PREDICTING OIL PRODUCTION FROM STOCHASTIC AVA INVERSION ATTRIBUTES

by

Hayden A. Powers

A thesis submitted to the Faculty and the Board of Trustees of the Colorado School of Mines in partial fulfillment of the requirements for the degree of Master of Science (Geophysics).

Golden, Colorado

Date _____

Signed: _____
Hayden A. Powers

Signed: _____
Dr. Trainor-Guitton
Thesis Advisor

Golden, Colorado

Date _____

Signed: _____
Dr. Bradford
Professor and Head
Department of Geophysics

ABSTRACT

Interpreting seismic information for future drilling locations is vital to the success of many oil and gas companies. These interpretations are primarily made on time or depth volumes or from derived attributes. However, inversions of these seismic data sets can solve for Earth properties of the reservoir and surrounding formations. This dissertation analyzes results from stochastic amplitude versus angle (AVA) inversions to train machine learning algorithms with the end goal of predicting cumulative oil production. We utilize two AVA inversions, the first from the SEAM Life of Field Model (SEAM) and the second from field data acquired over an offshore reservoir in West Africa (WAF). The two main algorithms we use are a Naive Bayesian Classifier (NBC) and Multidimensional Scaling (MDS), to make production predictions in both fields. MDS is an unsupervised method often used to understand how similar information is in a low-dimensional space. The NBC is a supervised technique we employ to predict high or low oil producing reservoir locations. Previously known wells in the reservoir determine the decision to classify locations or wells as high or low. The wells are split into high and low groups based upon their cumulative oil productions or total injected water. Sensitivity testing on the accuracy of the classifier is a major project focus, and requires evaluating many inputs to the NBC. We capitalize on the bulk accuracy from cross validation to evaluate the experiment parameters. The execution of the cross validation is exhaustive and based upon the number of omitted wells. In general, SEAM has a lower bulk accuracy than WAF, but has lower variance associated with changing the boundaries between high and low producing wells. We validate the boundaries by investigating the lower-dimensional space to understand specific misclassifications. Lastly, we incorporate the RMS error from the SEAM AVA inversion to aid in the quality control of the full reservoir classification results. The final reservoir classifications are accurate and prove the merit of using the NBC as a classification algorithm.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS

production classes	A
prior distribution	$P(A)$
AVA attribute distribution	B
marginal distribtution	$P(B)$
marginal distribution weight	W_m
conditional probability	$P(A \mid B)$
test attribute PDF	$P(Att)$
relative distance	d_{ij}

LIST OF ABBREVIATIONS

Society of Exploration Geophysicists . . . . . SEG

SEG Advanced Modeling Life of Field Model . . . . . SEAM

West African Reservoir . . . . . WAF

Naïve Bayesian Classifier . . . . . NBC

Multidimensional Scaling . . . . . MDS

Artificial Neural Network . . . . . ANN

Probability Density Function . . . . . PDF

ACKNOWLEDGMENTS

I would like to thank Chevron for providing the data and funding which allowed this project to happen. I would also like to thank my advisor Dr. Trainor-Guitton for always pushing me and thinking of creative pathways for this research to follow. Lastly, I want to give a thanks to my friends and family for all of the love and support the last two years, I could not have done this without you.

CHAPTER 1

INTRODUCTION

The primary goal of geophysical information in oil and gas exploration is to understand a reservoir and its surrounding geology. The knowledge gained from this understanding can be applied in a multitude of ways, from risk analysis to well planning. In this disseration a methodogy is developed and proved using statistical learning to predict future high grade drilling locations. A primary goal is for this work to be applied to new fields, where the well count is lower to aid the interpreter in the process of selecting better areas on interest. The work accomplished requires geophysical information as inputs, specifically a stochastic AVA inversion.

Amplitude variation with offest (AVO) or amplitude variation with angle (AVA) inversions have been popular in the oil and gas industry to determine earth properties of specific formations. These inversions are based upon approximations used to describe the wavefield response to certain media, like the Knott and Zoeppritz equations [1]. The theory for these equations dates back to the eary 1900s ([2], [3]). The equations are based on the primary wave (P-Wave) and shear wave (S-Wave) velocities, and densities of media to determine reflection information. These equations were later simplified to describe the reflection amplitudes and their relationship with incident angle [4]. Modern AVA analysis has reached as far as lithology predictions, hydrocarbon indicators, and fluid analysis. All of these are possible because of the change in pysical properties at boundaries in the Earth [5]. However, the inversions computed on the seismic information are often deterministic and describe the best fit to the data. There is a limitation to these inversions because they represents only a single point in the model space of the inverse problem.

This issue motivated the application of stochastic geophysical inversions, where the earth properties are sampled based upon set distributions and constrained to adhere to the ob-

served seismic data [6]. These methods have also been proven for use in joint inversions to include CSEM information [7]. Markov chain Monte Carlo methods can generate thousands of more random models. The main advantage of this process is increased sampling of the model space compared to a single solution. These solutions can also traverse through multi-modal functions of the model space, and not be stuck in local minimuma [6]. [8] describes a stochastic inversion solving for reservoir properties using AVA information and their results define the first step of this work.

The inversion process is summarized as follows. First, rock physics relationships are defined from related well logs in the field of interest. The primary relationships in the inversion exist between density, porosity, the ratio of P wave velocity over S wave velocity (Vp/Vs), and P wave impedance. These relationships are then compared with the reservoir formations to allow for the generation of lithotypes (brine sand, oil sand, shale, etc.). These defined relationships used for the inversion are converted to be a Bayesian statistical model. Using the statistical model allows for more realism concerning the earth, as many anomalies or exceptions can exist.

Markov Random Fields are used to provide prior probabilities of cells having the same lithotype as their neighbors. Markov chain Monte Carlo (MCMC) methods are used to generate many samples from the posterior distribution of the unknown parameters in the inversion. The sampling of the lithotypes uses a variety of methods from Metropolis-Hastings to Swendsen-Wang to allow for better convergence and increased efficiency in complex posterior distributions. The draws of the reservoir properties do differ a bit from the sampling of the lithotypes. For example, porosity is drawn using a Gibbs sampler because the full conditional probabilities can be derived analytically. The six-step procedure for MCMC sampling is described specifically in [7].

The number of samples for the inversion can be defined to be in the tens of thousands. Many steps in the chain are needed to reach convergence and generate realistic models. However, at the end of the inversion, there can be thousands of attribute models which

fit the original seismic information equally. Commonly done was computing P(10), P(50), P(90) models or calculating the median or mean of the attributes at each location. While this does provide a summary of the model space explored in the inversion, it eliminates some of the inherent advantages. The model space exploration is where the motivation of the work described in this dissertation is influenced. For the fields we had inversions for, there are hundreds to thousands of attribute values for all locations. A methodology is developed to use these many attribute models to predict future oil production in the reservoirs. In general machine learning has the potential for use in many areas of research due to its flexibility in application [9].

This thesis is primarily composed of a single research paper on the topic of this defined methodology. The paper begins with an in-depth introduction to the information available to be used in the statistical classifications. Then is a discussion about previous uses and the recent history of machine learning methodology. The two specific methods defined are a Naïve Bayesian Classifier, and Multidimensional Scaling. The theory for both of these are broken down, and each uses an example of their calculation for clarity of the procedure. The methods discussed are applied to an inversion from the SEAM Life of Field Model, and a West African Reservoir. The results for both fields are displayed in depth, with a focus on sensitivity testing and trying to understand the related uncertainty. The results are discussed, and conclusions are drawn to understand how well the selected classification algorithm performed.

Publications

Powers, H., Trainor-Guitton, W., and Hoversten, G. Michael. Predicting Oil Production with Machine Learning onStochastic AVA Inversions for the SEAM Life of Field Modeland a West African Reservoir: Geophysics, *To be submitted.*

Conference Paper Proceedings

Powers, H., Trainor-Guitton, W., and Hoversten, G. Michael. Classification of Total Oil Production of Wells in SEAM Life of Field from Stochastic AVA Inversion Attributes via Machine Learning: SEG Anaheim 2018.

CHAPTER 2

PREDICTING OIL PRODUCTION WITH MACHINE LEARNING ON STOCHASTIC AVA INVERSIONS FOR THE SEAM LIFE OF FIELD MODEL AND A WEST AFRICAN RESERVOIR

Hayden Powers, Whitney Trainor-Guitton, and G. Michael Hoversten

To be submitted to *Geophysics*

In the oil and gas industry, seismic surveying and subsequent interpretation hold substantial value when used for reservoir characterization and well planning. In addition to interpreting the migrated time and depth volumes, inversions on these data are used to determine earth properties in and around the reservoir formation. For this study, we utilize results from stochastic inversions of amplitude versus angle (AVA) gathers to train machine learning algorithms and predict cumulative oil production. The first inversion is from synthetic data in the SEAM Life of Field Model and the second from a survey in offshore West Africa. We use two algorithms, Multidimensional Scaling (MDS) and a Naï ve Bayesian Classifier (NBC), on the results from the inversions for the predictions and analysis. MDS is an unsupervised method which reduces dimensionality based on the similarities in the spatial structures around each wellbore. We implement the NBC as the supervised method to classify reservoir locations as high or low producing based upon known cumulative oil production values. Input to the NBC are the attributes from the AVA inversions. We use exhaustive cross-validation to calculate a bulk accuracy for each experiment. In general, the SEAM inversion has a lower bulk accuracy than WAF, but less variance associated with changing the cutoff to being a high producer. We use the results from MDS to evaluate the cutoffs in 2D space to understand the correlation between the decision boundary and the classification accuracy of each well or location. Overall, the combination of MDS and

a NBC prove to be effective for classifying total oil production at a limited computational expense.

2.1 Introduction

The use of Amplitude Variation with Offset (AVO) and Amplitude Variation with Angle (AVA) seismic data methods are commonly applied to differentiate fluid content of reservoir formations, with theoretical work coming from Knott and Zoeppritz at the beginning of the 20th-century [5]. A primary use of AVA is to determine lithological layers, i.e., gas bearing and non-gas bearing sands [10]. Currently, the advancements of AVO and AVA inversions extend to the prediction of specific earth attributes such as porosity and acoustic impedance [11]. However, doing so requires an input rock physics model to form proper predictions [12]. These types of inversions can be carried out stochastically to generate an ensemble of models from the seismic data and rock physics relationships [8]. An advantage of stochastic inversions is the increased sampling of the converged model space. However, even with hundreds and thousands of reservoir models, commonly a moment of the posterior distribution (e.g., Mean, Mode, or Median) is used, or the P10, P50, and P90 models selected. Machine learning offers techniques that make use of all the samples from the posterior distribution.

Naïve Bayes is often considered the baseline approach and has been proven in many different applications like spam detection and Alzheimer's prediction ([13], [14]). Naïve Bayes assumes all features are independent; however, this approximation is not detrimental to its accuracy [15]. Naïve Bayes is supervised, as the classes of the inputs are defined before training and testing. Naïve Bayes is known and used for its computational efficiency. Here we use it to predict high and low producing wells within a reservoir, if there is sufficient information to train from [16]. The second method considered in this study is Multidimensional Scaling (MDS). The primary objective of MDS it to reduce the dimensionality of information by computing a relative distance between data samples. MDS has been applied previously to understand differences between seismic responses of lithotypes in randomized channel fea-

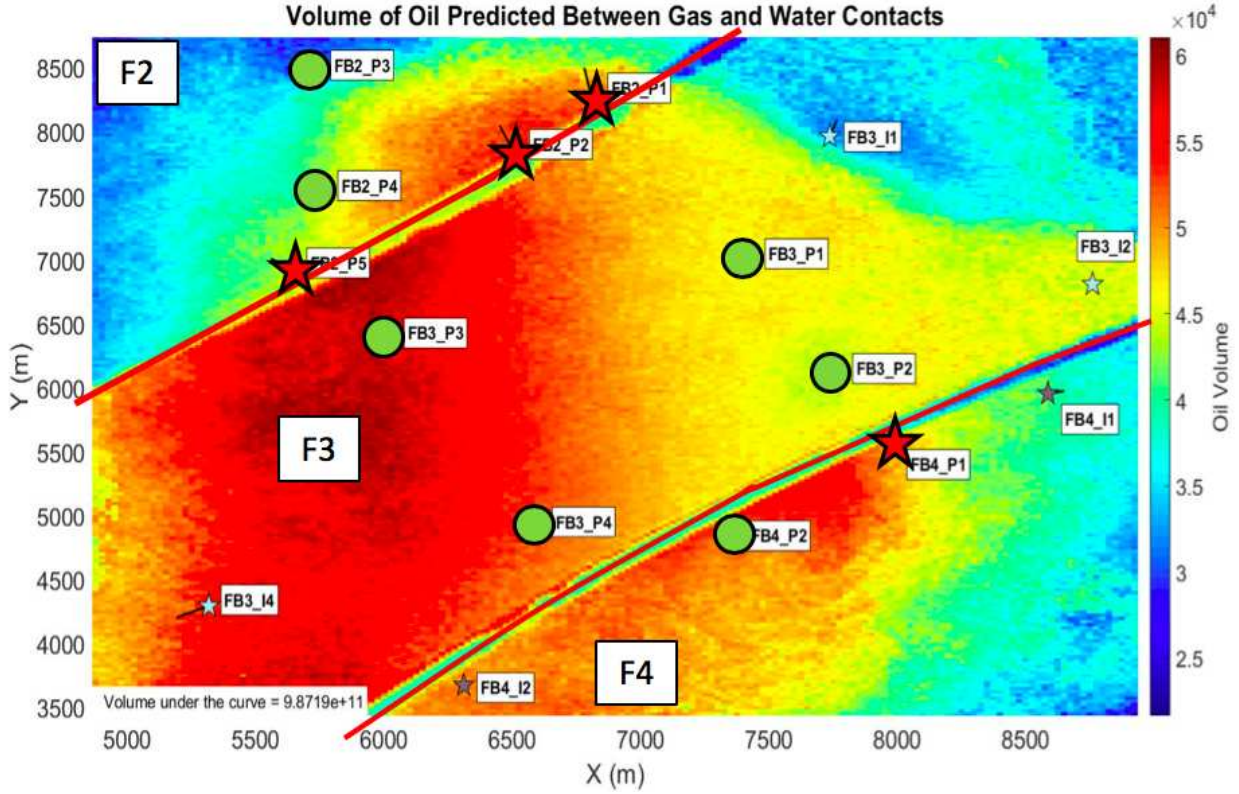

Figure 2.1 All production and injection wells in SEAM overlaid on total oil volume. Total oil volume calculation is done using the pore volume in the oil window of the reservoir. The three different fault blocks are also indicated by F2, F3, and F4. Wells covered by red stars indicate problem wells in the NBC classification, discussed later.

tures [17], or to explain uncertainty in alluvial transport systems [18]. A decomposition is performed on the input distance matrix to generate the best fit in a new lower dimensional space [19]. For example, the distance calculations (explained in the MDS section below) can be done within oil and gas reservoirs using earth density and porosity. An extension of this is to compare higher and lower producing wells in the low dimensional space to understand how similar or dissimilar sections of the reservoir are.

In this study we use the results from two separate inversions. Every model generated from the inversion is considered a sample or iteration; thus these two terms are used interchangeably. The first is from synthetic data on the SEAM Life of Field model (SEAM) and the second is from field data in offshore West Africa (WAF). The WAF data and AVA

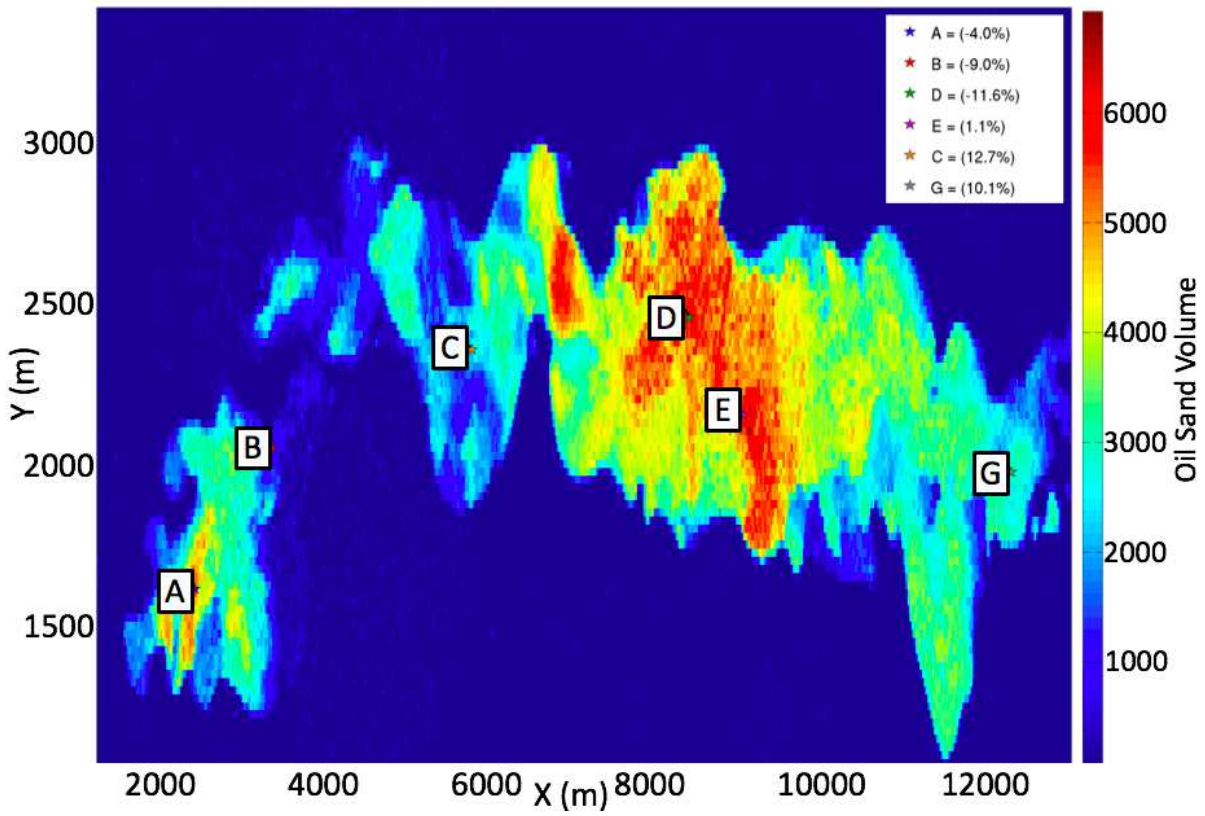

Figure 2.2 All West African Field wells labeled across the reservoir. The wells locations are displayed over oil in place, based upon interval thickness and porosity from the mode of the posterior distribution from the stochastic AVA inversion. Percentage labels show predicted error compared to log intervals.

algorithm are described in [20]. The SEAM inversion used seven angle stacks between 10.0 and 38.5 degrees, while WAF used five angle stacks between 12.5 and 52.5 degrees. Input to the machine learning algorithms are iterations confined to the oil producting zone from the chains of stochastic AVA inversion results. There are 5,000 samples from the chain for SEAM and 736 for WAF. Each iteration in the MCMC chain contains four attributes, density, porosity, Vp/Vs, and acoustic impedance. Density, Vp/Vs, and acoustic impedance are directly solved for, while porosity is drawn from a multi-Gaussian defined by the previous three [8]. MDS evaluates all four attributes individually across all samples. The Naïve Bayes Classifier uses each attribute simultaneously for classification. A cutoff of cumulative oil production is set to separate the high and low producing wells. The well locations are

displayed over maps of predicted oil volume in Figure 2.1 and Figure 2.2. The oil volume is calculated by summing the mode of the porosity distributions multiplied by the oil sand lithotype predicted by the stochastic AVA inversion [20]. The WAF training set is extended to include injection wells (assuming a relationship exists between high injectivity and high production) to provide sufficient data for training. Lastly, the volume of oil map can indicate potential new drilling locations, but we consider how a low-cost machine learning algorithm does.

The primary objective of this work is to predict future oil production of new wells at potential locations from two different fields by utilizing the 5,000 (SEAM) and 736 (WAF) samples available. First, the theory of the two machine learning algorithms is presented, with examples to provide context to the calculations. The defined algorithms are first applied to SEAM, and then to WAF in the results. Additionally, for SEAM the relative AVO error values over the entire reservoir are available for all iterations. These error values help in the quality control measures used to evaluate the results. Lastly, we conclude on the viability of using the NBC and MDS to make predictions for future drill locations in the respective reservoirs. The parameters of the algorithms for each reservoir are set to allow for comparison between the two.

2.2 Methods and Theory

This study uses two machine learning algorithms. The first is multi-dimensional scaling (MDS), an unsupervised method that accounts for spatial information, to transform attributes around each wellbore into a low-dimensional space for comparing and understanding fundamental inter-relationships. MDS receives no information for how to classify, unlike the Naïve Bayesian Classifier (NBC). Input to the NBC are stochastic inversion samples which are labeled based upon their associated production class (High, Low). Labeling of high and low is required to generate the Bayesian posteriors for each producing class. We base the labels of high and low production on the known cumulative oil production values from the available wells. Additionally, MDS allows a visual explanation for why wells are classified

appropriately or not. The two methods are picked to complement each other in the analysis.

2.2.1 Multi-Dimensional Scaling (MDS)

The objective of MDS is to take data with many dimensions and transform it into a low-dimensional space for interpretation [21]. Similar to Principal Component Analysis (PCA), MDS seeks to reduce the complexity of the data [22]. Additionally, MDS accounts for spatial variations by comparing distances between identical model cells of the grids around each well. We normalize the depth axis for each well to allow the respective oil producing zones to be directly compared. The calculated distances are used to transform the respective data into the new space [19]. The final number of dimensions is a defined input to the algorithm before decomposition.

2.2.1.1 Distance Calculation

The first step in MDS is the distance calculation. Many different distance equations are available, ranging from Hausdorff to cosine, or Euclidean [23]. For this study, we use Euclidean distance because of its innate simplicity and effectiveness. The Euclidean distance equation is shown in Eq 2.1.

$$d(\mathbf{a}, \mathbf{b}) = \sqrt{\sum_{i=1}^{n} (a_i - b_i)^2} \tag{2.1}$$

An example of **a** is the density values from Well FB2_P2 at iteration 1, while **b** would be the densities from FB2_P4 at iteration 53. We run through this for the samples of each well, so the number of **a**'s and **b**'s used for the calculation is equal to the iteration count multiplied by the number of wells used. The computation is done at each index through these vectors shown by a_i and b_i. The calculations are exhaustive through all wells and iterations for a given attribute. In this study, we never calculate distances between different attributes. A two-dimensional matrix houses the distances for all wells and their respective iterations. This two-dimensional matrix is the sole input to the MDS algorithm for decomposition [19].

The diagonal of this matrix is zero, where distances between identical wells and iterations are calculated.

An example of a final output from this process is displayed in Figure 2.3. The inputs to generate this figure consist of inversion iterations of gridded density values around nine wells from two areas of a reservoir. Each well has its iterations cluster closest to each other. The clustering of each well's iterations shows they are more similar to one another than they are to another well. A black dotted line is displayed as an example decision boundary which separates the wells from different areas. The axes shown have no physical meaning and are labeled as Y1 and Y2. The solution shown can be rotated or shifted without decreasing its accuracy.

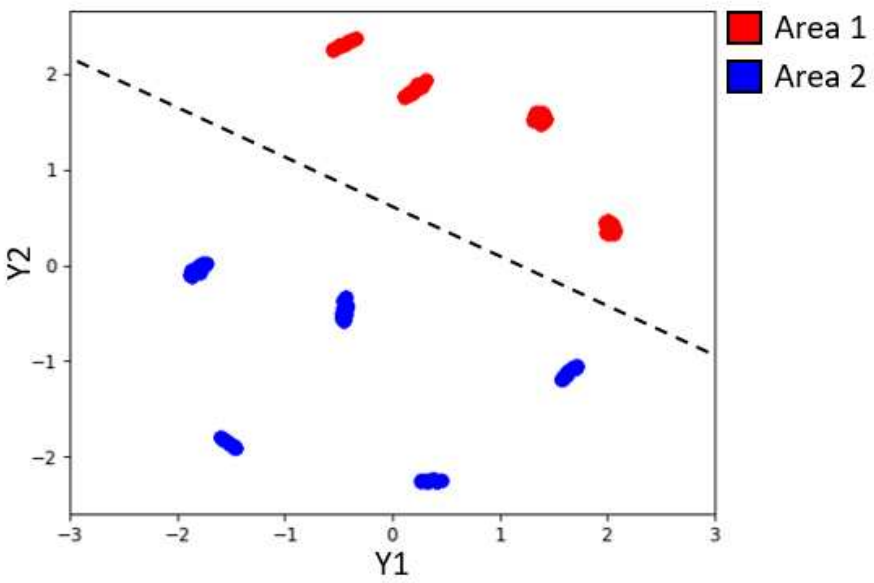

Figure 2.3 The 2D representation of distances calculated from density values of two areas in a reservoir. Areas 1 and 2 are represented by red and blue respectively. The black dotted line represents an example decision boundary that can divide the information in this space. The axes are labeled as Y1 and Y2 to represent the new non-physical dimensions.

2.2.2 Naïve Bayesian Classifier (NBC)

Naïve Bayesian Classification (Naïve Bayes) is a computationally efficient solution to many classification-based problems, like spam detection [13] or predicting production as we

do here. The key assumption of Naïve Bayes is that training sets and features are independent. While this assumption is incorrect for Earth attributes (e.g., density and acoustic impedance are correlated), it can still generate accurate results. Naïve Bayes is supervised and requires the input data to be labeled to train the posterior distributions [15]. Naïve Bayes generates a separate posterior for each input attribute. The likelihood functions for classification are probability density functions (PDFs) of the attribute values. The posteriors determine the class probabilities for each attribute, using a set of test likelihoods. We multiply the class probabilities from each attribute together to determine a classification value. The predicted class is set based on the index of the maximum classification value [15].

2.2.2.1 Generating Bayesian Posteriors

The basis of the NBC lies in Bayesian statistics and exploiting knowledge of the conditional probabilities and prior distributions. The generation of Bayesian posteriors for each input attribute used for classification is started by labeling data into the specific classes [15]. The first equation used in the NBC algorithm is shown in Eq 2.2. For this study, A is defined as the producing class (High, Low) and B is information about one of the attributes (density, porosity, Vp/Vs, and P-impedance). $P(A \mid B)$ is the conditional probability of having either production class based upon some given attribute information.

$$P(A = a_i \mid B = b_j) = \frac{P(B \mid A)\, P(A_w)}{W_m\ P(B)}, \quad A \in \{Low, High\},\ B \in [1, J] \tag{2.2}$$

$P(A_w)$ is the prior distribution of information from each producing class and can be varied to test the sensativity of the posterior distribution. $P(B)$ is the marginal probability of the given data from all the training information. B is discritized into J number of bins for calculation. The calculated weight

$$W_m = \frac{P(A_w)\, N_c}{n_c} \tag{2.3}$$

is used in the calculation of the posteriors for a given attribute and all defined classes are then calculated using Eq 2.2. For this, n_c is the number of data in its class, and N_c is the total number of data used to train the posterior.

An example of a posterior distribution created from this process is shown in Figure 2.4. These posteriors are specific to the density values for the defined iterations around all the training wells. In Figure 2.4 yellow represents $P(A = Low \mid B)$, and purple is $P(A = High \mid B)$, where B respresents the density bins. This display allows efficient evaluation of which bins are more related to each production class. For example, a density of 2.1g/cc would have just over a 60% probability of being related to a high producing well. These same types of posterior distributions are calculated for all attributes.

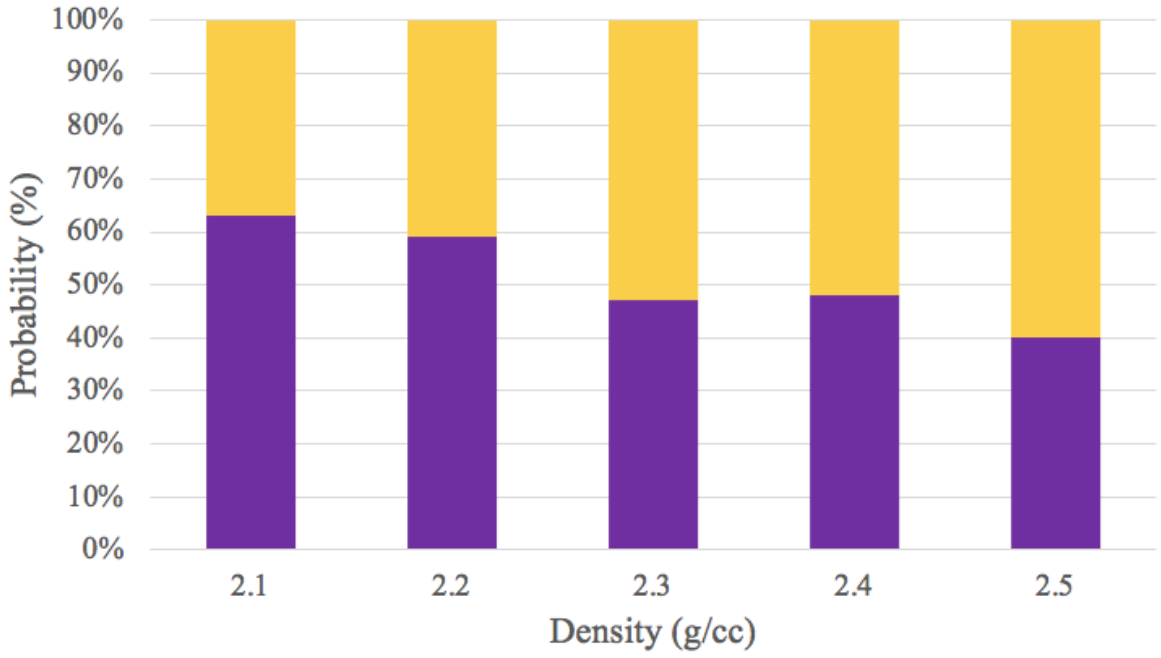

Figure 2.4 Example posterior distribution for two classes over five density bins. Purple represents the high producing class, $P(High)$, and the yellow is the low, $P(Low)$.

2.2.2.2 Classification

Determining the class of a new set of data is a straightforward process in NBC, as it only requires knowledge of the likelihood functions of the data for the attributes. These likelihood functions are multiplied with the previously calculated posterior distributions, creating probabilities of being a certain class based upon a single feature. The specific calculation is shown in Eq 2.4, where i is the class, j is the bin location, J is the total number of attribute bins, $P(A = a_i \mid B = b_j)$ is the calculated Bayesian posterior for one attribute at bin j, and $P(Att)$ is the probability density function (PDF) for the same attribute at a test location. We choose the notation of $P(Att)$ to make it more clear it is a test likelihood, and not part of the posterior information.

$$P(A_i) = \sum_{j=1}^{J} P(A = a_i \mid B = b_j) \cdot P(Att_j) \tag{2.4}$$

We calculate the classification values by multiplying the probabilities of each class from all the features. The classification of the data in question is based upon the index of the maximum classification value. The NBC is deterministic, so the classification probability of the maximum index is set to 1, while the rest are set to 0.

To continue the example set in Figure 2.4, we consider a set of likelihood functions for the attributes around an area of interest. The related density likelihood function is shown in Figure 2.5. As discussed, the likelihoods are multiplied and summed for each class shown in Eq 2.4. These are then used to generate the classification values used for the final prediction. The values from the example, with the additions of the other attributes, are shown in Table 2.1. We multiply these probabilities together to create the classification values in the final row. In this example, the final well classification for this example is a high producer.

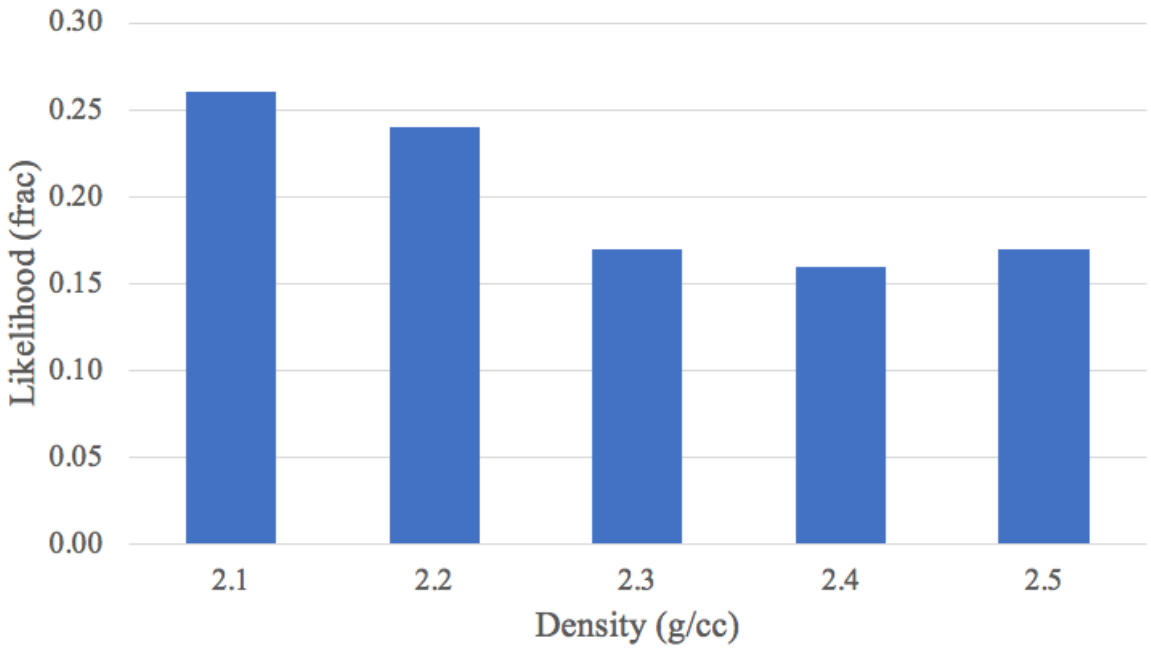

Figure 2.5 Example likelihood function of density values for a test location, $P(Att)$. The information is sorted into the same bins as the posterior distributions, B, and then normalized into a PDF.

Table 2.1 Results from the example classification across all four attributes. The highest final classification value is in bold to show the prediction.

Attribute	**P(High)**	**P(Low)**
Density	53.1%	46.9%
Porosity	55.4%	44.6%
Vp/Vs	46.2%	53.8%
Acoustic Impedance	51.1%	48.9%
Classification Value	**0.069**	0.055

2.2.2.3 Cross Validation

An important aspect of the NBC is determining the accuracy under a certain set of conditions. To accomplish this, we calculate a bulk accuracy through cross-validation to use in our assessment. Cross-validation is done by omitting a portion of the information available and using the rest for training. The cross-validation is repeated many times for different combinations of training data and predictions [26]. An example of this can be

omitting a certain number of wells in the production classification problem. The omitted wells are removed from the training and are subsequently classified. The NBC marks these classifications as correct or incorrect each time to determine the bulk accuracy. Additionally, if more than one well is omitted, then many different combinations exist which can be tested to solidify the classification accuracy further. We choose to perform the cross-validation exhaustively through all combinations to prevent bias in the bulk accuracy. A typical SEAM experiment omitted 4 of the 11 available wells, resulting in a total of 330 combinations to tested in the cross-validation, while West Africa has a total of 15 combinations from 2 of the 6 wells being omitted.

2.3 Results and Discussion

The results and discussion are in two sections, to separate the synthetic model and field data. We apply a NBC and MDS to both inversions, allowing for direct comparison. To reiterate, we use the suite of attributes from the stochastic AVA inversions as the inputs to generate our results. The attribute locations we use are extracted within a defined radius from each wellbore inside of the oil-producing window.

2.3.1 SEAM Life of Field (SEAM)

In SEAM, there are 16 total wells (11 producers, 5 injectors) distributed across three fault blocks. The two faults separating these blocks cut through the entire reservoir interval. For the analysis, we use only the oil-producing section of the reservoir. There are 5,000 samples (or iterations) from the converged chain of the inversion taken in a cylindrical region about each well, allowing every location within these cylinders to have up to 5,000 different values for use in MDS and the NBC. For SEAM we also have the last 100 iterations for the entire reservoir for classification. Included with this is the RMS data misfit by position from the inversion to help evaluate the results from the classifications. The normalized production values (RBBL) for the SEAM wells are shown in Figure 2.6.

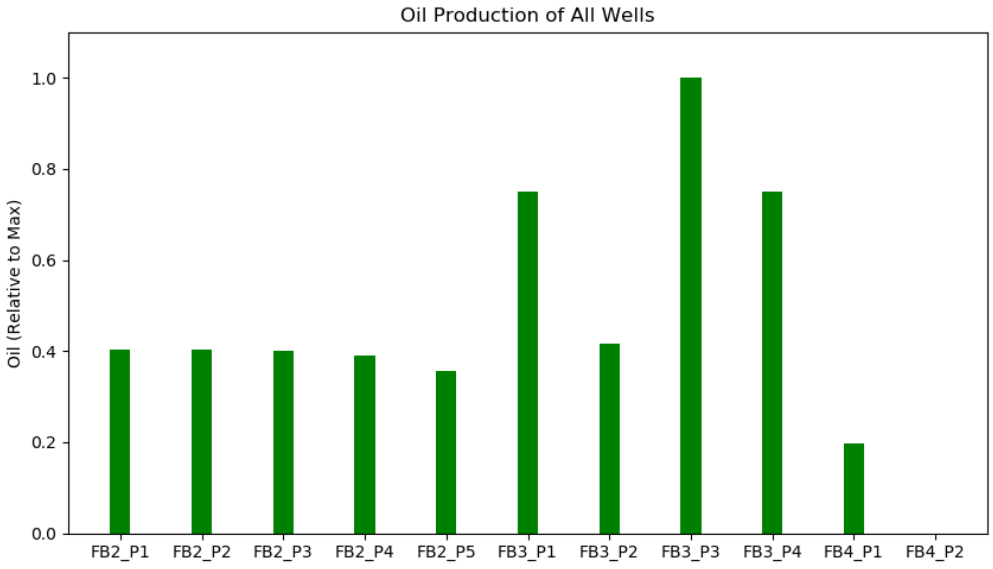

Figure 2.6 Normalized total oil production values for the life of the wells in SEAM. Wells in FB2 have similar production values. FB3 has the highest accumulations, while FB4 has the lowest.

2.3.1.1 Multidimensional Scaling (MDS)

We use the unsupervised approach of MDS to understand how the attributes within a 200m radius of the wells compared to each other in SEAM. In Figure 2.7 two key attributes, density (left) and porosity (right), are shown. Blue, green, and red indicate are fault block 2 (FB2), fault block 3 (FB3), and fault block 4 (FB4) respectively. Both plots show clusters at each well location, which represent 400 of the 5,000 inversion samples at each well. All 5,000 would have been used simultaneously, but the required memory exceeded our machines capabilities. Each of these 400 samples is within the converged chain of the inversion, and the clusters highlight the sampling in the converged model space.

Figure 2.7 has each well from SEAM labeled near its cluster. The coloring of the wells by fault block shows the separation between the different blocks in the reservoir. In both density and porosity, the wells from FB2 have higher values in Y2 than those from FB3. A linear decision boundary can be made to separate FB2 and FB3 in this space. Of the three

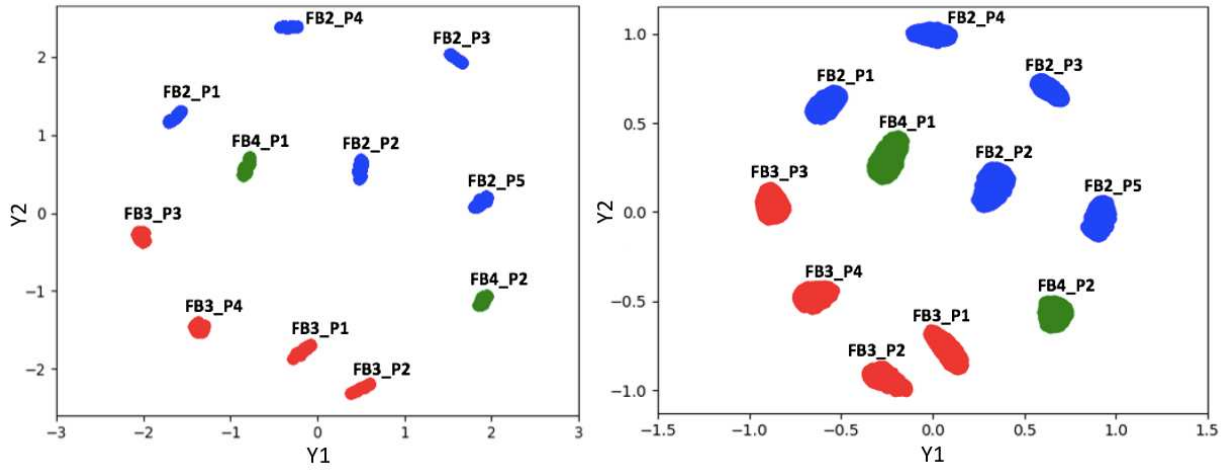

Figure 2.7 First 400 density samples around all wells within 200m compared to each other. Left uses the density values from the inversion iterations, right uses porosity. The colors of the clusters indicate their related fault blocks with blue, red, and green representing F2, F3, and F4 respectively.

fault blocks, the wells in FB4 have the largest average distance between them. This increase in distance indicates they are not similar when compared in this simplistic way, and would most likely be classified differently.

Apart from the comparison between all wells, we can refine the analysis down to a single well. An example of this for density and porosity across all samples at FB3_P3 is shown in Figure 2.8. By only evaluating a single well, we can extend the analysis to use all 5,000 inversion samples. The color of the plot indicates the sample number from the inversion with dark blue color as the first sample and dark red as the 5,000th. The horseshoe-like pattern is consistent among all wells and attributes, with porosity as an exception. For density, there are sets of iterations which possess more similar features than the others. These sections can be seen in the orange and cyan color of Figure 2.8 where the apparent notch cuts towards the center. The porosity plots with much higher variance around the edges but has the same smooth transition in color.

In the AVA inversion density, Vp/Vs, and acoustic impedance are solved for at each step in the MCMC chain. The porosity values are drawn from distributions generated based

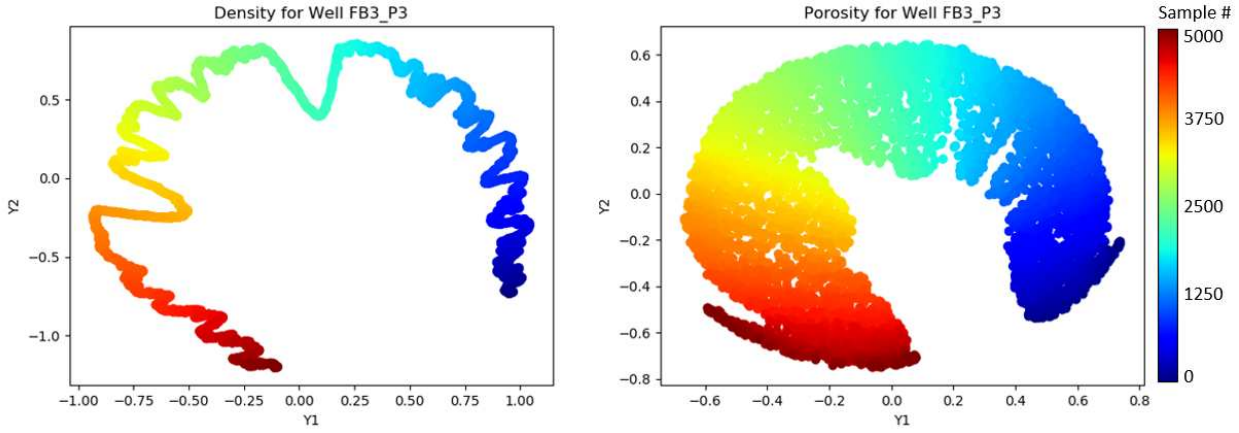

Figure 2.8 MDS on only FB3_P3 for density and porosity colored by sample number. Dark blue is the lowest, and the dark red is the highest. The samples appear to chain together in this two-dimensional space showing how the stochastic inversion utilizes information from the previous iteration to build the next model.

upon prior information about the other three attributes. The drawing of the porosity introduces some additional randomness in the result and explains the higher variance in the porosity values shown in Figure 2.8. The higher variance allows information from later inversion samples to be more similar to the initial ones. The results from both Figure 2.8 and Figure 2.7 will be important to understanding the classifications from our supervised method.

2.3.1.2 Naïve Bayesian Classifier (NBC)

The NBC classifies known wells and new locations as high and low producing. The wells are split into classes based on their total (normalized) production values shown in Figure 2.6. For this study, we determine a baseline experiment for comparison as parameters are adjusted to see their effect on bulk accuracy. The baseline test conditions for SEAM consist of using all 5,000 samples, a maximum radius of 100m from the well-bores, all four attributes, a decision boundary of 0.403 RBBL of cumulative oil (below 0.403 is classified as a low producer, above 0.403 is classified as a high producer), and omitting four wells in the cross-validation.

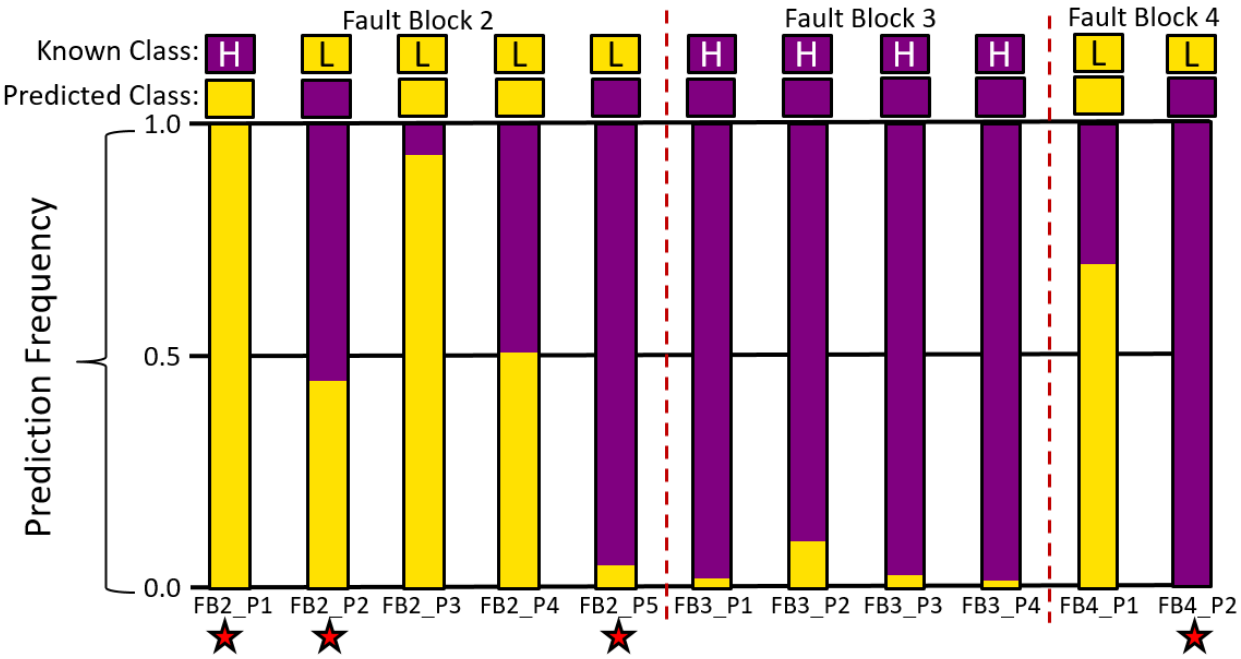

Figure 2.9 Proportion of the time each well is predicted as high or low producer from the baseline cross-validation. The purple denotes predictions of high production, while yellow indicates low. The known class is shown in the uppermost row, while right below it shows which class it is most often predicted as. Lastly, the four wells existing closest to the faulting planes are indicated with red stars at the bottom.

Figure 2.9 shows the classification results from the baseline cross-validation across all wells. The bulk accuracy of the baseline experiment is 59.2%. Each well has its classification proportions shown allowing us to evaluate where any misclassifications originate. The four wells with less than 50% accuracy are marked with red stars. All four of these wells are drilled near the fault planes in the reservoir. Two of the four wells have zero correct predictions and account for nearly half of the incorrect predictions. FB2_P3 is the furthest from the fault plane and the NBC classifies correctly in the cross-validation more than any of the wells in FB2. The wells within FB3 are all high producers in the baseline experiment and combine for over 90% prediction accuracy. FB3_P2 classifies as a low producer the most out of any wells in FB3 but is also the lowest cumulative oil producer in the block. It is important to note these results reflect only a single set of parameters for the cross-validation. We now shift our focus to understanding how the bulk accuracy varies with the parameters.

Table 2.2 First column shows the parameter(s) being changed, the second is the bulk accuracy, and third shows the specific wells with ≤50% accuracy. The baseline case uses a 100m radius, a cutoff of 0.403 RBBL, four wells omitted in the cross-validation, and using all four attributes.

Experiment	**Acc.**	**Wells with ≤50% accuracy**
Baseline	59.2%	FB2(P1, P2, P5,), FB4(P2)
No Density	60.0%	FB2(P1, P2, P5,), FB4(P2)
No Porosity	58.1%	FB2(P1, P2, P5,), FB4(P2)
No Vp/Vs	59.4%	FB2(P1, P2, P5,), FB4(P2)
No Acoustic Impedance	60.2%	FB2(P1, P2, P5,), FB4(P2)
50m radius	57.9%	FB2(P1, P2, P5,), FB4(P2)
200m radius	60.7%	FB2(P1, P2, P5,), FB4(P2)
0.300 RBBL cutoff	59.4%	FB2(P1, P2, P3), FB4(P1, P2)
0.600 RBBL cutoff	78.0%	FB2(P5), FB3(P2), FB4(P2)
0.600 RBBL, 200m	76.2%	FB2(P5), FB3(P2), FB4(P2)
0.600 RBBL, 50m	77.7%	FB2(P5), FB3(P2), FB4(P2)

Table 2.2 has results from a variety of experiments testing the different parameters set in the cross-validation. The left column describes the specific changes made, middle shows the bulk accuracy of the cross-validation, and the right column displays the wells predicted correctly less than 50% of the time. The first row of Table 2.2 shows the results of the baseline experiment, discussed prior. The first set of tests consists of omitting each attribute, leaving only three attributes for classification. Each one of the attributes, besides porosity, when removed increased the bulk accuracy. However, all of these changes in accuracy were less than a single percentage point.

The next parameter we test is the radius used for the attribute extractions around the wellbores. Increasing the radius to 200m raises the bulk accuracy to 60.7% while decreasing the radius to 50m reduced the accuracy to 57.9%. Adjusting the radius to these two values resulted in a greater change in accuracy than removing the attributes did, showing importance in the radius chosen.

Next, we evaluate two new decision boundaries by lowering it to 0.300 RBBL and raising it to 0.600 RBBL. These tests allow us to understand how the boundary affects the accuracy.

At the lower boundary of 0.300 RBBL the change in bulk accuracy from the baseline is minimal, but an additional well is predicted less than 50% of the time correctly. For this lower boundary, the only low producing wells are from FB4, and the NBC predicts both poorly. The high number of mispredictions on these two wells shows they are not good predictors for one another, and this would not be an ideal boundary for classification of the entire reservoir. At the higher boundary of 0.600 RBBL, the bulk accuracy of the cross-validation increases to 78.0%. At this new boundary we also test two additional radii, 50m and 200m. For the baseline decision boundary of 0.403RBBL, the accuracy increased when the radius was increased; however, at the higher boundary, we observe a different pattern. At 200m the bulk accuracy decreases slightly to 76.2%, and at 50m the bulk accuracy also decreases but to 77.7%.

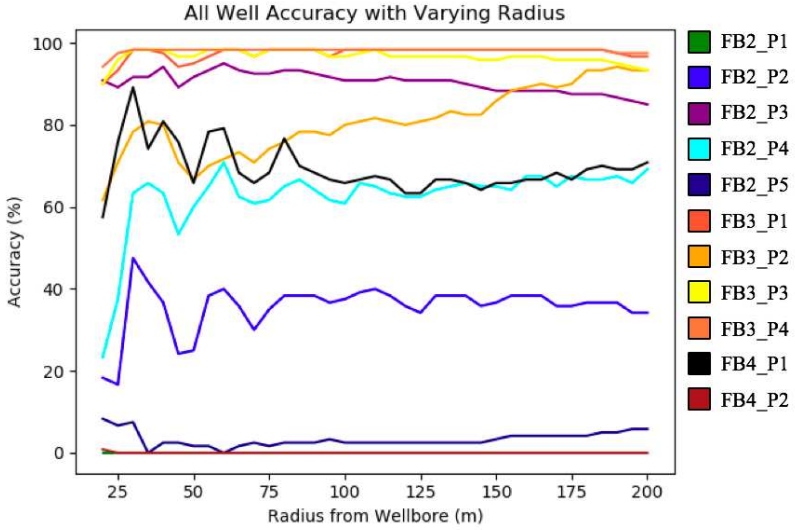

Figure 2.10 Bulk accuracy of each well in the baseline case. The radii are sampled in 5m intervals from 20m up to 200m.

The inconsistency in the relationship between radius and bulk accuracy shows it is more complex than a linear trend. Figure 2.10 shows the results from the baseline case but broken down by well across many radii. The radii in this plot range between 20m and 200m, sampled every 5m. Three of the wells from FB3 have the highest classification accuracy across all the

radii tested. The exception from this pattern is FB3_P2, which was also noted in Figure 2.9 as the well with the least prediction accuracy in the baseline case. However, the prediction accuracy of FB3_P2 on average increases with radius, after 50m. FB3_P2 is in the highest producing fault block in the reservoir, and increasing the radius adds more information into the likelihoods and posteriors during classification. The lower accuracy at the smaller radii shows the local information around FB3_P2 is indicative of being a low producer. While changing the radius showed an impact on the accuracy, the independence assumption of the NBC is another issue we explore.

Table 2.3 Bulk accuracy of results for each combination of two attributes for the baseline case on SEAM. The first row is the baseline case for comparison. The accuracies are in the middle column, and the absolute value of the mean correlation is in the last column.

Combination	**Bulk Accuracy (%)**	**Abs(Mean Correlation)**
All	59.24	N/A
Den and Por	60.23	0.9384
Den and Vp/Vs	58.78	0.9154
Den and Imp	**57.87**	**0.9654**
Por and Vp/Vs	61.59	0.8612
Por and Imp	60.30	0.8827
Vp/Vs and Imp	59.09	0.8938

Highlighted in Table 2.3 is a set of experiments to mitigate the independence assumption through feature reduction by using only a pair of attributes for classification [27]. We perform these experiments using the parameters from the baseline case. The left column shows the attributes used, the middle is the bulk accuracy of the experiment, and the right is the correlation between the two attributes. The correlations are included because higher correlation can lead to more overlap in the information they provide for the classification. This concept is highlighted in the table as many of the combinations with higher correlation have lower bulk accuracy when paired. In bold, the combination of density and acoustic impedance has the highest correlation and the lowest bulk accuracy. Conversely, porosity and Vp/Vs have the lowest correlation and the highest bulk accuracy of all the combinations.

The difference in accuracy between the best and worst scenarios is 2.81%, which is similar to the effect of changing the radius. Knowing how much these experiments cause the bulk accuracy to vary is important when interpreting the full reservoir classifications.

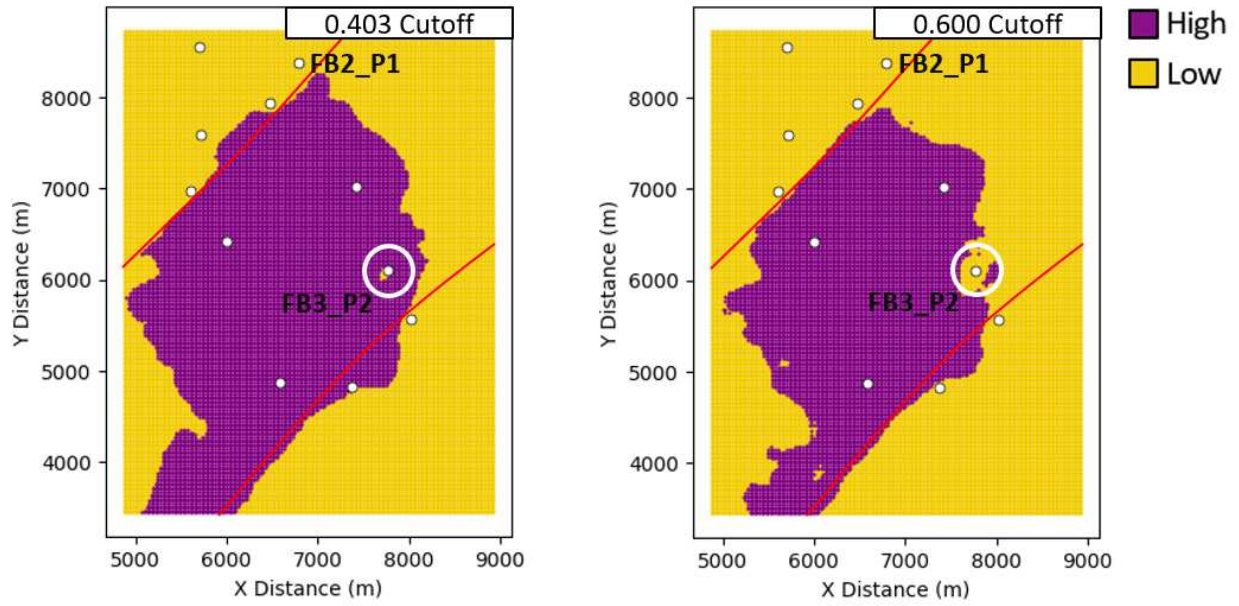

Figure 2.11 The classifications for SEAM using two decision boundaries. Left uses a cutoff of 0.403 RBBL to be a high producer and classifies much of the middle as high producing locations. The right uses a cutoff of 0.600 RBBL, which moves the two labeled wells from the high producing class and down to the low producing category. The most noticeable difference between the two is on the eastern flank of the high producing locations, where the higher cutoff has more conservative estimates. FB3_P2 is surrounded by low producing predictions and has the lowest cumulative oil produced of the wells in FB3. The bulk accuracy of each boundary in the cross-validation is 59.2% and 78.0% for the left and right respectively.

We perform full reservoir classiciations on SEAM by using the final 100 iterations from the stochastic inversion. The NBC uses a 100m radius around the wellbores and all 11 wells to build the posterior distributions for classification. For the prediction, we assume a perfectly straight wellbore through the oil sands of the reservoir. We classify using two decision boundaries, one at the baseline of 0.403 RBBL and the second at 0.600 RBBL. These two boundaries reflect the baseline case, and the higher accuracy boundary discussed prior, both are shown in Table 2.2. The results for the two classifications are shown in Figure 2.11.

Left shows the results of using the lower boundary and right is the results from the higher boundary, with yellow and purple indicating low and high predictions respectively. The white dots are plotted at the average wellbore location in the producing zone. We specifically mark wells FB2_P1 and FB3_P2 for discussion. The 0.403 RBBL boundary (left) has both FB2_P1 and FB3_P2 as high producing wells. In the baseline cross-validation well FB2_P1 was predicted correctly 0% of the time. The area around the average location of FB2_P1 is composed of many low-class predictions, but there is a small high producing lobe that extends close to it. When we increase the boundary, FB2_P1 becomes a low producer. As a result, the nearby high-producing predictions move further back. Second, we evaluate FB3_P2 which classifies the same as FB2_P1 for both boundaries.

The lower cutoff of 0.403 does yield accurate predictions for FB3_P2, but an interesting characteristic is noted. The white circles drawn on both sides of Figure 2.11 are to highlight the changes around FB3_P2 for each boundary. When the boundary is set to 0.403 RBBL, there are low-producing classifications around FB3_P2. The presence of low-class predictions indicates some pattern existing in the reservoir that is restricting oil production. When we increase the boundary to 0.600 RBBL, the area around FB3_P2 is almost entirely low-class predictions. As described before (see Figure 2.6) FB3_P2 has the lowest cumulative oil production in FB3. The NBC is able to pick up on a pattern related to the attributes and able to make predictions accordingly. Finally, in the 0.600 RBBL cutoff all the average wellbore locations displayed exist in locations which predict their production correctly.

2.3.1.3 AVA RMS Data Misfit

For SEAM we consider the RMS data misfit values across the reservoir for all 5,000 iterations. The error related to the AVA response of the inversion is important to consider because it gives us insight to locations where the inversion is less successful at fitting the data.

Plotted in Figure 2.12 is all SEAM producing wells and their related RMS values within 200m across all 5000 iterations from the inversion. The wells in the display are colored based

upon their predictions in the baseline decision boundary discussed previously. The red lines indicate wells which were predicted more than 50% of the time inaccurately, while the dark green indicates the opposite. The three wells with the largest RMS values, and where the RMS is still decreasing with sample number, are also wells which are predicted poorly in the cross-validation. There is one well which is red at the bottom of the RMS, indicating an outlier to the pattern we observe. The three red wells at the top of the figure also have a similar slope to their RMS values, as they all decrease for the iteration number. All four of these issue wells are near fault planes where it is more difficult to image due to diffractions and lateral velocity changes. The RMS vs. iteration number for other 8 wells are much flatter, suggesting their values are within the converged model space.

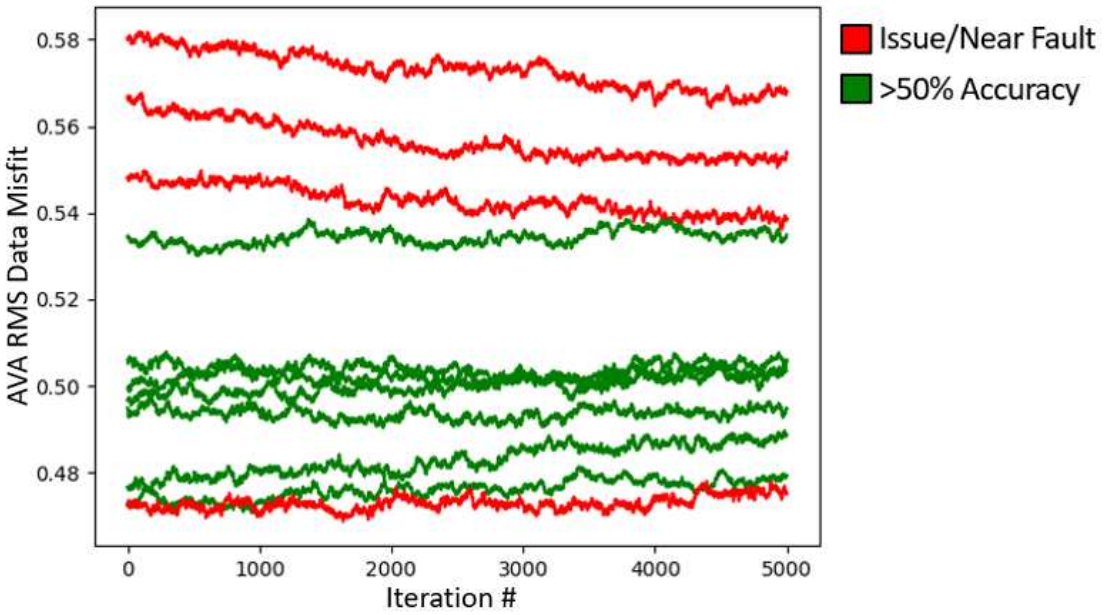

Figure 2.12 Curves for each well across all iterations of the RMS information. These values were calculated by taking a mean within 200m of the average wellbore location. The red lines indicate the wells which were classified poorly, with the green lines representing the opposite. All three red lines in the high RMS values are from FB2, while the red line near the bottom is from FB4.

The stochastic inversion for this data set was complete prior to the design of this study. In retrospect, it is clear that the convergences of the Markov chains at each well should be

considered if this type of analysis is to be performed since three of the four wells with the most inaccurate production predictions have RMS values that have not converged.

2.3.2 West African Reservoir (WAF)

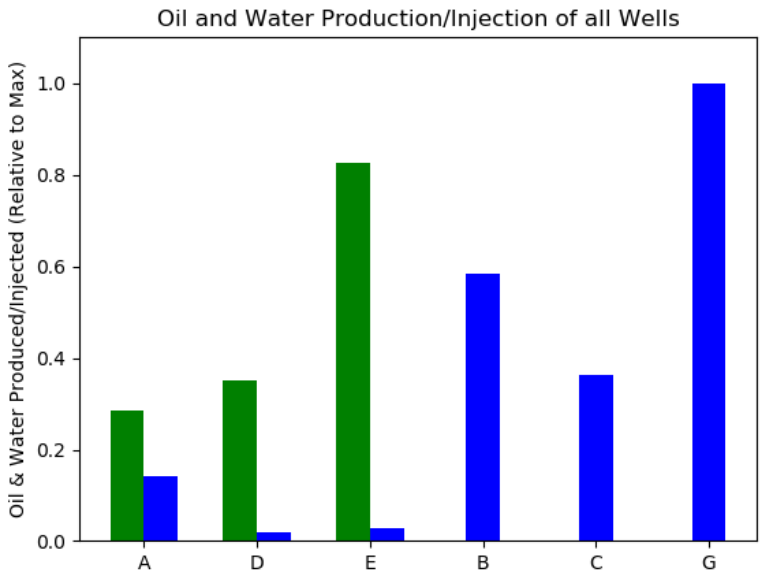

Figure 2.13 Normalized cumulative oil and water production/injection values from WAF. Productions wells D and E have minimal water cut, with Well A as an outlier. Injection well G has the largest cumulative volume of water injected.

The AVA inversion for WAF used the same algorithm as SEAM, only with field data and a different rock physics model. There are 6 wells available to us in this field (3 producers, 3 injectors). The normalized production values (RBBL), as of late 2017, are shown in Figure 2.13. This field is also less structurally controlled and consists of more stratigraphic trapping in the channeling systems present. This inversion has 736 samples from the converged chain for use. A significant challenge with WAF is only having three producing wells, as it is impossible to define a fully testable decision boundary for cross-validation. Since three producing wells is too few, we expand the training set to include the injection wells. The expansion of the training wells assumes injectivity and productivity would be directly related if oil was present at the injectors. By mixing injectors and producers, the predictions

become a reservoir quality prediction rather than a production prediction. Adding the three wells allows us to calculate a meaningful cross-validation.

2.3.2.1 Multidimensional Scaling (MDS)

The wells for WAF are not cut off by large fault blocks like SEAM but exist in different sections of reservoir sand from a channel system. Of interest were the porosity and acoustic impedance results, shown in Figure 2.14. Left is acoustic impedance and right is porosity. The structures of the clusters between the two plots vary significantly. The P-impedance has a more stacked structure as related to the producing class (high producers all with low Y1 values, low producers all with high Y1 values), while the porosity has a more circular pattern. Porosity has Well G significantly far from the other two high producers in Y2, as it sits closer in Y1 to the other two injection wells, B and C. Well C is plotted with the lowest values in Y2 of all the low producers, with respect to acoustic impedance, and is closest in porosity to the high injection well. These two factors provide a preliminary warning for future mispredictions as we observe in the NBC results.

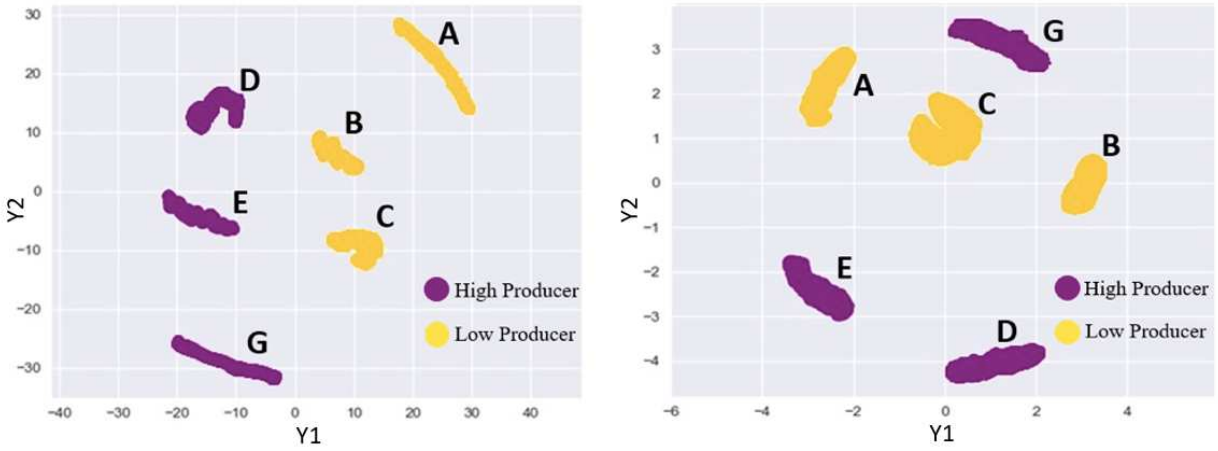

Figure 2.14 MDS applied to all 736 inversion samples of the six wells. Left uses the acoustic impedance. Right is calculated from porosity. Low and high producing wells are shown in yellow and purple respectively.

2.3.2.2 Naïve Bayesian Classifier (NBC)

The lower well count of the WAF data had a significant impact on our abilities to test the boundaries of the NBC. As discussed previously, the injectors are included as pseudo-producers. We also could not define a single decision boundary, but rather one for the producers and one for the injectors. It is apparent Well A has a significantly higher water cut, while D and E are still primarily producing oil (see Figure 2.13). The separation of A from the others is in the water cut and ultimately leads us to classify it as a low producer. Using injecting wells as pseudo-producers is justifiable from the relationship between injection rate and permeability/porosity of the reservoir formation. Well G is by far the best injection well and we therefore classify it as a high producer. The results of these decision boundaries and a 100m radius are shown in Figure 2.15.

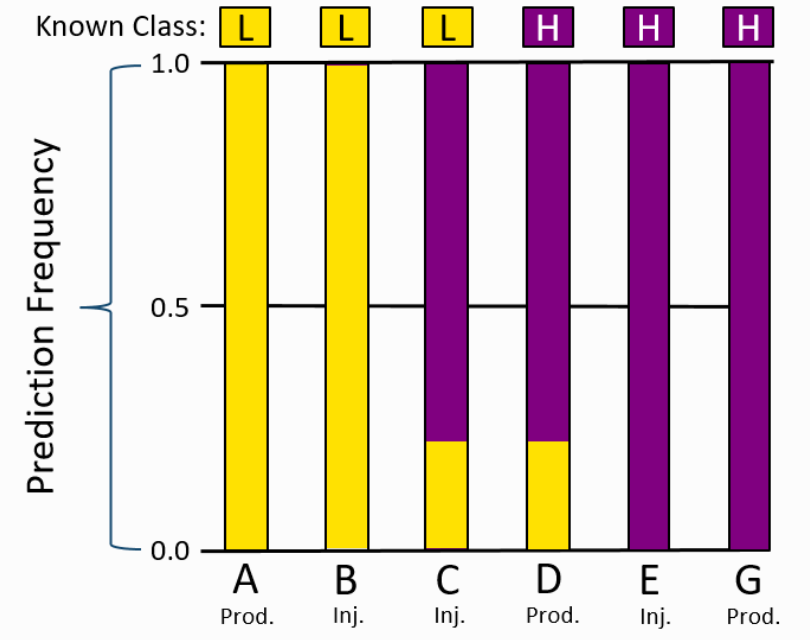

Figure 2.15 Individual well predictions for the cross-validation on the WAF inversion. The bulk accuracy for this experiment was 83.3%, with error spread across only two wells. Injector C is predicted as a high producer 80% of the time, and Producer D is predicted as a low producer 20% of the time.

This figure displayed uses the same color scheme as SEAM to allow for direct comparison. For WAF, we omit two of the six wells for the corss validation and the bulk accuracy for this baseline scenario is 83.3%. In this test of the NBC, only one producer and one injector have incorrect predictions. The prediction accuracy of the producers is over 90%. The majority of the incorrect predictions are related to Well C. We note in the results from MDS that Well C does not cluster convincingly with the other low producers and we see that reflected in Figure 2.15. Overall, using the injection wells as pseudo-producers is not perfect, but the assumption maintains a high level of accuracy in the NBC.

2.4 Conclusion

This study demonstrates the use of two machine learning methods used to evaluate high-grade drilling locations in producing reservoirs. The inputs to the machine learning algorithms are four attributes derived from stochastic AVA inversion. At every location within the models, 5,000 (SEAM) and 736 (WAF) samples from the converged chain are available. MDS proved valuable in understanding how the attributes within the oil window change across the reservoir. The key aspect of the MDS is how it incorporates relative spatial information. When implemented for SEAM, significant clustering was shown for both FB2 and FB3. However, the two producing wells from FB4 were more separated. The second algorithm we considered, Naïve Bayesian Classifier, is used to predict wells locations as high or low producing. The decision boundary between high and low producing wells is based upon the total cumulative oil production of the wells and yielded accuracies as high as 78%. In WAF, the inclusion of the injection wells as pseudo-producers requires a second, separate, boundary to be defined. The injecting wells proved useful for training and classification in the NBC and resulted in an 83.3% accuracy. Including injectors indicates there is the possibility of extending this approach for predicting reservoir quality to newly producing fields, where the well count is low.

For both SEAM and WAF, the sensitivity of NBC to many of its parameters was tested to determine the variance in the accuracy of the classifier. The higher well count in SEAM

allowed the testing to yield more robust results. The NBC did have a change in accuracy with respect to the parameter changes, but none compared to the impact the decision boundary made. The general quality of each decision boundary led us back to the MDS results, where changes were often explained. Overall, allowing the two different algorithms to work together to generate results truly was beneficial in giving meaning to the accuracy seen in the cross-validation.

Despite the success of the results, there is considerable future work to be done. Multidimensional Scaling was the unsupervised method that guided the understanding of the decision boundaries used for the NBC. However, both algorithms are on the lower end of computational expense. Additionally, the accuracy of the NBC may not be good enough when implemented to guide multi-million-dollar decisions. Future work will consider more complex, and computationally intense, classification algorithms, such as Artificial Neural Networks and generative-adversarial networks. These approaches can use both attribute values and location information as features to predict total oil production. We see this as the next step to highlight potential drilling locations from the outputs of the stochastic inversion.

2.5 Acknowledgements

We would like to thank Chevron for their continued funding and support on this project. They have allowed this work to be done by giving us access to their data for the analysis.

CHAPTER 3

CONCLUSION AND FUTURE WORK

This thesis developed a method of taking the output attributes from a stochastic AVA inversion and using them in machine learning algorithms to predict future oil production. The methodology was tested in both field and synthetic data cases. The synthetic field (SEAM) was the primary test field for the algorithms. SEAM had 11 producing wells, all of which had associated cumulative oil production values. The 11 wells proved sufficient for training and classification in the Naïve Bayesian Classifier when applying an exhaustive cross-validation. However, when breaking down the cross-validation results on a well-by-well basis, it was observed that only a few of the wells were responsible for the majority of the misclassifications.

The unsupervised MDS method provided an explanation to why these wells were not being classified accurately. The wells ended up falling between the many of the other wells of the alternative class in the low-dimensional space, showing that they were similar on a more basic level. Additionally, these wells often landed near reservoir fault planes. The inversion does not explicitly account for fault planes when calculating the covariance during a given step in the chain. The calculation allows lithological information from different fault blocks to affect each other out to the chosen search radius. Secondly, when we include information for the training our algorithms, we do not account for fault planes and can include information across the plane when describing certain well locations. When we classified the entire reservoir using the NBC, this became apparent as there were some predictions of the high producing class where we would not expect.

A full analysis was also carried out on the field inversion, WAF, but due to a few limitations, the study was limited. WAF had three producing wells, which are still producing to this day. The cumulative production information used to make the decision boundaries

was based on the current rates and the total production, but not a lifetime value like we had for SEAM. Additionally, the training and testing set for WAF was extended to include the injection wells to have enough data points to run a meaningful cross-validation. The addition of the injection wells as pseudo-producers did yield accurate results and suggests this to be a viable option in future fields with low well counts. Unfortunately, the full AVA inversion results for WAF were unavailable, so classification of the entire reservoir was not possible.

Overall, there are a few directions this project could go for future work. While Naïve Bayes is not the most sophisticated supervised machine learning algorithm, it proved viable for making binary predictions. The NBC also gives us an excellent baseline for how accurate supervised algorithms should run. By having a baseline for the expected accuracy, a value can be placed upon the computation time of the NBC and its future competitor to determine additional computational overhead. There has been some early experimentation done utilizing an artificial neural network to perform binary classification (See Appendix A). The initial results show improvement over the NBC; however, the randomness of the initialization of the network is proving to be an issue for generating consistent results. Continuing work on the neural network would allow for a direct comparison to the NBC, and potential for better results, but at the cost of an increased computational expense.

In summation, the NBC proved to be able to accurately classify locations in the reservoir as high and low producing despite its assumption of independence. Multidimensional scaling was shown to be valuable in understanding why certain decision boundaries do or do not work. In the case of SEAM, coloring the MDS results for a single well gave us insight to the sampling of the inversion through each iteration as jumps were noticed when the RMS increased near the wellbores. A Naïve Bayesian Classifier is a legitimate supervised algorithm for predicting high and low producing locations from the stochastic AVA inversion, giving the hundreds and thousands of samples from the reservoir more value.

REFERENCES CITED

[1] Robert E Sheriff and Lloyd P Geldart. *Exploration seismology*. Cambridge university press, 1995.

[2] Cargill Gilston Knott. Iii. reflexion and refraction of elastic waves, with seismological applications. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 48(290):64–97, 1899.

[3] K Zoeppritz. On the reflection and propagation of seismic waves. *Gottinger Nachrichten*, 1(5):66–84, 1919.

[4] Roo Bortfeld. Approximations to the reflection and transmission coefficients of plane longitudinal and transverse waves. *Geophysical Prospecting*, 9(4):485–502, 1961.

[5] Hong Feng and John C Bancroft. Avo principles, processing and inversion. *CREWES Research Report*, 18:1–19, 2006.

[6] Klaus Mosegaard and Albert Tarantola. Monte carlo sampling of solutions to inverse problems. *Journal of Geophysical Research: Solid Earth*, 100(B7):12431–12447, 1995.

[7] Jinsong Chen, G Michael Hoversten, Donald Vasco, Yoram Rubin, and Zhangshuan Hou. A bayesian model for gas saturation estimation using marine seismic ava and csem data. *Geophysics*, 72(2):WA85–WA95, 2007.

[8] Jinsong Chen and G Michael Hoversten. Joint inversion of marine seismic ava and csem data using statistical rock-physics models and markov random fields. *Geophysics*, 77 (1):R65–R80, 2012.

[9] Hua Wang, Cuiqin Ma, and Lijuan Zhou. A brief review of machine learning and its application. In *2009 International Conference on Information Engineering and Computer Science*, pages 1–4. IEEE, 2009.

[10] WJt Ostrander. Plane-wave reflection coefficients for gas sands at nonnormal angles of incidence. *Geophysics*, 49(10):1637–1648, 1984.

[11] Daniel P Hampson, Brian H Russell, and Brad Bankhead. Simultaneous inversion of pre-stack seismic data. In *SEG Technical Program Expanded Abstracts 2005*, pages 1633–1637. Society of Exploration Geophysicists, 2005.

[12] Simmons and Backus. Waveform-based avo inversion and avo prediction-error. *GEOPHYSICS*, 61(6):1575–1588, 1996. doi: 10.1190/1.1444077. URL https://doi.org/10.1190/1.1444077.

[13] Wei Wei, Shyam Visweswaran, and Gregory F Cooper. The application of naive bayes model averaging to predict alzheimer ' s disease from genome-wide data. *Journal of the American Medical Informatics Association*, pages 370–375, 2011.

[14] Jeremy J Eberhardt. Bayesian spam detection. *Scholarly Horizons: University of Minnesota, Morris Undergraduate Journal*, 2(1):2, 2015.

[15] Irina Rish. An empirical study of the naive Bayes classifier. *Empirical methods in artificial intelligence workshop, IJCAI*, pages 41–46, 2001. ISSN 14639076. doi: 10.1039/b104835j. URL http://www.cc.gatech.edu/home/isbell/classes/reading/papers/Rish.pdf.

[16] Hayden Powers, Whitney Trainor-Guitton, and G. Michael Hoversten. *Classification of total oil production of wells in SEAM Life of Field from stochastic AVA inversion attributes via machine learning*, pages 2131–2135. SEG, 2018. doi: 10.1190/segam2018-2998397.1. URL https://library.seg.org/doi/abs/10.1190/segam2018-2998397.1.

[17] Céline Scheidt, Cheolkyun Jeong, Tapan Mukerji, and Jef Caers. Probabilistic falsification of prior geologic uncertainty with seismic amplitude data: Application to a turbidite reservoir case. *Geophysics*, 80(5):M89–M12, 2015.

[18] Céline Scheidt, Anjali M Fernandes, Chris Paola, and Jef Caers. Quantifying natural delta variability using a multiple-point geostatistics prior uncertainty model. *Journal of Geophysical Research: Earth Surface*, 121(10):1800–1818, 2016.

[19] Patrick JF Groenen, Michel van de Velden, et al. Multidimensional scaling by majorization: A review. *Journal of Statistical Software*, 73(8):1–26, 2016.

[20] GM Hoversten, A Royle, J Chen, and D Myer. Mcmc inversion of offshore west africa ava data. In *79th EAGE Conference and Exhibition 2017*, 2017.

[21] Ingwer Borg and Patrick Groenen. Modern multidimensional scaling: Theory and applications. *Journal of Educational Measurement*, 40(3):277–280, 2003.

[22] Hao Guo, Kurt J Marfurt, and Jianlei Liu. Principal component spectral analysis. *Geophysics*, 74(4):P35–P43, 2009.

[23] Kai Yang and Jayant Trewn. *Multivariate statistical methods in quality management*, volume 21. McGraw-Hill New York, 2004.

[24] Joseph B Kruskal. Nonmetric multidimensional scaling: a numerical method. *Psychometrika*, 29(2):115–129, 1964.

[25] Jan De Leeuw and Willem J Heiser. Multidimensional scaling with restrictions on the configuration. *Multivariate analysis*, 5(1):501–522, 1980.

[26] Frederick Mosteller and John W Tukey. Data analysis, including statistics. *Handbook of social psychology*, 2:80–203, 1968.

[27] Saurabh Mukherjee and Neelam Sharma. Intrusion detection using naive bayes classifier with feature reduction. *Procedia Technology*, 4:119–128, 2012.

[28] By A M Turing. Computing machinery and intelligence. *Computing Machinery and Intelligence*, 49, 1950.

[29] Leo Gugerty. Newell and simon's logic theorist: Historical background and impact on cognitive modeling. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 50(9):880–884, 2006. doi: 10.1177/154193120605000904.

[30] Murray Campbell. Deep blue. *Artificial Intelligence*, 134(1):57 – 83, 2002.

[31] Imad Basheer and Maha N Hajmeer. Artificial neural networks: fundamentals, computing, design, and application. *Journal of microbiological methods*, 43 1:3–31, 2000.

[32] F Rosenblatt. The perceptron : A probabilistic model for information storage and organization. *Physcological Review*, 65(6):386–408, 1958.

[33] Kunihiko Fukushima. Cognitron: A self-organizing multilayered neural network. *Biological Cybernetics*, 20(3):121–136, Sep 1975. doi: 10.1007/BF00342633.

[34] Elizabeth Gibney. Google ai algorithm masters ancient game of go. *Nature News*, 529 (7587):445, 2016.

[35] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[36] Robert V Hogg, Elliot A Tanis, and Dale L Zimmerman. *Probability and statistical inference*, volume 993. Macmillan New York, 1977.

APPENDIX

ARTIFICIAL NEURAL NETWORK APPLICATION

A.1 Origin and Background of Neural Networks

The ideas for artificial intelligence have been around for much longer than the modern applications we see today. In the early 1940s and 1950s when computers were in their infancy, the potential for their capabilities was already being tested. During this time Alan Turing theorized how a computer could mimic human characteristics. The related idea would need a form of validation to see if a computer was capable of such a feat, known as the Turing Test [28]. Work continued in the field of artificial intelligence (AI), and in the late 1950s, Newell and Simon invented what is believed to be the first AI program, the Logic Theorist [29]. The goal of the program was to follow the decision making of the brain and solve complex problems. These early examples had one major drawback, which was the limited processing power of available computers.

With increasing computational power came some astounding results in the AI world. A significant milestone came when IBM's Deep Blue won a game of chess against the current world champion in 1997 [30]. Nobel advances would continue over a couple of decades as computers became faster and more capable. Within all of the AI advancements, there is a specific area of study in neural networks. Neural networks work to match how neurons in the brain fire based upon input information. These algorithms are capable of solving complex problems, but can be expensive to run.

The evolution of neural networks followed a similar advancement timeline as AI. The specific goal of neural networks is to match neurological functions existing in the brain [31]. The brain has many complexities, but the principle of repeatedly activating an electrical signal to perceive input information is something that can be mimicked by a computer. One of the first examples of a neural network was Rosenblatt's Perceptron from 1957. The goal

of the Perceptron was to take in images and perform a simple classification [32]. However, increasing the number of classes required a more complex network with additional layers, which created the concept of a multilayered perceptron. The first example of an unsupervised multilayered network named Cognitron came almost two decades later in 1975 [33]. The application for neural networks has grown immensely and consists of many categories. The types of problems vary from image classification, forecasting, optimization, and clustering analyses [31]. A recent achievement in 2016 was Google's Brain defeating a professional player at a game of Go. Go was considered unsolvable for a long time due to its inherent complexity and many moving pieces [34]. The feat accomplished highlights how far neural networks have come, and shows the potential for future applications especially those in earth sciences.

A.2 Why Use an Artificial Neural Network (ANN)

The goal of using an ANN is to take advantage and test a more computationally expensive supervised method. The previous supervised method of a Naïve Bayesian Classifier (NBC) worked with 78.0% accuracy on SEAM and 83.3% on WAF. While the NBC was able to predict the production of the wells accurately for WAF, it was not able to for SEAM. Additionally, the NBC was extremely cheap to run, taking only a matter of seconds on a local processor. However, this efficiency comes at the cost of assuming independence between all features and lumping of information. Independence is needed when calculating the classification value, and lumping is used for generation of the posterior distributions. Both are far from ideal and are what motivate the future switch to using an ANN.

The data used as inputs for the ANN are the same as what was used for the NBC and MDS. To reiterate, the inputs consist of reservoir attributes from stochastic AVA inversions over two producing fields, SEAM and WAF. The iterations, or samples, come from the converged chain of the inversion. Each inversion sample can have different algorithms for the selection of the attributes, as discussed in Chapter 1. Each iteration is different from the previous and the following one, with greater variance noted as the iteration count increases.

In NBC, many samples were used simultaneously to train and predict oil production. However, when using an ANN, we can use each iteration from the inversion individually during the training of the network, keeping much more physical continuity in the algorithm. As was done for the NBC, all model locations within a defined radius of the wellbores are extracted.

Wells are divided to be high and low producers based upon a cumulative oil production value, or total water injected. The split wells are then used to train the ANN and validate the accuracy over a set number of epochs (see below). It is important to note that with each well having hundreds of samples, the wells are split before the iterations are decided for training and validation. If Well A is a validation well, no iterations from Well A are used for training, allowing for the network to fully blind to that Well A's reservoir information.

This Appendix makes, a more in-depth discussion of the data. This discussion includes how the information is formatted for input into the ANN. While this formatting is similar to the previous methods, there are some specific and necessary changes made for the ANN. Next, the general theory and methods for the ANN are explained. These include many of the parameters which set to create the ANN. The uncertainty and confidence evaluations are also discussed, which consist of a k-fold cross-validation and initialization redundancy. Next, many of the experiments are shown in the results. A few main networks were tested, with the primary focus being on the bulk accuracy and sensitivity testing different parameters. These results are evaluated to explain observed phenomena. Creating a fully functional neural network for a complex problem is a non-trivial task, so an extended discussion is necessary. Lastly, a future work plan is presented to show potential changes which could be made or tested on the ANN.

The information used for the training, validation, and subsequent classification in the ANN consist of attributes from stochastic AVA inversions over two reservoirs. The first inversion was from the synthetic SEAM field. The second was from field data in offshore West Africa, WAF. The inversion predicts four reservoir attributes, density, porosity, Vp/Vs, and P-impedance. Attributes are extracted from the output model locations based upon a

radial distance from each wellbore in the respective reservoirs. These wellbores often deviate slightly, and the direction is based on nearby geologic structures. The locations of the output models have a 25m spacing in the X and Y directions, while the Z is variable but averages about 2m. The Z thickness is variable to allow it to follow reservoir bedding patterns with more fluidity. However, between the changing Z cell thickness and the wellbore deviations, the number of locations within the defined radius various greatly between the wells.

The attribute locations around each well are adjusted to fit inside of a normalized grid to allow all wells to have the same number of model locations when used as an input to the ANN. This workflow is similar to what is done for MDS, but is slightly more complex. This move into the grid is accomplished in two steps. The first step is to straighten the wellbore inside of the normalized grid. The Z locations are broken down into 2m intervals. At each Z interval, the X and Y locations are moved to be centered at [0, 0]. Once centered, the desired locations can be extracted based upon the radius set. The second step in the normalized gridding process is to interpolate these locations into the defined grid. This interpolation is primarily set to be linear. The specific grids used are defined in the results section, as they can vary between experiments. However, the X and Y location spacing commonly has minimal interpolation, while the Z dimensionality can be reduced by up to nearly an order of magnitude. While summarizing the information into this grid is unideal, it is important that the same number of features for each well are input into the network.

Finally, one major adjustment made to the information for input to the network is feature normalization. All features are normalized to [-1, 1]. The necessity for this is outlined in the following section, but the process is as follows. The gridded attribute information is broken down into vectors, which each attribute concatenated onto the end to form an array of length nX*nY*nZ*nAtt. Next, the training information is decided, and each of those locations in the vector is normalized between [-1, 1], based upon the other like locations in the training set. The variance and mean used to normalize the training information are also used to normalize the testing and validation sets. The validation and test sets are not used

for the normalization, which is strictly based upon the respective training set. This workflow is more cumbersome compared to that of the other two algorithms, but is necessary for the network to operate properly.

A.3 Theory and Methods

The theory and methods behind a neural network can be broken down into the many network components. First, discussed is the input layer of the network where the information is formatted for forward propagation. The second section runs through the process of determining quantitiy and distribution of layers. The number of nodes and layers is directly related to the computation time and how complicated of a boundary the network can define. Each layer has a set activation function used in forward propagation. There are many activation functions available, but it often comes down to preference and input data formating. Next, the loss function is determined for the network, which quantifies network solution accuracy. The chosen loss function does have flexibility, but is more often determined by the problem being solved.

The final network piece is the optimizer, which uses the related loss from the forward propagation to adjust the biases on each node and reweight the interlayer connections. The optimizers are a bit complex and are often set based upon preference. For this study, we choose to use the popular Adam optimizer [35]. There is one additional section at the end discussing network cross-validation and the reinitializations done to obtain a classification variance. Both of these are done as a quality control check for the specific network.

A.3.1 Inputs

The scientist running the network determines the inputs to the neural network. However, there is some common terminology and practices that need to be defined. The first has to do with feature selection for input to the neural networks. One chooses whether alll the original values should be used, or just selected data attributes like mean and median values for a training set. The difference between these two types of selection demonstrates the boundary

between deep learning and traditional machine learning. For the work done on this project, the choice was made to attempt deep learning by using the original attribute values as the input features. Regardless of whether deep learning with the network is being done, parameter normalization is necessary to allow for more accurate activation of the nodes.

The normalization is done across all of the same input features to the network using the training information. This normalization formats all feature information to be in [-1, 1] based upon the mean and standard deviation of the other feature information. The validation information is normalized using the mean and standard deviation of the training information, as to not influence the training of the network.

The input information in this project consists of gridded attribute data from the wells in both reservoirs. The specifics to the input information are described in the Data Used subsection.

A.3.2 Number of Layers and Nodes

A typical deep learning ANN consists of an input, an output, and many hidden layers. The input layer is described in the previous subsection. The output layer is predetermined by the problem being solved. For example, if a binary classification is desired, then there could be two output nodes representing classification probabilities. If a single continuous output is desired, then there would only be a single node in the output layer. The more complex problem to solve is the number of hidden layers to use, and how many nodes for each hidden layer. There is a significant amount of variability in these for different problems. The mathematics behind how the ANN functions is a series of matrix operations between layers. These operations are done through node activations and weighting of the connections between each of the nodes for the layers. For example, if there are two hidden layers of four nodes each, there would be 4x4 or16 total connections. If in the same example there was a layer of six nodes connected to a layer of 2, then there would only be 2x6 or 12 total connections. An important consideration is that even with the same number of nodes in the network, a significant difference in the number network of weights can exist. The weight for

each connection is set randomly when initializing the network. How these layer interactions occur is based upon the activation functions.

A.3.3 Activation Functions

The ANN discussed in this section, uses four distinct activation functions with each one having its advantages and disadvantages. The first is illustrated in Figure A.1a, which shows a linear relationship between the inputs and outputs. A linear activation function is elementary and states that the total weighted inputs are equal to the output. While it is simple, it can take advantage of inputs of any magnitude across all real numbers. The second activation is the rectified linear unit function, or RELU, in Figure A.1b. RELU is unique and composed of two different functions based on whether the input is positive or negative. Positive values are scaled linearly by 1, while any negative values are set to 0. There is a second function called a leaky RELU, where negative values are given a small gradient for scaling. Next, shown in Figure A.1c is the hyperbolic tangent function. The hyperbolic tangent function is very similar to the sigmoid, but is more compressed on the input axis and output values between [-1, 1]. The hyperbolic tangent function is popular when there are multiple hidden layers because its outputs fit well for input into another hyperbolic tangent layer. Lastly, shown in Figure A.1d is the sigmoid function. The sigmoid can have inputs across real numbers as well, but anything past an absolute value of 6 does not provide much value. The output of the sigmoid function also only exists between 0 and 1.

A.3.4 Loss Function

The loss function choice is vital for running a proper neural network. There are many choices for the loss function, but it is often determined by the problem being solved. This study performs a binary or categorical classification. The loss function of choice is called categorical cross entropy and is given by Eq A.1.

$$H(p,q) = -\sum_{x} p(x) log(q(x)) \tag{A.1}$$

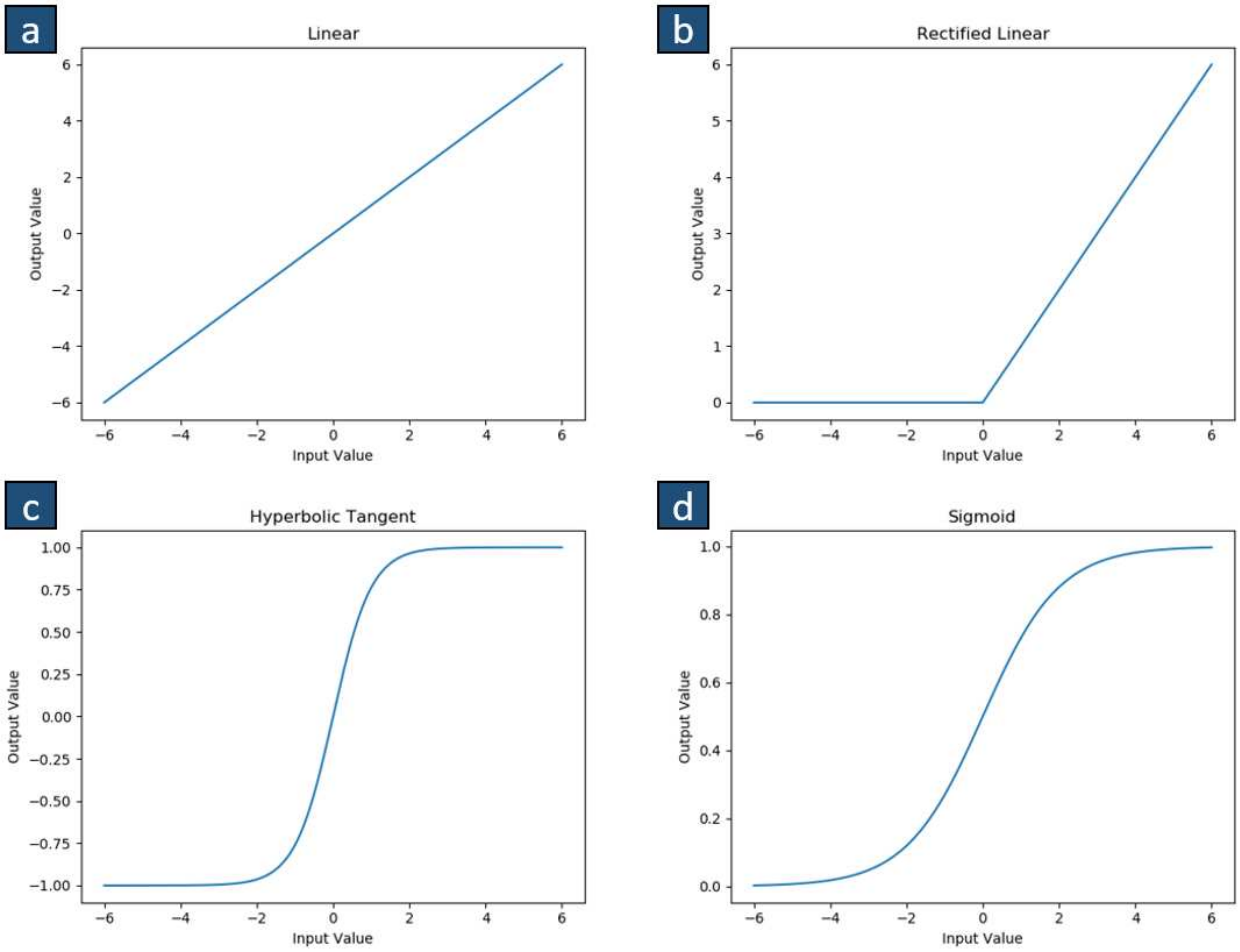

Figure A.1 Four activation functions that can be used for the neural network. (a) linear activation, (b) rectified linear function, (c) hyperbolic tangent, and (d) sigmoid. Each function can be used in any of the hidden layers in the network.

Where p is the true classification, which in our case is either a 0 or 1, and q is the calculated value from the forward propagation. The loss is then calculated across all of the training sets and summed to determine a total loss for the given epoch. The value calculated here is what is used for the optimization of the weights and biases for the next epoch.

A.3.5 Cross-Validation and Repeatability

Network cross-validation and repeatability are two important concepts to define to understand the confidence in our results for a given experiment. First, we define the network cross-validation. For a given setup a certain portion of the training information is separated

in the data pool to be used for validation of the network. In our case, we set aside a defined number of wells and all of their related iterations for validation. We test the network with these wells as the validation set and calculate and accuracy across the epochs. Next, a new set of wells is chosen as a validation set, and the network is rerun. While this testing is traditionally done randomly, we instead define every combination of wells and run through all possibilities exhaustively. For example, if there are 11 total wells and we decide to omit 4 of them for the cross-validation the total number of experiments is 330. However, because the network does random initialization of the weights for layers in the network, we also repeat these combinations up to 30 times to calculate variance in the classification accuracy [36].

A.4 Results and Discussion

The artificial neural network has been through many trial and error phases by testing many different setups and combinations. However, this section discussed just to two networks of interest. The first, known as the baseline network, consists of two layers of 50 and 15 nodes. The second, know as the flatter network, designed to contain the same number of nodes to keep the bias count equal, consists of more layers with nodes counts of 15, 15, 15, and 20. While the number of total biases remains constant, the flatter network does have more connections to weight during the optimization and backpropagation. Lastly, the activation function for the baseline network was a hyperbolic tangent for both of the hidden layers. The flat network was tested using both the sigmoid function and tanh.

A.4.1 Baseline Network

Figure A.2 shows the first network tested for this analysis, which was trained across 15 epochs and omits four wells in the cross-validation. The green lines indicate the average validation accuracies at each epoch for all the combination tested in the cross-validation. The red lines show the training accuracy and loss. The results for this network are interpreted as follows. The training accuracy is shown in Figure A.2a converges extremely fast, often after no more than tow epochs. The same pattern can be seen in Figure A.2c where the

loss function for the training set converges immediately as well. The combinations used for validation had a wide spread of accuracy and loss. The spread is highlighted in Figure A.2b and Figure A.2d where the standard deviation extends more than 10% in the validation accuracy. Lastly, the mean loss function in Figure A.2d is increasing with the epoch count indicating the network is overconstraining to the training information too quickly.

Figure A.3 shows the testing results of the baseline network the number of iterations used increased from 35 to 4999. The goal of this experiment was to see how using all of the information available would affect the classifications. Many lines in the validation accuracy are nearly flat, showing the additional epochs are ineffective. The training accuracy and loss are converged to a minima/maxima within the first epoch. The network has trained very quickly, but the mean accuracy for the validation sets is roughly equal to the first network. Lastly, the mean loss in the validation for this network is nearly four times larger than the baseline network.

A.4.2 Flatter Network

The next set of networks tested were formatted to be flatter, meaning an increased number of layers, but with the same number of total nodes. Figure A.4 shows the results from one of these flat networks using a hyperbolic tangent activation function for all hidden layers and uses 35 inversion iterations for training and validation. The results from this network can be described nearly identically to those seen in Figure A.2. There is very little difference in the results between the two networks.

Figure A.5 displays the last network with identical structure and iteration count to Figure A.4, but uses a sigmoid function for activation rather than hyperbolic tangent. The results of this network are unique compared to the other three networks. The training accuracy does not immediately converge to 1, and instead takes nearly 15 epochs to come close to convergence in some of the combinations. The training loss also decreases much slower, but the validation loss is also not strictly increasing with more epochs. These indicate the network is training slow and not able to immediately find a local minimum for the training

information.

The four networks displayed each has a distinct character, besides the two using a hyperbolic tangent activation and 35 inversion iterations. The goal for the neural network is to have higher accuracy than what we see in the NBC from Chapter 2, which we do have consistently. However, the loss functions for three of the networks increase with each epoch, which is suboptimal. We want to see the loss decrease for some epochs and then start increasing. This training is much too fast and is indicative of getting stuck in a local minimum. The network with the sigmoid function appears to be an extremely weak solution to the problem. While using a sigmoid is a reasonable choice in many cases, feeding a sigmoid into itself does not seem to make much sense. The sigmoid can take in values from our input which are normalized between -1 and 1. However, it outputs values between 0 and 1 so forward feeding this into another sigmoid does not make sense. Proving this, is the training and validation accuracies for this network greatly vary. The loss function appears to be much more reasonable as the loss does not strictly continue to increase with more epochs. The cause of this is most likely the network unable to find a reliable solution as the feeding of sigmoids into each other is not ideal.

A.5 Conclusion and Future Work

Overall, the neural network does have more accurate results than the NBC discussed in Chapter 2. The increase in accuracy does not come as a surprise as the networks can evaluate more complex information than a NBC. The ANN also does not have the independence assumption among the different features. The only summarization of the information for the network was the gridding process to get all the information into a similar space. The ANN does have room to improve, and we believe the maximum obtainable accuracy in the cross-validation is much higher than what we have observed so far. For the generation of better results, there needs to be additional work completed on refining the network and doing more dynamic testing of the training and validation information. Overall, these initial results are promising, and many options are available to continue improving on these results.

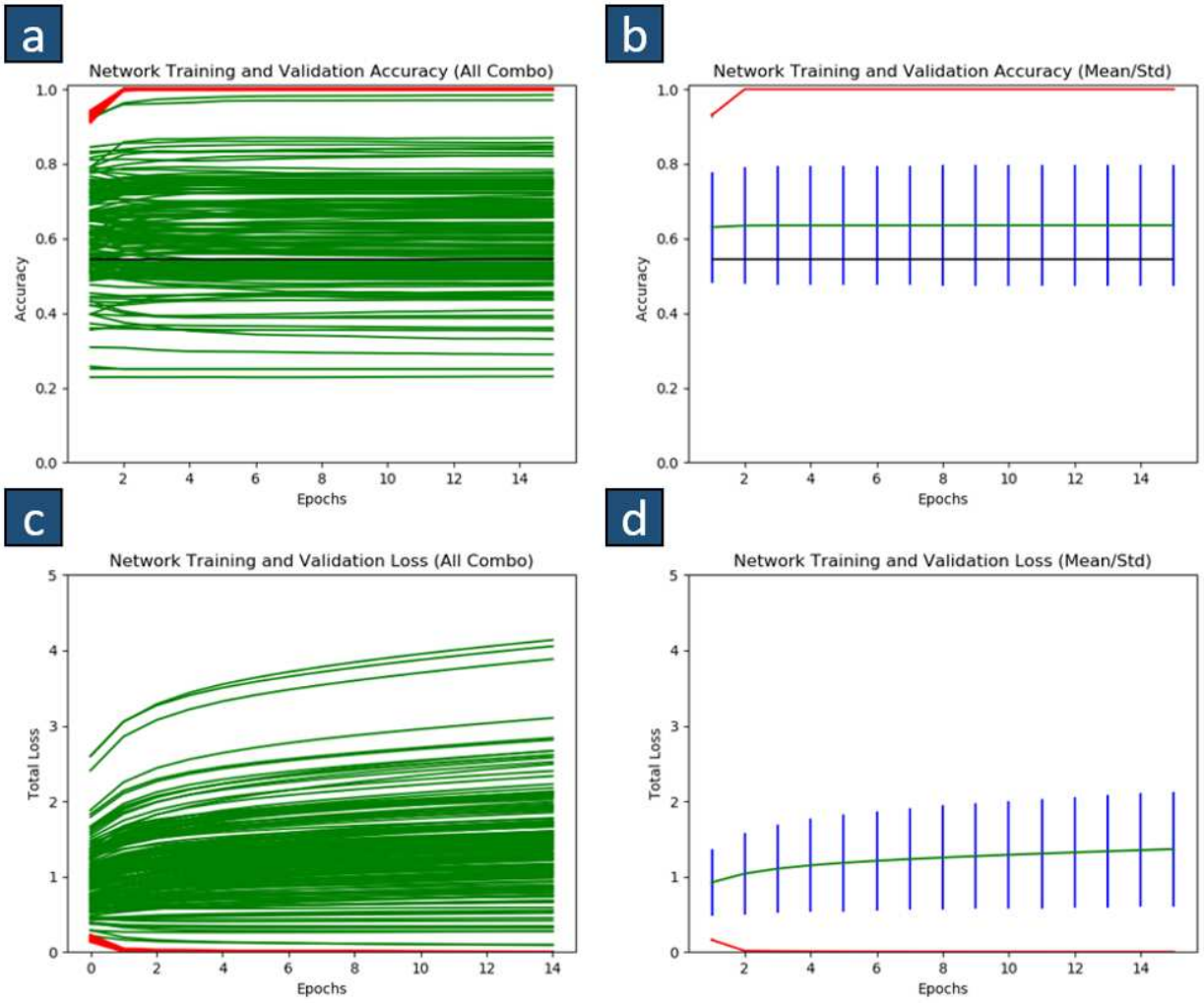

Figure A.2 Information related to the network loss and accuracy. The network displayed consists of two hidden layers, with 50 and 15 nodes respectively. Both of the hidden layers used the Tanh activation function. This network also used 35 of the iterations from the inversion for each well's training or validation set. The green lines in (a) and (c) represent the validation accuracy for the different combinations in the cross-validation. The red lines are the training accuracy and loss; however, they are often close to the same value and hard to delineate. (b) and (d) show the same information but in a simpler form. The green line indicates the mean of all the combinations at each epoch, with the blue line extending to one standard deviation on each side of the mean. The red line shows the training accuracy and loss, but are often so similar that it is hard to see the plotted standard deviation.

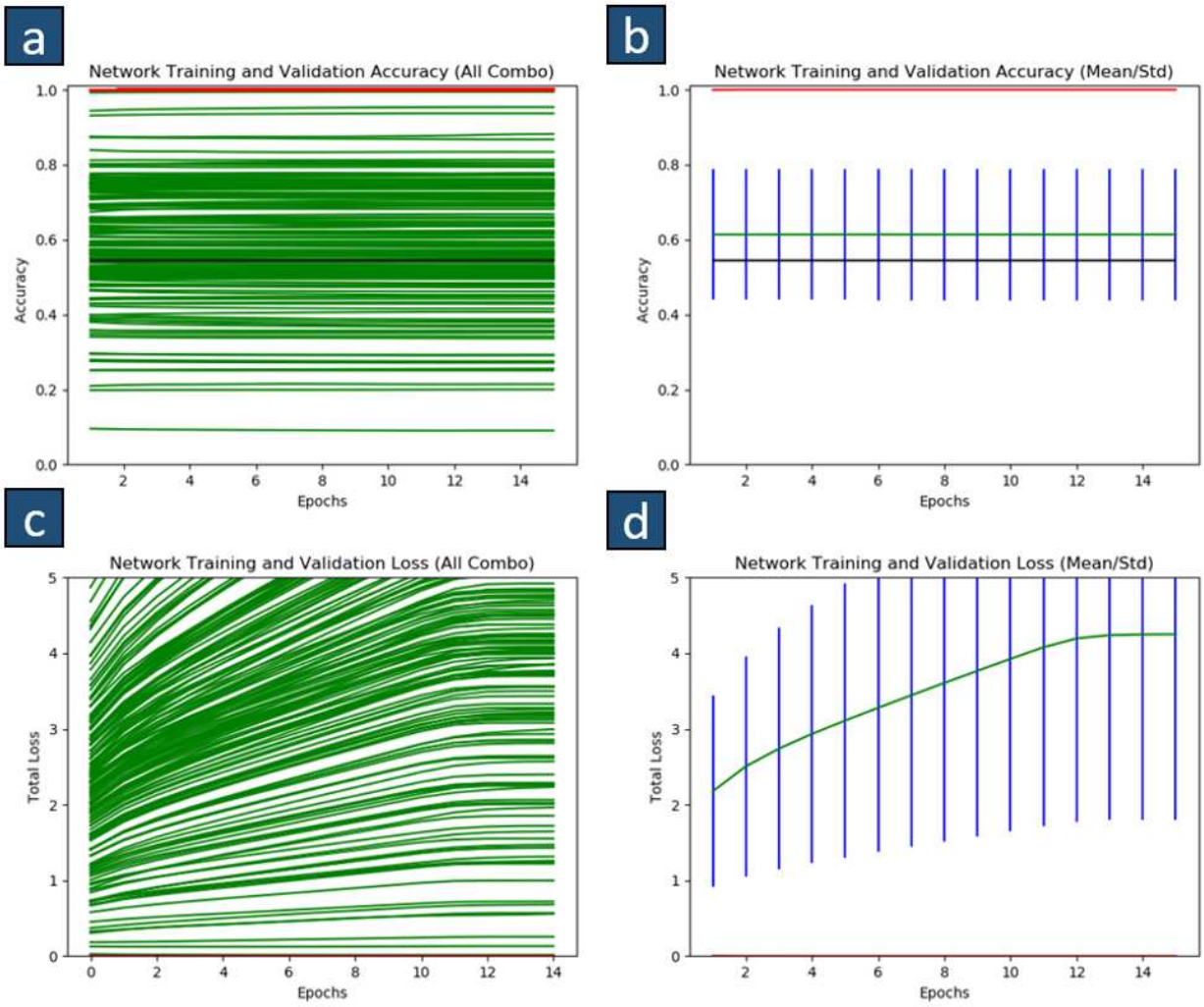

Figure A.3 Four subplots with the same structure as Figure A.2. The network is the exact same structure, only now all 4999 samples from the inversion are used for training and validation.

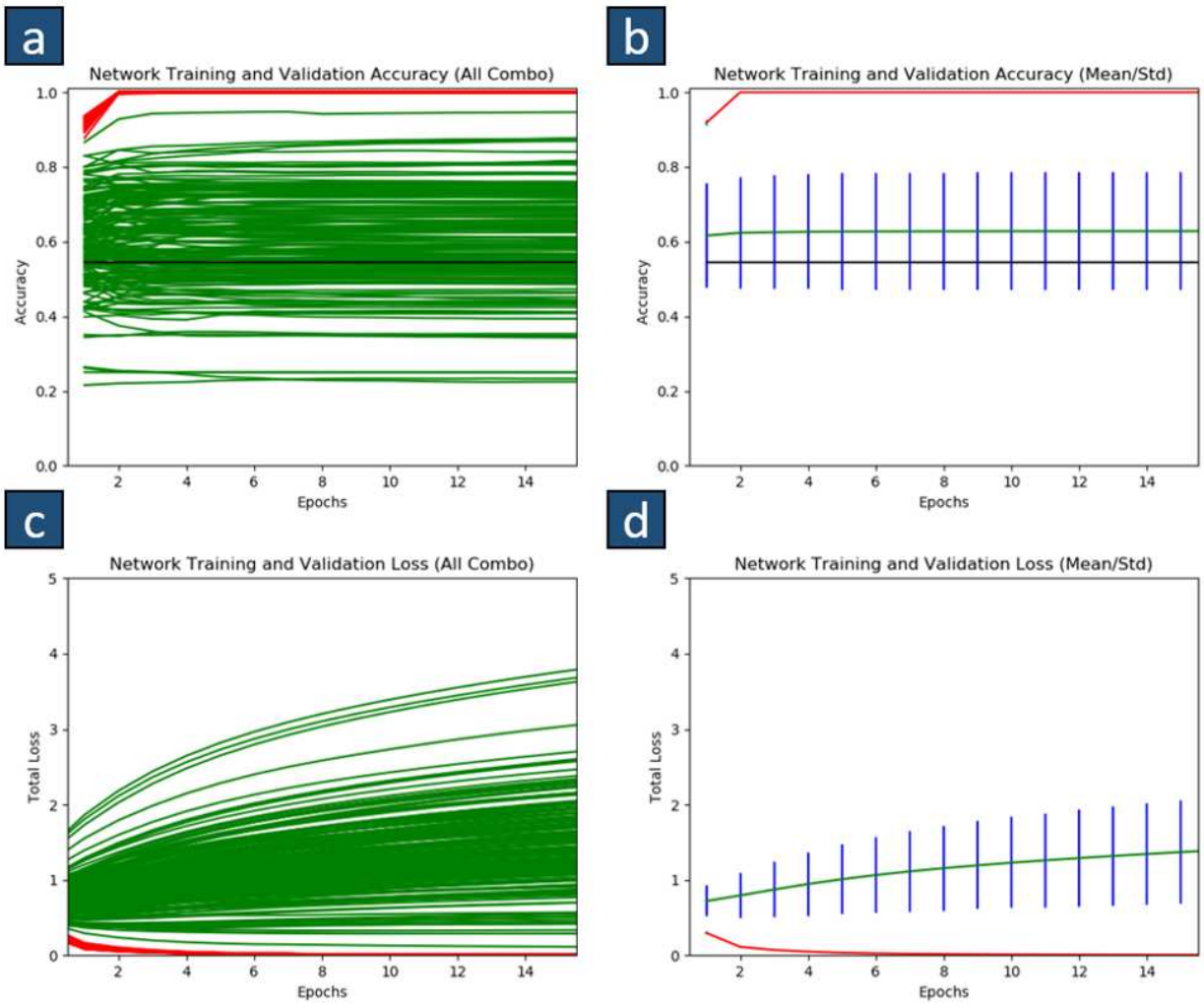

Figure A.4 Four subplots with the same structure as Figure A.2. The network structure is the same as Figure A.5, but the activation function is now set to be hyperbolic tangent.

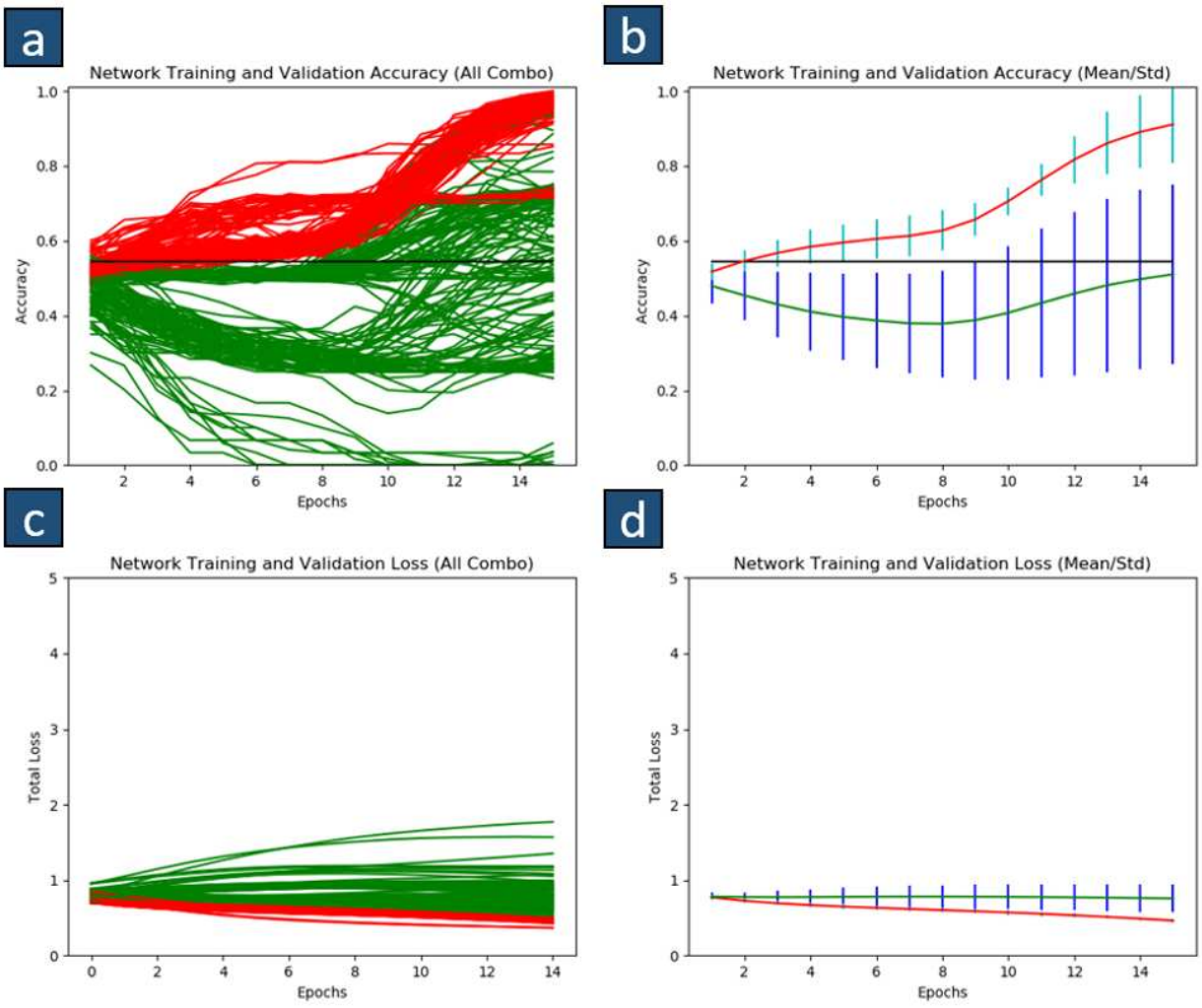

Figure A.5 Four subplots with the same structure as Figure A.2. The network structure is changed to consist of four hidden layer of 20, 15, 15, and 15 nodes respectively. The activation function is changed to be a sigmoid function, but the number of iterations from the inversion used is kept constant at 35.