ABSTRACT

Application of Machine Learning and Magnetotellurics to Aid in Subsurface Characterization of Petroleum and Geothermal Reservoirs

Elisabeth G. Rau, Ph.D.

Chairperson: Stacy C. Atchley, Ph.D.

Energy is the foundation of society and with future energy demand expected to increase significantly over the next few decades, solutions contributing to future energy resources are of high interest scientifically, geopolitically, and economically. Data analytics and machine learning provide useful tools to more efficiently and cost-effectively produce petroleum and geothermal resources vital for our energy future. Supervised and unsupervised machine learning can aid in the prediction of sedimentological and reservoir attributes in wells lacking core control to better and more efficiently characterize subsurface petroleum reservoirs. Using tree-based machine learning models, core-observed depositional attributes from the Late Devonian Duvernay Formation in Alberta, Canada may be predicted in wells lacking core control when class proportion and thickness conditions are met. Unsupervised machine learning technique, non-negative matrix factorization with *k*-means clustering (NMFk), automatically identifies reservoir significance, undetected through the traditional deterministic modelling, within the Duvernay Formation without calibration to core observations. The

application of NMFk with petrophysical data may assist in highlighting intervals of interest in advance of core descriptions reducing observer inconsistency and bias and enhancing the quality and relevance of core description for reservoir correlation and mapping. Machine learning methods provide precise, consistent, and objective petrophysical interpretations and reservoir characterization, and increases the consistency and accuracy of resource assessment for petroleum exploration and production. Unsupervised machine learning and magnetotellurics are useful analytical tools to assess prospective geothermal resources in the Tularosa Basin of south-central New Mexico based on heat flow, temperature, porosity, and permeability. The unsupervised machine learning method, NMFk, identifies locations with the highest likelihood of geothermal success, and the passive geophysical method, magnetotellurics can detect subsurface geothermal prospects. The integration of NMFk and MT can provide a 3D assessment of heat flow, temperature, and permeability for geothermal exploration. This research provides innovative methods to aid in the development of efficient and cost-effective approaches for future energy exploration and production.

Application of Machine Learning and Magnetotellurics to Aid in Subsurface Characterization of Petroleum and Geothermal Reservoirs

by

Elisabeth G. Rau, B.S., M.A.

A Dissertation

Approved by the Department of Geosciences

Joe Yelderman, Ph.D., Chairperson

Submitted to the Graduate Faculty of
Baylor University in Partial Fulfillment of the
Requirements for the Degree
of

Doctor of Philosophy

Approved by the Dissertation Committee

Stacy C. Atchley, Ph.D. Chairperson

Steve Dworkin, Ph.D.

James Fulton, Ph.D.

Jeff Donahoo, Ph.D.

Kathy Breen, Ph.D.

Accepted by the Graduate School
August 2022

J. Larry Lyon, Ph.D., Dean

Page bearing signatures is kept on file in the Graduate School.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

To my advisor, Dr. Stacy Atchley, for providing the opportunity to pursue a Ph.D. at Baylor. I am forever grateful for his mentoring and guidance during my time at Baylor and helping me grow as a scientist. To Dr. Scott James and his student Dr. Kathy Breen, for opening me up to the world of machine learning and modelling. I would like to acknowledge and thank Dr. Lisa Stright, my undergraduate mentor at Colorado State University, who first introduced me to petroleum geology. To Matador Resources and Los Alamos National Laboratory for providing me hands-on experience. I would like to thank the Department of Geosciences and Graduate School at Baylor University for their financial support.

Special thanks to Bart and Kaelin Yeates, Anna Thorson, Marilyn Wisler, E460 et al., The Scheaf Family, CREDO Women's Book Club, and fellow St. Jerome Core Team Members.

Lastly, to my parents, John and Alyson Rau, for providing the constant support and encouragement throughout this journey and to my siblings, Mary Cathryn and Jack, for always being there and having my back.

DEDICATION

To Dr. Scott C. James

CHAPTER ONE

Introduction

Energy is a primary factor for economic growth, and reliable access greatly impacts quality of life. Within the United State for example, the industrial revolution accelerated energy demand and the associated production of fossil fuels that provided a high standard of living that was envied worldwide (Jones, 2001). Currently, developing countries with burgeoning economies are experiencing a similar wave of industrialization and associated acceleration in energy demand (Rahman et al., 2021). This rapidly increasing energy demand among developing nations, combined with the continuously expanding demands of developed nations sets the stage for a global energy deficit. During 2021 global energy consumption was 601 quadrillion BTU, and energy consumption is projected to increase by 42% to 886 quadrillion BTU by 2050 (U.S. Energy Information Administration, 2021).

A sustained supply of energy is required to satisfy societal aspirations for a high standard of living. Since the 20th century, fossil fuels have been the primary global energy source and currently accounts for approximately 80% of global energy consumption (U.S. Energy Information Administration, 2021). Fossil fuels will continue to contribute toward global energy access and security; however, fossil fuel use will proportionately decrease because they are non-renewable and of finite supply, and because of the environmental priority to increasingly employ noncarbon, renewable energy resources (U.S. Energy Information Administration, 2021).

The results of this dissertation research contribute to the development of efficient and cost-effective approaches for energy exploration and production. With the increase in computational power and growth of data availability, data analytics and modelling are increasingly used to understand and predict subsurface geology and associated energy resources. In this research, machine learning and geophysical models are employed to aid in the characterization of high potential hydrocarbon and geothermal reservoirs.

Oil and natural gas are derived from organic material preserved in the subsurface. Through deep geologic time organic material subjected to high temperature and pressure transforms into hydrocarbons that migrate and accumulate in subsurface reservoirs. These reservoirs are the target for hydrocarbon exploration and production. Chapters 2 and 3 of this dissertation demonstrate the applicability of machine learning statistical techniques in the exploration and development of hydrocarbons in the Late Devonian Duvernay Formation, a shale reservoir within the Western Canada Sedimentary Basin. Chapter 2 evaluates supervised, tree-based machine learning methods in the prediction of core-calibrated facies and/or facies associations from wireline logs and investigates how thickness, proportion, and distinguishability impact class performance and facies association prediction. Chapter 3 compares two statistical approaches for the prediction of core-observed reservoir and non-reservoir facies within wireline well logs: 1) an unsupervised machine learning algorithm, and 2) a traditional deterministic approach. The comparison details how the unsupervised model provides more precise, consistent, and objective predictions. Both chapters 2 and 3 demonstrate how the application of advanced statistical models more reliably predict the occurrence of hydrocarbon reservoirs beneath the earth’s surface.

Geothermal energy is a nonintermittent, renewable resource that has the potential to help alleviate concerns associated with growing global energy demand while also mitigating carbon emissions that result from the burning of fossil fuels. The radioactive decay of unstable isotopes in the deep subsurface generates thermal energy that may heat and convect groundwater. The hot water and steam recovered through drilling is used to turn turbines that generate electricity. Chapter 4 demonstrates how the integration of unsupervised machine learning and passive geophysical methods can provide a 3D assessment of subsurface geothermal potential through the combined evaluation of heat flow, temperature, porosity, and permeability. The unsupervised machine learning approach identifies subsurface locations with the highest likelihood of geothermal success, whereas the geophysical method identifies specific subsurface geothermal prospects.

Attributions

Chapter two was published. Elisabeth G. Rau carried out the primary research and composed the manuscript. Dr. Kathy Breen and Dr. Scott C. James contributed to model development and editorial comments. Dr Stacy C. Atchley provided guidance and oversight of the project as well as editorial comments. Anna M. Thorson and David W. Yeates contributed to data collection and analysis and editorial comments.

Chapter three is submitted to AAPG Geohorizons. Elisabeth G. Rau carried out the primary research and composed the manuscript. David W. Yeates and Dr. Kathy Breen contributed technical expertise and editorial comments. David W. Yeates and Anna M. Thorson contributed to data collection. Dr. Stacy Atchley provided guidance and oversight of the project as well as editorial comments.

Chapter four will be submitted for publication to Geothermics and the listing authors will be: Elisabeth G. Rau, Bulbul Ahmmed, David W. Yeates, Velimir V. Vesselinov, Stacy C. Atchley, and Satish Karra. Elisabeth G. Rau carried out primary research and composed the manuscript. Bulbul Ahmmed contributed technical expertise, NMFk analysis, and editorial comments. David W. Yeates contributed to data analysis and editorial comments. Velimir V. Vesselinov and Satish Karra contributed technical expertise and Satish Karra provided guidance and oversight of the project. Stacy C. Atchley contributed to data analysis and editorial comments.

CHAPTER TWO

Applicability of Decision Tree-based Machine Learning Models in the Prediction of Core-Calibrated Shale Facies from Wireline Logs in the Late Devonian Duvernay Formation, Alberta, Canada

This chapter published as: Rau, E. G., James, S. C., Breen, K., Atchley, S. C., Thorson, A. M., Yeates, D. W., Applicability of decision tree-based machine learning models in the prediction of core-calibrated shale facies from wireline logs in the late Devonian Duvernay Formation, Alberta, Canada: Interpretations, 10, 1-45.

Abstract

Well logs provide insight into stratigraphically-compartmentalized rock properties and are a cost-effective alternative to core. The identification of reservoir (and non-reservoir) facies in core, and their calibration to well log response, has traditionally relied on expert domain knowledge and is inherently inconsistent. Such analyses are time consuming, tedious, error-prone, and often biased due to a lack of objectivity. Automated lithological interpretations from wireline logs appear to be a promising solution to identifying and understanding depositional complexity within a reservoir. Using the Duvernay Formation in the Western Canada Sedimentary Basin as a case study, we evaluate the applicability of decision tree-based, machine learning methods in the prediction of core-calibrated facies and/or facies association distributions within wireline logs. We use three independent, decision tree-based machine learning models to predict (1) facies (FACM), (2) facies associations (FAM), and (3) reservoir rock (RESM) from wireline-logs. Model accuracies are 60.3%, 88.1%, and 88.1% for FACM, FAM, and RESM respectively, but individual class F1 scores range from 0 to 0.92. We attribute discrepancies in individual class performance to interval thickness, sample proportion of

training data, and distinguishability of the output class. Classes thicker than 3m and encompassing at least 16% of the training dataset have F1 scores greater than 0.60. We attribute exceptions to these general cutoffs to the ability to recognize diagnostic sedimentologic features observed in core. Results from this study help in understanding stratigraphic complexity in absence of core aiding in subsurface characterization of reservoirs.

Introduction

Facies and facies associations are used to describe and interpret stratigraphic architectural complexity within a reservoir. Facies consist of a recurring assemblage of depositional attributes and are combined into associations that consist of more broadly varying, but genetically similar facies, i.e., facies associations. Together, facies and facies associations account for the variable distribution of rock properties within a stratigraphic framework. As pertains to unconventional shale reservoirs, a determination of the spatial distribution of organic-rich facies and/or facies association is essential in the efficient development of hydrocarbon resources.

Machine learning (ML) algorithms belong to a family of stochastic, data-driven algorithms. Supervised-ML classification techniques generate a mapping ($\mathbf{\Theta}$) to transform inputs ($\mathbf{X}$) such that the discrete output labels ($\mathbf{Y}$) are predicted with minimal error. Both Baldwin et al. (1990) and Rogers et al. (1992) demonstrated early the applicability of using ML to predict lithological changes in well logs. Recent studies have used a variety of ML algorithms to determine facies classifications from wireline log data (Dubois et al., 2007; Wang and Carr, 2013; Saneifar et al., 2015; Bhattacharya et al., 2016; Bhattacharya and Carr, 2019; Imamverdiyev and Sukhostat, 2019). As ML techniques gained popularity in the oil and gas industry, advances in computational hardware have facilitated the ascent of increasingly complex models that improve overall classification performance.

Decision trees are a type of supervised, ML algorithm that recursively splits the input data into smaller subsets until a prediction is possible (Quinlan, 1990 and Loh, 2011). The utility of decision trees is well documented in the literature as an effective ML model for classification tasks (Breiman et al., 1984; Quinlan, 1990; Geurts et al., 2006; Kotsiantis, 2007, 2013; Loh, 2011). Decision tree-based algorithms are common supervised ML algorithms that combine multiple decision trees to optimize model performance (Ho, 1998 and Geurts et al., 2006). Decision tree-based algorithms have been successful in the classification of facies from wireline logs (Hall and Hall 2017). Specifically, tree-based algorithms such as Extra Trees and Random Forests have been used for lithological identification resulting in overall model accuracies between 62% to 96% (Hall, 2016; Bestagini et al., 2017; Hall and Hall, 2017; Sun et al., 2019; Tewari et al., 2019; Bressan et al., 2020; Holotel et al., 2020; Ippolito et al., 2021). Although these results are encouraging for an automated system, there are still major discrepancies between the ability to classify specified classes. For example, Bestagini et al. (2017) reported a median accuracy of 62.3% in their prediction of 9 facies; however, classification accuracy for individual facies ranged from 12 – 77%. Bressan et al. (2020) had 4 different class lithologies with classification accuracies ranging from 71 – 86%. Sun et al. (2019) and Ippolito et al. (2021) had higher classification accuracy for individual classes of 70 – 95% and 82 – 100%, respectively.

Geologic setting

The Late Devonian (Frasnian) Duvernay Formation is a source rock for the conventional reservoirs that have accounted for the historically prolific hydrocarbon production within the Western Canada Sedimentary Basin (WCSB). The formation is composed of organic rich-mudrocks and carbonates and has TOC values ranging from 3-5 wt. % (Preston et al., 2016). As a results, since 2011 the Duvernay Formation has been developed as a shale reservoir through horizontal multistage fracturing (Preston et al., 2016). The Duvernay Formation accumulated as a restricted intracratonic basinal deposits contemporaneous with the shallow marine platform carbonates of the Leduc and Grosmont Formations. The Duvernay is overlain by the basinal mudrock of the Ireton Formation which commonly serves as the caprock seal for conventional Leduc reservoirs within the basin (Fig. 2.1) (Stoakes, 1980; Witzke and Heckel, 1988; Switzer et al., 1994; Weissenberger and Potma, 2001; Blakey, 2011; Weissenberger et al., 2016; Wong et al., 2016). During deposition, the WCSB was partitioned into the East and West Shale Basins by a linear Leduc Reef complex known as the Rimbey-Meadowbrook trend (Fig. 2.2) (Porter et al., 1982; Allan and Creaney, 1991). Late Cretaceous and Early Paleogene Laramide deformation resulted in crustal thrust-loading immediately west of the study area. This caused asymmetrical regional subsidence and southwestern dip of the Duvernay Formation (Porter et al., 1982; Stoakes and Creaney, 1984; Weissenberger and Potma et al., 2001). Trends of Duvernay thermal maturity mimic structural depths within the asymmetric basin (Stoakes and Creaney, 1984; Rokosh et al., 2012). Recent studies have evaluated the reservoir potential of the Duvernay Formation (Dunn et al., 2016; Preston et al., 2016; Chopra et al., 2017; Wong et al., 2016; Bauman, 2018; Datta, 2018;

Harris et al., 2018; Dong et al., 2019; Knapp et al., 2019; Li et al., 2020) including the identification and mapping of Duvernay facies based upon the evaluation and integration of whole core and wireline log data (Thorson, 2019).

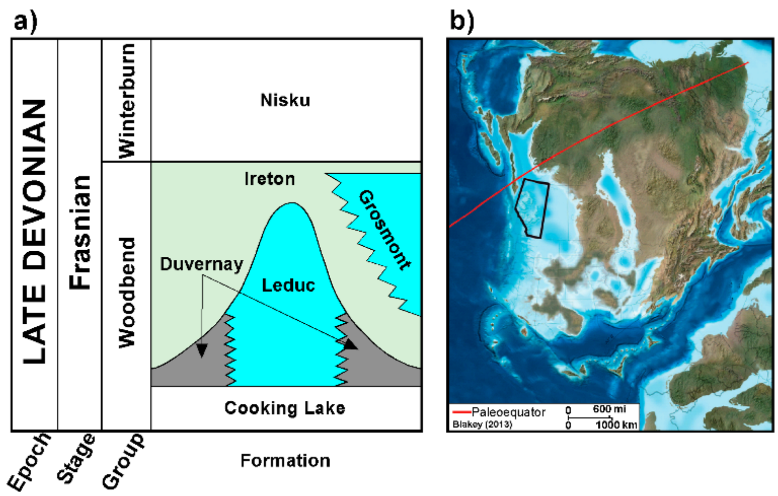

Figure 2.1. (a) Late Devonian (Frasnian) stratigraphic correlation chart for south-central Alberta, Canada showing relationships between the contemporaneous Duvernay, Leduc, Ireton and Grosmont Formations (modified from Chow et al., 1995). (b) Late Devonian paleogeography of North America (Wong et al., 2016) with the position of present-day Alberta outlined in black (modified from Blakey, 2013).

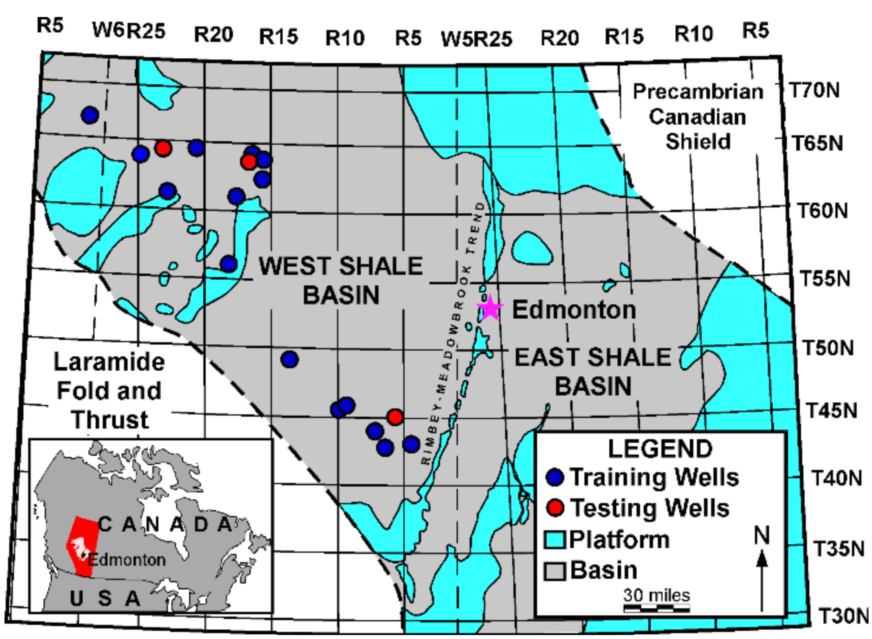

Figure 2.2. Late Devonian paleogeography of Alberta, Canada (modified from Switzer et al., 1994, Rokosh et al., 2012, Preston et al., 2016, Wang et al., 2016). The Rimbey-Meadowbrook trend is a linear stromatoporoid barrier reef trend that partitions the West and East Shale Basins. The Duvernay Formation was deposited as basinal fill within both basins. Circles identify well locations used in this study.

Objective

Using the Late Devonian Duvernay Formation as a test case, the objective of this study is to further evaluate the applicability of tree-based ML methods in the prediction of core-calibrated facies and/or facies association distributions using wireline logs, and investigate how thickness, proportion, and distinguishability impact the performance of classes. Proportion is the percentage of the total dataset the respective class represents. It is important there be an adequate number of training samples for the model to effectively learn relationships and apply them to unseen data. Distinguishability is the ability to differentiate a specific class based on their sedimentological attributes. Classes with distinguishable sedimentological attributes have less overlap in log responses with other classes, and thus easier to predict using a tree-based ML model.

Methods

We use three different datasets for (1) facies, (2) facies associations, and (3) reservoir/non-reservoir designation that are based on the core-derived facies model of Thorson (2019). Three decision tree-based models, one for each dataset, are implemented by modifying the available code from the top-performing teams of the 2016 ML competition (*sensu* Hall and Hall, 2017) with the objective of automating and refining the log-based classification of facies (FACM), facies associations (FAM), and reservoir/non-reservoir units (RESM). Code used in this project is available at https://github.com/elisabethrau/FaciesClassificationMachineLearning.

Decision tree-based machine learning

In a supervised classification problem, input features and known output labels are used to train a model to predict unknown outputs from new values not used in training (Loh, 2011). Ideally, input features cluster into k distinct areas, A_1, A_2 , … A_k, with each area representing the output class k (Loh, 2011). If new input features cluster into A_k, then the predicted output label belongs to class k. Decision trees for classification tasks establish distinct areas by recursively splitting the data into smaller subsets one input feature at a time (Loh, 2011). The Gini Index and Entropy is used to determine optimal features and feature values for each split (Breiman et al., 1984). For this study, input features are the various well log data types, and output features are respective model output class labels, i.e., facies, facies association, and reservoir/non-reservoir designations.

Care must be taken to mitigate the tendency of decision trees to overfit the training data, where patterns in data are “memorized” as opposed to “learned” (Ho, 1998; Bestagini et al., 2017). Decision tree-based methods such as Extra Trees and Random Forest classifiers combine multiple decision-tree predictions adding randomness to the model which reduces overfitting (Breiman et al., 1984; Ho, 1998; Geurts et al., 2006). In this study tree-based pipeline optimization (TPOT) determines the top-performing decision tree-based algorithm and associated hyperparameters (Olson and Moore, 2016). Extra Trees classifier is the classifier for the FACM, and Random Forest classifier is the classifier for the RESM and FAM. Table 2.1 lists the optimized hyperparameters for each model as a result of TPOT.

Table 2.1. List of hyperparameters for each model as a result of TPOT.

Hyperparameter	RESM	FAM	FACM
Classifier	Random forest	Random forest	Extra trees
n_estimators	100	100	100
criterion	entropy	entropy	gini
min_samples_split	2	7	5
min_samples_leaf	6	4	1
max_features	0.10	0.95	0.45

Data acquisition and preprocessing

A total of 18 wells in the West Shale Basin are used in this study (Fig. 2.2, Table 2.2). Each well has digital wireline log data ensuring modern, high-quality, open-hole well logs. Wireline logs include gamma ray, deep resistivity, photoelectric effect, neutron-density porosity and bulk density. Log samples are selected every tenth of a meter and are the input features for the model.

Table 2.2. List of West Shale Basin wells used in this study.

Training	Testing
00-01-12-064-25W5	00-13-23-064-23W5
00-01-18-061-17W5	00-13-32-063-16W5
00-01-24-061-23W5	02-13-07-045-05W5
00-01-35-045-10W5	
00-02-17-043-04W5	
00-04-21-064-16W5	
00-04-32-064-20W5	
00-06-15-056-18W5	
00-08-05-043-06W5	
00-13-05-064-15W5	
00-14-10-044-07W5	
00-14-19-062-15W5	
00-15-18-049-13W5	
02-04-09-046-09W5	
02-16-28-066-02W6	

In total, the dataset includes 9,858 wireline log samples and associated class labels. The dataset is split into training (75%) and testing (25%) subsets with similar well log data distributions. The training dataset is comprised of samples used for learning a mapping between wireline log data and class labels (Ripley, 1996). In this study, 15 wells (7,386 samples) are used for training and three wells (2,472 samples) are used for blind testing as shown in Figure 2.2 and Table 2.2. Prior to training, three wells are removed from the training dataset for cross validation and are used to validate and optimize the trained model.

Input features are preprocessed to remove outliers and erroneously high or low values commonly present at the beginning and end of a wireline run caused by surface casing and bottom hole effects. Outliers in this study are statistically defined as datapoints falling 1.5 interquartile ranges below or above the 1st and 3rd quartile, respectively. Some datapoints identified statistically as outliers are not discarded based on expert knowledge. For example, based on the outlier definition, gamma ray values above 221 API are statistically outliers; however, those values are not removed because geologically they are true observations that commonly characterize the best reservoir-quality rock (Thorson, 2019). Once the dataset is cleaned, each petrophysical log is scaled to the standard normal distribution.

Additional input features are used to enhance model performance and include the outputs from RESM and FAM and engineered features. The three models used in this study are independent of each other. Looking at them compared to scale, RESM encompasses the largest scale, and FACM encompasses the smallest scale with FAM falling in between. The outputs from the larger scale models are used as inputs for the smaller scale models. The RESM outputs are inputs for both smaller scale FAM and FACM and the FAM outputs are inputs for the smallest scale FACM. Furthermore, because the vertical distribution of facies is not random, additional input features relating to gradients of petrophysical responses are engineered from wireline logs following the methods outlined in Bestagini et al. (2017).

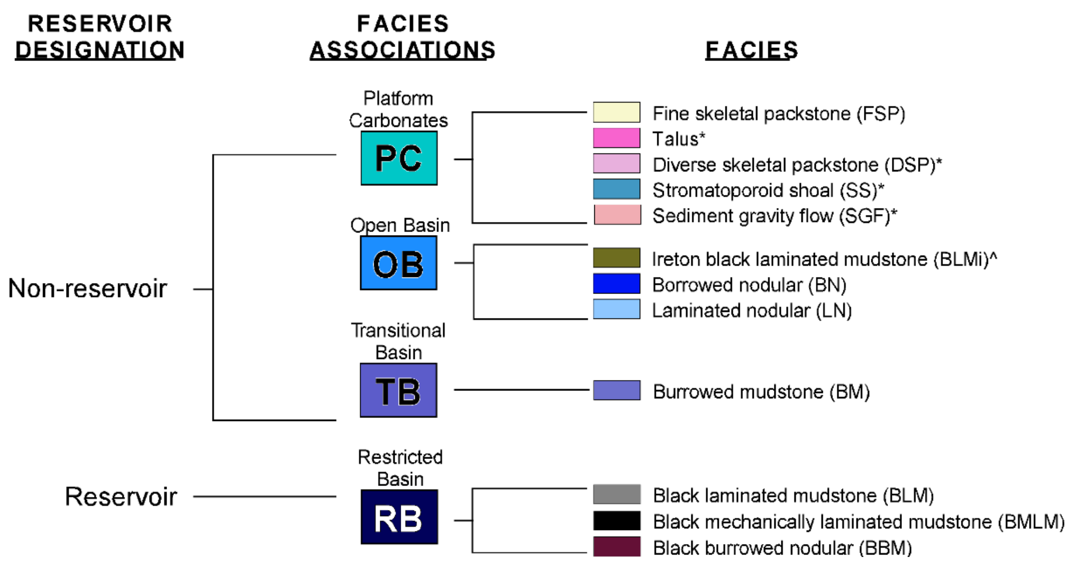

Figure 2.3. Organization chart showing groupings of facies into facies associations and facies associations into reservoir/non-reservoir designations. Several of the platform carbonate facies are only present in one or two wells and are thus not included in the model. These are indicated by an asterisk (*). The caret (^) indicates facies that are not contemporaneous with the Duvernay Formation. A complete listing of facies and their diagnostic attributes is provided within (Thorson et al., 2019).

Output labels comprise facies, facies association, and reservoir/non-reservoir designation. Figure 2.3 displays an organizational chart of the relationship between facies, facies associations and reservoir/non-reservoir designations previously established by Thorson et al. (2019). For this research, 18 continuously cored wells (1,500 m (4,921 ft) of total core length) were described in detail to account for the occurrence of facies, texture, allochem type and abundance, mechanical and biological structures, and fracture density. Core descriptions were digitized and merged with their corresponding petrophysical data using synthetic gamma ray total counts derived from uranium, thorium, and potassium concentrations (*sensu* Crain, 2018) collected from core at 1-meter resolution using a Bruker hand-held XRF. Facies designations (9 total; Thorson et.al. (2019)) were grouped into the following four associations: (1) open basin, (2) transitional basin, (3) restricted basin, and (4) platform carbonates (Fig. 2.3). Some facies comprising the platform carbonates only appear in one or two wells and thus could not be evaluated for their predictive potential (Fig. 2.3). The open basin, transitional basin, and restricted basin represent Duvernay-specific facies, whereas platform carbonates are associated with the contemporaneous Leduc and Grosmont Formations. The restricted basin association has the highest total organic carbon (TOC) estimates, and accordingly, most favorable hydrocarbon production potential (Thorson, 2019). Table 2.3 show the median log responses for each facies.

Table 2.3. Median log responses for each corresponding facies.

Facies	Facies Association	Gamma Ray (API)	Deep Resistivity (ohm-m)	Neutron Porosity (fraction)	Density Porosity (fraction)	RHOB kg/m^3	PE (barns/electron)
BM	Transitional Basin	109.8	245	0.12	0.05	2662	4.36
LN	Open Basin	47.4	41	0.07	0.003	2705	5.19
BLM	Restricted Basin	113.6	42	0.16	0.08	2538	4.04
BMLM	Restricted Basin	114.7	406	0.14	0.09	2555	3.65
BN	Open Basin	41.3	70	0.06	0.01	2691	5.00
BBM	Restricted Basin	151.9	79	0.16	0.08	2583	4.28

Oversampling. For ideal supervised classification tasks, class labels are evenly distributed within classes, i.e., each class contains approximately the same number of samples. Observational datasets, including the data used in this study, often have unbalanced classes where certain classes are more frequent (majority class) than others (minority class). The histogram of the original data in Figure 2.4 shows for each model at least one class more abundant than all others. Black laminated mudstone (BLM), restricted basin (RB), and reservoir (Res) rock comprised 45%, 83%, and 83% of their training datasets, respectively (Figure 2.4). Training an ML model with unbalanced class labels can lead to skewed accuracy metrics as accuracy may be artificially high because the majority class has a heavy influence, i.e., the majority class is predicted accurately and more often than classes with fewer training samples (Chawla et al., 2002; He and Garcia, 2009). For example, 83% of the RESM dataset is reservoir and if a model were to assign every sample as reservoir, the accuracy would look promising at 83%, but in reality, the model cannot distinguish classes.

Oversampling techniques, adaptive synthetic sampling (ADASYN) and synthetic minority oversampling technique (SMOTE) are applied to create synthetic samples of minority classes to balance label frequency as shown in the "oversampling" histogram in Figure 2.4 (Chawla et al., 2002; He et al., 2008; Krawczyk, 2016; Gosain and Sardana, 2017; Cahyana et al., 2019). SMOTE generates minority samples by finding k-nearest neighbors of existing classes, drawing a line between the neighbors of the same class, and selecting random points along that line (Chawla et al., 2002; Gosain and Sardana, 2017). Like SMOTE, ADASYN generates synthetic samples along a straight line between k-nearest neighbors. ADASYN differs in that it generates more samples for minority classes with more majority samples as neighbors to reduce the learning bias from the original imbalanced dataset (He et al., 2008). ADASYN is used in the RESM and FAM datasets. For FACM, SMOTE is better suited because each class size is relatively small and can be difficult to find the nearest neighbors needed for ADASYN.

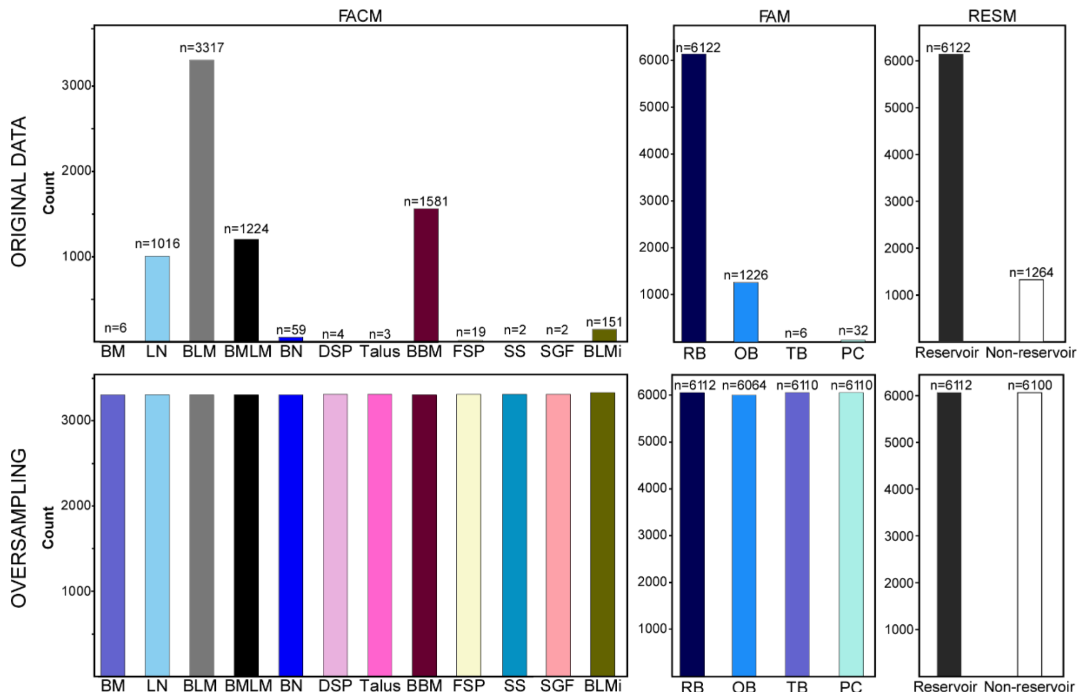

Figure 2.4. Histograms comparing the number of samples in each class for the original dataset and for the results of oversampling techniques (ADASYN or SMOTE) for facies, facies associations, and reservoir/non-reservoir designation training datasets. The original data histograms display unbalanced datasets compared to the oversampling histograms that show balanced datasets.

Evaluation metrics

Classification accuracy was evaluated using F1 scores for each class. An F1 score is calculated as a weighted average

$$F1 = \frac{2 \times p \times r}{p + r}$$

where *p* is the classification precision and *r* is the recall. Precision quantifies the number of correct predictions for a class and recall (sensitivity) quantifies a model's ability to correctly predict a class. The best performing classes have F1 scores above 0.90. Accuracy, the fraction of correctly classified output classes, is used to evaluate and compare overall model performances. It is important to note that while useful, accuracy values can be artificially high because of the influence from the majority class.

Results

Three test wells excluded from model training and validation are used to assess model performance (Fig. 2.2). For each model, 1000 predictions are made to analyze a distribution of results. Table 2.4 displays specific output labels' F1-scores for the best prediction and corresponding average thickness and proportion of training dataset. Table 2.5 displays statistical descriptions of the results from each model. Furthermore, displays confusion matrices for the best predictions and are used to quantitatively summarize the performance of the individual model classifications. No platform carbonate (PC) facies are included in the testing set due to their sparsity.

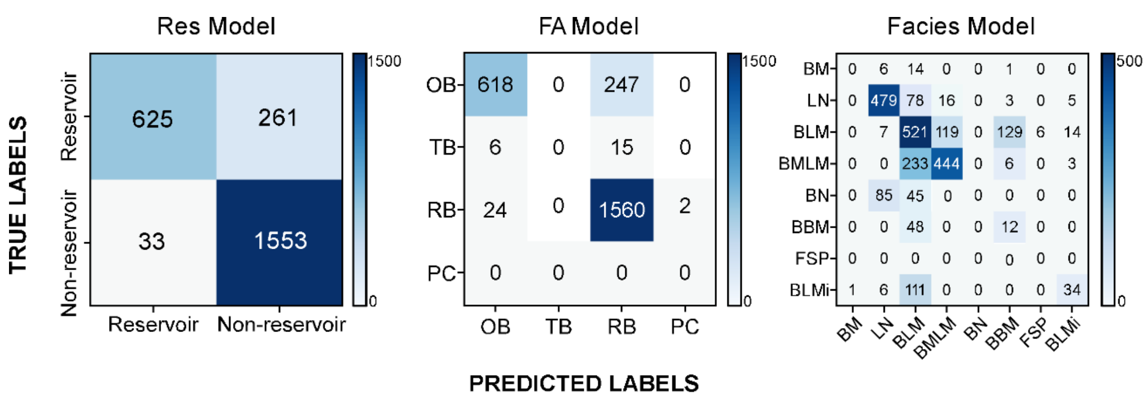

Figure 2.5 Confusion matrix providing a summary of the best classification results for each respective model. The correctly classified samples are along the diagonal, and the misclassified samples are off-diagonal. For example, for RESM, 625 reservoir and 1553 non-reservoir samples are correctly identified. On the contrary, 261 reservoir samples are misclassified as non-reservoir and 33 non-reservoir samples are misclassified as reservoir.

Table. 2.4. Facies-specific model test results for F1-scores, average thickness, and proportion of training dataset values. Classes are ranked based on F1-scores.

Class	**Precision**	**Recall**	**F1-score**	**Thickness (m)**	**Proportion of Training Dataset (%)**
RB	0.86	0.98	0.92	13.68	83.1
Reservoir	0.86	0.98	0.91	13.68	82.9
LN	0.82	0.82	0.82	4.18	13.8
OB	0.95	0.71	0.82	4.15	16.6
Non-reservoir	0.95	0.71	0.81	3.16	17.1
BMLM	0.77	0.61	0.68	3.80	16.6
BLM	0.5	0.65	0.56	3.71	44.9
BLMi	0.61	0.22	0.32	4.38	2.0
BBM	0.06	0.2	0.09	2.12	21.4
BN	0	0	0	1.41	0.8
TB	0	0	0	0.23	0.1
BM	0	0	0	0.23	0.1

Table 2.5. Predictive accuracy statistics for FACM, FAM and RESM based on 1000 model runs

Statistic	RESM	FAM	FACM
Highest	88.1%	88.1%	60.3%
Average	87.0%	87.6%	58.3%
Minimum	86.0%	87.1%	56.5%
Median	87.1%	87.7%	58.3%
Variance	1.1×10^{-5}	3.1×10^{-6}	3.9×10^{-5}
Error	0.22	0.22	0.40

Reservoir model

RESM correctly predicts 88.1% of the classes. Both classes receive relatively high F1 scores with non-reservoir rock being 0.81 and reservoir rock being 0.91 (Table

2.4). According to and Figure 2.5, 12% of samples are misclassified with most being reservoir rock misclassified as non-reservoir. Although the model correctly identifies general changes in reservoir and non-reservoir rock, discrepancies between the predicted and actual classifications are more likely to occur within thin beds at or below well log resolution, i.e., typically less than 1m. Additional discrepancies occur at class change boundaries. For example, as seen in Figure 2.6, the model recognizes that the deeper section of well 02/13-07-045-05W5 transitions from reservoir to non-reservoir at 2,978 m which is 2.5 m deeper than observed in core.

Facies association model

FAM achieves an accuracy of 88.1 % (Table 2.5), and F1 scores for each class are provided in Table 2.4. The confusion matrix in Figure 2.5 indicates the RB facies association has the fewest misclassifications with only 2% of the RB misclassified. Most (92% or 24 samples) misclassifications for RB are erroneously assigned as OB, with only two samples misclassified as carbonate. As shown in Figure 2.5 and Figure 2.6, the OB facies association has 29% of samples misclassified as RB. The TB facies association samples are thin and infrequent in the training dataset, so no test samples are correctly identified with 72% being misclassified as RB and 28% misclassified as OB. Similar to the RESM, the FAM also successfully identifies general facies association changes. For example, as shown in Figure 2.6 in well 02/13-07-045-05W5 at 2,942 m and 2,661 m and in well 00/13-32-063-16W5 at 2,865 m there are changes in facies association, and the change in facies association is predicted using the model. The FAM generally identifies the change to either RB or OB but fails to identify the transition to TB.

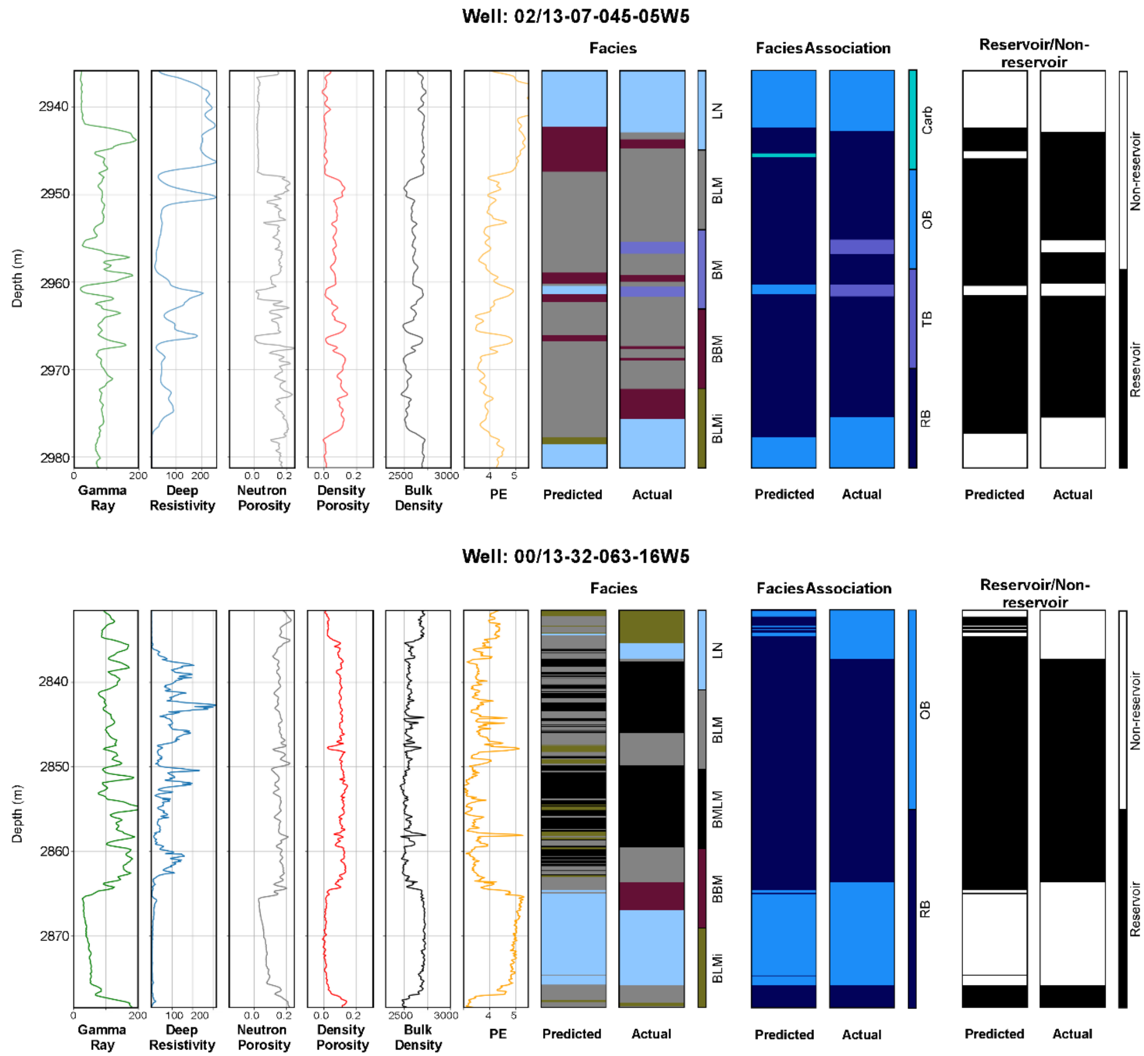

Figure 2.6. Comparison of log response with core-observed and model-predicted distributions of facies, facies associations and reservoir versus non-reservoir for wells for wells 02/12-07-045-05W5 and 00/13-32-063-16W5.

Facies model

FACM correctly identified 60.3% of the test set for all classes (Table 2.5). As listed in Table 2.4, the highest F1 scores are the laminated nodular (LN), black mechanically laminated mudstone (BMLM), and black laminated mudstone (BLM) facies with scores of 0.82, 0.68, and 0.56, respectively. F1 scores for Ireton black laminated mudstone (BLMi) and black burrowed mudstone (BBM) are 0.32 and 0.09 (Table 2.4). According to Table 2.4, burrowed mudstone (BM) and burrowed nodular (BN) both represent less than 1% of the training dataset and are not able to be classified. The confusion matrix in Figure 2.5 indicates overall 40% (982) of the test set is misclassified. Of the misclassifications, 54% (529) are predicted as BLM, 14% (139) predicted as BBM, and 14% (135) predicted as BMLM. Figure 2.6 shows FACM generally can identify the LN, BLM, and BMLM facies.

Feature importance

Feature importance scores are determined using the classification and regression trees (CART) algorithm in Scikit-learn. Figure 2.7 illustrates the feature importance scores for the input features in each respective model. The higher the relative score the more relevant the feature is to the desired output. For RESM there is little difference (~0.03) between the feature importance scores indicating all input features have some importance to the reservoir non-reservoir designation. For FAM and FACM, the reservoir designation and the facies association are the most important input features, respectively.

For all three models, there is little difference between the relative feature importance of each well log indicating all input logs are useful in the prediction of facies, facies associations, and reservoir non-reservoir designation. Recent works (Zhang et al., 2018; Jian et al., 2020; Hossain et al., 2020) establish methods to handle missing well log data that can be common in larger datasets, and Thorson (2019) established a semi-quantitative methodology to assign Duvernay facies using limited well logs.

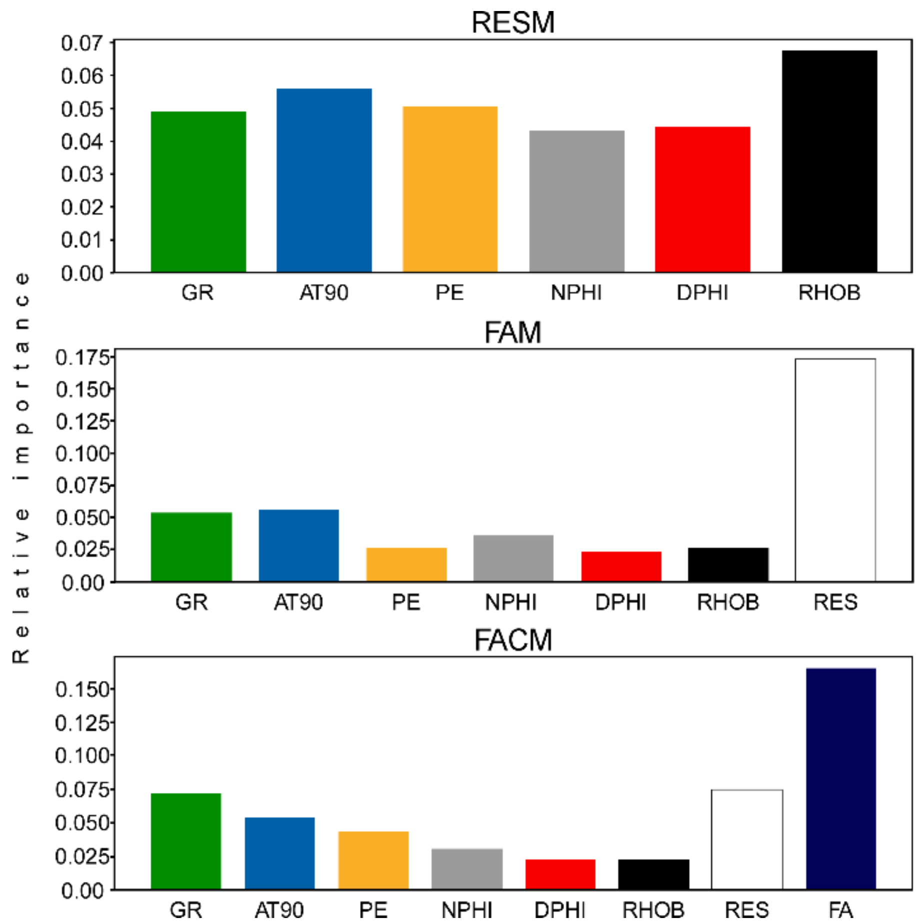

Figure 2.7. Feature importance scores for RESM, FAM and FACM. For each model, the features with larger importance values are more useful in classifying model output. GR = gamma ray, AT90 = deep resistivity, PE = photoelectric effect, NPHI = neutron porosity, DPHI = density porosity, RHOB = bulk density. Res = predicted outputs from the RESM used as inputs in FAM and FACM, and FA = predicted outputs from the FAM used as an input in FACM.

Discussion

These results demonstrate that it is possible to predict some facies, facies associations, and reservoir rock from wireline log using the method described here. Although overall performance accuracy exceeds 60%, some classes are misidentified.

The reason some classes are more successfully predicted than others is a combination of sample interval thickness and vertical resolution of the petrophysical data, sample proportion, and the distinguishability of the specific output class. Figure 2.8 shows two scatter plots detailing the relationship between F1 scores versus thickness and proportion respectively and the expected results. Classes falling above the trendline overperformed and classes falling below the trendline underperform.

Thickness

Petrophysical data used is limited by the vertical resolution of the instrumentation, with each log type having a finite resolution as determined by variables such as sonde design and sensor position, data sampling rate, logging speed, and data processing methods (Passey et al., 2006; Bond et al., 2010; Diniz Ferreira and Torres-Verdin, 2012). In terms of the well log data used in this study, the deep resistivity has the highest vertical resolution (0.6m [2 ft]), and therefore, defines the lower limit of attribute thickness resolved by the models (Passey et al., 2006). According to Figure 2.8, higher F1 scores are associated with thicknesses greater than approximately 3m (9 ft), a value greater than the resolution limit of deep resistivity. The discrepancy may be due to factors such as high petrophysical data sample rate (1 sample/0.1m) and subsequent interval averaging across bed edges, errors in core to well log depth correction, and incorrect bed boundary identification in core. The presence of thick (greater than 3m) beds mitigates the aforementioned factors and enhances the prediction of less distinguishable classes.

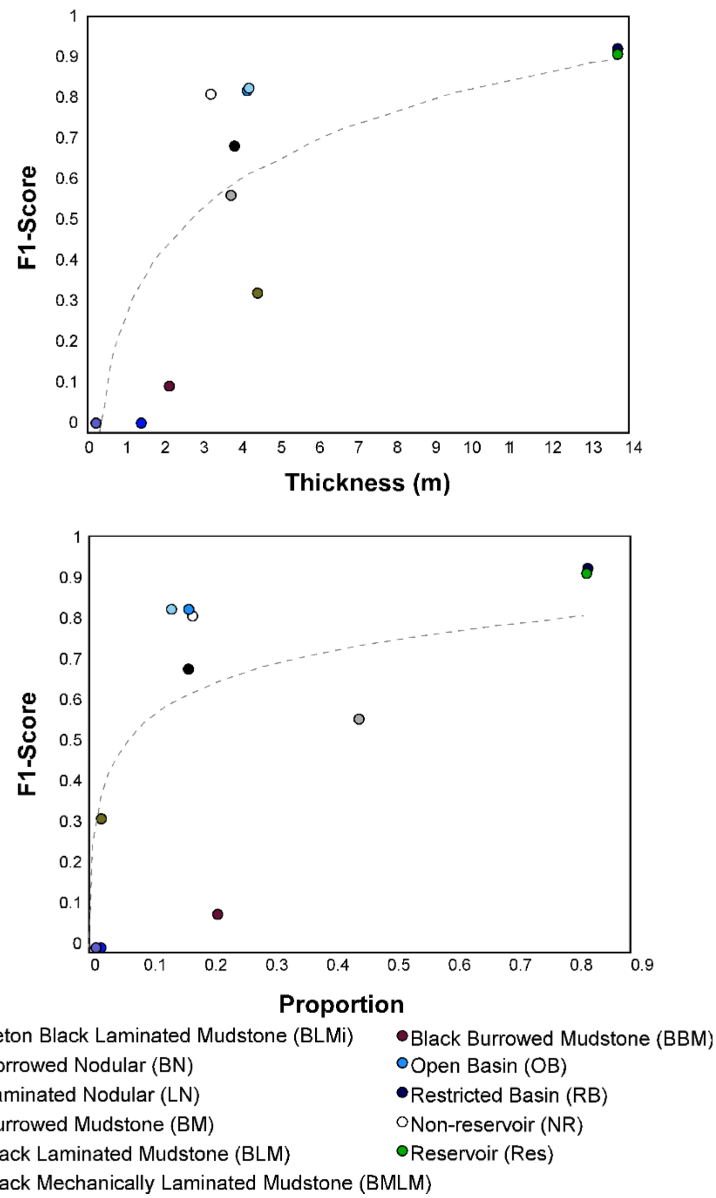

Figure 2.8. Scatter plots of thickness and proportion versus F1-score. The trend line is used as a baseline for expected results. In general, F1-scores increase as thickness and proportion increase.

Proportion

An adequate number of samples must be available for each class for the model to learn and apply relationships to unseen data. Figure 2.8 shows in general, as class representation increases, F1 score increases. At some lower limit, however, the sample proportion unsuccessfully predicts a given class. As listed in Table 2.4, BM and BN facies represent 0.10% and 0.80% of their respective training datasets. Each of their F1 scores are 0 and indicate a failure to identify using a tree-based ML model. As indicated in Figure 2.8, in general, F1-scores notably increase above a proportion of 16%.

Distinguishability

Classes that do not follow the proportion and thickness trends discussed above are likely affected by distinguishability, i.e., the human ability to label the class correctly and consistently in core. Figure 2.9 shows representative core photos of facies affected by distinguishability. Facies that are more distinguishable have higher F1 scores than expected for their respective thickness and proportion. According in Figure 2.8, LN overperforms based on the facies thickness and proportion. The presence of diagnostic sedimentologic features, such as the trace fossil *Chondrites* allow LN to be readily identified in core as seen in Figure 2.9d. Conversely, some facies lack diagnostic sedimentologic attributes and result in inconsistent facies designations amongst trained geologists. Less distinguishable facies underperform relative to their proportion and thickness. For example, as listed in Table 2.4, BLM has a proportion of 44% and an average thickness of 3.71m, both above the respective cutoffs, and yet has an F1 score of 0.56. The homogeneous dark color and ubiquitous and ambiguous millimeter lamina of BLM make differentiation from BBM difficult as shown in Figure 2.9. Furthermore, being the transitional facies between the BMLM and BBM end members, BLM has a wider range of petrophysical responses thus making it more difficult for the model to establish unique petrophysical cutoffs. Because the sedimentologic variability of BLM effects both a geologist’s facies assignment and petrophysical response, the BLM facies is commonly misclassified as BMLM or BBM, thereby lowering its precision and associated F1-score.

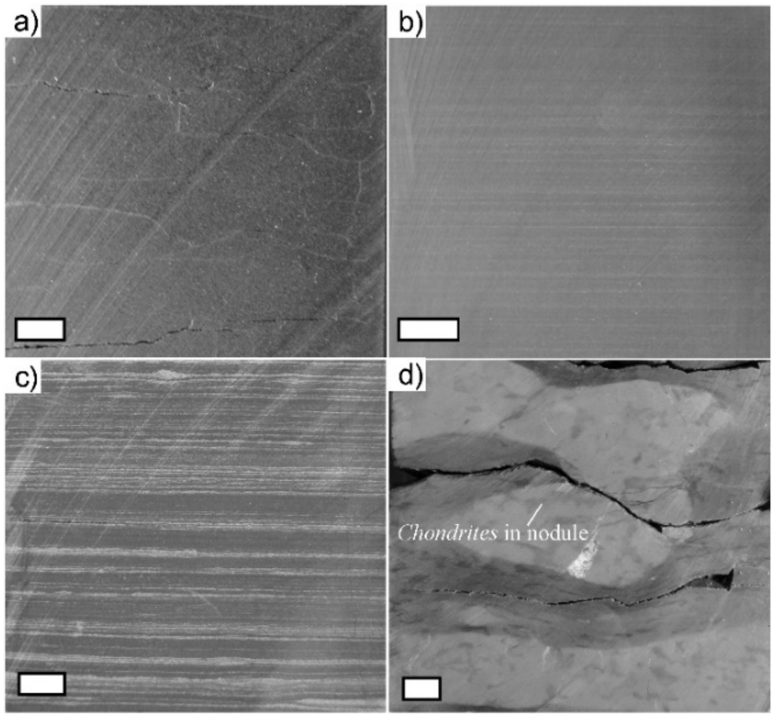

Figure 2.9. Representative core photo of (a) BBM, (b) BLM, (c) BMLM, and (d) LN. Scale bar is 1 cm.

Conclusion

With the increase in computation power and growth of data availability, ML has been increasingly used in the oil and gas industry to understand subsurface rock attributes. Decision tree-based algorithms are common ML methods used for lithological classification tasks. In this study, three decision tree-based models are applied to core-calibrated petrophysical data to predict facies, facies associations, and reservoir rock within the Late Devonian Duvernay Formation. Analyses are based upon 18 post-1980 vintage vertical wells with continuous whole core. Core sedimentologic features are digitized and depth-adjusted to their corresponding petrophysical data that includes gamma ray, deep resistivity, photoelectric effect, neutron-density porosity and bulk density. Using this dataset, each of the three models ran 1000 times. Models for both reservoir rock (RESM) and facies associations (FAM) performed the best with accuracies of 88.1%. The model for depositional facies (FACM) had an accuracy of 60.3%. Relative feature importance scores for each model indicate that all well logs are useful for the prediction of facies, facies associations, and reservoir/non-reservoir designation.

For decision tree-based techniques to be useful for subsurface rock-type predictions from wireline logs, it is important to know the model's predictive capabilities and limitations. In the case of the Duvernay Formation, some output classes are easier to classify using decision tree-based models than others. In general, classes thicker than 3m (10 ft) and encompassing at least 16% of the training dataset have a greater likelihood to be predicted. Exceptions to these cutoffs are attributed to diagnostic sedimentological features observed in core. Facies with ambiguous features have lower than expected F1 scores, whereas facies with more distinctive characteristics have higher than expected F1 scores. Overall, facies designations that are confidently assigned based upon conspicuous sedimentological features observed in whole core play a large role in the prediction of a desired class, because such facies are the most reliable for a model to learn and interpret. For facies lacking diagnostic attributes, and therefore lower confidence in their classification, increases in thickness and proportion improve the models' predictive ability by alleviating errors associated with inconsistent core observations and description and bed boundary identification. Results from this research suggest that core-observed depositional attributes may be predicted in wells lacking core control by having the necessary well logs, and as such, may be a valuable tool in petroleum reservoir correlation and mapping.

Acknowledgments

This study was completed in association with the Baylor University Applied Petroleum Studies Program. Pioneer Natural Resources, Anadarko Petroleum, and Angel Resources provided sponsorship. Petrophysical data was acquired with the help of IHS Markit AccuMap and geoLOGIC Geoscout software and GLJ Petroleum Consultants. Data analytics were completed using Scikit-learn machine learning library in Python 3.7.

References

Allan, J., and S. Creaney, 1991, Oil Families of the Western Canada Basin: Bulletin of Canadian Petroleum Geology, 39, 107–122.

Baldwin, J. L., R. M. Bateman, C. L. Wheatley, 1990, Application of a neural network to the problem of mineral identification form well logs: The Log Analyst, 31, 279-293.

Bhattacharya, S., and T. R. Carr, 2019, Integrated data-driven 3D shale lithofacies modeling of the Bakken Formation in the Williston basin, North Dakota, United States: Journal of Petroleum Science and Engineering, 177, 1072-1086, doi: 10.1016/j.petrol.2019.02.036.

Bhattacharya S., T. R. Carr, and M. Pal, 2016, Comparison of supervised and unsupervised approaches for mudstone lithofacies classification: Case studies from the Bakken and Mahantango-Marcellus Shale, USA: Journal of National Gas Science and Engineering, 33, 1119-1133, doi: 10.1016/j.jngse.2016.04.055.

Bauman, P., 2018, East Duvernay Shale Basin, https://chinookpetroleum.com/devonian-duvernay-formation-east-shale-basin-alberta/, accessed January 19, 2021.

Bestagini, P., V. Lipari, and S. Tubaro, 2017, A machine learning approach to facies classification using well logs: Society of Exploration Geophysicists Technical Program, SEG, Expanded Abstracts, 2137–2142.

Blakey R.C., 2011. Paleozoic Maps: Late Devonian – 375 Ma (380-370):http://deeptimemaps.com/north-american-key-time-slices-series/, accessed December 1, 2017.

Bond, L. J., K. M. Denslow, J. W. Griffin, G. E. Dale, R. V. Harris, T. L. Moran, D. M. Sheen, and T. Schenkel, 2010, Evaluation of Non-Nuclear Techniques for Well Logging: Technology Evaluation, PNNL-19867: US Department of Energy, doi:10.2172/1006309.

Breiman, L., J. Friedman, R. Olshen, and C. Stone, 1984, Classification and regression trees: Wadsworth and Brooks.

Bressan, T. S., M. K. de Souza, J. T. Girelli, and F. C. Junior, 2020, Evaluation of machine learning methods for lithology classification using geophysical data: Computers and Geosciences, 139, 104475, doi: 10.1016/j.cageo.2020.104475.

Cahyana, N., S. Khomsah, and A. S. Aribowo, 2019, Improving Imbalanced Dataset Classification Using Oversampling and Gradient Boosting: 5th International Conference on Science in Information Technology (ICSITech), 217–222, doi:10.1109/ICSITech46713.2019.8987499.

Crain, E.R., 2018, Natural gamma ray and spectral gamma ray logs: Crain's Petrophysical Handbook,https://www.spec2000.net/07-grlog.htm, accessed July 12, 2018.

Chawla, N. V., K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, 2002, SMOTE: Synthetic Minority Over-sampling Technique: Journal of Artificial Intelligence Research, 16, 321–357, doi:10.1613/jair.953.

Chopra, S., R. K. Sharma, A. K. Ray, H. Nemati, R. Morin, B. Schulte, and D. D'Amico, 2017, Seismic reservoir characterization of Duvernay shale with quantitative interpretation and induced seismicity considerations — A case study: Interpretation, 5, 185–197, doi:10.1190/INT-2016-0130.1.

Datta, A., 2018, The Resurgence of Interest in the Duvernay: A Review of Pertinent Economic and Environmental Issues: Canadian Energy Research Institute.

Diniz-Ferreira, E. L., and C. Torres-Verdín, 2012, Improved estimation of pore connectivity and permeability in deepwater carbonates with the construction of multi-layer static and dynamic petrophysical models: Society of Paraphysicists and Wel Log Analysts 53rd Annual Logging Symposium.

Dong, T., N. B. Harris, J. M. McMillan, C. E. Twemlow, B. R. Nassichuk, and D. L. Bish, 2019, A model for porosity evolution in shale reservoirs: An example from the Upper Devonian Duvernay Formation, Western Canada Sedimentary Basin: AAPG Bulletin, 103, 1017–1044, doi:10.1306/10261817272.

Dubois, M. K., G. C. Bohling, and S. Chakrabarti, 2007, Comparison of four approaches to a rock facies classification problem: Computers & Geosciences, 33, 599–617, doi:10.1016/j.cageo.2006.08.011.

Dunn, L., G. Schmidt, K. Hammermaster, M. Brown, R. Bernard, E. Wen, R. Befus, and S. Gardiner, 2016, The Duvernay Formation (Devonian): Sedimentology and Reservoir Characterization of a Shale Gas/Liquids Play in Alberta, Canada: Annual Convention and Exhibition, AAPG, Extended Abstract.

Geurts, P., D. Ernst, and L. Wehenkel, 2006, Extremely randomized trees: Machine Learning, 63, 3–42, doi:10.1007/s10994-006-6226-1.

Gosain, A., and S. Sardana, 2017, Handling class imbalance problem using oversampling techniques: A review: International Conference on Advances in Computing, Communications and Informatics, 79–85, doi:10.1109/ICACCI.2017.8125820.

Hall, B., 2016, Facies classification using machine learning: The Leading Edge, 35, 906–909, doi:10.1190/tle35100906.1.

Hall, M., and B. Hall, 2017, Distributed collaborative prediction: Results of the machine learning contest: The Leading Edge, 36, 267–269, doi:10.1190/tle36030267.1.

Halotel, J., V. Demyanov, and A. Gardiner, 2020, Value of geologically derived features in machine learning facies classification: Mathematical Geosciences, 52, 5–29, doi:10.1007/s11004-019-09838-0.

Harris, N. B., J. M. McMillan, L. J. Knapp, and M. Mastalerz, 2018, Organic matter accumulation in the Upper Devonian Duvernay Formation, Western Canada Sedimentary Basin, from sequence stratigraphic analysis and geochemical proxies: Sedimentary Geology, 376, 185–203, doi:10.1016/j.sedgeo.2018.09.004.

He, H. and E. A. Garcia, 2009, Learning from imbalance data: IEEE Transactions on knowledge and data engineering, 21, 1263–1284.

He, H., Y. Bai, E. A. Garcia, and S. Li, 2008, ADASYN: Adaptive synthetic sampling approach for imbalanced learning: Institute of Electrical and Electronics Engineers International Joint Conference on Neural Networks, IEEE World Congress on Computational Intelligence, 1322–1328, doi:10.1109/IJCNN.2008.4633969.

Ho, T. K., 1998, The random subspace method for constructing decision forests: Institute of Electrical and Electronics Engineers Transactions on Pattern Analysis and Machine Intelligence, 20, 832–844, doi:10.1109/34.709601.

Hossain, T. M., J. Watada, Z. Jian,, H. Sakai, S. Rahman, and I. A. Aziz, 2020, Missing well log data handling in complex lithology perdition: an NIS apriori algorithm approach: International Journal of Innovative Computing, Information and Control, 16, 1077-1091, doi: 10.24507/ijicic.16.03.1077.

Imamverdiyev, Y., and L. Sukhostat, 2019, Lithological facies classification using deep convolutional neural network: Journal of Petroleum Science and Engineering, 174, 216–228, doi:10.1016/j.petrol.2018.11.023.

Ippolito, M., J. Ferguson, and F. Jenson, 2021, Improving facies prediction by combining supervised and unsupervised learning methods: Journal of Petroleum Science and Engineering, 200, 108300, doi: 10.1007/s10708-021-10502-6.

Jian, H., L. Chenghui, C. Zhimin, and M. Haiwei, 2020, Integration of deep neural networks and ensemble learning machines for missing well logs estimation: Flow Measurement and Instrumentation, 101748, doi:10.1016/j.flowmeasinst.2020.101748.

Knapp, L. J., N. B. Harris, and J. M. McMillan, 2019, A sequence stratigraphic model for the orangic-rich Upper Devonian Duvernay Formation, Alberta, Canada: Sedimentary Geology, 387, 152–181, doi:10.1016/j.sedgeo.2019.04.008.

Kotsiantis, S. B., 2013, Decision trees: a recent overview: Artificial Intelligence Review, 39, 261–283, doi:10.1007/s10462-011-9272-4.

Kotsiantis, S. B., 2007, Supervised Machine Learning: A Review of classification techniques: Emerging artificial intelligence applications in computer engineering, 160, 3–24.

Krawczyk, B., 2016, Learning from imbalanced data: open challenges and future directions: Progress in Artificial Intelligence, 5, 221–232, doi:10.1007/s13748-016-0094-0.

Li, G., K. Luo, and D. Shi, 2020, Key technologies, engineering management and important suggestions of shale oil/gas development: Case study of a Duvernay shale project in Western Canada Sedimentary Basin: Petroleum Exploration and Development, 47, 791–802, doi:10.1016/S1876-3804(20)60094-5.

Loh, W. Y., 2011, Classification and regression trees: Wiley interdisciplinary reviews: data mining and knowledge discovery, 1, 14–23, doi:10.1002/widm.8.

Olson, R. S., and J. H. Moore, 2016, TPOT: A Tree-based pipeline optimization tool for automating machine learning: Proceedings of the Workshop on Automatic Machine Learning, 64, 66–74.

Passey, Q. R., K. E. Dahlberg, K. B. Sullivan, H. Yin, R. A. Brackett, Y. H. Xiao, and A. G. Guzmán-Garcia, 2006, Petrophysical evaluation of hydrocarbon pore-thickness in thinly bedded clastic reservoirs: AAPG Archie Series, 1, doi:10.1306/A11157.

Porter, J. W., R. A. Price, and R. G. McCrossan, 1982, The Western Canada Sedimentary Basin: Philosophical Transactions of the Royal Society of London. Series A: Mathematical and Physical Sciences, 305, no. 1489, 169–192.

Preston, A., G. Garner, K. Beavis, O. Sadiq, and S. Stricker, 2016, Duvernay reserves and resources report: A comprehensive analysis of Alberta’s foremost liquids-rich shale resource: Alberta Energy Regulator.

Quinlan, J. R., 1990, Decision trees and decision-making: Institute of Electrical and Electronics Engineers Transactions on Systems, Man, and Cybernetics, 20, 339–346, doi:10.1109/21.52545.

Ripley, B. D., 1996, Pattern Recognition and Neural Networks: Cambridge University Press, doi:10.1017/CBO9780511812651

Rogers, S. J., J. H. Fang, C. L. Karr, and D. A. Stanley, 1992, Determination of lithology from well logs using a neural network: AAPG Bulletin, 76, 731-739.

Rokosh, C.D., S. Lyster, S. D. A. Anderson, A. P. Beaton, H. Berhane, T. Brazzoni, D. Chen, Y. Cheng, T. Mack, C. Pana and J. G. Pawlowicz, 2012, Open file report: Summary of Alberta’s shale- and siltstone-hosted hydrocarbon resource potential: Alberta Geological Survey.

Saneifar, M., A. Aranibar, and Z. Heidari, 2014, Rock classification in the Haynesville Shale based on petrophysical and elastic properties estimated from well logs: Interpretations, 3, 65- 75, doi:10.1190/INT-2013-0198.1.

Stoakes, F. A., 1980, Nature and control of shale basin fill and its effect on reef growth and termination: Upper Devonian Duvernay and Ireton Formations of Alberta, Canada: Bulletin of Canadian Petroleum Geology, 28, 345–410.

Stoakes, F. A., and S. Creaney, 1984, Sedimentology of a carbonate source rock: the Duvernay Formation of Alberta Canada: Carbonates in Subsurface and Outcrop: Canadian Society of Petroleum Geologists, 132-147.

Sun, J., Q. Li, M. Chen, L. Ren, G. Huang, C. Li, and Z. Zhang, 2019, Optimization of models for a rapid identification of lithology while drilling – A win-win strategy based on machine learning: Journal of Petroleum Science and Engineering, 176, 321-341, doi:10.3390/en13040888.

Switzer, S. B., W. G. Holland, D. S. Christie, G. C. Graf, A. S. Hedinger, R. J. McAuley, R. A. Wierzbicki, and J. J. Packard, 1994, Devonian Woodbend - Winterburn Strata of the Western Canada Sedimentary Basin: Geological Atlas of the Sedimentary Basin, 165–202.

Tewari, S., and U. Dwivedi, 2019, Ensemble-based big data analytics of lithofacies for automatic development of petroleum reservoirs: Computers and Industrial Engineering, 128, 937-947, doi: 10.1016/j.cie.2018.08.018.

Thorson, A. M., 2019, Stratigraphic Partitioning and Distribution of Reservoir Attributes within the Late Devonian Duvernay Formation, Western Canada Sedimentary Basin: M.S. thesis, Baylor University.

Wang G., and T. R. Carr, 2013, Organic-rich Marcellus Shale lithofacies modeling and distribution pattern analysis in the Appalachian Basin: AAPG Bulletin, 97, 2173 – 2205, doi:10.1306/05141312135.

Weissenberger, J. A. W., and K. Potma, 2001, The Devonian of western Canada–aspects of a petroleum system: Introduction: Bulletin of Canadian Petroleum Geology, 49, 1–6.

Weissenberger, J. A. W., P. K. Wong, and M. G. Gilhooly, 2016, Stratigraphic architecture of the Frasnian South Jasper Basin, North-Central Alberta front ranges: New advances in Devonian carbonates: Outcrop Analogs, Reservoirs, and Chronostratigraphy, SEPM Special Publication 107, doi:10.2110/sepmsp.107.07.

Witzke, B. J., and P. H. Heckel, 1988, Paleoclimatic Indicators and Inferred Devonian Paleolatitudes of Euroamerica: Canadian Society of Petroleum Geologists, 14, 49–63.

Wong, P. K., J. A. W. Weissenberger, and M. G. Gilhooly, 2016, Revised Regional Frasnian Sequence Stratigraphic Framework, Alberta Outcrop and Subsurface: New advances in Devonian carbonates: Outcrop analogs, reservoirs, and chronostratigraphy, SEPM Special Publication 107,37–85, doi:10.2110/sepmsp.107.05.

Zhang, D., T. Chen, and J. Meng, 2018, Synthetic well logs generation via Recurrent Neural Network: Petroleum Exploration and Development, 45, 629 – 639,

CHAPTER THREE

Deterministic vs. Stochastic Facies Modelling within the Late Devonian Duvernay Formation, Western Canada Sedimentary Basin, Alberta

This chapter is submitted as: Rau, E. G., Yeates, D. W., Breen, K., Thorson, A. M., Atchley, S. C, Deterministic vs. stochastic facies modelling within the Late Devonian Duvernay Formation, Western Canada Sedimentary Basin, Alberta: AAPG Geohorizons

Abstract

Facies interpretation from wireline logs has traditionally been performed through the comparison of core-observed facies distributions and associated log response from which log-based, deterministic algorithms are developed for the prediction of facies in wells lacking core control. In contrast, the unsupervised learning stochastic approach analyzes and automatically clusters recurring well log data associations without calibration to core observations. From petrophysical and core observation data collected from the Late Duvernay Formation of Alberta, Canada, this study investigates whether unsupervised machine learning detects lithologic, and hence, reservoir attributes at a higher resolution than the deterministic approach. The unsupervised machine learning methodology non-negative matrix factorization with *k*-means clustering (NMFk) is applied to petrophysical data to determine and assign groupings independent of core observations. Results from the NMFk model are then compared to the predicted petrofacies from the deterministic approach. Four NMFk groups are identified: three groups coincide with varying shale lithologies and one group with carbonate lithologies. The NMFk model differentiates lithologic associations with reservoir significance that were undetected through deterministic modelling. The three "shale" groups successfully

discriminate the Duvernay into intervals of high, intermediate, and low reservoir quality. Such a differentiation is unrecognized by the deterministic approach.

Introduction

Facies and facies associations commonly account for the variably distributed rock properties within a petroleum reservoir. For unconventional shale reservoirs, understanding the spatial distribution of organic-rich facies and/or facies association is essential in the efficient development of hydrocarbon resources. Although facies distributions are ostensibly identified at high resolution through detailed core description and analysis, continuous core is uncommonly collected, the data collection process is time consuming, and the data collected are oftentimes inconsistent between workers. Well log data, on the other hand, are widely available, and once calibrated to core observations, are oftentimes used to predict facies occurrence in wells lacking core control. Such predictions are based upon deterministic analysis of well log data where the characteristic range of well log response for each facies and/or facies association is identified from univariate statistical analysis, and algorithms are subsequently developed that predict facies and/or facies association occurrence in non-cored wells (e.g., Atchley et al., 2010; Atchley et al., 2018). A similar deterministic approach was successfully applied to the Duvernay Formation within the Western Canada Sedimentary Basin (Thorson, 2019). Based upon a combinations of gamma ray, density and neutron porosity, photoelectric effect, and deep resistivity logs, Thorson et al. (2022) developed algorithms for the prediction of four facies associations with 75% accuracy.

With advances in computational power, stochastic machine learning methods are increasingly used in various geoscience applications including petrophysical facies prediction where the methodology has proven to be efficient, objective and reliable (Bhattacharya et al. 2016; Feng et al., 2020; Hussein et al., 2021; Wu et al., 2021, Rau et al., *in review*). Unsupervised learning is a subset of machine learning that analyzes and clusters unlabeled data. Non-negative matrix factorization with k-means clustering (NMFk) is a novel unsupervised machine learning algorithm that combines two unsupervised machine learning algorithms: 1) non-negative matrix factorization (NMF), and 2) k-mean clustering (Iliev et al., 2018; Vesselinov et al., 2018). In both methods, the optimal number of clusters, *k*, is unknown. However, by combining NMF with k-means clustering, *k* is automatically estimated (Ahmmed et al., 2021). NMFk has been successful in a variety of geoscience applications such as geothermal exploration (Siler et al., 2021), carbon sequestration (Ahmmed et al., 2021), hydrogeology (Alexandrov and Vesselinov, 2014) and contaminate transport (Vesselinov et al., 2018). Siler et al. (2021) used NMFk to identify key geologic factors that control hydrothermal circulation within a shallow geothermal reservoir, and Ahmmed et al. (2021) used NMFk to understand mineral-trapping mechanisms due to carbon injection. Alexandrov and Vesselinov (2014) applied NMFk to identify the source of transient pressure fluctuations in monitoring wells, and Vesselinov et al. (2018) demonstrated the capability of NMFk in the identification of aquifer contaminant sources. Because of the demonstrated versatility and effectiveness of NMFk, this study uses NMFk to classify reservoir-defining petrofacies based on the analysis of a variety of well log data. Although deterministic approaches are successful at identifying reservoir facies, the objective of this study is to investigate if NMFk applied to well log data independent of core observations provides a more objective approach to facies detection and reservoir characterization of the

Duvernay Formation, and at a higher resolution than the recent deterministic approach of Thorson et al. (2022).

Geologic Setting

The Late Devonian (Frasnian) Duvernay Formation accumulated within the Western Canada Sedimentary Basin (WCSB) as a euxinic, organic-rich basinal deposit contemporaneous with marine platform carbonates of the Leduc and Grosmont Formations and is overlain by inorganic basinal mudrock of the Ireton Formation (Fig. 3.1). During the Late Devonian, the WCSB was partitioned into the East and West Shale Basins by a linear reef complex known as the Rimbey-Meadowbrook trend (Fig. 3.2). The WCSB is presently bounded to the east by the Precambrian Canadian Shield and to the west by the Laramide fold and thrust complex (Fig. 3.2). The western portion of the WCSB, including the study area, is an asymmetrical foreland basin originating from cratonic thrust-loading associated with the northwest-trending Laramide fold-and-thrust belt (Porter et al., 1982; Weissenberger and Potma, 2001; Weissenberger et al., 2016). Duvernay thermal maturity trends mimic present-day structural depths within the WCSB foreland basin (Stoakes and Creaney, 1984; Rokosh et al., 2012).

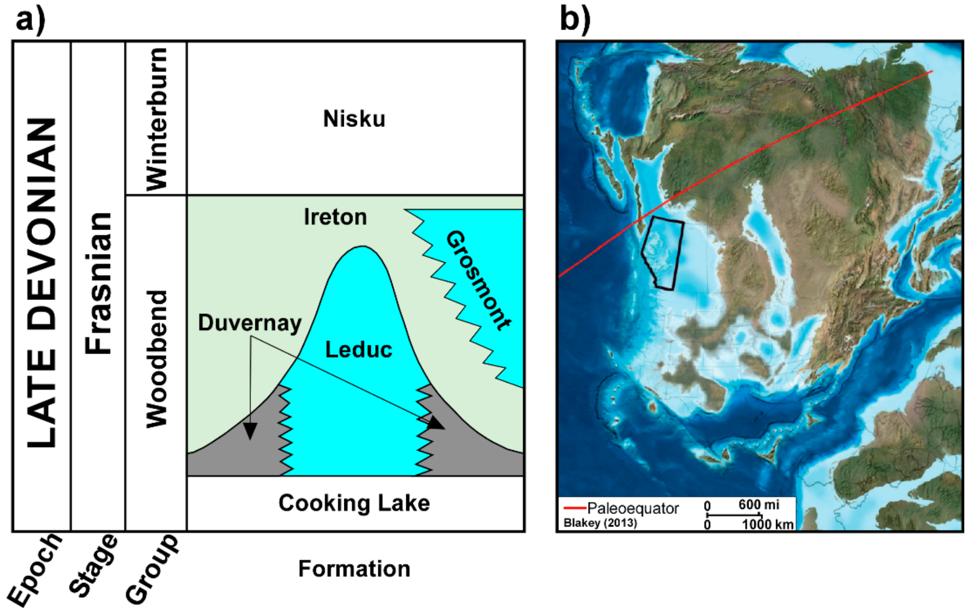

Figure 3.1. a) Late Devonian stratigraphic correlation for south-central Alberta, Canada showing relationships between the contemporaneous Duvernay, Leduc, Ireton and Grosmont Formations (modified from Chow et al., 1995). b) Late Devonian paleogeography of North America (Wong et al., 2016) with the position of present-day Alberta outlined in black and the paleoequator in red (modified from Blakey, 2013).

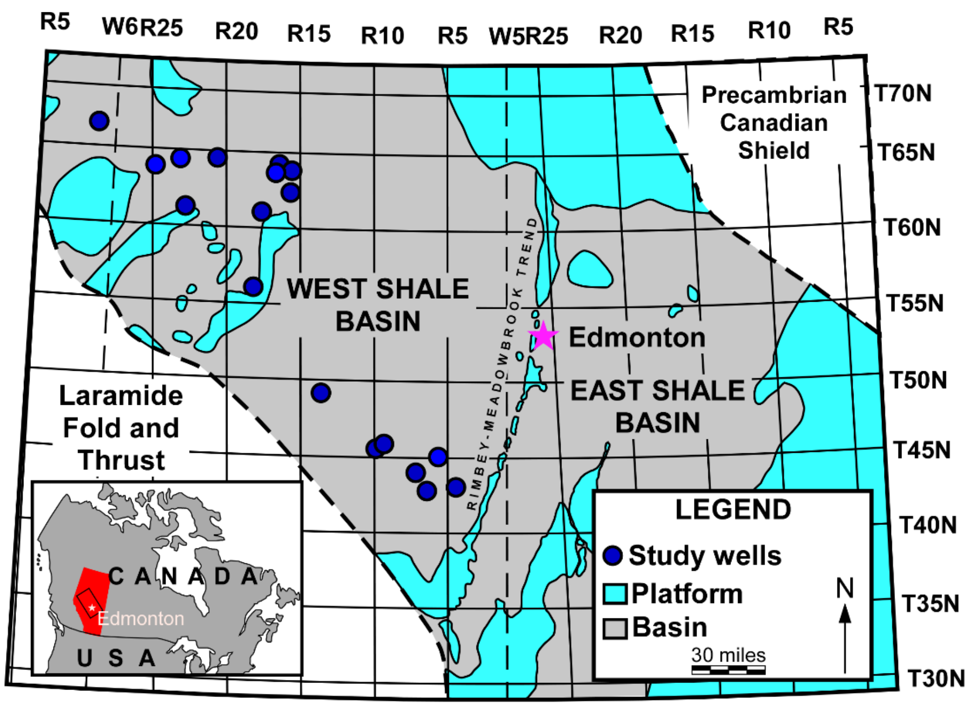

Figure 3.2. Late Devonian paleogeography of Alberta, Canada (modified from Switzer et al., 1994, Rokosh et al., 2012, Preston et al., 2016, Wang et al., 2016). The Duvernay was deposited as basinal fill within both the East and West Shale basins that are partitioned by the Rimbey-Meadowbrook trend. Blue circles identify well locations used in this study.

The Duvernay Formation is a major source rock for the conventional reservoirs that accounted for prolific hydrocarbon production within the WCSB during the 20th century. Since 2011, the Duvernay has been targeted as an unconventional reservoir and extensively studied for its reservoir potential (Dunn et al., 2016; Preston et al., 2016; Chopra et al., 2017; Wong et al., 2016; Bauman, 2018; Datta, 2018; Harris et al., 2018; Dong et al., 2019; Knapp et al., 2019; Thorson, 2022; Li et al., 2020). Duvernay facies have been recently classified, correlated, and mapped within the West and East Shale Basins based upon the evaluation and integration of whole core and wireline log data (Thorson, 2019). The facies and associated reservoir quality framework of Thorson (2019) is used in the analyses and results of this study.

Methods

Petrophysical Data

A total of 18 wells in the West Shale Basin with continuous core through the Duvernay Formation are used in this study (Fig. 3.2). Each well includes the following modern (post-1980 vintage) digital wireline log types: gamma ray, deep resistivity, bulk density, neutron-density porosity (calibrated to limestone matrix), and photoelectric effect. In total, the dataset includes 9,858 wireline log samples. The vertical resolution for each well log type is as follows: gamma ray (0.6 m), deep resistivity (2.1 m), bulk density (0.5 m), neutron-density (0.6 m), and photoelectric effect (0.05 m) (Alberty, 1992). A theoretical and operational summary of each log is provided within Alberty (1992).

Gamma ray. The gamma ray log (GR) measures natural gamma radiation derived from uranium, potassium, and thorium that are commonly associated with clay minerals. GR, therefore, is useful in the differentiation of shale (clay-rich) and non-shale sedimentary lithologies. Uranium is oftentimes highly concentrated in organic-rich shales and results in uncommonly high gamma ray activity, as is the case with the Duvernay Formation (Thorson, 2019).

Deep resistivity. The measurement of deep resistivity is influenced by rock matrix composition, porosity and associated pore fluid type and saturation, and the occurrence of organic matter (Passey et al., 1990; Alberty, 1992). Within thermally mature, organic-rich mudrock successions, as commonly occur within the Duvernay Formation, resistivity is particularly high due to the occurrence of both kerogen and associated pore-filling hydrocarbons.

Bulk density and compensated neutron density and porosity. Bulk density and derivative density porosity are influenced by both the composition of rock matrix material and the associated pore volume and fluid fill. Because organic matter has a much lower density than matrix minerals, the presence of organic matter causes a significant reduction in measured bulk density and a corresponding increase in calculated porosity (Meyer and Nederlof, 1982; Passey et al., 2010). The neutron porosity log measures the occurrence of hydrogen that may occur in association with pore-filling water or hydrocarbon, organic matter, or clay-bound *hydroxyl* (OH-) ions. Estimates of neutron porosity use the occurrence of hydrogen as a proxy for porosity, and as such, mimic density porosity estimates (Passey et al., 2010). Total organic carbon (TOC) rich intervals in the Duvernay correlate with higher neutron and density porosity.

Photoelectric effect. The photoelectric effect (PE) is recorded as a compliment to bulk density measurements and uses the bulk density radioactive source and detector. The PE measures a rock interval's ability to absorb induced gamma rays, which corresponds closely with mineralogy and related lithology (Alberty, 1992). The Duvernay is influenced by the surrounding carbonates of the Leduc and Grosmont Formations and results in variably calcareous facies and variably high PE values.

Unsupervised Machine Learning

In this study, NMFk is used to automatically detect the number of petrophysical groups identifiable in the petrophysical data. NMFk combines two unsupervised machine learning methods, non-negative matrix factorization (NMF) and k-means clustering (Alexandrov et al., 2014; Iliev et al., 2018). Specifically, the petrophysical groups are identified by analyzing the reconstruction error from NMF and the silhouette width from k-means clustering. Model input data X of size (n, m) is the petrophysical data where n is the borehole measured depth and m is the number of input features which in this study are the well logs.

NMF projects input data into a lower dimension to extract meaningful features (Ropes and Ribeiro, 2015). The NMF portion of NMFk decomposes the X matrix into matrices W of size (n,k) and H of size (k,m) as:

$$X = W \times H + \varepsilon(k)$$

where W is the depth matrix, H is the features matrix and $\varepsilon(k)$ is the error (Iliev et al., 2018; Ahmmed et al., 2021). The W matrix is commonly called the mixing matrix which in this study represents how each depth is measuring a mixture of petrophysical groups established by NMF. The H matrix depicts the relationship between the petrophysical groups and each well log. The loss function ($\mathcal{L}$) is minimized using Frobenius norm for a specified k resulting in non-negative W and H matrices. In NMF the value for k is unknown, therefore, NMF is performed for $k = 2,3, \ldots, d$ where d is the maximum number of well logs. The optimal number of signals cannot be more than the number of input features. For example, in this study the number of well logs used for input features is 6 which is the maximum possible number of petrophysical groups. For each k, NMF solves for 1000 random solutions of W and H matrices. The lowest value of $\mathcal{L}$ for a given k is considered the best reconstruction error.

Results from NMF are coupled with k-means clustering to identify the optimal number of petrofacies. The k-means clustering algorithm divides the data into a specific number of groups, k, such that the sum of squared distances between the data and center of each cluster is minimized (Hartigan and Wong, 1979). Similar to NMF, the value for k must be specified to execute the k-means clustering algorithm so k-means clustering is iteratively performed using a specified range of k values. For each iteration of k, similarities between the clusters are evaluated using the silhouette width (Rousseeuw et al., 1987). The silhouette width quantifies how well a point fits within its assigned cluster compared to neighboring clusters. The value ranges from -1 to 1 with higher positive values indicating that the point is very well clustered. Silhouette width typically declines after the optimal number of *k* is reached. The optimal number of petrophysical groups has low reconstruction error from NMF and higher silhouette values in *k*-means clustering.

Core Description Data

All cored wells (750 m [2,460 ft] of total core length) were described in detail to account for the occurrence of facies, texture, allochem type and abundance, mechanical and biological structures, and fracture density. Core descriptions were digitized and merged with their corresponding petrophysical data. Each digitized core data value was depth-shifted to coincide with well log measured depths through comparison with a core-derived synthetic gamma ray log composed of gamma ray total counts derived from uranium, thorium, and potassium concentrations (*sensu* Crain, 2018) measured within core at 1-meter resolution using a Bruker hand-held X-ray fluorescence (XRF). The core synthetic gamma ray was plotted and depth-shifted to coincide with graphical trends observed on the well log gamma ray. The resulting depth shift was applied to all digitized core description data.

Based on the recurrence of sedimentologic features, 10 facies are identified and grouped into 3 facies associations (Thorson, 2019): open basin, transitional basin, and restricted basin. Of the 10 facies, 6 are observed within the study wells (Fig. 3.3, Table 3.1). The open basin consists of the burrowed nodular (BN) and laminated nodular (LN) facies, the transitional basin of the burrowed mudstone (BM) facies, and the restricted basin of the black laminated mudstone (BLM), black mechanically laminated mudstone (BMLM), and the black burrowed mudstone (BBM). The restricted basin association has the highest total organic carbon (TOC) estimates, and accordingly, most favorable hydrocarbon production potential (Thorson, 2019). Three non-Duvernay facies are also present in the 18 wells used in this study. The fine skeletal packstone (FSP) and diverse skeletal packstone (DSP) from the contemporaneous Leduc and Grosmont Formations are present, but in small quantities. The black laminated mudstone of the Ireton Formation (BLMi) overlies the Duvernay Formation.

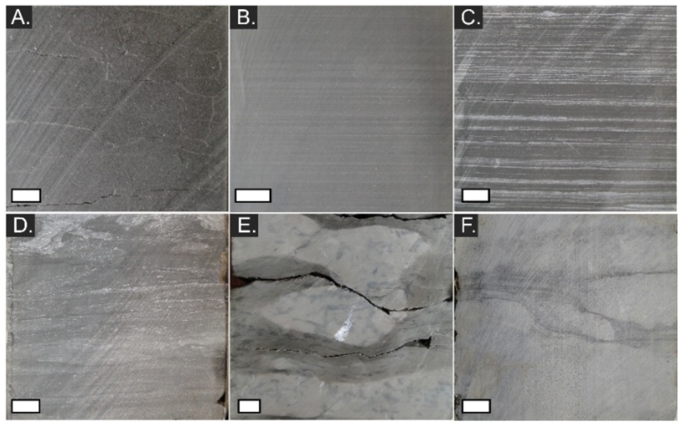

Figure 3.3. Representative core photo of (A) black burrowed mudstone (BBM), (B) black laminated mudstone (BLM), (C) black mechanically laminated mudstone (BMLM), (D) burrowed mudstone (BM), (E) laminated nodular (LN), and (F) burrowed nodular (BN). Scale bar is 1 cm.

Deterministic Facies Association Prediction

Duvernay petrofacies are based on the aforementioned facies associations and are predicted in non-cored wells based on their characteristic well log responses (Thorson, 2019). Using univariate statistical characterization of well log response from 42 cored Duvernay wells located within the West and East Shale Basins, Thorson et al. (2022) establishes well log cutoffs that are diagnostic of the open basin, restricted basin, Ireton Shale (BLMi) and undifferentiated Leduc/Grosmont “platform to slope” carbonates (Table 3.2). These cutoffs are used to predict the occurrence of petrofacies (identified in Table 3.2) in all 18 wells included in this study, although as prescribed by Thorson (2022), predictions are primarily based on gamma ray and deep resistivity responses and secondarily by the other log responses.

Table 3.1. Duvernay Formation facies summary table for the study area (modified from Thorson, 2019).

Facies Association	**Open Basin**		**Restricted Basin**			
Facies Name	Burrowed Nodular (BN)	Laminated Nodular (LN)	Burrowed Mudstone (BM)	Black Burrowed Mudstone (BBM)	Black Mechanically Laminated (BMLM)	Black Laminated Mudstone (BLM)
Environment	peri-platformal	basinal	basinal	basinal	basinal (restricted)	basinal (restricted)
Texture	mudstone (wackestone)	mudstone	mudstone (local packstone)	mudstone	mudstone	mudstone
Grains	brachiopods, crinoids, SK, intraclasts, peloids	brachiopods, crinoids, intraclasts, SK, peloids, gastropods	brachs., crinoids, SK, few mudstone-textured intraclasts and lithoclasts, peloids	few: SK, brachiopods, crinoids	few: *Amphipora*, crinoids, brachiopods, SK	few: brachiopods, crinoids, SK (<0.5mm), intraclasts
Sedimentary Features	TH, PL, AST, GLOSSI, hardground /firmground, irregular mudstone nodules	elongate horiz. mudstone nodules, cm-lamina, (few) mm-lamina between mudstone nodules & burrows, CH, TH, PL	TH, PL, GLOSSI, CH, mm-lamina, few cm-lamina, firmground/ hardground	massive/dark mm-lamina, dark color mottling, TH, PL, occasional BLM/BMLM interbeds, imbricate sed. gravity flows	mm to cm-lamina: alternating silt-sized carbonate grains and mud, TH, PL, Z, calcite concretions	mm-lamina, few TH, GLOSSI, PL, event beds (silt-sized CO_3 grains), firmg./hardg. West Basin: dark grey to black, poker-chip recovery
Average ichnofabric index (0-6)	6	5	6	3	1	1
Average hardness (R values)	37	32	33	20	20	19
Average fracture density (frac./m or /ft)	5	13	4	4	1	3
Representative core photo	Figure 3.3F	Figure 3.3E	Figure 3.3D	Figure 3.3A	Figure 3.3C	Figure 3.3B

Table 3.2. Well log cutoffs established by Thorson (2019) for the prediction of Duvernay Formation petrofacies.

Facies Association	**Gamma Ray (API)**	**Neutron Porosity (fraction)**	**Density Porosity (fraction)**	**PE (barns/electron)**	**Deep Resistivity (ohm-m)**	**Comments**
Ireton	75-100	0.0-0.2	0.0-0.2	3.8-4.0	0.2-20	Resistivity values take precedence in distinguishing from RB FA.
OB/TB FA: BN, LN, BM	20-85	0.0-0.5	0.0-0.3	4.2-5.1	200-1000	Gamma-ray and PE values take precedence in distinguishing from RB FA.
RB FA: BBM, BMLM, BLM	75-100	0.1-0.4	0.1-0.3	3.7-4.7	700-1000	Higher gamma-ray, resistivity, and porosity values than OB/TB FA.
Platform to Slope Carbonates	0-15	0.0-0.3	0.0-0.2	4.7-5.0	10-100	Lowest gamma-ray values of all petrofacies

Results

From NMFk stochastic analysis, the optimal number of recurring petrophysical groups is 4 because the silhouette width decreases after 4 and the reconstruction error is relatively low (<0.25) (Fig. 3.4). The model assigns one of the 4 groups to each of the 9,858 wireline log samples. The interquartile range for log responses for each group are listed in Table 3.3, and box and whisker plots for each group versus well log responses are provided in Figure 3.5. TOC weight percent is calculated using the Schmoker and Hester (1983) equation based on bulk density values from wireline logs. Group A is characterized by the highest gamma ray activity and TOC, and relatively high deep resistivity values. These attributes indicate organically enriched shale, and the associated low bulk density and PE values suggest a high proportion of matrix quartz. Group B is characterized by relatively high gamma ray values and low deep resistivity values indicative of inorganic shale, and the comparatively elevated PE values suggest a higher proportion of matrix carbonate than associated with either Group A or D. Group C has low gamma ray and relatively high bulk density and PE values that suggest a high proportion of matrix carbonate. Group D is characterized by the highest gamma ray activity, high deep resistivity, low bulk density and associated high density and neutron porosity that collectively indicate organic-rich mudrock. The associated low bulk density and PE values indicate a comparatively high proportion of quartz-rich matrix.

Table 3.3 The interquartile range for log response of each NMFk group and their associated well log interpretation.

Group	Gamma Ray (API)	Deep Resistivity (ohm-m)	Neutron Porosity (fraction)	Density Porosity (fraction)	Bulk Density (gm/cc)	PE (barns/electron)	Calculated TOC (wt%)	Interpretation
A	107-161	45-254	0.14-0.18	0.09-0.12	2.51-2.56	3.3-3.9	3.0-4.2	Quartz-rich, hydrocarbon bearing shale, good reservoir quality
B	99-145	8-68	0.14-0.19	0.03-0.07	2.60-2.67	4.0-4.5	0.6-2.1	Inorganic shale, poor reservoir quality
C	36-60	22-94	0.04-0.09	0.0-0.01	2.69-2.71	4.8-5.2	0.0-0.1	Allochthonous carbonates
D	97-148	597-911	0.12-0.16	0.08-0.11	2.53-2.57	3.5-3.9	2.8-3.8	Quartz-rich, hydrocarbon bearing shale, excellent reservoir quality

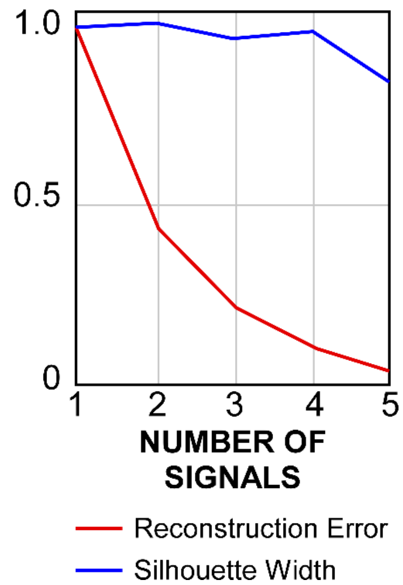

Figure 3.4. NMFk reconstruction error (red line) and silhouette width (blue line) for different numbers of clusters *k.* The optimal *k* value has low reconstruction error and higher silhouette values. The optimal number of groups is 4 because silhouette width decreases after 4 and reconstruction error is relatively low (<0.25).

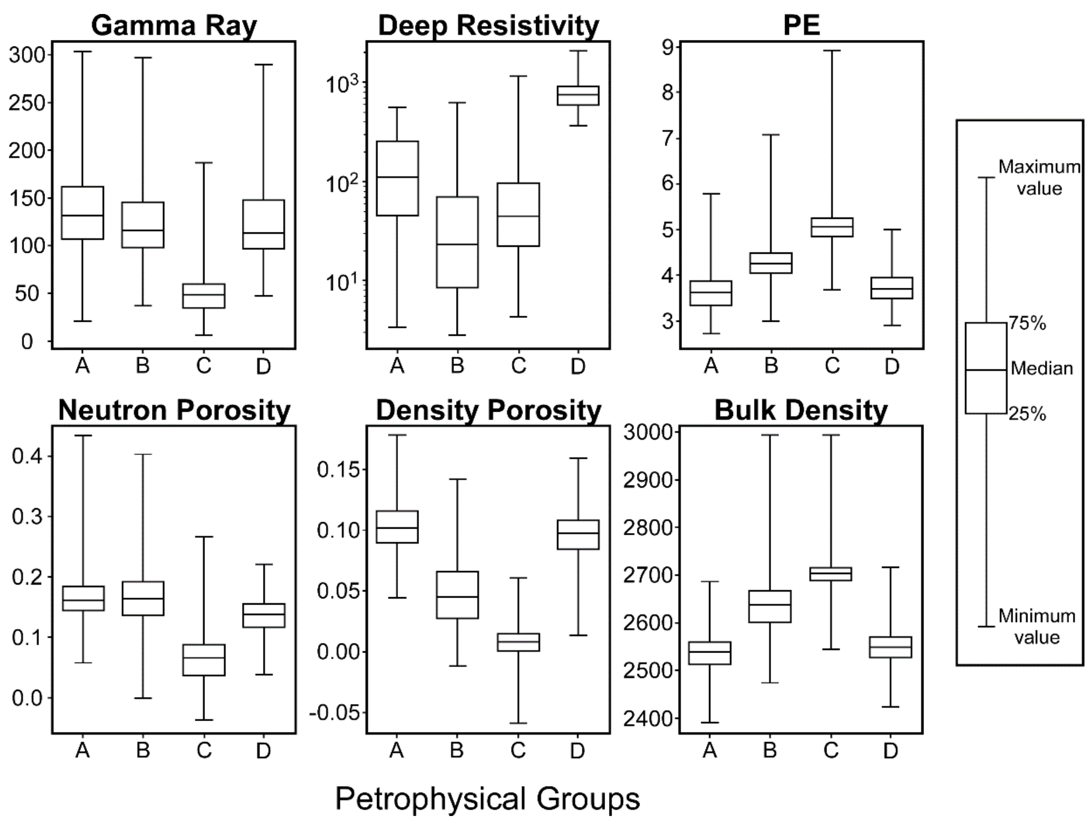

Figure 3.5. Box and whisker plot of well log type versus NMFk groups. Gamma ray and PE responses indicate Group A, B, and D are clay dominated lithologies and C is carbonate-dominated lithologies.

Sedimentologic Characterization of NMFk Groups

A comparison of the four NMFk groups with the core-observed depositional facies of Thorson (2019) indicates that the stochastic NMFk petrophysical groupings correspond with a combination of core-observed facies (Fig. 3.6). The BLM, BMLM, and BBM depositional facies all occur within Groups A, B, and D and indicates that NMFk identifies the restricted basin facies association, but it does not differentiate the specific restricted basin facies. About 93% of core-observed restricted basin facies (i.e., facies association) are assigned as either Group A, B, or D. Group A is characterized by all three restricted basin facies. BLM, BMLM, and BBM represent 51%, 29% and 17% of Group A, respectively. Group B includes BLM (52%) or BBM (28%), and Group D is dominated by BLM (35%) and BMLM (63%). In contrast, 89% of core-observed open basin facies associations coincide with Group C. LN and BN account for 66% and 8% of Group C (Fig. 3.6). The close correspondence of BLM, BMLM, and BBM with Groups A, B, and C and LN and BN with Group C indicates that objective NMFk model results coincide with depositionally-controlled lithologic attributes, which in this case is the open basin (non-reservoir) and restricted basin (reservoir) facies associations of the Duvernay.

Comparison of Deterministic and NMFk Approaches

Deterministic and NMFk petrofacies predictions were made on all 18 wells, and Figure 3.7 associates the well log response with core-observed depositional facies, petrofacies identified from the deterministic approach, and unsupervised groups assigned by the stochastic NMFk approach for two of the wells. For the deterministic model, the occurrence of "Ireton facies", "open basin and transitional basin facies association" (combined), and "restricted basin facies association" closely corresponds with the occurrence of core observed facies associations which are predicted with 88% accuracy (Fig. 3.7). The distribution of NMFk groups A, B, and C mimic the distribution of both core-observed facies associations and petrofacies predicted by the deterministic model (Fig. 3.7). Of the predicted restricted basin facies association from the deterministic model, 93% coincides with either Group A, B, and D. Groups A, B and D, however, occur as relatively thin intercalations within the comparatively homogenous, blocky occurrence of the deterministic restricted basin facies association (Fig. 3.7). Well 00/01-35-045-10W5 has two predicted restricted basin intervals that extend from 3290 m (10,794 ft) to 3301 m (10,830 ft) and from 3306 m (10,846 ft) to 3330 m (10,925 ft), and well 00/13-32-063-16W5 has one predicted restricted basin interval extending from 2836 m (9304 ft) to 2864 m (9396 ft). In all instances, deterministically-derived restricted basin facies association intervals are partitioned into thin alternations of NMFk groups A, B and D (Fig. 3.7).

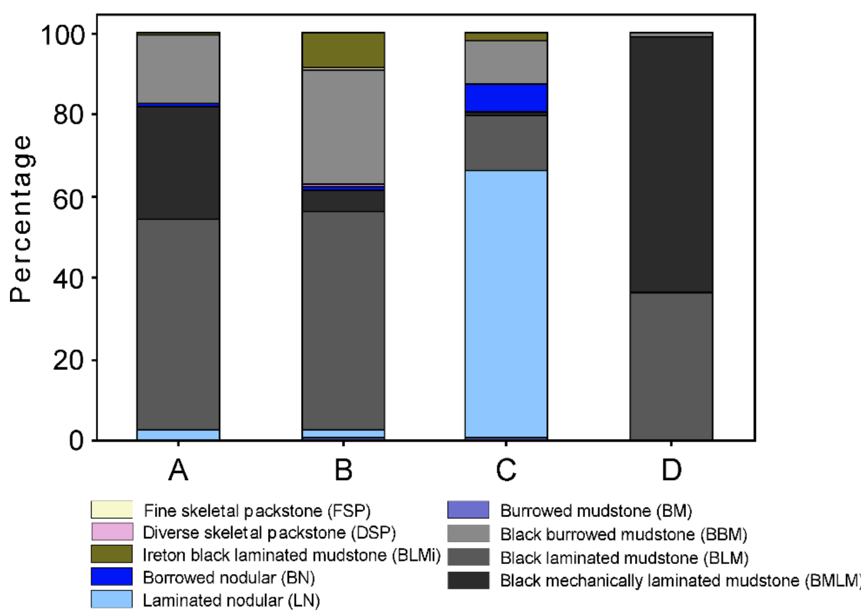

Figure 3.6. The proportion of core-observed depositional facies (all 18 wells) that occur within each NMFk group.

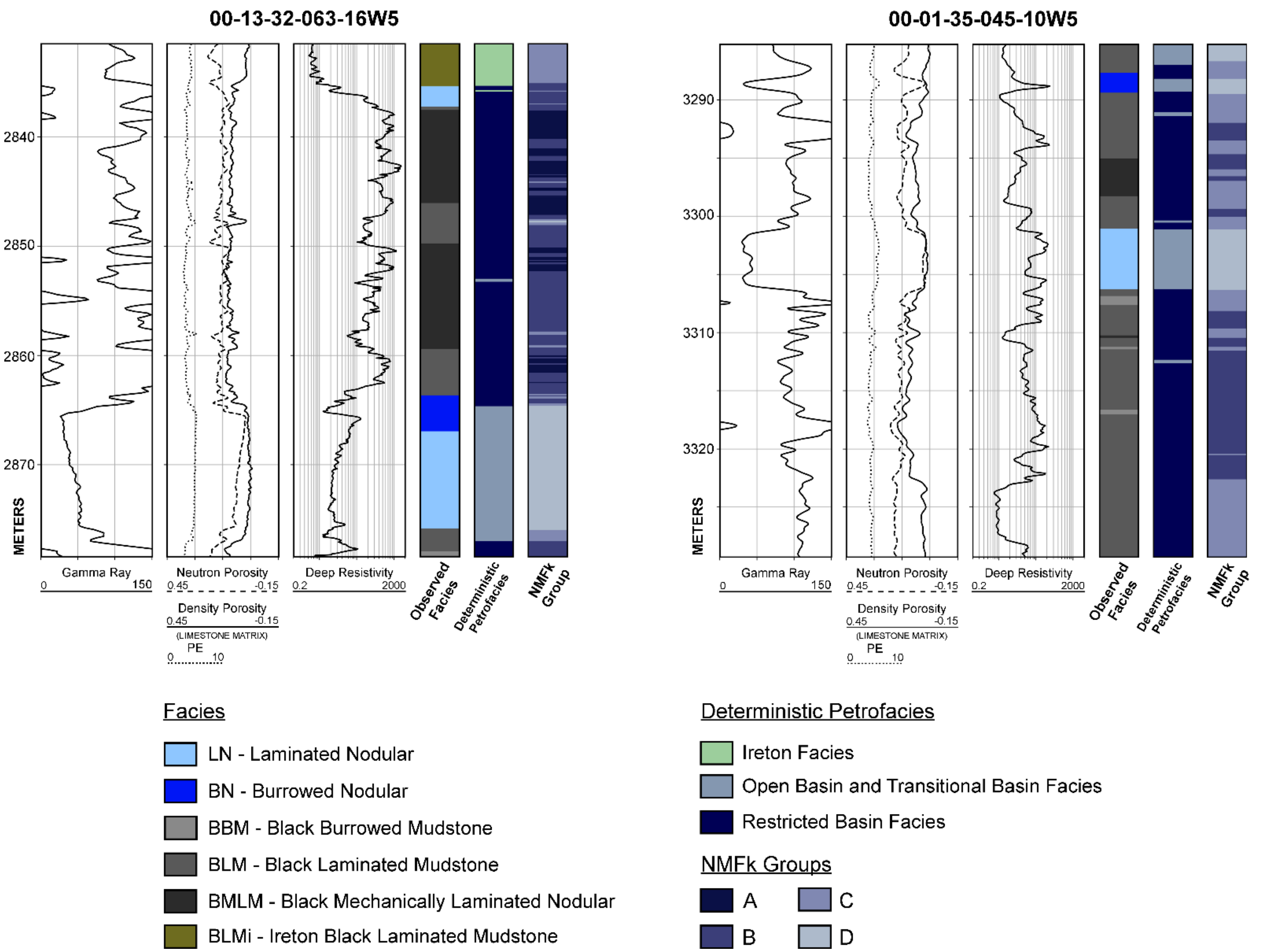

Figure 3.7. Well log responses for wells 00/13-32-063-16W5 and 00/01-35-045-10W5 annotated with the distribution of core-observed facies and facies associations, deterministic model predicted petrofacies, and NMFk groups

Discussion

The highly interstratified distribution of NMFk groups (i.e., D, highest reservoir potential, A, intermediate reservoir potential, and B, lowest reservoir potential) within the otherwise homogenous deterministic “restricted basin” intervals of the Figure 3.7 wells suggests that NMFk analysis detects meaningful variability in rock properties at a higher resolution than the deterministic approach. Fluid type, mineralogy, and TOC are factors affecting shale reservoirs that are readily detected by wireline logs, and therefore influence stochastic group assignments by NMFk analysis (Guo et al., 2017; Hou et al., 2021, Luo et al., 2021). Group D, for example, is characterized by high resistivity values induced by rock attributes such as the occurrence of organic matter and associated porosity and pore-filling hydrocarbon fluids. Although all restricted basin depositional facies within the restricted basin facies association are sedimentologically similar, subtle variability in both organic matter richness and fluid type cause variability in measured resistivity that is within the range deterministically-defined as the restricted basin association. Group D strata, which include the restricted basin depositional facies association, are categorized by NMFk on the basis of the petrophysical response to all rock and fluid attributes. Conversely, depositional facies are classified on the basis of rock attributes alone. Since variations in fluid type and abundance were not explicitly used in the classification of depositional facies, but do influence log response, facies and associated deterministic petrofacies assignments accordingly do not reliably account for fluid-induced variability in resistivity measurements. NMFk group assignments, however, do. As such, NMFk analysis discriminates reservoir defining attributes encoded within the petrophysical response that are difficult to recognize in core without detailed sampling for geochemical and/or petrographic analysis.

NMFk analysis results in a more detailed accounting of lithologic variability and provides more precise and objective recognition of varying reservoir attributes. This has implications on the mapping of reservoir attributes and assignment of rock volume and associated reserve estimates at the exploration and development scale. When used during production well planning, NMFk results may more accurately depict the location of the most prospective reservoir zones, and therefore optimize the placement of lateral boreholes. Furthermore, Results from NMFk analysis may also be used to anticipate and document lithologic variability during core description and sampling, thereby minimizing the inconsistency and error that is inherent to subjective human core description, particularly when core description involves multiple individuals within a project.

Conclusion

1. Four recurring petrophysical groups are identified through stochastic NMFk analysis of gamma ray, deep resistivity, bulk density, neutron-density porosity, and photoelectric effect log responses collected from the Late Devonian Duvernay Formation within the Western Canada Sedimentary Basin. Groups A, B, and D coincide with the restricted basin depositional facies association, but Groups A, B, and D are characterized by intermediate, low and high reservoir quality respectively. Group C coincides with the open basin (non-reservoir) depositional facies association.
2. NMFk resolves detail within the Duvernay restricted basin facies association that is not detected from deterministic petrofacies modeling. The petrophysical groups

established through NMFk discriminate reservoir defining attributes detected in the petrophysical response that are difficult to recognize in core or require further core analyses.

3. NFMk analysis provides objective petrophysical groups, thus minimizing human-based subjectivity during core description and analysis. This provides more precise, consistent, and objective petrophysical interpretation and reservoir characterization, and increases the consistency and accuracy of resource assessment and borehole placement and completion.
4. In shale reservoirs, the application of NMFk petrophysical analysis may assist in highlighting intervals of interest in advance of core description. This enhances the quality and relevance of the core description, and reduces observer inconsistency and bias, particularly when multiple individuals are participating in the description and analysis of core.

Acknowledgements

This study was completed in association with the Baylor University Applied Petroleum Studies Program. Pioneer Natural Resources, Anadarko Petroleum, and Angel Resources provided sponsorship. Petrophysical data was acquired using IHS Markit AccuMap and geoLOGIC Geoscout software and GLJ Petroleum Consultants. Data analytics were completed using the programming language Julia and open-source code from SmartTensors LLC.

References

Alberty, M.W., 1992, Basic open hole tools, in Development Geology Reference Manual, D. Morton-Thompson and A.M. Woods, eds., AAPG Exploration in Methods Series, v. 10, p. 144-149.

Ahmmed, B., S. Karra, V. V. Vesselinov, and M. K. Mudunuru, 2021, Machine learning to discover mineral trapping signatures due to CO_2 injection: International Journal of Greenhouse Gas Control, v. 109, p. 103382.

Alexandrov, B. S., and V. V. Vesselinov, 2014, Blind source separation for groundwater pressure analysis based on nonnegative matrix factorization: Water Resources Research, v. 50, no. 9, p. 7332–7347, 10.1002/2013WR015037.

Atchley, S. C., N. H. Ball, and L. E. Hunt, 2010, Reservoir characterization and facies prediction within the Late Cretaceous Doe Creek Member, Valhalla Field, west-central Alberta, Canada. American Association of Petroleum Geologists Bulletin, v. 94, no. 1, p. 1-25, doi:10.1306/10261817272.

Atchley, S. C., S. C. James, H. Harlow, C. Leslie, J. Jiang, G. Carrelli, W. Bindon, and J. Beal, 2018, Geologic and resource assessment of the Upper Devonian Grosmont and Upper Ireton Formations, central Grosmont shelf complex, Alberta, Canada: AAPG Bulletin, v. 102, no. 5, p. 731–759, 10.1306/0711171612816514.

Bauman, P., 2018, East Duvernay Shale Basin, https://chinookpetroleum.com/devonian-duvernay-formation-east-shale-basin-alberta/, accessed January 19, 2021.

Bhattacharya S., T. R. Carr, and M. Pal, 2016, Comparison of supervised and unsupervised approaches for mudstone lithofacies classification: Case studies from the Bakken and Mahantango-Marcellus Shale, USA: Journal of National Gas Science and Engineering, v. 33, p. 1119-1133, 10.1016/j.jngse.2016.04.055.

Chopra, S., R. K. Sharma, A. K. Ray, H. Nemati, R. Morin, B. Schulte, and D. D’Amico, 2017, Seismic reservoir characterization of Duvernay shale with quantitative interpretation and induced seismicity considerations — A case study: Interpretation, v. 5, no. 2, p. 185–197, 10.1190/INT-2016-0130.1.

Crain, E.R., 2018, Natural gamma ray and spectral gamma ray logs, https://www.spec2000.net/07-grlog.htm, accessed July 12, 2018.

Datta, A., 2018, The Resurgence of Interest in the Duvernay: A Review of Pertinent Economic and Environmental Issues: Canadian Energy Research Institute.

Dong, T., N. B. Harris, J. M. McMillan, C. E. Twemlow, B. R. Nassichuk, and D. L. Bish, 2019, A model for porosity evolution in shale reservoirs: an example from the Upper Devonian Duvernay Formation, Western Canada Sedimentary Basin: AAPG Bulletin, v. 103, no. 5, p. 1017–1044, 10.1306/10261817272.

Dunn, L., G. Schmidt, K. Hammermaster, M. Brown, R. Bernard, E. Wen, R. Befus, and S. Gardiner, 2016, The Duvernay Formation (Devonian): Sedimentology and Reservoir Characterization of a Shale Gas/Liquids Play in Alberta, Canada: Annual Convention and Exhibition, AAPG, Extended Abstract.

Feng, R., N. Balling, and D. Grana, 2020, Lithofacies classification of a geothermal reservoir in Denmark and its facies-dependent porosity estimation from seismic inversion: Geothermics, v. 87, p. 101854, 10.1016/j.geothermics.2020.101854.

Guo, X., D. Hu, Y. Li, Z. Wei, X. Wei, and Z. Liu, 2017, Geological factors controlling shale gas enrichment and high production in Fuling shale gas field: Petroleum Exploration and Development, v. 44, no. 4, p. 513–523, 10.1016/S1876-3804(17)30060-5.

Harris, N. B., J. M. McMillan, L. J. Knapp, and M. Mastalerz, 2018, Organic matter accumulation in the Upper Devonian Duvernay Formation, Western Canada Sedimentary Basin, from sequence stratigraphic analysis and geochemical proxies: Sedimentary Geology, 376, p. 185–203, 10.1016/j.sedgeo.2018.09.004.

Hou, L., Z. Yu, X. Luo, S. Lin, Z. Zhao, Z. Yang, S. Wu, J. Cui, and L. Zhang, 2021, Key geological factors controlling the estimated ultimate recovery of shale oil and gas: A case study of the Eagle Ford Shale, Gulf Coast Basin, USA: Petroleum Exploration and Development, v. 48, no. 3, p. 762–774, 10.1016/S1876-3804(21)60062-9.

Hussein, M., R. R. Stewart, D. Sacrey, J. Wu, and R. Athale, 2021, Unsupervised machine learning using 3D seismic data applied to reservoir evaluation and rock type identification: Interpretation, v. 9, no. 2, p. T549–T568, 10.1190/INT-2020-0108.1.

Iliev, F. L., V. G. Stanev, V. V. Vesselinov, and B. S. Alexandrov, 2018, Nonnegative matrix factorization for identification of unknown number of sources emitting delayed signals: PLOS ONE, v. 13, no. 3, p. e0193974, 10.1371/journal.pone.0193974.

Knapp, L. J., N. B. Harris, and J. M. McMillan, 2019, A sequence stratigraphic model for the oranic-rich Upper Devonian Duvernay Formation, Alberta, Canada: Sedimentary Geology, 387, p. 152–181, 10.1016/j.sedgeo.2019.04.008.

Lee, D. D., and H. S. Seung, 1999, Learning the parts of objects by non-negative matrix factorization: Nature, 401, no. 6755, p. 788–791, 10.1038/44565.

Li, G., K. Luo, and D. Shi, 2020, Key technologies, engineering management and important suggestions of shale oil/gas development: Case study of a Duvernay shale project in Western Canada Sedimentary Basin: Petroleum Exploration and Development, 47, p. 791–802,10.1016/S1876-3804(20)60094-5.

Luo, C., H. Lin, Y. Peng, H. Qu, X. Huang, N. Yin, W. Liu, and X. Gao, 2021, Factors controlling shale reservoirs and development potential evaluation: a case study: Geofluids, v. 2021, p. 1–13, 10.1155/2021/6661119.

Meyer, B.L., and M.H. Nederlof, 1984, Identification of source rocks on wireline logs by density/resistivity and sonic transit time/resistivity crossplots: AAPG Bulletin v. 68, p. 121-129.

Preston, A., G. Garner, K. Beavis, O. Sadiq, and S. Stricker, 2016, Duvernay Reserves and Resources Report: A Comprehensive Analysis of Alberta's Foremost Liquids-Rich Shale Resource: Alberta Energy Regulator, Calgary, Alberta.

Rokosh, C.D., S. Lyster, S.D.A. Anderson, A.P. Beaton, H. Berhane, T. Brazzoni, D. Chen, Y. Cheng, T. Mack, C. Pana, and J.G. Pawlowicz, 2012, Summary of Alberta's shale- and siltstone-hosted hydrocarbon resource potential: Energy Resources Conservation Board, ERCB/AGS Open File Report 2012-06, 1-327.

Rousseeuw, P. J., 1987, Silhouettes: A graphical aid to the interpretation and validation of cluster analysis: Journal of Computational and Applied Mathematics, 20, p. 53–65, 10.1016/0377-0427(87)90125-7.

Siler, D. L., J. D. Pepin, V. V. Vesselinov, M. K. Mudunuru, and B. Ahmmed, 2021, Machine learning to identify geologic factors associated with production in geothermal fields: a case-study using 3D geologic data, Brady geothermal field, Nevada: Geothermal Energy, v. 9, no. 1, p. 1-17, 10.1186/s40517-021-00199-8.

Switzer, S.B., W.G. Holland, G.C. Graf, A.S. Hedinger, R.J. McAuley, R.A. Wierzbicki, and J.J. Packard, 1994, Devonian Woodbend-Winterburn strata of the Western Canada Sedimentary Basin, *in* G. Mossop and I. Shetson, eds., Geological Atlas of the Western Canada Sedimentary Basin: Calgary, Alberta, Canada, Canadian Society of Petroleum Geologists and Alberta Research Council, 165-202.

Thorson, A. M., 2019, Stratigraphic Partitioning and Distribution of Reservoir Attributes within the Late Devonian Duvernay Formation, Western Canada Sedimentary Basin: M.S. thesis, Baylor University.

Vesselinov, V. V., B. S. Alexandrov, and D. O'Malley, 2018, Contaminant source identification using semi-supervised machine learning: Journal of Contaminant Hydrology, v. 212, p. 134–142, 10.1016/j.jconhyd.2017.11.002.

Vesselinov, V. V., M. K. Mudunuru, S. Karra, D. O'Malley, and B. S. Alexandrov, 2019, Unsupervised machine learning based on non-negative tensor factorization for analyzing reactive-mixing: Journal of Computational Physics, 395, p. 85–104, 10.1016/j.jcp.2019.05.039.

Wendte, J.C. and F. A. Stoakes, 1982, Evolution and corresponding porosity distribution of the Judy Creek reef complex, Upper Devonian, central Alberta, *in* W.G. Cutlr, ed., Canada's Giant Hydrocarbon Reservoirs, Core Conference Manual Canadian Society of Petroleum Geologists, p. 63–81.

Wong, P. K., J. A. W. Weissenberger, and M. G. Gilhooly, 2016, Revised Regional Frasnian Sequence Stratigraphic Framework, Alberta Outcrop and Subsurface: New advances in Devonian carbonates: Outcrop analogs, reservoirs, and chronostratigraphy, Society for Sedimentary Geology Special Publication 107, p. 37–85, doi:10.2110/sepmsp.107.05.

Wu, G., G. Chen, Q. Cheng, Z. Zhang, and J. Yang, 2021, Unsupervised machine learning for lithological mapping using geochemical data in covered areas of Jining, China: Natural Resources Research, v. 30, no. 2, p. 1053–1068, 10.1007/s11053-020-09788-z.

CHAPTER FOUR

Geothermal play assessment in the Tularosa Basin, south-central New Mexico using integrated unsupervised machine learning and geophysical methods

Abstract

With increasing global energy demand, there is a need to explore and develop alternative forms of energy. As a nonintermittent resource, geothermal energy has the potential to help satisfy growing energy demands while simultaneously mitigating the emission of greenhouse gases produced from the burning of fossil fuels. Using the unsupervised machine learning technique, non-negative matrix factorization with *k*-means clustering (NMFk) and the passive geophysical method, magnetotellurics (MT), this study identifies prospective geothermal targets within the Tularosa Basin of south-central New Mexico based on heat flow, temperature, porosity, and permeability. Two locations in the Tularosa Basin are assessed for their geothermal potential, White Sands Missile Range and McGregor Range at Fort Bliss. NMFk analysis identifies spatial locations at White Sands Missile Range with the highest likelihood of geothermal success, whereas MT is used to detect subsurface geothermal prospects at McGregor Range based on resistivity. Four groups (Signals A, B, C and D) are established from NMFk analysis at White Sands and evaluated for their geothermal resource potential. Of these, Signal A is identified as having the highest geothermal potential based upon the co-occurrence of relatively high heat flow, reservoir temperatures, and vertical permeability. Through application of the MT method at McGregor Range, three resistivity layers (L1, L2 and L3) and two resistivity structures (RS1 and RS2) are

identified. Based on MT analysis, the highest likelihood for geothermal resources occurs in the western portion of McGregor Range where thick, low resistivity earth materials are present. Results from this study demonstrate how the integration of NMFk and MT can provide a 3D assessment of heat flow, temperature, and permeability for geothermal exploration.

Introduction

The United States Energy Information Administration projects a 50% increase in global energy consumption between 2020 and 2050 (U.S. Energy Information Administration, 2021). Geothermal energy is a nonintermittent, renewable resource which has the potential to alleviate concerns associated with growing global energy demand while mitigating carbon emissions attributed to the burning of hydrocarbons. Specifically, in the United States, geothermal electric power capacity has the potential to increase from 2.3 GW_e in 2019 to 60 GW_e by 2050 (Hamm et al., 2021; Tester et al., 2021). Thermal energy is generated from the decay of naturally occurring radioactive elements deep in the Earth. Even though there is an inexhaustible thermal energy supply in the subsurface, much of the heat is unevenly distributed, seldomly concentrated, and too deep to be economically exploited (Barbier et al., 2002). However, through hydrothermal convection, hot water may flow through naturally occurring vertical fractures and faults to locations accessible by drilling (Barbier et al., 2002, Jolie et al., 2021). These locations are the targets for geothermal exploration.

Productive geothermal systems have a combination of three major factors: high heat flow, high temperature and high permeability. Geothermal resources are confined to regions of high heat flow (up to 600 – 800 mW m^{-2}) and temperature (125 – 225 °C) as a result of magmatism and/or crustal thinning (Ussher et al., 2000, Blackwell et al., 2006, Elders and Moore, 2016; Jolie et al., 2021). Vertical fractures and faults play a key role in the potential permeability of a geothermal resource. Faults and fractures facilitate geothermal fluid flow from hot rock units located deep within the earth's crust, to faulted and fractured rock units located at shallower depths more suitable for geothermal energy exploitation (Jolie et al., 2021).

Integration of unsupervised machine learning and geophysical techniques are useful in the assessment of heat flow, temperature, porosity and permeability for geothermal exploration and development. Unsupervised learning is a subset of machine learning that analyzes and clusters unlabeled data. Non-negative matrix factorization with k-means clustering (NMFk) is a novel unsupervised machine learning algorithm that can cluster data related to heat flow, temperature, and permeability to establish signals with geothermal resource significance (Iliev et al., 2018; Vesselinov et al., 2018). NMFk has been successful in a variety of geothermal applications including the identification of spatial locations of potential geothermal resources (Vesselinov et al., 2021; Ahmmed and Vesselinov et al. 2021) and geologic factors associated with geothermal production (Siler et al., 2021). Magnetotellurics (MT) is a passive geophysical technique used for measuring electrical resistivity structures in the subsurface (Vozoff, 1991) and is commonly used to characterize geothermal resources (Johnson et al., 1992; Arnason et al., 2000; Cumming, 2007; Muñoz et al., 2014, Coppo et al., 2015; Cherkose and Mizunaga, 2018; Han et al., 2021; Miri et al., 2021). Resistivity is one of the most useful indicators in the identification of geothermal resources, varying considerably with porosity (Arnason et al., 2000; Ussher et al., 2000; Barbier, 2002). In general, high potential geothermal systems are characterized by low resistivity because of the occurrence of conductive geothermal fluids within the pore space. As such, geothermal systems are commonly associated with faulted and fractured rocks and associated secondary pore networks filled with low resistivity, high salinity geothermal fluids (Ussher et al., 2000 and Muñoz, 2014). The correlation between low resistivity and geothermal resources makes MT surveys ideal for geothermal resource exploration and development.

The objective of this study is to evaluate geothermal targets in the Tularosa Basin of south-central New Mexico based on heat flow, temperature, porosity and permeability using NMFk and MT. NMFk analysis identifies the spatial location with the highest likelihood of geothermal success based on heat flow, temperature, and permeability, and MT provides insight into subsurface porosity and the potential for thermal convection of associated water. A play fairway analysis (PFA) performed by Ruby Mountain Inc. and the Energy and Geoscience Institute at the University of Utah identifies two prospective geothermal locations in the Tularosa Basin of New Mexico: White Sands Missile Range and the McGregor Range at Fort Bliss (Bennett et al., 2020). White Sands Missile Range is used to demonstrate the ability of NMFk to characterize geothermal targets, whereas the McGregor Range is used to demonstrate the ability of MT to identify geothermal prospects from resistivity in the subsurface. This study aims to show how integrating NMFk and MT can provide a 3D assessment in terms of heat flow, temperature, and permeability of potential geothermal resources in the Tularosa Basin.

Geologic Background

The Tularosa Basin is located on the eastern flank of the Late Paleogene Rio Grande Rift (Seager and Morgan., 1979; Sinno et al., 1986). The Rio Grande Rift occurs as a north trending, intermontane graben within south-central New Mexico and is bounded to the east by the Sacramento Mountains and to the west by the Organ and San Andreas Mountains (Fig. 4.1). Faults associated with the Rio Grande rift have several thousand feet of displacement and separate the basin from the surrounding uplifted mountains (Sandeen, 1954). Paleogene rifting induced high heat flow within southwestern New Mexico, and therefore, makes the southern portion of the Tularosa Basin favorable for geothermal exploration (Blackwell et al., 2011; Nash and Bennett, 2015). In the southern part of the basin, temperatures recorded from drilled wells range from 170°C to 200°C (Finger and Jacobson, 1997; Blackwell et al., 2011; Nash and Bennett, 2015) and clay mineral analysis indicate temperatures as high as 225°C (Ussher et al., 2000; Barker et al., 2014).

The Tularosa Basin is filled with strata of Paleozoic to Tertiary age (Finger and Jacobson, 1997; O'Donnell Jr. et al., 2001, Broadhead et al., 2002; Barker et al., 2014) (Fig. 4.2). Bedrock consists primarily of Paleozoic carbonates, including Ordovician and Silurian dolomite, Devonian and Mississippian interbedded chert-rich shales and limestones, and Pennsylvanian limestone with thinly bedded shales. Tertiary felsic intrusions commonly cross-cut the Paleozoic bedrock, and Quaternary graben fill overlies the bedrock and is composed of gravel, sand, silt, and clay derived from prograding alluvial fans originating from the edge of the rift valley.

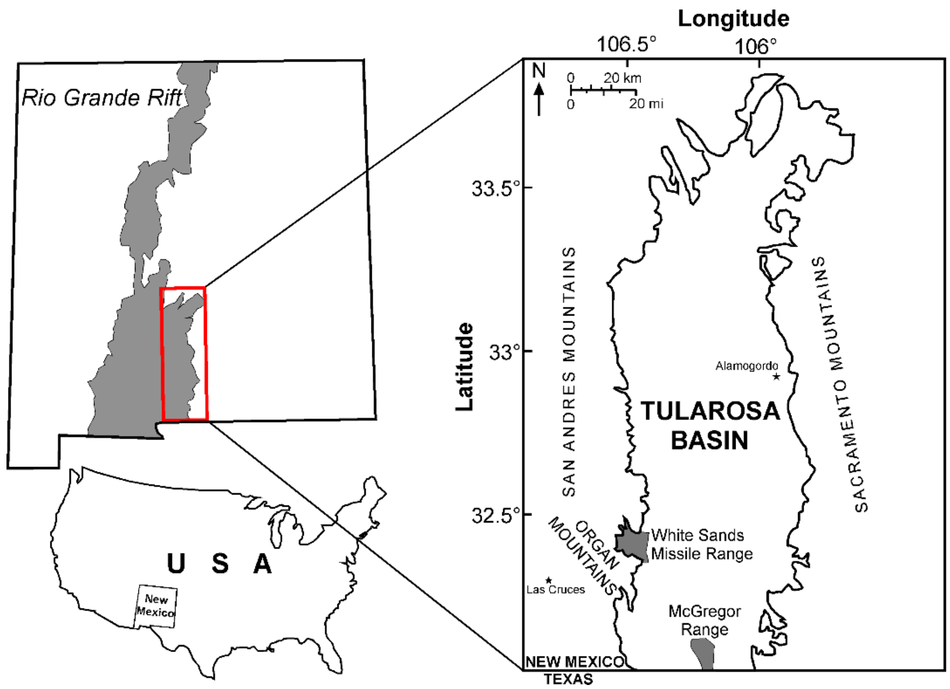

Figure 4.1. Location of the White Sands Missile Range and McGregor Range study areas within the Tularosa Basin in southern New Mexico. The Tularosa Basin is an intermontane graben located on the eastern flank of the Rio Grande Rift zone.

	Stratigraphic Unit	Generalized Lithological Description
CENOZOIC	Quaternary	Alluvial basin fill
	Tertiary	Felsic sills and dikes
PALEOZOIC	Pennsylvanian	Black lime mudstones and wackestones; cherty limestones interbedded with grey shale
	Mississippian	Limestone interbedded with black shale
	Devonian	Black shale
	Silurian	Cherty dolomite
	Ordovician	Dolomite

Figure 4.2. Generalized stratigraphic succession of the Tularosa Basin sedimentary and igneous fill.

Methods

NMFk

Non-negative matrix factorization with customized k-means clustering (NMFk) is used to determine the high geothermal resource prospective areas (Ahmmed, 2021, Siler et al., 2021) by automatically identifying the optimal number of groupings/signatures within a geothermal dataset. NMFk couples two unsupervised machine learning techniques called non-negative matrix factorization (NMF) and k-means clustering. NMF decomposes/factorizes the data matrix, $X_{m \times n}$, into $W_{m \times k}$ and $H_{n \times k}$ (Lee and Seung, 1999). The optimal number of signatures is represented by k. $W_{m \times k}$ and $H_{n \times k}$ matrices represent groupings in location and attribute, respectively. K-means clustering measures the goodness of each NMF solution ($W_{m \times k}$ and $H_{n \times k}$ matrices) using silhouette width (Rousseeuw, 1987, Vesselinov et al., 2019) that helps explaining the outputs. Mathematical details and its application to different datasets can be found at Vesselinov et al., 2014; 2018; Vesselinov et al., 2019.

NMFk was implemented using GeoThermalCloud, an open-source tool, available at https://github.com/SmartTensors/GeoThermalCloud.jl. GeoThermalCloud capabilities include (1) analyzing large field datasets, (2) assimilating model simulations (large inputs and outputs), (3) processing sparse datasets, (4) performing transfer learning (between sites with different exploratory levels), (5) extracting hidden geothermal signatures in the field and simulation data, (6) labeling geothermal resources and processes, (7) identifying high-value data acquisition targets, and (8) guiding geothermal exploration and production by selecting optimal exploration, production, and drilling strategies.

The following 10 attributes are used as input data for NMFk analysis to assess heat flow, temperature, and permeability of potential geothermal resources from 120 locations (Fig. 4.3): heat flow, gravity, temperature at 2m depth (temperature@2m), NaK-Giggenbach geothermometer, K-Mg geothermometer, NaK-Fourneir geothermometer, silica geothermometer, fault distance, quaternary fault density, and Lithium concentration. Heat flow measurements are from the 2011 “SMU Geothermal Laboratory Heat Flow Map of the Coterminous United States” (Blackwell et al., 2011). Heat flow can furthermore be assessed through gravity data as positive gravity anomalies correlate with high heat flow (Atef et al., 2016). Data from shallow temperature surveys and various geothermometers are used to evaluate subsurface temperature. Temperature surveys at 2 meters are effective in detecting thermal anomalies (Coolbaugh et al., 2007; Zehner et al., 2012), whereas the various geothermometers estimate subsurface temperatures based on elemental concentration distributions in ground water (Fournier, 1977). Permeability is evaluated based on distance to the nearest fault (fault distance) and the number of quaternary faults per square meter (faults density), since faults can act as conduits for water transmission and storage. Li concentrations are an indicator of vertical permeability and associated hydrothermal convection because Li is derived from deeper magmatic water flow. (Wang et al., 2020).

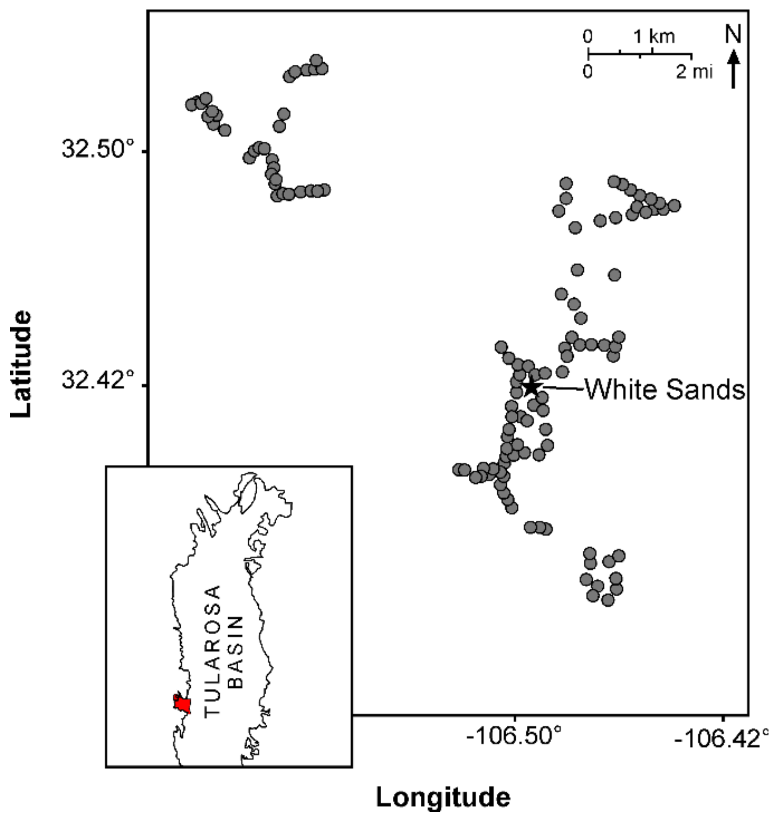

Figure 4.3. Locations (120 total) near White Sands selected as input data for NMFk. At each site, 11 geothermal attributes are collected and used as input into the NMFk model.

With an observational dataset, it can be difficult to obtain values for all 10 attributes at each location. In this study the only attribute available at all 120 locations is temperature@2m. For locations lacking the other attributes, we applied interpolation techniques so that all 10 attributes are included at each location. For heat flow, geothermometers, gravity, and Li concentration, three interpolation methods were applied and evaluated: block mean, kriging, and inverse distance weighting. Evaluation metrics, computed R^2 scores based on interpolated values and real values, were equivalent for the three interpolation methods. Block mean was ultimately selected due to its low computational time. ArcMap was used to interpolate fault distance and fault density values. Specifically, the near coverage tool was used to find the distance from location to nearest fault and the kernel density function was used to calculate fault density.

Magnetotellurics

Magnetotellurics (MT) is a passive geophysical technique used for measuring electrical resistivity structures (Vozoff, 1991). Solar winds and lightning from thunderstorms cause natural variations in the earth's magnetic field which penetrates the subsurface and induces an electrical current (Coppo et al., 2015) The electromagnetic fields (EM) from an MT survey are recorded at frequencies ranging generally from 0.001 kHz to 10 kHz (Yadav et al., 2020). The low frequency response originates from solar winds and the high frequency response originates from worldwide lightning strikes (Coppo et al., 2015; Cherkose and Mizunaga 2018).

In MT data a time series of the 2 components of the electric field (Ex and Ey) and 3 components of the magnetic field (Hx, Hy, and Hz) are measured on the earth's surface (Fig. 4.4). The ratio between the electric and magnetic field components (E/H) is called the impedance tensor (Z). As a proportion of the electric and magnetic fields the impedance tensor can be written as:

$$\boldsymbol{Z} = \frac{\boldsymbol{E}}{\boldsymbol{H}}$$

The horizontal components of electrical and magnetic fields are related to Z as follows:

$$\begin{pmatrix} E_x \\ E_y \end{pmatrix} = \begin{pmatrix} Z_{xx} & Z_{xy} \\ Z_{yx} & Z_{yy} \end{pmatrix} \begin{pmatrix} H_x \\ H_y \end{pmatrix}$$

The impedance tensor, Z, is used to determine the apparent resistivity and phase (Coppo et al., 2015; Cherkose and Mizunaga, 2018). The following equations use the components of Z to calculate apparent resistivity (ρ_{ij}) and phase (ϕ_{ij}):

$$\rho_{ij} = \frac{1}{\omega\mu_0} \left| Z_{ij} \right|^2$$

$$\phi_{ij} = tan^{-1}\left(\frac{Im(Z_{ij})}{Re(Z_{ij})}\right)$$

Where Im and Re are the imaginary and real parts of the impedance component, respectively. Both apparent resistivity and phase are commonly plotted as a function of frequency for MT data analysis of subsurface structures.

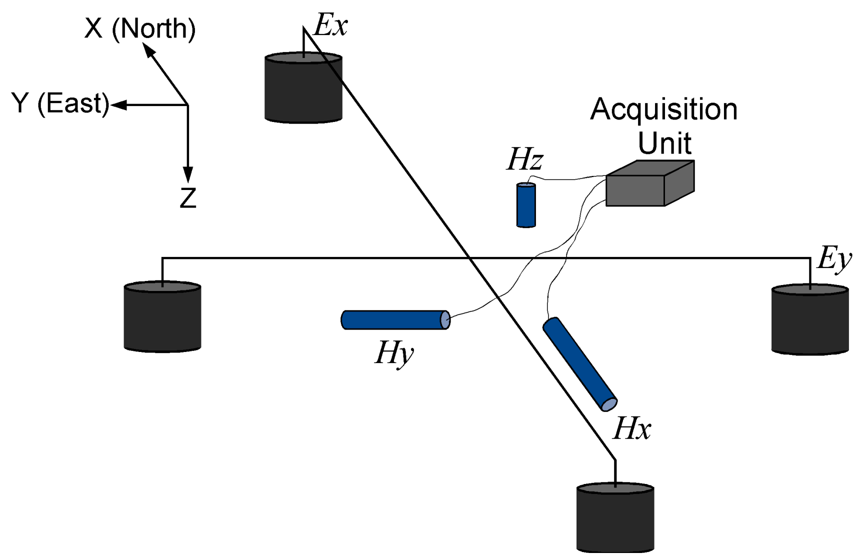

Figure 4.4. Schematic of the arrangement and setup of electrodes and coils in the field during MT data acquisition (modified from Grimm et al., 2021).

A 56 station MT survey was conducted at the McGregor Range by Quantec Geoscience and the inversion modelling was accomplished by the Energy and Geoscience Institute at the University of Utah (Fig. 4.5). Details about the inversion are described by Bennett and Nash (2020).

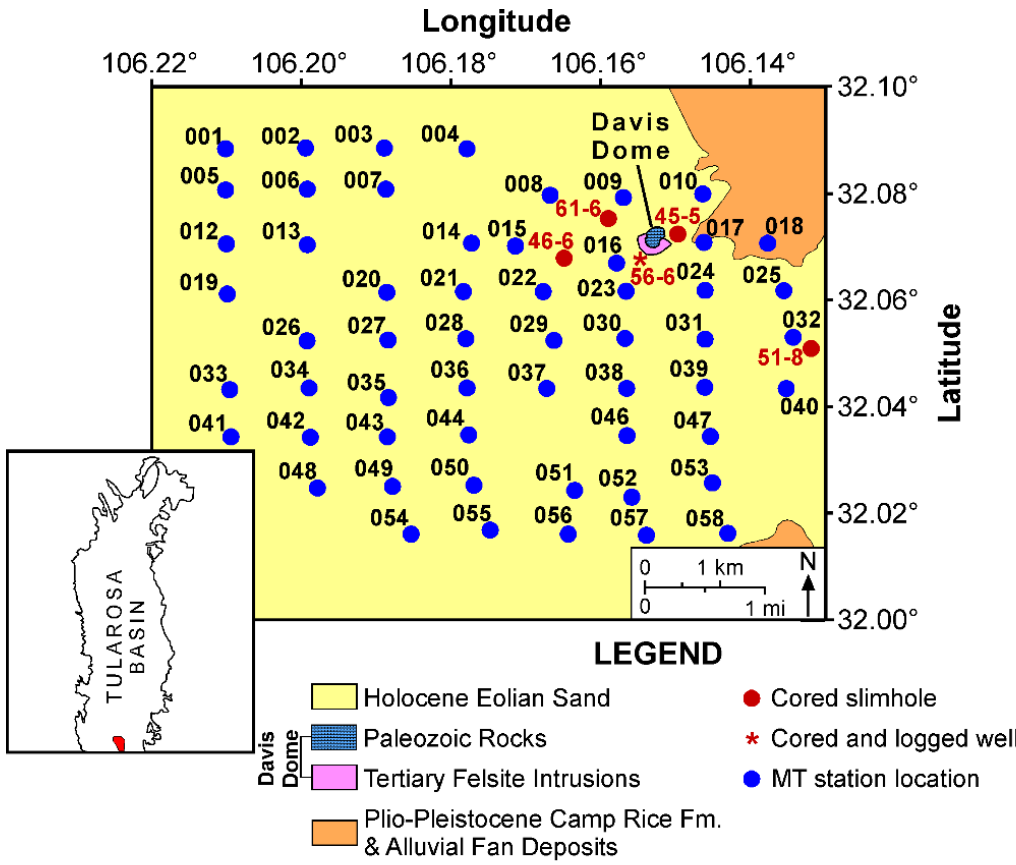

Figure 4.5. Geologic map of McGregor Range with MT station locations (blue) and slimhole core locations (red). McGregor is largely covered by Recent eolian sands, although Paleozoic and Tertiary outcrops occur in the northeast portion of the study area at Davis Dome.

Results

Geothermal Characterization of NMFk Signals

NMFk helps determine the optimal solution for k by comparing reconstruction error $O(k)$ and average silhouette width $S(k)$ (Fig. 4.6). Optimal solutions have low $O(k)$ and high $S(k)$values. Generally, low $O(k)$ and $S(k)$>0.25 are acceptable solutions (Ahmed et al., 2021). NMFk was ran for 2 to 10 signals and k=4 solution is found to be optimal solution because of its low $O(k)$ and high $S(k)$ values. The solution with k<4 is an underfitting representation of data whereas k>4 is an overfitting representation of data.

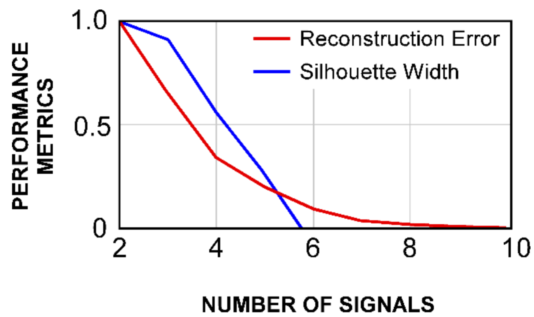

Figure 4.6. NMFk reconstruction error (red line) and silhouette width (blue line) for different numbers of clusters *k.* The optimal *k* value has low reconstruction error and higher silhouette values. In this study, the optimal number of signals is 4.

Geothermal signal heatmaps identify dominant attributes in each signal (Fig. 4.7). The warm colors represent a high weight between the signal and attribute and the cool colors represent a relatively low weight. Furthermore, for the geothermal attributes the warm colors correlate to high values and the cool colors correlate to lower values. The dominant attributes of signal A are heat flow, K-Mg geothermometer, silica geothermometer and quaternary fault density indicating high heat flow, subsurface temperature, and permeability. Similar to signal A, signal B is characterized by high heat flow and temperature@2m. Furthermore, the high Li concentration indicates that signal B is characterized by high vertical permeability. No geothermometer had a significant contribution to signal B. Fault distance is the major attribute in signal C. This indicates locations assigned as C have lower potential vertical permeability because they are relatively far from faults which act as conduits for fluid flow. Signal C is characterized by lower heat flow and temperature relative to the other signals. The dominant attributes for signal D are Na-K Giggenbach geothermometer and NaK-Fourneir geothermometer indicating high subsurface temperatures. Moderate weights for quaternary fault density and Li concentrations in signal D indicate relatively high permeability. Heat flow and temperature@2m have a relatively low contribution to signal D.

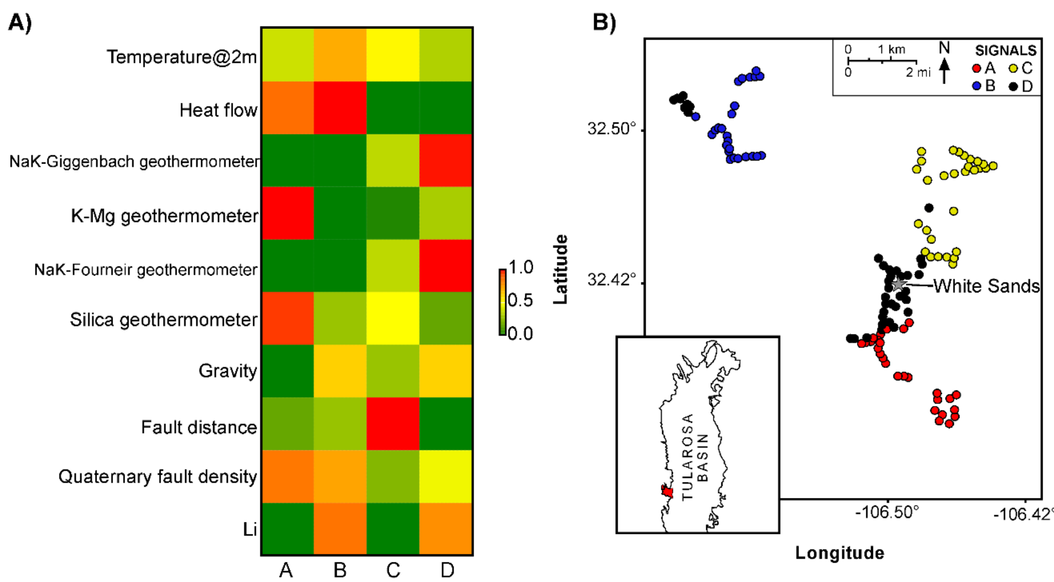

Figure 4.7. Results from the NMFk model. A) Heatmap identifying the dominant geothermal attributes in each signal. The warmer the color the more dominant the attribute for a particular signal. B) Spatial distribution of signals for the 120 locations at White Sands.

Subsurface Characterization of Potential Geothermal Locations

Data imaging and analysis. Apparent resistivity and phase curves display resistivity trends using period as a proxy for depth (longer periods correspond to increased depth). Congruent MT apparent resistivity curves of Z_{xy} and Z_{yx} indicate 1D resistivity structure, whereas separation indicates more complicated 2D or 3D resistivity structure (Coppo et al., 2015, Zhang et al., 2015). For example, MT apparent resistivity curves for station MT017 located in the northeast section of the survey show separation between the two curves at shorter periods, i.e., shallower depths (Fig. 4.8). This corresponds to geological structures related to Davis Dome, a small intra-bolson horst near station MT017 (Fig. 4.5 and Fig. 4.8) (Barker et al., 2014).

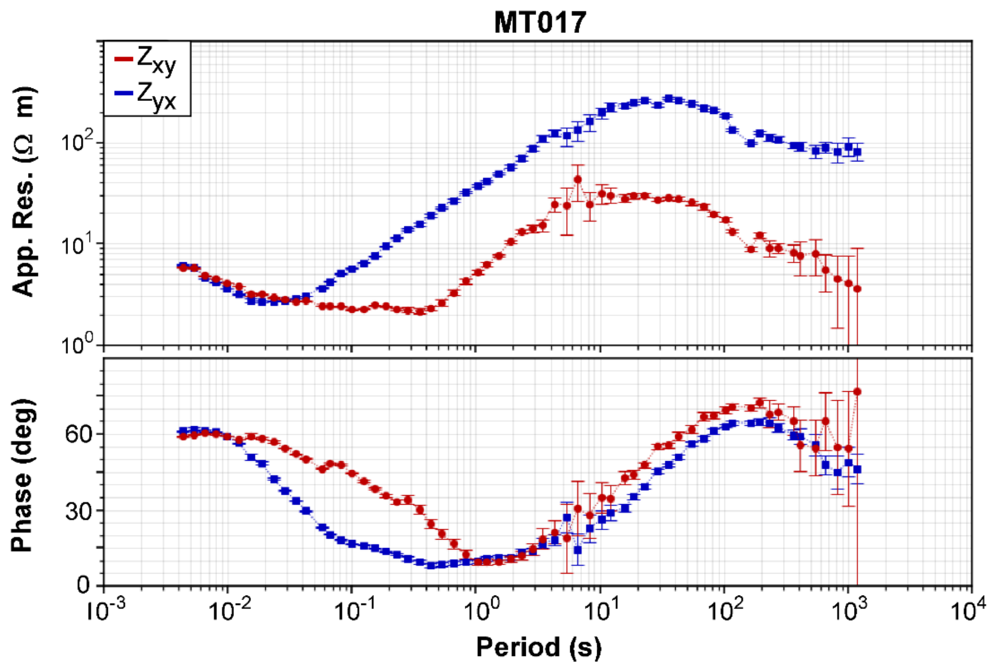

Figure 4.8. Apparent resistivity curves for MT station 017. Period is a proxy for depth, i.e., longer periods are deeper depths. Separation of Z_{xy} and Z_{yx} curves at 0.05 s indicate 2D or 3D resistivity structures.

MT apparent resistivity and phase curves of Z_{xy} and Z_{yx} from all 56 MT sites are shown in Figure 4.9. The apparent resistivity values show a cyclic trend from shorter to longer periods (shallower to deeper depths). At shallower depths, the apparent resistivity gradually decreases. Between 1 s and 100 s the apparent resistivity increases. At deeper depths, longer than 100s, the apparent resistivity decreases. Furthermore, at longer periods, the Z_{xy} and Z_{yx} curves diverge indicating complex, 3D resistivity structure at deeper depths. The depth of the low apparent resistivity varies from east to west. For MT sites 019, 022, and 025 the troughs for apparent resistivity occur at 1 s, 0.3 s, and 0.1 s, respectively (Fig. 4.10). The longer period to the west indicates the low resistivity unit occurs deeper in the west than the east.

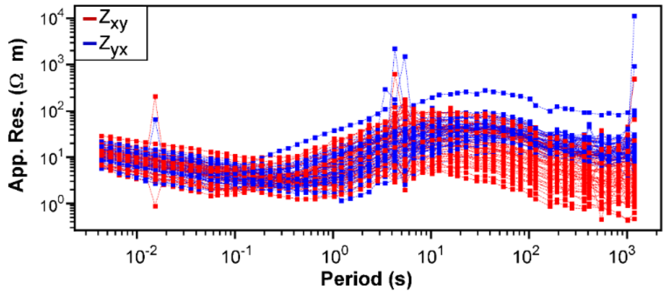

Figure 4.9. Apparent resistivity curves from all 56 MT stations. From shorter to longer periods the general apparent resistivity trend is lower at shorter periods, increases at medium periods, and then decreases at longer periods for a low-high-low trend. Furthermore, at longer periods, the Z_{xy} and Z_{yx} curves separate indicating 2D or 3D structure at deeper depths.

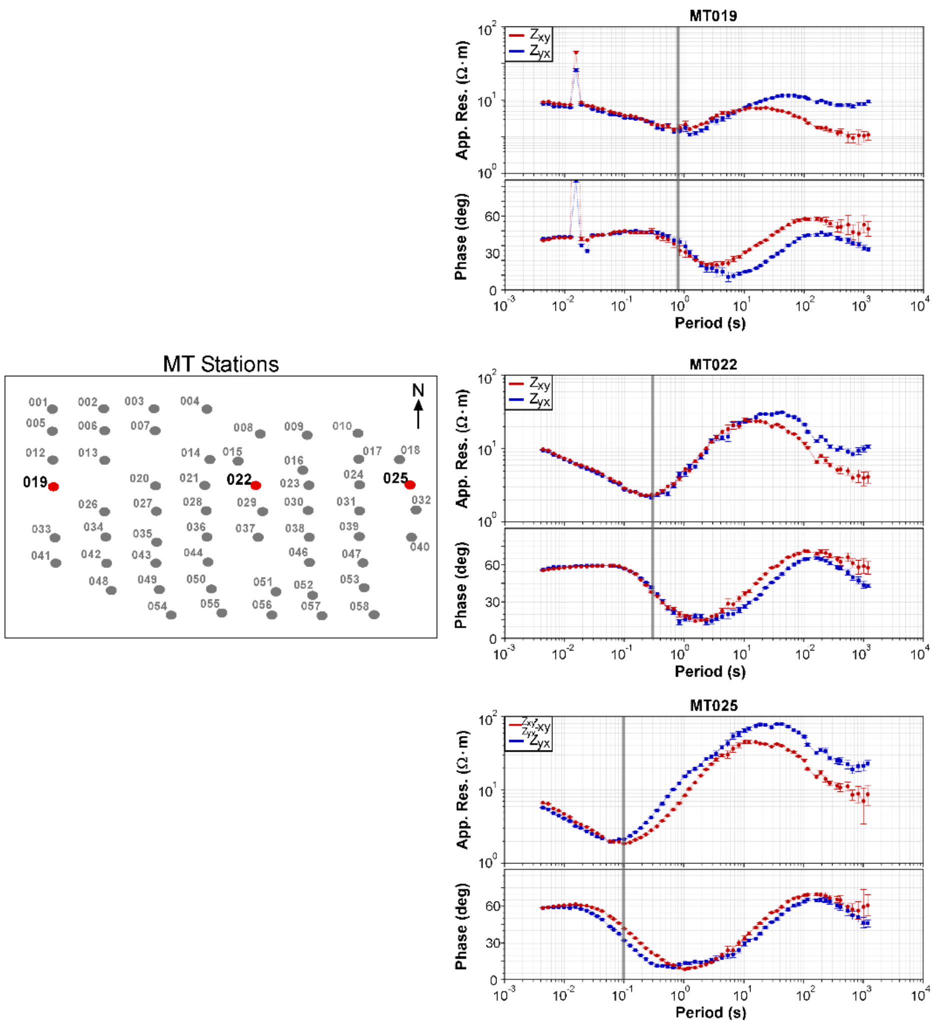

Figure 4.10. Apparent resistivity curves for MT sites 019, 022, and 025. The curves show a change in the low resistivity unit depth from west to east. The low apparent resistivity trough of the Z_{xy} and Z_{yx} curves for MT019 occur at 1 s. The trough of the Z_{xy} and Z_{yx} curves for MT022 occur at 0.3 s. The trough of the Z_{xy} and Z_{yx} curves for MT025 occur at 0.1 s. The decrease in periods from west to east indicates the low resistivity unit occurs deeper in the west than the east.

Phase tensor analysis. Dimensionality of the resistivity structure is determined by phase tensors. One-dimensional resistivity structures indicate a natural change in resistivity with depth due to compaction (James et al., 1987; Caldwell et al., 2004). Phase tensors are useful in the identification of lateral variations (2D and 3D resistivity structures) in the underlying regional resistivity (Caldwell, et al., 2004). Lateral variations in resistivity are a result of changes in porosity due to fault- and/or fracture-related diagenesis, and/or changes in lithology. A phase tensor (ϕ) is the ratio of the real (*X*) and imaginary parts (*Y*) of the complex impedance tensors (*Z*).

$$\boldsymbol{\phi} = X^{-1}Y$$

where

$$\boldsymbol{Z} = X + iY$$

The phase tensor is commonly plotted as an ellipse with a minimum and maximum principal axis and skew angle, β (measure of asymmetry) (Fig. 4.11A). In Figure 4.11B, ellipses are colored based on the skew angle. Yellow colors indicate a skew angle of 0 and the red and blue colors indicate larger skew angles (± 5°). The larger the skew angle, the more asymmetric the phase tensor indicating higher dimension resistivity structures. For 1D resistivity structures, the minimum and maximum principal axes are the same ($\phi_{max} = \phi_{min}$) resulting in the phase tensor characterized by a yellow, circular shape. The phase tensor of a 2D resistivity structure is characterized with an elliptical shape and a skew angle close to zero (± 3°). For 3D resistivity structures the phase tensor is asymmetric, and hence, the phase tensor is characterized by blue or red color. Furthermore, a rapid direction change in the phase tensor's principal axes between sites is indicative of 3D resistivity structure (Cherkose and Mizunaga, 2018).

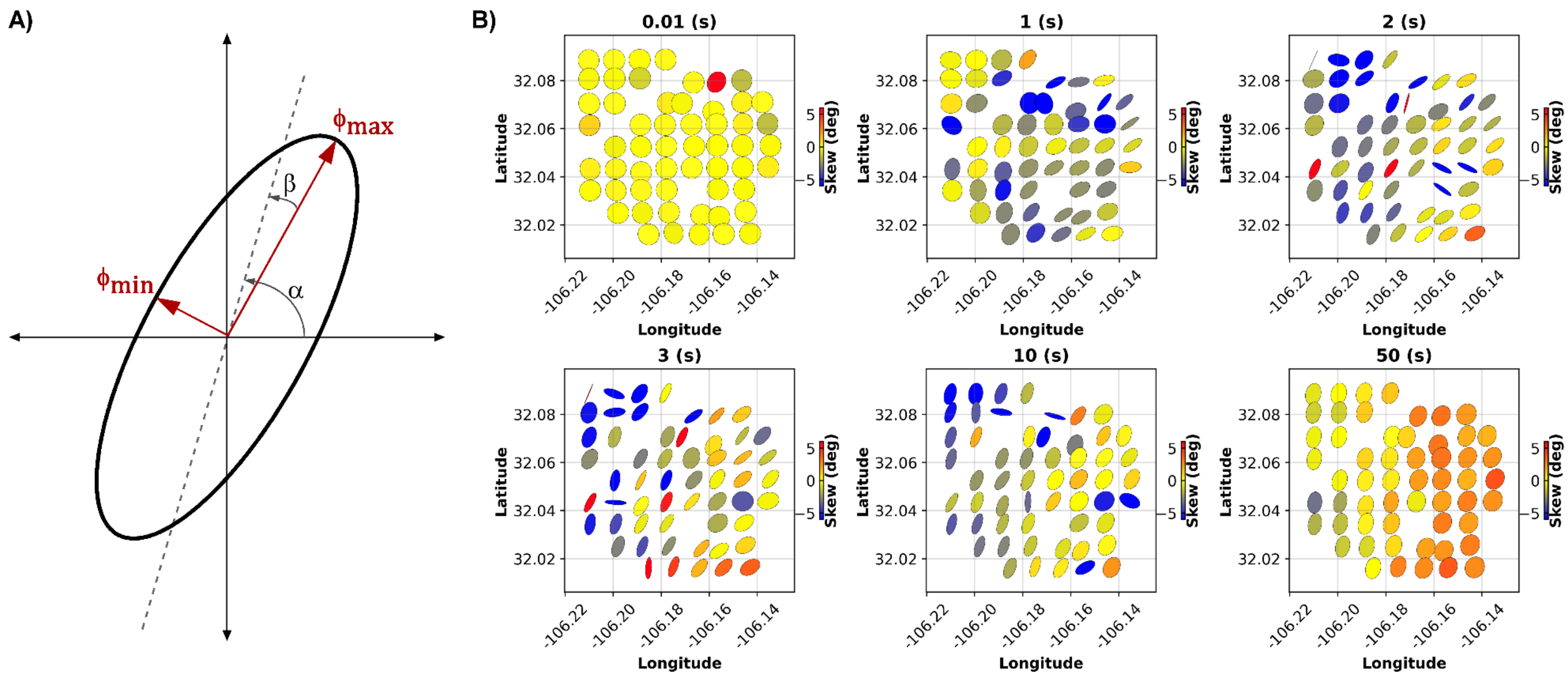

Figure 4.11. A) Graphical representation of the phase tensor (from Caldwell et al., 2004). B) Phase tensor maps at 0.1, 0.01, 1, 10, 50 and 100 s indicated the spatial distribution of resistivity structures with depth.

In general, across the study area the shorter periods are characterized by 1D resistivity structures and then higher dimension 2D and 3D resistivity structures with depth (Fig. 4.11B). This observation is consistent with the separation in apparent resistivity curves at longer periods (Fig. 4.8). Specifically, at 0.01 s and 0.1 s the phase tensors are characterized by 1D structures as indicated by the yellow circles. An exception to this observation is the northeastern corner of the study area where the shape of the tensors is more elliptical, and the skew angle is higher indicating lateral variation in the resistivity structure. This increase in dimensionality is consistent with shallow structural features and northwest trending faults associated with Davis Dome (O'Donnell et al., 2001). The shape of the ellipses and the red, blue, and orange colors at periods greater than 1 s indicate 2D or 3D resistivity structures. Specifically, the abrupt changes in the ellipses shape at 10 s suggest possible faulting. Caution must be taken when interpreting phase tensors at longer periods as they are more affected by attenuation.

Discussion

Geothermal Resource Potential of Signals

Temperature, heat flow, and permeability are the main geothermal attributes driving geological success (Jolie et al., 2021). Signal A has high geothermal resource potential because of the characteristically high heat flow, high K-Mg and silica geothermometers, and medium to high quaternary fault density. Therefore, signal A has a high likelihood of possessing higher temperature, heat flow and permeability compared to the other signals. Signal D has moderate geothermal potential because of the combination of high NaK-Giggenbach and NaK-Fourneir geothermometers values and low temperature@2m and heat flow. Signal B has moderate geothermal potential because of high temperature@2m, heat flow, quaternary fault density and Li concentrations; however, low values for the geothermometers suggest subsurface temperatures may not be as suitable as other signals. Signal C has the lowest geothermal resource potential because no geothermal attributes have a major contribution to the signal.

MT Inversion Interpretation

Once a spatial location is determined through NMFk, MT can be a valuable tool to aid in the subsurface characterization of a potential geothermal resource by analyzing resistivity trends. We interpret the MT data of the McGregor geothermal system as having 3 mappable resistivity layers and 2 resistivity structures (Fig. 4.12).

Layer 1 (L1) is characterized by the lowest resistivity (<8 Ωm) and is confined generally to the upper 500 m of the study area. L1 is thickest to the west and thins to about 300 m in the east (Fig. 4.12). L1 is thinnest in the northeast corner near Davis Dome. This regional low resistivity cap is most likely attributed to basin fill deposits. O'Donnell, Jr., et al. (2001) performed a seismic reflection survey over the same study area and observed a wedge-shaped feature above the bedrock that was attributed to alluvial fan deposits shed from the surrounding mountains. The observed thickening of L1 to the west in the MT data is consistent with the wedge-shaped feature observed in the seismic survey (Fig. 4.12).

Layer 2 (L2) is a low resistivity (10 – 100 Ωm) layer with the top located 200 – 600 m beneath the surface. In general, L2 is shallower to the east (~300 m) and deeper to the west (~ 600 m). Wells drilled in the northeast portion of the study area suggest the top of L2 corresponds to Paleozoic (Pennsylvanian limestone) bedrock (Finger and Jacobson, 1997; Barker et al., 2014). Finger and Jacobson (1997) observed and measured fracture permeability in core in nearly all Paleozoic units. Phase tensor analysis in the western part of the study area indicate 2D resistivity structure that suggests the presence of a possible fault system (Fig. 4.11). The thicker and lower resistivity L2 in the west may be attributed to an increase in fractures and/or faults that act as storage or conduits for geothermal fluids decreasing resistivity, i.e., L2 in the west is influenced by higher fractured and/or faulted units.

A low resistivity structure (RS1) is present below MT stations 039, 047, 051, 052, and 053 in the southeast section of the study area (Fig. 4.12). The structure has similar resistivity as L2 but extends to 2000 m. The lower resistivity of RS1 is interpreted to be related to a deformation observed in surrounding wells. A thrust fault and overturned beds are observed in core from well 51-8 located to the northeast of cross section EW 3 suggesting deformation in the area (Fig. 4.12) (Finger and Jacobson, 1997). Units related to this structure are pervasively fractured and may provide a conduit for fluid flow and associated lower resistivity in RS1 as observed in the Figure 4.12 (Finger and Jacobson, 1997; O'Donnell et al., 2001).

Layer 3 (L3) is characterized by the highest resistivity values (>100 Ωm) and the top is located about 250 – 2000 m beneath the surface (Fig. 4.12). In the east, the top of L3 is shallower (~500 m) and in the west the top of L3 is deeper (~1800 m).

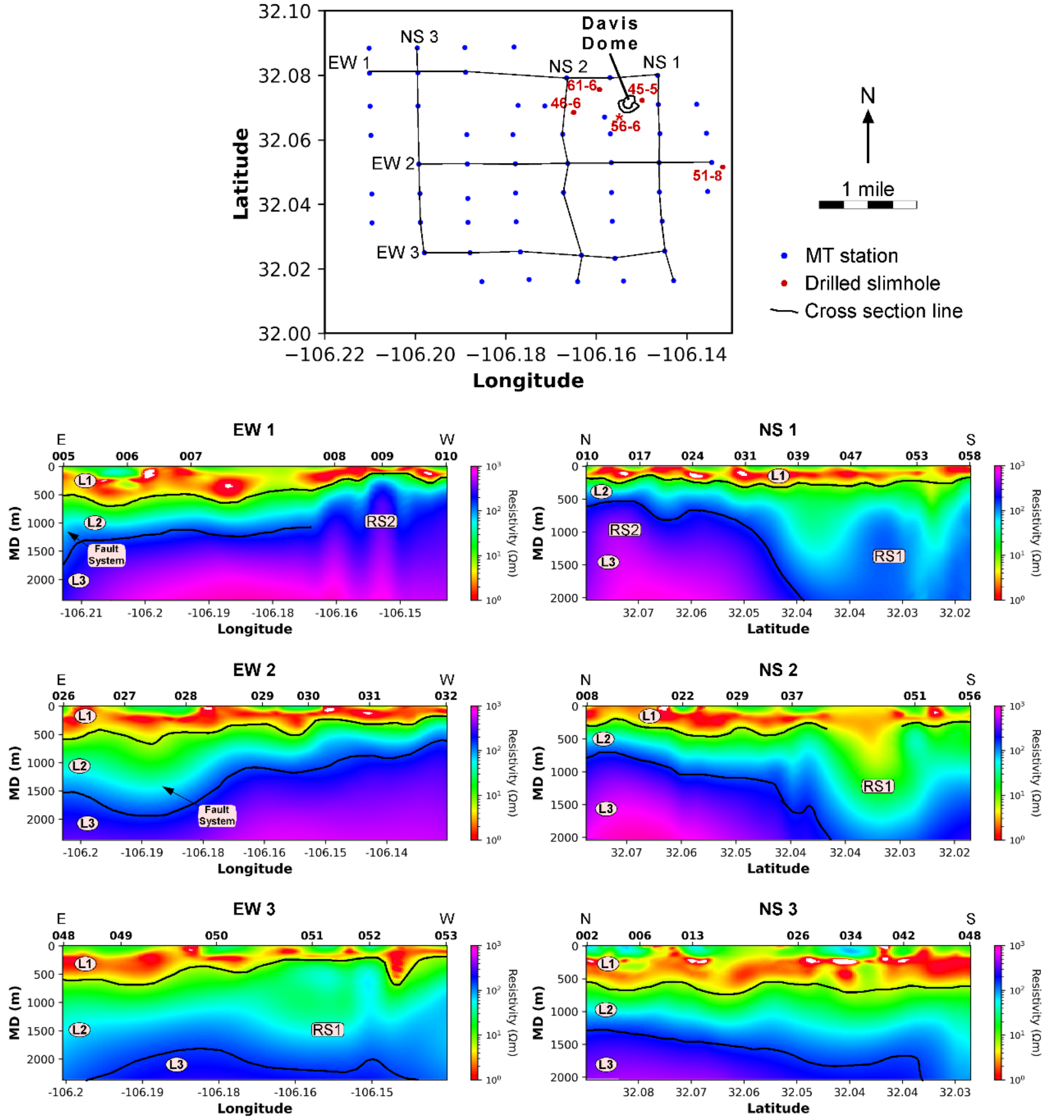

Figure 4.12. Three north-south and east-west MT cross sections with interpreted resistivity layers and structures. The low resistivity to the west is interpreted as a fault system. The faults, fractures, and possible dissolution as a result of geothermal fluids increase porosity thus decreasing resistivity. Assuming temperature is consistent with a geothermal reservoir, the west-central part of the McGregor Range has the highest geothermal potential because of the increase in porosity and associated permeability attributed to the interpreted fault system.

A high resistivity structure (RS2) is present in the northeast portion of the study area and has similar resistivity values to L3 (Fig. 4.12). The spatial location and cored wells in the area (45-5, 46-6, and 61-6) suggest RS2 coincides with structures related to

Davis Dome, an intrusive igneous laccolith (Fig. 4.12). Cored wells encounter felsite sills, a felsite laccolith, and Mississippian limestone and shale at relatively shallow depths between 360 m and 530 m (Finger and Jacobson, 1997; O'Donnell Jr. et al., 2001). The thin L2 layer above RS2 is most likely fractured Paleozoic strata, and high resistivity RS2 is most likely a low permeability felsic body associated with the Davis Dome intrusion (O'Donnell Jr. et al., 2001). These interpretations are consistent with a structural high from a laccolith intrusion observed in seismic, velocity and gravity models from O'Donnell Jr, et al. (2001).

The west-central section of L2 is interpreted as a possible fault system and has the highest geothermal potential. Geothermal reservoirs tend to have resistivity values between 10 – 60 Ωm similar to those observed in L2 (Cherkose and Mizunaga, 2018, Johnston et al., 1992, Yadav et al., 2020). The location where L2 is the thickest coincides with north-northwest trending, anomalously high thermal gradients (up to 140°C/km) delineated by Roy and Taylor (1980). According to Henry and Gluck (1981) the anomaly may be due to geothermal waters rising along a common fault zone or fractured bedrock adjacent to the fault zone which is consistent with the highly faulted and/or fractured units observed in L2 to the west (Fig. 4.12). Furthermore, the westward thickening of L2 suggests the possibility for a corresponding increase in reservoir transmissivity and increase in well productivity (Hurter and Schellschmidt, 2003; Augustine, 2014).

Limitations of MT

MT data is limited by its hectometer-scale vertical resolution. Resistivity is measured in well 56-6 using wireline logs that have a vertical resolution of 0.6 m (Passey et al., 2006). When compared to inverted MT resistivity, the well logs provide more detailed variations in resistivity (Fig. 4.13). For example, from 90 – 220 m, well log resistivity is characterized by high variability due to thinly interbedded limestones and shales that are not detected in the MT resistivity. Only general interpretations of fluid saturation and porosity can be made with MT data because of the low vertical resolution.

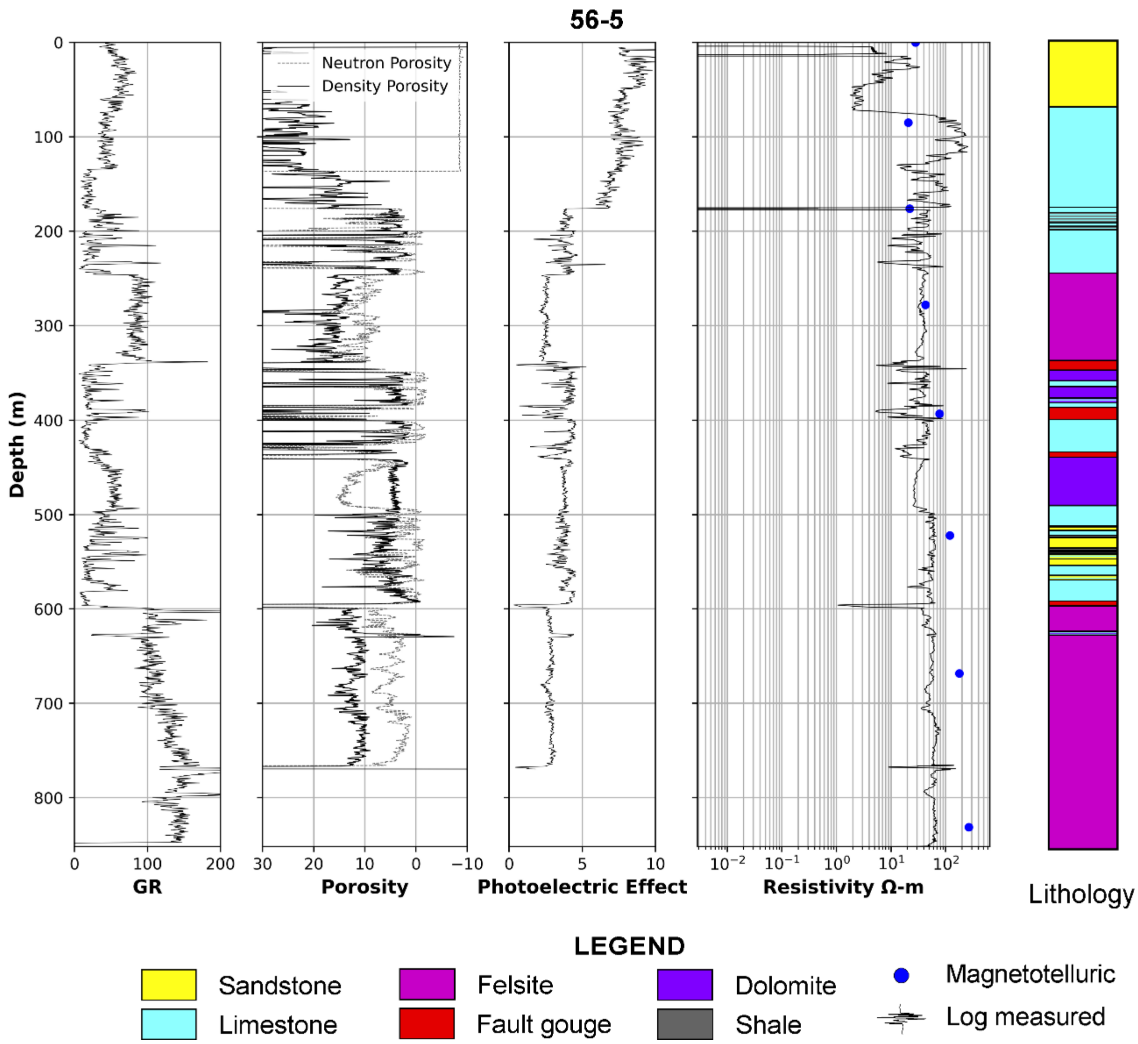

Figure 4.13. Comparison of resistivity from wireline logs and MT. The resistivity from well logs provides more detailed variations in resistivity that are not detected from lower resolution, inverted MT resistivity data. Lithologies are based on petrophysical interpretations from accompanying gamma ray (GR), neutron and density porosity, photoelectric effect (PE), and deep resistivity logs (Alberty, 1992) and core cutting descriptions.

Lithologic interpretations from MT inversions are difficult to make since resistivity is highly influenced by porosity and associated pore-filling fluid (Ussher et al., 2000). Because all rock matrices are highly resistive across the study area and saturated with similarly saline water, the primary control on resistivity variations is porosity. In general, lithification increases with depth and is associated with a decrease in porosity and permeability that is consistent with the observed increase in MT resistivity from L1 to L3 (Fig. 4.12) (James et al., 1987). Correlations between the four cored wells in the northeast portion of the study area indicate that L1 coincides with Quaternary basin fill that is under-compacted and highly porous and permeable, and therefore, characterized by low resistivity. Older strata associated with L2 and L3 are highly compacted and cemented and characterized by lower porosity and permeability and higher resistivity. The transition from L2 to L3 is controlled by porosity rather than lithology. For example, as seen in EW2, L2 thickness increases to the west suggesting an increase in porosity. The phase tensors in the west show 2D resistivity structures with increasing depth suggestive of a possible fault system (Fig. 4.12). Secondary pore networks derived from fluid-rock interactions induced by the high permeability fault system are interpreted to be filled with high salinity, low resistivity fluids. Also, cored wells 61-6, 45-5, 46-6, 56-6, and 51-5 indicate that L2 and L3 coincide with Paleozoic bedrock composed primarily of resistive carbonates. The resolution limitations of MT resistivity measurements mean that small-scale changes in carbonate lithologies are not detected and suggests that differences between L2 and L3 are not related to lithology (Fig. 4.13) (Finger and Jacobson, 1997).

Conclusion

1) NMFk is a useful analytical tool to assess prospective geothermal regions through the evaluation of variability in heat flow, porosity, permeability, and temperature. In the southwestern portion of the Tularosa Basin at White Sands Missile Range, 4 signals were established through NMFk that were evaluated for their geothermal resource potential. Signal A is interpreted to have the highest geothermal potential due to a combination of high heat flow, reservoir temperatures, and comparatively high porosity and permeability. Signals B and D have moderate potential because of their relatively low heat flow and temperature. Signal C has the lowest geothermal resource potential because no geothermal attributes have a major contribution to the signal.

2) MT inversions detect subsurface geothermal prospects based on resistivity. MT provides insight into relative porosity and associated permeability that is related to the subsurface resistivity trends detected in the MT inversion. From a MT survey from McGregor Range, three resistivity layers (L1, L2 and L3) and 2 resistivity structures (RS1 and RS2) are identified. The layers are inferred to be related to a combination of depth-related compaction and lithification effects and the resistivity structures are related to Davis Dome, an intrusive igneous laccolith, and differential faulting and fracturing. A fault system is interpreted in the western portion of the study area as indicated by the thickening of L2. Because low resistivity is a defining characteristic of geothermal prospects, the western portion of the McGregor MT survey has the highest geothermal potential.

3) The low vertical resolution of MT data, in contrast with high-resolution borehole resistivity measurements, make it difficult to relate lithological variability and associated rock attributes with MT inversions. MT is limited in that the interpreted resistivity layers only provide insight into relative porosity and do not correlate with lithological or stratigraphic units. Only large-scale characterization of porosity and associated permeability can be made when interpreting MT inversions.

References

Ahmmed, B., Vesselinov, V.V, 2021. Prospectivity Analyses of the Utah FORGE Site using unsupervised machine learning, Geothermal Resource Council Transactions 45.

Alberty, M.W., 1992, Basic open hole tools, in Development Geology Reference Manual, D. Morton-Thompson and A.M. Woods, eds., AAPG Exploration in Methods Series 10, p. 144-149.

Alexandrov, B.S., Vesselinov, V.V., 2014. Blind source separation for groundwater pressure analysis based on nonnegative matrix factorization. Water Resource. Research 50, 7332–7347. https://doi.org/10.1002/2013WR015037

Arnason, K., Karlsdottir, R., Eysteinsson, H., Flovenz, O.G., Gudlaugsson, S.T., 2000. The resistivity structure of high-temperature geothermal systems in Iceland. Proceedings of the World Geothermal Congress, May–June, Japan, 923–928.

Atef, H., Abd El Gawad, A., Abdel Zaher, M., Farag, K., 2016. The contribution of gravity method in geothermal exploration of southern part of the Gulf of Suez–Sinai region, Egypt. NRIAG Journal of Astronomy and Geophysics 5, 173–185. https://doi.org/10.1016/j.nrjag.2016.02.005

Augustine, C., 2014. Analysis of sedimentary geothermal systems using an analytical reservoir model, Geothermal Resource Council Transactions 38, 641–648.

Barbier, E., 2002. Geothermal energy technology and current status: an overview. Renewable and Sustainable Energy Reviews 6, 3–65. https://doi.org/10.1016/S1364-0321(02)00002-3

Barker, B., Moore, J., Segall, M., Nash, G., Simmons, S., Jones, C., Lear, J., Bennett, C., 2014. Exploration drilling and technology demonstration at Fort Bliss. 39th Workshop on Geothermal Reservoir Engineering, February, California, 1–12.

Bennett, C.R., Nash, G.D., Barker, B., 2020. The convergence of heat, groundwater & fracture permeability: Innovative play fairway modelling applied to the Tularosa Basin. Final Project Report. Ruby Mountain Inc.

Blackwell, D.D., Negraru, P.T., Richards, M.C., 2006. Assessment of the enhanced geothermal system resource base of the United States. Natural Resource Research 15, 283–308. https://doi.org/10.1007/s11053-007-9028-7

Blackwell, D. D., Richards, M.C., Frone, Z.S., Batir, J.F., Williams, M.A., Ruza, A.A., Dingwall, R.K., 2011. SMU Geothermal Laboratory Heat Flow Map of the Coterminous United States.

Broadhead, R.F., 2002. Petroleum geology of the McGregor Range Otero County, New Mexico. AAPG Southwest Section Meeting, June, New Mexico, 40–55.

Caldwell, T.G., Bibby, H.M., Brown, C., 2004. The magnetotelluric phase tensor. Geophysical Journal International 158, 457–469. https://doi.org/10.1111/j.1365-246X.2004.02281.x

Cherkose, B.A., Mizunaga, H., 2018. Resistivity imaging of Aluto-Langano geothermal field using 3-D magnetotelluric inversion. Journal of African Earth Sciences 139, 307–318. https://doi.org/10.1016/j.jafrearsci.2017.12.017

Coolbaugh, M., Sladek, C., Zehner, R., Kratt, C., 2014. Shallow temperature surveys for geothermal exploration in the Great Basin, USA, and estimation of shallow aquifer heat loss, Geothermal Resource Council Transaction 38, 115–122.

Coppo, N., Baltassat, J.M., Girard, J.F., Wawrzyniak, P., Hautot, S., Tarits, P., Jacob, T., Martelet, G., Mathieu, F., Gadalia, A., Bouchot, V., Traineau, H., 2015. 3-D Magnetotelluric investigations for geothermal exploration in Martinique (Lesser Antilles). characteristic deep resistivity structures, and shallow resistivity distribution matching Heliborne TEM results. Proceedings World Geothermal Congress, April, Australia, 1–11.

Cumming, W., Mackie, R., 2007. 3D MT resistivity imaging for geothermal resource assessment and environmental mitigation at the Glass Mountain KGRA, California. Geothermal Resources Council Transaction 31, 331–334.

Elders, W.A., Moore, J.N., 2016, Geology of geothermal resources, *in* DiPippo, R., (Ed.), Geothermal Power Generation: Developments and Innovation, Woodhead Publishing, Woodhead publishing series in energy 97, 7–28.

Finger, J.T. and Jacobson, R.D., 1997. Fort Bliss exploratory slimholes: Drilling and testing No. SAND-97-3075. Sandia National Lab, Albuquerque, NM.

Fournier, R.O., 1977. Chemical geothermometers and mixing model for geothermal systems. Geothermics 5, 41–50.

Grimm, R., Nguyen, T., Persyn, S., Phillips, M., Stillman, D., Taylor, T., Delory, G., Turin, P., Espley, J., Gruesbeck, J., Sheppard, D., 2021. A magnetotelluric instrument for probing the interiors of Europa and other worlds. Advances in Space Research 68, 2022–2037. https://doi.org/10.1016/j.asr.2021.04.011

Hamm, S.G., Anderson, A., Blankenship, D., Boyd, L.W., Brown, E.A., Frone, Z., Hamos, I., Hughes, H.J., Kalmuk, M., Marble, A., McKittrick, A.M.W., Metcalfe, E., Morse, L.J., Nieto, A., Payne, J., Porse, S.L., Prisjatschew, A., Stutz, G.R., Tasca, C., Vandermeer, W., Watson, G., Weathers, M., Winick, J., 2021. Geothermal Energy R&D: An overview of the U.S. Department of Energy's geothermal technologies office. Journal of Energy Resources Technology 143, 100801. https://doi.org/10.1115/1.4049581

Han, Q., Kelbert, A., Hu, X., 2021. An electrical conductivity model of a coastal geothermal field in southeastern China based on 3D magnetotelluric imaging. Geophysics 86, B265–B276. https://doi.org/10.1190/geo2019-0446.1

Hurter, S., Schellschmidt, R., 2003. Atlas of geothermal resources in Europe. Geothermics, Selected Papers from the European Geothermal Conference 2003 32, 779–787. https://doi.org/10.1016/S0375-6505(03)00070-1

Iliev, F.L., Stanev, V.G., Vesselinov, V.V., Alexandrov, B.S., 2018. Nonnegative matrix factorization for identification of unknown number of sources emitting delayed signals. PLoS one 13, e0193974. https://doi.org/10.1371/journal.pone.0193974

James, E.D., Hoang, V.T., Epperson, I.J., 1987. Structure, permeability, and production characteristics of the Heber, California Geothermal Field, 12th Workshop on Geothermal Reservoir Engineering, January, California, 267–271.

Johnston, J.M., Pellerin, L., Hohmann, G.W., 1992. Evaluation of electromagnetic methods for geothermal reservoir detection. Geothermal Resource Council Transactions 16, 241–245.

Jolie, E., Scott, S., Faulds, J., Chambefort, I., Axelsson, G., Gutiérrez-Negrín, L., Regenspurg, S., Ziegler, M., Ayling, B., Richter, A., Zemedkun, M., 2021. Geological controls on geothermal resources for power generation. Nature Reviews Earth & Environment 2, 324–339. https://doi.org/10.1038/s43017-021-00154-y

Lee, D.D., Seung, H.S., 1999. Learning the parts of objects by non-negative matrix factorization. Nature 401, 788–791. https://doi.org/10.1038/44565

Miri, H., Dehkordi, B.H., Payrovian, G., 2021. Oil field imaging on the Sarab Anticline, southwest of Iran, using magnetotelluric data, Journal of Petroleum Science and Engineering 202, 108497. https://doi.org/10.1016/j.petrol.2021.108497

Muñoz, G., 2014. Exploring for geothermal resources with electromagnetic methods. Surveys in Geophysics 35, 101–122. https://doi.org/10.1007/s10712-013-9236-0

Nash, G.D., Bennett, C.R., 2015. Adaptation of a petroleum exploration tool to geothermal exploration: preliminary play fairway model of Tularosa Basin, New Mexico, and Texas, Geothermal Resource Council Transactions 8, 743-750.

O'Donnell, T.M., Jr., Miller, K.C., Witcher, J.C., 2001. A seismic and gravity study of the McGregor geothermal system, southern New Mexico. Geophysics 66, 1002–1014. https://doi.org/10.1190/1.1487048

Passey, Q.R., Dahlberg, K.E., Sullivan, K.B., Yin, H., Brackett, R.A., Xiao, Y.H., Guzmán-Garcia, A.G., 2006. Petrophysical evaluation of hydrocarbon pore-thickness in thinly bedded Clastic reservoirs. American Association of Petroleum Geologists, 3. https://doi.org/10.1306/A11157

Rousseeuw, P.J., 1987. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. Journal of Computational and Applied Mathematics 20, 53–65. https://doi.org/10.1016/0377-0427(87)90125-7

Sandeen W. M., Geology of the Tularosa Basin, New Mexico: New Mexico Geological Society in 5th Field Conference Guidebook, 81-88.

Seager, W.R., Morgan, P., 1979. Rio Grande rift in southern New Mexico, west Texas, and northern Chihuahua. Rio Grande rift: Tectonics and magmatism 14, 87-106.

Siler, D.L., Faulds, J.E., Hinz, N.H., Dering, G.M., Edwards, J.H., Mayhew, B., 2019. Three-dimensional geologic mapping to assess geothermal potential: examples from Nevada and Oregon. Geotherm Energy 7, 2, 1-32. https://doi.org/10.1186/s40517-018-0117-0

Sinno, Y.A., Daggett, P.H., Keller, G.R., Morgan, P., Harder, S.H., 1986. Crustal structure of the southern Rio Grande Rift determined from seismic refraction profiling. Journal of Geophysical Research: Solid Earth 91, 6143–6156. https://doi.org/10.1029/JB091iB06p06143

Tester, J.W., Beckers, K.F., Hawkins, A.J., Lukawski, M.Z., 2021. The evolving role of geothermal energy for decarbonizing the United States. Energy and Environmental Science 14, 6211–6241. https://doi.org/10.1039/D1EE02309H

U.S. Energy Information Administration (EIA), 2021. International energy outlook 2021 (IEO2021) reference case. May 30, 2022. https://www.eia.gov/outlooks/ieo/pdf/IEO2021_Narrative.pdf.

Ussher, G., Harvey, C., Johnstone, R., Anderson, E., 2000, Understanding the resistivities observed in geothermal systems, Proceedings of the World Geothermal Congress, May–June, Japan, 1915–1920.

Vesselinov, V., Ahmmed, B., Mudunuru, M., Karra, S., Middleton, R., 2021. Discovering the hidden geothermal signatures of southwest New Mexico, Proceedings World Geothermal Congress, April-May, Iceland, 1-10.

Vesselinov, V.V., Alexandrov, B.S., O'Malley, D., 2018. Contaminant source identification using semi-supervised machine learning. Journal of Contaminant Hydrology 212, 134–142. https://doi.org/10.1016/j.jconhyd.2017.11.002

Vesselinov, V.V., Mudunuru, M.K., Karra, S., O'Malley, D., Alexandrov, B.S., 2019. Unsupervised machine learning based on non-negative tensor factorization for analyzing reactive-mixing. Journal of Computational Physics 395, 85–104. https://doi.org/10.1016/j.jcp.2019.05.039

Vozoff, K., 1991. The magnetotelluric method, in: Nabighian, M.N. (Ed.), Electromagnetic Methods in Applied Geophysics: Volume 2, Application, Parts A and B, Society of Exploration Geophysicists. https://doi.org/10.1190/1.9781560802686.ch8

Wang, C., Zheng, M., Zhang, X., Xing, E., Zhang, J., Ren, J., Ling, Y., 2020. O, H, and Sr isotope evidence for origin and mixing processes of the Gudui geothermal system, Himalayas, China. Geoscience Frontiers 11, 1175–1187. https://doi.org/10.1016/j.gsf.2019.09.013

Yadav, K., Shah, M., Sircar, A., 2020. Application of magnetotelluric (MT) study for the identification of shallow and deep aquifers in Dholera geothermal region. Groundwater for Sustainable Development 11, 100472. https://doi.org/10.1016/j.gsd.2020.100472

Zehner, R.E., Tullar, K.N., Rutledge, E., 2012. Effectiveness of 2-meter and geoprobe shallow temperature surveys in early stage geothermal exploration. Geothermal Resource Council Transactions 36, 835–842.

Zhang, L., Hao, T., Xiao, Q., Wang, J., Zhou, L., Qi, M., Cui, X., Cai, N., 2015. Magnetotelluric investigation of the geothermal anomaly in Hailin, Mudanjiang, northeastern China. Journal of Applied Geophysics 118, 47–65. https://doi.org/10.1016/j.jappgeo.2015.04.006

CHAPTER FIVE

Conclusions

Key findings of this dissertation project include the following.

1) Decision tree-based machine learning models applied to core-calibrated petrophysical data predict facies, facies associations, and reservoir rock within the Late Devonian Duvernay Formation. Models for both reservoir rock (RESM) and facies associations (FAM) performed the best with accuracies of 88.1%. The model for depositional facies (FACM) had an accuracy of 60.3%.
2) For decision tree-based machine learning models to be useful for Duvernay Formation subsurface rock-type prediction from wireline logs, classes should be thicker than 3m (10 ft) and encompassing at least 16% of the machine learning training dataset. Exceptions to these cutoffs are attributed to diagnostic sedimentological features observed in core. Facies with ambiguous features achieve lower than expected results, whereas facies with more distinctive characteristics have higher than expected results.
3) The unsupervised machine learning model, NMFk, identifies 4 reoccurring petrophysical groups within the Duvernay Formation. When compared to deterministic petrofacies modelling, the NMFk groups resolve more detail within the Duvernay. Specifically, NMFk discriminates reservoir-defining petrophysical attributes that are difficult to recognize in core or require further core analyses

4) The objectivity of NMFk provides a more precise and consistent petrophysical interpretation that can be used to anticipate and document lithologic variability during core description and sampling. This minimizes the inconsistency and error that is inherent to subjective human core description and results in more consistent reservoir characterization.
5) NMFk is a useful analytical tool to assess prospective geothermal regions through the evaluation of variability in heat flow, porosity, permeability and temperature. When applied to the Tularosa Basin in south-central New Mexico, 4 signals were established that each have varying geothermal potential.
6) MT detects subsurface geothermal prospects by providing insight into relative porosity and associated permeability that influence subsurface resistivity trends detected in the MT inversion. Low resistivity (associated with porous strata saturated with conductive groundwater) is a defining characteristic of geothermal prospects and is interpreted to occur within the western portion of the McGregor Range in the southern part of the Tularosa Basin. MT is limited in that resistivity only provides insight into relative porosity and does not correlate with lithological or stratigraphic units.

BIBLIOGRAPHY

Ahmmed, B., S. Karra, V. V. Vesselinov, and M. K. Mudunuru, 2021, Machine learning to discover mineral trapping signatures due to CO2 injection: International Journal of Greenhouse Gas Control, **109**, 103382.

Ahmmed, B. and V. V. Vesselinov, 2021, Prospectivity Analyses of the Utah FORGE Site using unsupervised machine learning, Geothermal Resource Council Transactions, **45**.

Alberty, M. W., 1992, Basic open hole tools, in Development Geology Reference Manual, D. Morton-Thompson and A.M. Woods, eds., AAPG Exploration in Methods Series, **10**, 144-149.

Alexandrov, B. S. and V. V. Vesselinov, 2014, Blind source separation for groundwater pressure analysis based on nonnegative matrix factorization. Water Resource. Research, **50**, 7332–7347, 10.1002/2013WR015037

Allan, J., and S. Creaney, 1991, Oil Families of the Western Canada Basin: Bulletin of Canadian Petroleum Geology, **39**, 107–122.

Arnason, K., R. Karlsdottir, H. Eysteinsson, O. G. Flovenz, and S. T. Gudlaugsson, 2000, The resistivity structure of high-temperature geothermal systems in Iceland. Proceedings of the World Geothermal Congress, May–June, Japan, 923–928.

Atchley, S. C., N. H. Ball, and L. E. Hunt, 2010, Reservoir characterization and facies prediction within the Late Cretaceous Doe Creek Member, Valhalla Field, west-central Alberta, Canada. American Association of Petroleum Geologists Bulletin, **94**, 1-25. 10.1306/10261817272.

Atchley, S. C., S. C. James, H. Harlow, C. Leslie, J. Jiang, G. Carrelli, W. Bindon, and J. Beal, 2018, Geologic and resource assessment of the Upper Devonian Grosmont and Upper Ireton Formations, central Grosmont shelf complex, Alberta, Canada: AAPG Bulletin, **102**, 731–759, 10.1306/0711171612816514.

Atef, H., A. Abd El Gawad, M. Abdel Zaher, and K. Farag, 2016, The contribution of gravity method in geothermal exploration of southern part of the Gulf of Suez–Sinai region, Egypt. NRIAG Journal of Astronomy and Geophysics, **5**, 173–185, 10.1016/j.nrjag.2016.02.005

Augustine, C., 2014, Analysis of sedimentary geothermal systems using an analytical reservoir model, Geothermal Resource Council Transactions, **38**, 641–648.

Baldwin, J. L., R. M. Bateman, and C. L. Wheatley, 1990, Application of a neural network to the problem of mineral identification form well logs: The Log Analyst, **31**, 279-293.

Barbier, E., 2002, Geothermal energy technology and current status: an overview. Renewable and Sustainable Energy Reviews, **6**, 3–65, 10.1016/S1364-0321(02)00002-3

Barker, B., J. Moore, M. Segall, G. Nash, S. Simmons, C. Jones, J. Lear, and C. Bennett, 2014, Exploration drilling and technology demonstration at Fort Bliss. 39th Workshop on Geothermal Reservoir Engineering, February, California, 1–12.

Bauman, P., 2018, East Duvernay Shale Basin, https://chinookpetroleum.com/devonian-duvernay-formation-east-shale-basin-alberta/, accessed January 19, 2021.

Bennett, C.R., G. D. Nash, and B. Barker, 2020, The convergence of heat, groundwater & fracture permeability: Innovative play fairway modelling applied to the Tularosa Basin. Final Project Report. Ruby Mountain Inc.

Bestagini, P., V. Lipari, and S. Tubaro, 2017, A machine learning approach to facies classification using well logs: Society of Exploration Geophysicists Technical Program, SEG, Expanded Abstracts, 2137–2142.

Bhattacharya S., T. R. Carr, and M. Pal, 2016, Comparison of supervised and unsupervised approaches for mudstone lithofacies classification: Case studies from the Bakken and Mahantango-Marcellus Shale, USA: Journal of National Gas Science and Engineering, **33**, p. 1119-1133, 10.1016/j.jngse.2016.04.055.

Bhattacharya, S., and T. R. Carr, 2019, Integrated data-driven 3D shale lithofacies modeling of the Bakken Formation in the Williston basin, North Dakota, United States: Journal of Petroleum Science and Engineering, **177**, 1072-1086, doi: 10.1016/j.petrol.2019.02.036.

Blackwell, D. D., M. C. Richards, Z. S. Frone, J. F. Batir, M. A. Williams, A. A. Ruza, and R. K. Dingwall, 2011. SMU Geothermal Laboratory Heat Flow Map of the Coterminous United States.

Blackwell, D. D., P. T. Negraru, and M. C. Richards, M.C., 2006, Assessment of the enhanced geothermal system resource base of the United States. Natural Resource Research, **15**, 283–308, 10.1007/s11053-007-9028-7

Blakey R. C., 2011, Paleozoic Maps: Late Devonian – 375 Ma (380-370):http://deeptimemaps.com/north-american-key-time-slices-series/, accessed December 1, 2017.

Bond, L. J., K. M. Denslow, J. W. Griffin, G. E. Dale, R. V. Harris, T. L. Moran, D. M. Sheen, and T. Schenkel, 2010, Evaluation of Non-Nuclear Techniques for Well Logging: Technology Evaluation, PNNL-19867: US Department of Energy,10.2172/1006309.

Breiman, L., J. Friedman, R. Olshen, and C. Stone, 1984, Classification and regression trees: Wadsworth and Brooks.

Bressan, T. S., M. K. de Souza, J. T. Girelli, and F. C. Junior, 2020, Evaluation of machine learning methods for lithology classification using geophysical data: Computers and Geosciences, **139**, 104475, 10.1016/j.cageo.2020.104475.

Broadhead, R.F., 2002, Petroleum geology of the McGregor Range Otero County, New Mexico. AAPG Southwest Section Meeting, June, New Mexico, 40–55.

Cahyana, N., S. Khomsah, and A. S. Aribowo, 2019, Improving Imbalanced Dataset Classification Using Oversampling and Gradient Boosting: 5th International Conference on Science in Information Technology (ICSITech), 217–222, 10.1109/ICSITech46713.2019.8987499.

Caldwell, T. G., H. M. Bibby, and C. Brown, 2004, The magnetotelluric phase tensor. Geophysical Journal International, **158**, 457–469, 10.1111/j.1365-246X.2004.02281.x.

Chawla, N. V., K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, 2002, SMOTE: Synthetic Minority Over-sampling Technique: Journal of Artificial Intelligence Research, **16**, 321–357, 10.1613/jair.953.

Cherkose, B. A., H. Mizunaga, 2018, Resistivity imaging of Aluto-Langano geothermal field using 3-D magnetotelluric inversion. Journal of African Earth Sciences, **139**, 307–318, 10.1016/j.jafrearsci.2017.12.017.

Chopra, S., R. K. Sharma, A. K. Ray, H. Nemati, R. Morin, B. Schulte, and D. D'Amico, 2017, Seismic reservoir characterization of Duvernay shale with quantitative interpretation and induced seismicity considerations — A case study: Interpretation, **5**, 185–197, 10.1190/INT-2016-0130.1.

Coolbaugh, M., C. Sladek, R. Zehner, and C. Kratt, 2014, Shallow temperature surveys for geothermal exploration in the Great Basin, USA, and estimation of shallow aquifer heat loss, Geothermal Resource Council Transaction **38**, 115–122.

Coppo, N., J. M. Baltassat, J. F. Girard, P. Wawrzyniak, S. Hautot, P. Tarits, T. Jacob, G. Martelet, F. Mathieu, A. Gadalia, V. Bouchot, and H. Traineau., 2015, 3-D Magnetotelluric investigations for geothermal exploration in Martinique (Lesser Antilles). characteristic deep resistivity structures, and shallow resistivity distribution matching Heliborne TEM results. Proceedings World Geothermal Congress, April, Australia, 1–11.

Crain, E. R., 2018, Natural gamma ray and spectral gamma ray logs: Crain's Petrophysical Handbook,https://www.spec2000.net/07-grlog.htm, accessed July 12, 2018.

Cumming, W. and R. Mackie, 2007, 3D MT resistivity imaging for geothermal resource assessment and environmental mitigation at the Glass Mountain KGRA, California. Geothermal Resources Council Transaction, **31**, 331–334.

Datta, A., 2018, The Resurgence of Interest in the Duvernay: A Review of Pertinent Economic and Environmental Issues: Canadian Energy Research Institute.

Diniz-Ferreira, E. L., and C. Torres-Verdín, 2012, Improved estimation of pore connectivity and permeability in deepwater carbonates with the construction of multi-layer static and dynamic petrophysical models: Society of Petrophysicists and Well Log Analysts 53rd Annual Logging Symposium.

Dong, T., N. B. Harris, J. M. McMillan, C. E. Twemlow, B. R. Nassichuk, and D. L. Bish, 2019, A model for porosity evolution in shale reservoirs: An example from the Upper Devonian Duvernay Formation, Western Canada Sedimentary Basin: AAPG Bulletin, **103**, 1017–1044, 10.1306/10261817272.

Dubois, M. K., G. C. Bohling, and S. Chakrabarti, 2007, Comparison of four approaches to a rock facies classification problem: Computers & Geosciences, **33**, 599–617, 10.1016/j.cageo.2006.08.011.

Dunn, L., G. Schmidt, K. Hammermaster, M. Brown, R. Bernard, E. Wen, R. Befus, and S. Gardiner, 2016, The Duvernay Formation (Devonian): Sedimentology and Reservoir Characterization of a Shale Gas/Liquids Play in Alberta, Canada: Annual Convention and Exhibition, AAPG, Extended Abstract.

Elders, W. A., and J. N. Moore, 2016, Geology of geothermal resources, in DiPippo, R., (Ed.), Geothermal Power Generation: Developments and Innovation, Woodhead Publishing, Woodhead publishing series in energy, **97**, 7–28.

Feng, R., N. Balling, and D. Grana, 2020, Lithofacies classification of a geothermal reservoir in Denmark and its facies-dependent porosity estimation from seismic inversion: Geothermics, **87**, 101854, 10.1016/j.geothermics.2020.101854.

Finger, J. T. and R. D. Jacobson, 1997, Fort Bliss exploratory slimholes: Drilling and testing No. SAND-97-3075. Sandia National Lab, Albuquerque, NM.

Fournier, R. O., 1977. Chemical geothermometers and mixing model for geothermal systems. Geothermics, **5**, 41–50.

Geurts, P., D. Ernst, and L. Wehenkel, 2006, Extremely randomized trees: Machine Learning, **63**, 3–42, 10.1007/s10994-006-6226-1.

Gosain, A., and S. Sardana, 2017, Handling class imbalance problem using oversampling techniques: A review: International Conference on Advances in Computing, Communications and Informatics, 79–85, 10.1109/ICACCI.2017.8125820.

Grimm, R., T. Nguyen, S. Persyn, M. Phillips, D. Stillman, T. Taylor, G. Delory, P. Turin, J. Espley, J. Gruesbeck, and D. Sheppard, 2021, A magnetotelluric instrument for probing the interiors of Europa and other worlds. Advances in Space Research, **68**, 2022–2037, 10.1016/j.asr.2021.04.011.

Guo, X., D. Hu, Y. Li, Z. Wei, X. Wei, and Z. Liu, 2017, Geological factors controlling shale gas enrichment and high production in Fuling shale gas field: Petroleum Exploration and Development, **44**, 513–523, 10.1016/S1876-3804(17)30060-5.

Hall, B., 2016, Facies classification using machine learning: The Leading Edge, **35**, 906–909, 10.1190/tle35100906.1.

Hall, M., and B. Hall, 2017, Distributed collaborative prediction: Results of the machine learning contest: The Leading Edge, **36**, 267–269, 10.1190/tle36030267.1.

Halotel, J., V. Demyanov, and A. Gardiner, 2020, Value of geologically derived features in machine learning facies classification: Mathematical Geosciences, **52**, 5–29, 10.1007/s11004-019-09838-0.

Hamm, S. G., A. Anderson, D. Blankenship, L. W. Boyd, E. A. Brown, Z. Frone, I. Hamos, H. J. Hughes, M. Kalmuk, A. Marble, A. M. W. McKittrick, E. Metcalfe, L. J. Morse, A. Nieto, J. Payne, S. L. Porse, A. Prisjatschew, G. R. Stutz, C. Tasca, W. Vandermeer, G. Watson, M. Weathers, and J. Winick, 2021, Geothermal Energy R&D: An overview of the U.S. Department of Energy's geothermal technologies office. Journal of Energy Resources Technology, **143**, 100801, 10.1115/1.4049581.

Han, Q., A. Kelbert, and X. Hu, 2021, An electrical conductivity model of a coastal geothermal field in southeastern China based on 3D magnetotelluric imaging. Geophysics, **86**, B265–B276, 10.1190/geo2019-0446.1.

Harris, N. B., J. M. McMillan, L. J. Knapp, and M. Mastalerz, 2018, Organic matter accumulation in the Upper Devonian Duvernay Formation, Western Canada Sedimentary Basin, from sequence stratigraphic analysis and geochemical proxies: Sedimentary Geology, **376**,185–203, 10.1016/j.sedgeo.2018.09.004.

He, H. and E. A. Garcia, 2009, Learning from imbalance data: IEEE Transactions on knowledge and data engineering, **21**, 1263–1284.

He, H., Y. Bai, E. A. Garcia, and S. Li, 2008, ADASYN: Adaptive synthetic sampling approach for imbalanced learning: Institute of Electrical and Electronics Engineers International Joint Conference on Neural Networks, IEEE World Congress on Computational Intelligence, 1322–1328, 10.1109/IJCNN.2008.4633969.

Ho, T. K., 1998, The random subspace method for constructing decision forests: Institute of Electrical and Electronics Engineers Transactions on Pattern Analysis and Machine Intelligence, **20**, 832–844, doi:10.1109/34.709601.

Hossain, T. M., J. Watada, Z. Jian,, H. Sakai, S. Rahman, and I. A. Aziz, 2020, Missing well log data handling in complex lithology perdition: an NIS apriori algorithm approach: International Journal of Innovative Computing, Information and Control, **16**, 1077-1091, 10.24507/ijicic.16.03.1077.

Hou, L., Z. Yu, X. Luo, S. Lin, Z. Zhao, Z. Yang, S. Wu, J. Cui, and L. Zhang, 2021, Key geological factors controlling the estimated ultimate recovery of shale oil and gas: A case study of the Eagle Ford Shale, Gulf Coast Basin, USA: Petroleum Exploration and Development, **48**, 762–774, 10.1016/S1876-3804(21)60062-9.

Hurter, S. and R. Schellschmidt, 2003, Atlas of geothermal resources in Europe. Geothermics, Selected Papers from the European Geothermal Conference 2003, **32**, 779–787, 10.1016/S0375-6505(03)00070-1.

Hussein, M., R. R. Stewart, D. Sacrey, J. Wu, and R. Athale, 2021, Unsupervised machine learning using 3D seismic data applied to reservoir evaluation and rock type identification: Interpretation, **9**, T549–T568, 10.1190/INT-2020-0108.1.

Iliev, F. L., V. G. Stanev, V. V. Vesselinov, and B. S. Alexandrov, 2018, Nonnegative matrix factorization for identification of unknown number of sources emitting delayed signals: PLOS ONE, **13**, e0193974, 10.1371/journal.pone.0193974.

Imamverdiyev, Y., and L. Sukhostat, 2019, Lithological facies classification using deep convolutional neural network: Journal of Petroleum Science and Engineering, **174**, 216–228, 10.1016/j.petrol.2018.11.023.

Ippolito, M., J. Ferguson, and F. Jenson, 2021, Improving facies prediction by combining supervised and unsupervised learning methods: Journal of Petroleum Science and Engineering, **200**, 108300, 10.1007/s10708-021-10502-6.

James, E.D., V. T. Hoang, and I. J. Epperson, 1987, Structure, permeability, and production characteristics of the Heber, California Geothermal Field, 12th Workshop on Geothermal Reservoir Engineering, January, California, 267–271.

Jian, H., L. Chenghui, C. Zhimin, and M. Haiwei, 2020, Integration of deep neural networks and ensemble learning machines for missing well logs estimation: Flow Measurement and Instrumentation, 101748, 10.1016/j.flowmeasinst.2020.101748.

Johnston, J. M., L. Pellerin, and G. W. Hohmann, 1992, Evaluation of electromagnetic methods for geothermal reservoir detection: Geothermal Resource Council Transactions, **16**, 241–245.

Jones, C. I., 2001, Was an Industrial Revolution Inevitable? Economic Growth Over the Very Long Run.: The BE Journal of Macroeconomics, **1**.

Jolie, E., S. Scott, J. Faulds, I. Chambefort, G. Axelsson, L. Gutiérrez-Negrín, S. Regenspurg, M/ Ziegler, B. Ayling, A. Richter, and M. Zemedkun, 2021. Geological controls on geothermal resources for power generation: Nature Reviews Earth & Environment, **2**, 324–339, 10.1038/s43017-021-00154-y

Knapp, L. J., N. B. Harris, and J. M. McMillan, 2019, A sequence stratigraphic model for the oranic-rich Upper Devonian Duvernay Formation, Alberta, Canada: Sedimentary Geology, **387**, 152–181, 10.1016/j.sedgeo.2019.04.008

Kotsiantis, S. B., 2007, Supervised Machine Learning: A Review of classification techniques: Emerging artificial intelligence applications in computer engineering, **160**, 3–24.

Kotsiantis, S. B., 2013, Decision trees: a recent overview: Artificial Intelligence Review, **39**, 261–283, 10.1007/s10462-011-9272-4.

Krawczyk, B., 2016, Learning from imbalanced data: open challenges and future directions: Progress in Artificial Intelligence, **5**, 221–232, 10.1007/s13748-016-0094-0.

Lee, D. D. and H. S. Seung, 1999, Learning the parts of objects by non-negative matrix factorization. Nature, **401**, 788–791, 10.1038/44565.

Li, G., K. Luo, and D. Shi, 2020, Key technologies, engineering management and important suggestions of shale oil/gas development: Case study of a Duvernay shale project in Western Canada Sedimentary Basin: Petroleum Exploration and Development, **47**, 791–802,10.1016/S1876-3804(20)60094-5.

Loh, W. Y., 2011, Classification and regression trees: Wiley interdisciplinary reviews: data mining and knowledge discovery, **1**, 14–23, 10.1002/widm.8.

Luo, C., H. Lin, Y. Peng, H. Qu, X. Huang, N. Yin, W. Liu, and X. Gao, 2021, Factors controlling shale reservoirs and development potential evaluation: a case study: Geofluids, **2021**, 1–13, 10.1155/2021/6661119.

Meyer, B. L., and M. H. Nederlof, 1984, Identification of source rocks on wireline logs by density/resistivity and sonic transit time/resistivity crossplots: AAPG Bulletin, **68**, 121-129.

Miri, H., B. H. Dehkordi, and G. Payrovian, 2021, Oil field imaging on the Sarab Anticline, southwest of Iran, using magnetotelluric data, Journal of Petroleum Science and Engineering, **202**, 108497,10.1016/j.petrol.2021.108497.

Muñoz, G., 2014, Exploring for geothermal resources with electromagnetic methods. Surveys in Geophysics, **35**, 101–122,10.1007/s10712-013-9236-0.

Nash, G. D. and C. R. Bennett, 2015, Adaptation of a petroleum exploration tool to geothermal exploration: preliminary play fairway model of Tularosa Basin, New Mexico, and Texas, Geothermal Resource Council Transactions, **8**, 743-750.

O'Donnell, T. M., Jr., K. C. Miller, and J. C. Witcher, 2001. A seismic and gravity study of the McGregor geothermal system, southern New Mexico. Geophysics, **66**, 1002–1014, 10.1190/1.1487048.

Olson, R. S., and J. H. Moore, 2016, TPOT: A Tree-based pipeline optimization tool for automating machine learning: Proceedings of the Workshop on Automatic Machine Learning, **64**, 66–74.

Passey, Q. R., K. E. Dahlberg, K. B. Sullivan, H. Yin, R. A. Brackett, Y. H. Xiao, and A. G. Guzmán-Garcia, 2006, Petrophysical evaluation of hydrocarbon pore-thickness in thinly bedded clastic reservoirs: AAPG Archie Series, **1**,10.1306/A11157.

Porter, J. W., R. A. Price, and R. G. McCrossan, 1982, The Western Canada Sedimentary Basin: Philosophical Transactions of the Royal Society of London. Series A: Mathematical and Physical Sciences, **305**, 169–192.

Preston, A., G. Garner, K. Beavis, O. Sadiq, and S. Stricker, 2016, Duvernay Reserves and Resources Report: A Comprehensive Analysis of Alberta's Foremost Liquids-Rich Shale Resource: Alberta Energy Regulator, Calgary, Alberta.

Quinlan, J. R., 1990, Decision trees and decision-making: Institute of Electrical and Electronics Engineers Transactions on Systems, Man, and Cybernetics, **20**, 339–346, 10.1109/21.52545.

Rahman, M. M., R. Nepal, and K. Alam, 2021. Impacts of human capital, exports, economic growth and energy consumption on CO_2 emissions of a cross-sectionally dependent panel: Evidence from the newly industrialized countries (NICs): Environmental Science and Policy, **121**, 24-36.

Ripley, B. D., 1996, Pattern Recognition and Neural Networks: Cambridge University Press, 10.1017/CBO9780511812651.

Rogers, S. J., J. H. Fang, C. L. Karr, and D. A. Stanley, 1992, Determination of lithology from well logs using a neural network: AAPG Bulletin, **76**, 731-739.

Rokosh, C. D., S. Lyster, S. D. A. Anderson, A. P. Beaton, H. Berhane, T. Brazzoni, D. Chen, Y. Cheng, T. Mack, C. Pana and J. G. Pawlowicz, 2012, Open file report: Summary of Alberta's shale- and siltstone-hosted hydrocarbon resource potential: Alberta Geological Survey.

Rousseeuw, P. J., 1987, Silhouettes: A graphical aid to the interpretation and validation of cluster analysis: Journal of Computational and Applied Mathematics, **20**, 10.1016/0377-0427(87)90125-7.

Sandeen W. M., 1954, Geology of the Tularosa Basin, New Mexico: New Mexico Geological Society in 5th Field Conference Guidebook, 81-88.

Saneifar, M., A. Aranibar, and Z. Heidari, 2014, Rock classification in the Haynesville Shale based on petrophysical and elastic properties estimated from well logs: Interpretations, **3**, 65- 75, doi:10.1190/INT-2013-0198.1.

Seager, W.R. and P. Morgan, 1979, Rio Grande rift in southern New Mexico, west Texas, and northern Chihuahua. Rio Grande rift: Tectonics and magmatism, **14**, 87-106.

Siler, D. L., J. D. Pepin, V. V. Vesselinov, M. K. Mudunuru, and B. Ahmmed, 2021, Machine learning to identify geologic factors associated with production in geothermal fields: a case-study using 3D geologic data, Brady geothermal field, Nevada: Geothermal Energy, **9**, 1-17, 10.1186/s40517-021-00199-8.

Siler, D. L., J. E. Faulds, N. H. Hinz, G. M. Dering, J. H. Edwards, B. Mayhew, 2019, Three-dimensional geologic mapping to assess geothermal potential: examples from Nevada and Oregon. Geotherm Energy, **7**, 1-32, 10.1186/s40517-018-0117-0.

Sinno, Y.A., P. H. Daggett, G. R. Keller, P. Morgan, and S. H. Harder, 1986, Crustal structure of the southern Rio Grande Rift determined from seismic refraction profiling. Journal of Geophysical Research: Solid Earth, **91**, 6143–6156, 10.1029/JB091iB06p06143.

Stoakes, F. A., 1980, Nature and control of shale basin fill and its effect on reef growth and termination: Upper Devonian Duvernay and Ireton Formations of Alberta, Canada: Bulletin of Canadian Petroleum Geology, **28**, 345–410.

Stoakes, F. A., and S. Creaney, 1984, Sedimentology of a carbonate source rock: the Duvernay Formation of Alberta Canada: Carbonates in Subsurface and Outcrop: Canadian Society of Petroleum Geologists, 132-147.

Sun, J., Q. Li, M. Chen, L. Ren, G. Huang, C. Li, and Z. Zhang, 2019, Optimization of models for a rapid identification of lithology while drilling – A win-win strategy based on machine learning: Journal of Petroleum Science and Engineering, **176**, 321-341, 10.3390/en13040888.

Switzer, S. B., W. G. Holland, D. S. Christie, G. C. Graf, A. S. Hedinger, R. J. McAuley, R. A. Wierzbicki, and J. J. Packard, 1994, Devonian Woodbend - Winterburn Strata of the Western Canada Sedimentary Basin: Geological Atlas of the Sedimentary Basin, 165–202.

Tester, J.W., Beckers, K.F., Hawkins, A.J., Lukawski, M.Z., 2021. The evolving role of geothermal energy for decarbonizing the United States. Energy and Environmental Science, **14**, 6211–6241, 10.1039/D1EE02309H.

Tewari, S., and U. Dwivedi, 2019, Ensemble-based big data analytics of lithofacies for automatic development of petroleum reservoirs: Computers and Industrial Engineering, **128**, 937-947, 10.1016/j.cie.2018.08.018.

Thorson, A. M., 2019, Stratigraphic Partitioning and Distribution of Reservoir Attributes within the Late Devonian Duvernay Formation, Western Canada Sedimentary Basin: M.S. thesis, Baylor University.

U.S. Energy Information Administration (EIA), 2021. International energy outlook 2021 (IEO2021) reference case. May 30, 2022. https://www.eia.gov/outlooks/ieo/pdf/IEO2021_Narrative.pdf.

Ussher, G., C. Harvey, R. Johnstone, and E. Anderson, 2000, Understanding the resistivities observed in geothermal systems, Proceedings of the World Geothermal Congress, May–June, Japan, 1915–1920.

Vesselinov, V. V., B. S. Alexandrov, and D. O'Malley, 2018, Contaminant source identification using semi-supervised machine learning: Journal of Contaminant Hydrology, **212**,134–142, 10.1016/j.jconhyd.2017.11.002.

Vesselinov, V. V., M. K. Mudunuru, S. Karra, D. O'Malley, and B. S. Alexandrov, 2019, Unsupervised machine learning based on non-negative tensor factorization for analyzing reactive-mixing: Journal of Computational Physics, **395**, 85–104, 10.1016/j.jcp.2019.05.039.

Vesselinov, V. V., B. Ahmmed, M. Mudunuru, S. Karra, and R. Middleton, 2021, Discovering the hidden geothermal signatures of southwest New Mexico, Proceedings World Geothermal Congress, April-May, Iceland, 1-10.

Vozoff, K., 1991. The magnetotelluric method, in: Nabighian, M.N. (Ed.), Electromagnetic Methods in Applied Geophysics: Volume 2, Application, Parts A and B, Society of Exploration Geophysicists, 10.1190/1.9781560802686.ch8.

Wang G., and T. R. Carr, 2013, Organic-rich Marcellus Shale lithofacies modeling and distribution pattern analysis in the Appalachian Basin: AAPG Bulletin, **97**, 2173 – 2205, doi:10.1306/05141312135.

Wang, C., M. Zheng, X. Zhang, E. Xing, J. Zhang, J. Ren, Y. Ling, 2020, O, H, and Sr isotope evidence for origin and mixing processes of the Gudui geothermal system, Himalayas, China. Geoscience Frontiers, **11**, 1175–1187, 10.1016/j.gsf.2019.09.013.

Weissenberger, J. A. W., and K. Potma, 2001, The Devonian of western Canada–aspects of a petroleum system: Introduction: Bulletin of Canadian Petroleum Geology, **49**, 1–6.

Weissenberger, J. A. W., P. K. Wong, and M. G. Gilhooly, 2016, Stratigraphic architecture of the Frasnian South Jasper Basin, North-Central Alberta front ranges: New advances in Devonian carbonates: Outcrop Analogs, Reservoirs, and Chronostratigraphy, SEPM Special Publication, **107**, 10.2110/sepmsp.107.07.

Wendte, J.C. and F. A. Stoakes, 1982, Evolution and corresponding porosity distribution of the Judy Creek reef complex, Upper Devonian, central Alberta, in W.G. Cutlr, ed., Canada's Giant Hydrocarbon Reservoirs, Core Conference Manual Canadian Society of Petroleum Geologists, 63–81.

Witzke, B. J., and P. H. Heckel, 1988, Paleoclimatic Indicators and Inferred Devonian Paleolatitudes of Euroamerica: Canadian Society of Petroleum Geologists, **14**, 49–63.

Wong, P. K., J. A. W. Weissenberger, and M. G. Gilhooly, 2016, Revised Regional Frasnian Sequence Stratigraphic Framework, Alberta Outcrop and Subsurface: New advances in Devonian carbonates: Outcrop analogs, reservoirs, and chronostratigraphy, SEPM Special Publication, **107**, 37–85, 10.2110/sepmsp.107.05.

Wu, G., G. Chen, Q. Cheng, Z. Zhang, and J. Yang, 2021, Unsupervised machine learning for lithological mapping using geochemical data in covered areas of Jining, China: Natural Resources Research, **30**, 1053–1068, 10.1007/s11053-020-09788-z.

Yadav, K., M. Shah, and A. Sircar, 2020, Application of magnetotelluric (MT) study for the identification of shallow and deep aquifers in Dholera geothermal region. Groundwater for Sustainable Development, **11**, 100472. 10.1016/j.gsd.2020.100472.

Zehner, R. E., K. N. Tullar, and E. Rutledge, 2012, Effectiveness of 2-meter and geoprobe shallow temperature surveys in early stage geothermal exploration. Geothermal Resource Council Transactions, **36**, 835–842.

Zhang, D., T. Chen, and J. Meng, 2018, Synthetic well logs generation via Recurrent Neural Network: Petroleum Exploration and Development, **45**, 629 – 639.

Zhang, L., T. Hao, Q. Xiao, J. Wang, L. Zhou, M. Qi, X. Cui, and N. Cai, 2015, Magnetotelluric investigation of the geothermal anomaly in Hailin, Mudanjiang, northeastern China. Journal of Applied Geophysics, **118**, 47–65, 10.1016/j.jappgeo.2015.04.0