-1958.

Srivastava, N; Salakhutdinov, R, 2014, '*Multimodal Learning with Deep Boltzmann Machines'*, Journal of Machine Learning Research, **15**, 2949-2980.

Tarantola, A, 2005, '*Inverse problem theory'*, SIAM.

Tikhonov, A, 1963, '*Solution of incorrectly formulated problem and the regularization method*', Soviet Mathematics, **4**, 1035-1038.

Thomsen, L, 1986, '*Weak elastic anisotropy'*, Geophysics, **51**(10), 1954-1966.

Virieux, J; Asnaashari, A; Brossier, R; Metivier, L; Ribodett; A, Zhou, W, 2017, '*An introduction to full waveform inversion*', Encyclopedia of Exploration Geophysics, R1-1 to R1-40.

Voigt, V, 1910, '*Lehbuch der kristallphysik*', Teubner, Leipzig.

Walden, A; White, R, 1984, '*On errors of fit and accuracy in matching synthetic seismograms and seismic traces'*, Geophysical Prospecting, **32**(5), 871-891.

Wang, J; Chen, D; Zhang, H; Zhang, X; He, X; Wang, X, 2012, '*Studies on phase and group velocities from acoustic logging*', Applied Geophysics, **9**(1), 108-113.

Wang, L; Mendel, J, 1992, '*Adaptive minimum prediction-error deconvolution and source wavelet estimation using Hopfield neural networks'*, Geophysics, **57**(4), 670-679.

Wang, Y, 1999, '*Approximations to Zoeppritz equations and their use in AVO analyses'*, Geophysics, **64**(6), 1920-1927.

Wang, Y; Morozov, I, 2020, '*A simple approach to nonstationary source waveform estimation in seismic records*', Geophysics, **85**(3), V257-V268.

White, R; Simm, R, 2003, *'Tutorial: Good practice in well-ties'*, EAGE First Break, **21**, 75-83.

Yilmaz, Oz, 2001, '*Seismic Data Analysis'*, Society of Exploration Geophysicists.

Zhang, R; Sen, M; Srinivasan, S, 2013, '*Multi-trace basis pursuit inversion with spatial regularization'*, Journal of Geophysics and Engineering, **10**, 035012.

Zheng, Y, Zhang, Q, Yusifov, A, Shi, Y, 2019, '*Applications of supervised deep learning for seismic interpretation and inversion*', The Leading Edge, **38**(7), 526-533

Zoeppritz, K, 1919, *'Erdbebenwellen VII. VIIb. Über Reflexion und Durchgang seismischer Wellen durch Unstetigkeitsflächen*', Nachrichten von der Königlichen Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-physikalische Klasse, 66-84.

Vita

Son Phan received his B.S. degree in Geosciences and Mathematics from the University of Tulsa, Oklahoma in 2009, and M.S. degree in Geological Sciences/Geophysics from the University of Texas at Austin in 2011. From 2011-2017, he worked as a geophysicist for a petroleum operator in Vietnam. In January 2018, Son was readmitted to the UT Austin as a Ph.D. student in Geological Sciences/Geophysics with Dr. Mrinal Sen. During his study, he interned at EOG resources (2019), and BP America (2020). He is interested in designing and integrating different physics-guided machine learning algorithms into improving the seismic reservoir characterizations.

Email address: sonpdt@utexas.edu

This dissertation was typed by Son Phan.