Application of Artificial Neural Networks to geological classification: porphyry prospectivity in British Columbia and oil reservoir properties in Iran

A Thesis Submitted to the
College of Graduate and Postdoctoral Studies
In Partial Fulfillment of the Requirements
For the Degree of Master of Science
In the Department of Geological Sciences
University of Saskatchewan
Saskatoon

By
Moslem Azarpour

PERMISSION TO USE

In presenting this thesis in partial fulfillment of the requirements for a Postgraduate degree from the University of Saskatchewan, I agree that the Libraries of this University may make it freely available for inspection. I further agree that permission for copying of this thesis in any manner, in whole or in part, for scholarly purposes may be granted by the professor or professors who supervised my thesis/dissertation work or, in their absence, by the Head of the Department or the Dean of the College in which my thesis work was done. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Saskatchewan in any scholarly use which may be made of any material in this thesis.

Requests for permission to copy or to make other uses of materials in this thesis in whole or part should be addressed to:

Head of the Department of Geological Sciences
Geology Building
114 Science Place
University of Saskatchewan
Saskatoon, Saskatchewan S7N 5E2
Canada

OR

Dean

College of Graduate and Postdoctoral Studies

University of Saskatchewan

116 Thorvaldson Building, 110 Science Place

Saskatoon, Saskatchewan S7N 5C9, Canada

ABSTRACT

Seismic facies analysis aims to classify oil and gas reservoirs into geologically and petrophysically meaningful rock groups, or classes. An artificial neural network (ANN) is a versatile and efficient tool for classifying data or estimating subsurface properties from large geophysical datasets. This tool can provide critical information for oilfield development and reservoir characterization.

This study includes application of artificial neural networks on two different datasets: 1) geophysical characterization of an oil reservoir in Iran and 2) geological prospectivity for porphyry in British Columbia, Canada. In the first case study, I utilize seismic attributes, well-log data, and core data analysis and use supervised machine learning techniques to efficiently estimate the acoustic impedance and porosity of the reservoir and to classify it into four lithological classes. Seismic attributes as inputs for our techniques capture the lithological patterns or structural characteristics in the seismic amplitude, phase, frequency, and other complex seismic properties that cannot be directly seen in the original seismic images. Selection of an optimal set of input features from the vast number of possible mathematical transformations of seismic data is a critical task for reservoir property prediction and classification. This selection is performed by standard as well as innovative procedures employing properties of the target classes.

Three different supervised approaches to non-linear classification are used: 1) the so-called probabilistic neural network (PNN), 2) conventional ANN, and 3) an ANN with the new approach of optimal attribute selection. For each of these approaches, images of classification confidence levels and confidence-filtered class images are produced. Assessments of the robustness and accuracy of seismic facies classification is performed for each of these algorithms. The ANN classifiers are validated using validation and test data subsets. The proposed algorithm shows a higher performance, particularly in comparison with the PNN algorithm. Several visualization techniques are used to examine and illustrate the power of the ANN-based approaches to classify the seismic facies with high accuracy. However, the three approaches still provide significantly different levels of lateral continuity, frequency content, and classification accuracy. Therefore, some level of expert assessment is still required when using machine learning for reservoir interpretation.

In the second case study, I use an ANN to explore the prospectivity for porphyry within the Quesnel Terrane, BC, Canada. A purely data-driven approach based on geophysical, structural, and volcanic-age data results in a predictive prospectivity map which correlates well with known mineral occurrences and suggests new areas for potential exploration.

ACKNOWLEDGMENTS

I would like to express my sincere appreciation to Prof. Igor Morozov for his guidance, motivation, financial support and constant help throughout this research. I would especially like to thank him for his patience during this research.

My thanks also go to my graduate advisory committee members, Prof. Samuel Butler, the Head of the Department of Geological Sciences, and Prof. Jim Merriam.

A special thanks go to Chantal Strachan-Crossman, graduate program secretary, for her support and help.

I would also like to thank Song Hu, ICT support, for his support with computing services and updating the software licences.

I express my gratitude to all my officemates within my study at the University of Saskatchewan for their support and standing by for their assistance. Thanks for the friendship, encouragement and moral support!

DEDICATION

This dissertation is dedicated to the individuals who made me the person I am today, supported me, and believed in me, filled me with inspiration and motivation, my parents, my wife and my sisters.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SYMBOLS AND ABBREVIATIONS[1]

3D, 3-D	Three dimensions, three-dimensional
AI	Acoustic impedance
ANN	Artificial Neural Network
EF	Electrofacies
GRNN	Generalized Regression Neural Network
MLFN	Multi-layer feed-forward ANN
NPHI	Neutron Porosity
PNN	Probabilistic Neural Network
RBFN	Radial Basis Function Network
TMI	Total Field Magnetic Intensity

[1] Only abbreviations most commonly used throughout this thesis are listed.

CHAPTER 1: Introduction

In order to conduct geological or geophysical interpretation, a combination of multiple datasets such as geological maps, well logs, measurements of chemical and physical properties of core samples and in thin sections, and estimations of rock density, porosity, or permeability are used. Broadly, such interpretation tasks reduce to classification of the input data into certain classes, such as certain rock types, productive or non-productive zones within an oil or gas reservoir, or geographical areas of various levels of mineral prospectivity. With large volumes of the datasets and broad variety of input parameters, the classification procedure requires significant expertise and effort. In the recent years, machine learning techniques using computer algorithms called the artificial neural networks (ANN) have started getting to the forefront of many of these interpretation methods, particularly for automated interpretation of geophysical maps and making reservoir property estimations.

The need for ANN-aided approaches is particularly important in seismic data interpretation and reservoir characterization. By adding seismic data to geological parameters, an analyst has to deal with a large amount of data, which is time-consuming and often even impossible to do. A typical seismic reflection trace contains several thousand of time samples, with millions of traces in a three-dimensional (3-D) dataset, and numerous geophysical features (seismic attributes) need to be taken into account when making a parameter estimation or classification at each sample. However, with the help of machine learning, this process of detailed reservoir characterization can be aided and possibly even automated in the future.

The general approach to geological or geophysical data classification used in this thesis is as follows. Using seismic processing and interpretation, an expert geophysicist extracts the significant geological or geophysical features from the data and performs classifications on a set of ground truth and carefully studied examples (maps or well logs, called the training dataset). Further, a data scientist uses machine-learning techniques to learn the rules used by the expert, and then uses these rules to the valuable parameters or classify the complete dataset. In this thesis, I try the roles of both the geophysics expert and data scientist in applications to two case studies:

1) Investigate reservoir properties in an Iranian oilfield (Chapters 3, 4 and 5);

2) Exploration of prospectivity for mineral deposits in Quesnel area, BC, Canada (Chapter 6).

In the first of these applications, I also use utilize the inversion and machine learning methods to derive volumes of acoustic impedance and to estimate porosity within the reservoir (Chapter 4).

1.1 Objectives and contributions of this study

The general objective of this study consists in obtaining new interpretations of geological and geophysical datasets using machine learning techniques utilizing seismic attributes, well-log data, core data and thin-section analysis. The specific and practical objectives for the two datasets, and the corresponding contributions are:

1) For the reflection seismic dataset:
   a) Perform complete and accurate model-based inversion of acoustic impedances (AI). This procedure results in a new 3-D volume of AI, which is the seismic attribute most useful for seismic interpretation and correlation of seismic results with geology. The AI volume is also used further for estimating porosity and for obtaining final classification.
   b) Use machine learning to estimate porosity volume from the estimated AI and other seismic attributes. Similar to AI, prediction of porosity at any point within the subsurface is the key part of reservoir analysis and classification.
   c) Perform a feasibility study and quality control to investigate the relation between porosity and AI. The contributions from this objective consist in measuring the correlation between these two parameters and testing several machine learning techniques and making recommendations for the best algorithm porosity estimation.
   d) Apply and compare three machine learning techniques for producing classifications Perform an innovative feature selection approach and use neural network to obtain the highest possible validation accuracy. The proposed approach shows a significant improvement in the classification of seismic facies compared to previously used neural network and also current machine learning techniques used in a specific industrial software.

2) Train and apply an artificial neural network to model prospectivity of porphyry related deposits in Quesnel Terrane (BC, Canada) using a pattern-recognition algorithm and structural, geological and geophysical data.

In meeting the practical objectives for the seismic (principal) part of the thesis, significant methodological questions need to be addressed. There exist numerous seismic attributes, but many of them are mutually related and cannot be used for final classification of the reservoir. Selecting the best set of seismic attributes is the primary step of feature extraction, and it has a significant impact on the final parameter estimation or seismic facies classification. Selection of an optimal set of independent attributes should yields valuable information that could lead to a better understanding of the studied area and help detecting detailed features of the subsurface geology.

How to select the best method and the best structure of the ANN to predict and estimate reservoir properties from seismic data? What features should be selected, and which algorithms should be employed to address this problem? These are the main challenges for someone dealing with reservoir characterization from seismic data. In this thesis, I discuss these challenges and propose an approach to their solution using an innovative attribute selection approach. This is the key methodological contribution from this study (chapter 2). Another important new methodology of this thesis consists in the estimation of confidence levels of the classifications and in using these levels for plotting the classification sections for final interpretation (chapter 5).

1.2 Structure of this thesis

In the current Chapter 1, I briefly introduce the general approach of machine learning and summarize the goals of this study. In chapter 2, I describe the key methodologies which are common to most applications in this study. Also in chapter 2, I describe the new methodology of optimal selection of features for ANN classification developed in this study.

Chapters 3, 4, and 5 are devoted to the reservoir classification project. In chapter 3, I describe the geology of the study area, seismic data and processing, and the initial interpretation of the studied reservoir. In chapter 4, I describe the post-stack seismic acoustic-impedance inversion method and porosity estimation within the 3-D seismic volume using the Emerge method by Hampson and Russell. In chapter 5, I develop a rock-type classification model by using the

seismic data, borehole core data, and wireline logs. The model consists of four reservoir electrofacies, which represent group or clusters of geophysical measurements.

In chapter 6, I describe a different, geological classification study to classifying potential porphyry deposits in Quesnel area in BC, Canada. The chapter contains a brief overview of the data, methods of prospectivity prediction, and an application of ANN to several geophysics' fields and also structural and geological data.

Finally, in chapter 7, I present the conclusions from this research, discuss the applications of multiple machine learning techniques, and also make recommendations for further development of the methods of this thesis and for solving the outstanding problems.

CHAPTER 2: Methodology

This chapter summarizes the general methodology used in other chapters of this thesis. Inversion and porosity estimation are mostly performed using Hampson-Russell software. Some seismic attributes are extracted using HRS (Hampson-Russell) and some of them using interactive, open-source OpendTect software by dGB Earth Sciences. Machine learning techniques are performed mostly using MATLAB and Python.

Acoustic impedance estimation for entire seismic cube consists of multiple steps. I used the post-stack seismic inversion using HRS software. There are several inversion methods including model-based, bandlimited, sparse spike and colored inversion. In this thesis, I tested all methods and the best method in our case is model-based inversion which gives the highest correlation (Chapter 4). The most critical step in seismic inversion is the quality control which a geophysicist could use different visualization and numerical validation to confirm or reject a final output of the inversion process.

After confirming the estimated acoustic impedance using several quality control steps, I aimed to estimate porosity volume using seismic attributes, especially the estimated acoustic impedance in the previous step, and well-log data. I used Emerge tool in HRS which includes selecting the best subset of attributes from a range of several attributes using stepwise regression and then utilizing multi-attribute regression. After selecting the best subset of data using cross-validation plots. I used several supervised learning techniques such as the probabilistic neural network (PNN), multi-layer feedforward network (MLFN) and radial basis function neural network (RBFN) and after performing a careful comparison, the best technique with least validation error was selected (Chapter 4). These algorithms are artificial neural networks. Artificial neural networks (ANN) are multilayer perceptron (MLP) supervised neural networks, which are mathematical mechanisms designed to execute complicated pattern recognition tasks. They are utilized to classify the seismic facies into multiple classes due to their power to eradicate the need for intricate statistical techniques, noise control, and complex and extensive database management.

Then, the most meaningful seismic attributes were extracted from HRS and OpendTect to feed into different neural networks in order to classify our lithofacies from the well-log analysis

and core data interpretation. This step was done using MATLAB and Python and the results are compared to the PNN algorithm using HRS. In this study, I used an innovative approach to select the optimal number of attributes to avoid overfitting and obtain the highest validation score and the results are compared to PNN and regular ANN in chapter 5.

After standard or improved initial preconditioning of the data (section 2.1), this methodology can be subdivided into three broad tasks:

1) **Inversion**: Using continuously distributed data such as reflection seismic records, produce new continuous data such as the acoustic impedance or porosity.
2) **Prediction**: Using known occurrences of a certain class, derive a probabilistic model for predicting this class from values of a set of features measured from the data (seismic attributes).
3) **Classification**: Using a known classification of a part of the data into several discrete types, extend this classification to the complete dataset.

The Inversion task is considered in application to inversions for acoustic impedance (AI; section 2.2) and porosity (section 2.3). The AI inversion is based on knowledge of the specific physical relation between the wave reflectivity and acoustic impedance, and therefore the algorithm is also based on inverting a rigorous forward problem. By contrast, the prediction of porosity (section 2.3) utilizes no specific model for its dependence on seismic attributes. The prediction is obtained by training an artificial neural network (ANN) to reproduce the ground-truth observations in the available wells.

In section 2.4, I describe several approaches to the Prediction task above. This task is accomplished in the thesis by using a kernel-based interpolation method (chapter 4) or an ANN (chapters 5, and 6).

Section 2.5 describes the principal idea and several application methods for the Classification task above. With any type of classification, it is useful to additionally precondition the input data by selecting the combinations of features which are most sensitive to the expected classes. As shown in chapter 5, this selection significantly improves the classification results while reducing the computational effort. The approach to selecting an optimal set of features from the seismic dataset is described in section 2.6.

2.1 Data preconditioning

Before performing a statistical or machine-learning analysis, both the input and output data need to be preconditioned. Let us represent the feature vectors **x** (column matrices in Matlab) by transposed rows of data matrix **X**. For example, the dataset in chapter 5 consists of 692 vectors $\mathbf{x}_i$ in a 25-dimensional attribute space, with $i = 1, \ldots, 692$, and therefore **X** is a 692×25 matrix. The most common preconditioning method consists in applying the Z-score transform to the rows of this matrix by relation $z = (x - \mu)/\sigma$, where μ is the mean and σ is the standard deviation (Haykin, 2009). The j^{th} component (attribute) of i^{th} vector equals $(\mathbf{x}_i)_j = X_{ij}$. For data with zero mean $\langle \mathbf{x} \rangle = 0$, the covariance of the j^{th} and k^{th} attributes equal $(\mathbf{C}_x)_{ik} \equiv (\operatorname{cov}\mathbf{X})_{ik} = X_{ij}X_{ik}$ where summation over repeated indices (i) is implied. In matrix notation, this expression is $\mathbf{C}_x = \mathbf{X}^T\mathbf{X}$, where '^{T}' denotes the matrix transpose.

My first goal in data preconditioning is to transform the data attributes (features) **X** into new attributes **X′** which are uncorrelated and possess unit variance. This transformation can be represented by matrix multiplication of each row of matrix **X** with some matrix **M** as

$$\mathbf{X}' = \mathbf{X}\mathbf{M}. \tag{2.1}$$

The multiplication needs to be on the right side of **X**, so that each row is multiplied by **M** independently. Note that if dimensionality reduction is desired, matrix **M** may have fewer columns than rows (also the number of columns in **X**). After the transformation, the covariance matrix equals:

$$\mathbf{C}'_x = \mathbf{X}'^T\mathbf{X}' = (\mathbf{X}\mathbf{M})^T\,\mathbf{X}\mathbf{M} = \mathbf{M}^T\mathbf{C}_x\mathbf{M}. \tag{2.2}$$

To determine matrix **M**, let us first determine the normalized eigenvectors $\mathbf{e}_n$ and the corresponding eigenvalues σ_n^2 of the data covariance matrix $\mathbf{C}_x$ in Equation (2.23). If some of the values of σ_n^2 are considered small and insignificant for subsequent analysis, these eigenvectors can be excluded. Let us denote N the number of significant eigenvectors $\mathbf{e}_n$. Using its eigenvectors, the square, symmetric, and positive-definite matrix $\mathbf{C}_x$ can be represented by the sum as

$$\mathbf{C}_x = \sum_{n=1}^{N} \sigma_n^2 \mathbf{e}_n \mathbf{e}_n^T . \tag{2.3}$$

Using the eigenvectors of matrix $\mathbf{C}_x$, let us define the n^{th} new attribute (feature) as projection of the data onto the eigenvector $\mathbf{e_n}$ divided by the corresponding σ_n: $x'_n = \frac{\boldsymbol{x}^T \boldsymbol{e}_n}{\sigma_n}$. This relation means that the transformation matrix **M** in Equation (2.1) consists of *N* columns, with n^{th} column being the scaled eigenvector $\frac{\boldsymbol{e}_n}{\sigma_n}$.

Using equations (2.2) and (2.3), the covariance between the k^{th} and l^{th} transformed attributes equals

$$\left(\mathbf{C}_{x'}\right)_{kl} = \mathbf{M}^T \mathbf{C}_x \mathbf{M} = \sum_{n=1}^{N} \frac{\mathbf{e}_k^T \mathbf{e}_n}{\sigma_k} \sigma_n^2 \frac{\mathbf{e}_n^T \mathbf{e}_l}{\sigma_l} = \delta_{kl} \tag{2.4}$$

(Kronecker delta, equal one when $k = l$ and zero otherwise), which means that $\mathbf{C}_{x'} = \mathbf{I}$ as expected. Therefore, after this transformation, the new set of attributes are zero-mean, normalized, and mutually uncorrelated.

2.2 Model-based estimation of acoustic impedance

The acoustic impedance (AI) is the most important attribute derived from reflection seismic records and used in many types of geological and geophysical interpretation of reflection seismic data. At normal incidence, the AI is usually denoted by *Z* and given by equation

$$Z = \rho V , \tag{2.5}$$

where ρ is the mass density and *V* is the P-wave velocity. Because the impedance is proportional to both ρ and *V*, it reduces with increasing porosity (decrease of ρ) and increases with stiffness of the rock. For example, for P waves, the velocity depends on the elastic moduli λ and μ as $V = \sqrt{(\lambda + 2\mu)/\rho}$, and therefore the AI increases with both moduli and density: $Z = \sqrt{\rho(\lambda + 2\mu)}$ (Lindseth, 1979).

Because of such key importance of the AI, numerous studies analyse its properties and invert it from seismic data. Lindseth (1979) and Oldenburg et al (1983) gave the first inversions

of AI in a layered medium from stacked seismic data. These authors pointed out the fundamental uncertainty of this inverse (described below) and suggested practical approaches to its correction by interpretive analysis. Latimer et al. (2000) reviewed many practices of using AI in seismic interpretations. Connoly (1999) proposed an extension of the AI concept to offset (pre-stack) seismic data, which is called the elastic impedance (EI). Although the EI is not as rigorous physical property as the normal-incidence AI, Connoly (1999) and Whitcombe et al. (2002) showed that the offset dependencies of this quantity contain important indicators of fluid content within reservoirs. Multiple forms of AI and EI inversion approaches exist, most notably the 'colored' method by Lancaster and Whitcombe (2000), log-calibration (SILC) inversion by Morozov and Ma (2009), and the ray-path EI inversion by Santos and Tygel (2004). Mallick (1995) started a broad group of AI inversions using stochastic models and algorithms. In addition, a broad new area of research focuses on AI inversions for non-stationary (time- and depth-variant) source waveforms in the presence of seismic attenuation (e.g., Zou et al., 2022).

All of the above applications of the AI or EI consider the same general problem, which I explain below on the example of the normal-incidence post-stack AI. In a layered subsurface, the waves travel in the vertical direction, and the recorded P-wave reflection amplitude from the i^{th} layer boundary is related to the impedance ratio across the boundary as

$$r_i = \frac{Z_i / Z_{i-1} - 1}{Z_i / Z_{i-1} + 1}, \tag{2.6}$$

where Z_i and $Z_{i\text{-}1}$ are the impedances below and above the boundary, respectively. Using eq. (2.6), the impedance time series can be obtained starting from $Z = Z_0$ at the top of the model ($i = 0$) and proceeding downward:

$$Z_i = Z_{i-1} \frac{1 + r_i}{1 - r_i}. \tag{2.7}$$

This inversion formula is known as the recursive AI inverse (Lindseth, 1979). Representing the time series Z_i and r_i by vectors $\mathbf{Z}$ and $\mathbf{r}$, respectively, this formula can be written as $\mathbf{Z} = \mathbf{Z}_B + \mathbf{H}\mathbf{r}$, where $\mathbf{H}$ is a matrix implementing recursive eq. (2.7), and Z_B is the background impedance model.

The reflectivities r_i are related to the data values u_i in stacked seismic records by the source wavelet w_i which also contains a scaling factor arising from seismic processing:

$$\left(w*r\right)_i = u_i, \tag{2.8}$$

where '*' denotes the convolution operation for time series. In matrix form, this relation also becomes matrix multiplication:

$$\mathbf{Wr} = \mathbf{u}. \tag{2.9}$$

The wavelet matrix **W** is difficult to determine accurately but it is constant or only smoothly variable for different layer numbers (time samples in a seismic trace) *i*. Thus, the inversion for AI consists in solving equations and (2.6) and (2.7) for Z_i at each sample *i*. The outcome of this inversion is the physical property $Z = \rho V$ of each layer, which should be directly related to the rock type, porosity, fluid content, etc. By contrast, the measured reflection amplitudes u_i only refer to boundaries between different layers, and they contain an arbitrary scaling and effects of wavelet shape.

Inversion of eqs. (2.6) and (2.7) for Z_i is nonunique and often represents a complex procedure. The difficulties of this procedure are: 1) in the need to simultaneously estimate the wavelet w_i, 2) in the invariance of these equations with respect to arbitrary scaling of impedance $Z_i \to cZ_i$, 3) in the lack of low-frequency signal in seismic data u_i, which causes an uncertainty in the background model Z_B, and 4) in the effects of noise in the data. The many AI and EI inversion methods differ in the ways of treating these issues. Morozov and Ma (2009) argued that since the AI inversion problem is inherently under-determined (Menke, 1984; Morozov, 2021), the inverse typically fits the reflectivity data nearly accurately. Therefore, the quality of the result should be judged not by the data fit but by satisfying the criteria of geological interpretation.

To obtain geologically meaningful results also broadly accepted in the industry, I use the constrained least-squares inversion method included in the Hampson-Russell software STRATA. In this method, the wavelet w_i and the background impedance time series $\mathbf{Z}_B$ are estimated first, and the layered model beneath a given point is parameterized by reflectivity values r_i for all *i*. These time series are represented by matrix **W** (wavelet) and vectors $\mathbf{Z}_\mathrm{B}$ and **r**, respectively. Then, the reflectivity **r** is obtained by minimizing the cost function consisting of two terms (Menke, 1984):

$$J\left(\mathbf{r}\right) = w_1\left(\mathbf{L}-\mathbf{Hr}\right)^T\left(\mathbf{L}-\mathbf{Hr}\right) + \left(1-w_1\right)\left(\mathbf{u}-\mathbf{Wr}\right)^T\left(\mathbf{u}-\mathbf{Wr}\right), \tag{2.10}$$

where **u** is the seismic time series, **L** is some prior (initial, low-frequency) model for the impedance, and w_1 is a weight factor between 0 and 1. The low-frequency model **L** cannot be obtained from seismic data, and it has to be constructed based on geological and/or well-log considerations (Lindseth, 1979). The first term in this cost function penalizes deviations of predicted seismic-frequency impedance (**Hr**) from the prior model, and the second term penalizes deviations of the predicted reflection record (**Wr**) from the observed data **u**. If we set $w_1 = 0$, it means that the prior model is not used. If we set $w_1 = 1$, the inversion will attempt fitting only the prior model while ignoring the data.

The solution **r** minimizing the cost function $J(\mathbf{r})$ (eq. (2.10)) can be obtained from the requirement that at the optimum point, the gradient of $J(\mathbf{r})$ equals zero: $\partial J / \partial \mathbf{r}^T = \mathbf{0}$ (Menke, 1984). This equation gives

$$\mathbf{r} = \left[w_1 \mathbf{H}^T \mathbf{H} + (1 - w_1) \mathbf{W}^T \mathbf{W} \right]^{-1} \left[w_1 \mathbf{H}^T \mathbf{L} + (1 - w_1) \mathbf{W}^T \mathbf{u} \right]. \tag{2.11}$$

Using this **r**, the final AI model is obtained as $\mathbf{Z} = \mathbf{L} + \mathbf{Hr}$ (Russell and Hampson, 2006).

Thus, the solution for AI depends not only on the data **u** but also on the low-frequency model **L**, wavelet **W**, and weight w_1 with which these effects are intermixed. In addition, since seismic traces (**u**) always contain noise, the resulting model **r** is also affected by this noise. During the inversion analysis and testing, I compared different models for **L** and **W** and values of w_1, and selected the best model based on the accuracy of the AI estimated near the wells. This accuracy was measured by the normalized correlation coefficient (R) calculated between the estimated and original acoustic impedances at the well locations (Haykin, 2009):

$$R = \frac{Cov_{xy}}{\sigma_x \sigma_y}, \tag{2.12}$$

where Cov_{xy} is the covariance of the two variables (the estimated and well-log acoustic impedances), and σ_x^2 and σ_y^2 are the variances of these variables.

2.3 Porosity estimation

The total or effective porosity is another key petrophysical material property of great interest in reservoir studies. Porosity of the in-situ rock is routinely measured in various ways: using core samples in the laboratory; from well-log data, from interval velocities of seismic traces and from mathematical modelling (Willie et al., 1956). Well-log and core-sample observations yield high detail and spatial resolution of measurements, but they are only available at sparse locations. By contrast, seismic data allow obtaining continuous areal coverage between well locations. To obtain such coverage, one needs to find a combination of seismic attributes that can be calibrated at the wells and used as a proxy for porosity.

From physical considerations, rock porosity is most sensitive to the acoustic impedance (AI), and therefore empirical relations are always sought between these quantities. The density of porous rock equals $\rho = (1-\varphi)\rho_s + \varphi\rho_f$, where ρ_s is the density of the solid matrix, ρ_f is the density of the pore fluid, and φ is the porosity. The effect of porosity on the P-wave velocity is weaker, and it was estimated by a similar average of slownesses: $V^{-1} = (1-\varphi)V_s^{-1} + \varphi V_f^{-1}$ (Willie et al., 1956). From these relations, for different relations between ρ_s and ρ_f, and also V_s and V_f,, the logarithm of AI can be expected to linearly relate to $\log\left[\varphi/(1-\varphi)\right]$ (Kumar et al., 2016):

$$\log Z \approx \log Z_0 + b \log\left(\frac{\varphi}{1-\varphi}\right), \tag{2.13}$$

where $\log Z_0$ is the intercept and $b < 0$ is the slope of this dependence. An inverse of this relation gives a slightly nonlinear relation of φ to Z. Verma and Biswal (2012) investigated similar empirical relations between porosity and the extended elastic impedance evaluated at nonzero source-receiver offsets in seismic data (Whitcombe et al., 2002).

Based on near-linear empirical relations like eq. (2.13), Din and Hongbing (2019) and Ekone et al. (2020) created comprehensive models of hydrocarbon reservoirs using geostatistical methods such as cokriging, stochastic Gaussian simulations, and probabilistic neural networks. These methods allow making probabilistic estimates of reservoir porosities and their uncertainties from seismic and well-log data.

In this thesis, I also start by estimating a linear empirical approximation for porosity from seismic data but using not only AI but a several other seismic attributes. I use the multi-linear regression, which attempts presenting the desired outcome (porosity) as a linear combination of the input attributes. If we use N_f seismic features (columns of data matrix X_{ij}) to estimate porosity φ_i at i^{th} data point, we will have

$$\mathbf{y} = \mathbf{X}\mathbf{w} + \mathbf{w}_0 , \tag{2.14}$$

where $\mathbf{w}$ is a vector row of weights, $\mathbf{w}_0$ is the bias vector, and $\mathbf{y}$ is the vector of predicted porosity values at all n data points. If $\mathbf{X}$ and $\mathbf{y}$ represent the training dataset (seismic data near wells), then eq. (2.14) contains n equations for the unknown elements of $\mathbf{w}$ and $\mathbf{w}_0$. By convention in this thesis, I denote all these model variables by vector $\boldsymbol{\theta}$. Equations (2.14) can be solved for $\boldsymbol{\theta}$ in the least-squares sense by minimizing the cost function

$$J(\boldsymbol{\theta}) = \frac{1}{2n}\sum_{i=1}^{n}(\varphi_i - y_i)^2 , \tag{2.15}$$

where φ_i is the observed data (porosity) and y_i is the predicted value at sample i. Methods of linear and nonlinear inversion are well known (Menke, 1984) and are not discussed in this thesis.

With broad ranges of parameter variations such as porosity varying from 0.2 to 0.7 and AI values between 2500–6000 m/s·g/cm^3, the relation between φ and Z may become significantly nonlinear (Kushwaha et al., 2020). Kushwaha et al. (2020) used a small three-layer feed-forward to model such nonlinear porosity-AI relations. In this thesis, I also utilize an ANN to derive nonlinear predictions of a porosity volume but again combining the AI with multiple other seismic attributes. I tried several algorithms called the Radial Basis Function Network (RBFN), multi-layer feed-forward network (MLFN), and Generalized Regression Neural Network (GRNN) (Haykin, 2009). These algorithms are outlined in the next section. As described in chapter 4, the best results for porosity were obtained by using the GRNN.

2.4 Prediction

To create a predictive model that would fit the training data sufficiently accurately and produce predictor values normalized within the range [0,1], nonlinear and sufficiently flexible

forward models are needed. Such nonlinear models can be provided by the logistic regression and neural-network approaches. However, linear inverse lies at the core of all these approaches, and it is also useful for preforming initial feasibility studies and finding starting models for nonlinear inversions. In the following subsections, I briefly describe the linear and nonlinear approaches to property prediction used in this project.

2.4.1 Linear inverse

The simplest predictor that can be used for property prediction is a linear combination of the features with bias (intercept) (Haykin, 2009). The vector of predictors for property **y** (for example, porosity or probability of finding the k^{th} class of seismic facies) is sought in the form of a linear combination of the data with additional bias terms:

$$\mathbf{y} = \mathbf{X}\mathbf{w} + b\mathbf{I}_n, \tag{2.16}$$

where n is the number of data samples in the training set, **X** is the data (feature) matrix, **w** is the desired vector of weights, **b** is the bias value, and $\mathbf{I}_n$ denotes a column of n values equal one. This equation can be rewritten as a linear problem:

$$\mathbf{y} = \mathbf{K}\mathbf{m}, \tag{2.17}$$

where $\mathbf{K} = \begin{bmatrix} \mathbf{X} & \mathbf{I}_n \end{bmatrix}$ is the kernel matrix and the model vector is $\mathbf{m} = \begin{pmatrix} \mathbf{w} \\ b \end{pmatrix}$. The least-squares solution is obtained as (Menke, 1984)

$$\mathbf{m} = \left(\mathbf{K}^T\mathbf{K}\right)^{-1}\mathbf{K}^T\mathbf{y}. \tag{2.18}$$

After $\mathbf{m}_k$ is obtained, the predictors for any data point can be obtained by the matrix multiplication in eq. (2.17).

2.4.2 Logistic regression

Logistic regression is a popular method for mapping input feature vectors **x** onto scalar outputs y ($\mathbf{x} \rightarrow y$) so that the values of y are within the normalized range [0,1] (Haykin, 2009). The function $y = h(\mathbf{x})$ is called the hypothesis. In contrast to the general parameter prediction where

$h(\mathbf{x})$ is some deterministic function of $\mathbf{x}$ (section 2.2), in logistic regression, the hypothesis has a different meaning. This meaning consists in viewing the rows in data matrix as occurrences of a certain event. For a classification problem, this event may consist in the data point belonging to the k^{th} class. The logistic regression of the dataset obtains a directional vector $\boldsymbol{\theta}$ in the space of features such that if we measure the dot product of $\mathbf{x}$ with $\boldsymbol{\theta}$: $x_\beta = \boldsymbol{\theta}^T\mathbf{x}$, then the probability of finding this event in area $x_\beta \leq \xi$ equals some cumulative probability function $P(\xi)$. As a probability function, $P(\xi)$ must tend to zero for $\xi \to$ -∞ and to one for $\xi \to +\infty$. The usual practical choice for $P(\xi)$ is the 'logistic', or sigmoid function (Figure 2-1):

$$g(z) = \frac{1}{1+e^{-z}}. \tag{2.19}$$

Therefore, the hypothesis function for logistic regression is a sigmoid-shaped transformation of the linear projection of $\mathbf{x}$ onto vector $\boldsymbol{\theta}$:

$$h_\theta(\mathbf{x}) = g(\boldsymbol{\theta}^T\mathbf{x}). \tag{2.20}$$

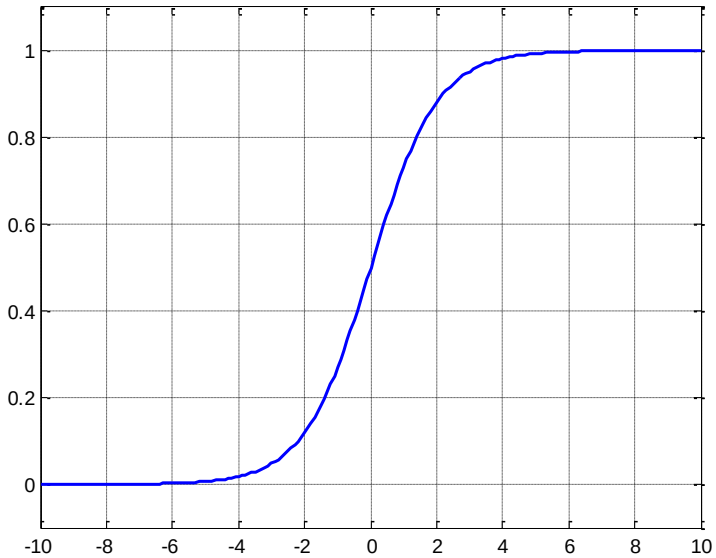

Figure 2-1: Sigmoid function

To perform logistic regression for a binary classification, we need to use nonlinear optimization to make the elements of vector $\boldsymbol{\theta}$ reach the global minimum of the following cost function. The cost function is given by a sum of two terms: $J(\boldsymbol{\theta}) = J_0(\boldsymbol{\theta}) + J_1(\boldsymbol{\theta})$, where J_0 and J_1

are contributions from all data points with $y = 1$ (occurrences of the event) and $y = 0$ (non-occurrences) respectively:

$$J_0(\boldsymbol{\theta}) = -\frac{1}{n}\sum_{i=1}^{n} \log\left[h_\theta(\mathbf{x}^i)\right]\Big|_{y^i=0}, \tag{2.21}$$

$$J_1(\boldsymbol{\theta}) = -\frac{1}{n}\sum_{i=1}^{n} \log\left[1-h_\theta(\mathbf{x}^i)\right]\Big|_{y^i=1}, \tag{2.22}$$

where log means logarithm and n is the total number of data samples. This minimization of the negative sum of logarithms of probabilities corresponds to finding the distribution with largest entropy. The sum of J_0 and J_1 can be written as one summation over all data as

$$J(\boldsymbol{\theta}) = -\frac{1}{n}\sum_{i=1}^{n}\left[y^i \log(h_\theta(\mathbf{x}^i)) + (1-y^i)\log(1-h_\theta(\mathbf{x}^i))\right]. \tag{2.23}$$

The algorithms for finding the minimum of $J(\boldsymbol{\theta})$ often require setting starting models and additional parameters. A good starting model for $\boldsymbol{\theta}$ can be obtained by fitting the same data with a linear inverse (subsection 2.4.1). Iterative algorithms such as gradient descent are commonly used for finding a global minimum of a function. Such algorithms usually contain an adjustable parameter α which controls the learning rate (size of update made during one iteration). Since the direction of the update and curvature of the cost function is constantly changing, with larger learning rate, we may miss the global minimum, and the process of convergence to the global minimum may become oscillatory or completely unstable. Small values of parameter $\alpha < 1$ reduce the steps and slow the inversion down, but the convergence to the minimum of $J(\boldsymbol{\theta})$ becomes more stable. The most commonly used rates range from $\alpha = 0.001$ to 0.1. However, generally, the value of α should not affect the final solution.

In addition to α, another algorithm-related parameter λ may be required for finding the optimal $\boldsymbol{\theta}$. As with other inverse problems, overfitting may occur from logistic regression (Haykin, 2009). Overfitting means producing accurate predictions for the examples in the training set but capturing too much of their detail. The model may contain high variability (be 'rough') and not generalizing well enough to make proper predictions on new (previously unseen) examples. This problem arises whenever there is a large number of features and limited input data in the

training dataset, i.e. the inverse problem is not very strongly determined. To solve the problem of overfitting, I use a procedure called regularization. The purpose of regularization is to keep fitting all the features but also to try reducing the values of parameters θ_j themselves. Smaller values of θ_j give larger gradient of the logistic function (Figure 2-1) and consequently sharper transitions between classes in the feature space. In Tikhonov regularization, this is achieved by adding the squared length of the model vector $\boldsymbol{\theta}$ to the cost function (Menke, 1984; Haykin, 2009):

$$J(\boldsymbol{\theta}) = \frac{1}{n}\left[-\mathbf{y}^T \log(\mathbf{h}) - (1-\mathbf{y})^T \log(1-\mathbf{h})\right] + \frac{\lambda}{2n}\boldsymbol{\theta}^T\boldsymbol{\theta}\,. \tag{2.24}$$

With larger values of λ, the regression approximates the probability function $P(y)$ less accurately but with smaller average values of θ_j, and the model is 'smoother'. However, regularization also creates problems of its own: 1) it introduces additional parameters like λ which have little physical meaning and may be difficult to set, 2) it introduces subjectivity of the selected form of regularization terms and 'prior' models, and 3) it reduces the accuracy of data prediction. Similarly, to α, parameter λ should be selected so that it does not affect the data prediction significantly.

2.4.3 Feed-forward Artificial Neural Network

Logistic regression can only be applied to hypotheses in the form of a single transformation of a linear function (eq. (2.20)). However, for more complex applications, nonlinear relations between multiple inputs x_i and outputs y_j are required. A general computational model for such relations simulating the memory and decision-making processes of a network of neurons was introduced in bioinformatics by McCulloch and Pitts (1943). The nonlinear transformation $\mathbf{x} \rightarrow \mathbf{y}$ depends on the connections between the neurons and is usually depicted graphically (Haykin, 2009). For example, Figure 2-2 shows a multi-layer feed-forward perceptron ANN (MLFN), in which the neurons (dots) are arranged into layers (yellow boxes), and the evaluation of the response is performed by a 'propagation' process from the input to the output (left to right in Figure 2-2). Layer $A^{(1)}$ is the input layer, $A^{(2)}$ and $A^{(3)}$ are called hidden layers, and $A^{(4)}$ is the output layer. This diagram represents a mathematical relation $\mathbf{x} \rightarrow \mathbf{y}$ evaluated by a series of matrix multiplications and nonlinear functions:

$$\begin{cases} \mathbf{y}^{(1)} = \mathbf{x} + \mathbf{b}^{(1)}, \\ \mathbf{y}^{(i+1)} = \mathbf{W}^{(i)}\mathbf{A}^{(i)}\left(\mathbf{x}^{(i)}\right) + \mathbf{b}^{(i+1)} \qquad \text{for} \quad i = 1 \ldots N-1, \\ \mathbf{y} = \mathbf{A}^{(N)}\left(\mathbf{x}^{(N)}\right) + \mathbf{b}^{(N)}. \end{cases} \tag{2.25}$$

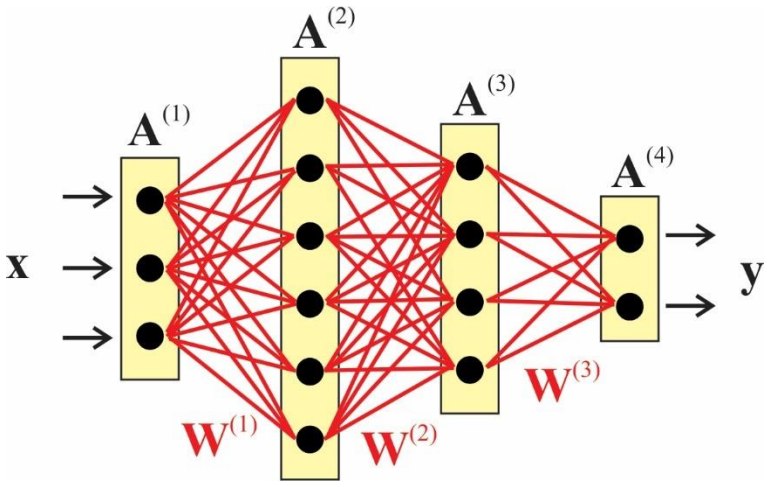

Figure 2-2: A general multi-layer feed-forward artificial neural network (ANN, MLFN) with four layers representing eqs. (2.23) (modified after Morozov (2021)). Yellow boxes labeled $\mathbf{A}^{(i)}$ are the layers, black dots are the neurons (layer elements) containing the activation functions and bias terms $\mathbf{b}^{(i)}$), and red lines are the ANN weights $\mathbf{W}^{(i)}$.

In these equations, $\mathbf{x}^{(i)}$ and $\mathbf{y}^{(i)}$ denote the inputs and outputs of the i^{th} layer, functions $\mathbf{y} = \mathbf{A}^{(i)}(\mathbf{x})$ represent the nonlinear elements, matrices $\mathbf{W}^{(i)}$ are the ANN weights (red lines in Figure 2-2), and terms $\mathbf{b}^{(i)}$ are called biases. By adding a constant input equal one into each vector $\mathbf{x}^{(i)}$, the biases can be viewed as elements of weight matrices. In geophysical applications (e.g., Kushwaha et al., 2020), functions $\mathbf{A}^{(i)}(\mathbf{x})$ are usually taken in the form of a smooth sigmoid function (Figure 2-1) applied to each element of $\mathbf{x}$ independently, or softmax function using all neurons in an ANN layer simultaneously. The softwax function outputs value from the i^{th} output is

$$y_i = \frac{e^{x_i}}{\sum_{k=1}^{K} e^{x_k}}, \tag{2.26}$$

where $i = 1,\ldots K$, K is the total number of outputs, x_i represents the values from the neurons of the output layer, and the exponentials act as the non-linear function. Because $\sum_{i=1}^{K} y_i \equiv 1$ and $y_i \to 1$

when the i^{th} input is much larger than all others, values y_i can be interpreted as class probabilities. Thus, for the i^{th} estimator, the softmax function normalizes the outputs in the same way as eq. (2.20).

The weights and biases comprise the information about mapping $\mathbf{x} \rightarrow \mathbf{y}$ stored in the selected ANN model. These parameters are determined by 'training' the ANN to reproduce a set of known examples ($\mathbf{x}$,$\mathbf{y}$). Denoting the vector of all model parameter $\boldsymbol{\theta}$ and the mapping function (eq. 2.23) by $\mathbf{y}_\theta(\mathbf{x})$, vector $\boldsymbol{\theta}$ is obtained by minimizing the misfit of the training data (cost function)

$$J(\boldsymbol{\theta}) = \frac{1}{2n}\sum_{i=1}^{n}\left[\mathbf{y}_i - \mathbf{y}_\theta(\mathbf{x}_i)\right]^T\left[\mathbf{y}_i - \mathbf{y}_\theta(\mathbf{x}_i)\right], \tag{2.27}$$

where $i = 1,\ldots n$ denotes the i^{th} training data point. In many cases (particularly with multiple and large hidden layers as in the example in Figure 2-2), the inverse for $\boldsymbol{\theta}$ becomes nonunique or unstable, and the training process is prone of overfitting. To control overfitting, regularization terms are added as in eq. (2.24):

$$J(\boldsymbol{\theta}) = \frac{1}{2n}\sum_{i=1}^{n}\left[\mathbf{y}_i - \mathbf{y}_\theta(\mathbf{x}_i)\right]^T\left[\mathbf{y}_i - \mathbf{y}_\theta(\mathbf{x}_i)\right] + \frac{\lambda}{2n}\boldsymbol{\theta}^T\boldsymbol{\theta}\,. \tag{2.28}$$

The above relations for cost functions are useful for predicting continuous output values, such as rock porosity in chapter 4. For classification problems such as considered in chapters 5 and 6, the training data $\mathbf{y}_i$ are binary vectors consisting of values of 0 or 1. In such cases, quantities $\mathbf{y}_\theta(\mathbf{x})$ predicted by the ANN are interpreted as probabilities of observing the different classes, and the cost function is modified by taking logarithms as in logistic regression (eq. (2.24)):

$$J(\boldsymbol{\theta}) = \frac{1}{2n}\sum_{i=1}^{n}\sum_{k}\left\{-y_{ik}^T \log\left[y_{\theta k}(\mathbf{x}_i)\right] - (1-\mathbf{y}_i)^T \log\left[1-y_{\theta k}(\mathbf{x}_i)\right]\right\} + \frac{\lambda}{2n}\boldsymbol{\theta}^T\boldsymbol{\theta}, \tag{2.29}$$

where k denotes the class number (k = 1,2, 3, 4 in chapter 5). As above, careful analysis should be performed to search for the optimum sizes of the hidden layers (one hidden layer is used in chapters 5 and 6) and the regularization factor λ.

Minimization of cost functions (2.22) or (2.26) is a nonlinear inverse problem, which can be solved by iterative linearization methods (Menke, 1984). Due to the usually (moderately) large number of weights, backpropagation, Newton, or conjugate-gradient algorithms are commonly

used (e.g., Kelley, 1960; Werbos, 1982). In these methods, derivatives of the cost function with respect to parameters θ are evaluated and the weights and biases are updated by proceeding from the output of the network to its input (back-propagation). By iteratively repeating these updates with different strategies for selecting update directions, the desired solution θ is obtained, and the ANN training is thereby completed.

Once all weights and biases in the network are determined, eqs. (2.23) (forward propagation) can be used for predicting continuous geophysical properties (chapter 4) or probabilities of classes (chapters 5 and 6) using the data not included in the training datasets. The prediction process represents an application of the trained neural network to new samples using the forward algorithm.

2.4.4 Other types of neural networks

In this section, I outline two other algorithms used in chapters 4 and 5: the Radial Basis Function Network (RBFN) and the Generalized Regression Neural Network (GRNN), also known as the Probabilistic Neural Network (PNN). In the machine-learning literature, mathematical expressions for any functions $\mathbf{y}(\mathbf{x})$ by using summations, multiplications, and nonlinear transformations (as in eq. (2.25)) are often referred to as neural networks (e.g., Haykin, 2009). The RBFNs and the GRNNs are of this kind. However, these expressions are based on pure spatial relations and do not employ the topology of connections between neurons. In other subject areas, such representations of functions $\mathbf{y}(\mathbf{x})$ are known as interpolation or smoothing kernels, kriging and cokriging, basis functions, solution functions, shape functions, or finite elements (e.g., Polyanin and Manzhirov, 2007).

Let us consider an input set of n samples $\mathbf{x}_i$ ($i = 1,..n$) in an M-dimensional space of attributes, which are, for example, geophysical log values in a borehole. Let us assume that we also have a function $f(\mathbf{x}_i) = y_i$ (e.g., porosity) measured at these samples, and our goal is to predict this function at an arbitrary other point at which the attributes equal $\mathbf{x}$. In basis-function methods, the output is sought as a weighted sum of some basis functions $\varphi_k(\mathbf{x})$ (termed activation functions in machine learning):

$$f(\mathbf{x}) = \sum_{k=1}^{K} W_k \varphi_k(\mathbf{x}), \tag{2.30}$$

where W_k are the weights.

The difference between the RBFN and GRNN is in the selection of the basis functions and weights. In RBFN, Gaussian functions of Euclidean distances between point **x** and a set of *K* basis centers $\boldsymbol{\mu}_k$ ($k = 1,\ldots K$) are used:

$$\varphi_k(\mathbf{x}) = \exp\left(-\frac{\|\mathbf{x}-\boldsymbol{\mu}_k\|^2}{2\sigma^2}\right), \tag{2.31}$$

where σ is the standard deviation of the distribution of attributes, or smoothing parameter. A small number of basis centers is used ($K \ll n$), and these centers and parameter σ are determined using some unsupervised clustering algorithm. After locating the clusters, weights W_k are found by inverting equations $f(\mathbf{x}_i) = y_i$ for the known (training) data by the least-squares method.

The RBFN approach is appropriate for large training datasets in which closely spaced points may contain noise or conflicting data. By using clusters as basis centers, a smaller set of significant features $\boldsymbol{\mu}_k$ is unidentified. These basis centers are spatially separated (in the attribute space).

In contrast to RBFN, the GRNN algorithm belongs to the Bayesian (probabilistic) Neural Networks, which are feed-forward networks without backpropagation, i.e., without using inverse algorithms for training. In the GRNN, each training data point is used as a basis-function center ($\boldsymbol{\mu}_k = \mathbf{x}_k$, and $K = n$) and training data are used as weights: $W_k = y_k$. The basis functions are obtained from Gaussian functions in eq. (2.31) by normalizing them so that $\sum_{k=1}^{n} \varphi_k(\mathbf{x}) = 1$ at any **x**:

$$\varphi_k(\mathbf{x}) = \frac{\exp\left(-\frac{\|\mathbf{x}-\mathbf{x}_k\|^2}{2\sigma^2}\right)}{\sum_{j=1}^{n} \exp\left(-\frac{\|\mathbf{x}-\mathbf{x}_j\|^2}{2\sigma^2}\right)}. \tag{2.32}$$

Thus, the predictions by a RBFN or GRNN can be viewed as spatial smoothing of the data (eq. 2.30) by using kernels (2.31) or (2.32), respectively. In both of these methods, the data at the training samples y_i are not predicted accurately, but the general variation of the data is captured. For application of both of these methods, it is important to carefully estimate parameter σ, which should be close to the characteristic distance between clusters in the data.

2.5 Classification

The key goal of most applications in this thesis consists in obtaining a discrete classification of the dataset, i.e. assigning an integer class number to each data point. Such assignment can be done based on any type of predictors described in the preceding sections. To classify the data into N_c classes, we need to split the training dataset into N_c subsets and construct N_c predictors $f_j(\mathbf{x})$. In most applications, the values of $f_k(\mathbf{x})$ will lie within the interval [0,1]. When training the algorithm or constructing another model for the k^{th} predictor, we need to specify the target output for the i^{th} data point equal $f_i = 1$ if the point belongs to k^{th} class and $f_i = 0$ otherwise. After training or inversion, the model will predict larger values $f_j(\mathbf{x}_i)$ for feature vectors $\mathbf{x}_i$ belonging to class j. Using these outputs, the predictors can be normalized as

$$p_k(x) = \frac{f_k(x)}{\sum_{i=1}^{K} f_i(x)}. \tag{2.33}$$

These normalized predictors have a sum of one, and consequently $p_k(\mathbf{x})$ can be interpreted as probabilities of finding class k at point $\mathbf{x}$. Note that this is the same normalization as used, for example, in the PNN method (section 2.3). In ANNs using the softmax activation function (eq. (2.26)), this normalization is included in the construction of the algorithm.

Based on the normalized predictors (2.33), the classification decision appears straightforward: for a feature vector $\mathbf{x}$, class k with the largest $p_k(\mathbf{x})$ should be selected (Haykin, 2009). Let us denote the corresponding highest p_k by P_{max}. However, clearly, this selection would still have a low degree of statistical confidence if other values of $p_j(\mathbf{x})$ are close to the selected one. Therefore, I include not only the most likely class label k but also the estimated probability $p_k(\mathbf{x})$ in the outputs of the classification procedure. In this way, I produce images of the

estimated probability of each selected class for each sample of the subsurface, and images of confidence of the classification (next subsection).

With regard to their use for classification, each of the predictor models considered in the preceding section has advantages and disadvantages. Because of its form and relatively small number of model parameters, linear inverse is simple, well-constrained, and efficient, but it usually results in high fractional classification errors. The fractional classification error shows the percentage of samples whose classification values were incorrectly predicted. The values of predictors may also become negative or larger than one, which makes it difficult to interpret them as estimates of class probability. Therefore, this method is generally not a proper technique for final classification of the data. However, this approach often captures the key relationships within the data. Linear prediction is also useful for forming starting and prior models for nonlinear classification.

The advantage of using ANN predictors for classification is in their nonlinearity, universality, and flexibility. Numerous ANN architectures can be used, with different numbers of layers, neurons, and shapes of the activation functions. With larger numbers of ANN weights and bias terms, training can be made more accurate. At the same time, the universality of the ANN model is also its disadvantage. The predictors are obtained by pure empirical fitting of the training data **x** consisting of zeros and ones, and the resulting values cannot be rigorously viewed as probabilities of the classes. With large numbers of ANN weights, the ANN can overfit the training data and allow data noise to influence the model.

If using logistic regression for obtaining class predictors, its advantage is in a clear physical meaning of the output quantity as the probability of finding the specified class k at point (feature set) **x** in the data. This probability is defined within the expected range [0,1], although unfortunately, the different classes do not exclude each other. A disadvantage of the standard logistic regression approach is in limited number of variable parameters and consequently limited accuracy of data fitting. In one pass of logistic regression, the probability of only one class is estimated. However, to solve for a multiple classification, we can choose one class k, consider all other classes as non-occurrences of k, and estimate the probability of class k as described above. This method is known as One-versus-All classification. After repeating this estimation for each class k, their predictors $f_k(\mathrm{x})$ can be combined by eq. (2.33).

Thus, a large number of choices for selecting prediction and/or classification algorithms is available, and the appropriate algorithm should be selected by trials and quality control. However, selection of the criteria of this model quality control may become a nontrivial issue in geophysical applications.

2.5.1 Probability and confidence of the predicted class

By using eq. (2.33), outputs of any ANN or similar prediction algorithms can be normalized into the range [0.1], which allows them to be interpreted as estimated probabilities (likelihoods) of finding these classes at a given data point **x** (Haykin, 2009). In many ANN algorithms, the softmax activation function (eq. (2.26)) is available, which performs this normalization automatically.

In many applications, it is useful to output not only the resulting class label *i* but also some measure of statistical confidence of this classification. For example, when using attributes not allowing differentiation between the classes, near-random values of some classes would be reported, but the confidence of these identifications should be low. Unfortunately, this issue of confidence is usually not considered in conventional approaches to AI classifications, but it is important for reservoir characterization.

Denoting the measure of classification confidence by $C(\mathbf{x})$, two useful expressions for it can be proposed using the values of normalized estimators p_i in equation (2.33). First, we can use the value of P_{max} itself:

$$C = P_{\max}. \tag{2.34}$$

This measure would indicate the inferred probability of observing the class *i* reported for data point **x**. Values of *C* close to one correspond to high confidence of identification of class *i* compared to all other classes, and values close to 1/*K* would mean low confidence.

In addition to $p_k(\mathbf{x})$ for the selected class, another measure of confidence can be obtained by taking the ratio of the largest p_k to the second largest one:

$$C = \frac{p_{\max}}{p_{next\,\max}}. \tag{2.35}$$

This value is always greater or equal one, and larger values indicate lower ambiguity of the

classification, i.e., situations when class k is more likely that the next competing alternative.

Using the probability of each prediction (2.33) or its confidence levels (2.34) and (2.35), I also filter the classification seismic sections to only show the zones with desired minimum confidence (chapter 5).

2.6 Construction of optimal features

As shown above, the accuracy and stability of classification critically depends on the selection of input features. In seismic data interpretation, a vast number of input time series (seismic attributes) is available, and selection of an optimal subset of them is a nontrivial task. In engineering and machine-learning applications, this selection optimal attributes (feature reduction) is performed by analyzing the statistical distribution of the data (Haykin, 2009; section 2.1 in this thesis). In this section, I propose an improved approach to this selection additionally utilizing the relations between the expected classes in the training dataset. The idea of this approach is in constructing combinations of input attributes which produce the largest validation scores in classification. In this sense, the procedure is analogous to the iterative cross-validation used for determining the optimum inputs in PNN-based parameter predictions (section 2.3).

Denoting the number of classes K and the number of features N_f, the proposed procedure of optimal feature selection is as follows:

1) Create up to N_f binary classifications of the data. In the first $k = 1 \ldots K$ of these classifications, include one class k (assign to it a target value of one) against all other classes (target value zero). For the next $k = (N_c + 1) \ldots 2K$ classifications, remove the k^{th} class and repeat the same procedure for the remaining $K - 1$ classes. Continue this procedure until a sufficient number of binary classifications is obtained.
2) Using logistic regression, calculate probabilities p_k for each of the binary classifications above. Sort the values of p_k in descending order. These values of k give the combinations of the original classes most classifiable by the training data.
3) In the k^{th} logistic regression, parameters vector θ_k is returned. This vector gives the direction in the feature space in which the k^{th} binary classification is most effective.

By evaluating the matrix product $\mathbf{u}_k = \mathbf{X}\boldsymbol{\theta}_k$ (where $\mathbf{X}$ is the data matrix), feature vectors $\mathbf{u}_k$ are obtained. Finally, by orthogonalizing and normalizing these vectors (section 2.1), the desired k^{th} optimal attribute $\mathbf{x}_k$ is obtained.

The above procedure outputs a set of features that are mutually uncorrelated, normalized, and optimized by their ability to differentiate between different combinations of classes. By selecting a subset of these features with p_k exceeding a certain threshold, feature selection (dimensionality reduction) can be performed.

CHAPTER 3: Data and Geology of Mansuri Oil Field

In this chapter, I describe the geological setting, data, previous results, and goals of the first study of this thesis, focusing on a characterization of reservoir quality in Mansuri oil field in Iran. This field was discovered in 1963 using a 2-D seismic exploration project. From 1986 to 2005, 65 wells were drilled using only regional geology studies and previous well geological information. In a previous study (Zahmatekesh et al., 2021), I with co-authors developed an initial classification of the reservoir. In this thesis, I keep this basic classification but improve and extend it using the methods described in chapter 2.

In section 3.1, I give an overview of the geology and results of stratigraphic interpretations of the Mansuri Oilfield in Iran, history of the reservoir, and petrophysical data. In section 3.2, the seismic dataset is described. In section 3.3, I overview the well-log geophysical data and describe the previous seismic interpretation of the study area. Further analysis of these datasets will be conducted in chapters 4 and 5.

3.1 Geology of the study area

In this study, I aim at performing reservoir characterization using seismic facies classification on Asmari Formation in the Mansuri oil field, located in southwest Iran and north Dezful embayment (Figure 3-1). This formation is considered the main oil-producing reservoir unit in Iran.

The Asmari formation mainly consists of interbedded carbonates, sandstones, and shales of the Oligocene to Miocene age. In Dezful Embayment of the Zagros fold belt, oil and gas are trapped in Asmari Formation in asymmetrical anticlines sealed by evaporites of the overlying Gachsaran Formation (Dashti and Sfidari, 2016; Zahmatkesh et al., 2017; Abdizadeh et al., 2017).

The Mansuri Asmari reservoir consists of two-thirds of limestones and dolomites and one-third of clastic sediments. The production has been and is still mainly from the sandstones. Based on well-core descriptions and wireline logs, the Asmari reservoir of Mansuri oil field is subdivided into eight zones indicated in Figure 3-2. In this figure, sandstones are shown by yellow color, and pink and cyan colors show carbonites. Among the eight zones, carbonates are dominant in zones

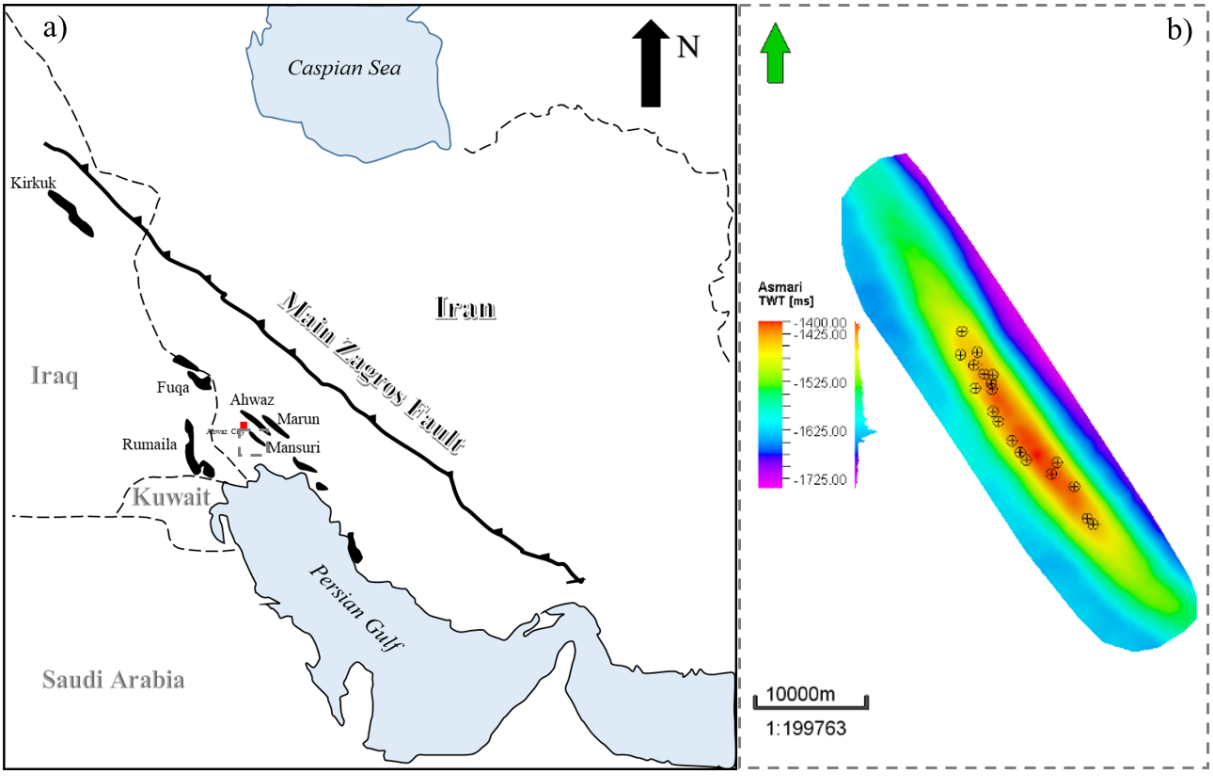

Figure 3-1: a) Regional location map of the Mansuri oil field (red box in (a)) in southwestern Iran. The large and important Tertiary and Cenomanian oil fields are also shown in the region. b) Time structure map (two-way time) of the Asmari Formation (Oligo-Miocene) in the study. Well locations are spotted as small circles with plus sign.

1, 6, 7, and 8, and other zones contain extensive sandstones. Thus, beneath the first zone which contains carbonates, the first half of the Asmari reservoir mainly contains sandstones (yellow color in Figure 3-2).

The stratigraphic zones in Figure 3-2 were obtained from three logged and cored key wells, along with additional 15 logged wells of the Mansuri oil field. After that, depth matching was performed, which included the correlation of the well-logs with the related core samples, thin-section analysis, and permeability measurements in the samples. The petrophysical parameters selected from wireline logs consist of bulk density (denoted RHOB), sonic transit time (DT), compensated gamma-ray (GR), neutron porosity (NPHI) and effective porosity (PHIE) (Figure 3-2). These well-logs parameters are directly related to reservoir quality and reveal the reservoir lithology.

The petrophysical classification which was developed in a previous study (Zahmatkesh et al., 2021) and further improved in this thesis (chapter 5) is based on the concept of electrofacies. One group of electrofacies represents a cluster of similar well-log measurements, demonstrating the physical and chemical characteristics of specific rocks and fluids involved in the volume. In addition, electrofacies need to be calibrated based on the laboratory analysis of the core plugs to confirm their geological consistency.

The principal use of electrofacies consists in their use as known labels for training machine learning algorithms to predict the geological (or reservoir) facies through the oil field, in areas where the core samples are unavailable. The importance of this concept consists in its being a link between pure numerical geophysical measurements in the well and the geological, petrophysical, and economic properties of the reservoir.

After calibrating the well logs using core-sample and petrophysical data and merging some electrofacies due to sparse distribution in the studied wells, the final electrofacies in the Asmari reservoir were subdivided into four subfacies, or classes denoted EF1 through EF4. Among these electrofacies, the first two classes are related to sandstone rocks (EF1 and EF2), and classes EF3 and EF4 are related to limestone rocks (Figure 3-3).

Zones with high permeability and high porosity are usually considered as having good reservoir quality. Therefore, electrofacies related to high-reservoir quality rocks should show a lower gamma-ray, higher neutron porosity, lower density, and higher sonic transit time (i.e., lower seismic-wave velocity). Thus, classes EF1 and FF3 contain rocks with the best reservoir quality, EF2 has a medium reservoir quality, and class EF4 with low porosity and high percentages of shale considered as non-reservoir (Figure 3-3; Zahmatkesh et al., 2021).

Since core data analysis, thin-section studies and petrophysical parameters were evaluated for each of the four electrofacies, we have an almost complete lithological information about each of them (Zahmatkesh et al., 2021). Class EF1 includes unconsolidated channel sandstones and medium to coarse-grained sandstones with dolomitic cement, in which interparticle porosity was regarded as the predominant pore type. Class EF2 included fine to very fine-grained argillaceous or dolomitic sandstones whose interparticle porosity was partially impeded by dolomitic cement

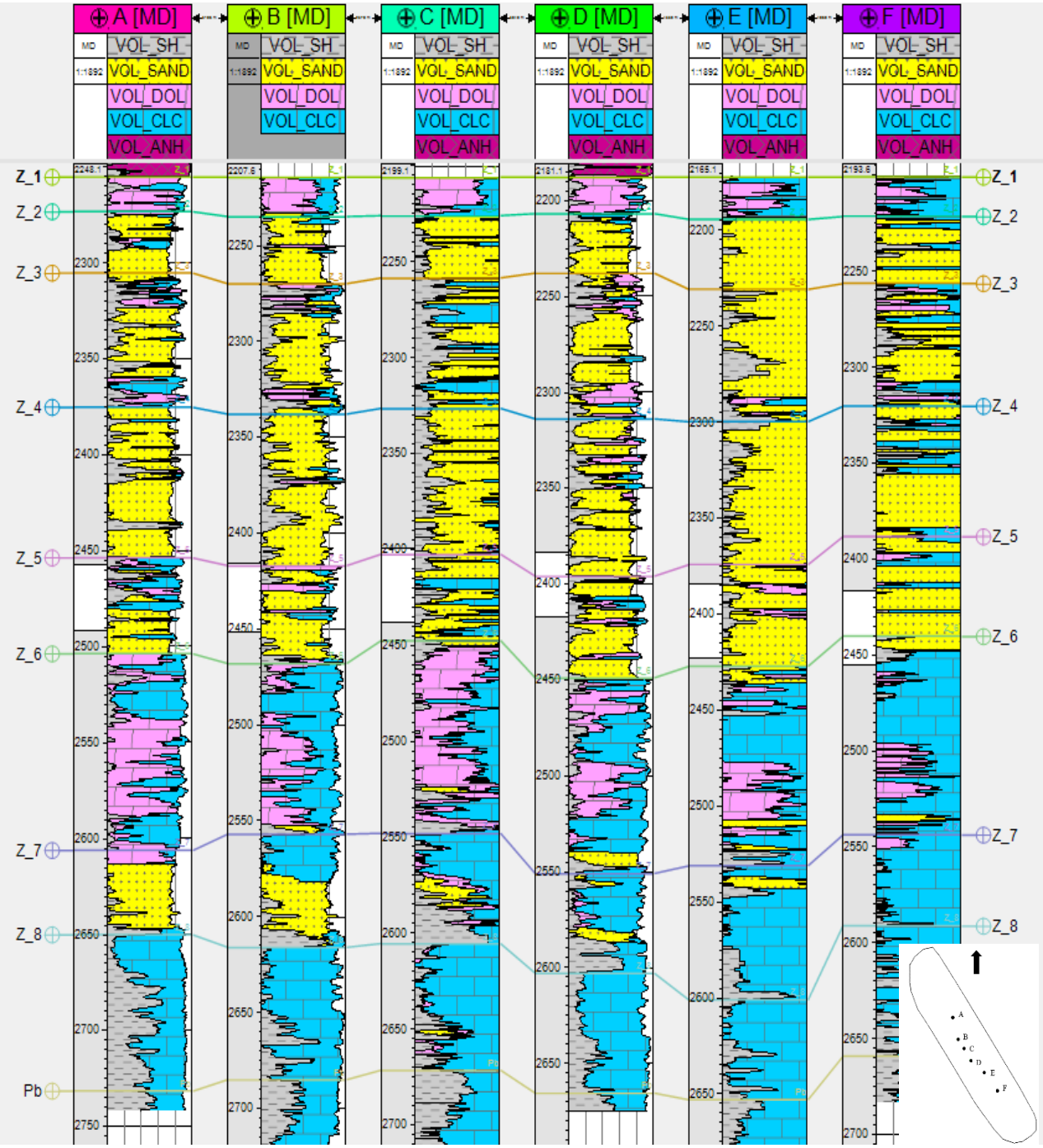

Figure 3-2: Lithology or stratigraphy log in six vertical wells in the studied area. Z_1 is the top and Pb is the bottom of the reservoir. Reservoir zones are shown in the left and right side of figure.

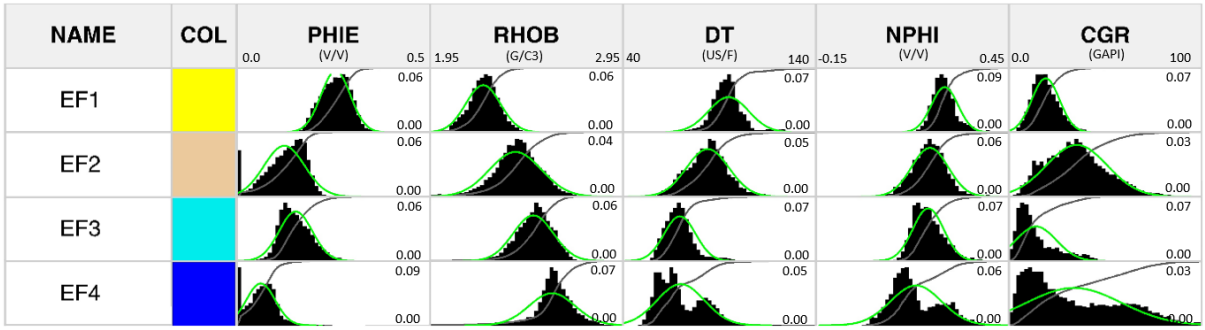

Figure 3-3: Final electrofacies model and histograms of wireline log values in Asmari reservoir, Mansuri field. The green line is the normal density function. In each histogram, the Y axis shows the frequency and the X axis the wireline logs values: PHIE = effective porosity; ROHB = bulk density; DT = sonic transit time; NPHI = Neutron porosity; CGR = compensated gamma-ray.

or clay matrix. Class EF3 is characterized by vuggy and intercrystalline porosity, including dolomitic skeletal packstone and grainstone. Finally, the carbonate samples of mudstone with microporosity and low reservoir quality were associated with the non-reservoir class EF4. shows the four classes of electrofacies and their characteristic porosity, density, gamma-ray wireline logs, lithology types, and examples of thin sections. Thus, the electrofacies are also significantly distinct by their lithologies, structure of the matrix, grain sizes, types of porosity, depositional textures, and also the characteristic values of some wireline logs.

The lithology log consists of several lithologies such as anhydride, shale, sandstone, calcite and dolomite (Figure 3-2). Sandstone, dolomite and calcite are dominant lithologies in the Asmari reservoir, which means that it can be subdivided into two primary lithologies: sandstone and carbonate. The sandstone compartments of the reservoir can be further subdivided into pure sandstone and sandstone with shale content. The sandstone layers contaminated with shale have poorer reservoir quality, with low porosity and high acoustic impedance. Carbonate rocks within the reservoir also have different reservoir qualities depending on calcite or dolomite being dominant, and the amount of shale content.

The electrofacies classes in this study give information about the lithology of each class, reservoir quality and also the specific type of porosity. Thus, the prior study was carried out, and electrofacies obtained from calibrating the well-logs and core and petrophysical data can be divided into six subfacies which contain information about the size of the grains, rock types, lithology, porosity type, matrix, depositional texture, reservoir quality and the specific wireline

log values (Figure 3-4). Then, electrofacies 3 and 6 with low distribution in the studied wells were combined with facies 2 and 5, respectively.

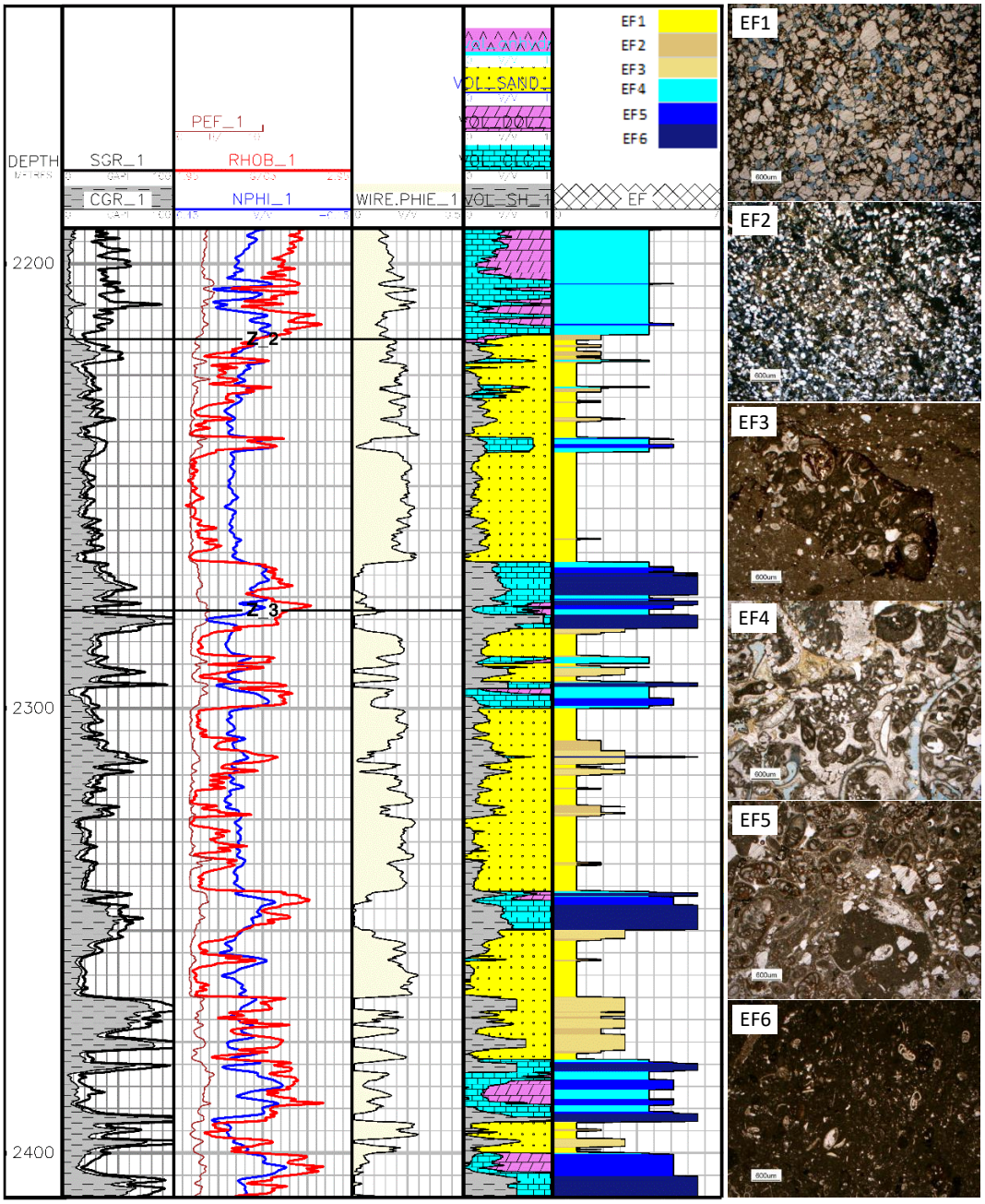

Figure 3-4: Well-log tracks for Asmari reservoir interval. From left to right: track 1: depth; track 2: spectral gamma ray and compensated gamma ray; track 3: photoelectric factor (PEF), bulk density (RHOB), and neutron porosity (NPHI); track 4: effective porosity (PHIE); track 5: estimates of volumetric concentrations of shale, calcite, dolomite, sand and anhydrite; track 6: electrofacies rock types (legend); track 7: thin section photomicrographs showing the main pore types and rock fabrics in each reservoir electrofacies

3.2 Seismic data

The 3-D seismic reflection dataset was provided by the National Iranian Oil Company. The dataset also included a set of well logs. The seismic acquisition was performed in 2005, with the main goal of identifying the fractures within reservoir, estimating its quality, and recommending targets for future drilling. Seismic acquisition parameters of this survey are presented in Table 3.1. Two source types (Vibroseis and dynamite) were used depending on the topography and seismic operation challenges. After processing the seismic data, a 3-D stacked seismic volume with 2040 inlines and 550 cross-lines was obtained.

Figure 3-5 shows an inline seismic section with the interpreted horizons. As you can see the structure of this oil field is gentle with asymmetric anticline.

Table 3.1: Seismic acquisition parameters.

PATCH GEOMETRY	
Receiver group interval	50 m
Receiver line interval	400 m
Source line interval	500 m
Number of channels per receiver line	180
Number of eecording lines	10
In-line fold	9.0
Cross-line fold	5.0
Total fold	45
Bin size	25m x 25m
Full fold area (approx.)	421
Steepest dips	10.0°
RECORDING PARAMETERS	
Number of seismic channels	1800 (Full spread)
Record length	7.0
Total number of shots	26368
Sampling interval	2 ms
Format	SEG-D
Low cut filter	Out
Anti alias filter	200Hz (0.8 Nyquist) Minimum Phase – Dynamite Zero Phase - Vibroseis

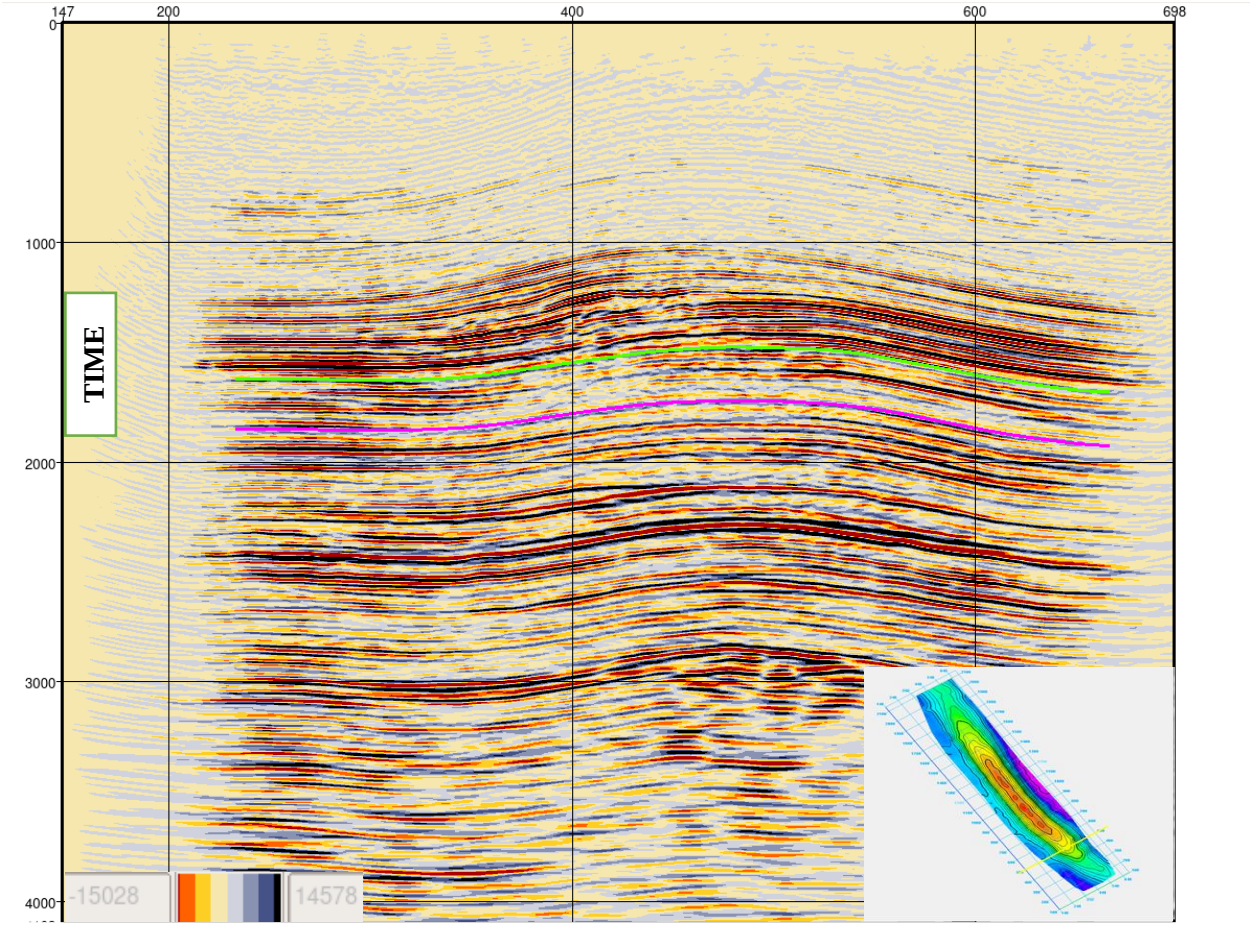

Figure 3-5: Interpreted seismic inline 500 of Mansuri oil field. At the top of Asmari formation is Gachsaran formation which is the most important cap rock of hydrocarbon reservoirs in Iran. The Asmari formation (As to Pb) is our studied reservoir is shown around 1500 ms to 2000 ms in the inset.

3.3 Previous results and problems to address

In the previous study by (Zahmatekesh et al., 2021), I used supervised and unsupervised machine learning techniques to classify the Asmari reservoir into the four electrofacies classes EF1 to EF4 described above. I used seismic attributes, the principal-component analysis and a supervised ANN learning algorithm in commercial software Stratimagic by Emerson E&P Software, which yielded a high validation score of 68% in the quality control step. The result of this classification applied on the selected seismic line which goes through 9 wells, is shown in

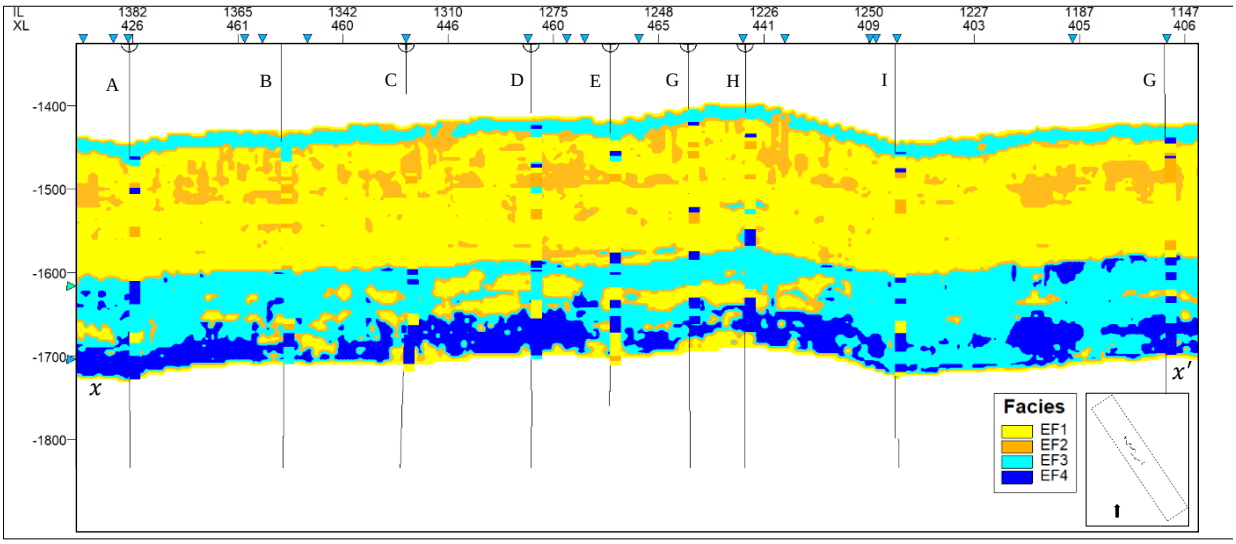

Figure 3-6: Reservoir classification by Zahmatekesh et al., 2021) along a line passing through the boreholes. Well logs facies are inserted at the well locations in order to compare the actual well logs facies with the estimated facies adjacent to the well locations. Wells "A" and "G" are blind wells used for testing the accuracy of classification.

Figure 3-6Figure 3-6. Wells “A” and “G” were blind wells, and they could be conceived as future wells in this oilfield which reasonably match the classification results in selected seismic section. These two wells were not used in the training step.

Despite the high validation score and the classification generally perceived as successful, the results in Figure 3-6 Figure 3-6 still contain several issues of concern and questions:

1) Lack of lateral continuity in the classification section, which makes it difficult to correlate it with the geological features within the reservoir. Usually, more continuous geological layers would be expected in an area with such gentle variation of stratigraphic layering (Figure 3-5).
2) I would like to investigate whether using additional seismic attributes could improve the classification.
3) At the same time, when using multiple input attributes, they must be carefully selected and preconditioned before training an ANN. The noisy appearance of the classification in Figure 3-6Figure 3-6 is likely due to the contributions from noise input attributes. In a previous study (Zahmatekesh et al., 2021), I used the conventional methods for attribute selection based on statistical data properties, such as covariances and

principal components. However, it appears that this procedure could be improved by making the selection of the attributes based not only on data but also on the desired classification.

4) Different parts of the image are classified with variable levels of statistical confidence. For example, in areas where identification of class EF4 appears patchy (noisy), the probability of this class may be insignificantly greater than those of other classes. However, confidence levels are not visible in the conventional classification section in Figure 3-6Figure 3-6. It is important to develop a visualization and interpretation technique in which the confidence level is displayed and utilized.
5) In addition to classification into discrete classes EF1 to EF4, it is useful to estimate other important reservoir properties such as porosity.

In this thesis, I address questions 2) to 5) above for the Mansuri oil field. These questions lie within the realm of machine learning. Question 1) appears to be more complex and related to the sensitivity of the attributes to the noise inherent in seismic records. However, as shown in chapter 5, by taking into account the confidence levels (i.e., solving question 4) above), the continuity of the images is greatly improved.

CHAPTER 4: Acoustic Impedance and Porosity

Evaluating petrophysical properties from seismic data and well logs is vital in hydrocarbon exploration. Well sampling is detailed and reliable, and one can utilize a vast range of information such as core data, geophysical logs, production data, and geological studies for developing the reservoir models. However, these models are limited to well locations and depths. By contrast, 3-D seismic data cannot be interpreted directly in terms of geological and petrophysical properties, but they can characterize the lateral reservoir continuity away from the wellbore (e.g., Yilmaz, 2001). Therefore, it is important to develop methods for relating the characteristics of seismic signals to the geological and petrophysical properties, and also to their variability. This task of producing effective well-log columns from reflection seismic records is generally known as prediction of pseudo-logs (Yilmaz, 2001). The ability to predict pseudo-logs is particularly crucial when there is an insufficient borehole coverage or when studying an extensive reservoir with relatively weak lateral variation.

A direct pseudo-log prediction approach using seismic attributes and artificial neural networks (ANN) was introduced commercially by Hampson et al. (2001). By using the flexibility of ANN training, adaptive nonlinear relations between seismic records and geophysical logs were developed at the well locations and applied to the entire 3-D volumes.

In this chapter, I describe applications of ANN and other inverse methods (chapter 2) to deriving 3-D volumes of the acoustic impedance (AI) and porosity for the Asmari reservoir. These inversions are obtained from seismic and well-log data using mathematical transformations and machine learning procedures. Thus, as in other reservoir characterization studies, I try combining a broad variety of data near the wells to derive the reservoir properties of interest in inter-well locations.

In section 4.1, I present the AI inversion. The AI contains the most valuable information about the subsurface geology, and this property is further used for lithology classification in this thesis (chapter 4). The AI (product of density and acoustic wave velocity) shows the geologic layers much clearer and more accurately than the original seismic data, and it is closely related to porosity, rock density, and hydrocarbon saturation. After obtaining the AI volume, I use it together

with other seismic attributes to further extract a 3-D porosity volume (section 4.2). In chapter 5, the obtained AI and porosity will be again combined with other seismic attributes to perform a classification of electrofacies within the reservoir.

4.1 Inversion for acoustic impedance

I tried several post-stack AI inversion methodologies in the Hampson-Russell software, and the method called the model-based inversion appeared to be the best. This method appears to be the most popular in the industry and is commonly used for porosity and permeability estimations, and also in geostatistical studies (e.g., Din and Hongbing, 2019; Kushwaha et al., 2020). Principles of the model-based AI inversion method are described in chapter 2. As most nonlinear AI inversion approaches, this method contains three key components: 1) inverting for the source wavelet, 2) building an initial model, and 3) iterative modifications of this model to achieve fitting the seismic reflectivity data.

The source wavelet was estimated by fitting the spectra of reflectivity synthetics to the data between the top and bottom of the reservoir in the vicinities of the selected wells. Because the phase spectrum of the wavelet cannot be accurately determined, a minimum-, zero-, or constant-phase wavelets are commonly used. In this inversion, I used a 120-ms length and a constant -7° phase of the wavelet (Figure 4-1). This selection produced a wavelet with little reverberations and spectral bandwidth from about 6 to 45 Hz (Figure 4-1).

Ideally, inverse methods should be independent of starting and ‘prior’ models. However, this is not the case for many problems in geophysics, and in particular to the inversion of seismic reflectivities for AI. Because seismic data lack low-frequency information, AI inversion methods are often sensitive to initial models and algorithm parameters (chapter 2). Therefore, for the remaining steps 2) and 3), I tried several approaches available in the Hampson-Russell STRATA software and selected solutions providing the best accuracy of the inverse. In the following subsections, I briefly summarize these two steps of model-based AI inversion.

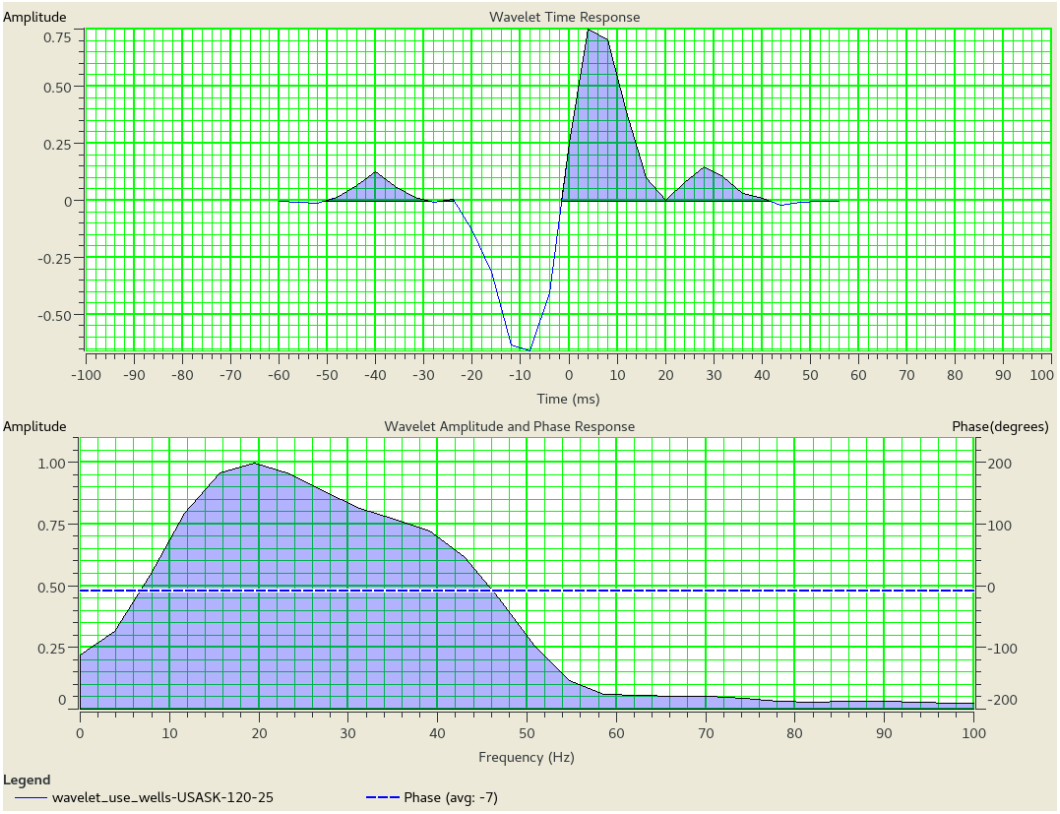

Figure 4-1: Wavelet inverted from seismic data in the reservoir interval near the 12 selected wells. Upper panel: wavelet shape, lower panel: amplitude spectrum (blue shading) and phase (dashed line, constant –7°).

4.1.1 Initial model

In model-based AI inversion, the initial model is obtained by interpolating the low-frequency component of the well logs (Russell and Hampson, 2006). For the present dataset, the best results were obtained by using the inverse-distance power spatial interpolation method. This well interpolation method weighs the contribution of each well inversely to its distance from the point being calculated (chapter 2). This procedure is controlled by a parameter called acceptance variance, which is the maximum smoothing variance (parameter σ in the method, chapter 2) which can be used by the mapping procedure. If this acceptance variance is too low, a circular 'bull's eye' pattern is often produced around each well and fades away with distance from the well location. A higher value of the acceptance variance smoothens the map and reduces the possibility of bull's eye artifacts. However, selection of a larger acceptance variance is also undesirable

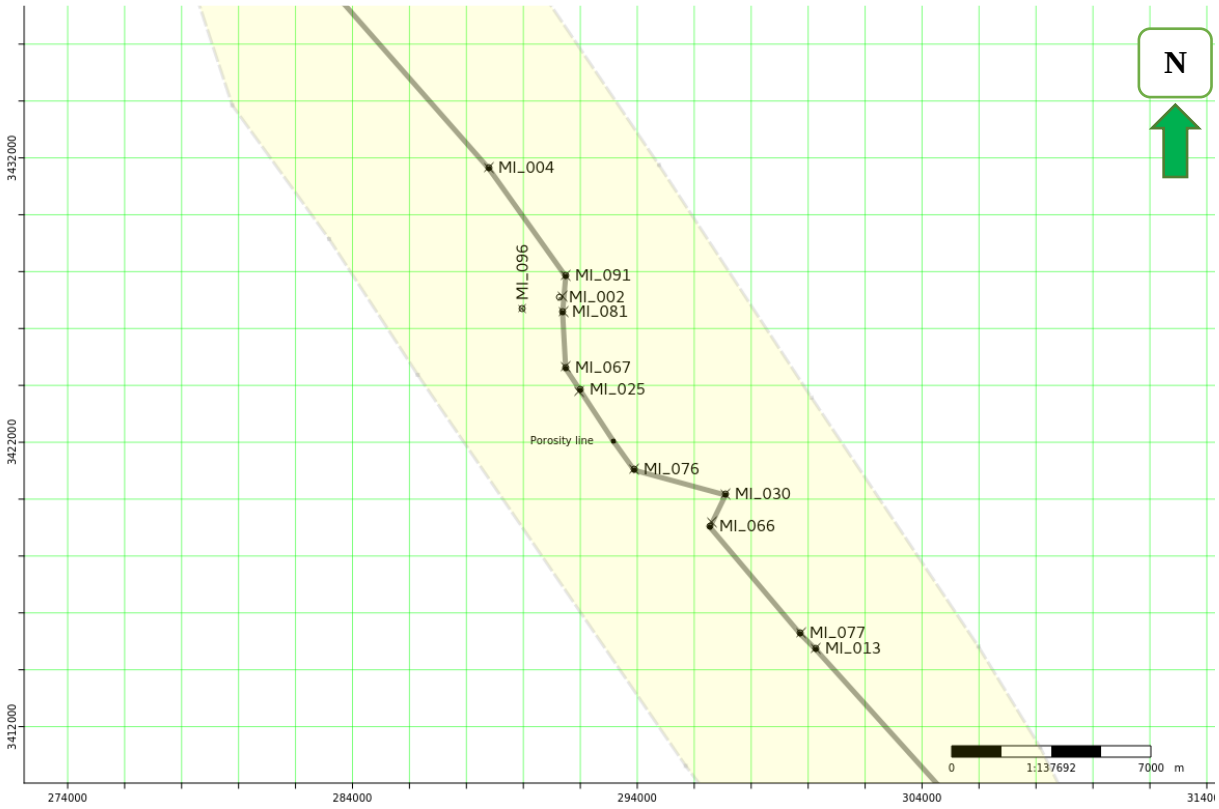

Figure 4-2: Seismic survey area (gray) and a line going through all 12 used for displaying the subsequent depth cross-sections of AI and porosity.

because it does not allow fitting the well data sufficiently closely. To prevent the bull's-eye artifact while achieving reasonable accuracy, I selected the acceptance variance in a well tie to equal 15%.

Figure 4-2 shows a map of the survey area with a custom (arbitrary) line drawn through the locations of all wells. This line will be used for presenting the various cross-sections from the 3-D volumes of the various models and seismic attributes. The cross-section of the 3-D initial model along the selected line (Figure 4-2) is shown in Figure 4-3. As shown in this figure, the AI model is created between two horizons (Asmari and Pabdeh) which are top and bottom of the reservoir. Above and below these horizons, the AI is extended by constant values. Eight zones within the reservoir (chapter 3) are labeled from Z-1 (the top of Asmari formation) to Pb (the top of Pabdeh formation). The model is close to well-log observations at all well locations. No depth

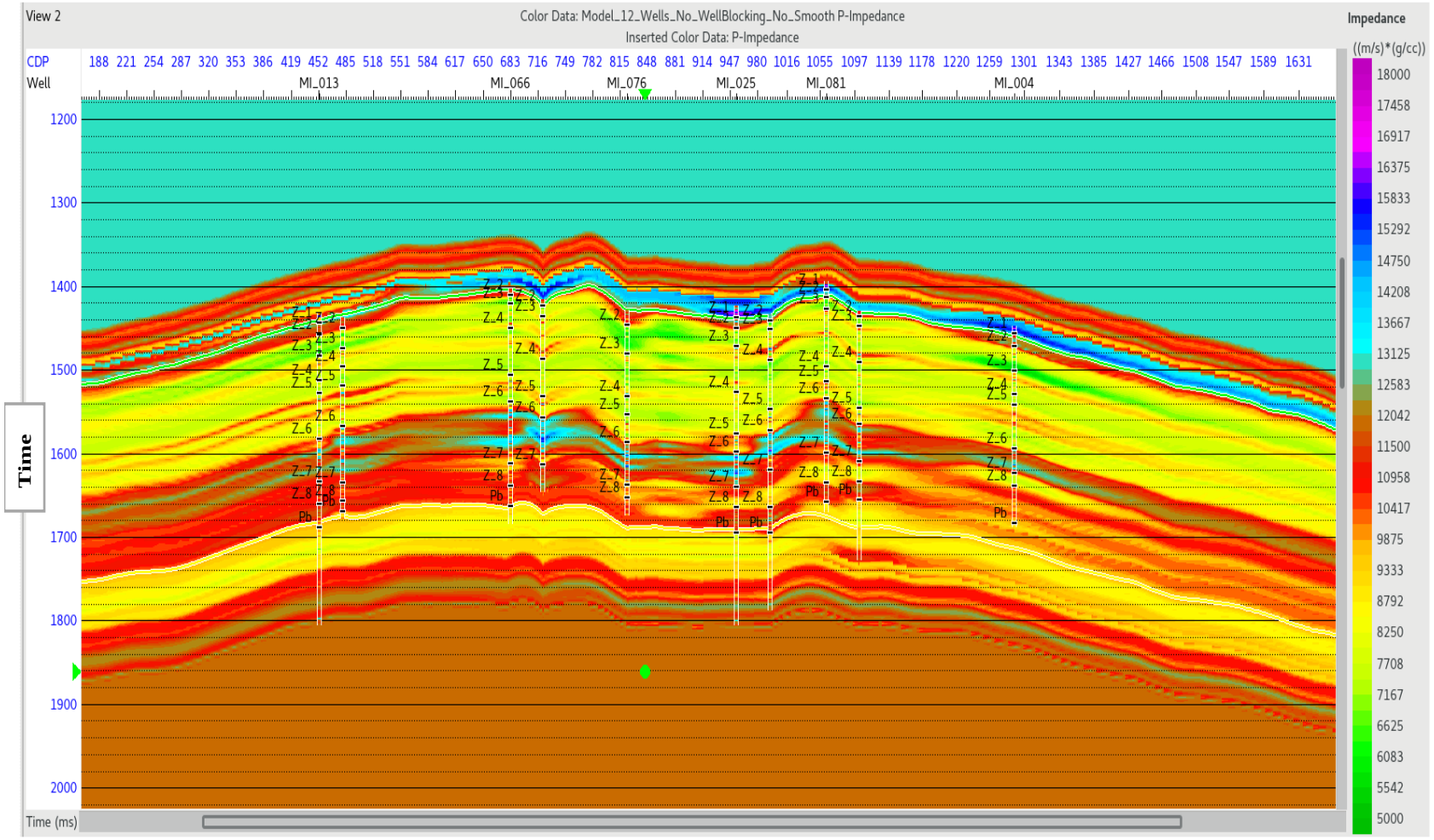

Figure 4-3: Initial AI model (color) and acoustic well logs (insets) along the line in Figure 4-2. Labels indicate the reservoir zones defined in chapter 3.

smoothing filter was used to create this initial model, and the layering in this model appears to be geologically reasonable and structurally close to the reality.

4.1.2 Acoustic impedance estimation

The next step of AI estimation consists in using the stacked seismic records to modify the initial model in Figure 4-3, so that in addition to the low-frequency velocity variation in the interpolated logs, it starts predicting the reflectivity observed in the seismic data. The model-based AI inversion (chapter 2) is known as the generalized linear inversion (GLI) algorithm, which attempts to adjust the model until the synthetic record derived from the AI model matches the seismic trace. Using the previously estimated source wavelet and an updated impedance, a synthetic seismogram is created and compared to the original seismic data in each seismic trace.

As described in chapter 2, the model-based inversion contains user-specified control parameters which determine how far the final model is allowed to deviate from the initial model and how closely it should match the seismic data. These control parameters represent constraints in the inversion, and their selection significantly affects the results. To select the control parameters and evaluate the performance of inversion, I compare the inverted AI near well locations to AI values from the well logs (Figure 4-4). As this figure shows, the correlation between the inverted AI and the original AI (from the well log nearby) equals 0.86, and the normalized error is 0.06, which shows a good performance with this method. The slope of regression for the inverted seismic AI with respect to the well-log AI equals 0.98. This value close to 1.0 shows that the model-based inverse correctly estimates the average high-frequency AI variation near the wells.

In another standard quality control test, I compare the synthetic seismograms modeled at well locations with waveform synthetics calculated from the reflectivities in the well logs (Figure 4-5). This correlation between synthetic seismograms from the inversion and seismograms modeled from well logs gives quality control to confirm the inversion process (Austin et al., 2018). In Figure 4-5 inversion analysis is illustrated for one of the wells (number 066).

As shown in Figure 4-5, the predicted seismic records (panel c) are close to the observed data (panel d), and their difference is small (panel e). The correlation measure between these time

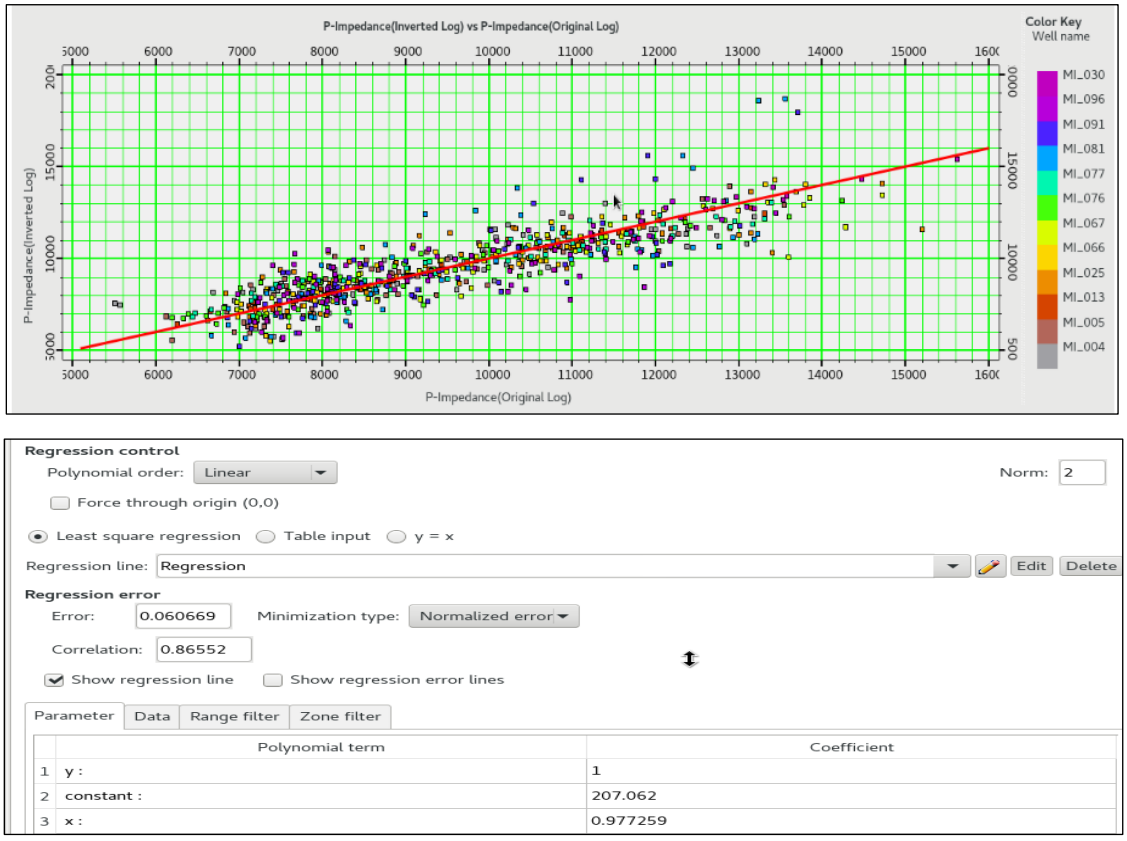

Figure 4-4: Inversion analysis cross plot between original and inverted acoustic impedance from STRATA software (top) and parameters of the correlation (red line) between the true and inverted AI (bottom).

series (labeled 'Correlation' in Figure 4-5) is the square of the correlation coefficient, R^2 (chapter 2), which is often called the goodness of fit. This measure equals 0.98 (and therefore $R \approx 0.99$), which shows a very high accuracy of predicting the reflection seismic signal. A high accuracy of predicting the reflectivity is typical for AI inversion, because this inverse problem is highly under-determined, and accurate data fitting is expected (Menke, 1984). The difficulty of under-determined inverses is in achieving stability and accuracy of AI models (Morozov and Ma, 2009), and the 0.86 correlation with independent well-log data (Figure 4-4, bottom panel) is a good indicator of inversion accuracy. This quantity was used as the primary quality criterion for selecting parameters of the algorithm.

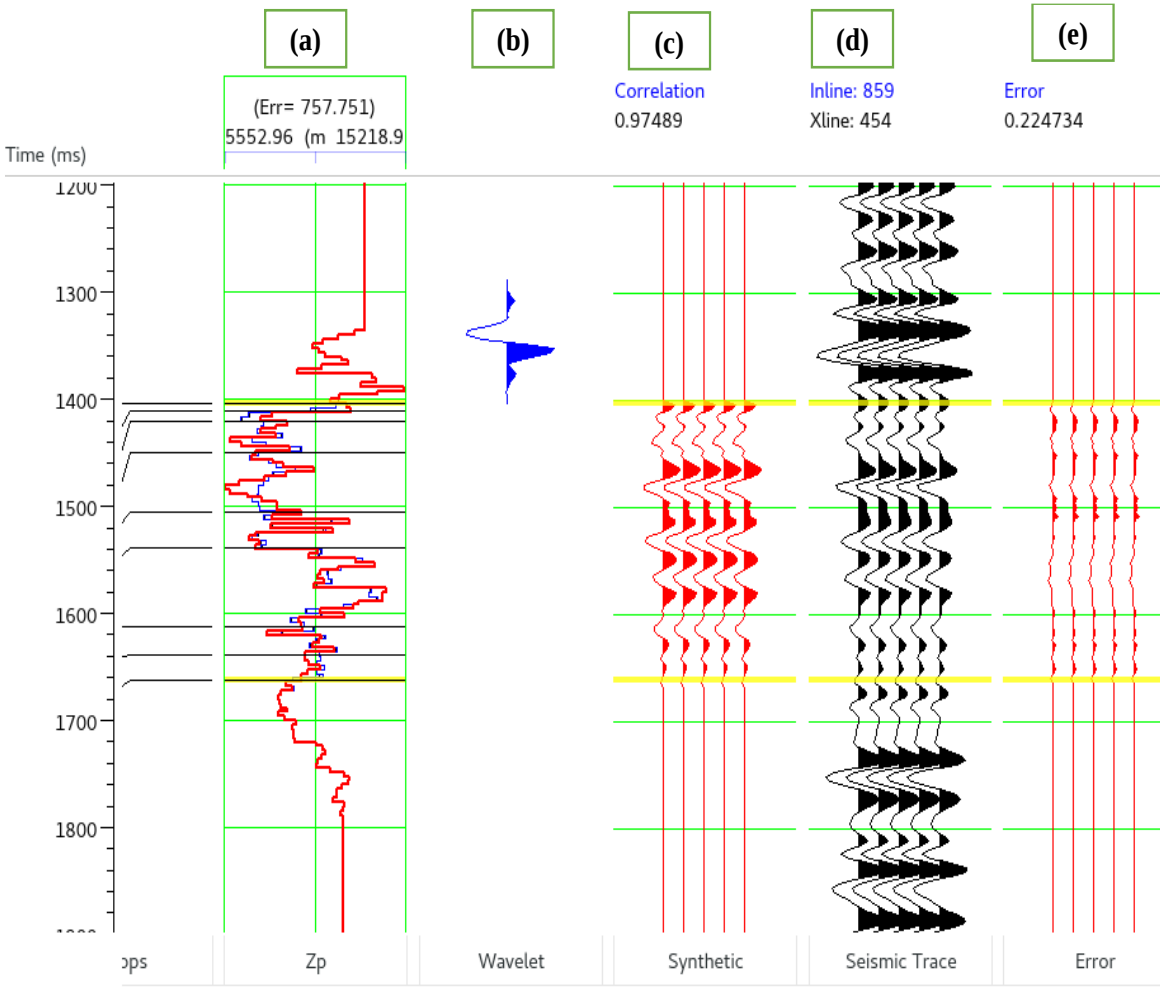

Figure 4-5: Analysis of the inversion results at well 066: a) The true AI at well location (blue) and the inverted AI (red); b) source wavelet; c) synthetic seismograms from the inversion (estimated reflectivity convolved with the wavelet); d) seismic records near the borehole; e) their error (difference between d) and c)).

The inverted AI along the selected line is shown in Figure 4-6, with AI columns calculated from well logs inserted for comparison. Systematic variations of the AI with depth and lateral coherence demonstrate a layering pattern and indicate different lithofacies within the reservoir interval (Eze et al., 2019). As in the initial model (Figure 4-3), there is a strong correlation between the AI values and well-log impedances at well locations. Between the wells, the inverted AI reveals the detail of geological layering within the reservoir.

The AI and estimated porosity are restricted to the reservoir because a common well-log interval is required to perform the inversion or machine learning techniques. Generally, the lower

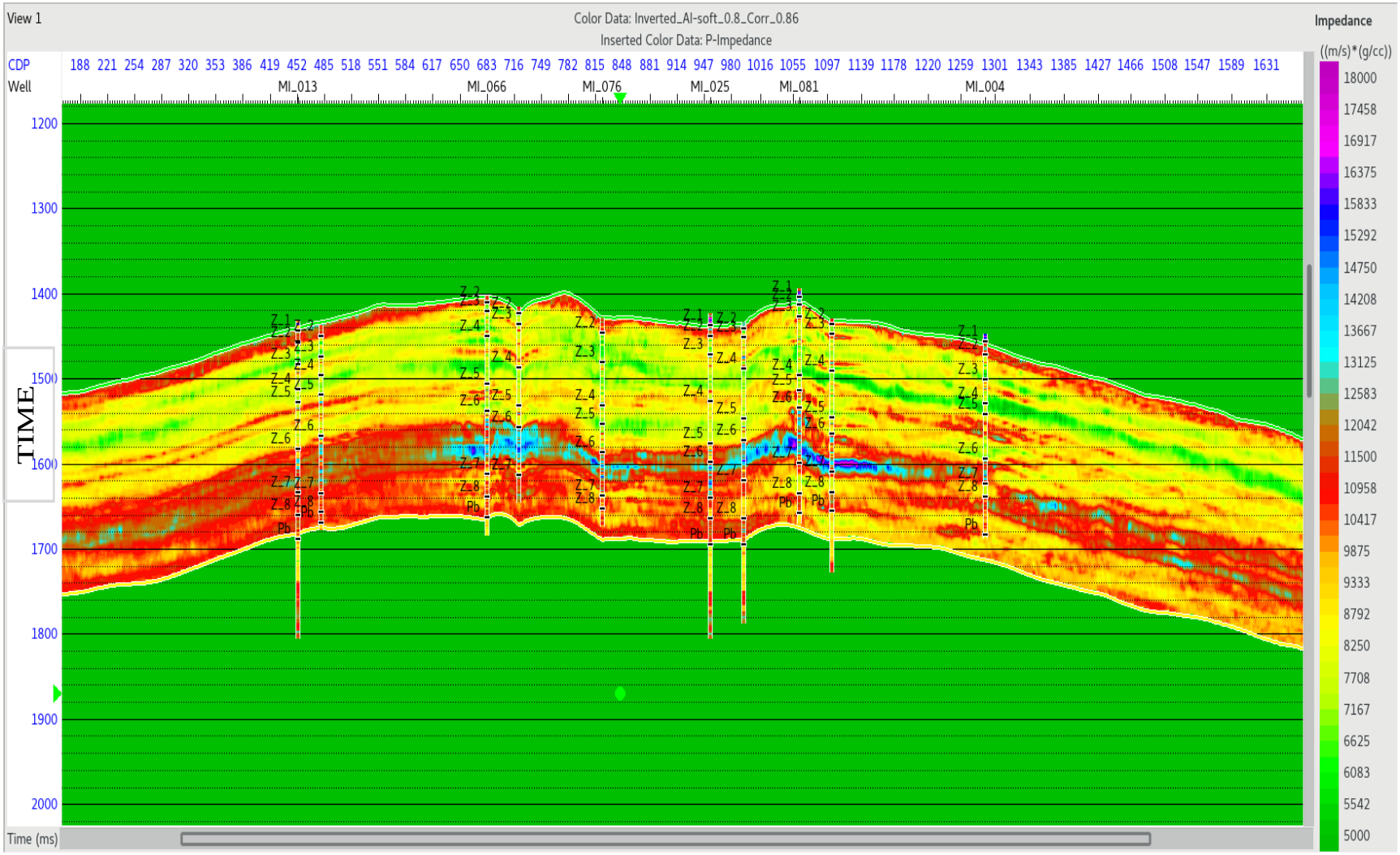

Figure 4-6: Cross-section through the inverted AI cube using model-based inversion method. Insets show the AI columns from well logs, and labels indicate the reservoir zones.

part of the estimated AI volume shows higher values compared to the upper part of the reservoir. Also, at the very top of the reservoir, there is a thin layer of higher acoustic impedances compared to the other part of the first half of the reservoir. This pattern is also expected to be seen in the porosity volume estimated in section 4.2.

4.2 Porosity estimation

To estimate a porosity volume from seismic data, I use the estimated acoustic impedance (AI) in combination with additional seismic attributes using the GRNN (PNN) algorithm in Hampson-Russell STRATA software. This methodology was successfully used to estimate porosity in many previous studies (e.g., Russell et al., 1997; Din and Hongbing, 2020).

The first step in this estimation is a feasibility study establishing that the dataset contains attributes which are sufficiently sensitive to porosity. This feasibility study is performed in subsection 4.2.1, and it shows that the AI should indeed be the primary attribute for porosity estimation (Ekone et al., 2020).

Porosity inversion using GRNN training involves the solution of a nonlinear inverse problem with many unknowns (chapter 2). To obtain stable and geologically meaningful results, it is important to avoid attributes which are noisy or do not contribute to the final estimate. In subsection 4.2.2, I describe the procedure for this selection of optimal attributes. Once the GRNN algorithm is trained, it can be applied to any location within the seismic cube, producing a 3-D volume of reservoir porosity. This porosity volume for Mansuri reservoir is derived in subsection 4.2.3.

4.2.1 Feasibility study

Prior to designing an ANN structure and performing training, it is useful to conduct a feasibility study, which is a series of tests showing whether the expected correlation between the input features and output values (porosity in this case) exist. To perform the feasibility study, the AI and true porosity values derived from the well logs are cross-plotted in Figure 4-7.As this plot shows, there is a significant anti-correlation between the AI and porosity, which indicates a

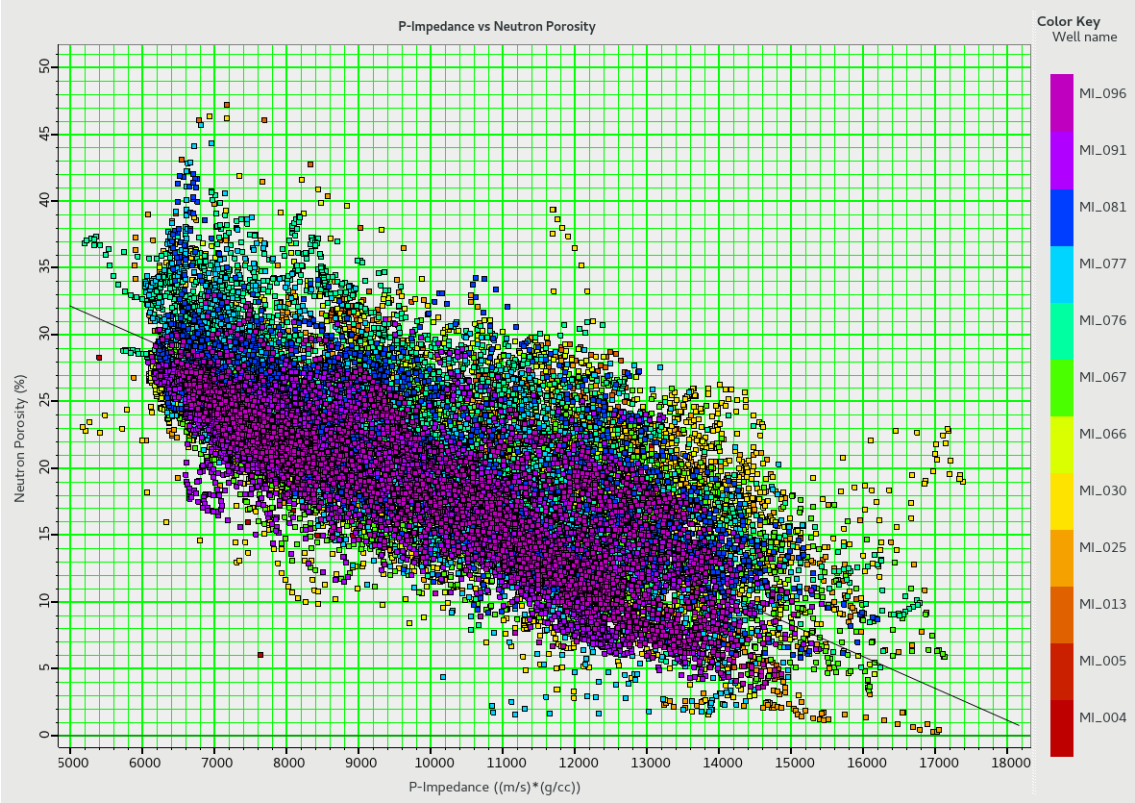

Figure 4-7: Cross plot of acoustic impedance (AI) and porosity in several selected wells (color bar). The *X*-axis is the acoustic impedance and *Y*-axis is porosity. Line indicates the least-squares fitting of porosity as function of the AI.

meaningful relation between these physical properties of geological layers. This relation is principally due to the rock density ρ decreasing with porosity for a fixed rock type, and the AI being proportional to density as $Z = \rho V$. The correlation coefficient between porosity and the AI is 0.73 and the normalized regression error is 0.07 using linear regression for 12 wells (Figure 4-7). Because of this significant linear trend, the AI should be included into estimation of porosity from seismic data.

Along with the general negative linear correlation between the porosity and AI, there still is a significant scatter of porosity values from this trend (Figure 4-7). This scatter may be due to multiple other factors, such as depth (deposition age) and layering. To investigate the porosity variation with depth, a cross-plot of porosity versus depth is shown in Figure 4-8. Across the entire reservoir interval, there is an about 5% drop in porosity, but it appears to occur near the middle of

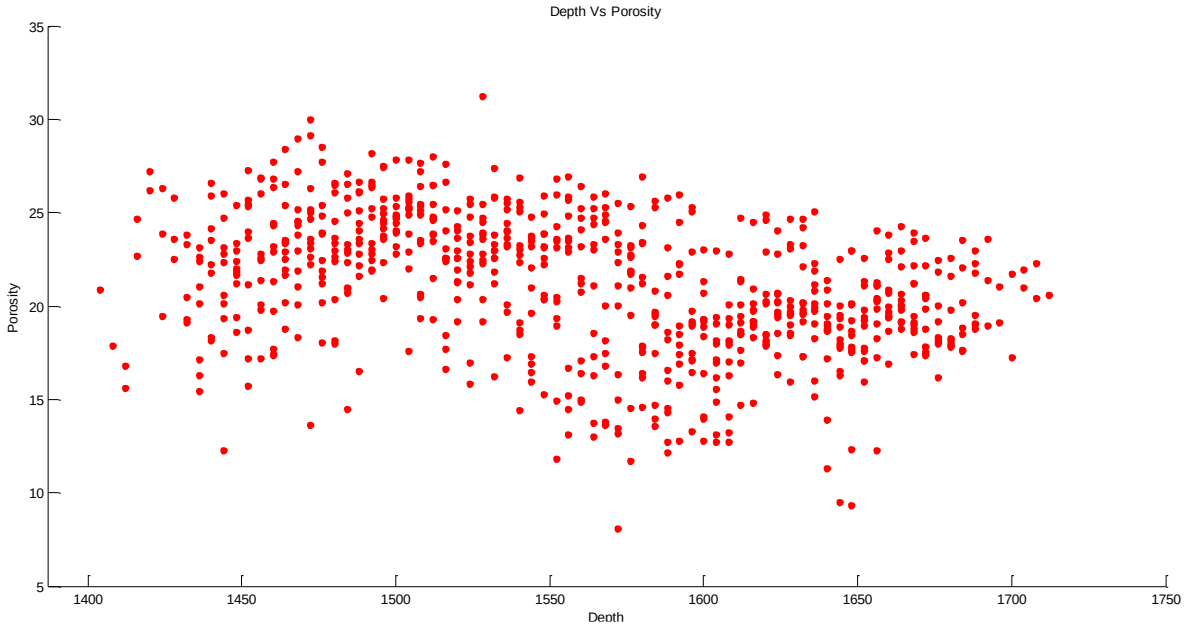

Figure 4-8: Scatter plot between porosity and depth in the reservoir.

the reservoir, due to different rock types dominating the upper and lower halves of the reservoir. This about 5% change in porosity may also contribute to the vertical offsets of the dots from the regression line in Figure 4-7. This contrast in porosity between the two halves of the reservoir will also be seen in the results from facies classification (chapter 5).

4.2.2 Selection of seismic attributes

The standard procedure for selecting input attributes for ANN training consists in analyzing the covariances of all attributes and discarding those which are valeither have too low variance or corelated with each other (Haykin, 2009). However, a better approach would be to select attributes based not only on their mutual covariances but on how they contribute to the final prediction of porosity. In Hampson-Russell software STRATA, such selection is done by a process called stepwise regression. In this process, one or several wells are removed from the dataset and used for validation, i.e. checking how the algorithm is capable of predicting known but 'blind' data. In an iterative process, the prediction (cross-validation) error is minimized to find the best attribute, then the best pair of attributes, then the best triplet, and so on. The cross-validation procedure is as follows:

1) Remove the target log and attributes for one well from the training data;
2) Calculate the multi-attribute coefficients without the removed well;
3) Apply the coefficients to the removed well;
4) Repeat steps 1) to 3) for each well in turn;
5) Evaluate and average the errors by repeating the procedure twelve times, each time treating one of the wells as the blind-predicted wells. The squared error for each test is defined as

$$E^2 = \frac{1}{N}\sum_{i=1}^{N} e_{vi}^2 , \tag{4.1}$$

where e_{vi} the validation error for the blind well, and N is the number of known points (i.e., time samples) in the analysis.

In other words, at the first step, all attributes are checked to find the one predicting the target value (porosity) best when used alone. At the next step, this first attribute is kept, and an additional attribute is scanned in the same way, to produce the best pair which yields the highest correlation with the target value in the removed wells. Then we search the best triplet of attributes keeping the previously selected pair of attributes.

In Figure 4-9, the performance of this procedure is illustrated graphically. The cross-validation error (error in predicting the porosity in the removed blind wells; red line) decreases until the number of used attributes equal nine (Figure 4-9). When adding more than nine attributes, the cross-validation error increases, showing that the additional attributes reduce the ability to predict known data. Thus, the optimal number of attributes equals nine, and they should be selected in the order of the above procedure. Using additional attributes continues to decrease the total training error (black line in Figure 4-9) but this reduction can be viewed as ‘overfitting’ since it compromises prediction in the removed cross-validation well(s).

The nine seismic attributes selected by the stepwise regression are:

1) AI (Acoustic Impedance) from model-based inversion process.

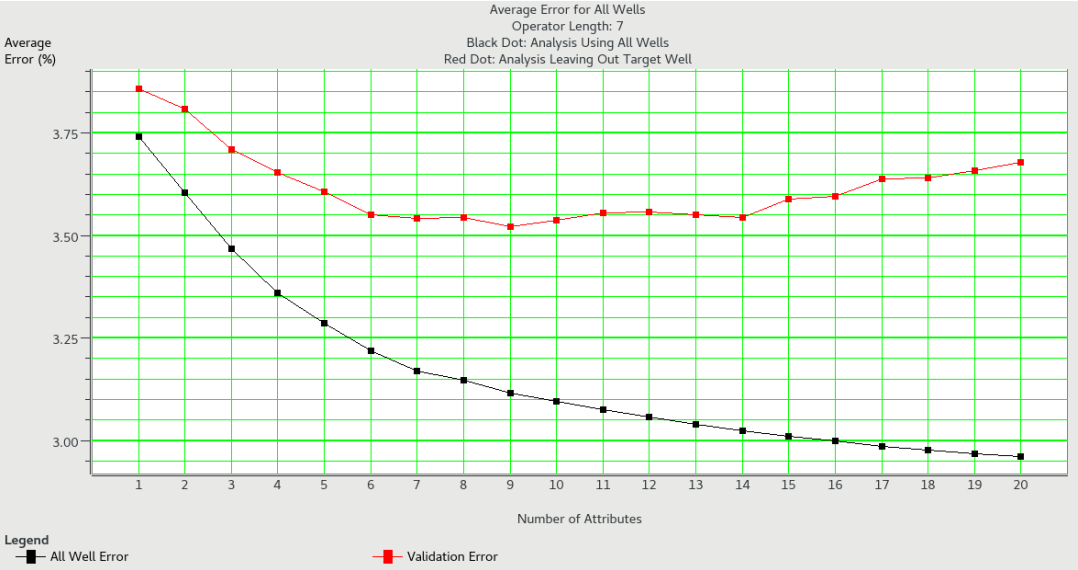

Figure 4-9: Measures of cross-validation error for different numbers of optimal input attributes used. The total training error (black) continuously decreases as the number of attributes increases, but the validation score (red) decreases till the optimal number of attributes (9 in this case) and then again increases. The validation error will always be higher than the training error since in the validation process there are always fewer data to use (in our case, fewer well to use) than in the training process.

2) Dominant frequency of the seismic signal determined from the maximum of the amplitude spectrum over a small window centered on each time sample.
3) Bandpass filtered seismic signal with the frequency range of 5/10–15/20 Hz, which means that the filter amplitude increases from zero to one between 5 and 10 Hz and tapers off from one to zero between 15 and 20 Hz.
4) "Derivative Instantaneous Amplitude", which is the derivative of the amplitude envelope of the input trace.
5) "Amplitude Weight Cosine Phase", which means the product of the amplitude envelope and the cosine of instantaneous phase.
6) Cosine of the instantaneous phase of the signal.
7) Average signed amplitude envelope.
8) Instantaneous frequency.
9) "Amplitude-Weighted Frequency", which is the product of the amplitude envelope and the instantaneous frequency.

By applying multi-linear regression using these attributes, the training correlation (cross-correlation between all data and predicted data at the well logs) is approximately 0.82, and the validation correlation (the same cross-correlation for blind wells) is about 0.75. Detailed observed and predicted porosity profiles in the training and validation phases are shown in Figure 4-10 and Figure 4-11 respectively.

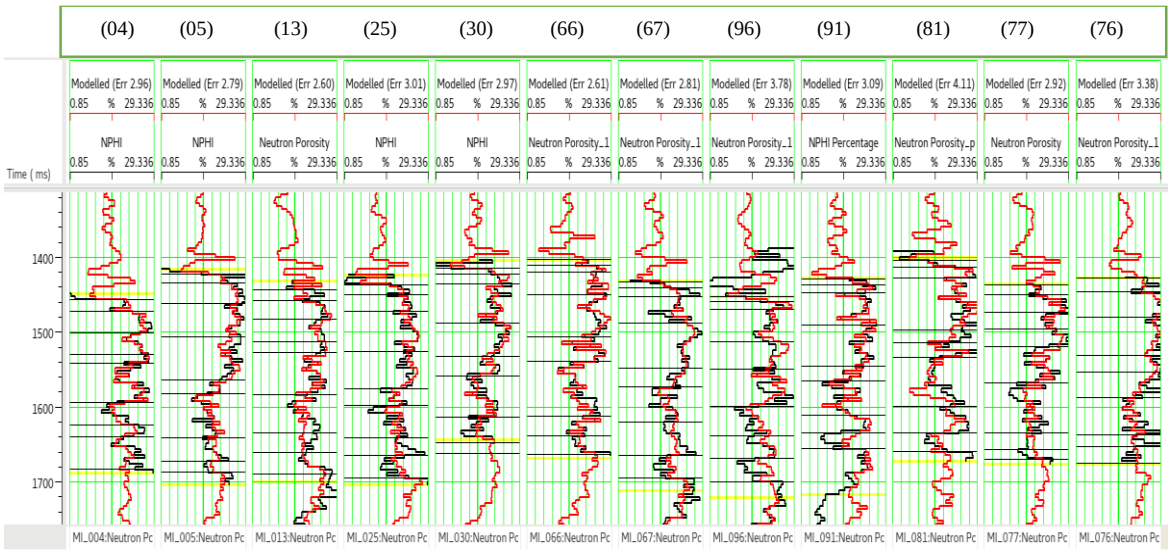

Figure 4-10: Porosity predicted in a training well (red) and the true porosity (black). Well numbers are shown in the labels (training correlation = 0.82).

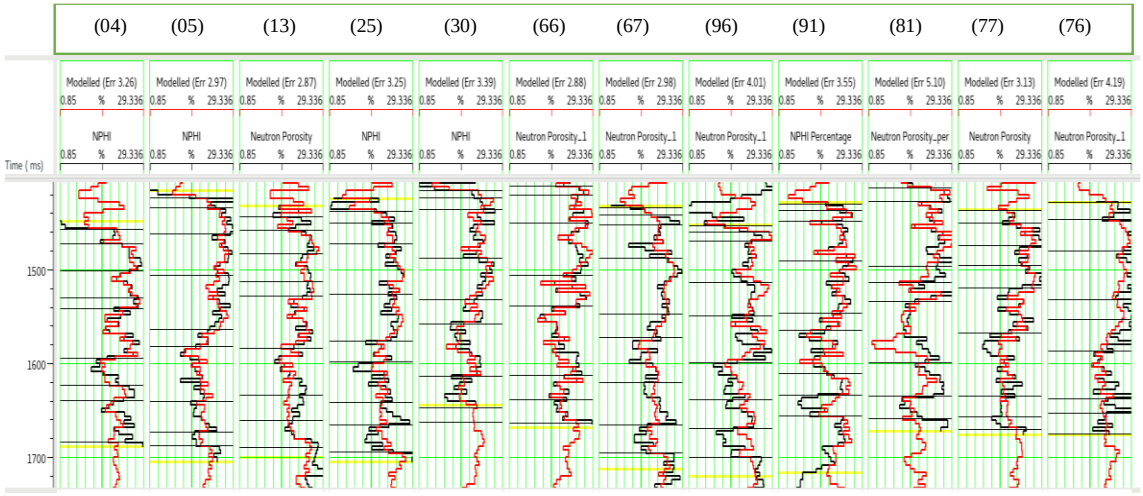

Figure 4-11: Predicted porosity (red) and true porosity (black) in a blind well (validation correlation = 0.75).

4.2.3 Estimation of porosity volume

After selecting the best subset of input seismic attributes, this subset is used to estimate the desired reservoir parameters using a multi-linear regression algorithm in Hampson-Russell software. Then, this subset is used for further machine learning techniques such as GRNN (PNN), RBFN, and MLFN (chapter 2). All these machine learning techniques were used and compared by validation on blind wells, and the GRNN was found to provide the highest accuracy. The results obtained by the GRNN are shown further in this section.

With any porosity prediction algorithm, cross-plotting of the observed and predicted porosity values is another way to check the quality of the results (e.g., Kushwaha et al., 2020). Figure 4-12 and Figure 4-13 show cross-plots of the predicted versus actual porosities, obtained by applying the multi-linear regression using the nine attributes of the preceding subsection. Figure 4-12 compares the porosity data in the training phase, and Figure 4-13 shows the results in the validation phase. From Figure 4-12, the training correlation is about 0.82. The validation

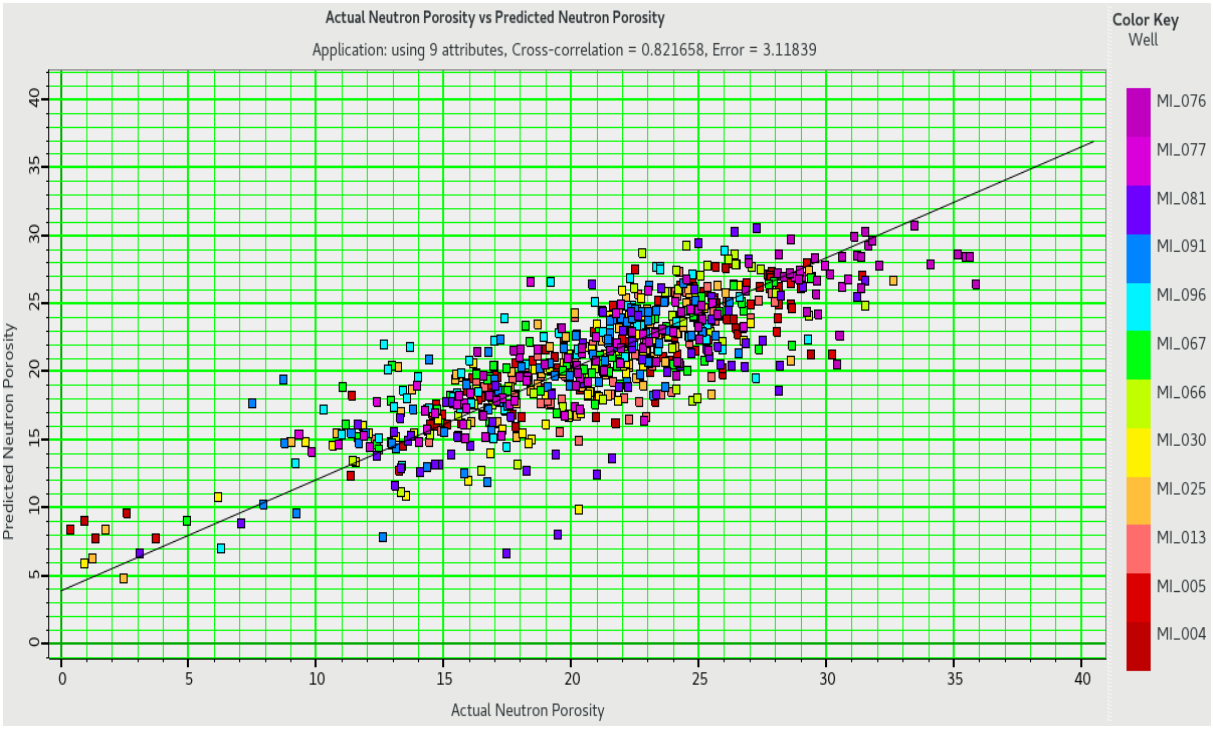

Figure 4-12: Cross-plot of the actual and predicted porosities using multi-linear regression in training step (training correlation = 0.82).

correlation is near 0.75 (Figure 4-13), which is a high accuracy for porosity estimations (e.g., Kushwaha et al., 2020).

Although the conventional quality control only checks for the correlation coefficient (0.82 in Figure 4-12), a closer look at the cross-plots also indicates a problem with this model and its solution. The regression line in Figure 4-12 corresponds to equation

$$\phi_{\text{inverted}} = 0.82\phi + 3.85. \tag{4.2}$$

Ideally, this relation is expected to be $\phi_{\text{inverted}} = \phi$, and therefore eq. (4.2) shows that the prediction contains an 18% underestimation of the porosity range and a 3.85% bias in it. These distortions are undesirable, and they can be corrected by adding an empirical transformation of the inverted porosity (inverse of eq. (4.2)).

$$\phi = 1.22\phi_{\text{inverted}} - 4.70. \tag{4.3}$$

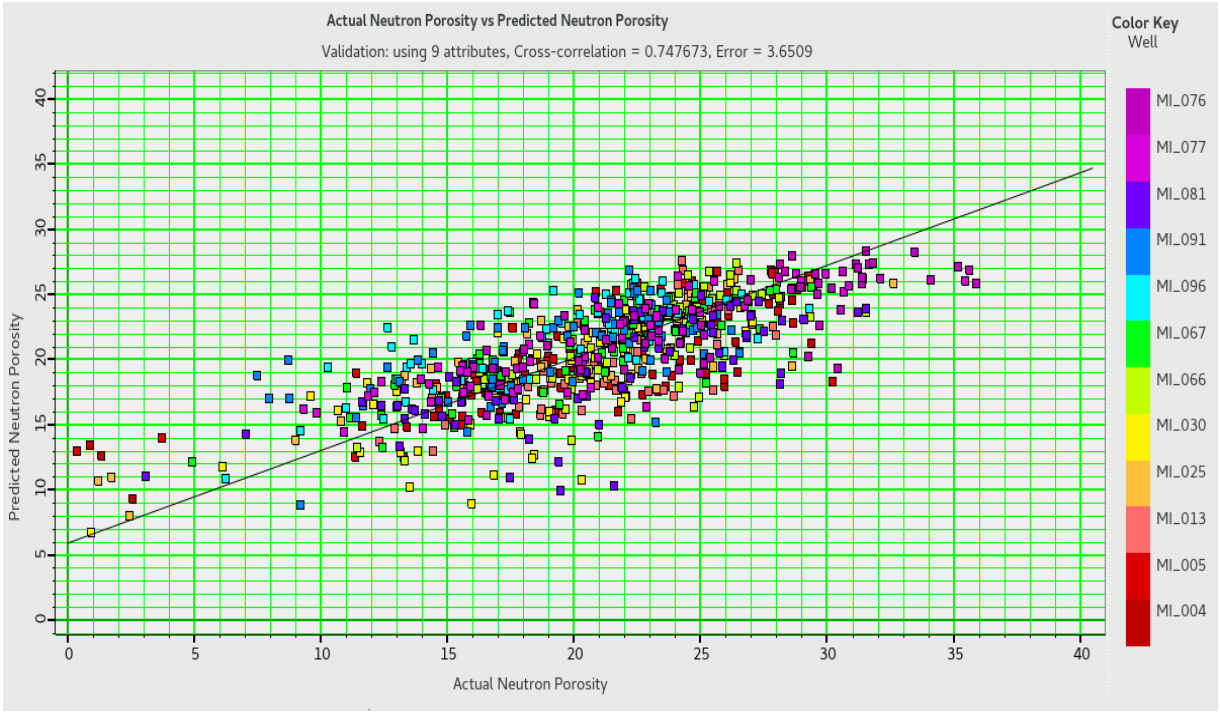

Figure 4-13: Cross-plot of the actual and predicted porosities using multi-linear regression in validation step (validation correlation = 0.75).

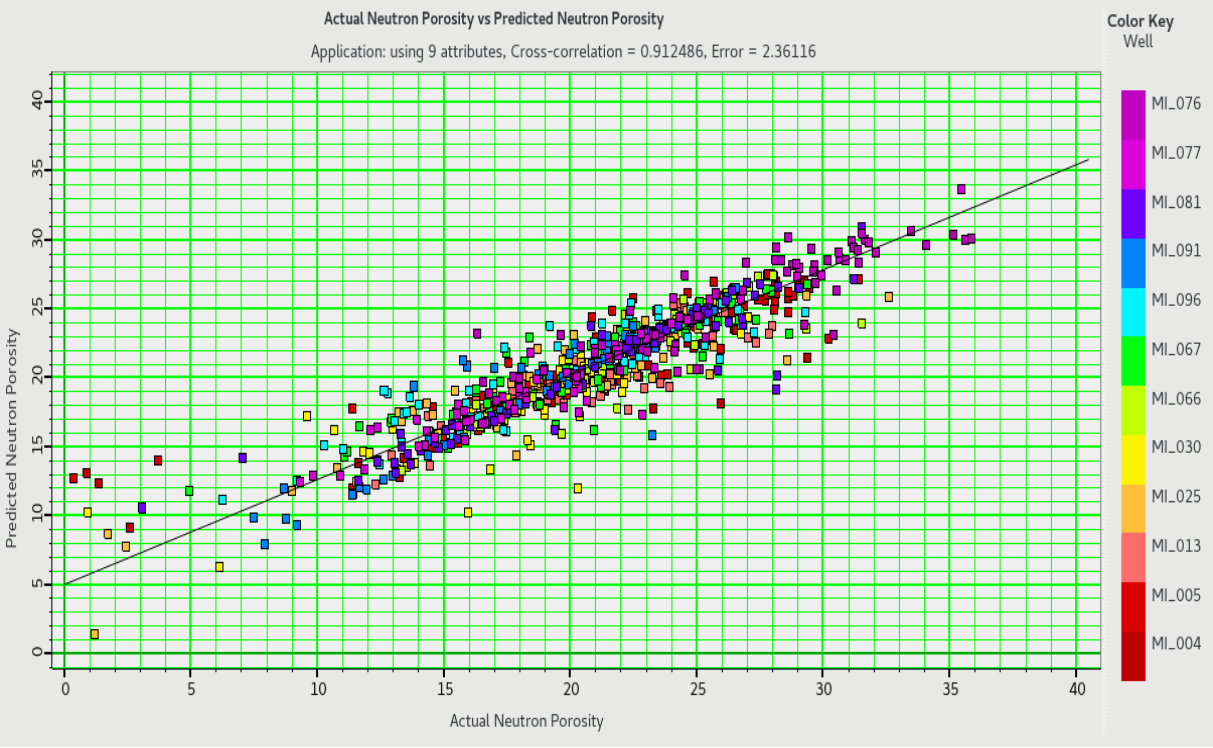

Figure 4-14: Cross-plot of the actual and predicted porosities in training step using GRNN (training correlation= 0.91). The x axis is actual porosity at well locations, and the y axis is the predicted porosity. Also, the color bar represents the well data, so the reader can find which well shows the highest correlation.

In order to verify the best method to predict the porosity, after applying the GRNN to predict the desired output, cross-plots of training and validation results are shown in Figure 4-14 and Figure 4-15 respectively. The GRNN method with 91% correlation at training and 77% correlation at the validation steps is preferred to the multi-linear regression with 82% at training and 75% at validation steps. At the same time, note that the simple multi-linear regression method yields an only 2.5% reduction of validation accuracy compared to the best (GRNN) method. This comparison shows that the limitation on the accuracy of porosity inversion principally comes from the data and can hardly be overcome by data-driven algorithms.

In addition, the GRNN algorithm yields the highest training and validation accuracy compared to other machine learning algorithms. The validation accuracy is used as the quality control to select the best algorithm to estimate porosity. In this step, 12 wells and 9 attributes are used to estimate the porosity using machine learning techniques.

The porosity section predicted by the GRNN method along the custom line is shown in Figure 4-16. This figure shows reasonable porosity values, with expected match between the

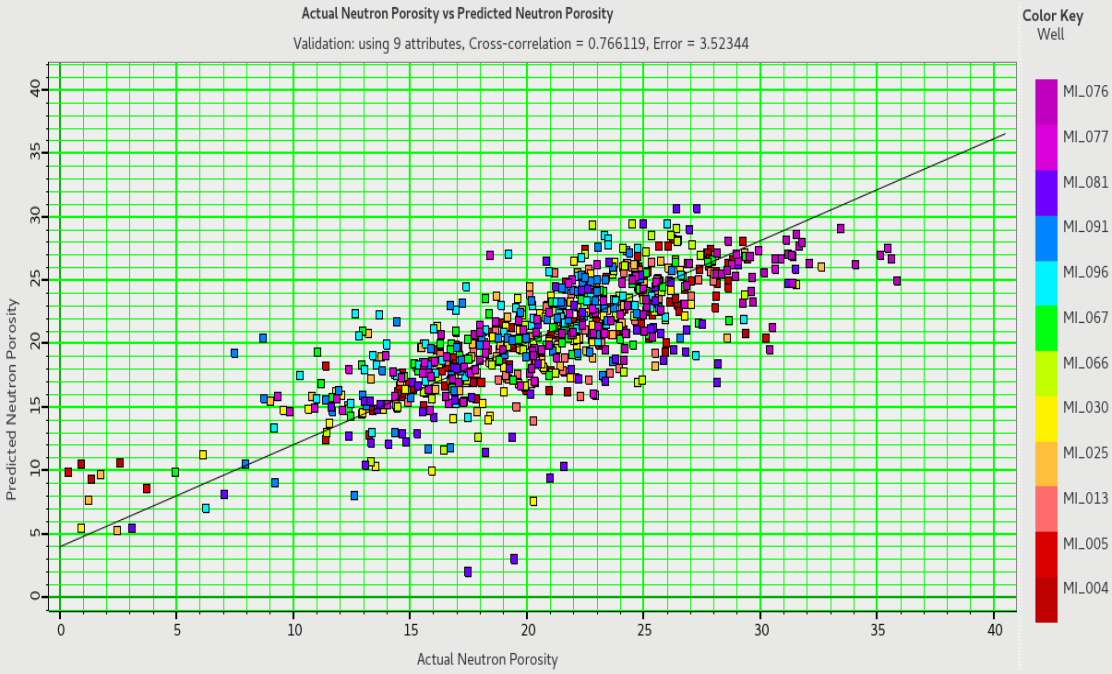

Figure 4-15: Cross-plot of the actual and predicted porosities in validation step using GRNN (validation correlation= 0.75).

estimated porosity section and porosities measured from well logs. The porosities are generally layered similarly to the AI layering, with noise seen as vertical bands at some midpoints. This noise comes from increased levels of noise in the seismic attribute sections. In chapter 5, I show details of some of these attribute sections. Clearly, since the porosity is derived purely from seismic data, it is also affected by all limitations of seismic data acquisition and processing. The quality and smoothness of seismic attributes could potentially be improved by more accurate data processing, which can be a subject of future research of this dataset (chapter 7).

Similar to the cross-plotting at well locations in subsection 4.2.1, it is useful to cross-plot the inverted porosity data versus the acoustic impedance (AI) for the whole 3-D volume or within the selected cross-section (Figure 4-17). As this cross-plot shows and as expected, the estimated porosity generally anti-correlates with the AI, with slope $-2.7 \cdot 10^{-3}$ (line in Figure 4-17) and correlation coefficient 0.76. The distribution has a generally parabolic shape as suggested by Kushwaha et al. (2020). However, this plot also suggests that several relations between porosity and the AI may actually be present in the reservoir. In particular, it appears that the upper half of the reservoir (warm colors in Figure 4-17) has lower values of porosity but steeper dependence on

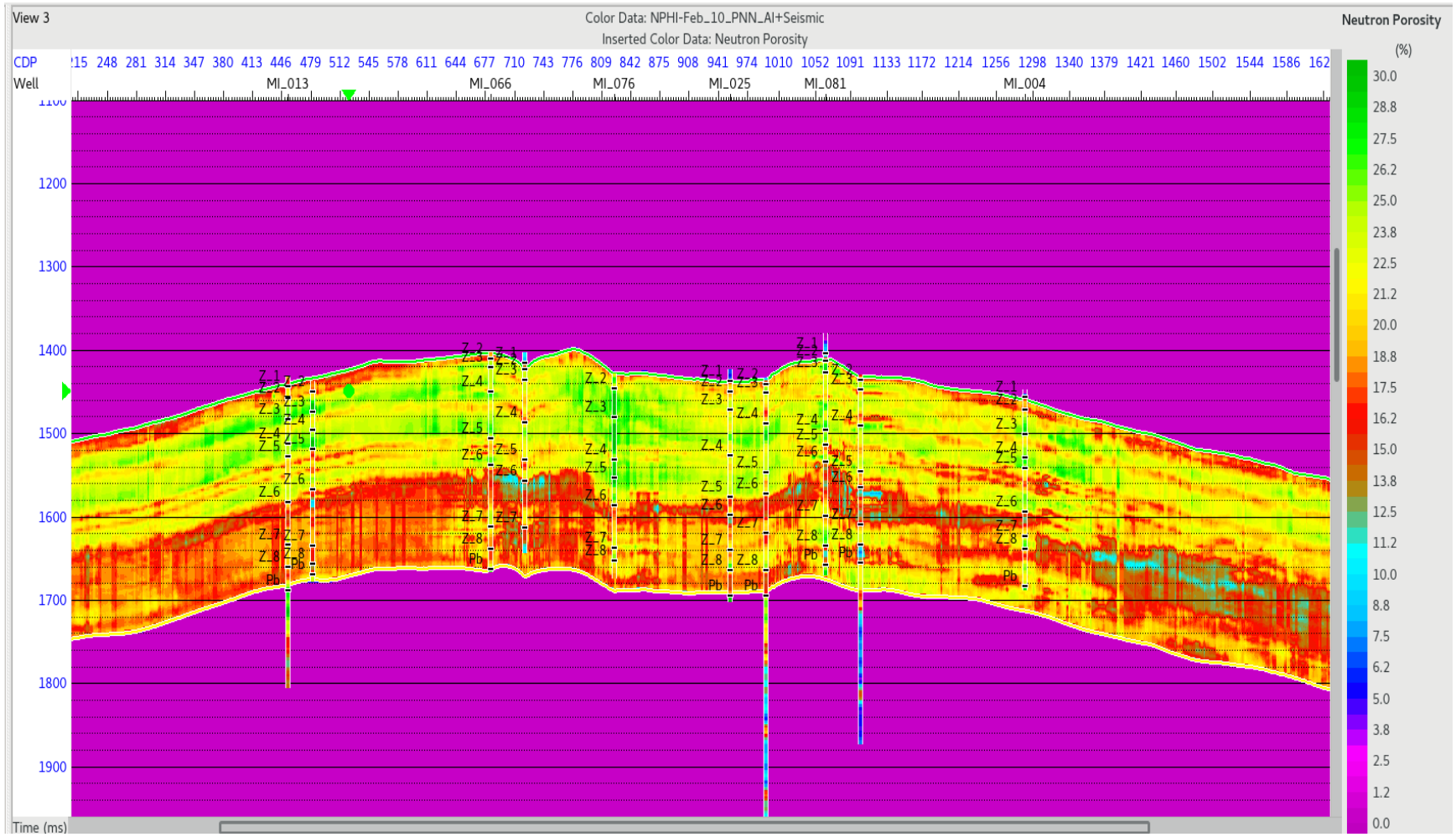

Figure 4-16: Inverted porosity cross-section along the same custom line as in Figure 4-6.

the AI. By contrast, the lower half of the reservoir (cold colors in Figure 4-17), seem to show higher porosity values but weaker dependence on the AI. In addition, these dependencies may be nonlinear, which means that some transformation of porosity ϕ, such as evaluation of $\log(\phi)$ may improve the correlation. These issues may also deserve further analysis in future studies.

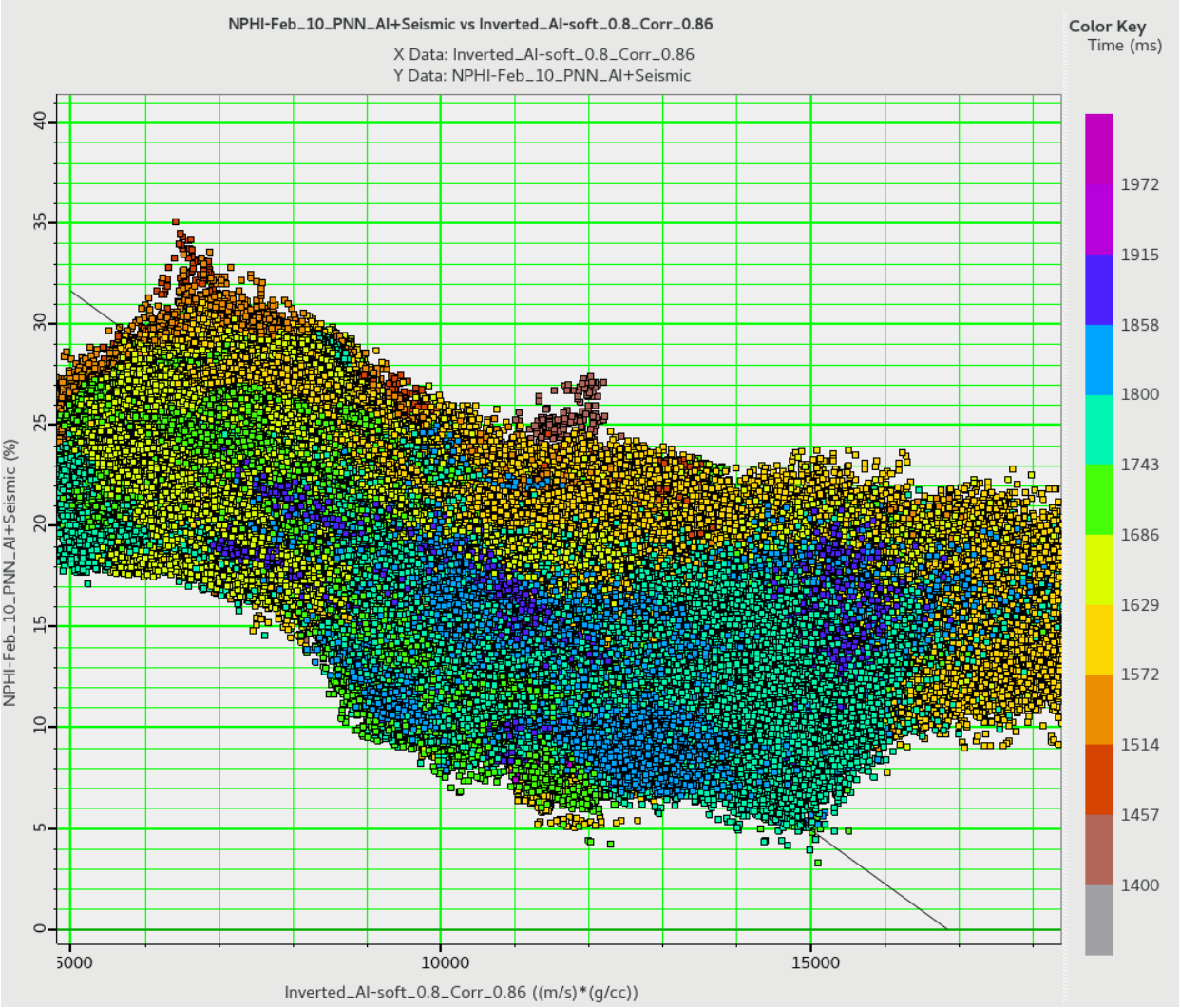

Figure 4-17: Cross-plot of the estimated porosity and inverted acoustic impedance within the cross-section along the custom line passing through all wells. Line shows a least-squares regression of porosity values with as a function of the AI.

CHAPTER 5: Classification of Seismic Facies

In this chapter, I continue the study of pseudo-log prediction in Mansuri data to classification of seismic facies. Seismic facies are defined as generally 3-D units composed of certain groups of seismic reflections, which are different from adjacent groups. Geologically, seismic facies represent sedimentary facies and/or certain structural and stratigraphic characteristics recognized within the patterns of seismic amplitudes, phases, or dips (Baaske et al., 2007). Broadly, these localized patterns in seismic data are represented by seismic attributes, which are secondary records extracted from seismic data by various mathematical transformations. When wisely extracted using specifically targeted techniques, seismic attributes can highlight specific geological and petrophysical characteristics (Amendola et al., 2017; Hampson et al., 2001). Examples of such complex transformations leading to attributes with the meanings of acoustic impedance and porosity were discussed in chapter 4. Seismic attributes were successfully applied in characterizing and evaluating reservoir properties in many studies (e.g., Ashraf et al., 2019; Du et al., 2015; Farzadi, 2006; Riedel et al, 2013a, 2013b; Roy et al., 2013; Wrona et al., 2018; Zhang et al., 2019; Zhao et al., 2015, 2018).

In modern practice, some form of quantitative interpretation of seismic attributes represents the only way to classify the facies of the reservoir. Since seismic data are continuous and contain a considerable amount of data, manual interpretation is a time-consuming and demanding task even for an experienced interpreter, and it can still be impractical for detailed interpretation of a 3-D survey. Therefore, automation and machine learning could be a proper solution to this problem. Recently, there has been a tremendous increase in the number of studies using multivariate regression and ANN applications for automating the more challenging tasks in seismic interpretation and seismic pattern recognition. (e.g., Baaske et al., 2007; Bagheri and Riahi, 2015; Col´eou et al., 2003; Feng et al, 2018a, 2018b; Linari et al., 2003; Liu et al., 2018; Marroquín, 2014; Roden et al., 2015; Saggaf et al., 2003; Sayago et al., 2012; Zahraa et al., 2017; Zahmatkesh et al., 2021).

In this chapter, I use the 3-D stacked seismic volume to characterize the rock facies distribution within the Mansuri reservoir described in chapter 3. In section 5.1, I describe the seismic attributes used for the classification and select an optimized set of attributes using a

conventional and a novel method described in chapter 2. Using these optimal attributes and a selected ANN structure, in section 5.3, I discuss selection of the regularization parameters for its training. In section 5.4, I present three forms of the resulting classification and discuss their statistical properties and significance.

For consistency with the previous study (Zahmatkesh et al., 2021), the seismic facies targeted in these classifications are the electrofacies EF1 to EF4 described in chapter 3. Electrofacies are identified by their distinctive physical and chemical characteristics related to a certain rock type, and also by fluid content within the volume assessed by the well-log analysis. Electrofacies are traditionally calibrated on the laboratory analysis of the core plugs in order to confirm their consistency based on geological and reservoir characterization (Zahmatkesh et al, 2021). Accordingly, they are used as training samples to predict the geological (or reservoir) facies based on the wireline logs where the core is not available (Euzen et al., 2010; Roslin and Esterle, 2016). In this chapter, I similarly use the electrofacies EF1 to EF4 identified within the well logs, and then extend this classification to the entire reservoir volume using 3-D seismic data.

An important part of the present study consists in comparing several supervised pattern recognition methods to determinate lateral and vertical differences in Asmari reservoir in terms of the seismic facies. The methodology outlined in this chapter can be viewed as a workflow using multiple seismic attributes to classify seismic facies for reservoir-scale characterization. I compare the three premier supervised classification methods and explore their differences and similarities with the same dataset. After careful comparisons using different quality control techniques, I conclude that the proposed ANN with targeted attribute selection (chapter 2) results in an improved classification compared to the conventional methods. Most importantly, the interpretation of these classifications changes significantly if we also visualize the statistical significance of the results.

5.1 Seismic attributes

Seismic attributes were already utilized for porosity estimation in chapter 4 in this thesis, but here, I analyze them more closely with regard to their contribution to classification of seismic facies. Generally, a seismic attribute is some local quantity extracted from seismic data, obtained by a mathematical transformation or an algorithm enhancing and quantifying features of interpretational interest. The classic attributes from which the modern culture of visualization has

started were the instantaneous amplitude (envelope), phase and frequency of the seismic signal (Taner et al., 1979). Hundreds of seismic attributes are in routine use today, which can be divided into several groups based on their usage. For example, there are attributes helping us to quantify the morphological component of seismic data (similarity, curvature, spectral decomposition, dip and variance) and those that help quantifying the reflectivity component of seismic data (amplitude envelopes, frequencies, acoustic impedance).

This study uses an integrated practice of combining multiple seismic attributes with four well-log electrofacies in order to evaluate the study area's lithofacies and reservoir quality distribution patterns (chapter 3; Zahmatkesh et all, 2021). To achieve this task, the relevant input attributes must be carefully selected out of the multitude of available transformations of the 3-D seismic data. Determining the optimal set of input attributes is one of the main challenges of the classification procedure. Clearly, selecting attributes with some geophysical and geologic values would be the best strategy compared with testing various pure mathematical transformations. However, in this project, I attempt a more abstract, "machine-learning" approach to selecting the final attributes based on evaluation of their metrics and sensitivity to the intended classification of seismic facies.

In the first pass of the above attribute selection, I extracted 23 attributes from the reflection seismic image plus the estimated acoustic impedance (AI) and porosity from chapter 4. These attributes with summaries of their meanings are listed in Table 5.1. Values of these attributes within the cross-section passing through eight wells are shown in Figure 5-1. This cross-section will be discussed in detail in section 5.4.

As Figure 5-1 shows, the attributes are broadly variable in terms of their correspondence to geological layering, depth resolution, and noise levels and patterns. In particular, the similarity attribute (number 4 in Table 5.1; Figure 5-1) contains a banded pattern which is likely to be an artifact of the algorithm. In addition, several of these attributes are derived from the AI by similar transformations, and therefore they may be mutually related. Nevertheless, I did not attempt reducing this list by analysing the algorithms or performing signal processing. Instead, I start with the entire 25-dimensional space of input data (Figure 5-1) and develop procedures for reducing this dimensionality to new 7 or 9 attributes, which are used in the final classifications. Two

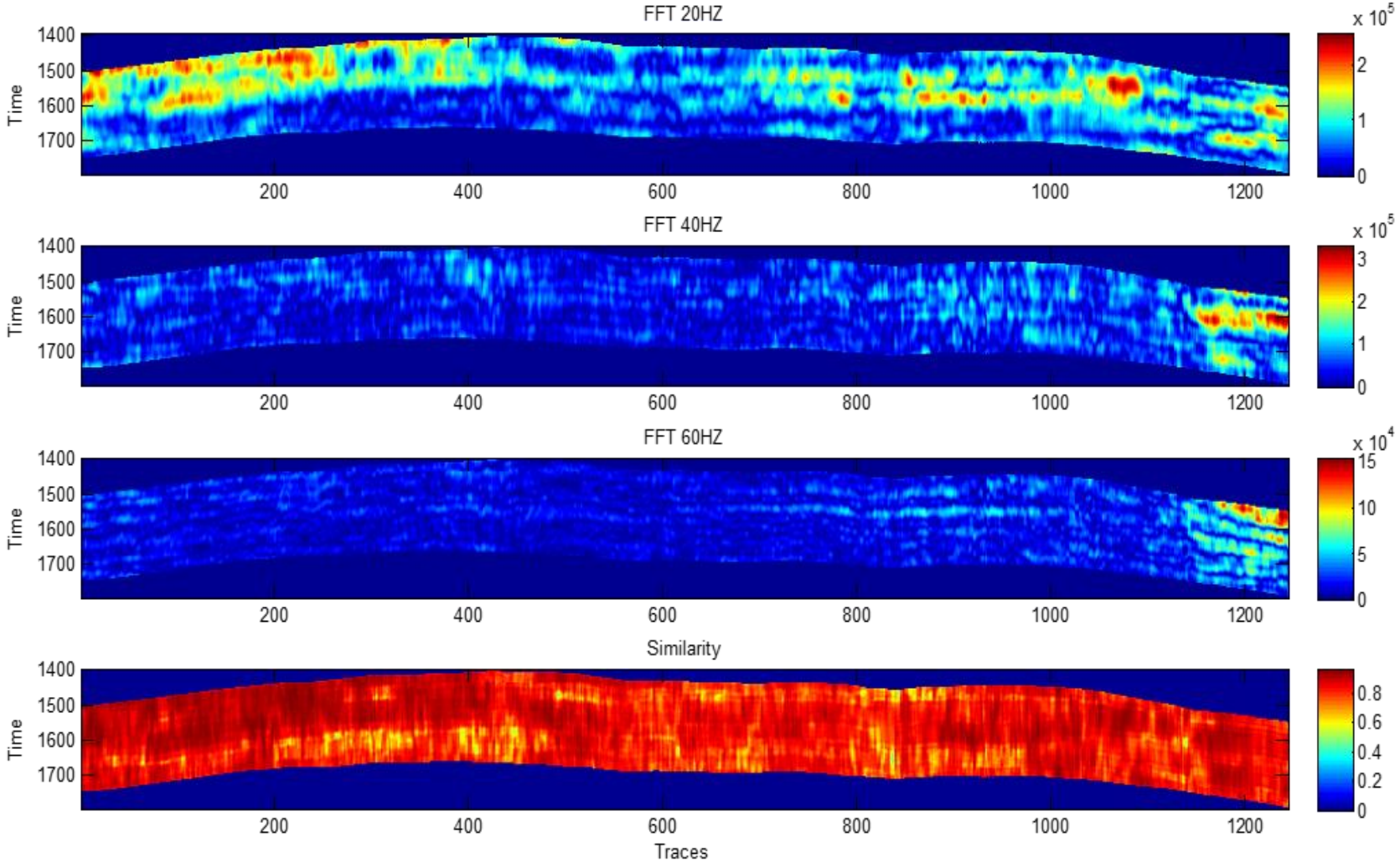

Figure 5-1: Seismic attributes within the selected line section.

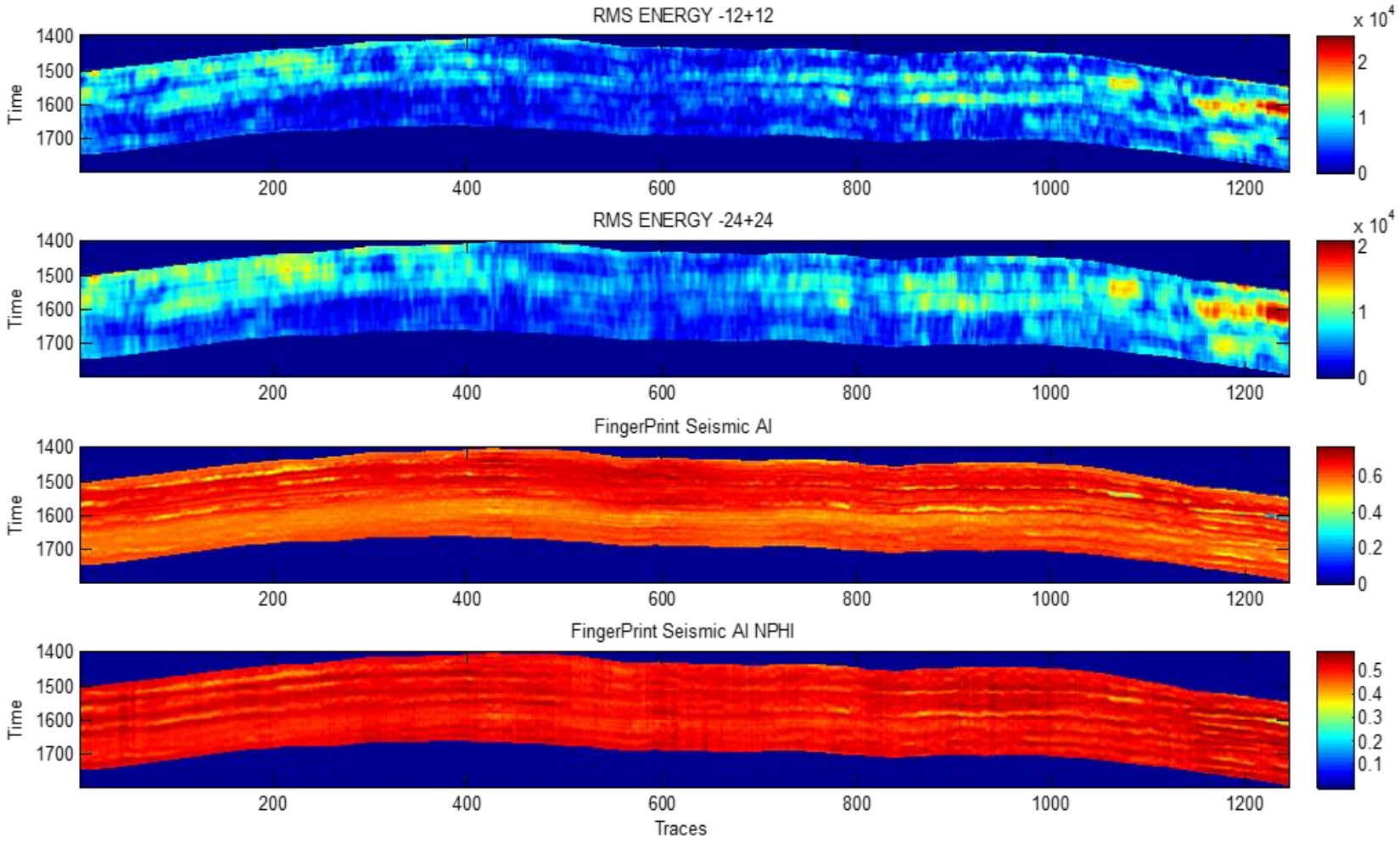

Figure 5-1, continued.

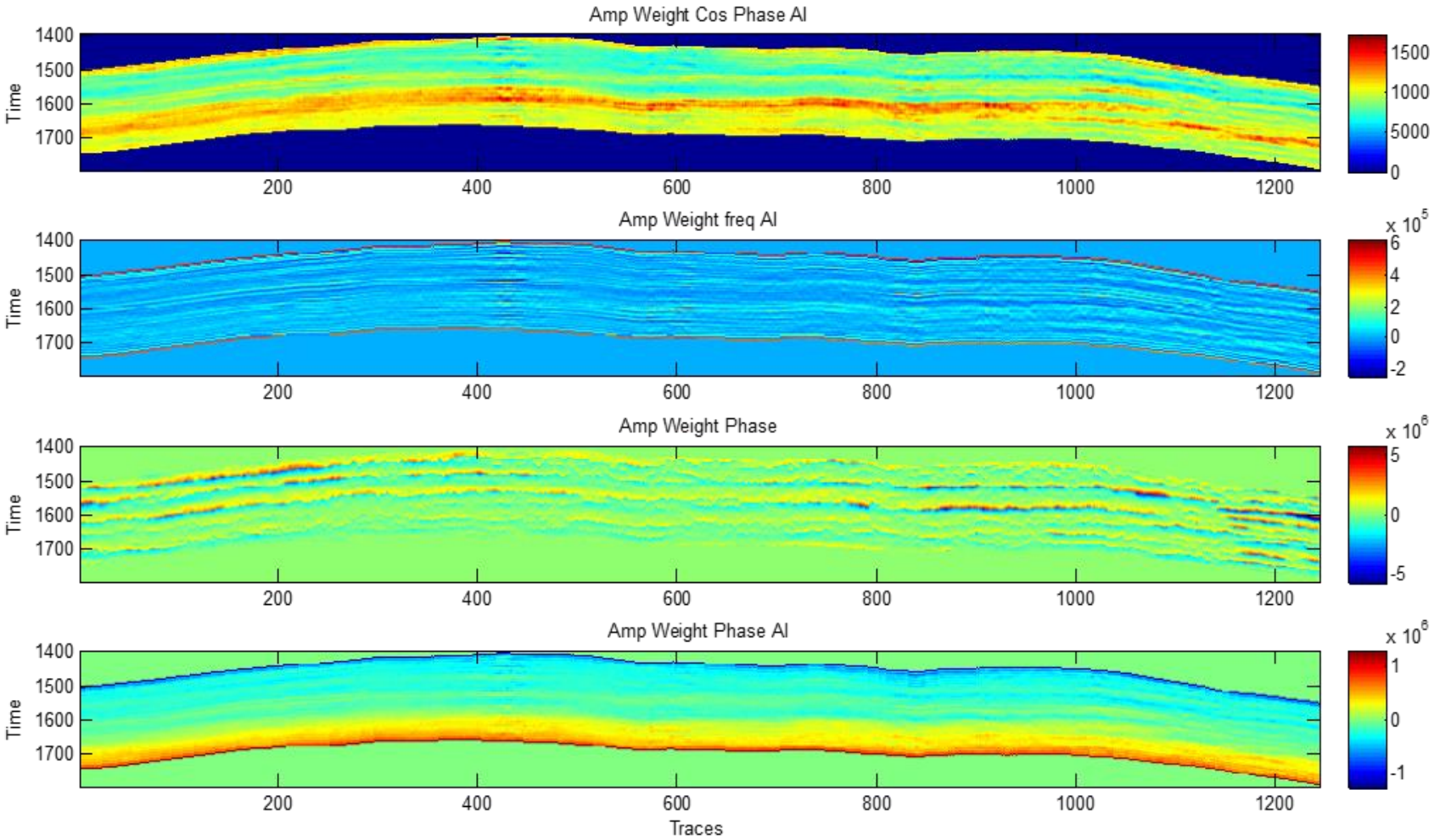

Figure 5-1, continued.

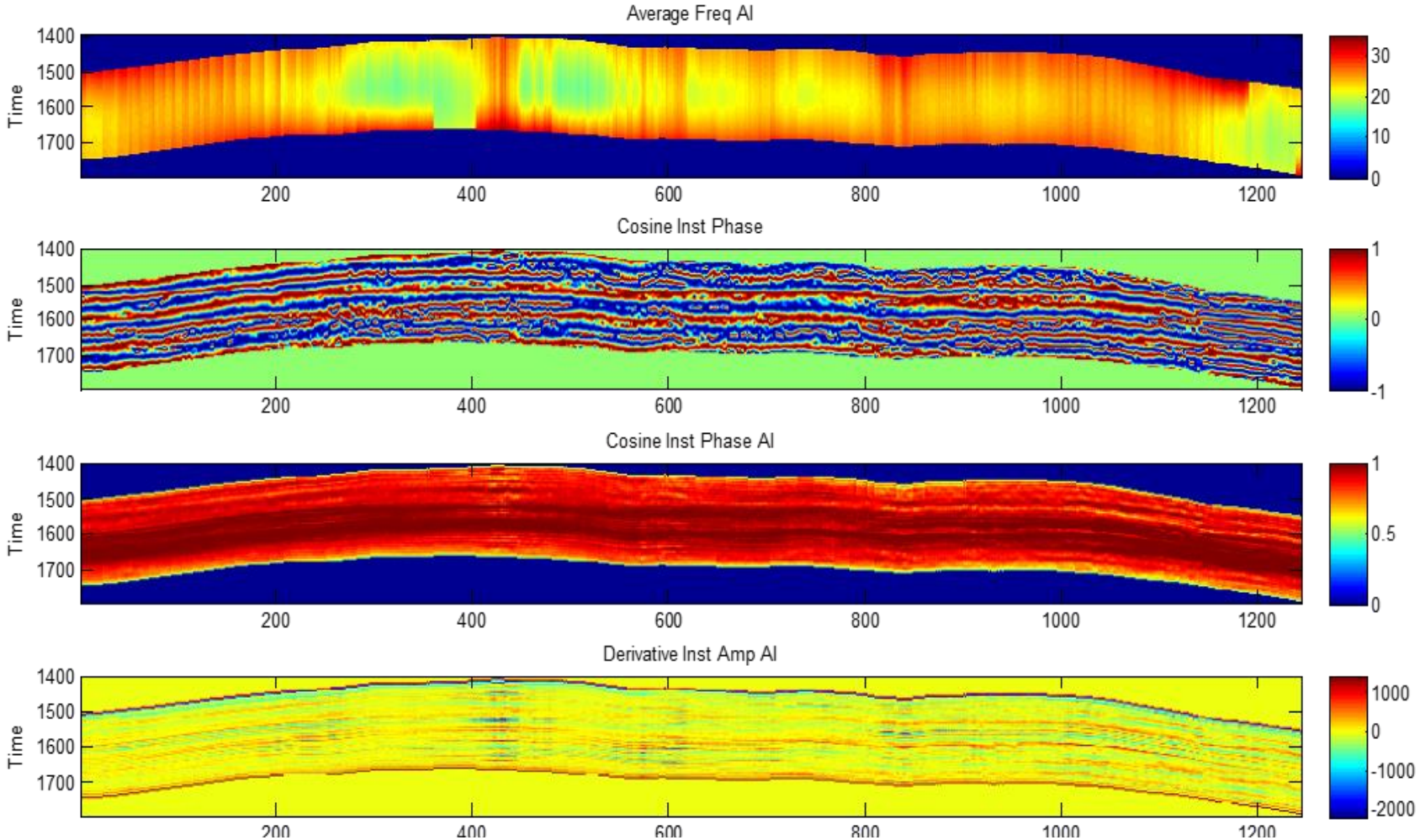

Figure 5-1, continued.

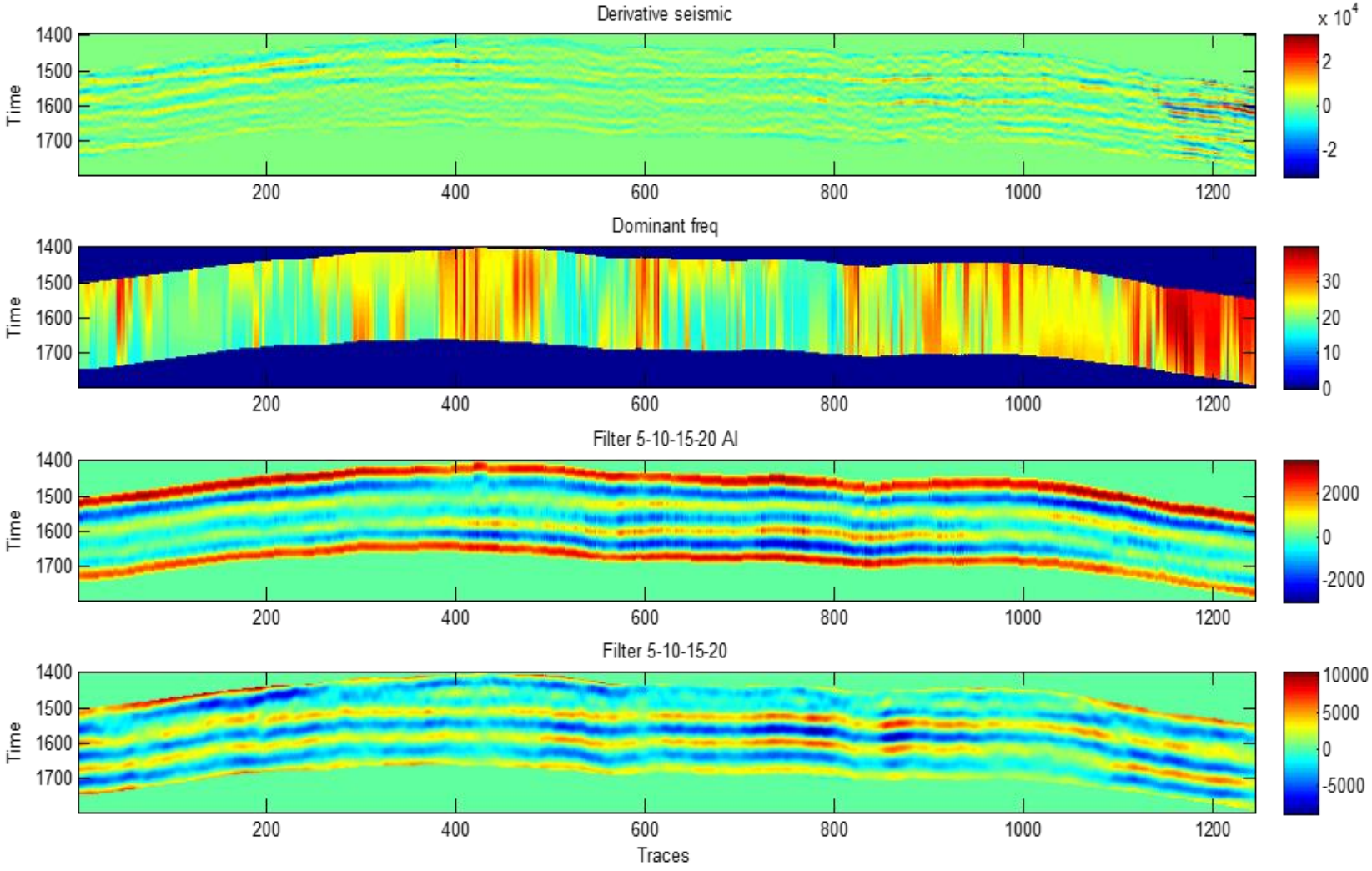

Figure 5-1, continued.

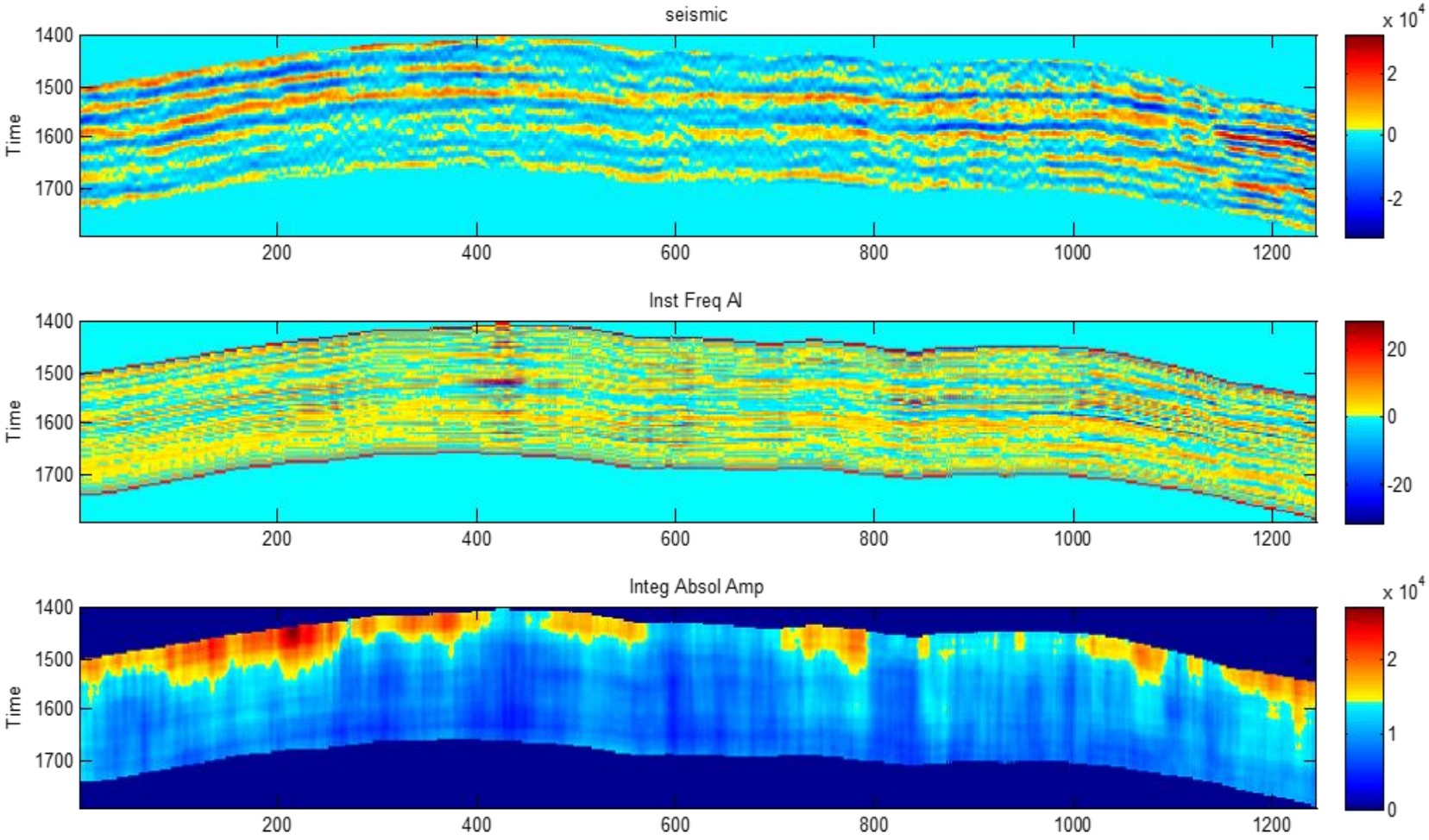

Figure 5-1, continued.

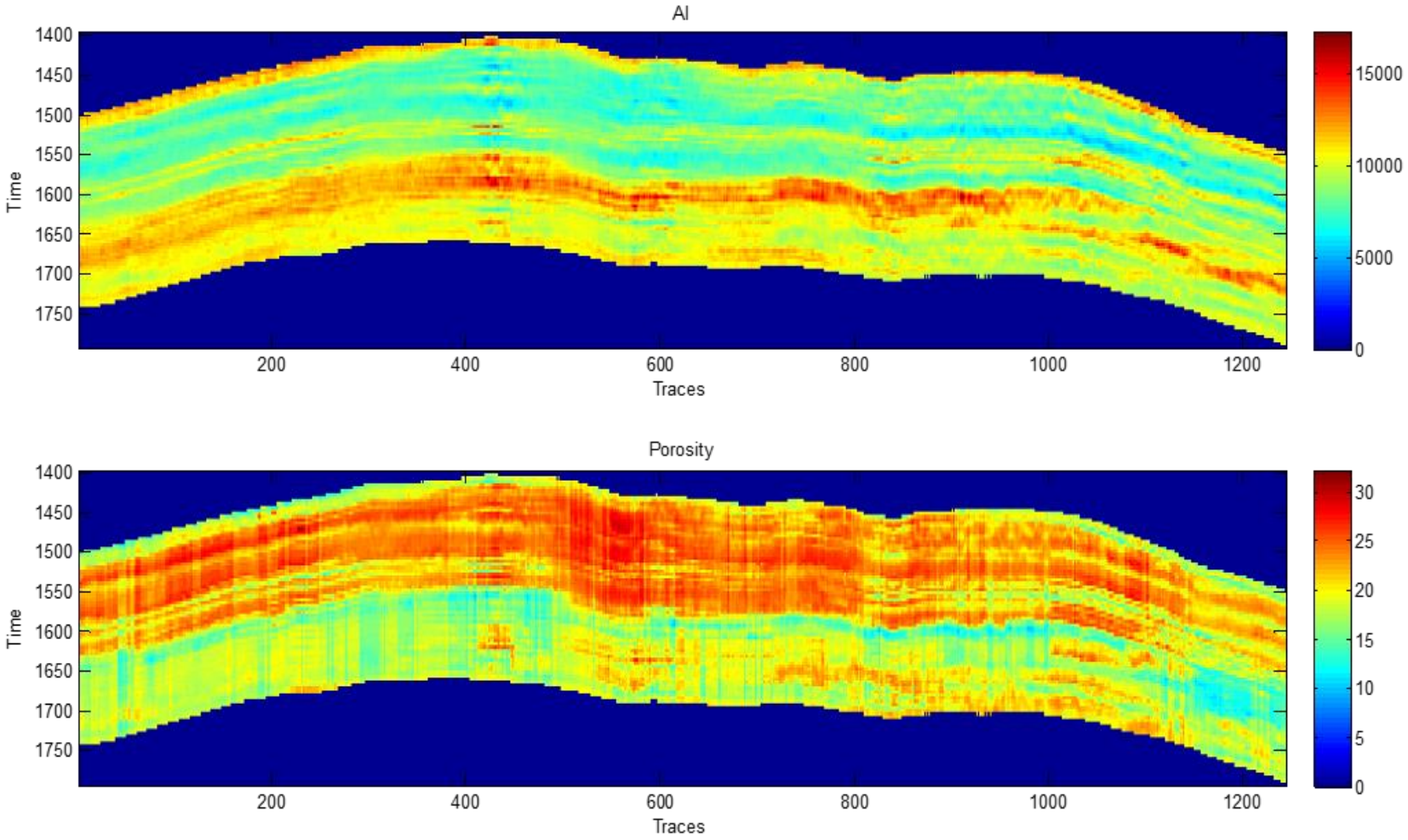

Figure 5-1, continued.

Table 5.1: Seismic attributes used in this study

Number	*Name*	*Description*	*Reservoir and geological implications*
1	FFT spectral amplitude at 20Hz	Signal amplitudes at the specified frequencies	- Useful for sequence boundaries and thin bed tuning - Correlates with lithology & porosity
2	FFT spectral amplitude at 40Hz		
3	FFT spectral amplitude at 60Hz		
4	Similarity seismic full block	Value indicating how much two or more trace segments look alike	-Lateral variable lithofacies, - Visualize channels, reef edges, karst features, fracture zones
5	RMS Energy with time-gate of (-12 +12)	Calculate the squared sum of the sample values in the specified time-gate divided by the number of samples in the gate	-Highlight hydrocarbons and thick sands - Identify stratigraphic features which are isolated from background features, like channels and bright spots
6	RMS Energy with time-gate of (-24 +24)		
7	Fingerprint Seismic & AI	Measure the similarity between the seismic signal and AI (or also NHI) taken at each sample position inside the cube, calculated by the normalized Euclidean distance between the two vectors. Values range from 0 (vectors are not identical at all) to 1 (vectors are 100% identical)	-Useful to find similar good reservoirs as encountered in certain wells
8	Fingerprint Seismic & AI & NPHI		
9	Amplitude-Weighted Cosine Phase AI	Product of the amplitude envelope and the cosine of the instantaneous phase of the AI	
10	Amplitude-Weighted Frequency AI	Product of the amplitude envelope and the instantaneous frequency of the AI	
11	Amplitude-Weighted Phase	Product of the amplitude envelope and the instantaneous phase (of the seismic signal or AI)	
12	Amplitude-Weighted Phase AI		
13	Average Frequency AI	Average of the amplitude spectrum of the AI over a small window around the time sample.	
14	Cosine Instantaneous Phase	Cosine of the instantaneous phase	Highlight bedding very well
15	Cosine Instantaneous Phase AI		
16	Derivative Instantaneous Amplitude AI	Derivative of the amplitude envelope of the AI	
17	Derivative Seismic	Derivative of the input trace, calculated as the difference between adjacent samples	
18	Dominant Frequency Seismic	Maximum of the amplitude spectrum over a small window around the time sample	

19	Filter 5-10-15-20 AI	5/10 - 15/20 bandpass filter of the AI or seismic record	
20	Filter 5-10-15-20 Seismic		
21	Seismic	Stacked seismic record	
22	Instantaneous Frequency AI	Time derivative of the phase of the AI	Abnormal attenuation and thin bed tuning, fluid content
23	Integrated Absolute Amplitude	Integration of the amplitude envelope of the input trace	Show bedding continuity, lithology, fluid content
24	AI	Acoustic Impedance derived from inversion of amplitude seismic data	Indicate lithological and porosity changes, discontinuities, unconformity surface, sequence boundaries and fluid content in the reservoir
25	Estimated Porosity	Porosity estimated in chapter 4	Primary petrophysical property of rock

approaches to this attribute selection procedure are described in the following section. Also, the physical units and ranges of values are different for most of these attributes, and therefore normalization and preconditioning are necessary for feeding them into classification algorithms.

5.2 Preconditioning and selection of optimal attributes

The procedure for optimizing the data inputs is called feature selection or dimensionality reduction in machine-learning applications. The goal of this procedure is to select a subset of extracted features that do not duplicate each other and provide the best predictive power. In this study, I tried several conventional approaches to feature selection described in subsection 5.2.1 and also a more complete data preconditioning procedure described in chapter 2. The distinctive property of all these approaches is in only using the statistical distributions of the data for feature selection, regardless of the target classes. In the subsection 5.2.2, I also apply a more advanced feature reduction method considering the target classes.

5.2.1 Standard approach

The commonly used approach to feature reduction is based on their pairwise correlation in the data (Riedel et al., 2013a, 2013b). By considering the Pearson's correlation matrix **R** between all features (time series of seismic attributes near well locations), a correlation matrix is obtained

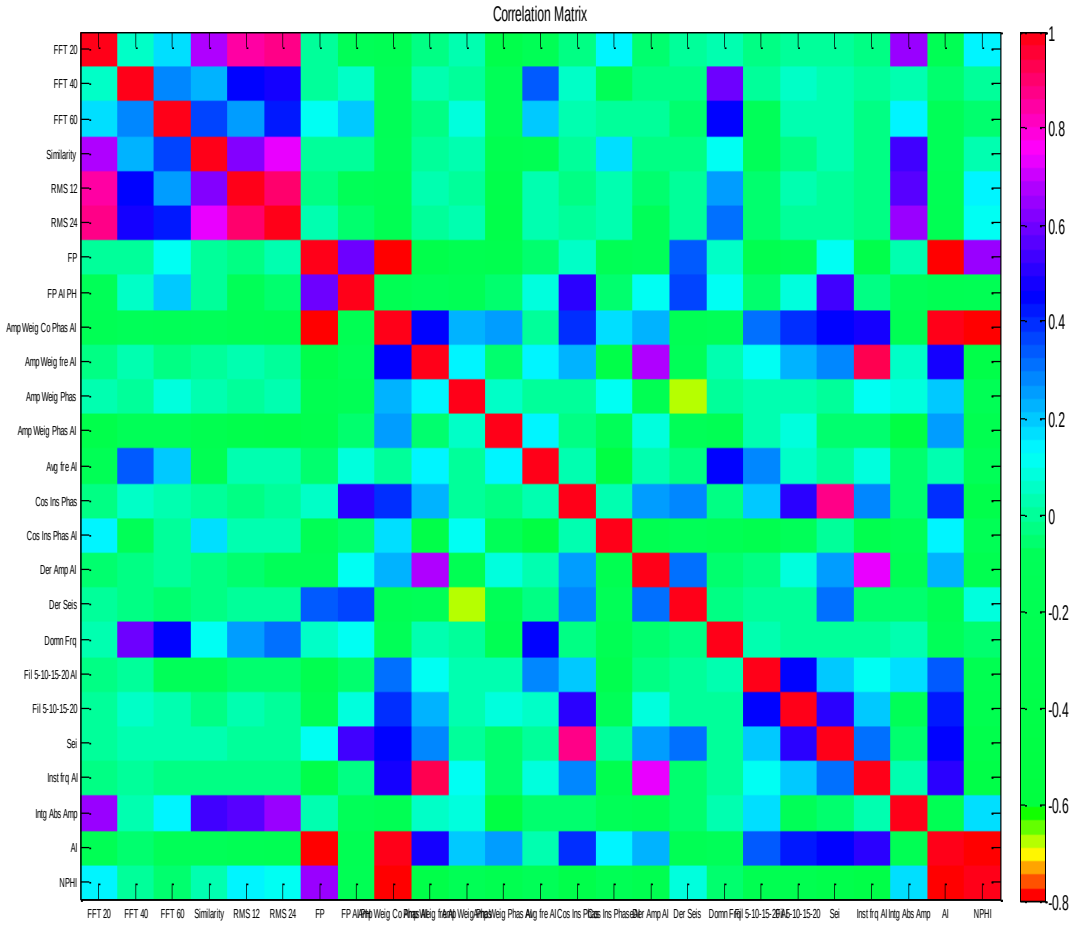

Figure 5-2: Correlation matrix of 25 input seismic attribute volumes. A correlation of ±0.7 or higher may indicate redundancy.

(Figure 5-2). If any two features are highly correlated, i.e. shown by purple or red colors in Figure 5-2, they are expected to contribute similarly to any predictor, and therefore one of them can be viewed as redundant and dropped from the classification model. For example, some of the attributes with the strongest pairwise correlations are (Figure 5-2):

1) Attribute #9 (Amplitude-Weighted Cosine Phase AI) has a very high correlation of 0.99 and 0.8 with attributes #24 (AI) and #25 (porosity), respectively;
2) Attributes #5 (RMS –12 +12) and #6 (RMS -24 +24) show a very high correlation of 0.93. The same level of correlation is between attributes #10 (Amplitude Weighted Frequency AI) and #22 (Instantaneous frequency AI);

3) Attribute #1 (FFT 20 Hz) also has correlation coefficients 0.84 and 0.86 with attributes #5 and #6, respectively;
4) Attribute #14 (cosine of the instantaneous phase) has a high correlation of 0.87 with attribute #21 (seismic record).
5) Attribute #24 (AI) and #25 (porosity) are highly correlated with of $R \approx -0.8$.

In each of these cases, the correlated attributes are either closely related mathematically or arise from a common inversion procedure (such as the estimated porosity and AI in Table 5.1).

One standard approach to feature selection consists in dropping from each pair of correlated attributes the one with lower correlation with the target and less relevance to the objective classification (Haykin, 2009). This approach was tried, but the final results were not acceptable since the performance of the algorithm was not good enough due to deleting some attributes without even using the smallest portion of the information that they might maintain. The second approach is applying a preconditioning step to keep all attributes set and make the most use of all input datasets. In each of these cases, the correlated attributes are either closely related mathematically or arise from a common inversion procedure (such as the estimated porosity and AI).

Another standard approach to feature selection consists in dropping from each pair of correlated attributes the one with lower correlation with the target and less relevance to the objective classification (Haykin, 2009). This approach was also tried, but the final results were not acceptable since the performance of the algorithm was not good enough because of deleting some attributes without even using the smallest portion of the information that they might maintain. The second approach is applying a preconditioning step to keep all attributes set and make the most use of all input datasets.

Figure 5-3 shows the distributions of each of the 25 attributes and Pearson's correlation coefficients between them (labels). Each of the off-diagonal subplot contains a scatterplot of a pair of variables with a least-squares regression line (red), with slope equal to the displayed correlation coefficient. The diagonal blocks of the plots contain histograms of the corresponding variable.

Some of the histograms in Figure 5-3 have bell-like shapes similar to the normal distribution, but others and particularly scatter plots show truncated distributions of the features

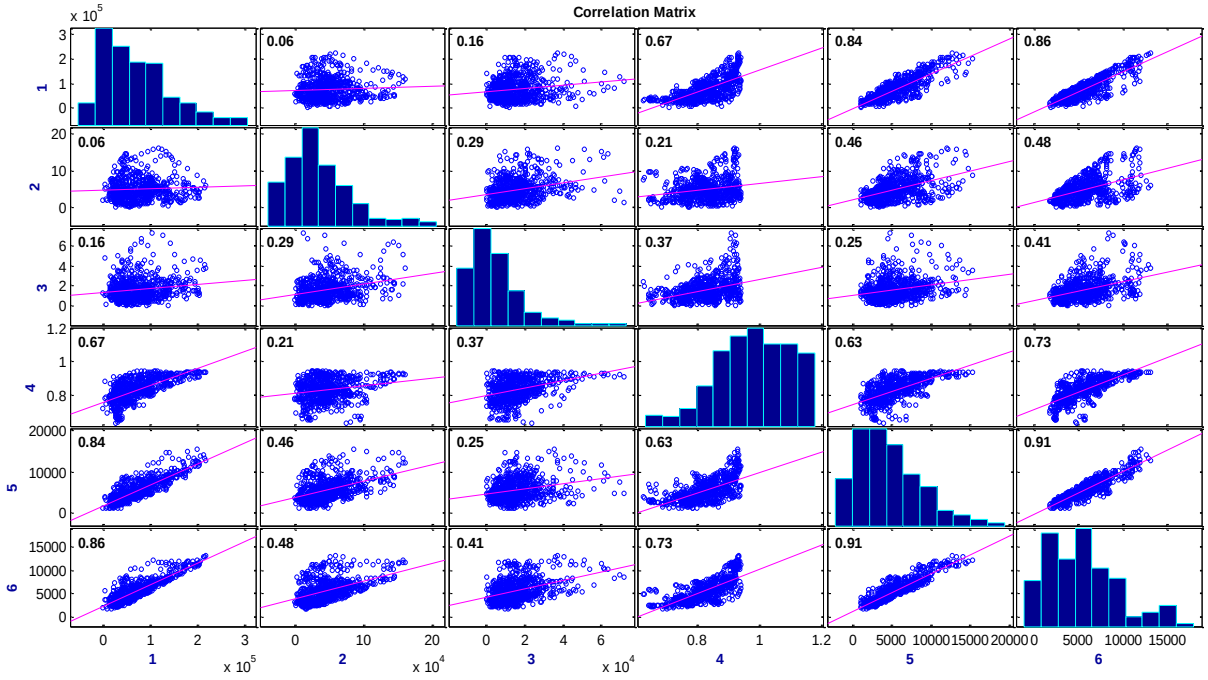

Figure 5-3: Cross-plots of pairs of the raw attributes. Attribute numbers are shown along the bottom and left of the grid of panels. Histograms of the distribution of the attributes are shown on the diagonal of the grid. Red lines indicate the correlations between pair of attributes.

(seismic attributes). Bulb-shaped scatterplots with almost horizontal correlation lines (pink) represent uncorrelated features which should be suitable for their use in classification. However, several attributes show significant correlations and complex distribution shapes (Figure 5-3). Therefore, careful data selection and preconditioning is required (next section).

To assess the feasibility and estimate accuracy of the expected classification, it is useful to plot histograms of the key attributes within each class for near-borehole locations. Figure 5-4 shows such histograms of neutron porosity within the four classes EF1 to EF4. This figure shows that class 1 and class 3 can be differentiated by porosity level $\phi \approx 21\%$. However, these distributions also show that a location with porosity $19\% < \phi < 21\%$ can still belong to class 1 with probability of about 15%. Despite the location of class 3 in the medium and lower parts of the porosity range, this class still represents high reservoir quality because fractures within these rocks create high permeability and high effective porosity of the formation.

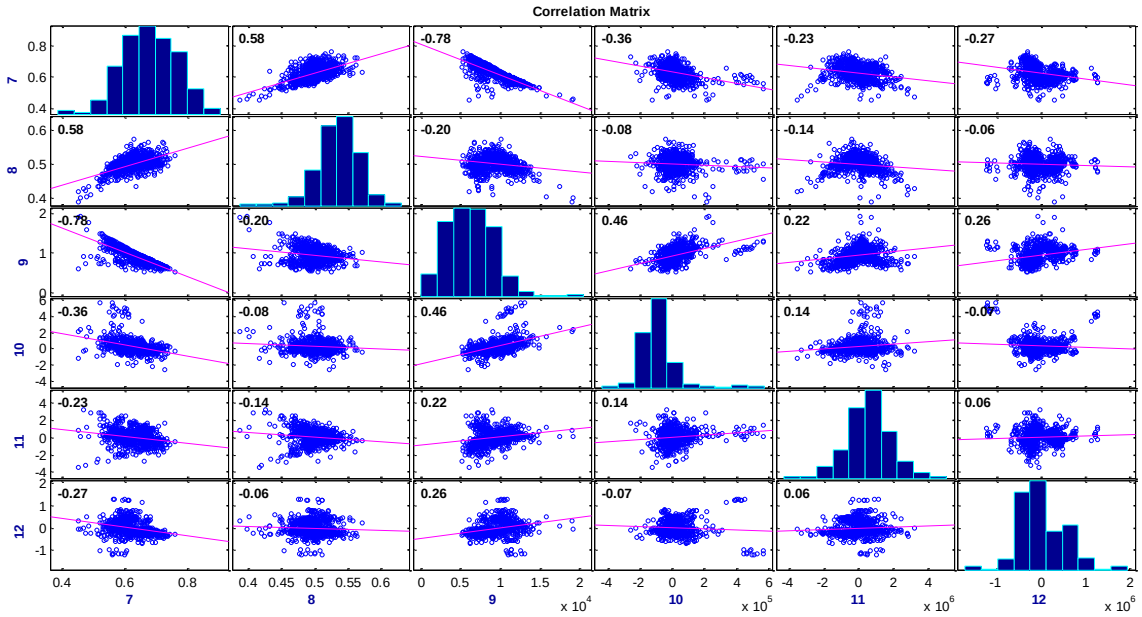

Figure 5-3, continued.

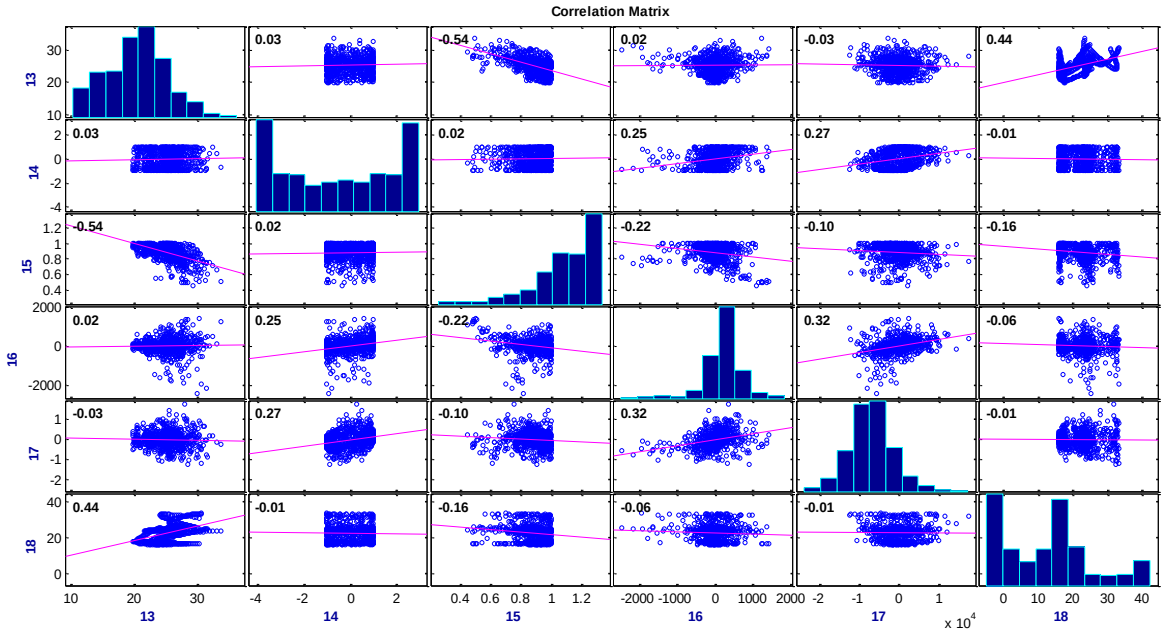

Figure 5-3, continued.

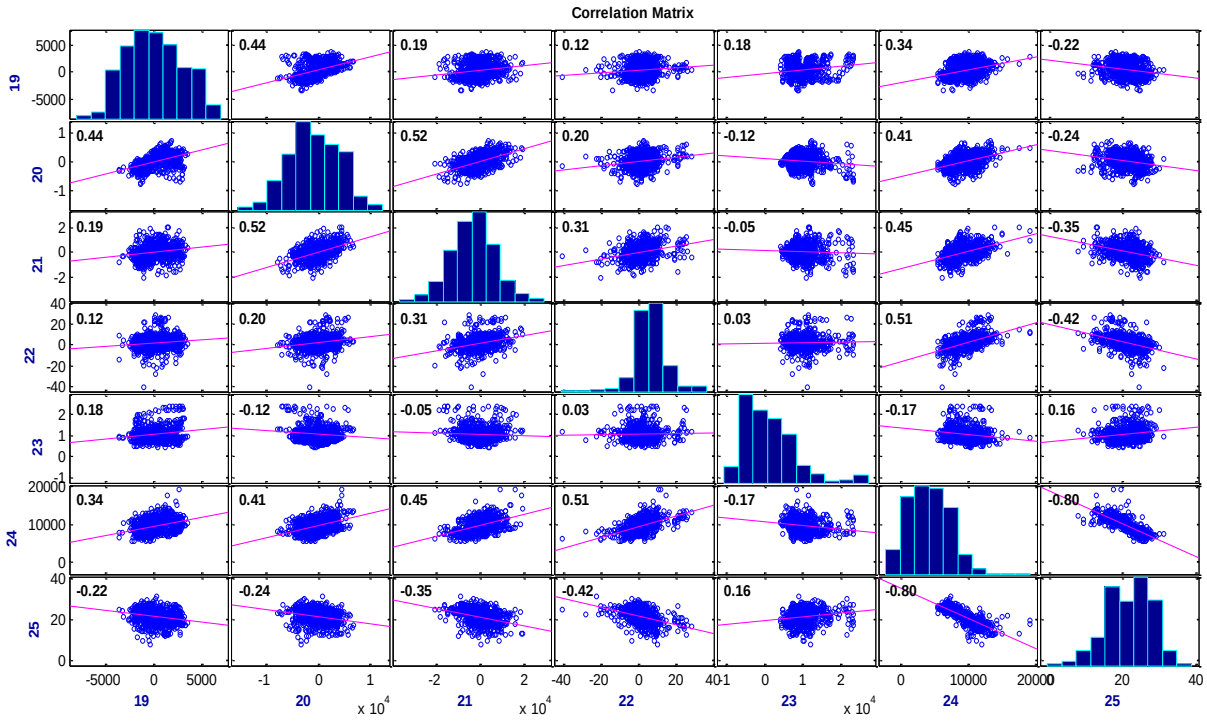

Figure 5-3, continued.

Histograms in Figure 5-4 also show the differences in the available statistics within the four classes. Classes 1 and 3 contain more data points and show the highest peaks in this plot, and class 2 is represented by the fewest number of readings in the dataset.

To visualize the distributions of data within classes using two attributes (for example, #24 and #25), I cross-plot them within each of the four classes (Figure 5-5). Because these attributes are highly correlated (at 80%, see Figure 5-3), observations from this plot are similar to those above. Classes 1 and 2 are at the upper-left of the cross-plot, which means that they have low acoustic impedance and high porosity. Classes 1 and 3 could be separated easily (again with about 15% uncertainty), but classes 1 and 2 are thoroughly mixed. From these attributes, a class 2 zone might be predicted as class 1 and vice versa.

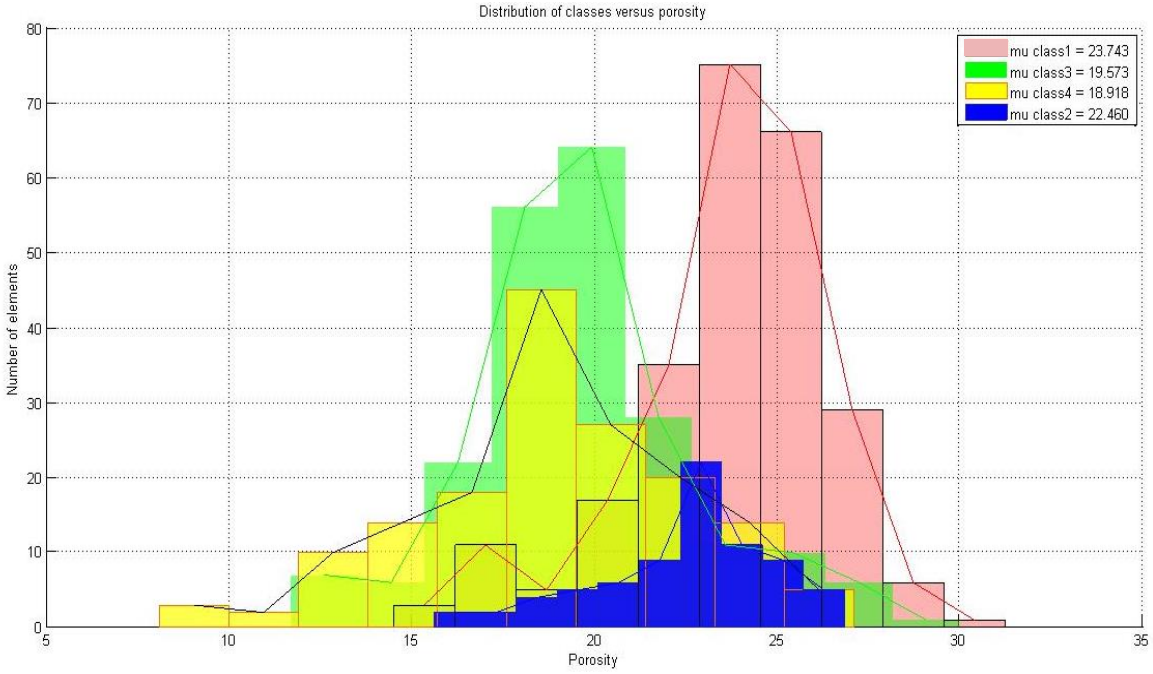

Figure 5-4: Histograms of porosity values for each target class in the wells (legend). Mean values of porosities are labeled in the legend.

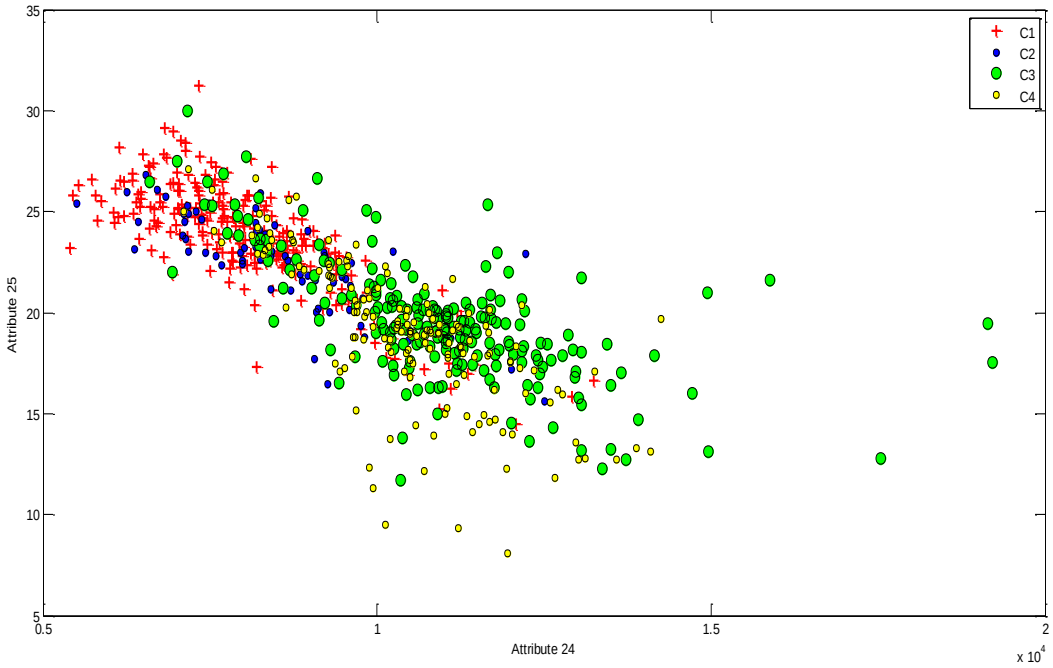

Figure 5-5: Cross-plot of attributes #24 (AI) and #25 (estimated porosity) within each class (legend).

5.2.2 Feature selection using target classes

In this subsection, I apply the optimal feature selection procedure from section 2.6 of chapter 2. As explained there, the advantage of this procedure is in constructing composite features which are uncorrelated, normalized, and have the strongest contributions to the desired classification. The application of this procedure to the Mansuri reservoir data near boreholes is summarized in the following tables. In these tables, column 'accuracy' contains the proportion of data points correctly predicted by the corresponding logistic regression, and the bold underlined accuracy values indicate lower-accuracy attributes excluded from the final classification process.

Table 5.2 shows the attributes most sensitive to one selected class. Table 5.3 shows the attributes obtained by removing one of the classes from the labels (targets) and the corresponding samples from the input (observations).

Table 5.2: Accuracy score of new attributes.

Attribute Number	Description	Accuracy
1	Class 1 =1 and other classes=0	80.3
2	Class 2 =1 and other classes=0	89.3
3	Class 3 =1 and other classes=0	**78.2**
4	Class 4 =1 and other classes=0	82.5
5	Class 1 & 2=1 and other classes=0	84.1
6	Class 1 & 3=1 and other classes=0	**72.3**
7	Class 1 & 4=1 and other classes=0	**68.8**

Table 5.3: Accuracy score of new attributes in absence of classes.

Attribute Number	Removed Class	Description	Accuracy
8	Class 1	Class 2 =1 and other classes (3 & 4) =0	87.8
9	Class 1	Class 3 =1 and other classes (2 & 4) =0	**71.2**
10	Class 1	Class 4=1 and other classes (2 & 3) =0	**74.1**
11	Class 2	Class 1 =1 and other classes (3 & 4) =0	86.1
12	Class 2	Class 3=1 and other classes (1& 4) =0	**77.3**
13	Class 2	Class 4=1 and other classes (1& 3) =0	80.56
14	Class 3	Class 1 =1 and other classes (2 & 4) =0	**78.2**
15	Class 3	Class 2=1 and other classes (1& 4) =0	84.62

16	Class 3	Class 4=1 and other classes (1& 2) =0	88.36
17	Class 4	Class 1 =1 and other classes (3 & 4) =0	77.9
18	Class 4	Class 2=1 and other classes (1& 3) =0	85.8
19	Class 4	Class 3=1 and other classes (1& 2) =0	86.7
20	Class 1 and Class 2	Class 3=1 and other classes (4) =0	**71.3**
21	Class 1 and Class 3	Class 2=1 and other classes (4) =0	86.7
22	Class 1 and Class 4	Class 2=1 and other classes (3) =0	84.3
23	Class 2 and Class 3	Class 1=1 and other classes (4) =0	88.7
24	Class 3 and Class 4	Class 1=1 and other classes (2) =0	**77.7**
25	Class 2 and Class 4	Class 1=1 and other classes (3) =0	85.2

To reduce the dimensionality of the feature space and possible overfitting while having good representation of all four classes, I dropped the attributes with accuracy below 80% (underlined in Table 5.2 and Table 5.3). The resulting dataset then contains 16 attributes. After orthogonalization and preconditioning, I further used the singular value decomposition to obtain the eigenvalues (variances) and eigenvectors (principal components), which represent the final new set of attributes. The cumulative sum of data variances (Figure 5-6) shows that the first three

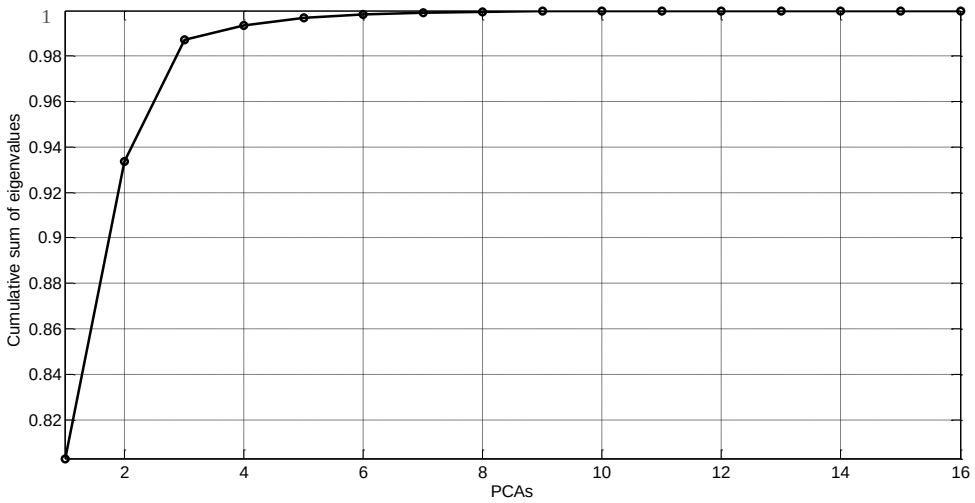

Figure 5-6: Relative cumulative data variance among the new attributes.

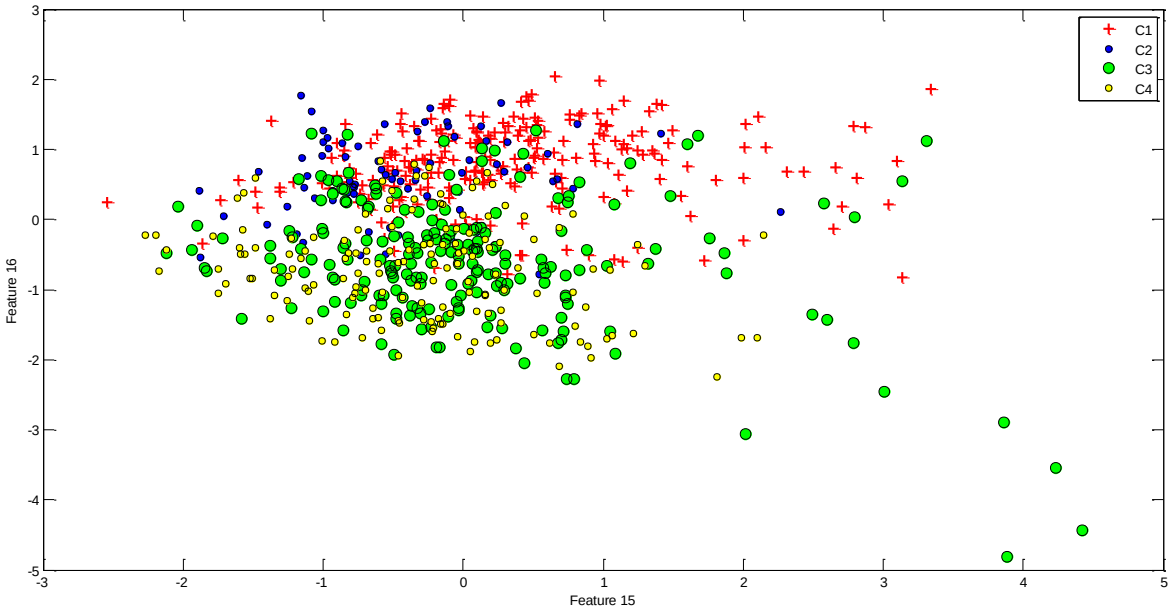

Figure 5-7: Distribution of the classes with respect to two new features.

of the new attributes retain more than 98% of the total data variance. Figure 5-7 shows the distribution of our classes versus the highest eigenvectors in our new dataset. These two new features explain more than 93% of the data variation.

The third standard approach to feature reduction is by using the principal-component analysis (PCA) (Haykin, 2009). This approach consists in finding the eigenvectors and eigenvalues of the covariance matrix of the data, or equivalently, by the singular value decomposition of the data matrix itself. The key output of the PCA is a set of eigenvectors $\mathbf{e}_i$ in the feature space (called principal components) and their variances σ_i^2. The variances are sorted in descending order, so that the first principal component is dominant in the data. The principal components are uncorrelated, and therefore if we select *N* first principal components, then they will account for the total variance in the data equal $\sigma_N^2 = \sum_{i=1}^{N} \sigma_i^2$. A ratio of this cumulative variance to the total variance $\sigma_N^2 / \sigma_{N=25}^2$ is shown as a function of *N* in Figure 5-8. This figure shows that by keeping only the first 10 eigenvectors in the data, we can account for almost 90% of the total data variance. By

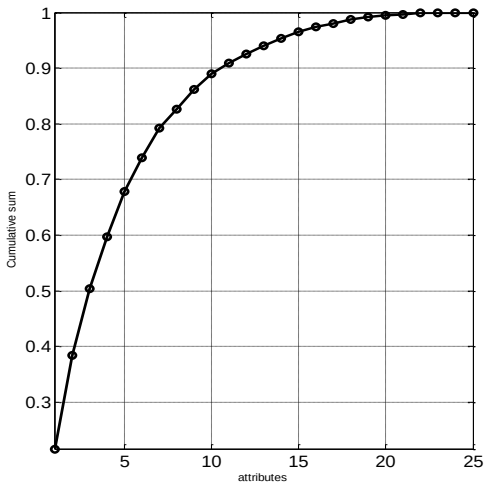

Figure 5-9: Relative cumulative sum of the variances for the 25 principal components of data.

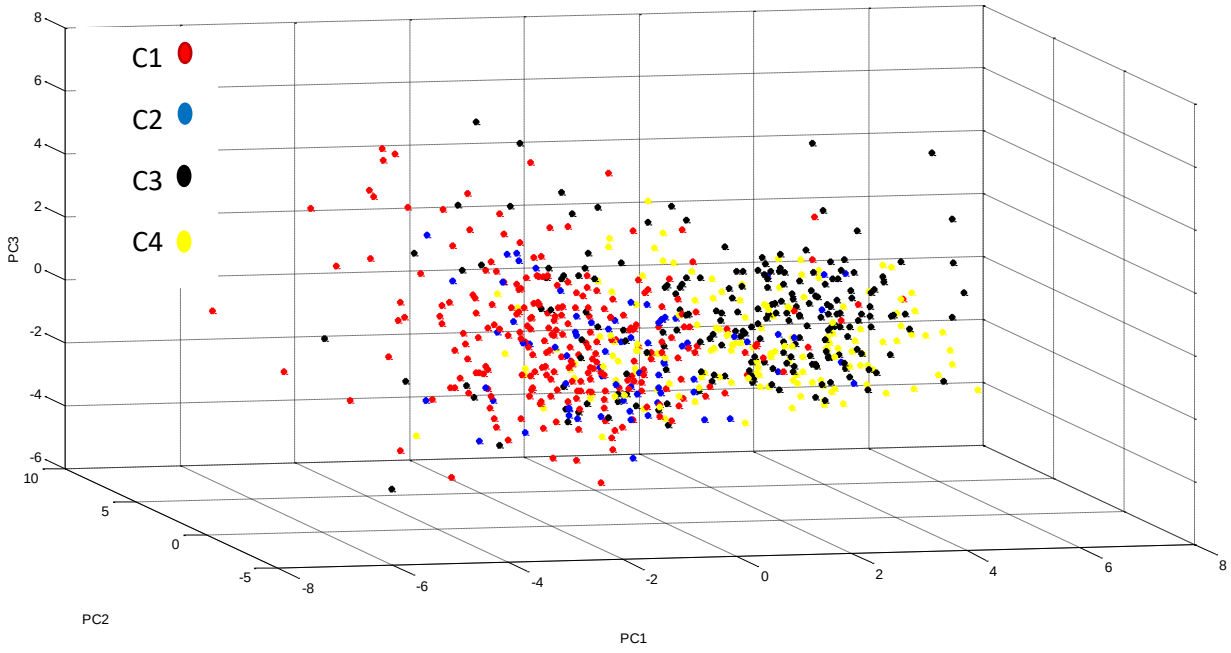

Figure 5-8: Distribution of classes in the space of the first three principal components of the data.

selecting the first three or seven principal components, 50% and 80% of the data variation would be retained, respectively. If using the first three principal components, a good visual separation between the four classes can be obtained in the 3-D space (Figure 5-8).

Finally, the preferred data preconditioning approach in this study was the complete feature othogonalization procedure described in chapter 2. This procedure includes the PCA and optionally, it can also include the correlation-threshold based feature rejection described above. After this preconditioning, the resulting features become mutually uncorrelated and orthogonal combinations of all original features, sorted in reverse order of their impact on the data. Similarly to PCA, the feature reduction consists only in selecting the desired number of final features. Figure 5-10 illustrates the fact that these new features are uncorrelated and bulb-shaped, which is indicated by the absence of average slope in the distribution of data points. This plot can be compared to Figure 5-5. The uncorrelated features make it easier to evaluate the separation between the target classes. Figure 5-11 shows the most important original features contributed to the most dominant principal component. As expected, the AI and porosity with "Amplitude weight cosine phase AI" attribute have the highest contributions compared to other original features.

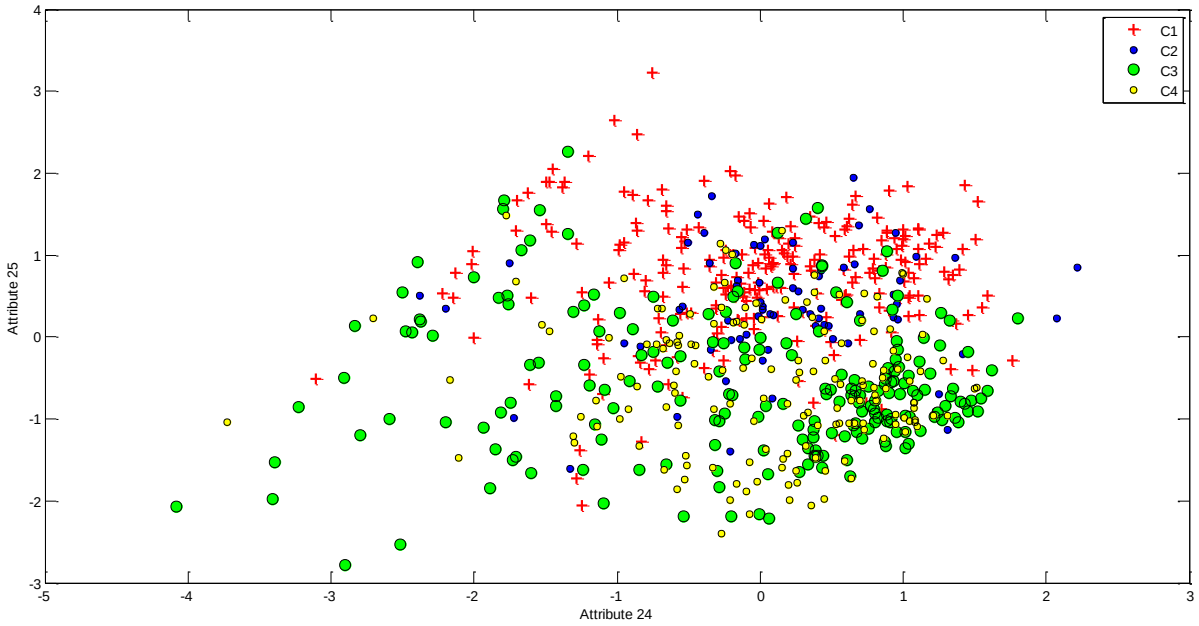

Figure 5-10: Distribution of classes using the transformed features #24 and #25 (legend).

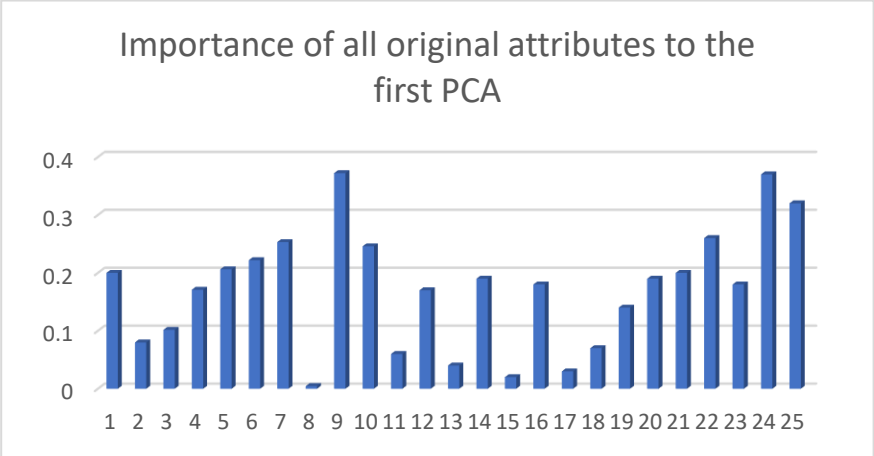

Figure 5-11: Importance of each original attribute to the dominant PCA.

5.3 Selection of algorithm parameters

Along with the preconditioning and selection of data features, results of classification depend on the types of algorithms, architectures of the ANNs, and regularization and tuning parameters included in them. Such regularization parameters were denoted λ in chapter 2. After several trials, I selected an ANN with a single hidden layer with 30 neurons. To test for the optimal λ value, I used the validation and accuracy scores shown in Figure 5-12.

Figure 5-12 shows the performance of the trained neural network as a function of λ when using the 16 selected features. In this figure, parameter λ is varied from 0 to 10, and a classification is performed using a subset of wells. For each value of λ, training and validation scores are measured using some blind wells (not included in training). The blue line represents the training score, which monotonically decreases with increasing λ. The red line is the validation score which oscillates for $\lambda < 1$, then slowly rises, reaches the maximum value near $\lambda = 3$, and remain stable to the end of the tested interval (Figure 5-12). The validation score never goes over the 65%. Thus, more effort is required to reach the benchmark 70% validation score. Therefore, I also test the performance of the algorithm using the 7 highest eigenvalues (Figure 5-13). The validation accuracy becomes over 73%, which is a significant improvement, and the optimal $\lambda = 0.5$, at which the validation accuracy is the highest and the training accuracy is about 77%. Thus, at this point,

the optimal number of features is seven, and the optimal lambda (factor in the regularization term) is 0.5 (Figure 5-14).

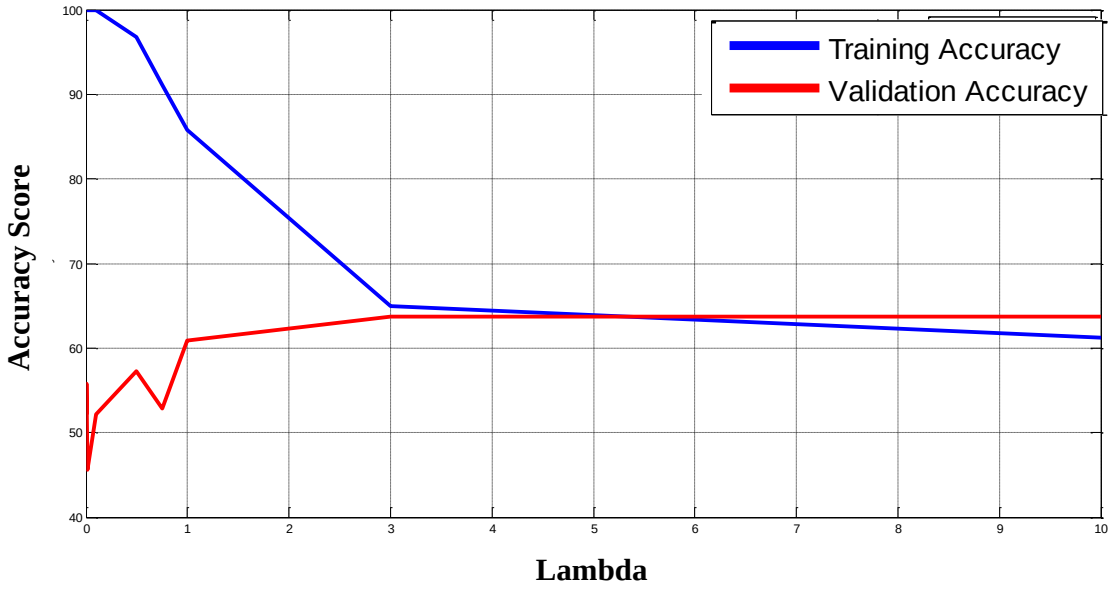

Figure 5-12: Training accuracy and validation accuracy using 16 principal components.

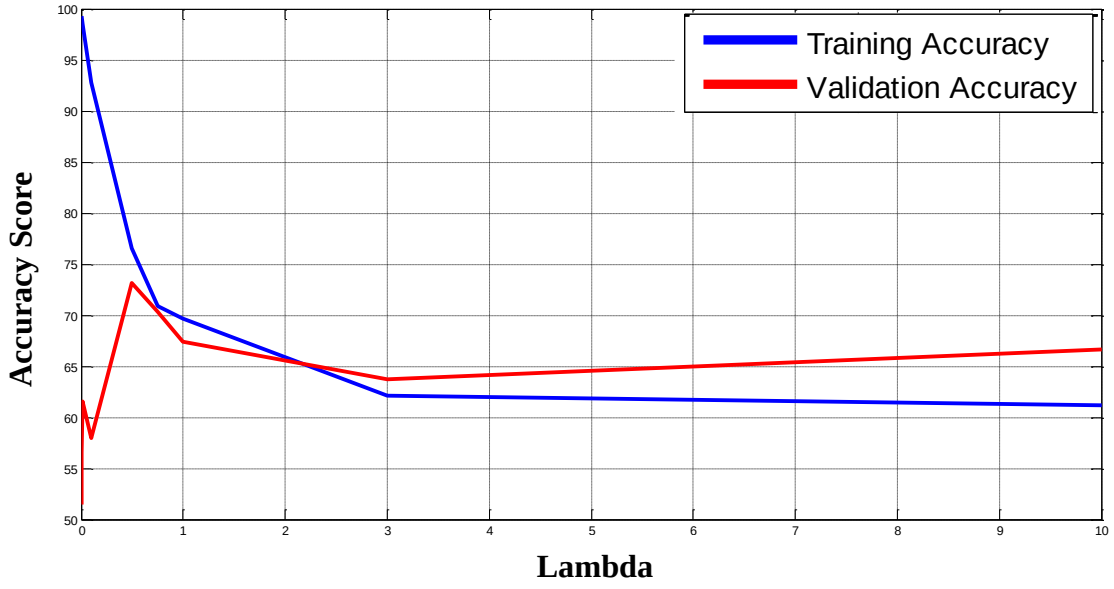

Figure 5-13: Training and validation accuracies versus regularization parameter λ for ANN training using seven principal components.

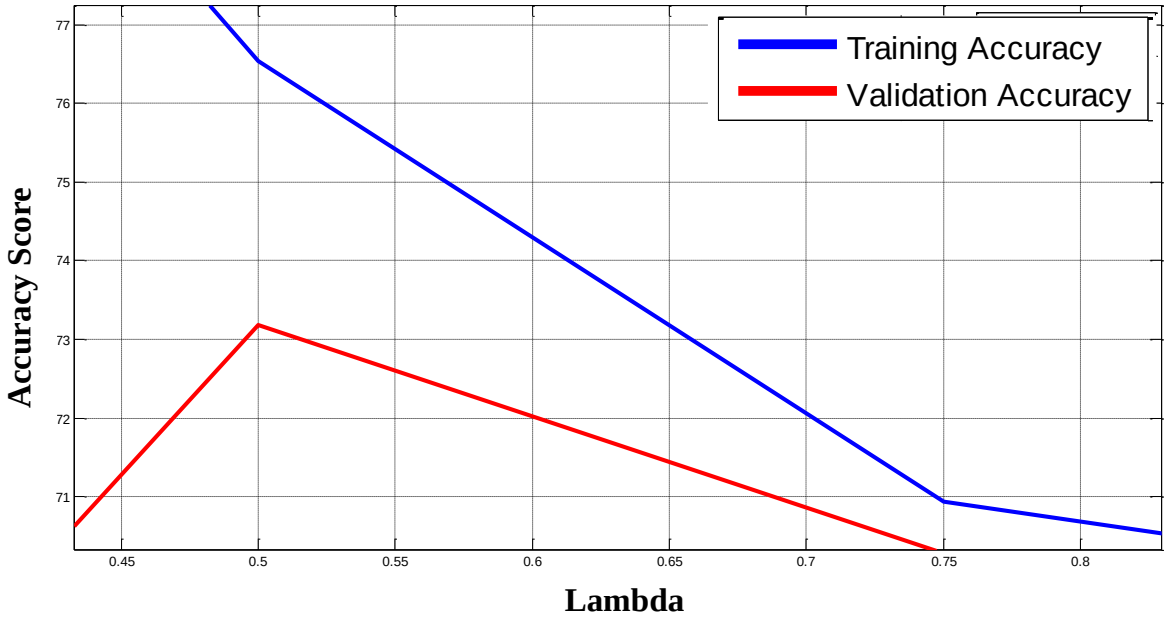

Figure 5-14: Training and validation accuracies vs. parameter λ using seven principal components of the data.

The next analysis is to investigate the optimal number of neurons or hidden-layer size. This selection can also be done based on the validation and training scores. Figure 5-5 shows these scores versus the hidden layer size from 7 to 100 neurons, determined by using the 7 selected features and $\lambda = 0.5$. As shown in this figure, the peak validation score just above 70% is attained with 30 neurons in the hidden ANN layer, but with 20 or 50-60 neurons, the validation performance is only slightly lower. Thus, the final algorithm parameters for the present project include seven input attributes, an ANN with the hidden-layer size of 30, and the regularization parameter $\lambda = 0.5$.

The regularization term (λ) is necessary to avoid overfitting. Although setting a high value for λ, causes a significant overdamping of the solution. With such strong regularization, the inversion is mostly focused on minimizing the ANN weights and not achieving accurate data prediction. On the other hand, selecting $\lambda = 0$ may still cause an ANN with highly variable weights causing overfitting the training data.

In all presented graphs, when λ=0, the training accuracy is 100%, which means the model is too complex and will learn too much about the data including noises, and it will not be able to

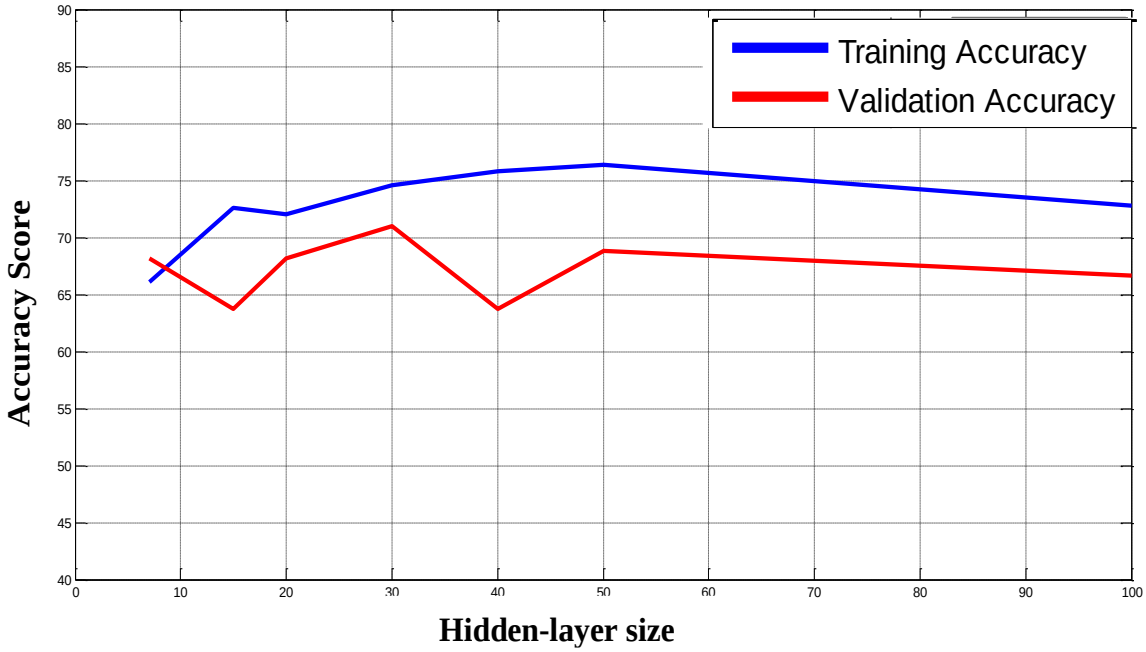

Figure 5-15: Searching for the optimal hidden layer size using seven principal components of the data.

generalize to unseen samples. By increasing the value of λ, the training accuracy will decrease gradually, but validation accuracy will increase at some point and decrease again. Thus, the optimal value of lambda will be at some point when the validation score is at the highest. After finding the optimal number of features, the optimal number of lambda and also the optimal number of neurons (size of hidden layer), the model will be trained based on optimal parameters and will be applied to unseen data.

5.4 Results

Using parameter selections described above, I derived three classifications for Asmari reservoir summarized in Table 5.4. The classifications are referred to as the "regular ANN" (using an ANN with conventional feature selection), the "proposed ANN", or P-ANN (an ANN with improved feature selection described in section 5.2), and the Probabilistic Neural Network (PNN) in Hampson-Russell software (section 5.5). Table 5.4 shows the parameters of these algorithms and the key statistics of their training and validation errors using two validation wells 91 and 96.

Table 5.4: Application of three different neural network on test data (Well 91 and 96).

	Proposed ANN (P-ANN)		Regular ANN		PNN	
Number of features	7 principal components		7 principal components		9 original features	
Type of preconditioning and feature selection	Orthogonalization, relevant features to classes, singular value decomposition		Orthogonalization, singular value decomposition		Step-wise regression	
Lambda value	0.5		1		-	
Validation score	73		67		57	
	Training score	Test score	Training score	Test score	Training score	Test score
Well 91	77.5	68	63.5	62.5	95	56
Well 96	77	72.5	74	66.6	86	50

5.4.1 Validation tests using blind wells

Figure 5-16 and Figure 5-17 show the results of application of the three methodologies to classification of the test data (blind wells 91 and 96). These images show that misclassifications mostly occur within short time intervals, which are likely due to the high-frequency noise present in seismic attributes. Near the bottom of the wells (particularly 96), the performance of PNN is poorer than that of the P-ANN.

Plots in Figure 5-18 and Figure 5-19 show the estimated confidence levels P_k (chapter 2, eq. (2.33)) for the class k identified at each sample in the preceding figures. The average level of confidence is the highest using the proposed ANN (P-ANN) and lowest when using the PNN algorithm. The confidence levels are moderate when the ANN is applied to the blind wells. Overall, P-ANN appears to be superior compared to the ANN and PNN algorithm, in terms of validation accuracy and also considering the confidence of classification. At the same time, interestingly, the classification predictors using PNN in well 91 (right panel in Figure 5-18) are dominated by low-frequency variations on about 100-ms scale, which are different from ANN-based results. The confidence estimates when using P-ANN and ANN reveal similar patterns with depth.

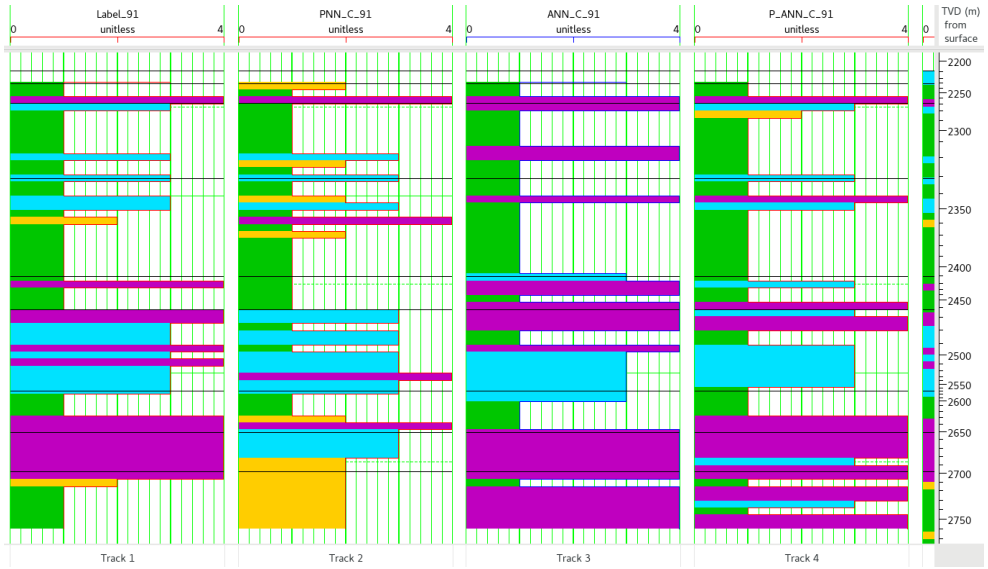

Figure 5-16: Results of classification of the blind well 91. The panels are, from left to right: 1) known classification in the well, 2) results by using the P-ANN, 3) ANN, 4) PNN algorithms. Green, orange, blue and purple color represent classes 1, 2, 3 and 4, respectively. The best match is between the label classes (far left) and the P-ANN (far right).

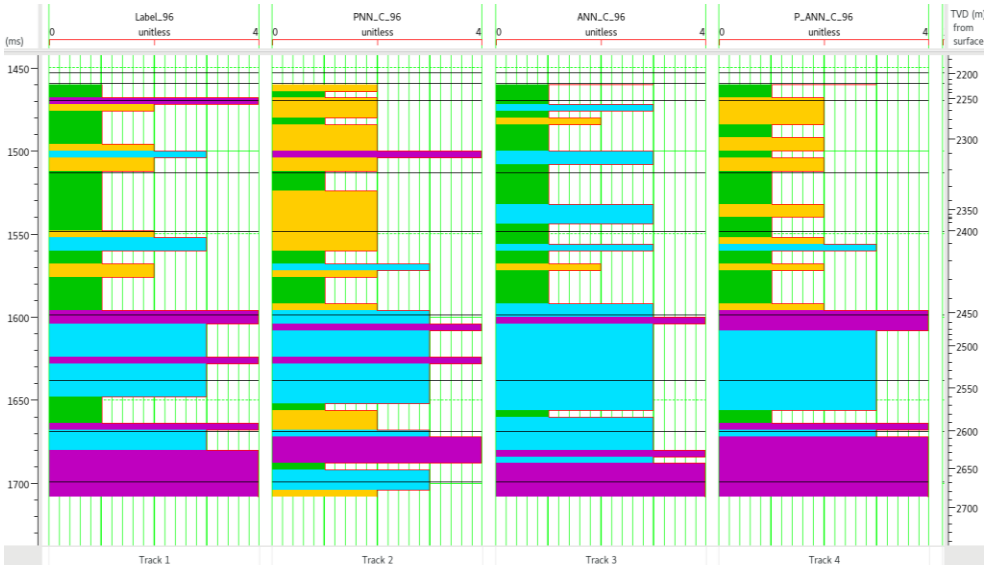

Figure 5-17: Results of classification of the blind well 96. The panels are, from left to right: 1) known classification in the well, 2) results by using the P-ANN, 3) ANN, 4) PNN algorithms. Green, orange, blue and purple color represent classes 1, 2, 3 and 4, respectively. The best match is between the label classes (far left) and the P-ANN (far right).

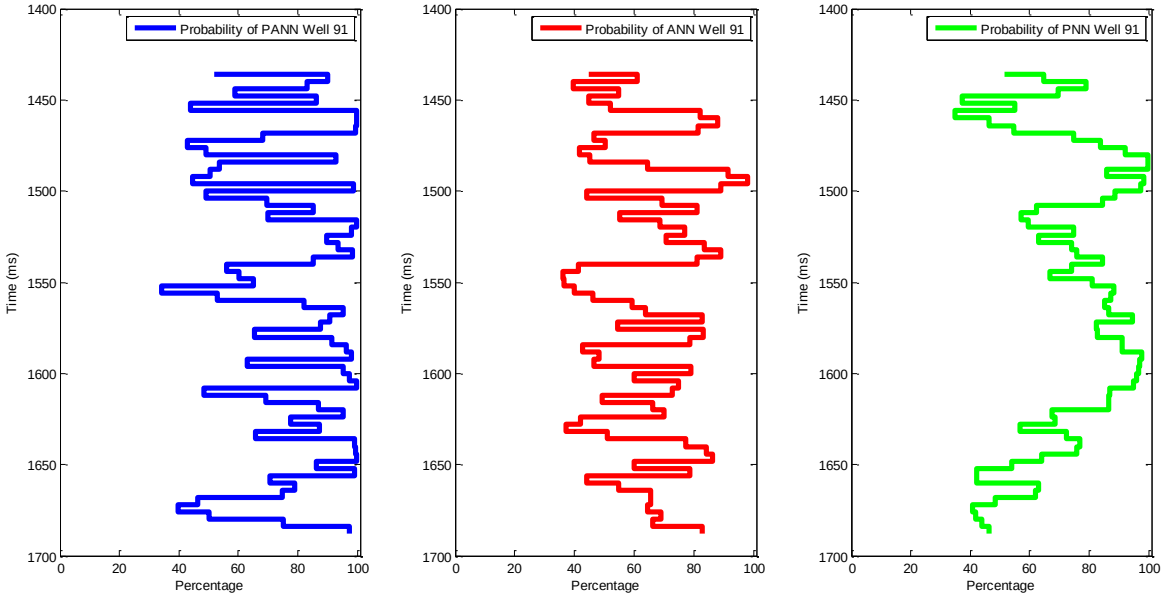

Figure 5-18: Probability of finding the class identified at each point of the subsurface of three different ANN on test data (well 91). The left panel is the results of PANN, the middle panel is the ANN algorithm, and the right panel is the results of PNN. Notice that the overall probability for proposed ANN is higher than ANN and PNN.

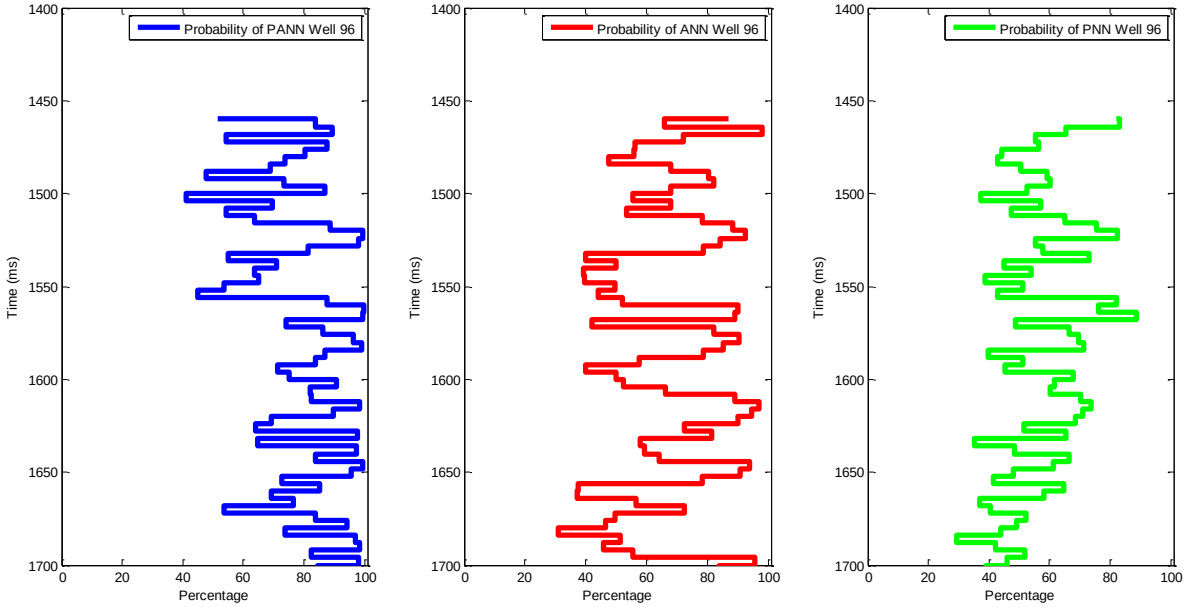

Figure 5-19: Probability of finding the class identified at each point of the subsurface of three different ANN on test data (well 96). The left panel is the results of PANN, the middle panel is the ANN algorithm, and the right panel is the results of PNN. Notice that the overall probability for proposed ANN is higher than ANN and PNN.

To summarize the classification confidence values, the pie chart in Figure 5-20 shows the distribution of confidence levels P_k (chapter 2) for classification of the validation dataset using the P-ANN approach. I consider class probabilities $P_k > 0.7$ as high confidence and $P_k < 0.4$ as low confidence. Note that in observational statistics, the range $P > 0.95$ is usually considered as high confidence, but such stringent requirement is impractical for the present application. With the $P > 0.7$ criterion, the increase of validation accuracy from my previous attempts from 67% to over 73% represents a significant rise in validation accuracy in machine-learning techniques.

For a more detailed assessment of the accuracy of class predictions, confusion matrices are often used (Haykin, 2009). The confusion matrix visualizes the performance of the classification model by comparing the actual and predicted classes. The j^{th} element of the k^{th} row of the confusion matrix represent the percentage of cases in which a data point from class k is classified by the algorithm as class j. Thus, the diagonal of the confusion matrix represents the percentages of correctly classified classes, and the off-diagonal values are the percentages of various

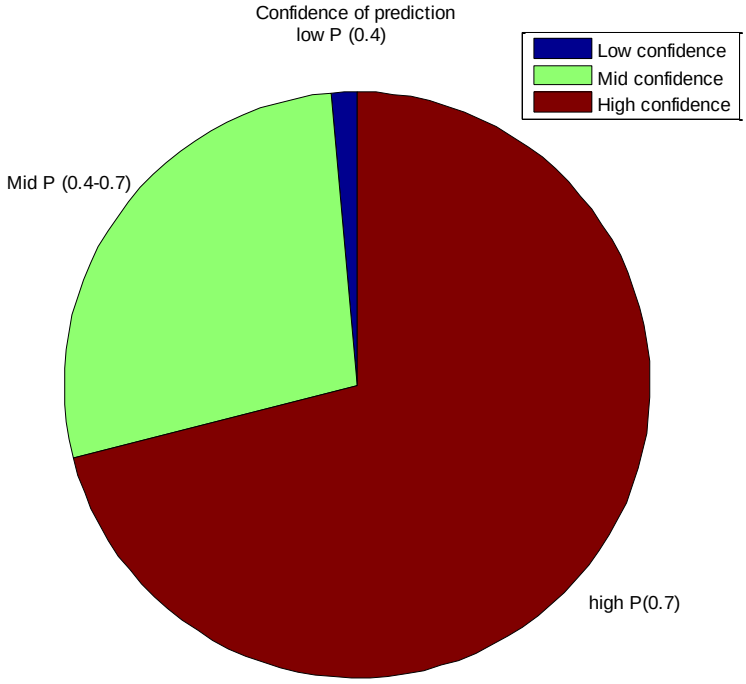

Figure 5-20; Confidence values for validation set using P-ANN.

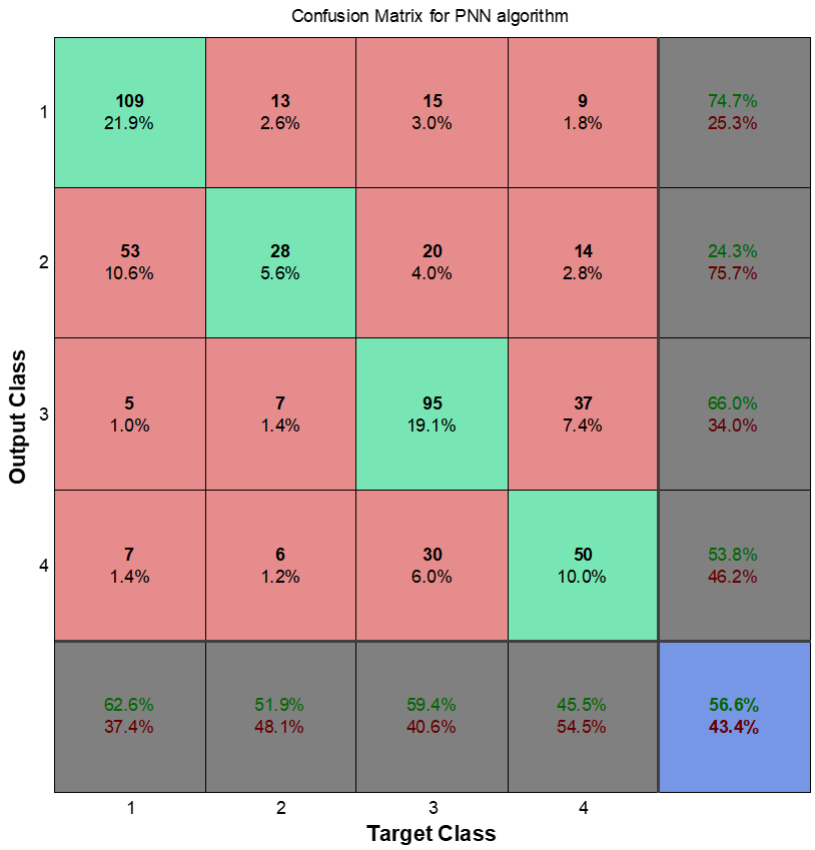

Figure 5-21: Confusion matrix for PNN algorithm.

misclassifications. Ideally, a perfect classifier would result in a confusion matrix equal identity matrix, i.e. have values equal one only on the diagonal.

Various metrics can be derived from confusion matrices to quantify the performance of machine learning algorithms. The overall accuracy is the most common performance metric used in classification problems. This metric is defined as the ratio of the number of correctly predicted examples to the total number of examples. Figure 5-21, Figure 5-22, and Figure 5-23 show confusion matrices for applications of the three algorithms to the validation set, and the overall accuracies are summarized in Table 5.5. The overall accuracy when using PNN is about 57%, which means that the fractional classification error is 43% for this dataset (Figure 5-21). This algorithm has the highest performance in finding class 1 with almost 63% accuracy and the lowest

performance for class 4 (45% accuracy). Class 4 is mainly classified as class 3, apparently because they are lithologically close and therefore difficult to differentiate in seismic data.

Table 5.5: Summary of validation accuracy for three classification algorithms.

Algorithms	Proposed ANN (P-ANN)	Regular ANN	PNN
Overall validation score	72.3	66.9	56.6
Class 1 accuracy	87.9	87.4	62.6
Class 2 accuracy	42.6	1.9	51.9
Class 3 accuracy	75	73.1	59.4
Class 4 accuracy	58.2	57.3	45.5

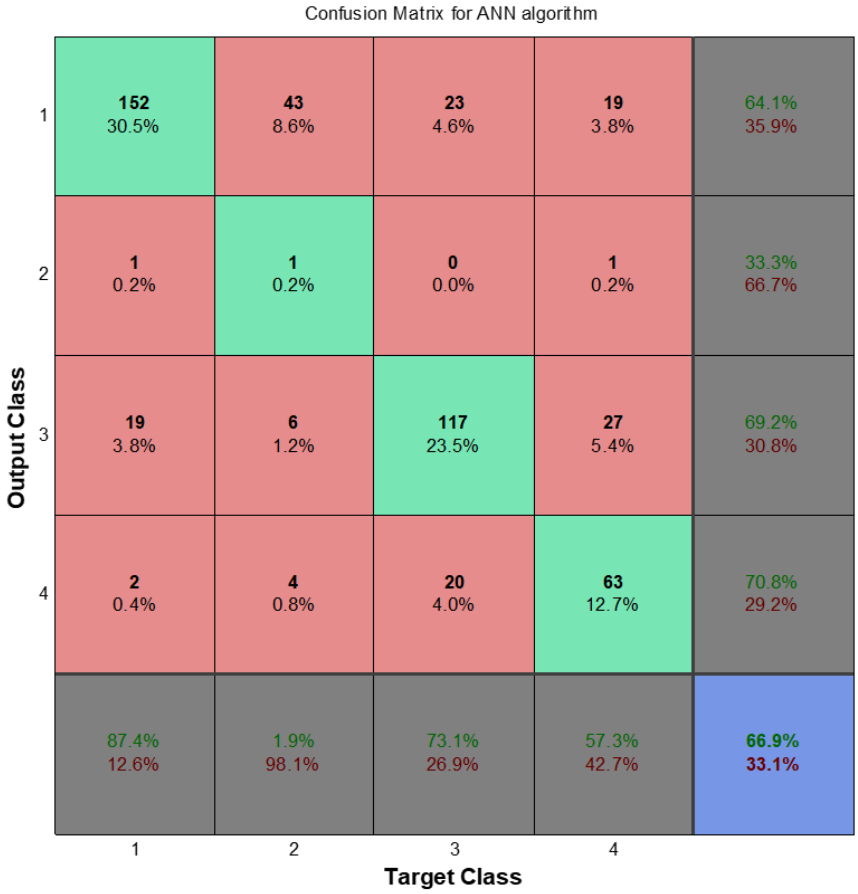

Figure 5-22: Confusion matrix for ANN algorithm.

For the ANN algorithm (Figure 5-22), the overall accuracy is almost 67%. (Table 5.5). This ANN algorithm has higher performance than PNN for sandstone and carbonate classification (classes 1 and 3) with almost 80% accuracy. It has a very low performance in classifying class 2 (2%) and mostly misclassifies it as class 1 (Figure 5-22). Since class 4 is non-reservoir, it is also crucial to classify this class carefully. With regard to this class, the ANN is more efficient than the PNN algorithm (Table 5.5).

The P-ANN algorithm (Figure 5-23) achieved the highest accuracy for validation data. The overall accuracy is about 72.3%, which is the highest among the techniques used in this study. This algorithm could be decisive in classifying class 1 with 88%, class 3 with 75%. The P-ANN also failed to classify class 2 with 43% correct predictions which is again higher than the ANN but a bit less than the PNN (Table 5.5). The main advantage of P-ANN could be its ability to classify class 4 with 58% accuracy compared to 45% for PNN.

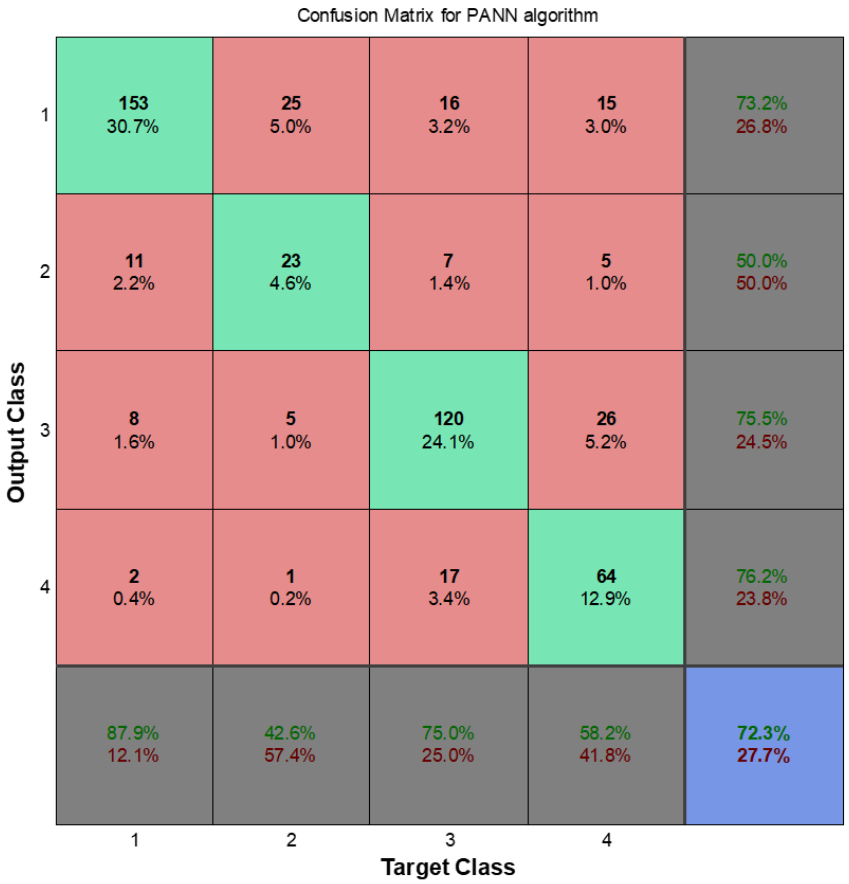

Figure 5-23: Confusion matrix when using P-ANN.

Based on the above confusion matrices and overall statistics, the above comparison shows that the P-ANN algorithm outperformed the PNN, and the regular ANN almost approaches P-ANN in terms of its ability to achieve the maximum discrimination accuracy for blind wells. Thus, the P-ANN supervised classification analysis provided a convenient and robust method, which can classify the reservoir to known classes which numerically is superior to the two other machine learning techniques.

5.4.2 Application to seismic data

To compare applications of the three classification algorithms to all seismic data, I use a cross-section of the 3-D model volume along the line shown in Figure 5-24. This selected seismic line goes through seven training and one blind well, which allows detailed evaluation of the

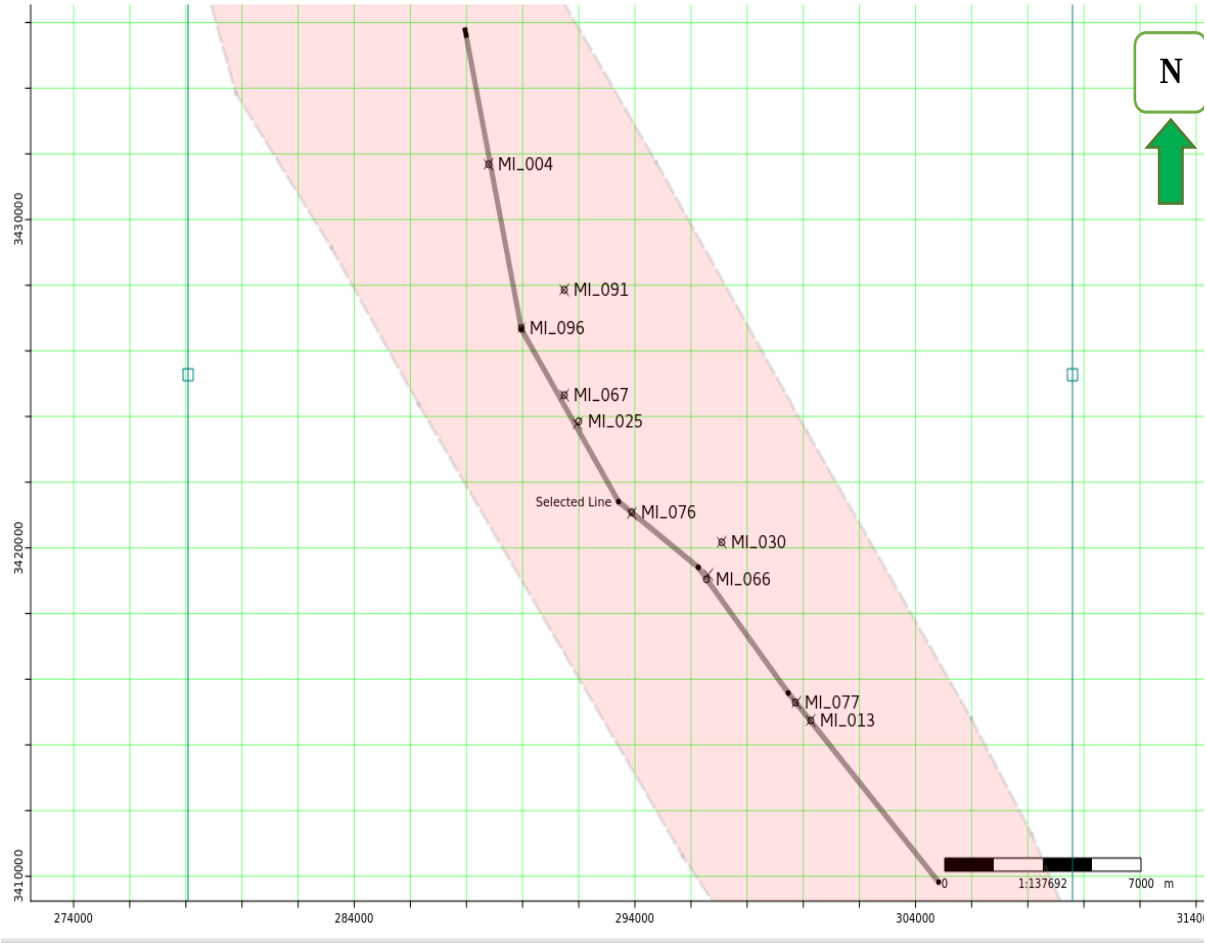

Figure 5-24: Selected line to compare the classifications in Table 5.5.

classification accuracy for each algorithm. The previously interpreted rock facies classes available at the eight well locations along this line are shown in Figure 5-25, and the resulting classifications by the three methods are shown in Figure 5-26.

Generally, the ground-truth (Figure 5-25) and all three classified seismic datasets (Figure 5-26) show two distinctive groups of classes within the upper and lower halves of the reservoir. They also show similar spatial relations between classes 1 and 2 within the upper half and classes 3 and 4 within the lower half of the reservoir. However, there are also some differences in the resulting classifications. Based on the known labels, we expect to see a near-continuous layer of class 3 at the top of the selected line (yellow in these figures), and also a continuous layer of class 4 (brown) at the bottom of the cross-section in Figure 5-26. The PANN and ANN successfully

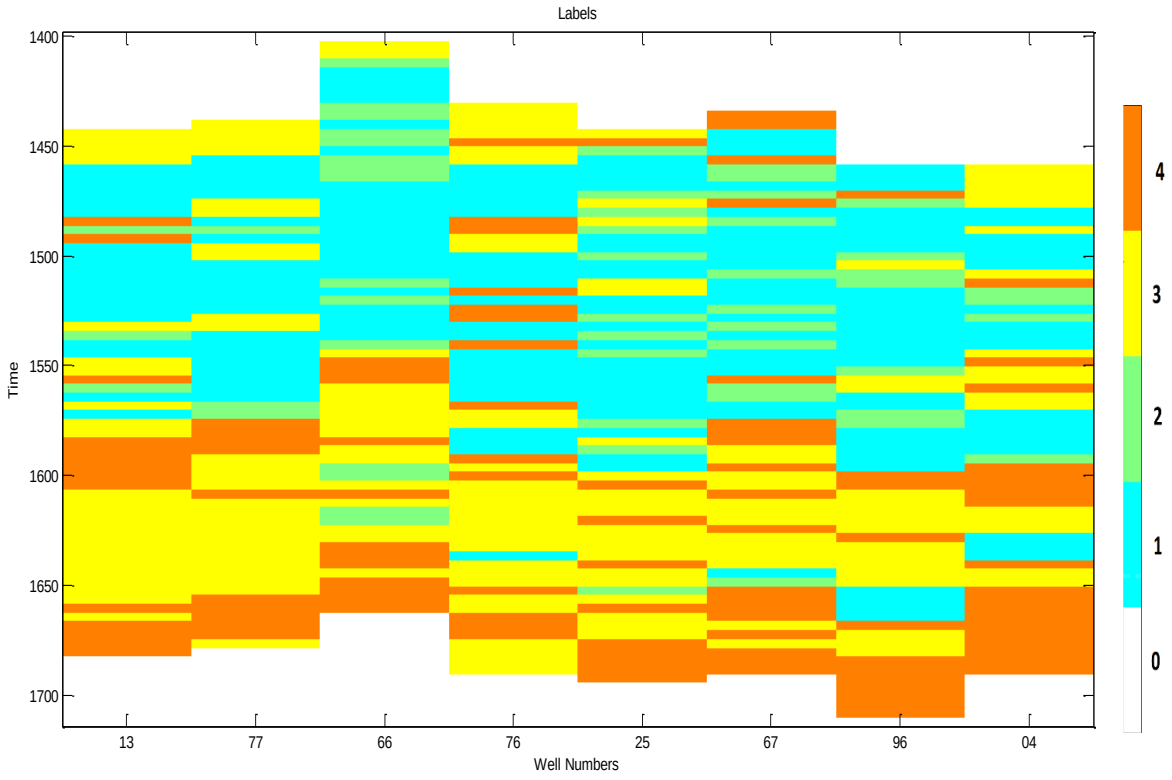

Figure 5-25: Known class labels (colors) at eight well locations. The X axis represents well numbers and Y axis is the two-way seismic reflection time in milliseconds.

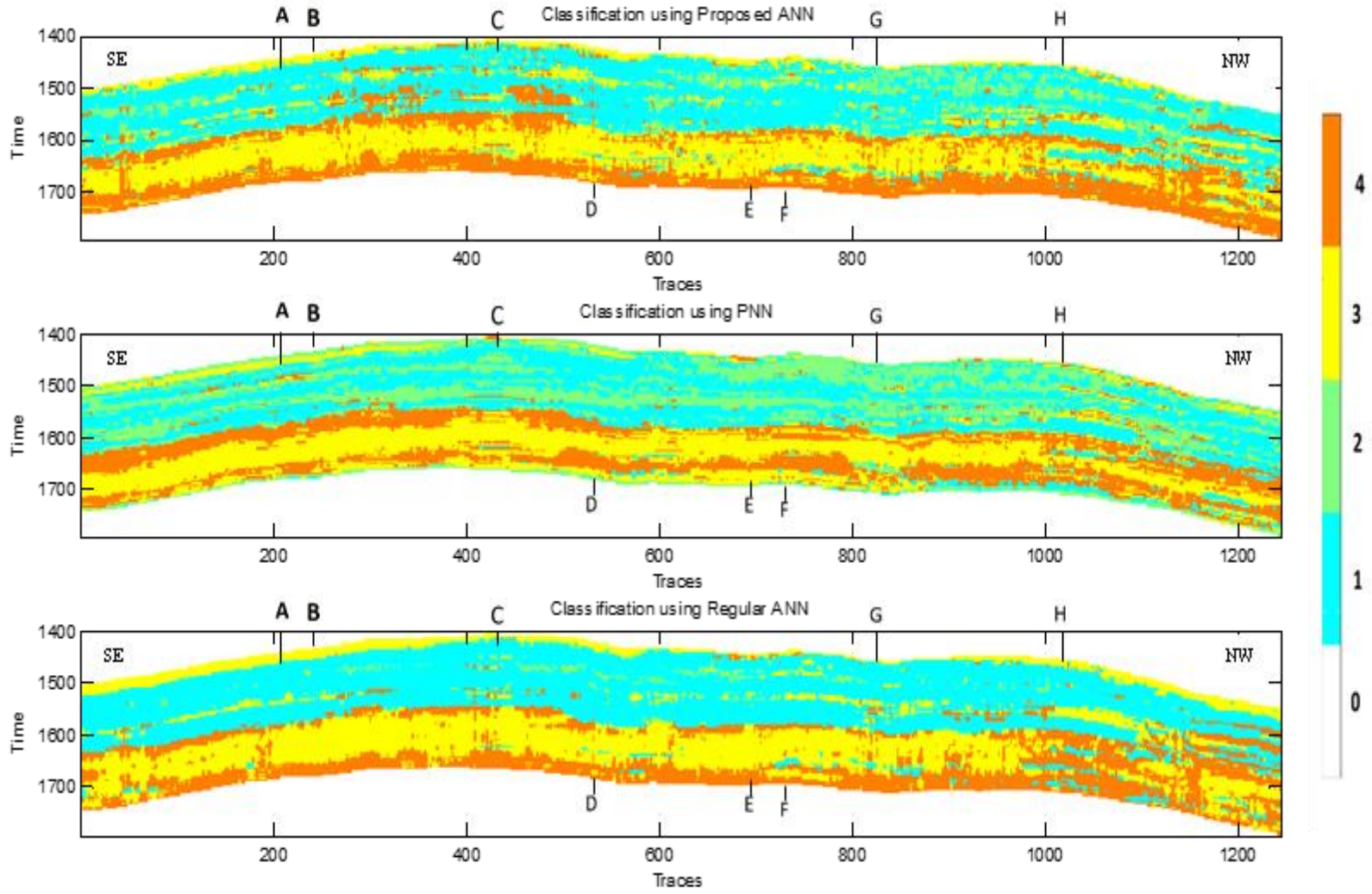

Figure 5-26: Classification cross-sections along the selected line using the P-ANN, ANN and PNN algorithms (headings of the panels). Color bar represents class numbers, as in Figure 5-25.

classify these zones with high accuracy (top and bottom panels in Figure 5-26), but the PNN results show neither continuous layer of class 3 at the top nor continuous class 4 at the bottom of the section (middle panel in Figure 5-26). At the same time, layering of seismic facies 1 and 2 within the upper half of the reservoir appears to be more pronounced and contiguous in the PNN classification.

To compare the predicted classes with known classifications in more detail, Figure 5-27 shows an overlay of the classification of well 004 with a small segment of the image around this well (Figure 5-26). For P-ANN results, this comparison shows a reasonable match the known and the surrounding predicted classes, although with some mismatches within the bottom part of the section. For PNN, there also is a reasonable match at this location, although the pattern of predicted

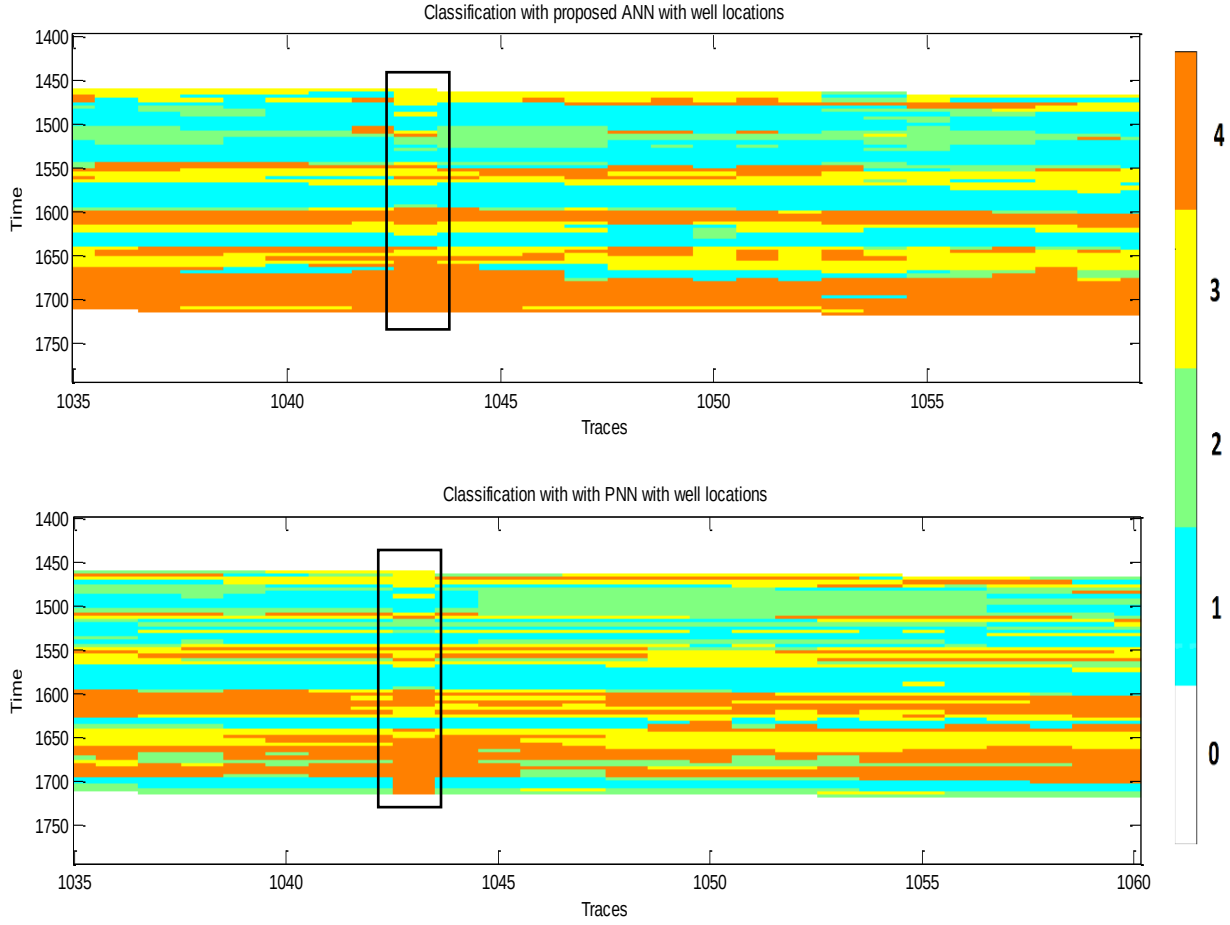

Figure 5-27: Zoom-in of the P-ANN and PNN images in Figure 5-26 near well 004 with known classes at this well location inserted (black rectangles).

layering appears to be of higher frequency than the one within the well. A layer of class 1 (cyan) is predicted at the bottom of the imaged zone, which is nevertheless absent in the well. In addition, the final classifications in Figure 5-26 consist of integer numbers determined from the estimated probabilities P_k of each of the classes and confidence levels C of the prediction (chapter 2). These probabilities contain important additional information which is, unfortunately, rarely utilized for characterizing hydrocarbon reservoirs. Nevertheless, this information may have a critical impact on interpretation. Figure 5-28 shows the confidence levels of the three classifications (Figure 5-26), also plotted in the form of continuous pseudo-log cross-sections.

As shown in these plots, the confidence values show clearer and more consistent layering within the reservoir than the images of interpreted classes (Figure 5-26). The confidence levels for P-ANN are higher than for PNN and particularly for ANN algorithms. However, the meanings of predictors P_k in the PNN algorithm are significantly different from outputs of an ANN, and therefore the confidence thresholds should likely be interpreted differently for these methods.

The estimated confidence levels can also be used to create enhanced and easier to interpret images of the data, any seismic attributes, of the classification. For example, to emphasize the subsurface areas with reliable classification, I create a filter (mask) for the image displaying the data only in areas with confidence $C \geq 70\%$ when using the ANN-based methods or $C \geq 50\%$ when using the PNN. The filtered displays of classifications are shown in Figure 5-29. This image emphasizes the layered character of the subsurface. Compared to Figure 5-26, the filtered section allows making several significant observations. The filtered classifications for P-ANN and ANN are not showing class 2, which means that all of its identification lies below the 70% confidence. For PNN, the 50% confidence threshold allows seeing all four classes nicely, except at the bottom of the section. Also, a significant observation with PNN is that class 4 at the bottom of the plot is intermittent due to lower confidence.

Also, notably in the filtered PNN image (Figure 5-29, middle), class 2 at the top and bottom of the section shows high confidence, but it is misclassified instead of the expected classes 3 and 4, respectively. This misclassification with high probability shows that despite the PNN algorithm creating images which look more geologically realistic, it still fails to classify the electrofacies properly.

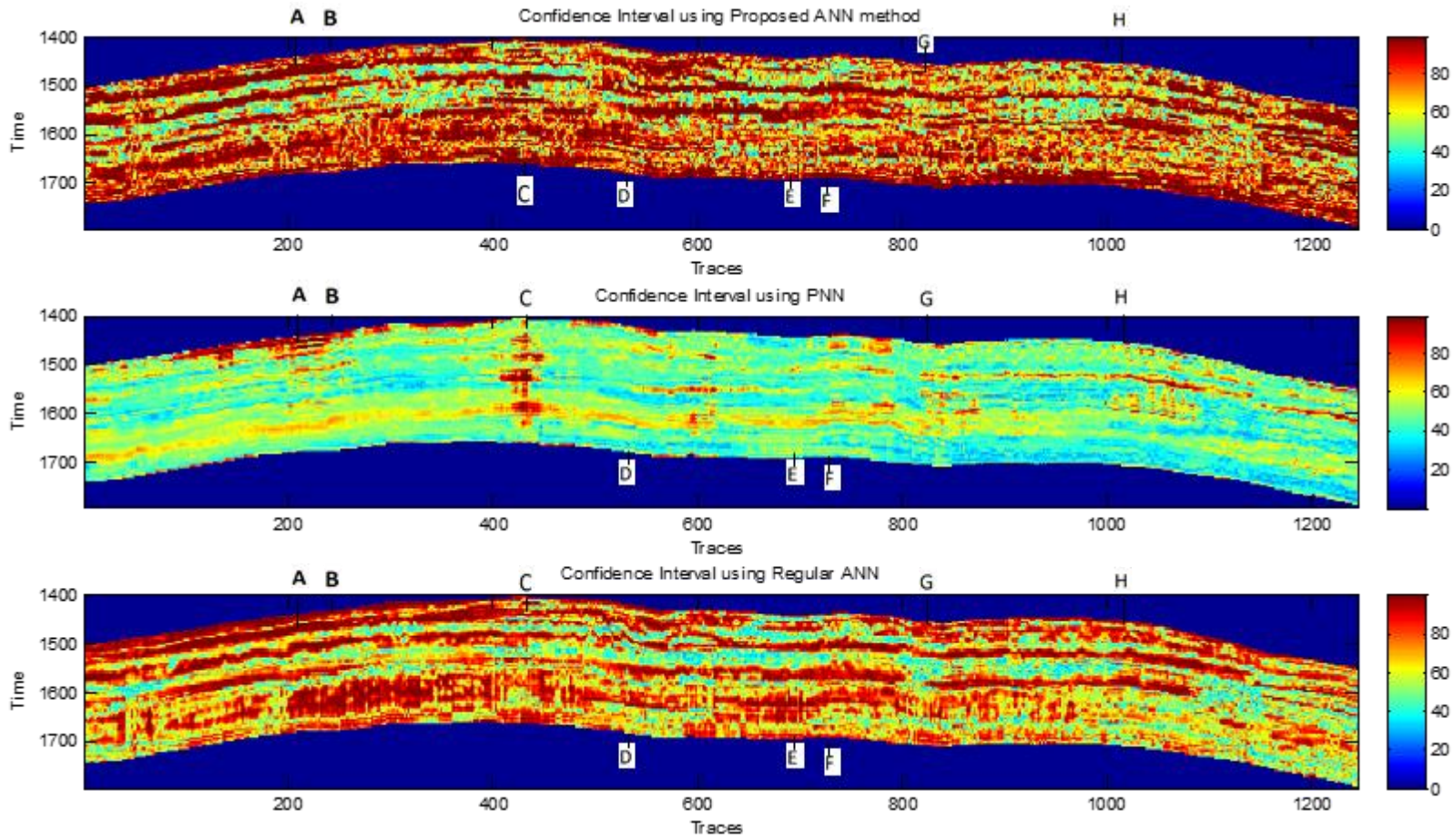

Figure 5-28: Confidence interval for three algorithms; P-ANN (top), PNN (middle) and ANN (bottom).

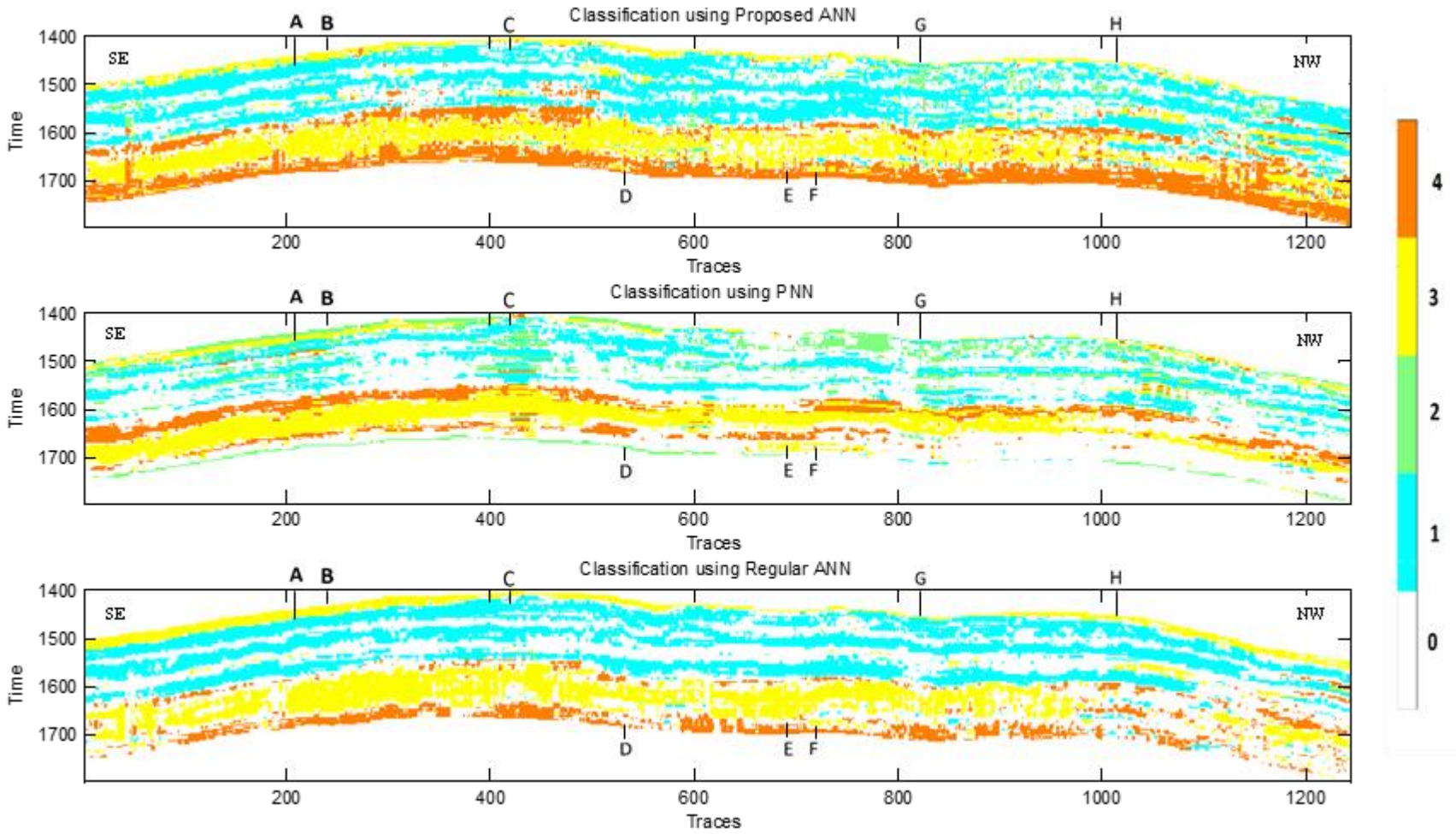

Figure 5-29: Filtered seismic facies classification using probability of the predicted class (The threshold for probability for ANN and proposed ANN is 70% and for PNN is 50%).

CHAPTER 6: Porphyry Prospectivity in Quesnel Terrane, British Columbia

This chapter describes a different application from the rest of the thesis application of machine learning to applied geology and geophysics. I use an Artificial Neural Network to explore the mapped occurrences of mineral deposits within the Quesnel Terrane (British Columbia) and to extend the prospectivity predictions to adjacent areas. To form the predictions, geophysical data are combined with time-constrained geological, structural, and tectonic data to produce quantitative estimates of the probability for the target mineral.

This chapter is based on my contribution to the report by the University of Saskatchewan "4-D Integration" team for the 2022 Frank Arnott - Next Generation Explorers Award by the Prospectors and Developers Association of Canada as the geophysicist and data analysis (https://www.pdac.ca/members/students/faa/2022-challenge). Our team received the Grand prize after competing in the finals against five teams from Canada, France, Peru, and Great Britain. In this project, Dène Tarkyth applied mineral systems expertise to simplify the categories of mineral occurrences and geological features in a manner appropriate for machine learning and assisted with geospatial information systems (GIS) and plotting tasks. My contribution to this project consisted in data integration and machine-learning data analysis and prospectivity modeling, as described in this chapter.

In the following section 6.1, I briefly describe the geology of the study area and its relation to economic prospectivity for minerals. In section 6.2, I outline the methodology of this project, in section 6.3, I describe the data, and in section 6.4, I assess the data for mutual correlations and relations to the target prospectivity. After feature analysis, in section 6.5, I apply neural-network based pattern recognition to evaluate a probability-type estimator for potential occurrences of porphyry, classify the data into 'prospective' and 'non-prospective', and produce a predictive prospectivity map. Further in section 6.5, I measure the classification accuracy and discuss the performance of the algorithm.

6.1 Geology background and goals of this project

The study area covers an about 400-km long segment of the Quesnel Terrane, in the Cariboo area of the Cariboo Chilcotin Coast region of British Columbia (BC), Canada (Figure 6-1). This terrane is rich in gold porphyry deposits, which are locally covered by dense layers of glacial deposits. In 2007, the BC Geological Survey started a major program of developing a modern geoscience data base over the Quesnel Terrane in order to facilitate the discovery of additional resources. This program contributed several geological, geophysical, and economic datasets on which innovative methods of discovery and data integration could be tested within the framework of the Next Generation Explorers competitions. An example of such data integration is presented in this chapter. My goal in this project is to utilize machine learning for evaluation of economic prospectivity for porphyry-related minerals using all available data.

The Quesnel Terrane (also called Quesnel Trough, or Quesnellia) and the Stikine Terrane (also called Stikinia) extend for about 2,000 km along the axis of the Canadian Cordillera and represent two Mesozoic-age volcanic arcs preserved within western Canada. These terrains are joined at their northern ends, but within the study area, they are separated by the Cache Creek Terrane consisting of relics of the Tethyan ocean basin and oceanic arc rocks.

The calc-alkalic and alkalic porphyry Cu-(Mo-Au) and Cu ± Au-Mo-Ag deposits occur in the Stikine-Quesnel arc terranes, which constitute the most significant economic metal deposits in British Columbia. Most of the economic metal endowment was emplaced during the mineralizing epoch is a 6-m.y. pulse centered on 205 Ma. There are several active mines in the study area, with the mines at Mt Milligan in the north and Mt. Polly in the south shown in Figure 6-1.

In the central part of the terrane, the quaternary cover is extensive and is estimated to reach over 50-m thicknesses locally. This sedimentary cover complicates geological mapping and sometimes reduces the effectiveness of geochemical techniques. For these reasons, it is important to integrate structural geology and geophysical data in order to evaluate the economic potential of ore deposits.

Tertiary surficial placers in Quesnel may be evidence of eroded orogenic gold. The Cariboo and Barkerville gold districts in the southern part of the Quesnel have a historical production of

approximately 1.2 million ounces of gold from lode deposits and 3.8 million ounces gold from placer deposits.

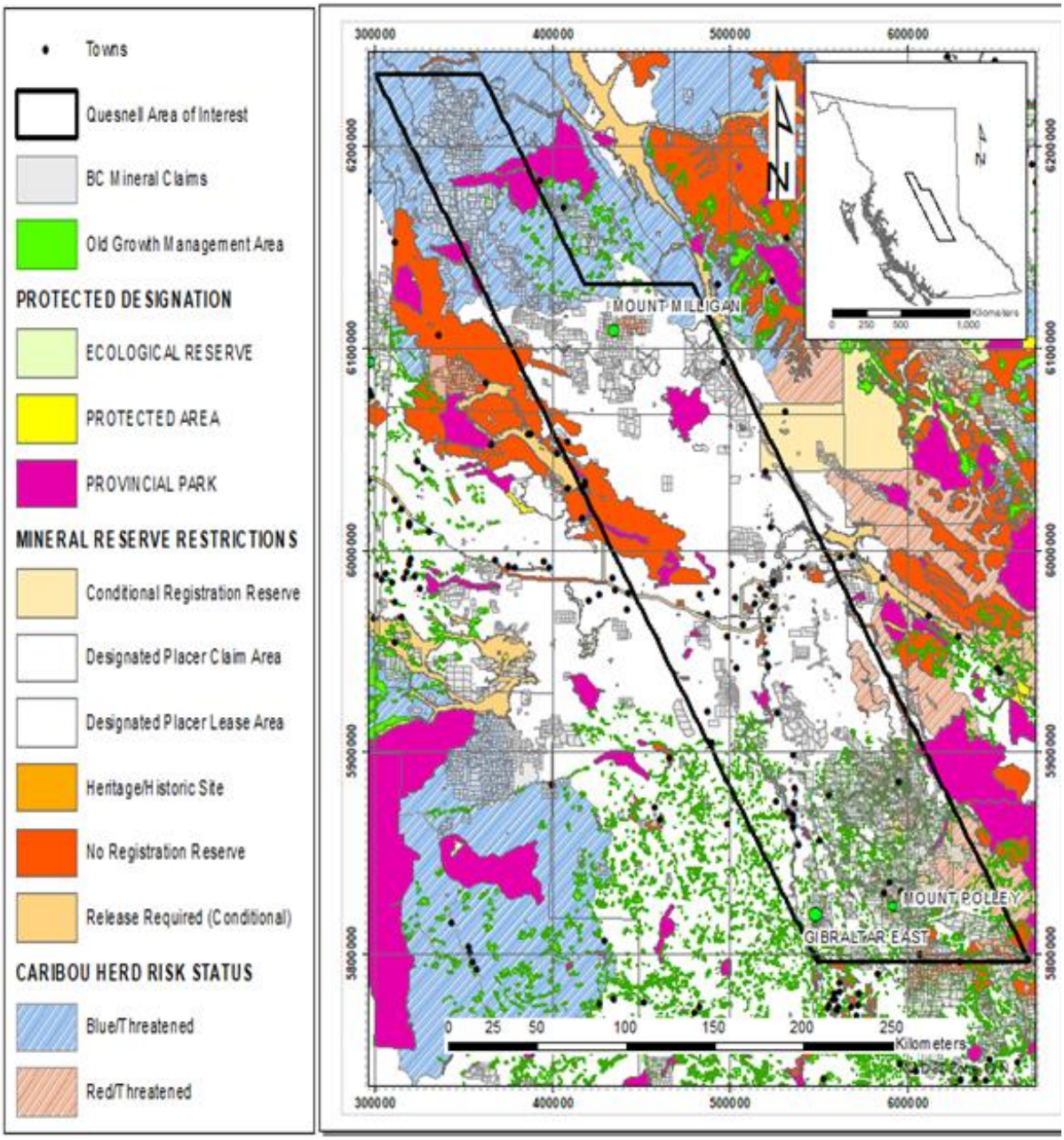

Figure 6-1: Study area in British Columbia, Canada. The map axes are the UTM northings and eastings in kilometers.

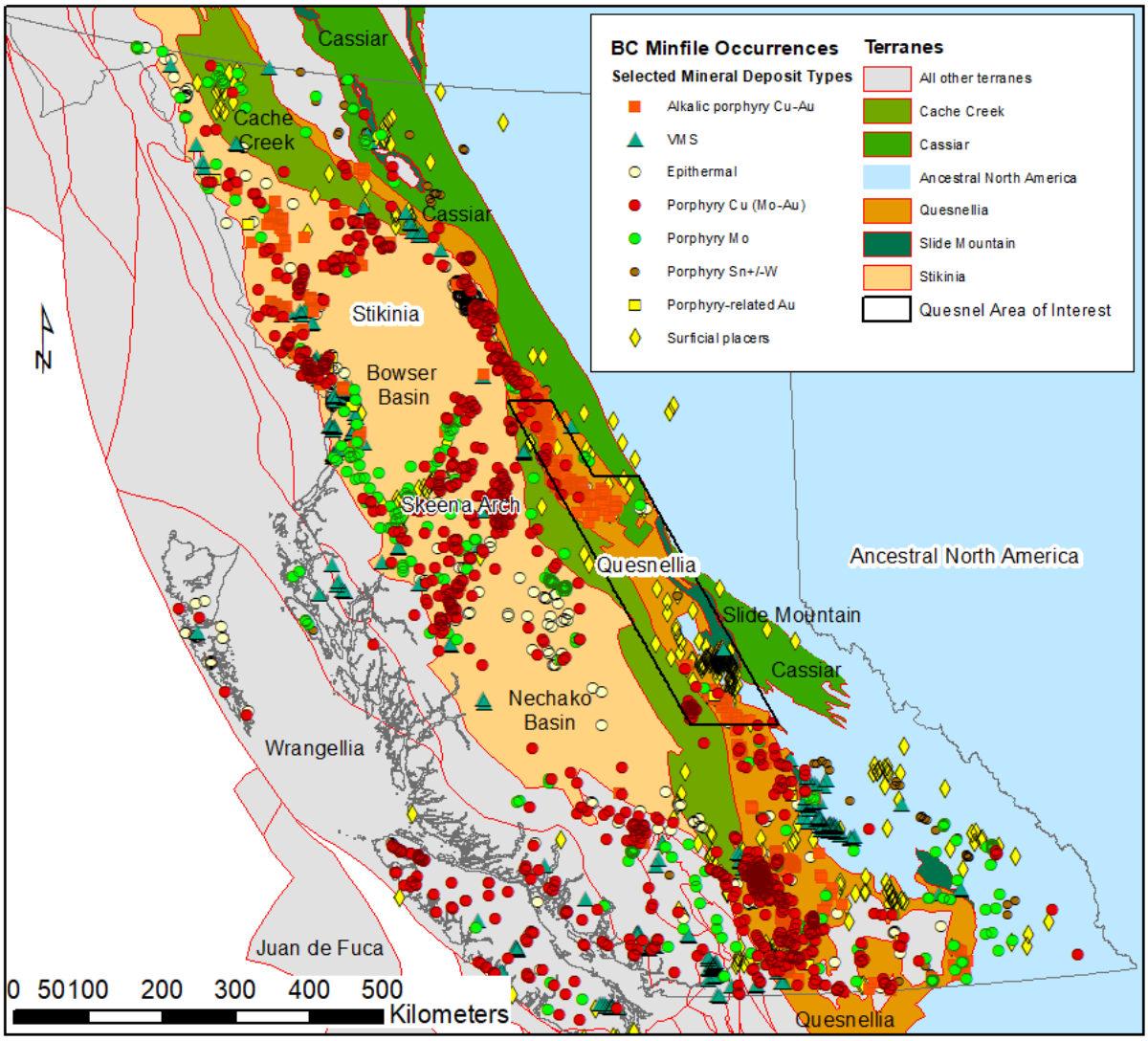

Figure 6-2; Regional geology of the central and northwestern British Columbia, with known porphyry mineral deposits (legend). Major tectonic units are labeled.

Figure 6-2 shows the porphyry-related mineral deposits in the study area. The porphyry deposits dominate the area of interest, and porphyry-related mineral systems express the best targets for explorers aiming for economic deposits with attractive size and grade potential.

The ages of mineralization are important for estimation of the potential mineral types and prospectivity. The principal porphyry Cu ± Au–Mo deposit formation occurred was happened in Quesnel and Stikine terranes from 210 – 175 Ma, and the major gold deposits appeared in Stikinia at around 190 Ma, during the switching from tectonic compression to extension. These time

relations are incorporated in the analysis by creating the relevant structural and tectonic data to feed into ANN.

The geophysical data in this study is provided by the Advanced Geophysical Interpretation Centre at Mira Geoscience. Maps of 3-D density contrasts, magnetic susceptibility, and inverted subsurface conductivity are the main geophysical data used in this study. These data were derived from multiple airborne gravity, airborne total-field magnetic, and airborne electromagnetic (EM) surveys, respectively. The data were collected as part of the Geoscience BC's QUEST Project, which was a program of regional geochemical and geophysical surveys designed to attract the mineral exploration industry to an under-explored region of British Columbia between Williams Lake and Mackenzie (https://www.geosciencebc.com/major-projects/quest).

6.2 Methodology

Ore prospectivity is an old and immensely important subject in geology, and a vast literature on this subject exists. In this chapter, I do not intend to provide a comprehensive review of this literature but only give a brief summary related to applications of artificial neural networks (ANN) to data-driven mapping of mineral prospectivity.

In areas with large numbers of prospects, the weights of evidence (WofE) modeling method is often applied to map the mineral potential (Carranza, 2004). The WofE method is based on a probabilistic model of prospectivity, which is also used in the present study. The area of interest is subdivided into equal-area cells, and probability of prospectivity P for some deposit D is sought for each cell. The estimation starts by measuring the prior probability as $P_{prior} = N(D)/N_c$, where N_c is the number of randomly selected cells and $N(D)$ is the number of observations of D among them. This prior probability is updated based on statistical correlations between the occurrences or non-occurrences of D (denoted $N(D)$ or $N(\bar{D})$, respectively), and similarly occurrences or non-occurrences of some additional evidence B ($N(B)$ or $N(\bar{B})$). This evidence can be, for example, proximity to a fault or gravity anomaly. The update of probability is based on the positive and negative weights of evidence defined as logarithms of likelihood ratios:

$$W^{+} = \log\left[\frac{N(B \cap D)/N(D)}{N(B \cap \bar{D})/N(\bar{D})}\right], \quad W^{-} = \log\left[\frac{N(\bar{B} \cap D)/N(D)}{N(\bar{B} \cap \bar{D})/N(\bar{D})}\right]. \tag{6.1}$$

The spatial pattern of weight W^{+} shows how an observation of *B* increases the likelihood of *D* in a given area, and W^{-} shows how $\bar{B}$ increases the likelihood of non-observation $\bar{D}$. Carranza (2015) suggested that a modification of this approach called the evidential belief modeling achieves similar performance for smaller sample sizes but is more stable with respect to missing data. Ford et al. (2016) further compared these two methods with fuzzy logic models.

Machine-learning (ML) methods attempt estimating the posterior probabilities *P*(*D*) directly from geochemical, geophysical, and other evidence data without evaluating the likelihood ratios (eq. 6.1). Harris and Pan (1999), Rodriguez-Galiano et al. (2015), and Ghezelbash et al. (2019) reviewed several ML techniques to obtain predictive maps for potential mineral deposits: ANN and probabilistic neural networks (PNN), generalized regression neural network (GRNN), radial basis function networks (RBFN), logistic regression, random forest (RF) algorithms, regression trees, discriminant analysis and support vector machines (SVM), and fractal models (Haykin, 2009). Based on examples of epithermal gold prospectivity in Spain, Rodriguez-Galiano et al. (2015), suggested that the RF methods the other ML techniques in terms of stability and success rates. At the same time, these authors suggested that when the ore deposit evidences are scarce, all ML techniques perform comparably. Maepa et al., (2021) combined the spatial statistical methods such as the WofE, SVM, and RBFN (chapter 2 in this thesis) for mapping gold deposits in Ontario. Yin and Li (2022) reported application of statistical techniques called Bayesian optimization and ensemble learning aimed at deriving of the probability distribution from mapped geological, geochemical, and geophysical data. Qin et al. (2021) used RF methods for 3-D modeling of ore prospectivity in eastern China.

As noted with regard to GRNN, PNN, and RBFN in chapter 2, many of the algorithms called neural networks in the preceding paragraph actually represent interpolation, spatial filtering and correlation analysis methods. Cluster analysis of the prospective and no-prospective points is a key part of the GRNN method (Haykin, 2009). Abedi and Norouzi (2012) directly used cluster analysis to prepare copper prospectivity maps.

6.3 Data

Similar to chapter 5, to perform data classification using an ANN, I used ten input features, which included values extracted from four geophysical maps, two types of geological data, and four measurements of the crustal structure. The geophysical attributes from Mira Geoscience BC Modelling Report (2009) were:

1) TMI (total-field magnetic intensity),
2) Electric conductivity,
3) Magnetic susceptibilities at the sea level, and
4) Density contrast at the sea level.

These data are shown in Figure 6-3.

The density at the sea level was extracted from 3-D inversion of terrain-corrected QUEST airborne gravity data acquired by Sander Geophysics AIRGRAV system. Flight lines spaced 2 km apart were oriented east-west across the survey area. Some more detailed data were acquired using 1-km flight-line traverses in the areas of the Gibraltar, Cariboo Bell and Mount Milligan deposits.

The QUEST airborne magnetic data were collected from multiple surveys by Mira Geoscience. The data were processed to the total magnetic field (TMI above), which was further inverted for a 3-D susceptibility model of the earth on a 500 × 500 × 250 m grid using the UBC-GIF MAG3D inversion code (Mira Geoscience BC Modelling Report, 2009).

Therefore, although the magnetic susceptibility is estimated from TMI data, it can be viewed as a different attribute emphasizing the lateral variation of the deeper geology. This use of secondary attributes is analogous to deriving 25 attributes from seismic reflection records in chapter 5.

The magmatic rocks are categorized by different ages (Tertiary, Cretaceous, younger Jurassic, Triassic Jurassic), as shown in Figure 6-4. These data were taken from multiple studies focusing on intrusive and volcanic rocks (Yousefi and Carranza, 2015; Pazand and Hezarkhani, 2015). Intrusive rocks are among of the key features of finding porphyry-Cu prospects because intrusive rocks provide the heat sources for this type of porphyry mineralization

Density contrast

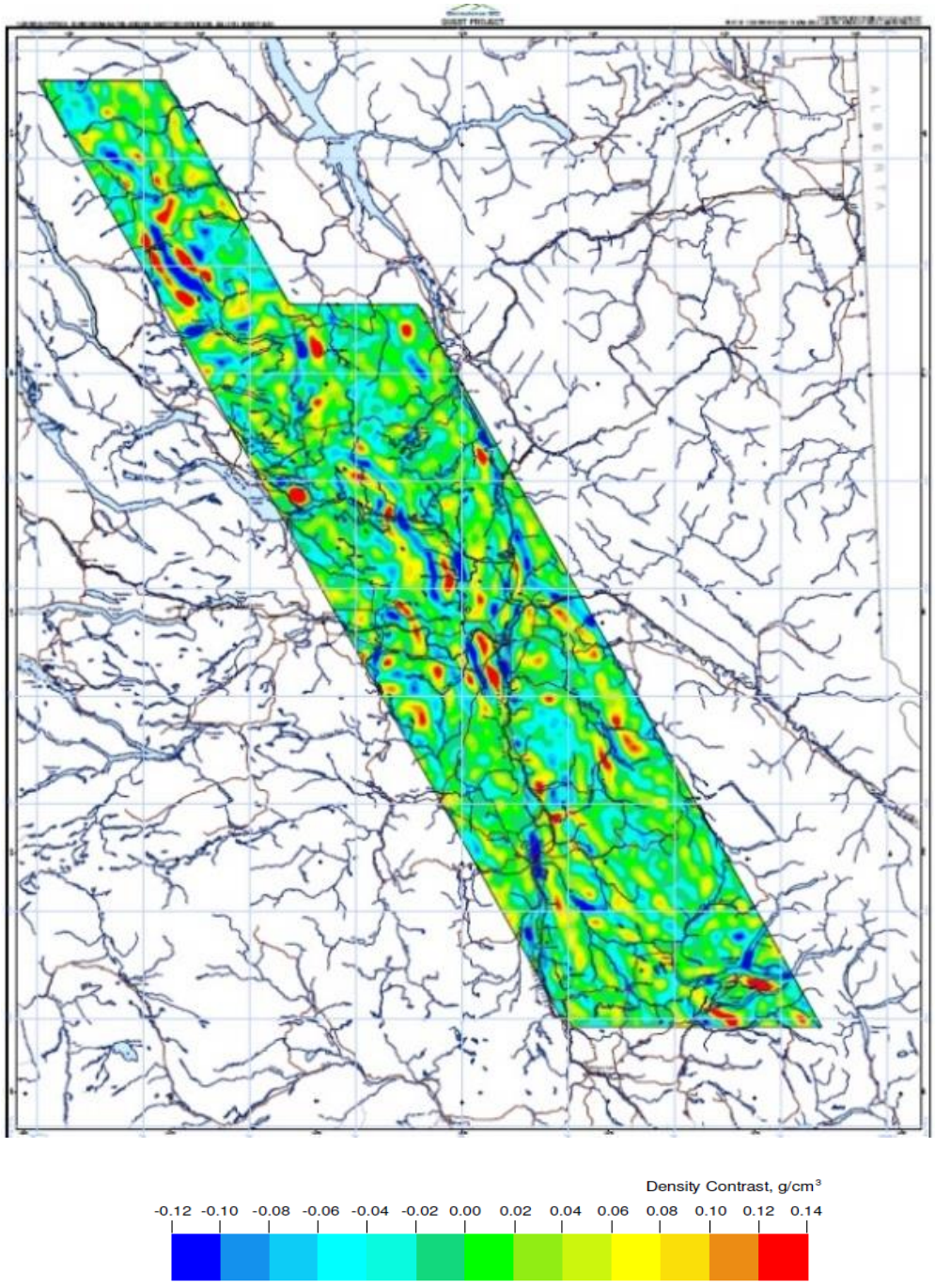

Figure 6-3: Geophysical data used in this study (Mira Geoscience report, 2009).

Electric conductivity

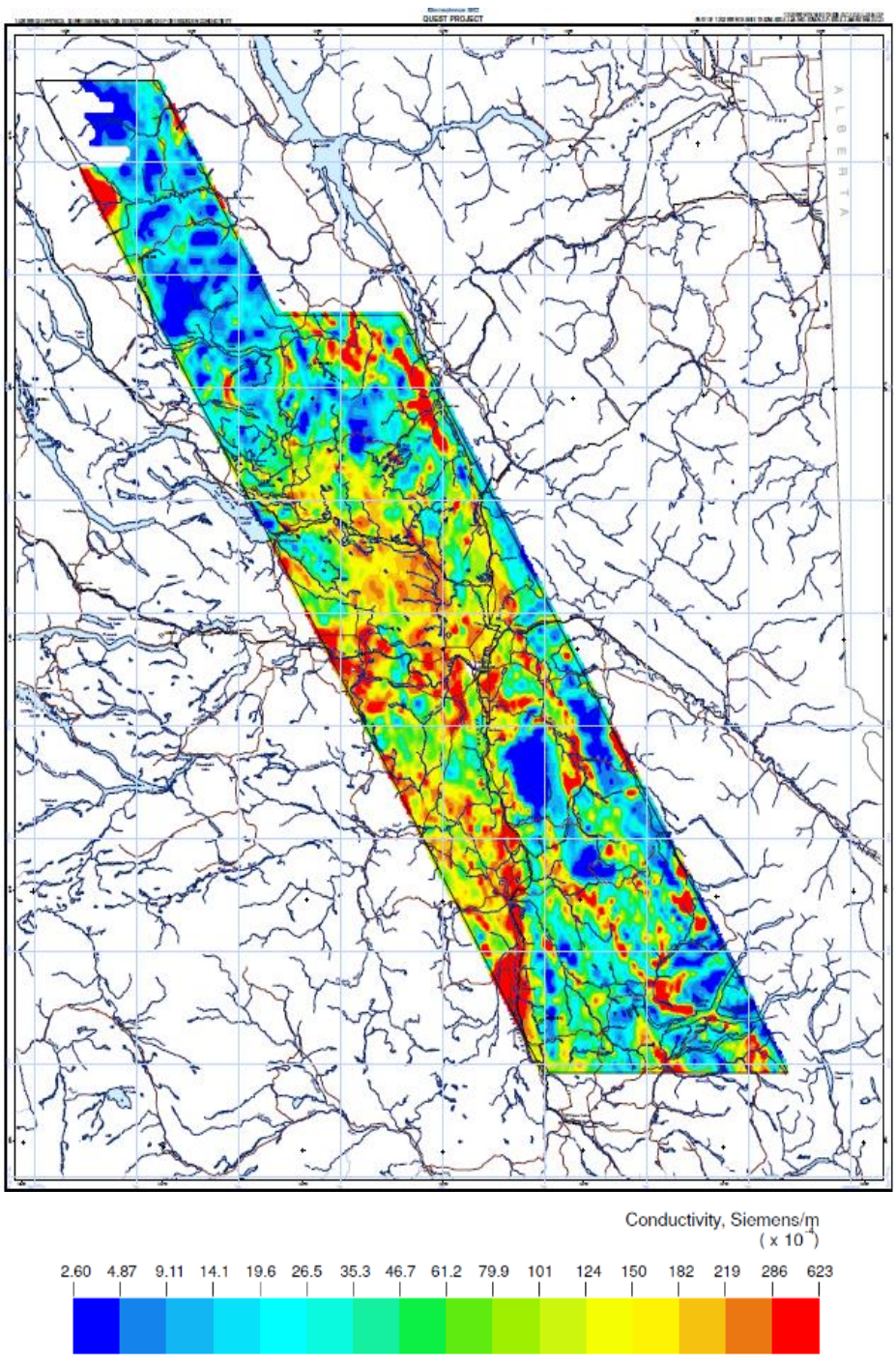

Figure 6-3; Continued.

Magnetic susceptibility

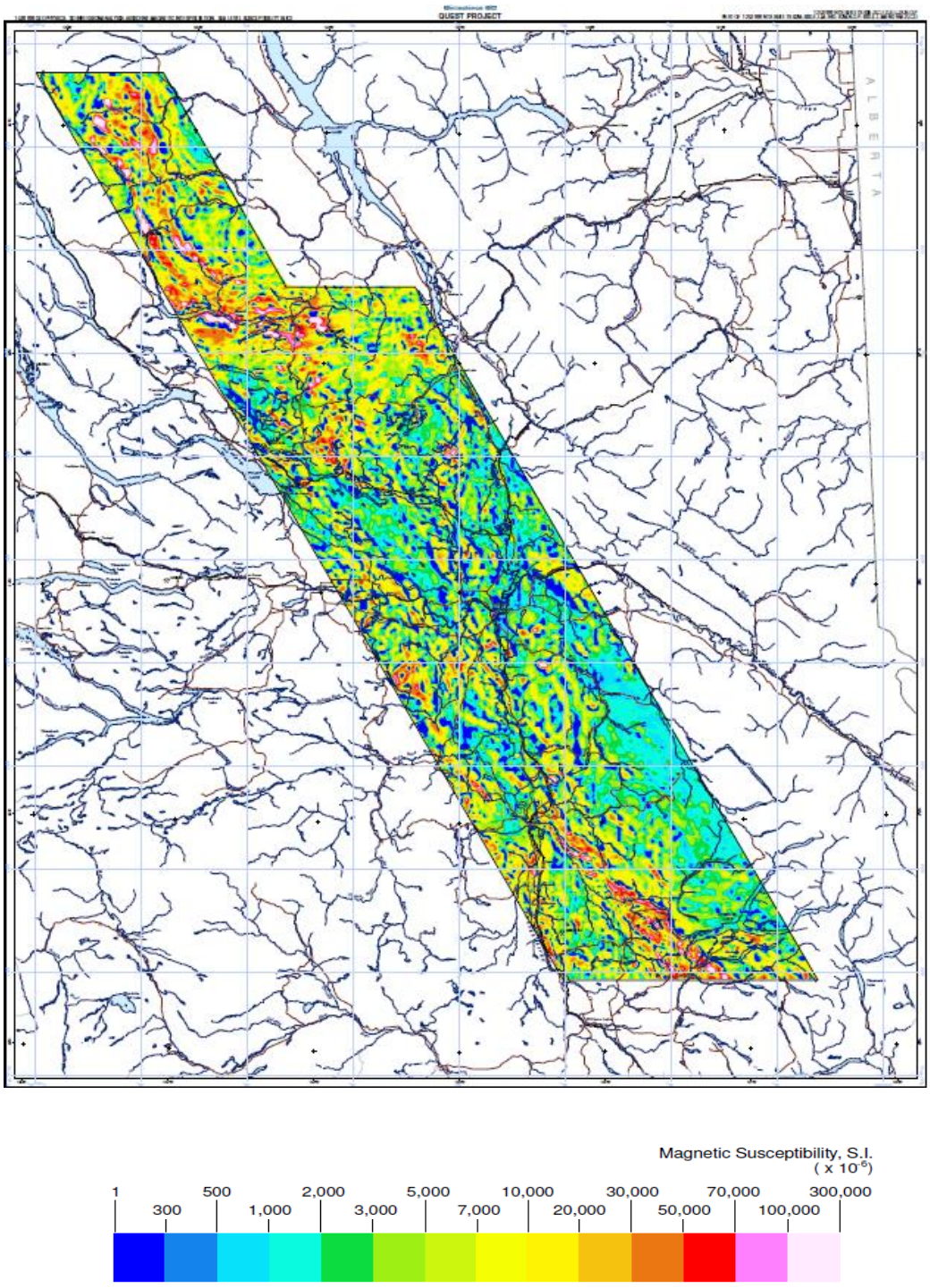

Figure 6-3: Continued.

Total magnetic field

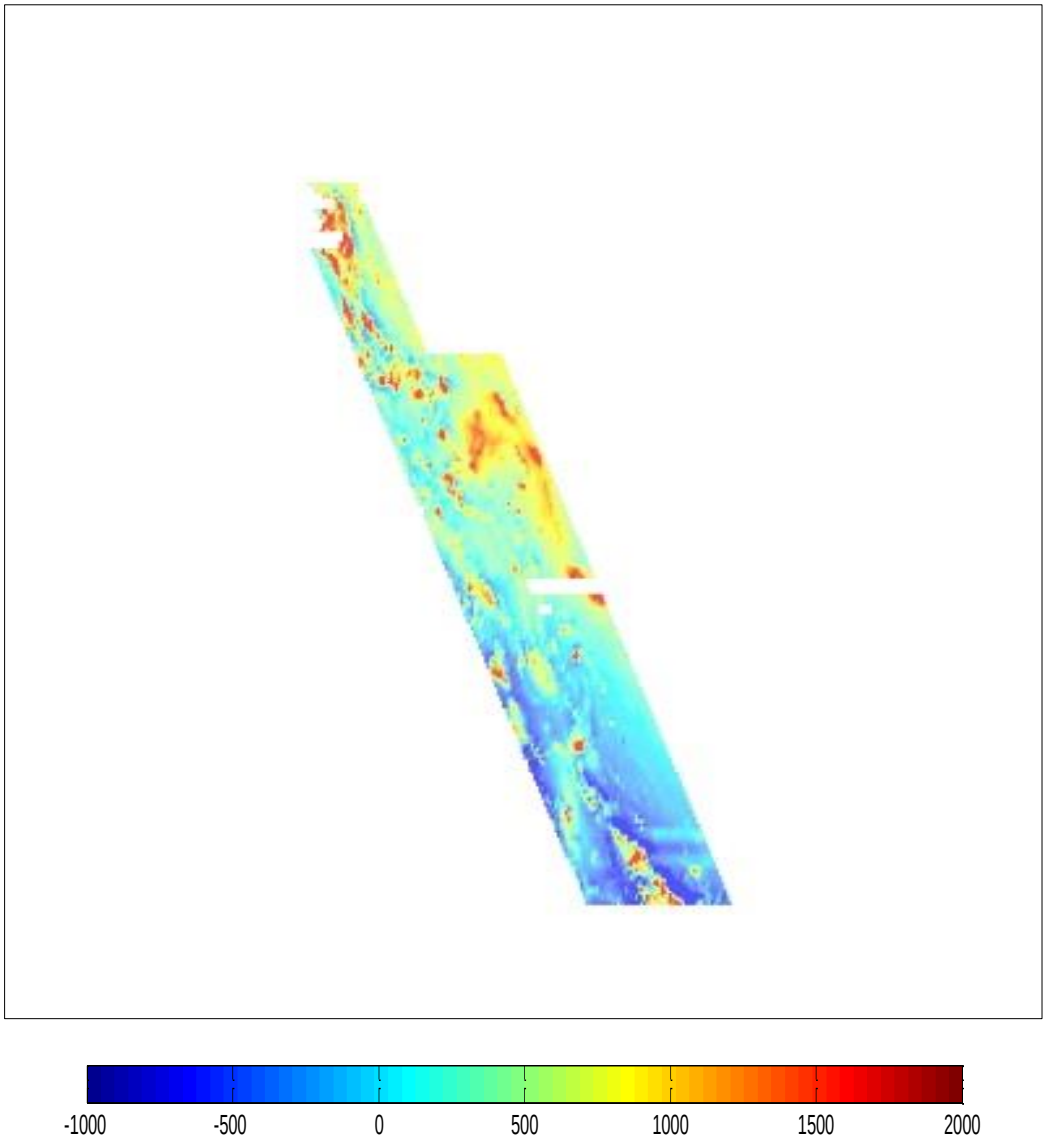

Figure 6-3: continued.

The two geological features were related to age attribution:

5) Intrusive age classes ranked from youngest to oldest, and
6) Volcanic age classes ranked from youngest to oldest.

Combining age attribute to intrusive and volcanic classes is a new way to perform data integration.

The remaining four structural features are calculated based on tectonic switches. At a regional scale at which there exist reasonable paleo-reconstruction models, an explorer could look for brief tectonic switches, considering the hypothesis that magmatism of this age would be most prospective for significant porphyry deposits. These switches are the tectonic triggers often invoked as critical factors in diverse mineral system models. Thus, in Quesnel area, a quick switch from compression, extension and again compression happened between 190 ma and 180 ma which

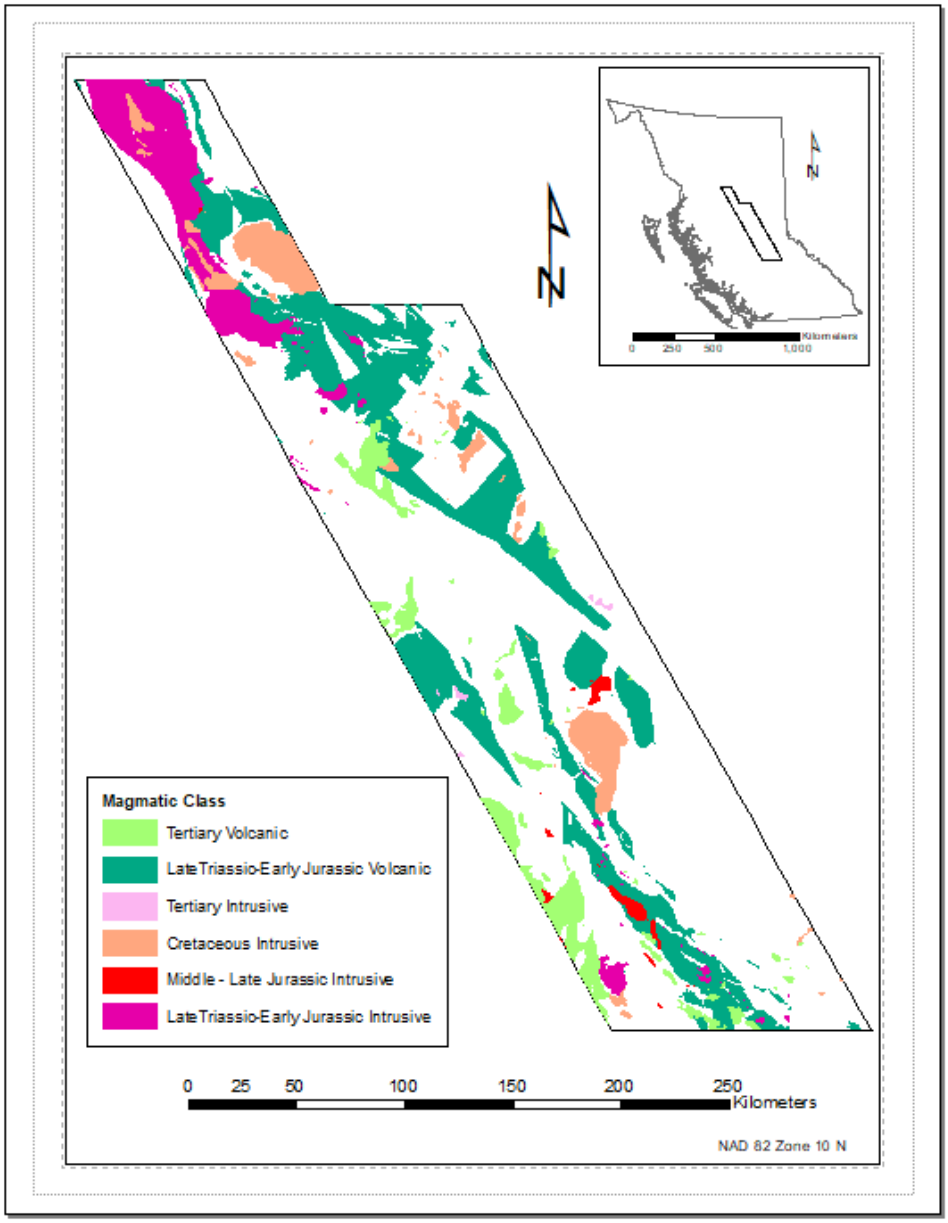

Figure 6-4. Outcrops of volcanic and intrusive rocks with age constraints.

approximately coincident with highest Cu and Au endowment. Therefore, the structural features input into the ANN are:

7) Distance in metres to the nearest fault in the regional fault database (all faults are part of database);
8) Direction (relative azimuth between 0 and 180 degrees) with respect to the orientation of the nearest fault;
9) Distance in metres to the nearest fault in the subset of faults that go from compression to extension and back into compression at ~190 Ma;
10) Direction (similar to 9)) from cell centroid to the nearest fault in the subset of faults that go from compression to extension and back into compression at ~190 Ma.

The faults used to evaluate these structural features are shown in Figure 6-5.

With regard to the prediction of porphyry at depth, different geophysical measurements give different depths of investigation. The depths of the estimated density contrast and magnetic susceptibility reach to about 2500 m below the sea level, but 3-D conductivity models are limited to about 700 m depths of investigation**.**

The combined observed data were geographic locations for which the attributes 1) through 10) above are evaluated and class labels, which were assigned values equal one for known porphyry-related mineral occurrences and label = 0 for no mineralization. The total number of data points (rows in this spreadsheet and selected cells on the map) was 455 including:

I. 142 non-prospective points which contain some kind of metallic mineral occurrences (Figure 6-2). Some geological processes or features characteristic of porphyry deposits may occur at these locations. Geophysical or geochemical evidence at these locations may look like potentially porphyry-related, although caused by different or unknown deposit styles. Therefore, target label = 0 was assigned to these points but false positive predictions are not ruled out at these locations.
II. 86 non-prospective points at which no mineralization of any kind has been observed or reported. I treated these points as a mix of true negatives and no data and assigned label = 0.

III. 227 prospective points with known porphyry deposits (Figure 6-2). These points were assigned label = 1.

A part of the resulting spreadsheet of input and output data is illustrated in Figure 6-6.

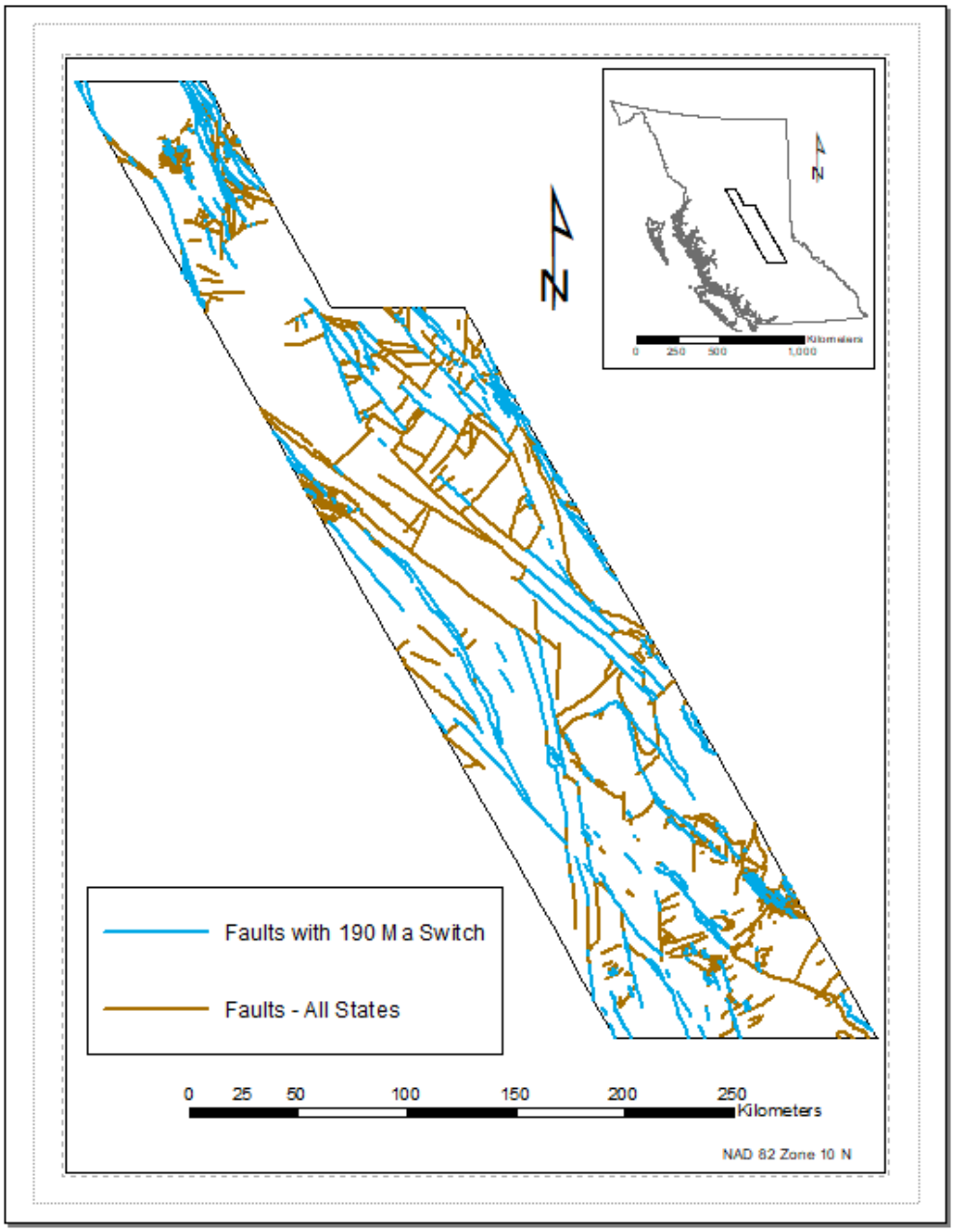

Figure 6-5: Mapped faults used for evaluating structural parameters.

GridID	TMI	Conductiv	Suscepti	Density	IntrC	VolcC	DIST_AnyFa	ANGLE_180_	NEAR_DIST_	ANGLE_180_	Labels
1274	56940.12	0.0185964704	13854	0.065785	1	2	1686.98071797	73.3634344559	1686.98071797	73.3634344559	1
1275	56942.5	0.0153311845	19848	0.05957100	1	4	1216.61748569	67.8229135972	1216.61748569	67.8229135972	1
1276	57416.38	0.0136734974	34054	0.05311800	1	4	763.484686371	66.0233549679	763.484686371	66.0233549679	1
1277	57246.59	0.0138750421	35375	0.04766700	1	4	310.956387383	66.0291528904	310.956387383	66.0291528904	1
1278	57090.22	0.0156847201	23461	0.041532	1	4	136.266469979	51.2786970824	136.266469979	51.2786970824	1
1279	57015.66	0.0186053924	15190	0.03409900	1	1	519.627440844	51.2844912735	519.627440844	51.2844912735	1
1280	57331.78	0.0220822897	16587	0.02612500	1	1	888.867332948	47.2524522855	888.867332948	47.2524522855	1

Figure 6-6: A sample of the input and output spreadsheet of the machine learning algorithm.

The use of a mesh with unique grid ID values (first column in Figure 6-6) facilitated collaboration between the team members working in Matlab and those working with ArcGIS and preparing input data. As shown in Figure 6-7, grid IDs allow converting the map data into spreadsheet forms such as Microsoft Access and other databases. Thus, after collecting the input data and selecting the mineral occurrences for training and validation step, a spatial grid was created relating points in the available rasters of geophysical data to vector data on plausible influences on prospectivity.

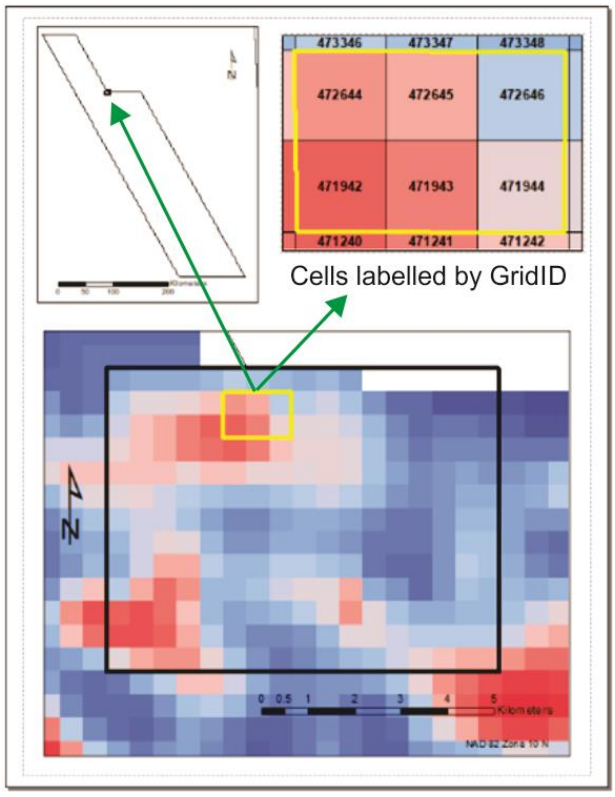

Figure 6-7: Example of mesh cells on the map with assigned grid IDs.

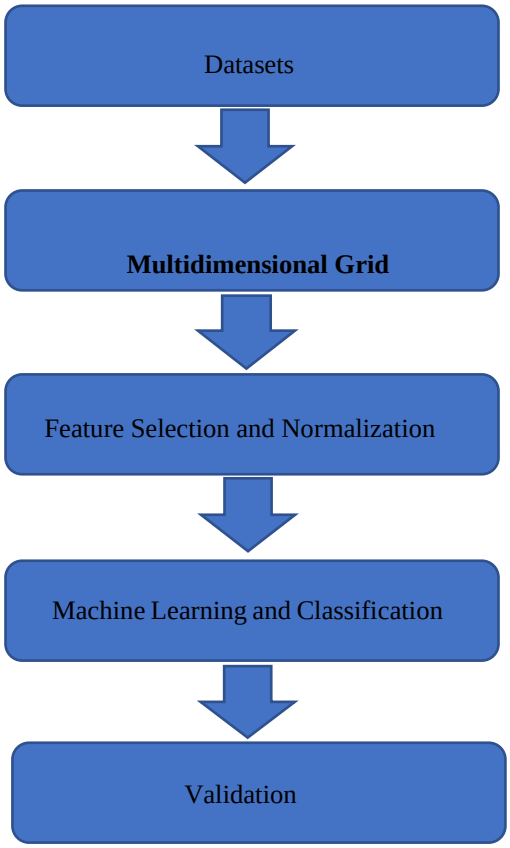

Figure 6-8: Procedure of applying machine learning techniques to mapping.

The procedure of using machine-learning techniques to find the best potential spots of porphyry related deposits is shown in Figure 6-8. The first steps consist in collecting the data and creating a spatial grid described above. Then, I test and select a subset of features with greatest correlation to porphyry-related mineral occurrences and normalize the data ranges to facilitate comparison. After this, the data are subdivided into training, validation, and testing datasets. The samples are randomly subdivided into three categories: 70% of the points were assigned to training, 15% to validation, and 15% to test data. In the validation step, the trained model is evaluated for accuracy of prediction, and parameters of the algorithm are adjusted to achieve the expected performance of the trained model.

6.4 Feature engineering

To examine the training dataset, I perform the Pearson correlation test in order to identify any linear correlations between the ten input features (Figure 6-9, Figure 6-10, and Figure 6-11). This test consists in evaluating the Pearson correlation coefficient R for each pair of input features and removing the features which are highly correlated, i.e. have an absolute value or R exceeding a certain threshold. In most studies, the cut-off threshold value on $|R|$ is taken equal 0.9 or 0.85 (Haykin, 2009). However, the map attributes used in this study come from significantly different observations and should therefore be not highly correlated. The highest correlation coefficient $R = 0.51$ among the geophysical features is between the TMI (feature 1) and magnetic susceptibility (feature 3; Figure 6-9). The strongest data correlation across the entire dataset equals $R = 0.79$ and occurs between the distances to the faults of different ages (features 7 and 9). These levels of correlation are insufficient for discarding the data in Pearson test.

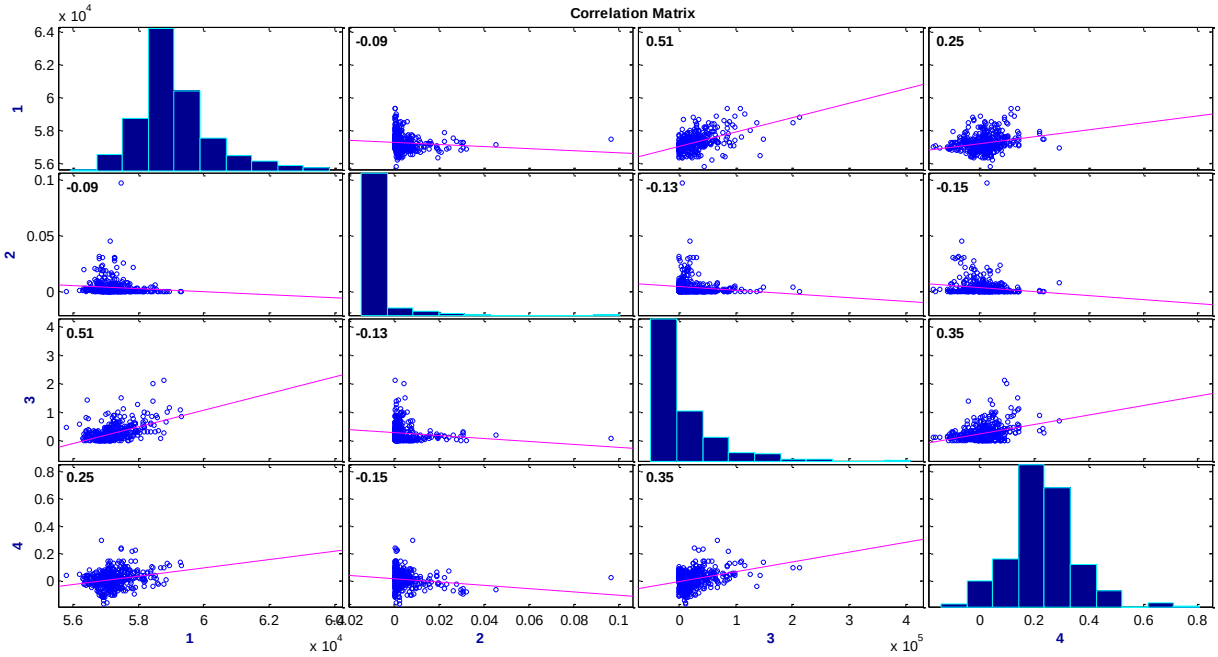

Figure 6-9: Correlation between the four geophysical features 1 to 4. Plots on the diagonal of this grid show the histograms of the distribution of the respective features (shown by rows and columns of this grid), and off-diagonal plots show scatterplots of pairs of features. Pink lines indicate the least squares regressions relating these pairs of features. Numbers in the upper-left corners of plot boxes are the Pearson's correlation coefficients.

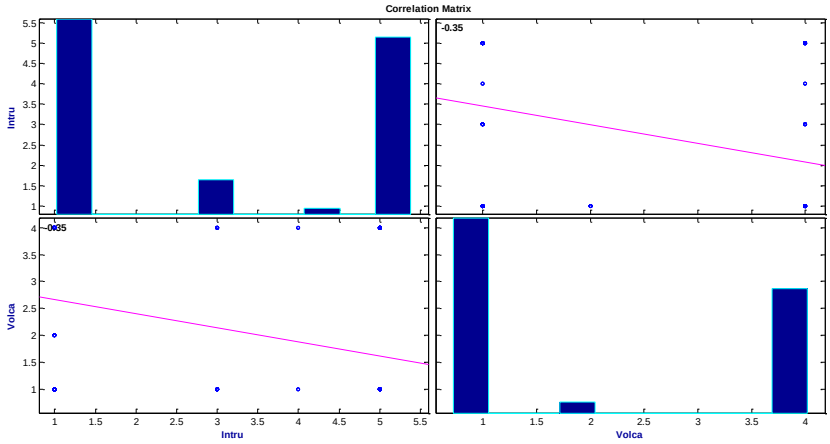

Figure 6-10: Correlation coefficients between geological features 5 and 6. Plotting style the same as in Figure 6-9.

The increased correlation $R = 0.79$ between features 7 and 9 (Figure 6-11) largely occurs because segments of the same or closely spaced faults are apparently listed as these features (and also as 8 and 10) in the database. These pairs of faults produce close distances and direction values, which can be seen by the distributions of dots on diagonal lines in the cross-plots (Figure 6-11), the located in generally the same area. In summary of the correlation analysis, Figure 6-12 shows the correlation-coefficient matrix between all ten features in the form of a heatmap plot.

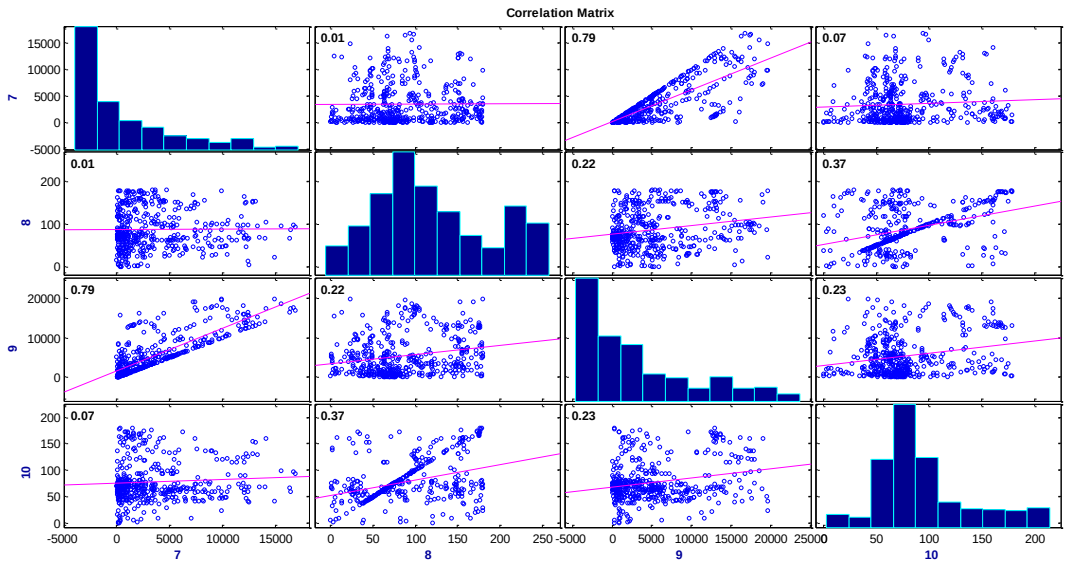

Figure 6-11: Correlation coefficients between the four structural features 7 to 10. Plotting style the same as in Figure 6-9.

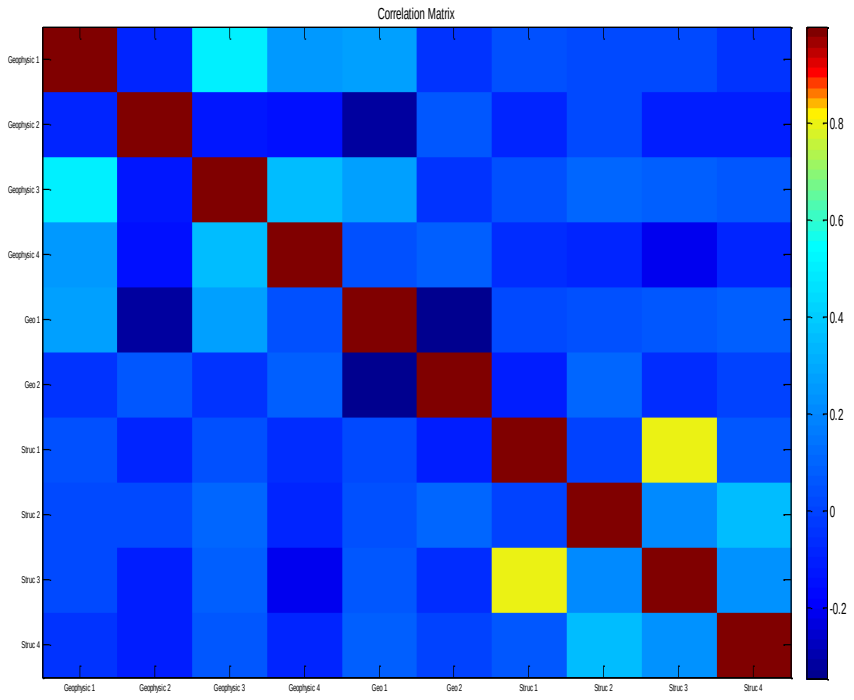

Figure 6-12: Heatmap for the Pearson correlation matrix for all input data features.

6.5 Results

Using the assembled dataset, I performed prediction of a 'prospectivity' property, or probability of porphyry deposits *P* ranging from zero to one. I used a two-layer feed-forward ANN with 30 neurons in the hidden layer. Sigmoid activation functions were used in the hidden and output layers (chapter 2).

During ANN training, the weights and bias values were updated using the scaled conjugate gradient backpropagation method. Although the validation and test set are kept away from training data, the validation set is used for tuning model parameters and the test set is used to evaluate the final tunned model performance. The training continued until the validation error stopped decreasing for six iterations (this selection is called the validation stop).

The result of the application of neural network is shown in Figure 6-13 in the form of a prospectivity map with similar approach as a case study in Iran (Asadi et al., 2016). In this figure, black dots are the known porphyry occurrences , light blue dots are no-identified occurrences, dark blue dots are Non-porphyry occurrences and colors in the map represent different categories of

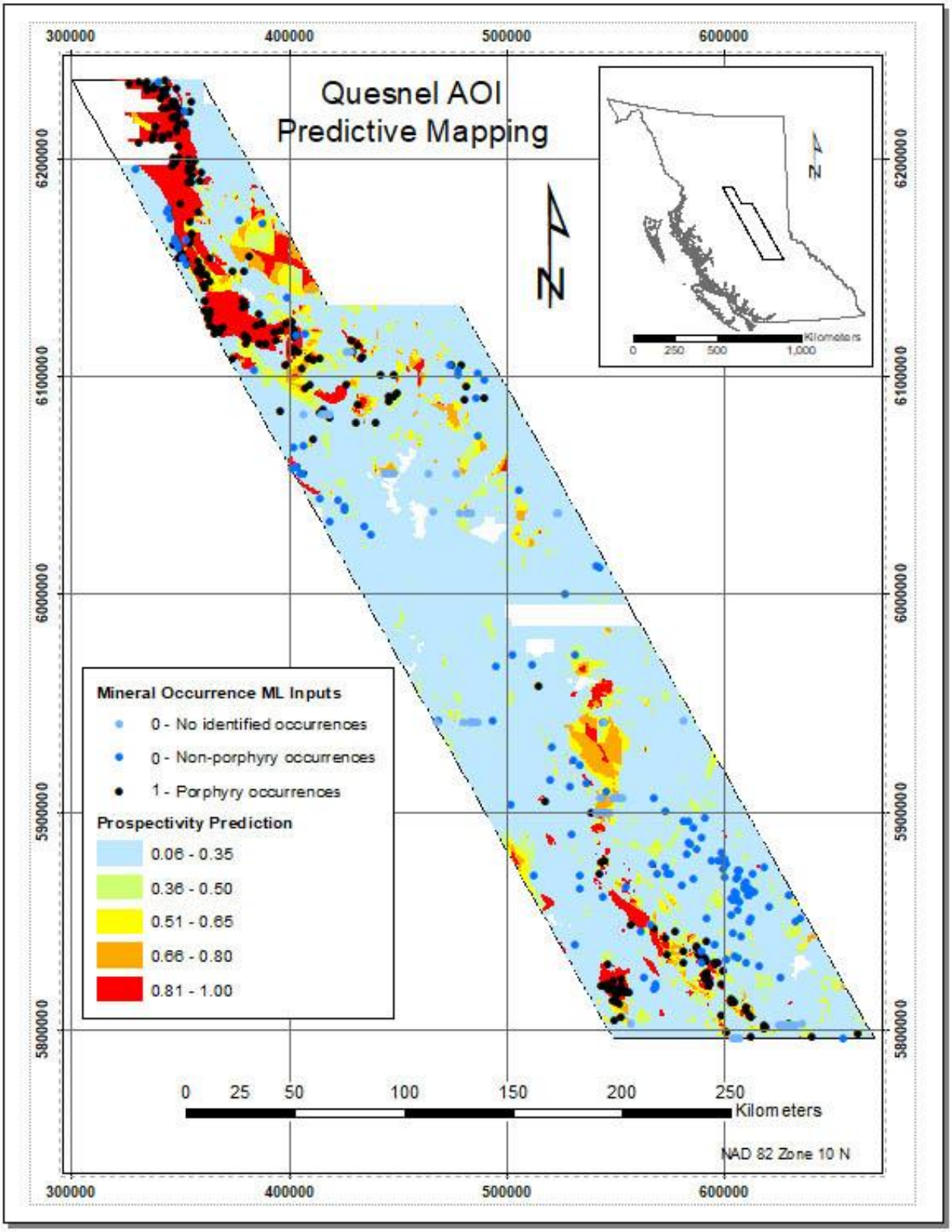

Figure 6-13: Predicted prospectivity map (plot courtesy of Dène Tarkyth). Colours show intervals of prospectivity values, and blue and black dots show the data labels (legend).

potential porphyry mineralization (ranges of *P*). The predicted porphyry class is the area that the probability is larger which means larger than 0.5. Thus, the yellow, orange and red areas are the potential spots for porphyry-related deposits. From red-color to blue-color regions, the probability of porphyry deposits decreases. As this figure shows, known deposits Quesnel area are concentrated in or around the red areas ($P \geq 80\%$), which suggests good quality of the prediction. At the same time, some of the red areas are not sampled by black dots, which suggests that additional exploration may be recommended in these areas.

As for all data-driven methods, the results shown in Figure 6-13 may contain a limitation, which is the exploration bias due to the nonuniform area for sampling exploration sites and mineral deposits (Bonham-Carter, 1994; Coolbaugh et al., 2007; Ford et al. 2016). In mappings by supervised algorithms, the final model is often biased toward known mineral occurrences, because such occurrences are reported more frequently than non-occurrences (Harris and Pan, 1999; Carranza, 2004, 2008, 2015; Carranza and Laborte, 2016). In the present case, the exploration bias should be moderate because a half of the database consists of non-prospective cases (228 vs. 227 prospective; section 6.3). Knowledge-driven methods (unsupervised, based on deterministic geological models) are not subject to such exploration bias, but their efficiency of prediction and success rate are usually much lower than for data-driven methods (McKay and Harris, 2016).

To analyze the performance of the algorithm in more detail, a confusion matrix for the dataset is shown in Figure 6-14. The confusion matrix contains the counts of data points correctly and incorrectly predicted by the ANN. As noted in section 6.3, 142 of these non-prospective cases may have prospective-like geophysical signatures, and therefore we might expect a bias toward false positive predictions for these cases. However, the confusion matrix is nearly symmetric (Figure 6-14) and shows no such bias.

Four useful summary metrics arising from the confusion matrix are shown in Table 6.1 (Haykin, 2009). Each of these metrics is defined within the range of [0,1].

Table 6.1: Validation scores of the applied neural network.

Precision	Accuracy	Recall	F1 Score
0.79	**0.78**	**0.77**	**0.78**

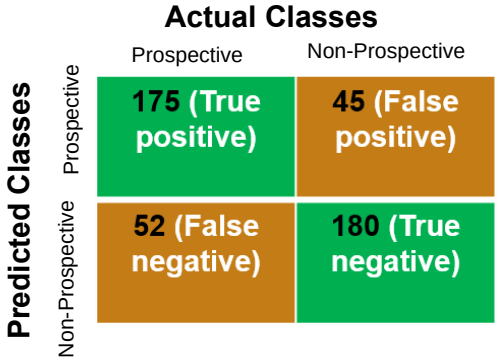

Figure 6-14: Confusion matrix of porphyry prospectivity prediction neural network.

In Table 6.1, the precision is the ratio of the number of true positive to the total of positive cases and the accuracy is the ratio of the number if true positive or negative cases to the total number of data (Haykin, 2009). The recall value (also called sensitivity, hit rate) is the ratio of the number of true positive to the number of real positive cases in the data. A system with high recall but low precision returns many results, but most of its predicted labels are incorrect when compared to the training labels. A system with high precision but low recall is just the opposite, returning very few results, but most of its predicted labels are correct when compared to the training labels. An ideal system with high precision and high recall will return many results, with all results labeled correctly. If we want to be very confident about predicting prospective cases, we try to get high precision, but if we want to avoid missing many prospective cases, we should try obtaining a higher recall. High scores for both show that the classifier is returning accurate results (high precision), as well as returning a majority of all positive results (high recall). The F1 score is defined as the harmonic mean of the precision and recall rates. A high F1 score means good prediction of both known positive and negative outcomes.

All confusion matrix-statistics are remarkably close to 0.78 for the present dataset and algorithm (Table 6.1). This suggests that the prediction algorithm contains no bias, and the 0.78 represents its level of confidence for any type of prediction. The remaining error probability 0.22 should be due to the errors and limitations of the data. By using statistical methods like cokriging or probabilistic neural networks (e.g., Din and Hongbing (2019), Ekone et al. (2020)) this estimate of statistical confidence can be improved in future studies.

CHAPTER 7: Conclusions

In this chapter, I summarize key results of the two case studies of this thesis (sections 7.1 and 7.2). The discussion is close to the sequence of the study objectives in chapter 1. In section 7.3, I outline some outstanding questions, limitations of the present work, and suggest recommendations for future research.

7.1 Results of reservoir characterization study

Using machine learning methodologies, three groups of reservoir characterization results were obtained from the well-log and 3-D seismic dataset from Mansuri oil field in Iran: inversion for acoustic impedance (AI), porosity estimation, and classification of the reservoir into four classes corresponding to electrofacies EF1 to EF4 identified in a previous study (Zahmatkesh et al, 2021).

AI inversion was performed using the industry-standard constrained model-based method in Hampson-Russell software and showed good accuracy in matching the seismic data and correspondence with well logs. The inverted impedance variations were at 98% level of those obtained within 13 well logs, with 86% statistical correlation. These levels of accuracy are considered good in AI inversion practice.

A feasibility study for porosity estimation showed that porosity has an about 73% correlation with the AI in the available well logs. Porosity estimation was performed using the inverted AI combined with additional seismic attributes. Feature selection using stepwise regression reduced the this set of seismic features to nine principal attributes. The resulting porosity volume is consistent with well logs and correlates with geological layering.

After confirming a meaningful relationship between reservoir parameters through feasibility study at the beginning, inversion analysis was implemented, and different inversion methodologies were checked to obtain the best estimated acoustic impedance. Seismic-well-tie was performed using 13 wells, and a set of different wavelets was tested to obtain the highest accuracy of seismic-well-tie. The initial model was created to be used in model-based inversion analysis. The results of estimated acoustic impedance using constraint-model-based inversion

were compared to other inversion algorithms and found to be the best after a series of quality control.

Porosity estimation was successfully implemented using seismic attributes and estimated acoustic impedance. The best seismic attributes were selected using stepwise regression, and different machine learning techniques such as PNN, MLFN and RBFN were used, and the final results were compared to pick the best methodologies based on performing a complete quality control process.

Using the inverted AI and porosity volumes combined with additional 23 seismic attributes, a classification of the reservoir into classes EF1 to EF4 was performed using three algorithms: the probabilistic neural network (PNN), the standard artificial neural network (ANN), and an ANN with optimized data inputs proposed in this thesis (P-ANN). The P-ANN algorithm showed the best performance in validation and training scores compared to the other algorithms. The PNN results showed a more continuous layered structure within the reservoir. The P-ANN results also revealed a continuous layered structure with some intermixing of other classes, particularly within the upper half of the reservoir. Thus, multiple algorithms still need to be used for classification of reservoirs.

In addition to class numbers (labels), probabilities of the classes and statistical confidence estimates were obtained from the classification algorithms. Plotting of these quantities as seismic sections or volumes is useful for reservoir characterization and interpretation. Cross-sections of classification confidence exhibit layered patterns related to those of AI and porosity. The P-ANN algorithm showed the largest and PNN the lowest confidence estimates; however, this difference was likely due to the strongly different models for class probabilities.

Analyzing the confidence levels allows enhancing many types of interpretations of seismic data. With a given classification algorithm, it is useful to use the confidence volumes for plotting class labels or other types of seismic data. By plotting zones with confidence levels above 0.7 (for P-ANN) or 0.5 (for PNN), I obtained images of reliable classifications of the reservoir.

After applying these three algorithms to the new samples from a selected seismic line, the PNN results show a more continuous layered structure through the Asmari reservoir in the Mansuri Oilfield. The results of the proposed ANN also show a continuous layered structure with some

intrusion from other classes rather than the dominant class, especially at the top half of the reservoir. I used a probability filter based on the probability of the selected class and applied that filter to the classification section made by the proposed ANN algorithm, and the classes with low confidence level were removed. Accordingly, these classes were mostly related to uncertain classes or were related to the selected classes with almost the same probability as the unpicked classes. The result of the applied probability filter on the PNN classification section shows a section that has lost most of the data at the bottom of the section, which implies a more uncertain classification in that part of the section.

7.2 Results of porphyry prospectivity study

In the mineral prospectivity study in central and northwester British Columbia (chapter 6), geophysical, geological mapping, and structural data were combined in an integrated ANN-based interpretation. Feature engineering showed that all ten types of data could be used in the interpretation. The result if this study was a new prospectivity map of porphyry deposits within a part of the Quesnel Terrane. The predicted prospectivity showed good correspondence with existing data, and the predicted porphyry deposits were considered geologically reasonable. In several areas, the estimated probability of prospectivity suggested directions for further field exploration.

7.3 Recommendations for future research

In this study, a variety of inversion, statistical, and machine learning approaches were applied and compared. Although these applications were successful and led to practical results such as cross-sections of acoustic impedance, porosity (chapter 4), classifications of seismic facies (chapter 5), or a map or porphyry perspective (chapter 6), the study also suggested a number of ways for further enhancements of the approaches and results. Both the geophysics and machine-learning parts of the work can be continued to enhance lateral continuity of the images.

This can be achieved by filtering the noise, improving the algorithms for attribute extraction, and also by changing the ANN optimization algorithms to honour lateral continuity. These enhancements can be subdivided into three categories: 1) improvements in the specifically geophysical aspects of inversions, and 2) improvement of the machine-learning

techniques, and 3) further use of these datasets for extraction of additional information and method development.

7.3.1 Geophysics aspects

In this thesis, the emphasis was placed on machine-learning methods whereas the input geophysical features were obtained by standard techniques. However, some limitations of these techniques were also revealed which affected the accuracy of the results. In particular, as noted chapter 4, the inversion for the AI from seismic data underestimated the seismic-frequency variation of the impedance within the wells by about 16%. Although this underestimation does not affect the classification of seismic facies (chapter 5), it is still a significant deviation of the AI from reality. This deviation also caused an underestimation of the reflectivity and porosity.

To overcome the above inaccuracy of the AI, different inversion methods need to be used. In particular, the method by Morozov and Ma (2009) is explicitly formulated to obtain an AI which would reproduce the spectral amplitude of reflectivity within the well logs. This method also uses a Delaunay triangulation method for spatial interpolation, which is free from the 'bull's eye' artifact mentioned in chapter 2.

In the study of mineral prospectivity (chapter 6), the geophysical aspect can also be enhanced in potential future studies by deriving maps of additional geophysical attributes. For example, pseudo-gravity, analytic signal, spatial derivatives, Euler derivatives, curvature and other attributes can be derived from the magnetic and gravity maps and used as inputs into classifications. These transformations enhance the spatial gradients, and they could be useful for detecting mineralized zones.

7.3.2 Machine-learning aspects

Regarding possible improvement of machine-learning techniques, it would be useful to use more complex for application of ANN to the data, such as convolutional or and recurrent neural networks. The ANN used in this study were applied to only one sample at a time (one time sample in a seismic record in chapter 5 or one grid cell in a map in chapter 6). This choice was made in order to keep the approach simple and tractable in this thesis. However, for seismic or geological

mapping applications, a more appropriate and accurate approach is to apply the ANN to a broader vicinity of the sample, i.e. to treat the ANN as a temporal or spatial convolutional filter. This type of ANN application is referred to as the convolutional in machine learning (Haykin, 2009).

In application of machine learning techniques to geological mapping (chapter 6), the present study was somewhat limited by the small number of proven porphyry deposits, and the known labels are available only at a small number of locations compared to the vast study area. Perhaps a larger and more complex dataset containing not only porphyry but other minerals such as gold could be used to constrain the target mineralizations better. Potentially, mutual relations between occurrences of different minerals could be developed and used for improving the confidence in their predictions.

A significant enhancement of the machine-learning results, and also of the AI and porosity inversions of this thesis could consist in producing more detailed and accurate estimates of their uncertainties. Such estimates can be obtained by inversions with randomized parameters, known as bootstrapping inversion or the Random Forest Decision Trees in machine learning (Haykin, 2009). These randomized methods could be used for feature ranking to find the most critical features in our 10-dimensional feature space, and also to test for the impacts of the algorithm regularization parameters. In addition, it would be interesting to obtain the so-called Naïve Bayes classifier and compare its results to the pattern recognition classifier in chapter 5.

7.3.3 Further use of the datasets

The potential of the datasets in this thesis is also not exhausted, and Since the Mansuri oil field has a relatively simple structure of a gentle anticline and its seismic data is of high quality, it would be a good case study to use different deep learning techniques and perform classification or other types of reservoir parameter estimation.

The proposed approach to constructing the optimal of attributes (chapters 2 and 5) could be used in other studies and could be compared to the results of other machine learning techniques. In particular, if using more geophysical attributes and convolutional neural networks in the mineral prospectivity mapping study, using optimized and orthogonal attributes could improve its results.

REFERENCES

Abdizadeh, H., Kadkhodaie, A., Ahmadi, A., & Heidarifard, M. H. (2017). Estimation of Total Organic Carbon from well logs and seismic sections via neural network and ant colony optimization approach: a case study from the Mansuri oil field, SW Iran. Geopersia, 7(2), 255-266.

Abedi, M., & Norouzi, G. H. (2012). Integration of various geophysical data with geological and geochemical data to determine additional drilling for copper exploration. Journal of Applied Geophysics, 83, 35-45.

Amendola, A., Gabbriellini, G., Dell'Aversana, P., Marini, A.J. (2017). Seismic facies analysis through musical attributes. Geophysical Prospecting, 65(S1), 49-58.

Asadi, H. H., Sansoleimani, A., Fatehi, M., & Carranza, E. J. M. (2016). An AHP–TOPSIS predictive model for district-scale mapping of porphyry Cu–Au potential: a case study from Salafchegan area (central Iran). Natural Resources Research, 25(4), 417-429.

Ashraf, U., Zhu, P., Yasin, Q., Anees, A., Imraz, M., Mangi, H.N., Shakeel, S., 2019. Classification of reservoir facies using well log and 3D seismic attributes for prospect evaluation and field development: a case study of Sawan gas field, Pakistan. Journal of Petroleum Science and Engineering, 175, 338-351.

Austin, O. E., Onyekuru Smauel, I., Ebuka, A. O., & Abdulrazzaq, Z. T. (2018). Application of model-based inversion technique in a field in the coastal swamp depobelt, Niger Delta. International Journal of Advanced Geosciences, 6(1), 122-126.

Baaske, U.P., Mutti, M., Baioni, F., Bertozzi, G., Naini, M.A. (2007). Using multi-attribute neural networks classification for seismic carbonate fades mapping: a workflow example from mid-Cretaceous Persian Gulf deposits. Geological Society, London, Special Publications, 277(1), 105-120.

Bagheri, M., Riahi, M.A. (2015). Seismic facies analysis from well logs based on supervised classification scheme with different machine learning techniques. Arabian Journal of Geosciences, 8(9), 7153-7161.

Bonham-Carter, G. F., & Bonham-Carter, G. (1994). Geographic information systems for geoscientists: modelling with GIS (No. 13). Elsevier.

Carranza, E. J. M. (2004). Weights of evidence modeling of mineral potential: a case study using small number of prospects, Abra, Philippines. Natural Resources Research, 13(3), 173-187.

Carranza, E. J. M. (2015). Data-driven evidential belief modeling of mineral potential using few prospects and evidence with missing values. Natural Resources Research, 24(3), 291-304.

Carranza, E. J. M., & Laborte, A. G. (2016). Data-driven predictive modeling of mineral prospectivity using random forests: A case study in Catanduanes Island (Philippines). Natural Resources Research, 25(1), 35-50.

Carranza, E.J.M., 2008. Geochemical anomaly and mineral prospectivity mapping in GIS. Handbook of Exploration and Environmental Geochemistry, 11. Elsevier, Amsterdam.

Col´eou, T., Poupon, M., Azbel, K. (2003). Unsupervised seismic facies classification: a review and comparison of techniques and implementation. Lead. Edge 22(10), 942–953.

Connolly, P. (1999). Elastic impedance. The Leading Edge, 18, 438–452.

Coolbaugh, M. F., Raines, G. L., & Zehner, R. E. (2007). Assessment of exploration bias in data-driven predictive models and the estimation of undiscovered resources. Natural Resources Research, 16(2), 199-207.

Dashti, A. L. I., & Sfidari, E. (2016). Physical properties modeling of reservoirs in Mansuri oil field, Zagros region, Iran. Petroleum Exploration and Development, 43(4), 611-615.

De Matos, M.C., Osorio, P.L., Johann, P. R. (2007). Unsupervised seismic facies analysis using wavelet transform and self-organizing maps. Geophysics, 72(1), P9-P21.

Din, N. U., & Hongbing, Z. (2020). Porosity prediction from model-based seismic inversion by using probabilistic neural network (PNN) in Mehar Block, Pakistan. Episodes Journal of International Geoscience, 43(4), 935-946.

Du, H., Cao, J. X., Xue, Y. J., & Wang, X. J. (2015). Seismic facies analysis based on self-organizing map and empirical mode decomposition. Journal of Applied Geophysics, 112, 52-61.

Ekone, NO., Ehirim, C. N., Nwosu, J. I., & Dagogo, T. Porosity modeling using acoustic impedance and well log data in EK field, Niger Delta.

Euzen, T., Delamaide, E., Feuchtwanger, T., & Kingsmith, K. D. (2010). Well log cluster analysis: an innovative tool for unconventional exploration. In Canadian unconventional resources and international petroleum conference. OnePetro.

Eze, S., Orji, O. M., Nnorom, S. L., & Ubogun, K. (2019). Model Based Inversion of Acoustic Impedance from Seismic Trace for Lithofacies Differentiation: an Application in Xy Field Offshore Niger Delta. Journal of Applied Sciences and Environmental Management, 23(9), 1677-1684.

Farzadi, P., (2006). Seismic Facies Analysis Based on 3D Multi-attribute Volume Classification, Dariyan Formation, SE Persian Gulf. Journal of Petroleum Geology, 29(2), 159-173.

Feng, R., Luthi, S.M., Gisolf, D., Angerer, E. (2018a). Reservoir lithology classification based on seismic inversion results by Hidden Markov Models: applying prior geological information. Marine and Petroleum Geology, 93, 218-229.

Feng, R., Luthi, S.M., Gisolf, D., Angerer, E. (2018b). Reservoir lithology determination by hidden Markov random fields based on a Gaussian mixture model. IEEE Transactions on Geoscience and Remote Sensing, 56(11), 6663-6673.

Ford, A., Miller, J. M., & Mol, A. G. (2016). A comparative analysis of weights of evidence, evidential belief functions, and fuzzy logic for mineral potential mapping using incomplete data at the scale of investigation. Natural Resources Research, 25(1), 19-33.

Ghezelbash, R., Maghsoudi, A., & Carranza, E. J. M. (2019). Performance evaluation of RBF-and SVM-based machine learning algorithms for predictive mineral prospectivity modeling: integration of SA multifractal model and mineralization controls. Earth Science Informatics, 12(3), 277-293.

Hampson, D.P., Schuelke, J.S., Quirein, J.A., (2001). Use of multiattribute transforms to predict log properties from seismic data. Geophysics, 66(1), 220-236.

Harris, D., & Pan, G. (1999). Mineral favorability mapping: a comparison of artificial neural networks, logistic regression, and discriminant analysis. Natural Resources Research, 8(2), 93-109.

Haykin, S. (2009). Neural networks and learning machines, 3rd ed. Pearson Prentice Hall, 905 p.

Johann, P., de Castro, D. D., & Barroso, A. S. (2001). Reservoir geophysics: Seismic pattern recognition applied to ultra-deepwater oilfield in Campos basin, offshore Brazil. In SPE Latin American and Caribbean Petroleum Engineering Conference, March 2001. OnePetro.

Kelley, H. J. (1960). Gradient theory of optimal flight paths. ARS Journal, 30 (10), 947–954.

Keykhay-Hosseinpoor, M., Kohsary, A. H., Hossein-Morshedy, A., & Porwal, A. (2020). A machine learning-based approach to exploration targeting of porphyry Cu-Au deposits in the Dehsalm district, eastern Iran. Ore Geology Reviews, 116, 103234.

Kumar, R., Das, B., Chatterjee, R., Sain, K. (2016). A methodology of porosity estimation from inversion of post-stack seismic data, Journal of Natural Gas Science and Engineering 28, 356–364.

Kushwaha, P. K., Maurya, S. P., Rai, P., & Singh, N. P. (2020). Porosity prediction from offshore seismic data of F3 Block, the Netherlands using multi-layer feed-forward neural network. Current Science, 119(10), 1652.

Lancaster, S., & Whitcombe, D. (2000). Fast-track 'colored' inversion. SEG Expanded Abstracts, 19, 1572.

Latimer, R. B., R. Davison, & van Riel, P. (2000). An interpreter's guide to understanding and working with seismic-derived acoustic impedance data. The Leading Edge, 19, 242–256

Linari, V., Santiago, M., Pastore, C., Azbel, K., & Poupon, M. (2003). Seismic facies analysis based on 3D multiattribute volume classification, La Palma Field, Maracaibo, Venezuela. The Leading Edge, 22(1), 32-36.

Lindseth, R. O. (1979). Synthetic sonic logs—A process for stratigraphic interpretation. Geophysics, 44(1), 3-26.

Liu, J., Dai, X., Gan, L., Liu, L., & Lu, W. (2018). Supervised seismic facies analysis based on image segmentation. Geophysics, 83(2), O25-O30.

Maepa, F., Smith, R. S., & Tessema, A. (2021). Support vector machine and artificial neural network modelling of orogenic gold prospectivity mapping in the Swayze greenstone belt, Ontario, Canada. Ore Geology Reviews, 130, 103968.

Marroquín, I.D. (2014). A knowledge-integration framework for interpreting seismic facies. Interpretation, 2(1), SA1-SA9.

McCulloch, W., & Pitts, W. (1943). A logical calculus of ideas immanent in nervous activity. Bull. of Math. Biophysics, 5, 115-133.

McKay, G., & Harris, J. R. (2016). Comparison of the data-driven random forests model and a knowledge-driven method for mineral prospectivity mapping: A case study for gold deposits around the Huritz Group and Nueltin Suite, Nunavut, Canada. Natural Resources Research, 25(2), 125-143.

Mallick, S. (1995), Model-based inversion of amplitude-variations with offset data using a genetic algorithm. Geophysics, 60, 939–954.

Menke, W. (1984). Geophysical data analysis: Discrete inverse theory, Academic Press, doi:10.1016/B978-0-12-490920-5.X5001-7

Morozov, I. B. (2021) Geophysical Inversion. University of Saskatchewan GEOL884 course notes.

Morozov, I. B., and Ma, J. (2009). Accurate poststack acoustic-impedance inversion by well-log calibration. Geophysics, 74(5), R59-R67.

Oldenburg, D. W., T. Scheuer, & S. Levy, S. (1983). Recovery of the acoustic impedance from reflection seismograms, Geophysics, 48, 1318–1337

Pazand, K., & Hezarkhani, A. (2015). Porphyry Cu potential area selection using the combine AHP-TOPSIS methods: a case study in Siahrud area (NW, Iran). Earth Science Informatics, 8(1), 207-220.

Phillips, N., Nguyen, T. N. H., Thomson, V., Oldenburg, D. (2009). 3D inversion modelling, integration, and visualization of airborne gravity, magnetic, and electromagnetic data: the Quest project, BC, Canada. Mira Geoscience modeling report.

Polyanin, A. D., & Manzhirov, A. V. (2007). Handbook of mathematics for engineers and scientists, Chapman & Hill/CRC

Pramanik, A.G., Singh, V., Vig, R., Srivastava, A. K., & Tiwary, D. N. (2004). Estimation of effective porosity using geostatistics and multiattribute transforms: A case study. Geophysics, 69(2), 352-372.

Santos, L.T., and M. Tygel, 2004, Impedance-type approximations of the P-P elastic reflection coefficient: Modeling and AVO inversion: Geophysics, 69, 592–598.

Qin, Y., Liu, L., & Wu, W. (2021). Machine learning-based 3D modeling of mineral prospectivity mapping in the Anqing Orefield, Eastern China. Natural Resources Research, 30(5), 3099-3120.

Riedel, M., Bahk, J.-J., Kim, H.-S., Yoo, D.-G., Kim, W.-S., Ryu, B.-J. (2013a). Seismic facies analyses as aid in regional gas hydrate assessments. Part-I: classification analyses. Marine and petroleum geology, 47, 248-268.

Riedel, M., Bahk, J.-J., Kim, H.-S., Scholz, N.A., Yoo, D.G., Kim, W.-S., Ryu, B.-J., Lee, S. R. (2013b). Seismic facies analyses as aid in regional gas hydrate assessments. Part-II: prediction of reservoir properties, gas hydrate petroleum system analysis, and Monte Carlo simulation. Marine and petroleum geology, 47, 269-290.

Roden, R., Smith, T., Sacrey, D. (2015). Geologic pattern recognition from seismic attributes: Principal component analysis and self-organizing maps. Interpretation, 3(4), SAE59-SAE83.

Rodriguez-Galiano, V., Sanchez-Castillo, M., Chica-Olmo, M., & Chica-Rivas, M. J. O. G. R. (2015). Machine learning predictive models for mineral prospectivity: An evaluation of neural networks, random forest, regression trees and support vector machines. Ore Geology Reviews, 71, 804-818.

Roslin, A., Esterle, J. S. (2016). Electrofacies analysis for coal lithotype profiling based on high-resolution wireline log data. Computers & Geosciences, 91, 1-10.

Roy, A., Dowdell, B.L., Marfurt, K.J. (2013). Characterizing a Mississippian tripolitic chert reservoir using 3D unsupervised and supervised multiattribute seismic facies analysis: An example from Osage County, Oklahoma. Interpretation, 1(2), SB109-SB124.

Russell, B., & Hampson, D. (2006). The old and the new in seismic inversion. CSEG Recorder, 31(10), 5-11.

Russell, B., Hampson, D., Schuelke, J., & Quirein, J. (1997). Multiattribute seismic analysis. The leading edge, 16(10), 1439-1444.

Saggaf, M.M., Toks¨oz, M.N., Marhoon, M.I. (2003). Seismic facies classification and identification by competitive neural networksSeismic Facies Mapping. Geophysics, 68(6), 1984-1999.

Sayago, J., Di Lucia, M., Mutti, M., Cotti, A., Sitta, A., Broberg, K., Przybylo, A., Buonaguro, R., & Zimina, O. (2012). Characterization of a deeply buried paleokarst terrain in the Loppa High using core data and multiattribute seismic facies classification. AAPG bulletin, 96(10), 1843-1866.

Taner, .T., Koehler, F., & Sheriff, R. E. (1979). Complex seismic trace analysis. Geophysics, 44(6), 1041-1063.

Werbos, P. J. (1982). Applications of advances in nonlinear sensitivity analysis. In System modeling and optimization (pp. 762-770). Springer, Berlin, Heidelberg.

Whitcombe, D., Connolly, P., Reagan, R., & Redshaw, T. (2002), Extended elastic impedance for fluid and lithology prediction. Geophysics, 67(1), 63–67.

Wyllie, M.R.J., Gregory, A.R., & Gardner, G.H.F. (1956). Elastic wave velocities in heterogeneous and porous media. Geophysics 21, 40e70.

Wrona, T., Pan, I., Gawthorpe, R.L., Fossen, H., (2018). Seismic facies analysis using machine learning. Geophysics, 83(5), O83-O95.

Yilmaz, Ö. (2001). Seismic data analysis: Processing, inversion, and interpretation of seismic data. Society of exploration geophysicists.

Yin, J., & Li, N. (2022). Ensemble learning models with a Bayesian optimization algorithm for mineral prospectivity mapping. Ore Geology Reviews, 145, 104916.

Yousefi, M., & Carranza, E. J. M. (2015). Prediction–area (P–A) plot and C–A fractal analysis to classify and evaluate evidential maps for mineral prospectivity modeling. Computers & Geosciences, 79, 69-81.

Zahmatkesh, I., Kadkhodaie, A., Soleimani, B., & Azarpour, M. (2021). Integration of well log-derived facies and 3D seismic attributes for seismic facies mapping: A case study from mansuri oil field, SW Iran. Journal of Petroleum Science and Engineering, 202, 108563.

Zahmatkesh, I., Soleimani, B., Kadkhodaie, A., Golalzadeh, A., Abdollahi, A.A.M. (2017). Estimation of DSI log parameters from conventional well log data using a hybrid particle swarm optimization–adaptive neuro-fuzzy inference system Journal of Petroleum Science and Engineering, 157, 842-859.

Zahraa, A., Zailani, A., Prasad Ghosh, D. (2017). Characterizing geological facies using seismic waveform classification in sarawak basin. In IOP Conference Series: Earth and Environmental Science (Vol. 88, No. 1, p. 012001). IOP Publishing.

Zhang, S.-H., Xu, Y., Abu-Ali, M., Teng, M.K. (2019). Seismic facies recognition and stratigraphic trap characterization based on neural networks. In International Petroleum Technology Conference. OnePetro.

Zhao, T., Jayaram, V., Roy, A., Marfurt, K.J. (2015). A comparison of classification techniques for seismic facies recognition. Interpretation, 3(4), SAE29-SAE58.

Zhao, W., Hu, S., Xu, Z., Zeng, H., Liu, W., Fu, Q., Shi, S., Wang, L., Jiang, L. (2018). Lithology mapping of a mixed siliciclastic− carbonate− evaporite system using 3D seismic and well data: Lower Triassic Jialingjiang Formation, Sichuan Basin, southwestern China. Marine and Petroleum Geology, 93, 422-436.

Zou, A., Wang,Y., & Wang, D. (2022). Blind inversion of multichannel nonstationary seismic data for acoustic impedance and wavelet. Pure and Applied Geophysics, 179 (2022), 2147–2166