

2016

Shale lithofacies modeling of the Bakken Formation in the Williston basin, North Dakota

Shuvajit Bhattacharya

Follow this and additional works at: <https://researchrepository.wvu.edu/etd>

Recommended Citation

Bhattacharya, Shuvajit, "Shale lithofacies modeling of the Bakken Formation in the Williston basin, North Dakota" (2016). *Graduate Theses, Dissertations, and Problem Reports*. 5207.
<https://researchrepository.wvu.edu/etd/5207>

This Dissertation is protected by copyright and/or related rights. It has been brought to you by the The Research Repository @ WVU with permission from the rights-holder(s). You are free to use this Dissertation in any way that is permitted by the copyright and related rights legislation that applies to your use. For other uses you must obtain permission from the rights-holder(s) directly, unless additional rights are indicated by a Creative Commons license in the record and/ or on the work itself. This Dissertation has been accepted for inclusion in WVU Graduate Theses, Dissertations, and Problem Reports collection by an authorized administrator of The Research Repository @ WVU. For more information, please contact researchrepository@mail.wvu.edu.

**SHALE LITHOFACIES MODELING OF THE BAKKEN FORMATION IN THE WILLISTON BASIN,
NORTH DAKOTA**

Shuvajit Bhattacharya

**Dissertation submitted to the
Eberly College of Arts and Sciences
at West Virginia University
in partial fulfillment of the requirements
for the degree of**

**Doctor of Philosophy
in
Geology**

**Timothy R. Carr, Ph.D., Chair
Thomas Wilson, Ph.D.
Dengliang Gao, Ph.D.
Amy Weislogel, Ph.D.
Ebrahim Fathi, Ph.D.**

Department of Geology & Geography

Morgantown, West Virginia

2016

**Keywords: 3D Geological Model; Bakken Formation; Machine Learning; Petrophysics; Shale
Lithofacies; Williston Basin**

Copyright 2016 Shuvajit Bhattacharya

ABSTRACT

SHALE LITHOFACIES MODELING OF THE BAKKEN FORMATION IN THE WILLISTON BASIN, NORTH DAKOTA

Shuvajit Bhattacharya

The Bakken petroleum system (Devonian-Mississippian) in the Williston basin of North Dakota and Montana in the United States, and Saskatchewan and Manitoba in Canada is one of the largest unconventional oil plays in North America. The Bakken Formation consists of three members: upper, middle, and lower. Both upper and lower members are shale (source rocks), whereas the middle member (reservoir rock) is composed of mixed lithologies, including sandstone, dolostone, and limestone. Underlying the lower Bakken shale member, the Three Forks Formation is another target for hydrocarbon exploration.

Although the middle Bakken member along with the Three Forks Formation have been the targets for horizontal drilling and hydraulic stimulation throughout the basin, several uncertainties remain, including facies variation due to depositional and diagenetic controls on mineral composition and organic matter content in the Bakken shale members, which could play a significant role in hydrocarbon generation and production. Although the Bakken shale members may look homogeneous in the appearance, they are significantly heterogeneous and complex mixture of quartz, smectite, illite, carbonate, pyrite, and kerogen in varying proportions. Improved characterization of the Bakken shale lithofacies is important to better understand depositional environment, lithofacies distribution, and their potential influence on hydrocarbon production.

The main objective of this work is to investigate vertical and lateral heterogeneities of the Bakken shale lithofacies, based on mineralogy and organic matter richness. Secondly, if the Bakken shale members are composed of different lithofacies, can they be associated with different depositional and/or diagenetic conditions, which could influence source, transportation, and preservation of organic matter and sediment in the Williston basin.

Core data (such as X-ray diffraction, X-ray fluorescence, and Total Organic Carbon content), conventional borehole geophysical logs (such as gamma, resistivity, bulk density, neutron porosity, and photo-electric factor), and advanced petrophysical logs (such as Spectral Gamma and Pulsed Neutron Spectroscopy) are used and integrated together to classify the Bakken shale lithofacies and build models of lithofacies distribution at multiple scales. Usually there are minimal core data, scattered advanced well logs, and ubiquitous conventional well log suites in a petroliferous basin, which hinders lithofacies analysis and petrophysical modeling. Therefore, a significant effort of this work is geared towards developing and applying cost-effective mathematical algorithms (such as Support Vector Machine and Artificial Neural Network etc.) and geostatistical techniques (such as Sequential Indicator Simulation) to classify, predict, and

interpolate shale lithofacies with high accuracy, using conventional well log-derived petrophysical parameters from several wells.

The results show that both upper and lower Bakken shale members are vertically and laterally heterogeneous at core, well, and regional scales. Bakken shale members can be classified as five different lithofacies, in terms of mineralogy and organic matter content. Organic-rich shale lithofacies are more dominant than organic-poor shale lithofacies. It appears several factors (such as source of minerals, paleo-redox conditions, organic matter productivity, and preservation etc.) controlled the Bakken shale lithofacies distribution pattern. Silica in the Organic Siliceous Shale (OSS) lithofacies near the basin center is hypothesized to be related to the presence of biogenic silica (e.g. radiolaria), whereas the portion of OSS lithofacies near the basin margin is believed to be associated with eolian action. High organic matter content in the Organic Mudstone (OMD) lithofacies near the basin margin could be interpreted due to the presence of algal matter. The borehole geophysical, petrophysical approaches, and the 3D lithofacies modeling techniques developed in this study can be applied to detailed studies of complex shale formations and exploration of hydrocarbon resources worldwide.

ACKNOWLEDGEMENTS

I completed the dissertation with the generous help and support of many organizations, mentors, and colleagues. U.S. Department of Energy, West Virginia University, American Association of Petroleum Geologists, and Society of Petrophysicists & Well Log Analysts provided support for this research. North Dakota Geological Survey, PDC Energy, and Newfield Exploration provided required core and well log data for the thesis. I thank CGG Jason (PowerLog™), IHS (Petra™), Paradigm (Geolog™), Schlumberger (Petrel™), WEKA, and Intelligent Solutions Inc. (IDEA) for providing necessary software packages for this work.

I would like to express my greatest gratitude to Dr. Timothy Carr, my advisor and mentor. Your academic knowledge and industrial experience helped me a lot to understand the potential of a value-added research in academia. I am immensely grateful to you for your patience, answering my countless “one more question”s. Thank you again for mentoring, and being a great friend of mine.

Special thanks go to all my committee members, including Dr. Thomas Wilson, Dr. Dengliang Gao, Dr. Amy Weislogel, and Dr. Ebrahim Fathi for their valuable assistance in my dissertation. I would like to thank Dr. Harvey Eastman (URS Corporation) for his help in analytical measurements. I really appreciate the invaluable help of Mr. Fred Jenson (CGG Jason) for instilling in me the passion for petrophysical and borehole geophysical modeling, and the sheer joy out of it.

I would like to thank all other faculty members and staff at Department of Geology & Geography for their genuine help and support. Thanks to all friends and family members for your well wishes throughout this journey.

Shuvajit Bhattacharya

March 18, 2016

TABLE OF CONTENTS

ABSTRACT.....	ii
ACKNOWLEDGEMENTS.....	iv
TABLE OF CONTENTS.....	v
LIST OF FIGURES.....	viii
LIST OF TABLES.....	xiv
PREFACE.....	xvi
CHAPTER 1.....	1
Integrated Petrofacies Characterization and Interpretation of Depositional Environment of the Bakken Shale in the Williston Basin, North America.....	1
Abstract	2
1.1 Introduction	3
1.2 Geological Background	5
1.3 Dataset and Methods	6
1.4 Results and Discussions	13
1.5 Conclusions	28
Acknowledgements	30
Nomenclature.....	30
References.....	31
CHAPTER 2.....	37
Comparison of Supervised and Unsupervised Approaches for Mudstone Lithofacies Classification: Case Studies from the Bakken and Mahantango-Marcellus Shale, USA	37
Abstract.....	38
2.1 Introduction.....	40
2.2 Principles of Mathematical Algorithms.....	42

2.2.1 Support Vector Machine (SVM).....	42
2.2.2 Artificial Neural Network (ANN).....	45
2.2.3 Self-Organizing Map (SOM).....	46
2.2.4 Multi Resolution Graph-based Clustering (MRGC).....	48
2.3 Bakken and Mahantango-Marcellus Shale Lithofacies.....	49
2.4 Materials and Methods.....	54
2.4.1 Pre-processing Input Petrophysical Dataset.....	55
2.4.2 Training and Testing the Classifiers.....	57
2.5 Results and Discussions.....	58
2.6 Conclusions.....	72
Acknowledgements.....	72
References.....	73
CHAPTER 3.....	78
Integrated 3D Shale Lithofacies Modeling of the Bakken Shale in the Williston Basin, North Dakota.....	78
Abstract.....	79
3.1 Introduction.....	81
3.2 Geologic Setting.....	83
3.3 Dataset and Methods.....	86
3.4 Results and Discussions.....	95
3.5 Suggestions for Further Study.....	112
3.6 Conclusions.....	112
Acknowledgements.....	114
References.....	114

SUMMARY.....122

LIST OF FIGURES

CHAPTER 1

Figure 1-1. Structural map of the Upper Bakken member showing location of two wells (marked by green circles) chosen for the study. Nesson anticline, which is a north-south oriented anticline present near the basin center has been shown on the map (marked by crisscross brown lines).

.....8

Figure 1-2. Cross-section showing thickness variation of different members (upper, middle and lower members in the Bakken Formation between well A and B. Well B shows larger thickness of Bakken units compared to well A due to its presence near the basin center. Similarity Threshold Modeling (STM) of common logs performed between A (reference well) and B (application well) shows similar class (value 3), ambiguous class (value 2) and dissimilar class (value 1) across the Bakken interval. GR log in Track 1 is shaded based on API value.....9

Figure 1-3. Multiwell neutron-density crossplot showing similarity of density and neutron logs color coded by STM log in the Bakken interval. Most of the corresponding data points between two wells are similar, apart from a few measurements (which can be considered as ambiguous or dissimilar). Dissimilarities may arise due to log measurements, bad borehole conditions affecting log quality, tight streaks, or boundary effects.....10

Figure 1-4. Ternary diagram showing mineralogical composition and TOC of Bakken shale members based on core and PNS log data. It can be observed that Bakken shale units are generally rich in quartz and clay compared to carbonate. Three lines (Quartz-to-Carbonate ratio 3 and 1/3) and clay (30%) have been drawn over the triangle to identify and characterize different shale facies clusters. Color scale is based on weight percent Total Organic Carbon (TOC)11

Figure 1-5. The schematic workflow utilized to classify shale petrofacies based on core analysis and PNS log data in the Bakken shale. Five different shale petrofacies could be classified based on available mineralogical composition and TOC data.....12

Figure 1-6. Thorium/Potassium vs. Thorium/Uranium plot showing most of the organic-rich shale units are illite/mixed clay in nature and were deposited in reducing condition compared to organic-poor gray shale samples, which were rather deposited in relatively “oxic” (dysoxic)

condition (modified after Doveton, 1994). Color scale is based on weight percent Total Organic Carbon (TOC).....14

Figure 1-7. Relation between GR and uranium from Spectral GR log showing higher GR values in Bakken shale are due to high uranium content, which can be attributed to presence of high organic matter richness (indicated by TOC color bar). The correlation coefficient is around 98%

.....15

Figure 1-8. Mineralogy and petrofacies of Bakken shale in the well B. PNS log derived mineralogy was used to supervise shale petrofacies classification. Five different shale petrofacies could be identified. It should be noted that Lodgepole Formation, middle Bakken and Three Forks Formation were not classified; they were assumed to be composed of one facies, using the geological knowledge for this study. Track 1 is Gamma-ray shaded by API, Track 2 is the multi-mineral model defined by PNS logs, Track 3 is the stochastic multi-mineral model with kerogen (kerogen could also include fluids), and Tracks 4 is the five organic-rich and organic-poor petrofacies determined from the stochastic mineralogy, using criteria in Figure 1-5.....16

Figure 1-9. Stochastic mineralogy and classified petrofacies of the Bakken shale in the well A. Five different shale petrofacies could be identified, similar to well B. Track 1 is Gamma-ray shaded by API, Track 2 is the stochastic multi-mineralogical model with kerogen (kerogen could also include fluids), and Track 3 are the five organic-rich and organic-poor petrofacies defined by criteria in Figure 5.....17

Figure 1-10. Different depositional environments of mudstone (modified after Newton, 2012). Trace element concentrations increase with sulfur level, which becomes the most in the euxinic environment.....18

Figure 1-11. (a) through (e) show trace element (molybdenum, uranium, vanadium, copper, and nickel) behavior with TOC in the shale units. Trace element richness is small in the dysoxic zone, whereas it exhibits linear relation with TOC in the anoxic environment. All of the trace elements show highest amount of enrichment and independence of increase in TOC during euxinic environment. Green line (at TOC 7%) indicates possible boundary of dysoxic and reducing (anoxic and euxinic) environment.....19

Figure 1-12. Petrofacies distribution, depositional environment interpreted based on major mineralogy, trace element and TOC data in the well A. The upper and lower Bakken shale members are characterized by trace element patterns (marked by red arrows). Track 1 is Gamma-ray shaded by API, Track 2 is the stochastic mineralogy, Track 3 is the petrofacies from stochastic mineralogy, Track 4 is the depositional environment (yellow, green, gray, and red bars indicating

oxic, dysoxic, anoxic, and euxinic conditions respectively), and Tracks 5 through 9 show trace element distribution.....25

Figure 1-13. Petrofacies distribution, depositional environment interpreted based on major mineralogy, trace element and TOC data in the well B. A gap in interpreting depositional environment in the upper Bakken portion resulted due to unavailable core data. Track 1 is Gamma-ray shaded by API, Track 2 is the stochastic mineralogy, Track 3 is the petrofacies from stochastic mineralogy, Track 4 is the depositional environment, and Tracks 5 through 9 show trace element distribution.....26

CHAPTER 2

Figure 2-1. An optimal hyperplane separating two classes of data in SVM model (Kordon, 2010).....43

Figure 2-2. SVM kernel function maps data from input space to feature space (modified after Kordon, 2010).....43

Figure 2-3. Schematic architecture of an ANN algorithm where input layer corresponds to different well logs and output refers to different lithofacies (modified after Manshad et al., 2015).....46

Figure 2-4. Study areas in the Williston and Appalachian basins marked by black rectangles49

Figure 2-5. Ternary diagrams (a and b) showing variation in major mineralogy and TOC content (all in percentage) in the Bakken and Marcellus shale formations respectively. Both shale formations are siliceous mudstone (modified after Bhattacharya and Carr, 2016).....50

Figure 2-6. Schematic methodology to classify shale lithofacies. Three criteria such as clay volume, TOC content, and quartz-to-carbonate ratio (Q/C) are used to classify shale lithofacies quantitatively (modified after Bhattacharya and Carr, 2016).....52

Figure 2-7. Integrated well log and original lithofacies model display in one of the training wells in the Bakken dataset. Upper Bakken and lower Bakken shale members are composed five different mudstone lithofacies, based on mineralogy and organic matter richness. Limestone, mixed lithology (calcareous siltstone), GMD, GMS, GSS, OSS, and OMD lithofacies are represented by numerical codes from 1 to 7 in a rainbow pattern.....53

Figure 2-8. Integrated well log and original lithofacies model display in one of the training wells in the Mahantango-Marcellus interval. The Mahantango interval is mostly composed of lower

order gray shale facies, whereas the Marcellus Formation is mostly composed of different types of organic shale lithofacies, along with carbonate interlayers. Limestone, GMD, GMS, GSS, OSS, and OMD lithofacies are represented by numerical codes from 1 to 6 respectively.....54

Figure 2-9. Histograms showing statistical feature of conventional well log suites (as model input) in the Bakken (a) and Mahantango-Marcellus (b) dataset. The green curve indicates Gaussian distribution and the black curve indicates cumulative distribution.....57

Figure 2-10. An example of 2D SOM maps representing all ten input parameters in lower-dimensional grid space (five neurons in each side) by representative log values (a) and lithofacies (b) respectively in the Bakken dataset.....62

Figure 2-11. The Bakken interval lithofacies are classified in a training well. The second track shows original lithofacies, and third, fourth, fifth, and sixth track are showing training results from four different algorithms. Lithofacies models classified by all three supervised techniques (SVM, ANN, and SOM) look significantly similar to the original lithofacies model.....63

Figure 2-12. The Bakken interval lithofacies are classified in a test well. The second track shows original lithofacies, and four other tracks showing predicted lithofacies results from four different algorithms.....64

Figure 2-13. The Mahantango-Marcellus interval lithofacies are predicted in a test well. The second track shows original lithofacies, and third, fourth, fifth, and sixth tracks show test results from four different algorithms.....65

Figure 2-14. Summary of lithofacies prediction by SVM, ANN, SOM and MRGC methods in the Bakken and Mahantango-Marcellus dataset. SVM appears to be the best method.....68

Figure 2-15. SVM Network parameter optimization by checking accuracy level at every gamma and C values (a and b).....69

Figure 2-16. Variation in lithofacies classification accuracies with change in data size.....70

CHAPTER 3

Figure 3-1. Paleogeographic location of the Williston basin in North America during the Late Devonian period (a) and study area in North Dakota, with available core and well log data (b). Figure 3-1b shows the present day upper Bakken structure map (modified after Blakey, 2011).

.....85

Figure 3-2. Lithostratigraphy of the Bakken Formation. The Bakken Formation is composed of three members- upper shale, middle mixed lithology, and lower shale (modified after Egenhoff et al., 2011).....86

Figure 3-3. Ternary diagram showing variation of mineralogy and TOC content of the Bakken shale members based on core and PNS log data. The Bakken shale members are rich in quartz and clay, compared to carbonate. Most of the samples have high TOC. Three lines (Quartz-to-Carbonate ratio 3 and 1/3) and clay (30%) are drawn over the triangle to identify and distinguish different shale lithofacies (Bhattacharya and Carr, 2016).....88

Figure 3-4. A general workflow to classify shale lithofacies using core and PNS log data. Six different shale lithofacies could be classified based on mineralogy and TOC data, however, organic mixed shale lithofacies is absent in the Bakken Formation.....89

Figure 3-5. Integrated well log and original lithofacies model display in one of the training wells in the Bakken dataset. Upper Bakken and lower Bakken shale members are composed of five different shale lithofacies. Limestone, mixed lithology (calcareous siltstone), GMD, GMS, GSS, OSS, and OMD lithofacies are represented by numerical codes from 1 to 7 in a rainbow pattern
.....91

Figure 3-6. Integrated display of lithofacies model, depositional conditions (paleo-redox) and geochemical data. A gap in interpreting depositional environment in the upper Bakken member resulted due to unavailable core data.....92

Figure 3-7. Histograms showing statistical feature of conventional well log suites in the Bakken dataset, used as input parameters for SVM and ANN model. Five basic well logs and five derived parameters are used as input shale lithofacies pattern recognition. The green curve indicates Gaussian distribution, whereas the black curve indicates cumulative distribution of well log data
.....94

Figure 3-8. Lithofacies in the Bakken interval are classified in a well. The third track shows original lithofacies, whereas fourth, and fifth track are showing training results from SVM and ANN algorithms. Lithofacies models generated by SVM looks significantly similar to the original lithofacies model, compared to the ANN model.....97

Figure 3-9. Lithofacies proportion in the upper and lower Bakken shale members. Both upper and lower shale members are classified into 50 different layers, to preserve the details of lithofacies distribution information during 3D modeling.....101

Figure 3-10. 3D shale lithofacies model of the Bakken Formation. AA' line represents the well cross-section in Figure 3-11.....102

Figure 3-11. An example of lithofacies upscaling in the Bakken Formation. Lithofacies logs are upscaled separately for the upper Bakken, and the remaining part of the Bakken (middle plus lower members) separately. The first and second track show the GR log and original lithofacies, whereas third and fourth track show upscaled lithofacies logs. The well cross-section is flattened on the upper Bakken top and it is oriented along east-west (AA') (Figure 3-10).....103

Figure 3-12. 3D shale lithofacies model of the upper and lower shale members in “simbox” view104

Figure 3-13. 3D shale lithofacies model validated using 90% of total number of wells. The validation of the 3D models is done for both upper and lower Bakken members (Figures 3-12c and 3-12d).....106

Figure 3-14. Lithofacies cross-section along east-west direction. Nesson Anticline can be observed near the central portion of the basin. Middle Bakken member (colored blue) is assumed to be composed of one lithofacies- mixed lithology.....107

Figure 3-15. Net Organic Siliceous Shale (OSS) lithofacies thickness map of the upper and lower Bakken shale members. A major portion of thick OSS lithofacies is present near the central portion of the basin, along with the northern margin of the study area. Area with OSS lithofacies thickness less than 5 feet are colored white, to better understand the subtle depositional signal of shale lithofacies.....108

Figure 3-16. Silica versus titanium crossplot showing variation of silica content with titanium in the upper, lower Bakken members, and Three Forks Formation.....109

LIST OF TABLES

CHAPTER 1

Table 1-1. Pearson’s correlation coefficients were calculated to understand covariation among GR, TOC, different major and trace elements in the euxinic system. Bold colors indicate highly positive correlation.....23

Table 1-2. Pearson’s correlation coefficients were calculated to understand covariation among GR, TOC, different major and trace elements in the anoxic system. Bold colors indicate highly positive correlation.....23

Table 1-3. Chi-square test between core/log derived shale petrofacies and depositional environment for both wells.....28

CHAPTER 2

Table 2-1. Confusion matrix illustrates accuracy of SVM algorithm in lithofacies prediction in the Bakken test dataset. Overall accuracy by SVM is 87.3%.....59

Table 2-2. Confusion matrix illustrates accuracy of ANN algorithm in lithofacies prediction in the Bakken test dataset. Overall accuracy by ANN is 83.7%.....60

Table 2-3. Confusion matrix illustrates accuracy of 2D SOM algorithm in lithofacies prediction in the Bakken test dataset. Overall accuracy by SOM is 84.03%.....60

Table 2-4. Confusion matrix illustrates accuracy of MRGC algorithm in lithofacies prediction in the Bakken test dataset. Overall accuracy by MRGC is 71.3%.....61

Table 2-5. Confusion matrix illustrates accuracy of SVM algorithm in lithofacies prediction in the Mahantango-Marcellus dataset. Overall accuracy by SVM is 82.46%.....66

Table 2-6. Confusion matrix illustrates accuracy of ANN algorithm in lithofacies prediction in the Mahantango-Marcellus dataset. Overall accuracy by ANN is 78.75%.....66

Table 2-7. Confusion matrix illustrates accuracy of 2D SOM algorithm in lithofacies prediction in the Mahantango-Marcellus dataset. Overall accuracy by SOM is 69.64%.....67

Table 2-8. Confusion matrix illustrates accuracy of MRGC algorithm in lithofacies prediction in the Mahantango-Marcellus dataset. Overall accuracy by MRGC is 64.42%.....67

CHAPTER 3

Table 3-1. Confusion matrix illustrating accuracy of SVM algorithm in lithofacies prediction in the Bakken test dataset. Overall accuracy by SVM algorithm is ~87.3%.....98

Table 3-2. Confusion matrix illustrating accuracy of ANN algorithm in lithofacies prediction in the Bakken test dataset. Overall accuracy by ANN technique is ~83.7%.....98

Table 3-3. Summary of the mineralogical features, TOC content, and conventional well log responses of the five shale lithofacies defined from core and advanced well logs in the Bakken shale of the Williston basin.....100

APPENDICES

Table A-1. Total organic Carbon (TOC) content and trace Element Geochemical data for the Sikes State Well. TOC data were measured at NETL laboratory, Morgantown, whereas trace element data were measured by Actlabs, Canada.....120

PREFACE

This research is a multi-scale study, integrating various types of geological, geochemical, petrophysical, and geophysical data at core, well, and regional scales. The study was completed and written as three papers, described as three chapters. Chapter 1 describes the methodology for shale lithofacies classification and interpretation of depositional conditions, using core data ("ground truth") and stochastic petrophysical modeling. Chapter 2 discusses application of different supervised and unsupervised mathematical algorithms (such as Support Vector Machine, Artificial Neural Network, Self-Organizing Map, and Multi Resolution Graph-based Clustering) for shale lithofacies pattern recognition and prediction, and it compares their results at well scale. All these algorithms are also being applied in the Marcellus shale of the Appalachian basin, North America to test their robustness and broader applicability in different geological domains. The research culminates with Chapter 3, describing 3D shale lithofacies modeling approaches at regional scale, and detailed interpretation of depositional and diagenetic environment of different shale lithofacies in the Bakken Formation. The Appendix A presents core data used in this work.

The first, second, and third chapters have been submitted as manuscripts to the Petrophysics journal, Journal of Natural Gas Science & Engineering, and Interpretation journal respectively. All the journals have their own standards for publications, and therefore, a few figures have been altered to meet their criteria. Also, a cross-section has been included in the Appendix (A-2) of this study to illustrate stratigraphic relationships among different geological formations and members.

CHAPTER 1

Integrated Petrofacies Characterization and Interpretation of Depositional Environment of the Bakken Shale in the Williston Basin, North America

Shuvajit Bhattacharya¹ and Timothy R. Carr¹

¹ Department of Geology and Geography

West Virginia University, Morgantown

WV- 26506, USA

(304)-777-9885

Email: subhattacharya@mix.wvu.edu; tim.carr@mail.wvu.edu

Abstract

The study demonstrates applications of core and advanced well logs on computing general log based stochastic multi-mineral solutions to build detailed 1-D shale petrofacies model and integrate with chemostratigraphy to better decipher depositional environments of the Bakken Shale units in the Williston basin of North Dakota, USA. In particular, relations among trace element geochemical data and organic matter coupled with well log derived crossplots and solutions are explored to understand vertical and areal heterogeneity of the shale members in the Bakken Formation. A methodology based on mineral composition and organic matter richness derived from well logs and core data is proposed for facies classification in the Bakken mudstone units. The results show that Bakken shale members are heterogeneous, in terms of mineralogy and organic matter, which can be classified as five different petrofacies, reflective of changes in depositional and diagenetic environment. Highly organic-rich shale facies units were deposited in euxinic environment, whereas relatively organic-poor shale units were deposited in anoxic and dysoxic conditions. Statistical analyses suggest that trace element geochemical data can be applied to a significant degree of confidence to compare with log derived facies model to characterize different shale petrofacies and construe the depositional environment in detail.

1.1 Introduction

Lithofacies or petrofacies classification, assigning a rock type to specific rock samples on the basis of petrography or measured petrophysical properties, is fundamental to subsurface investigations. Clastic and carbonate petrofacies have been studied extensively for depositional and diagenetic environment studies. However, research in black shale petrofacies is relatively rare, most being based on either single well study or descriptive analysis (Bhattacharya et al., 2015; Egenhoff and Fishman, 2013; Hickey and Henk, 2007; Schieber, 1999). A case study from the Bakken Formation in the Williston basin in North America has been chosen for this study.

The Bakken petroleum system (Devonian-Mississippian) in the Williston basin of North Dakota and Montana is one of the largest unconventional oil plays in North America with an estimated 7.4 billion barrels of undiscovered oil reserve (Pollastro et al., 2013). The Bakken Formation consists of two world class source rocks (upper Bakken and lower Bakken shale) that sandwich the reservoir middle Bakken, which is composed of mixed lithologies including sandstone, dolomite and limestone. Most of the geological and petrophysical studies in the Bakken play, ranging from depositional history, facies analysis, and reservoir characterization, have been focused on the middle Bakken member and Three Forks Formation (LeFever et al., 2011; Sesack, 2011; Simenson, 2011; Warner, 2011). Source rock potential of different mudstone reservoirs (including the Bakken Formation), in terms of hydrocarbon production and chemostratigraphy using X-ray Fluorescence (XRF) and Total Organic Carbon (TOC) data has been performed recently towards better understanding of geological processes controlling source and transportation of organic matter and mud to depositional site, preservation of organic matter and diagenesis of mud after deposition (Maldonado, 2012; Nandy et al., 2014; Rowe et al., 2012;

Schieber et al., 1998a, b). In this study we attempt to identify different petrofacies in the Bakken shale units based on petrophysical and geochemical characteristics to better understand their depositional and diagenetic controls on mineral composition and TOC.

Although Bakken along with the Three Forks formations have been the target for horizontal drilling and hydraulic stimulation throughout the basin, several uncertainties remain including facies variation due to depositional and diagenetic controls on mineral composition and organic matter content in the Bakken shale members, which could play a significant role in hydrocarbon production. Improved characterization of Bakken shale petrofacies is important to better understand depositional environment, facies distribution and their influence on hydrocarbon production. In this study, we claim that both Bakken shale members are heterogeneous and complex mixture of quartz, smectite, illite, carbonate, pyrite and kerogen in varying proportions based on core and well log derived petrophysical interpretation, which can be classified as different petrofacies. Five different petrofacies could be identified and correlated in the Mississippian- Devonian shale interval. Contrary to the popular beliefs of shale facies deposition in anoxic environment, we “deconstruct” that major portion of both upper and lower shale members were rather deposited in euxinic environment (high sulfur content in the ocean) with limited portion being restricted to anoxic and dysoxic conditions based on trace element geochemical study.

In this study proper well log conditioning and similarity modeling between the wells were performed before application and interpretation for facies clustering. Core X-ray Diffraction (XRD) and TOC data has been calibrated with conventional and advanced well log (such as Pulsed

Neutron Spectroscopy, PNS) derived petrophysical parameters to classify shale petrofacies at core and log scales. A general methodology integrating core and log analysis for shale petrofacies classification in the Bakken Formation has been presented. Next, covariance of GR, TOC logs, various trace elements (such as Mo, U, V, Cu and Ni), and clay (Al and K) has been investigated, in an attempt to tie geochemical signatures from core data with log derived stochastic petrofacies models to better delineate association of different shale petrofacies with respective depositional environment with a significant degree of confidence.

1.2 Geological Background

The Bakken Formation in the Williston basin spans over portions of North Dakota, South Dakota, Montana in the United States and Saskatchewan, Manitoba in Canada (~200,000 mi², ~500,000 Km²). The Williston basin is a large intracratonic sedimentary basin (Pitman et al., 2001). Deposition of sediments in the basin began in the Cambrian; however, increased subsidence and sedimentation took place from the Ordovician to the Devonian when the Bakken Formation was deposited (Smith and Bustin, 2000; Webster, 1984). Both shale members in the Bakken Formation are considered to be deposited during the basin wide transgression and/or basin subsidence at the end of Three Forks Formation deposition, whereas middle Bakken member (composed of mixed lithology such as limey siltstone and dolomitic siltstone) has been interpreted to have been deposited as a result of dramatic sea-level drop at the close of the Devonian (Dumonceaux, 1984; LeFever et al., 1991; Meissner, 1978; Sesack, 2011; Sonnenberg and Pramudito, 2009; Steptoe, 2012). A major north-south structural feature, Nesson Anticline is

present near the center of the basin, where all Bakken members are thickest (>50 feet, 15m) in the basin.

1.3 Dataset and Methods

Two wells (A and B), around 26 miles (42 km) apart in McKenzie County, North Dakota were chosen for this study. Well B is closer towards the basin center (Figure 1-1), which is why it has thicker Bakken interval compared to Well A (Figure 1-2). The dataset was ideal for this study as both of the wells have conventional well logs (GR, resistivity, neutron porosity, bulk density, and photoelectric) for the selected Bakken interval and trace element geochemical data was collected in a consistent manner. Apart from that, well B possesses PNS log, core spectral GR, core-derived XRD and Rock-Eval pyrolysis (TOC) data. While upper Bakken portion of well B had limited core available for experimental studies, Well A had full core available. XRD and TOC data were measured in laboratory. Two different sources of trace element geochemical data were available: Na-Peroxide-treated Inductively Coupled Plasma-Mass Spectrometry (ICP-MS) analysis and X-ray Fluorescence (XRF). Identification of both upper and lower Bakken shale members was done based on

High GR (~ 400 to 900 API)

High resistivity (~ 25 to 500 Ω -m)

High neutron-porosity (~ 30 to 45%)

High TOC (~ 2 to 20 wt%)

Quality control of well logs is critical before applying them for advanced interpretation and facies clustering. Both of the wells were drilled using oil-based mud, which resulted in good quality of well logs. A density correction curve already available with the well dataset was used to correct bulk density log.

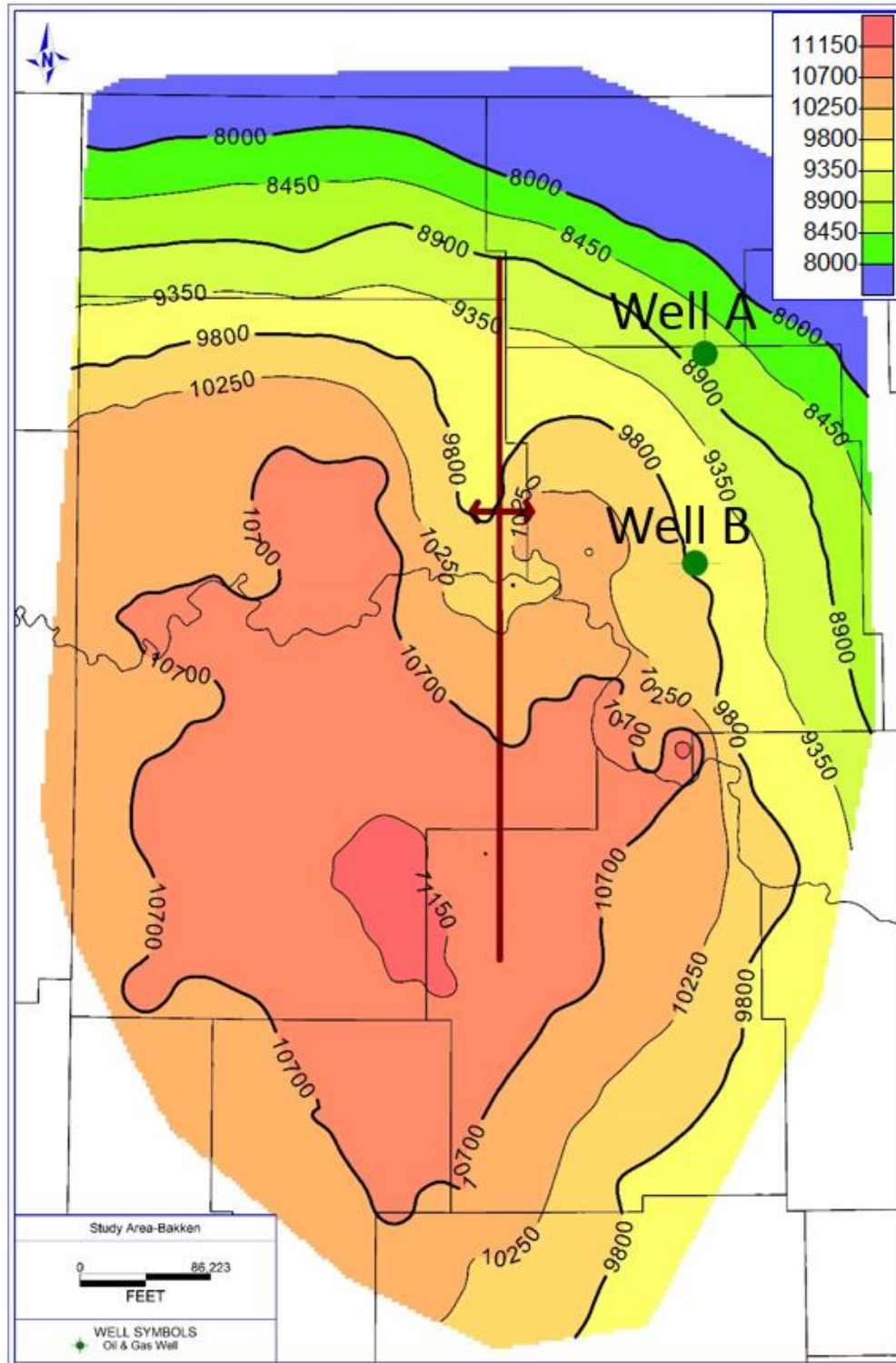


Figure 1-1. Structural map of the upper Bakken member showing location of two wells (marked by green circles) chosen for the study. Nesson Anticline, which is a north-south oriented anticline present near the basin center has been shown on the map (marked by crisscross brown lines).

Similarity Threshold Modeling (STM) was applied to compare the similarity of well log measurements in both wells in the selected interval of study (Thevoux-Chabuel et al., 1997). STM technique can be used to insure quality control by determining the similarity of well logs in the reference well with the application well, which can be represented either in depth section or crossplot (Figures 1-2 and 1-3). Dissimilarities may arise due to bad borehole conditions affecting log quality, tight streaks, or boundary effects.

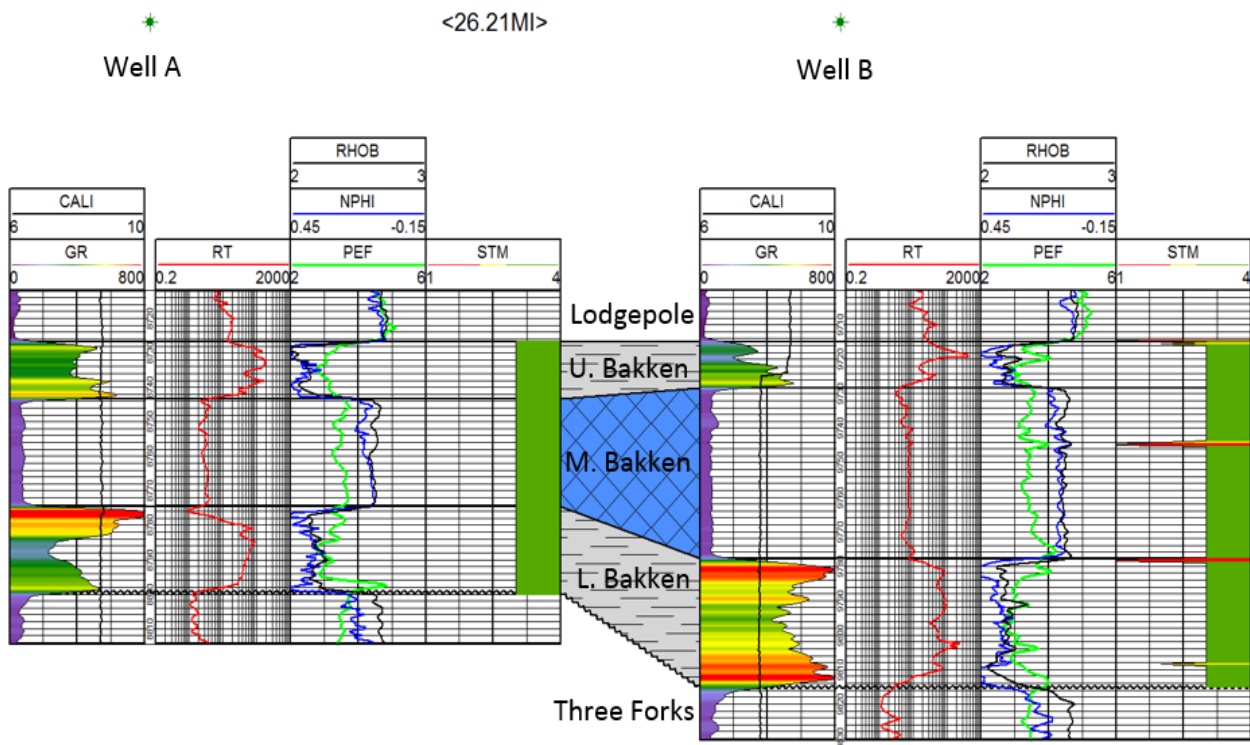


Figure 1-2. Cross-section showing thickness variation of different members (upper, middle and lower members in the Bakken Formation) between well A and B. Well B shows larger thickness of Bakken units compared to well A due to its presence near the basin center. Similarity Threshold Modeling (STM) of common logs performed between A (reference well) and B (application well) shows similar class (value 3), ambiguous class (value 2) and dissimilar class (value 1) across the Bakken interval. GR log in Track 1 is shaded based on API value.

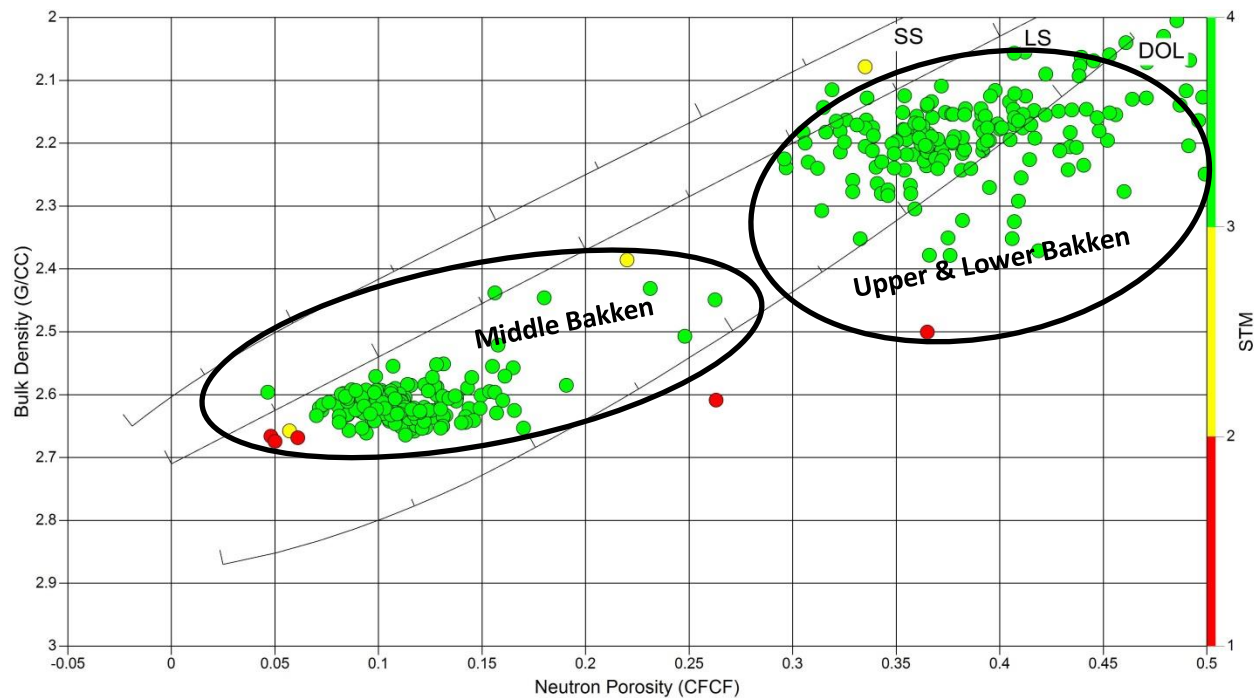


Figure 1-3. Multiwell neutron-density crossplot showing similarity of density and neutron logs color coded by STM log in the Bakken interval. Most of the corresponding data points between two wells are similar, apart from a few measurements (which can be considered as ambiguous or dissimilar). Dissimilarities may arise due to log measurements, bad borehole conditions affecting log quality, tight streaks, or boundary effects.

Corrected density log was used to calculate TOC by Schmoker’s method (1983) and volume of clay was computed using average Vclay from neutron-density (N-D) crossplot and uranium free spectral gamma log (or Computed Gamma Ray, CGR) for better calibration with XRD and TOC data. The equations used are as follows:

$$\text{TOC}_{\text{Schmoker}} = (154.497/\text{Bulk Density}) - 57.261 \text{ and}$$

$$\text{Vclay}_{\text{Final}} = (\text{Vclay}_{\text{N-D}} + \text{Vclay}_{\text{CGR}})/2$$

Next mineralogy and TOC data from core and PNS logs were combined and plotted in a ternary diagram to understand heterogeneity of the shale members (Figure 1-4) and a general

scheme has been proposed for Bakken shale facies classification based on the data pattern (Figure 1-5). Due to the relative scarcity of core points and unavailability of PNS log in well B, stochastic mineral solutions were obtained for both wells using linear inversion technique (Kulyapin and Sokolova, 2014; Mitchell and Nelson, 1988; Moss and Harrison, 1985; Savre, 1963).

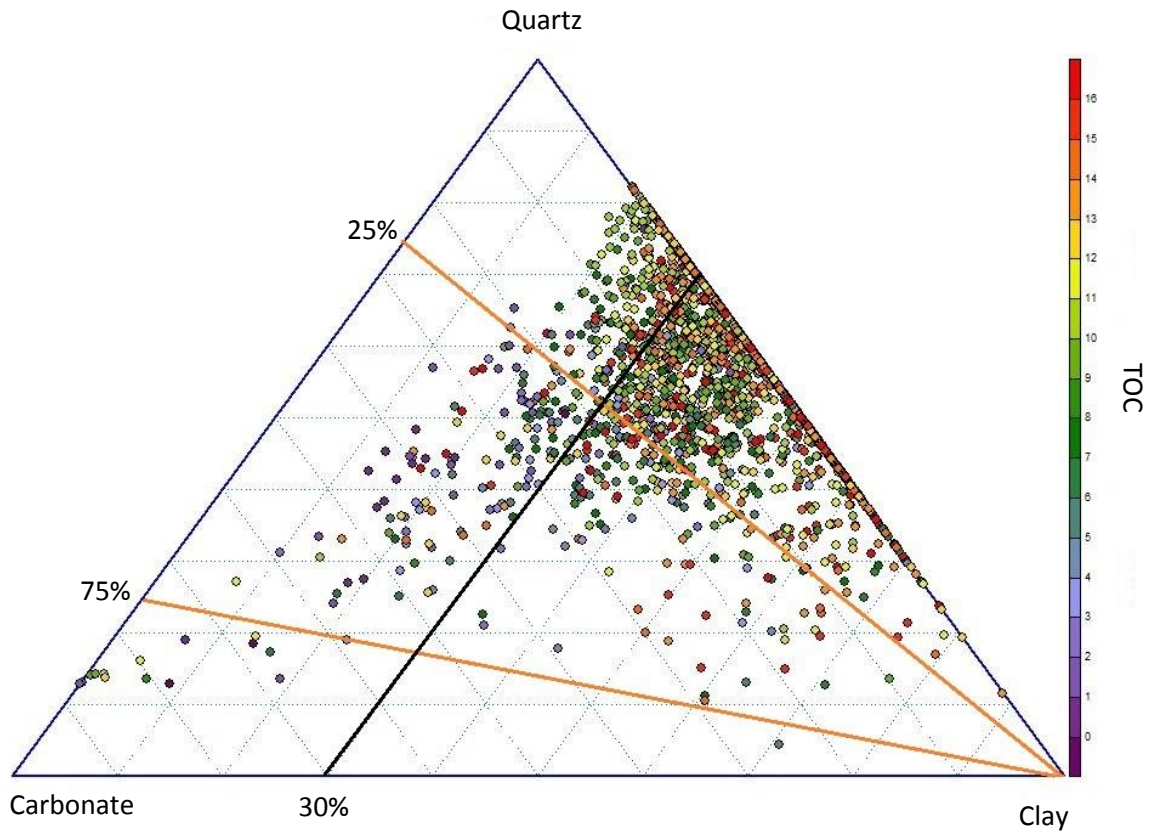


Figure 1-4. Ternary diagram showing mineralogical composition and TOC of Bakken shale members based on core and PNS log data. It can be observed that Bakken shale units are generally rich in quartz and clay compared to carbonate. Three lines (Quartz-to-Carbonate ratio 3 and 1/3) and clay (30%) have been drawn over the triangle to identify and characterize different shale facies clusters. Color scale is based on weight percent Total Organic Carbon (TOC).

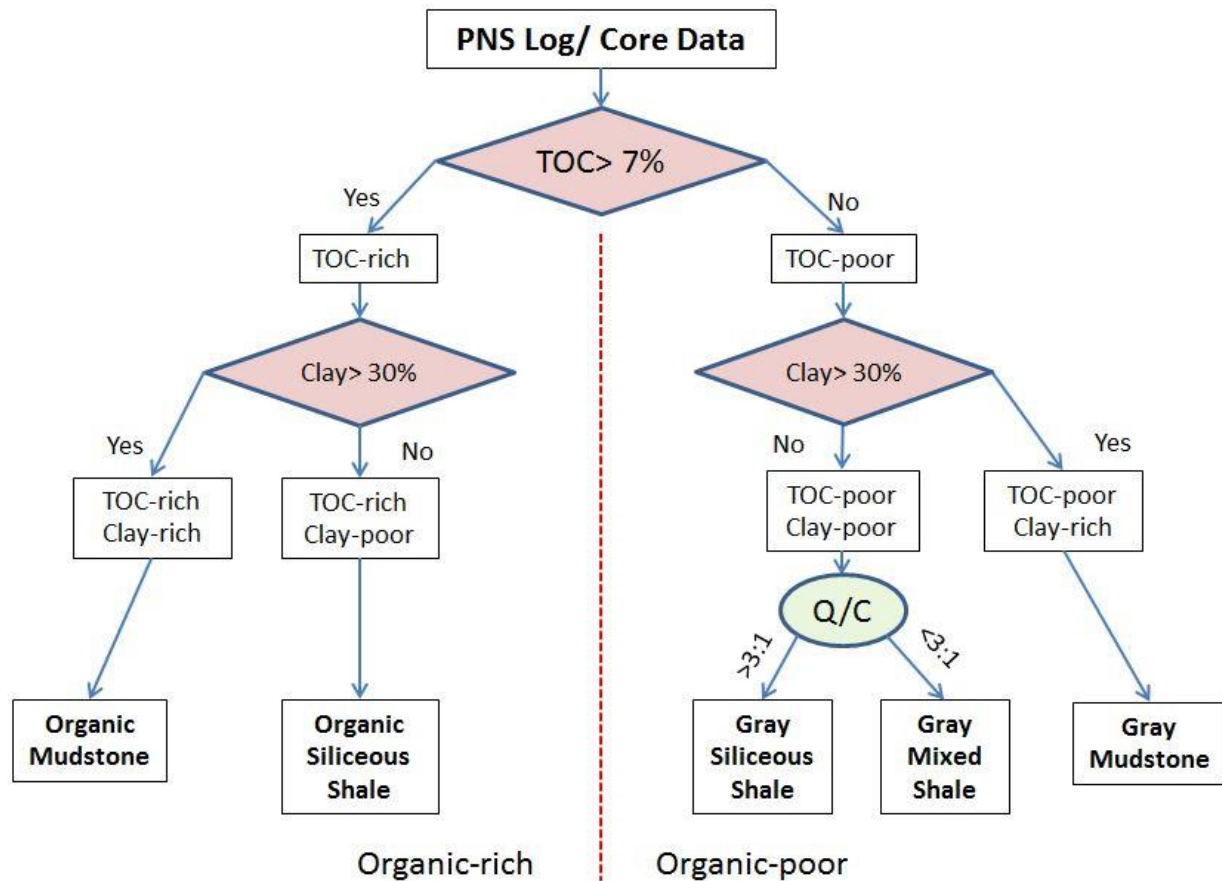


Figure 1-5. The schematic workflow utilized to classify shale petrofacies based on core analysis and PNS log data in the Bakken shale. Five different shale petrofacies could be classified based on available mineralogical composition and TOC data.

Input well logs selected for stochastic mineralogy solutions using PowerLog™ are: GR, neutron-porosity, bulk density, and Umaa (Umaa is the product of photoelectric and density log corrected by apparent total porosity). The final mineralogical solution was constrained by Vclay. Calculated output curves are volumetric proportions of quartz, clay, calcite, dolomite, kerogen and pyrite. The kerogen estimation could also include any free oil and water that is present in the system. The calculated proportions of log derived quartz, clay, carbonate (calcite + dolomite)

and TOC data were used to classify different petrofacies in the shale units based on mineralogy and TOC cutoff criteria as defined from core and PNS logs for rock-typing scheme.

After well log based shale facies classification, trace element geochemical data were cross-plotted with each other and well logs (e.g. GR and TOC etc.) to understand their interrelationship and variability, in terms of depositional environment and different statistical analyses (Pearson's correlation coefficient and Chi-square test) were performed to find association of core and log defined shale petrofacies with chemostratigraphy based depositional environment.

1.4 Results and Discussion

It can be observed from the spectral GR component plots that the clay type is mostly illite (Figure 1-6). Spectral GR log was used for this study as it helps to understand contribution of uranium component on total radioactivity across organic-rich black shale and found to correlate well with TOC (Figure 1-7), because uranium is insoluble in reducing condition and soluble in oxidizing condition (Adams and Weaver, 1958).

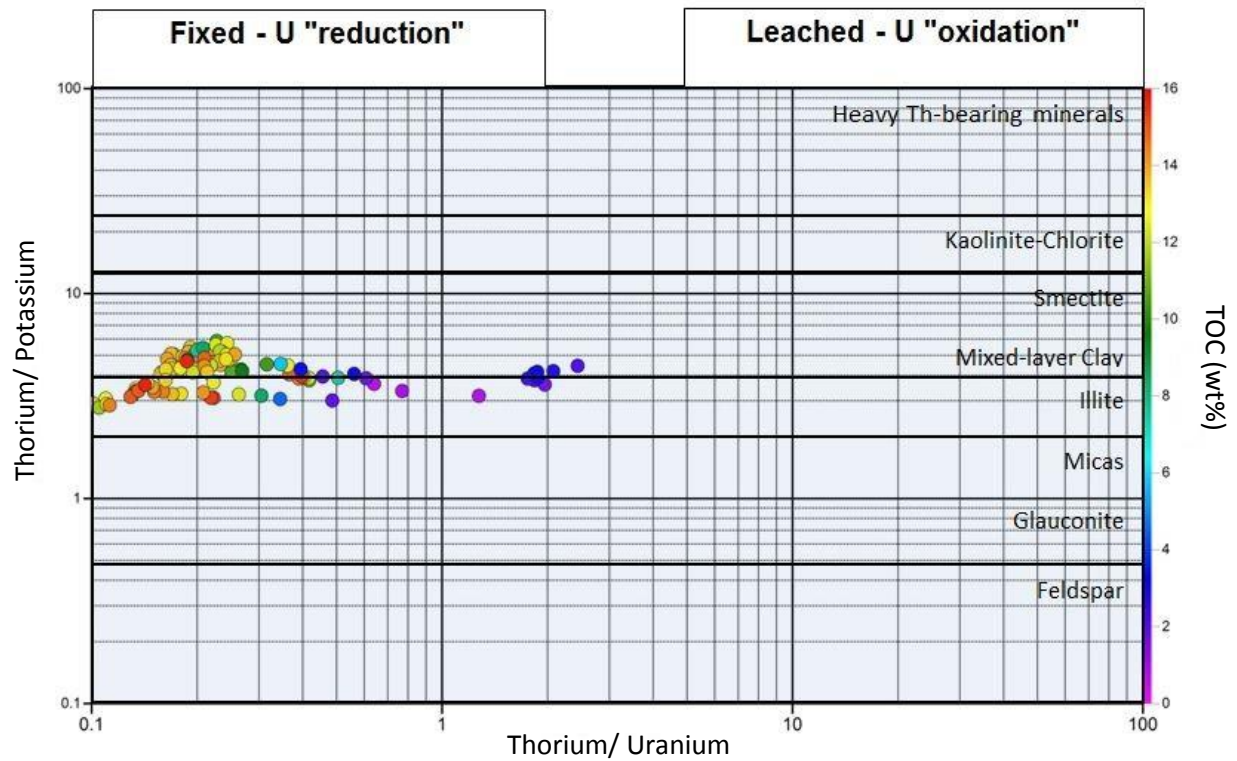


Figure 1-6. Thorium/Potassium vs. Thorium/Uranium plot showing most of the organic-rich shale units are illite/mixed clay in nature and were deposited in reducing condition compared to organic-poor gray shale samples, which were rather deposited in relatively “oxic” (dysoxic) condition (modified after Doveton, 1994). Color scale is based on weight percent Total Organic Carbon (TOC).

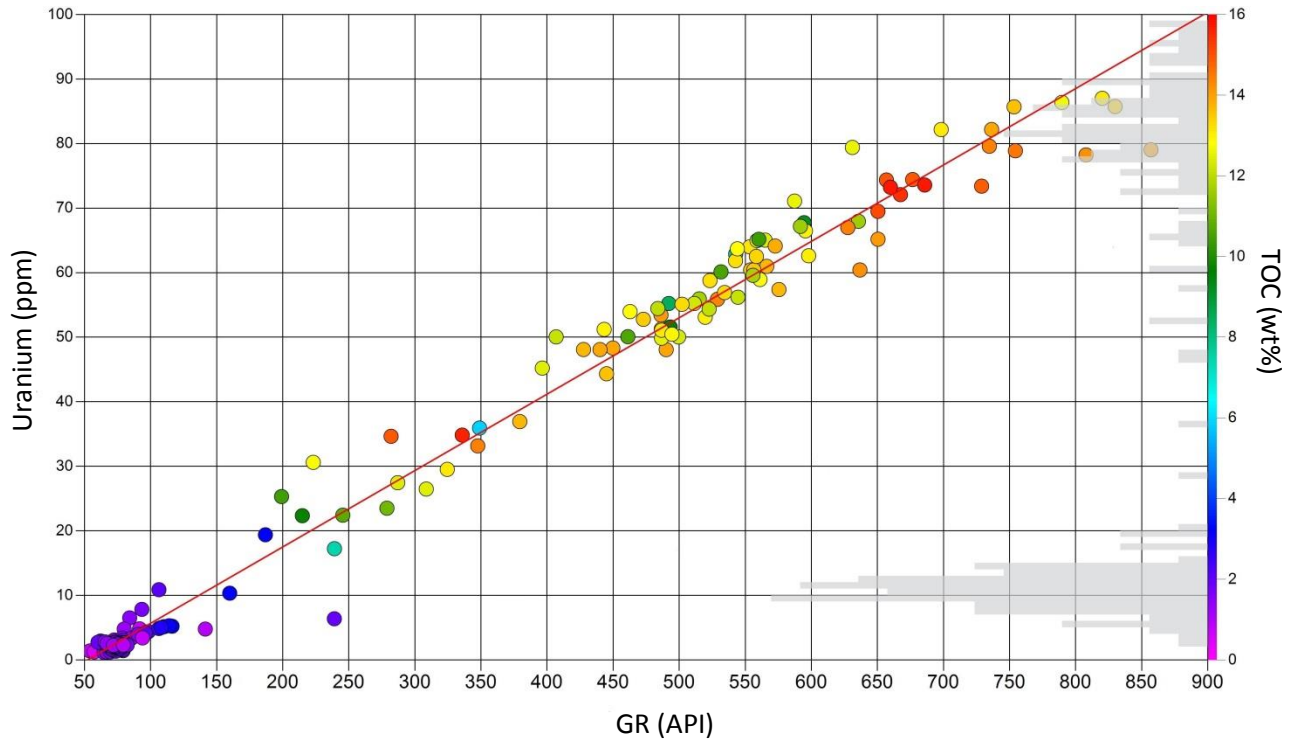


Figure 1-7. Relation between GR and uranium from Spectral GR log showing higher GR values in Bakken shale are due to high uranium content, which can be attributed to presence of high organic matter richness (indicated by TOC color bar). The correlation coefficient is around 98%.

The ternary plot of major mineralogy and TOC data reveals heterogeneity of upper and lower shale members, both of which are rich in quartz and clay compared to carbonate. Concentrating on the points having TOC more than 7% made it easier to distinguish different rock clusters. Three criteria were used to classify shale petrofacies: TOC (cutoff 7%), clay volume (cutoff 30%) and quartz to carbonate ratio (cutoff 3).

Evaluation of stochastic mineral solutions based on above-mentioned shale facies classification criteria was performed to define five different shale petrofacies: two are organic-rich (organic mudstone and organic siliceous shale) and the remaining three are organic-poor shale petrofacies (gray siliceous shale, gray mixed shale and gray mudstone) (Figures 1-8 and 1-

9). It is evident that two organic-rich petrofacies are dominant compared to three organic-poor petrofacies. The three organic-poor petrofacies could be present due to shoulder bed effects, but they appear to be related to core observations, trace element geochemical data, and the expected pattern of facies at the geological transition between anoxic and oxic environments in the Bakken interval.

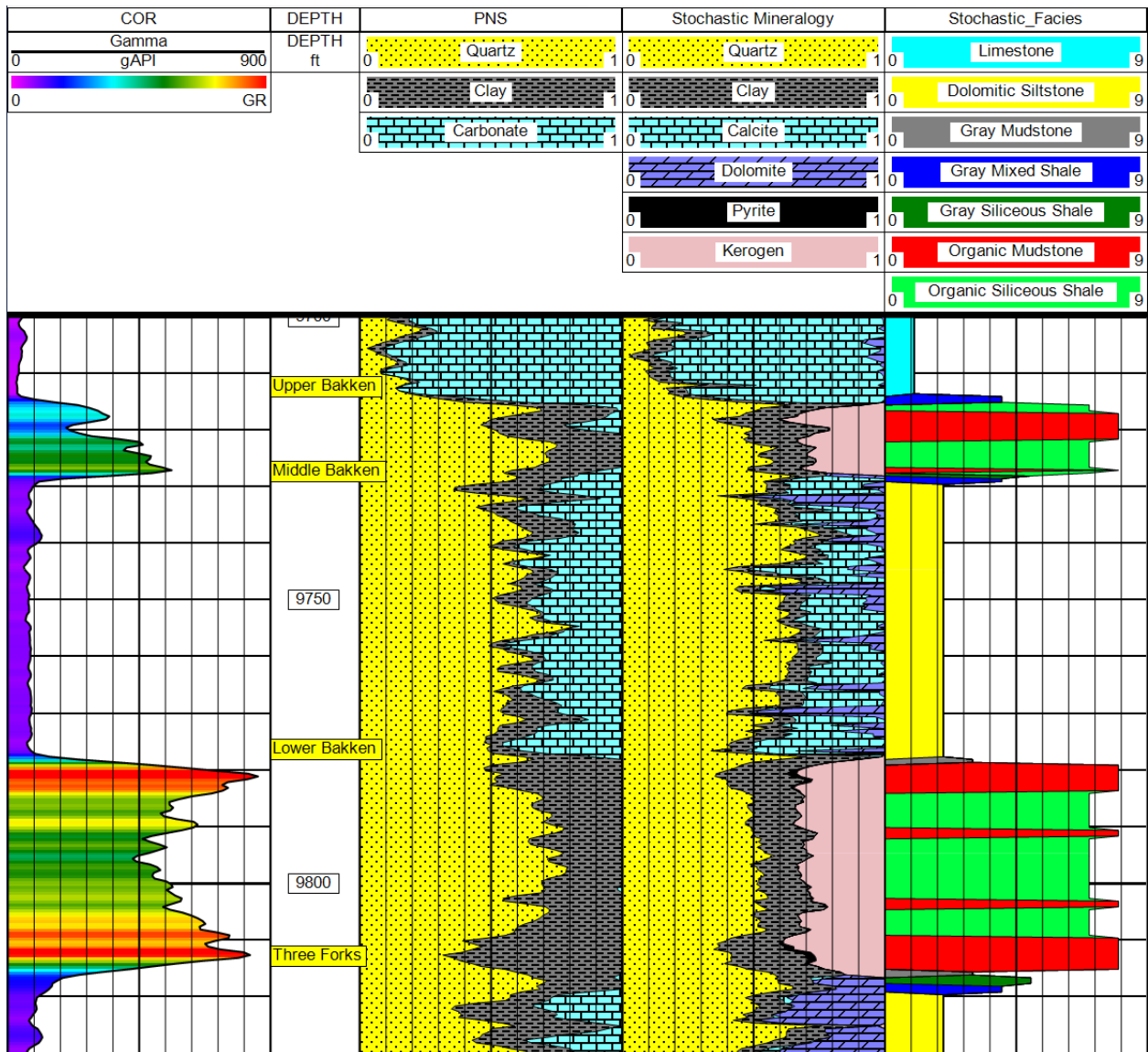


Figure 1-8. Mineralogy and petrofacies of Bakken shale in the well B. PNS log derived mineralogy was used to supervise shale petrofacies classification. Five different shale petrofacies could be

identified. It should be noted that Lodgepole Formation, middle Bakken and Three Forks Formation were not classified; they were assumed to be composed of one facies, using the geological knowledge for this study. Track 1 is Gamma-ray shaded by API, Track 2 is the multi-mineral model defined by PNS logs, Track 3 is the stochastic multi-mineral model with kerogen (kerogen could also include fluids), and Tracks 4 is the five organic-rich and organic-poor petrofacies determined from the stochastic mineralogy, using criteria in Figure 1-5.

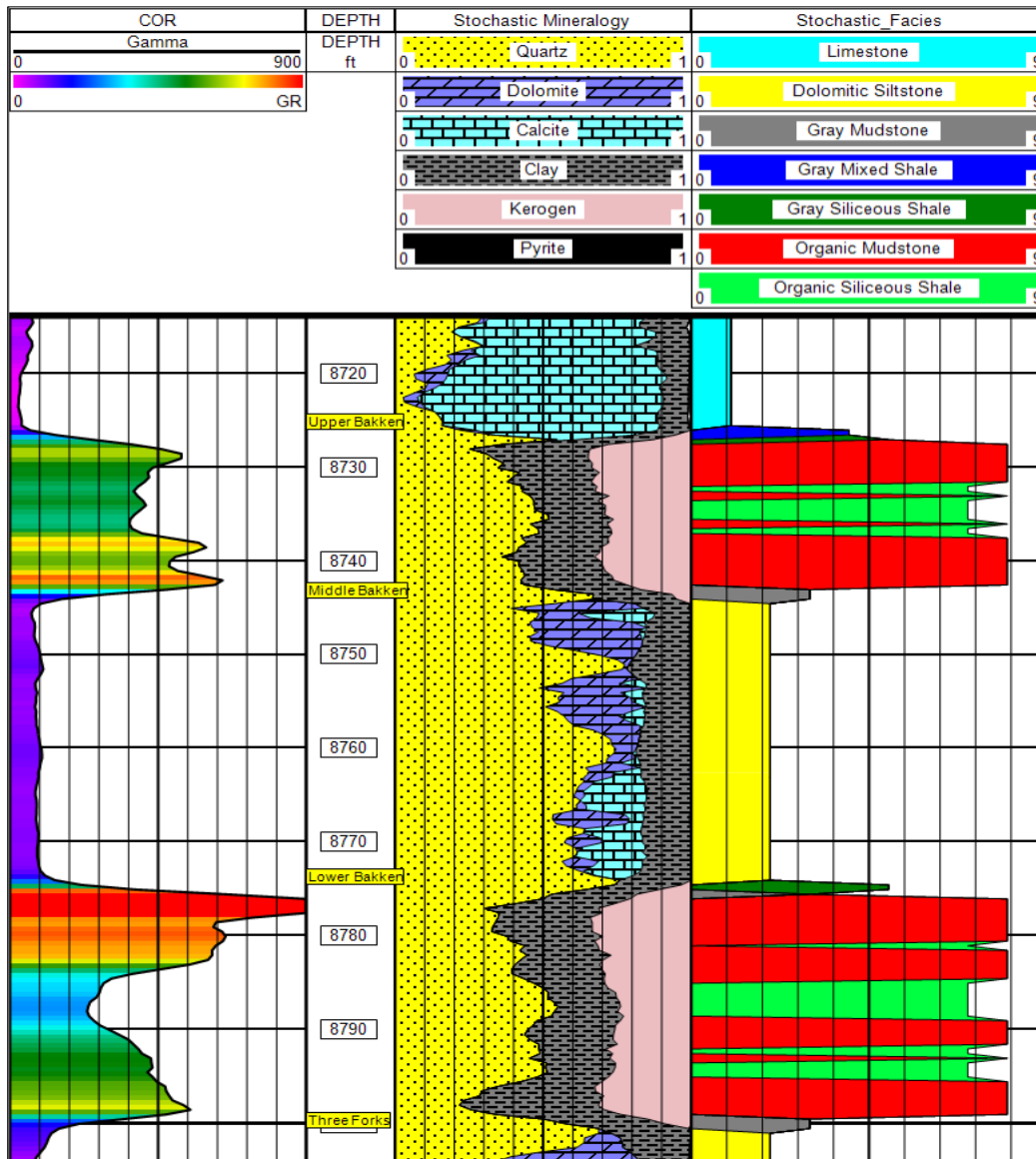


Figure 1-9. Stochastic mineralogy and classified petrofacies of the Bakken shale in the well A. Five different shale petrofacies could be identified, similar to well B. Track 1 is Gamma-ray shaded by API, Track 2 is the stochastic multi-mineralogical model with kerogen (kerogen could also include fluids), and Track 3 are the five organic-rich and organic-poor petrofacies defined by criteria in Figure 5.

Trace element concentrations can be used as geochemical proxies to understand depositional environment in detail and their influence on shale facies distribution (Algeo and Maynard, 2004; Newton, 2012) (Figure 1-10). Bakken shale members are highly enriched in redox-sensitive trace elements (Figures 1-11a-e) such as molybdenum (Mo), uranium (U), vanadium (V), copper (Cu), and nickel (Ni). Covariation among trace elements and TOC depicts three different patterns:

1. Clusters with low TOC (<7%) have relatively small trace element concentrations,
2. Clusters with moderate TOC (7- 12%) have relatively higher trace element concentrations; showing linear interrelationship and
3. Clusters with high TOC (>12%) are overwhelmingly rich in trace element concentrations; showing non-linear positive relationship for both shale members.

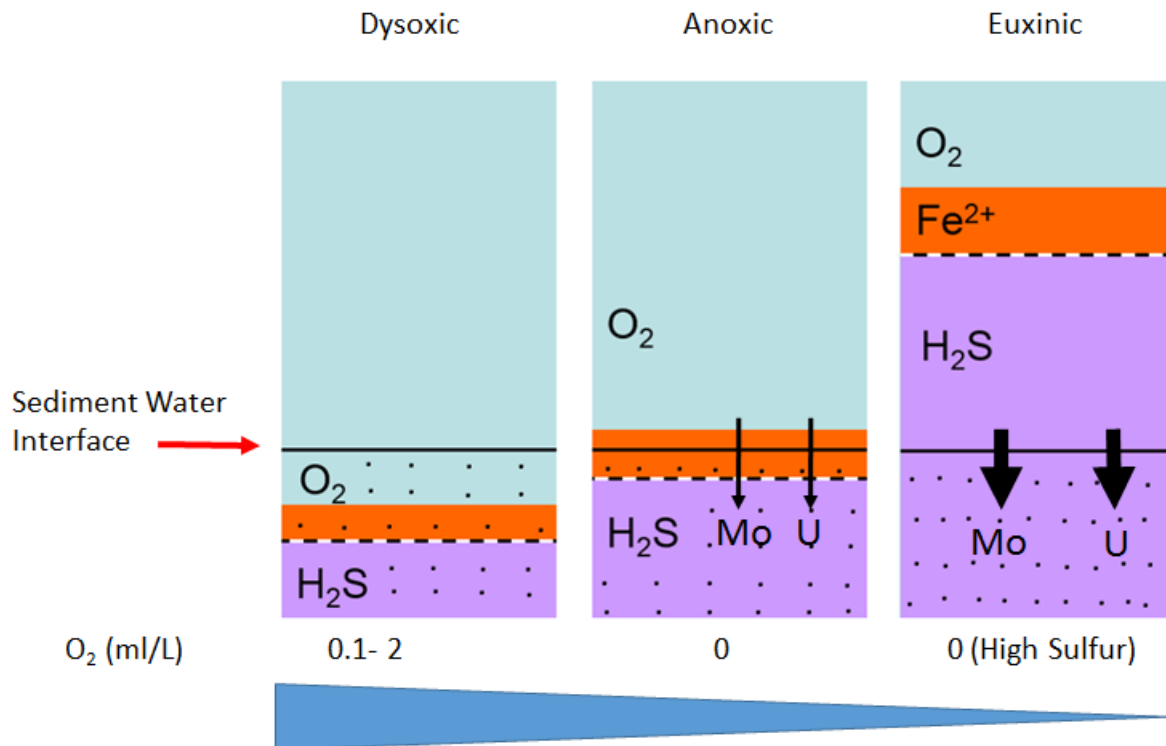
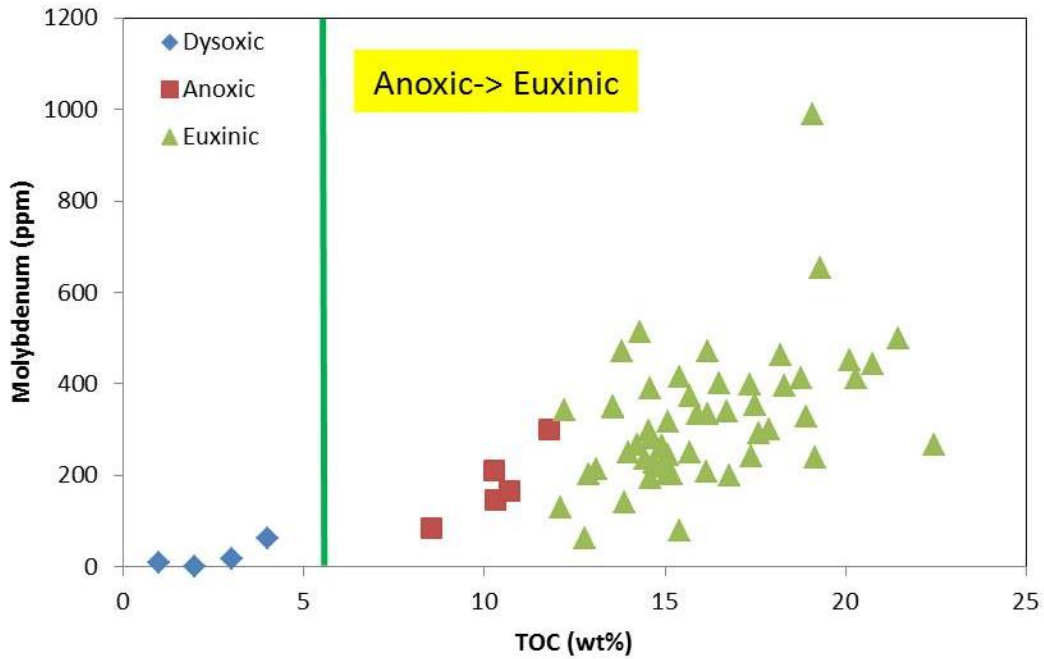
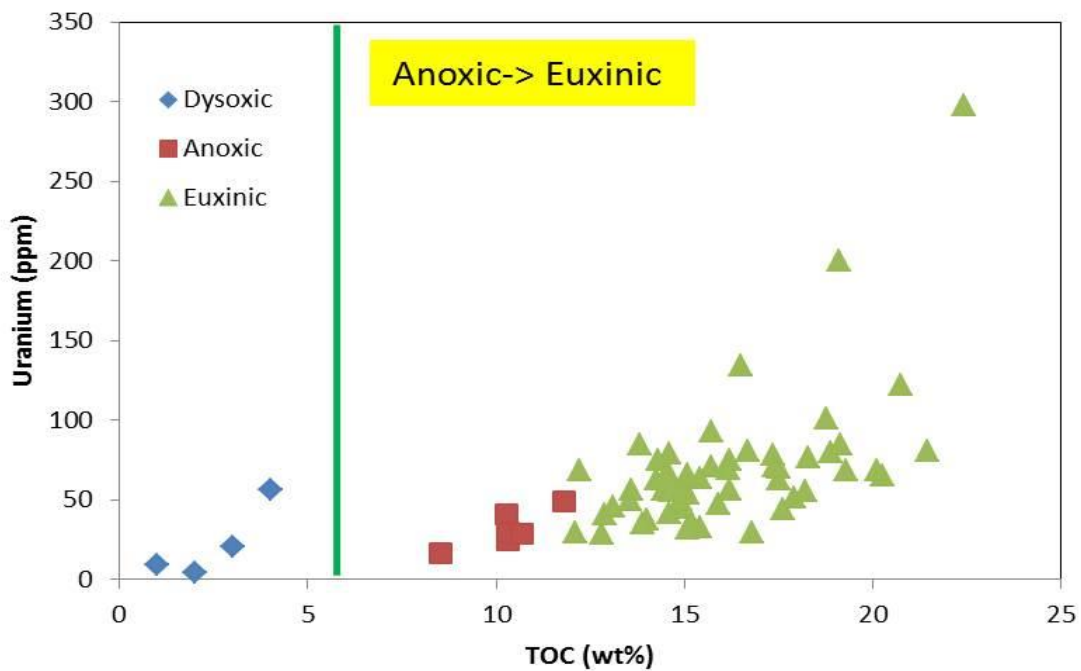


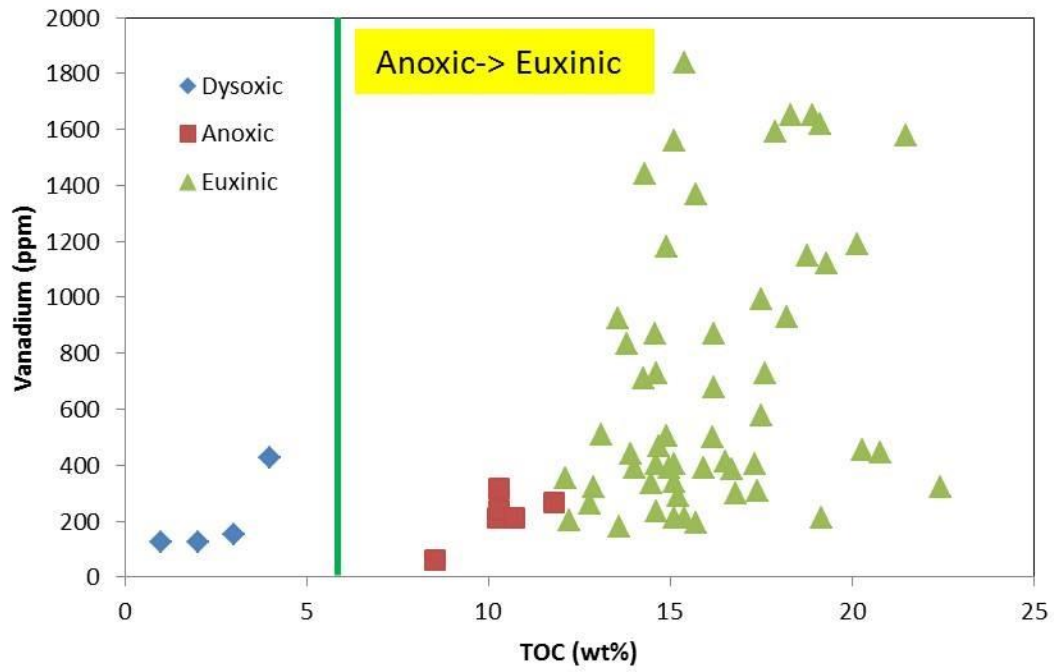
Figure 1-10. Different depositional environments of mudstone (modified after Newton, 2012). Trace element concentrations increase with sulfur level, which becomes the most in the euxinic environment.



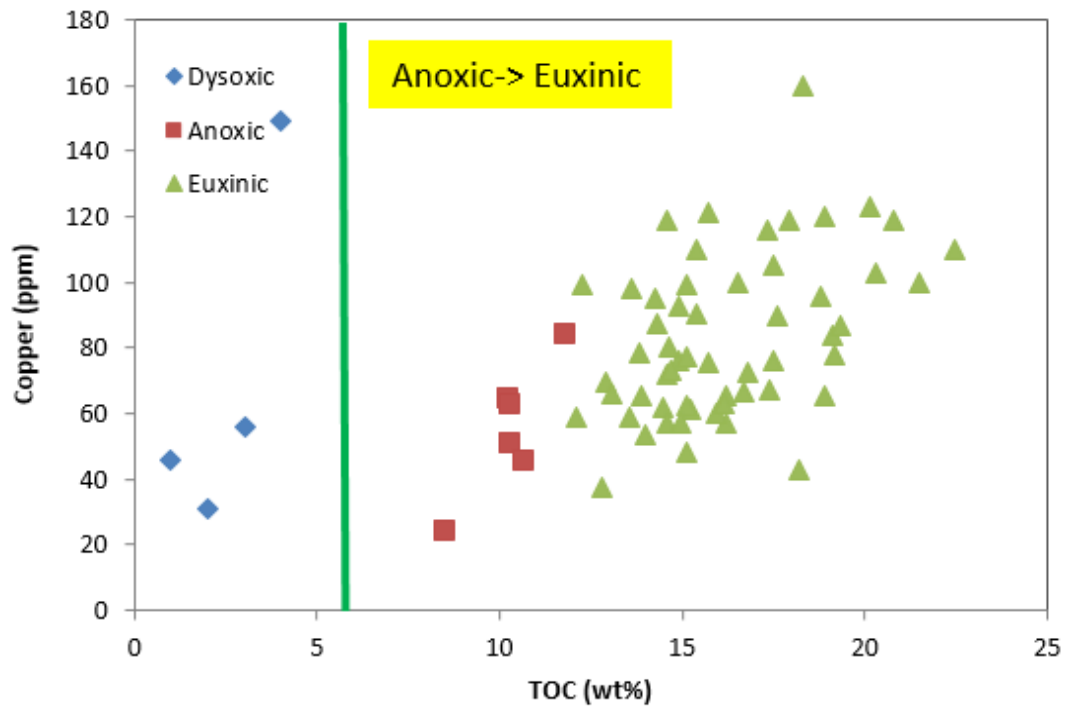
(a)



(b)



(c)



(d)

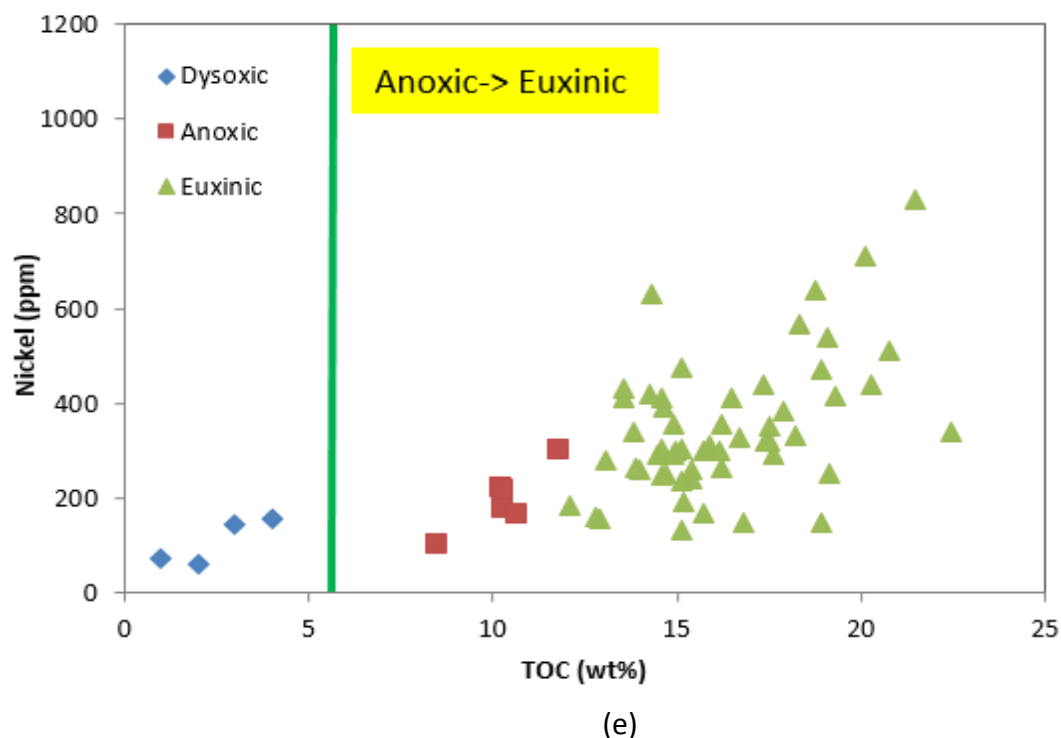


Figure 1-11. (a) through (e) show trace element (molybdenum, uranium, vanadium, copper, and nickel) behavior with TOC in the shale units. Trace element richness is small in the dysoxic zone, whereas it exhibits linear relation with TOC in the anoxic environment. All of the trace elements show highest amount of enrichment and independence of increase in TOC during euxinic environment. Green line (at TOC ~7%) indicates possible boundary of dysoxic and reducing (anoxic and euxinic) environment.

Trace elements generally exhibit higher enrichment in euxinic marine environment when sulfur concentration in the ocean is significantly high (Brumsack, 1986; Dahl et al., 2013). All of the trace elements show strong euxinic behavior when TOC crosses the threshold of 12-13%, whereas they show relatively smaller enrichment when TOC ranges from 7-12% and at relatively lower sulfur level. This pattern can be associated with non-sulphidic anoxic condition. The other pattern where TOC falls far below 7% and at very low sulfur level, shows poor concentration of trace elements (transported by detrital influx), prevailing directly from the top and base of upper

and lower shale members are gray shale members. The observations suggest gray shale units were deposited in relatively higher level of dissolved oxygenated water column or dysoxic condition, whereas the dominant portion of the shale units were deposited in relatively oxygen depleted condition, because redox-sensitive trace elements are less soluble under reducing condition than that of oxic condition based on the behavior of multiple geochemical proxies and TOC (Nandy, 2014; Tribovillard et al., 2006).

It was inferred that three different depositional environments prevailed during Bakken shale deposition, which are dysoxic, anoxic and euxinic. All of the dysoxic, anoxic and euxinic events were sorted out of the core sample database, then Pearson's correlation coefficient was calculated separately for euxinic and anoxic-influenced samples to better understand GR, TOC versus trace element behavior (Tables 1-1 and 1-2). It was observed that presence of silica (Si) decreases GR and TOC responses, whereas clay elements such as aluminum and potassium (Al and K) and trace elements (Mo, U, V, Cu and Ni) enrich them.

Table 1-1. Pearson's correlation coefficients were calculated to understand covariation among GR, TOC, different major and trace elements in the euxinic system. Bold colors indicate highly positive correlation.

	GR	TOC	Si	Al	K	Ti	Mo	V	U	Ni	Cu	S
GR	-	0.49	-0.18	0.52	0.27	-0.04	0.69	0.22	0.55	0.47	0.11	0.2
TOC	0.49	-	-0.28	0.53	0.48	0.09	0.61	0.44	0.51	0.62	0.41	0.77
Si	-0.18	-0.28	-	-0.41	-0.75	-0.2	-0.18	-0.45	-0.35	-0.57	-0.48	-0.6
Al	0.52	0.53	-0.41	-	0.76	0.43	0.43	0.44	0.5	0.6	0.42	0.69
K	0.27	0.48	-0.75	0.76	-	0.46	0.24	0.56	0.38	0.6	0.67	0.72
Ti	-0.04	0.09	-0.2	0.43	0.46	-	-0.13	0.22	-0.006	0.07	0.43	0.66
Mo	0.69	0.61	-0.18	0.43	0.24	-0.13	-	0.36	0.6	0.6	0.33	0.55
V	0.22	0.44	-0.45	0.44	0.56	0.22	0.36	-	0.4	0.67	0.51	0.4
U	0.55	0.51	-0.35	0.5	0.38	-0.006	0.6	0.4	-	0.56	0.41	0.89
Ni	0.47	0.62	-0.57	0.6	0.6	0.07	0.6	0.67	0.56	-	0.48	0.55
Cu	0.11	0.41	-0.48	0.42	0.67	0.43	0.33	0.51	0.41	0.48	-	0.71
S	0.2	0.77	-0.6	0.69	0.72	0.66	0.55	0.4	0.89	0.55	0.71	-

Table 1-2. Pearson's correlation coefficients were calculated to understand covariation among GR, TOC, different major and trace elements in the anoxic system. Bold colors indicate highly positive correlation.

	GR	TOC	Si	Al	K	Ti	Mo	V	U	Ni	Cu
GR	-	0.3	-0.76	-0.52	-0.51	-0.78	0.69	-0.07	0.75	0.55	0.52
TOC	0.3	-	-0.05	0.05	-0.2	-0.52	0.89	0.77	0.84	0.89	0.88
Si	-0.76	-0.05	-	-0.05	-0.08	0.3	-0.4	0.02	-0.54	-0.4	-0.4
Al	-0.52	0.05	-0.05	-	0.94	0.76	-0.2	0.49	-0.13	0.09	0.1
K	-0.51	-0.2	-0.08	0.94	-	0.86	-0.35	0.32	-0.27	-0.03	-0.02
Ti	-0.78	-0.52	0.3	0.76	0.86	-	-0.73	-0.06	-0.71	-0.5	-0.5
Mo	0.69	0.89	-0.4	-0.2	-0.35	-0.73	-	0.52	0.97	0.93	0.91
V	-0.07	0.77	0.02	0.49	0.32	-0.06	0.52	-	0.57	0.75	0.79
U	0.75	0.84	-0.54	-0.13	-0.27	-0.71	0.97	0.57	-	0.94	0.93
Ni	0.55	0.89	-0.4	0.09	-0.03	-0.5	0.91	0.79	0.93	-	1
Cu	0.52	0.88	-0.4	0.1	-0.02	-0.5	0.91	0.79	0.93	1	-

Elemental interrelationships vary considerably in euxinic environment compared to anoxic environment. Pearson's correlation coefficient (r_p) was computed to understand strength of multivariate relationships in anoxic and euxinic systems. Most of the trace elements show significantly strong correlation ($r_p > 0.30$), between themselves (especially, Cu and Ni in anoxic system than that of euxinic system) and with GR and TOC logs also. This pattern manifested by trace elements, TOC, and GR log indicate trace elements residing in the organic matter and getting deposited as organometallic complexes in non-sulphidic anoxic condition, whereas they are drastically enriched in the euxinic condition and independent of increase in TOC concentration after a certain threshold of organic matter richness (TOC~12%). Sulfur content increased during this time and trace elements were directly precipitated in the basin as authigenic phases such as metal sulphides (e.g. pyrite) (Algeo and Maynard, 2004). Pyrites are mostly observed across euxinic events, because of higher sulfur content during that time.

Trace element behaviors match well with GR, TOC and proportions of major stochastic mineralogy such as clay (Figures 1-12 and 1-13). With increase in GR (associated with high TOC), all of the redox-sensitive elements (such as Mo, U, V, Cu and Ni) exhibit positive change in respective concentrations. Integrated multi-well study of well logs and elemental data reveals that upper Bakken is defined by two major patterns (upward decreasing trend followed by an increasing upward pattern), whereas lower Bakken can be characterized by three different patterns such as "upward decreasing", "no visible change", and "upward decreasing" trend of trace element richness, which can be also observed in the gradational changing pattern of GR, TOC, and stochastic mineralogical volumes.

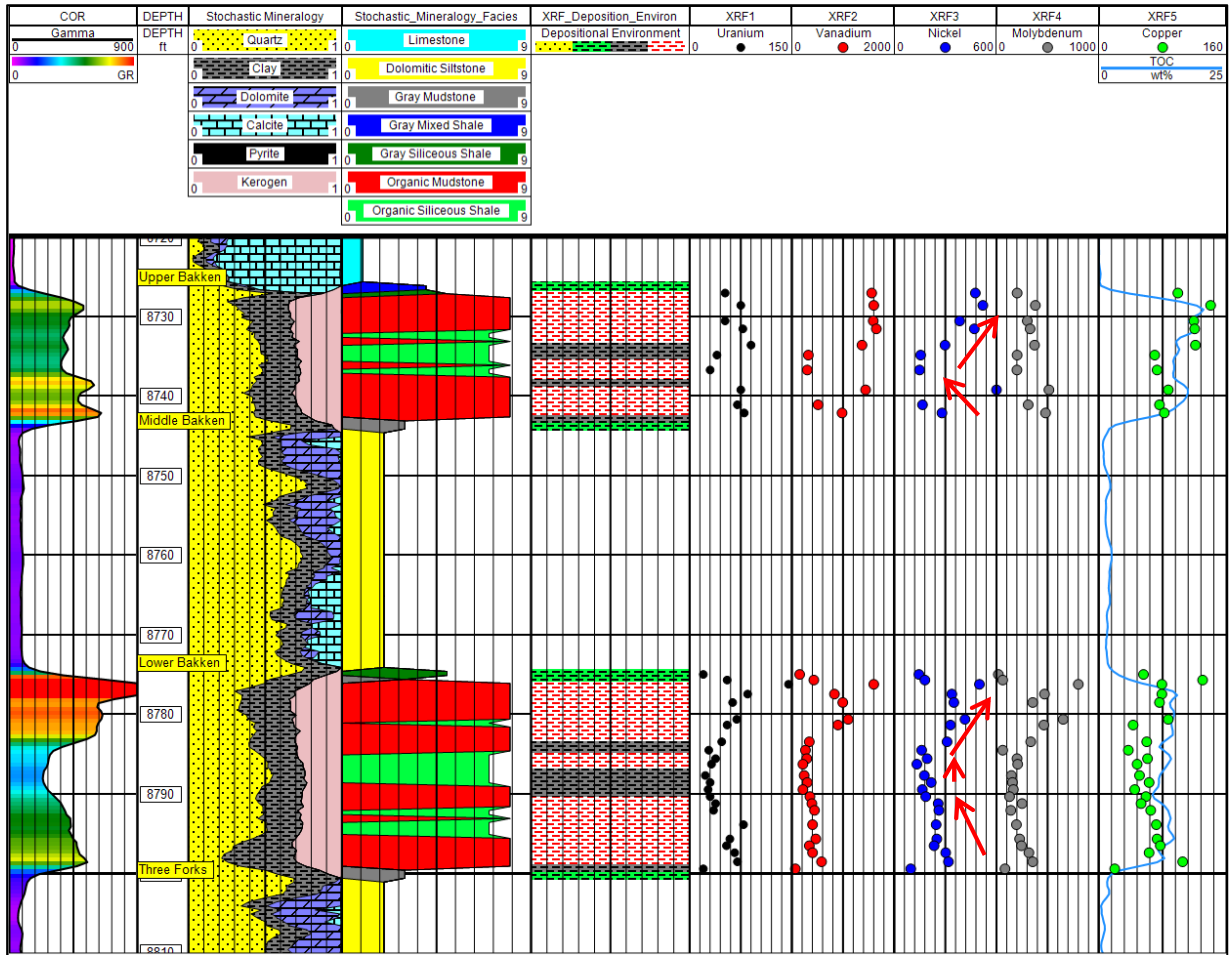


Figure 1-12. Petrofacies distribution, depositional environment interpreted based on major mineralogy, trace element and TOC data in the Well A. The upper and lower Bakken shale members are characterized by trace element patterns (marked by red arrows). Track 1 is Gamma-ray shaded by API, Track 2 is the stochastic mineralogy, Track 3 is the petrofacies from stochastic mineralogy, Track 4 is the depositional environment (yellow, green, gray, and red bars indicating oxic, dysoxic, anoxic, and euxinic conditions respectively), and Tracks 5 through 9 show trace element distribution.

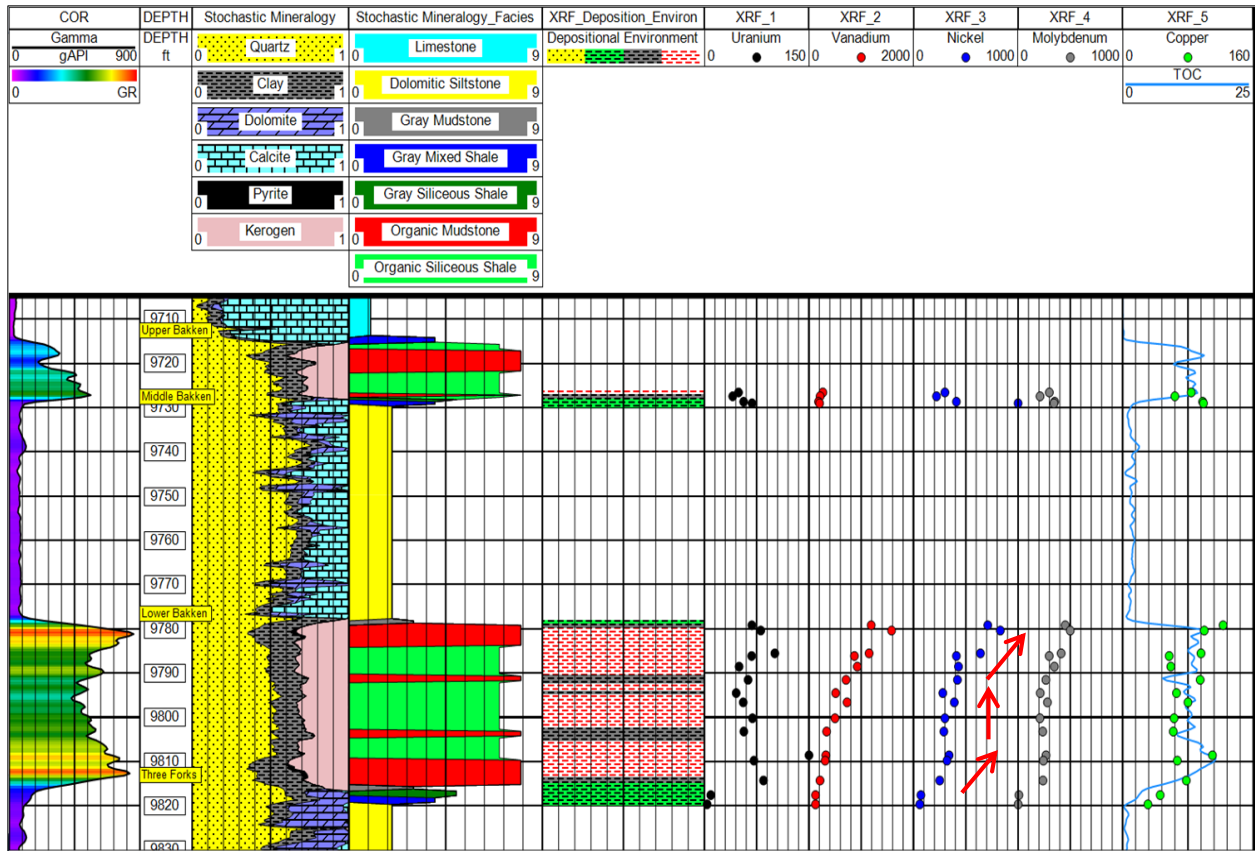


Figure 1-13. Petrofacies distribution, depositional environment interpreted based on major mineralogy, trace element and TOC data in the Well B. A gap in interpreting depositional environment in the upper Bakken portion resulted due to unavailable core data. Track 1 is Gamma-ray shaded by API, Track 2 is the stochastic mineralogy, Track 3 is the petrofacies from stochastic mineralogy, Track 4 is the depositional environment, and Tracks 5 through 9 show trace element distribution.

Based on the above results, five different petrofacies are believed to be present in the upper and lower Bakken shale members, which are associated with characteristic chemostratigraphy. While possible shoulder effects may be present, the boundaries between upper and lower shale units with respective overlying and underlying non-shaly formations appear to fall into broad category of organic-poor (TOC < 7%) gray shale petrofacies (gray siliceous shale, gray mixed shale and gray mudstone), which were interpreted as deposited in dysoxic condition. All of these gray shale facies are bioturbated facies, due to less preservation of kerogen in sediment (possible

effect of burrowing by organisms) and dilution. During this transitional period (between oxic and reducing) relatively higher level of oxygen and lower sea level did not help preservation of kerogen; hence low GR, TOC and poor concentration of trace elements. Apart from gray shale petrofacies units, major portion of both shale members are dominated by cyclic depositional pattern of organic siliceous shale and organic mudstone (TOC > 7%), which are enriched in trace elements and interpreted to be deposited in both anoxic and euxinic environment, indicating higher levels of paleoproductivity and redox condition in the basin.

It was hard to assign particular depositional environment to each of these two organic-rich petrofacies with TOC greater than 7% based on trace element behavior and TOC, because both facies were deposited in anoxic and euxinic environments. So a Chi-square test (Table 1-3) was performed to understand influence of respective depositional environments (category_1) on classified shale petrofacies (category_2). We stated the hypothesis as below:

Null Hypothesis (H_0): Depositional environment and petrofacies are independent.

Alternative Hypothesis (H_a): Depositional environment and petrofacies are not independent.

We constructed a 5x3 contingency table (five petrofacies and three depositional environments) with degrees of freedom equal to 8 [(number of category_1 - 1)*(number of category_2 - 1)]. A robust statistical significance level was chosen (0.0001).

Table 1-3. Chi-square test between core/log derived shale petrofacies and depositional environment for both wells

Facies Environment	Gray Mudstone	Gray Mixed Shale	Gray Siliceous Shale	Organic Siliceous Shale	Organic Mudstone	Total
Euxinic	0	0	0	58	38	96
Anoxic	0	0	0	10	21	31
Dysoxic	4	5	4	0	0	13
Total	4	5	4	68	59	140

The test revealed that Chi-square P value is much less than the chosen significance level of 0.0001; hence H_0 was rejected, which implies deposition of different shale petrofacies is related to corresponding geological environment with a high level of accuracy.

1.5 Conclusions

(1) Both upper and lower shale members of the Bakken Formation are vertically and laterally heterogeneous, in terms of mineralogy and organic matter content. Stochastic mineralogical solutions of both wells reveal presence of five different petrofacies in the shale members, which are recognizable, and predictable at core, log scales, and could potentially provide a methodology to map organic-rich and organic-poor mudrock petrofacies in the shale members of the Bakken Formation at the regional scale. Clay proportion, TOC content and quartz-to-carbonate ratio are the key identifiers for Bakken shale petrofacies classification. The average values of quartz (40-70%), clay (15- 35%) and TOC (10- 15%) in the upper Bakken are different from the average values of quartz (35- 60%), clay (15- 40%) and TOC (10- 17%) in the Lower Bakken.

(2) The uncertainty of stochastic formation evaluation technique can be reduced with the help of existing core and/or PNS logs. If supporting “ground truth” (e.g. core and PNS logs etc.)

information is not available for multi-well studies, proper input log conditioning and STM technique can be applied to compare similarity of corresponding log responses among all wells for classified facies propagation at regional scale.

(3) Application of spectral GR tool proved to be useful to understand high values of total GR response and its association with high kerogen content in shale units, which was also used to determine refined clay volume from well logs.

(4) Geochemical proxies have been tied to log defined shale petrofacies successfully, which was used to interpret depositional environment in detail. Organic-poor (TOC<7%) gray shale members (gray siliceous shale, gray mixed shale and gray mudstone), which exhibit insignificant covariation of GR, TOC, major elements and trace elements were deposited in dysoxic condition, whereas organic-rich samples (TOC>7%), showing high degree of covariation of GR, TOC, major elements and trace elements were deposited in reducing condition. Organic-rich parts of upper and lower shale members indicate cyclic deposition of organic siliceous shale and organic mudstone petrofacies, both of which were deposited in episodic non-sulfidic anoxic (TOC: 7-12%) and euxinic (TOC>12%) condition.

(5) Statistical analyses indicate petrofacies distribution in the Bakken shale members are controlled by depositional environment. Certain elements such as TOC, aluminum, Mo, U, Ni, Cu and V (to an extent) are present in higher concentration in well B compared to well A for both anoxic and euxinic conditions. It may be due to the fact that Bakken shale units in the well B were deposited in a deeper elevation (near the basin center) compared to well A, which can be attributed to prevalent bottom water anoxia suitable for organic-rich black shale deposition and

preservation. The benefit of this detailed analysis by churning out a plethora of geoscientific data is that the three different depositional conditions, owed to changes in paleoenvironment could not be interpreted from well log defined shale petrofacies only and the characteristic association of different well logs with trace element data can be utilized to predict elemental behavior in other wells without core data.

Acknowledgements

We would like to thank Julie LeFever of North Dakota Geological Survey for providing access to well log database and core library for sampling. We thank Shared Research Facilities at West Virginia University, National Energy Technology Laboratory in Morgantown and Actlabs for analytical measurements. Special thanks go to Dr. Harvey Eastman (URS Corporation), Ted Holden (CGG Jason) and Fred Jenson (CGG Jason) for providing helpful suggestions concerning the use of PowerLog™. We would also like to express our thanks to Jason, IHS and Paradigm for providing necessary computing facilities at West Virginia University. The authors wish to thank two reviewers, whose comments help to strengthen the manuscript.

Nomenclature

Al = Aluminum (percent)

Cu = Copper (ppm)

GR = Gamma ray log, gAPI

K = Potassium (percent)

Mo = Molybdenum (ppm)

Ni = Nickel (ppm)

NPHI = Neutron porosity log (fraction)

PEF = Photoelectric log, barns/e

RHOB = Bulk density log, gcm^{-3}

r_p = Pearson's correlation coefficient

S = Sulfur (percent)

Si = Silica (percent)

STM = Similarity Threshold Modeling

Ti = Titanium (percent)

TOC = Total Organic Carbon, weight percent

U = Uranium (ppm)

V = Vanadium (ppm)

Vclay_CGR = Clay volume estimated from uranium free computed gamma ray log

Vclay_Final = Final clay volume estimated from neutron-density crossplot and uranium free
computed gamma ray log

Vclay_N-D = Clay volume estimated from neutron-density crossplot

XRD = X-ray Diffraction

XRF = X-ray Fluorescence

References

Adams, J.A.S., Weaver, C.E., 1958, Thorium to Uranium Ratios as Indications of Sedimentary Processes- example of Concept of Geochemical Facies, AAPG Bulletin, **42**, 387-430.

Algeo, T.J., Maynard, J.B., 2004, Trace-element Behavior and Redox Facies in Core Shales of Upper Pennsylvanian Kansas-type Cyclothems, Chemical Geology, **206**, 289-318.

Bhattacharya, S., Carr, T.R., Wang, G., 2015, Shale Lithofacies Classification and Modeling: Case Studies From the Bakken and Marcellus Formations, North America, presented at the AAPG Annual Convention and Exhibition, Denver, Colorado, USA, 31 May-3 June.

Brumsack, H.-J., 1986, The Inorganic Geochemistry of Cretaceous Black Shales (DSDP Leg 41) in Comparison to Modern Upwelling Sediments from the Gulf of California. In: Summerhayes, C.P., Shackleton, N.J. (Eds.), North Atlantic Paleocenography, Geol. Soc. Spec. Publ., **21**, 447 – 462.

Dahl, T.W., Ruhl, M., Hammarlund, E.U., Canfield, D.E., Rosing, M.T., Bjerrum, C.J., 2013, Tracing Euxinia by Molybdenum Concentrations in Sediments using X-ray Fluorescence Spectroscopy (HHXRF), Chemical Geology, **360-361**, 241-251.

Doveton, J.H., 1994, Geologic Log Interpretation, SEPM Short Course Notes No. **29**, 123-125.

Dumonceaux, G., 1984, Stratigraphy and Depositional Environments of the Three Forks Formation (Upper Devonian), Williston Basin, North Dakota, Unpublished M.S. Thesis, University of North Dakota, Grand Forks.

Egenhoff, S.O., Fishman, N.S., 2013, Traces in the Dark—sedimentary Processes and Facies Gradients in the Upper Shale Member of the Upper Devonian–Lower Mississippian Bakken Formation, Williston Basin, North Dakota, U.S.A, Journal of Sedimentary Research, **83**, 803-824.

Hickey, J.J., Henk, B., 2007, Lithofacies Summary of the Mississippian Barnett Shale, Mitchell 2 T.P. Sims Well, Wise County, Texas, AAPG Bulletin, **91**, 437-443.

LeFever, J.A., 1991, History of Oil Production from the Bakken Formation, North Dakota, *in* Hanson, W.B., ed., 1991 Guidebook to Geology and Horizontal Drilling of the Bakken Formation: Billings, Montana Geological Society, 3–17.

LeFever, J. A., LeFever, R.D., Nordeng, S.H., 2011, Revised Nomenclature for the Bakken Formation (Mississippian-Devonian), North Dakota, *in* Robinson, J.W., LeFever, J.A., and Gaswirth, S.B., eds., The Bakken-Three Forks Petroleum System in the Williston Basin: Denver, Colo., Rocky Mountain Association of Geologists, 11–26.

Kulyapin, P., Sokolova, T.F., 2014, A Case Study about Formation Evaluation and Rock Physics Modeling of the Bazhenov Shale, *Petrophysics*, **55**(3), 211-218.

Maldonado, D.N., 2012, Chemostratigraphy and Geochemical Constraints on the Deposition of the Bakken formation, Williston Basin, eastern Montana and western North Dakota, Master's Thesis, University of Texas at Arlington, Arlington, Texas.

Meissner, F.F., 1978, Petroleum Geology of the Bakken Formation, Williston Basin, North Dakota and Montana, *in* Estelle, D., and Miller, R., eds., The Economic Geology of the Williston Basin, 1978 Williston Basin Symposium, Billings, Montana: Montana Geological Society, 207–230.

Mitchell, W.K., Nelson, R.J., 1988, A Practical Approach to Statistical Log Analysis, Paper S. Transactions, presented at the SPWLA 29th Annual Logging Symposium, San Antonio, Texas, USA, 5-8 June.

Moss, B., Harrison, R., 1985, Statistically Valid Log Analysis Method Improves Reservoir Description, Paper SPE-13981, presented at the OEC, Aberdeen, UK, 10-13 September.

Nandy, D., Sonnenberg, S., Humphrey, J.D., 2014, Application of Inorganic Geochemical Studies for Characterization of Bakken Shales, Williston Basin, North Dakota and Montana, Paper URTEC-1922974, presented at the Unconventional Resources Technology Conference, Denver, USA, 25-27 August, DOI 10.15530/urtec-2014-1922974.

Newton, R., 2012, Water Column Redox Proxies, presented at Life and Planet-NERC.

Pitman, J.K., Price, L.C., LeFever, J.A., 2001, Diagenesis and Fracture Development in the Bakken Formation, Williston Basin: Implications for Reservoir Quality in the Middle Member, U.S. Geological Survey Professional Paper 1653, U.S. Department of the Interior/U.S. Geological Survey, 19.

Pollastro R.M., Roberts, N.R., Cook, T.A., 2013, Geologic Assessment of Technically Recoverable Oil in the Devonian and Mississippian Bakken Formation, Chapter 5, http://pubs.usgs.gov/dds/dds-069/dds-069_w/contents/REPORTS/69_W_CH_5.pdf (accessed April 3rd, 2014).

Rowe, H.D., Hughes, N., Robinson, K., 2012, The Quantification and Application of Handheld Energy-dispersive X-ray Fluorescence (ED-XRF) in Mudrock chemostratigraphy and Geochemistry, *Chemical Geology*, **324-325**, 122-131.

Savre, W.C., 1963, Determination of a More Accurate Porosity and Mineral Composition in Complex Lithologies with the use of the Sonic, Neutron, and Density Surveys, *Journal of Petroleum Technology*, **15** (6), 945-959.

Schieber, J., 1999, Distribution and Deposition of Mudstone Facies in the Upper Devonian Sonyea Group of New York, *Journal of Sedimentary Research*, **69**, 909-925.

Schieber, J., Zimmerle, W., Sethi, P.V., 1998a, Shales and Mudstones, Volume 1: Basin Studies, Sedimentology, and Paleontology, E. Schweizerbart'sche Verlagsbuchhandlung, Stuttgart, Germany.

Schieber, J., Zimmerle, W., Sethi, P.V., 1998b, Shales and Mudstones, Volume 2: Petrography, Petrophysics, Geochemistry, and Economic Geology, E. Schweizerbart'sche Verlagsbuchhandlung, Stuttgart, Germany.

Schmoker, J.W., Hester, T.C., 1983, Organic Carbon in Bakken Formation, United States Portion of Williston Basin, *AAPG Bulletin*, **67**, 2165–2174.

Sesack, S.A., 2011, Sequence Stratigraphy, Depositional Environments, and Regional Mapping of the Late Devonian Interval, Upper Three Forks Formation, Sanish Member, and Lower Bakken Shale, U.S. Portion of the Williston basin, MS Thesis, West Virginia University, Morgantown, West Virginia.

Simenson, A., 2011, Depositional Facies and Petrophysical Analysis of the Bakken Formation, Parshall field, Mountrail County, North Dakota, MS Thesis, Colorado School of Mines, Golden, Colorado.

Smith, M.G., Bustin, R.M., 2000, Late Devonian and Early Mississippian Bakken and Exshaw Black Shale Source Rocks, Western Canada Sedimentary Basin: A sequence stratigraphic interpretation, *AAPG Bulletin*, **84**, 940-960.

Sonnenberg, S.A., Pramudito, A., 2009, Petroleum Geology of the Giant Elm Coulee Field, Williston basin, AAPG Bulletin, **93**, 1127–1153.

Steptoe, A., 2012, Petrofacies and Depositional Systems of the Bakken Formation in the Williston Basin, North Dakota, MS Thesis, West Virginia University, Morgantown, West Virginia.

Thevoux-Chabuel, H., Veillerette, A., Rabiller, P., 1997, Multiwell Log Data Coherence Characterization using the Similarity Threshold Method, presented at the SPWLA 38th Annual Logging Symposium Transactions, Houston, USA, 15-18th June.

Tribovillard, N., Algeo, T.J., Lyons, T., Riboulleau, A., 2006, Trace Metals as Paleoredox and Paleoproductivity proxies: An Update, Chemical Geology, **232**, 12-32.

Wang, G., 2012, Black shale lithofacies prediction and distribution pattern analysis of Middle Devonian Marcellus Shale in The Appalachian Basin, Northeastern U.S.A., Ph.D. Dissertation, West Virginia University, Morgantown, West Virginia.

Warner, T., 2011, Subsurface Horizontal Microfracture Propagation within the Middle Member of the Bakken Formation, Williston basin, North Dakota: Evidence and Implications, MS Thesis, West Virginia University, Morgantown, West Virginia.

Webster, R.W., 1984, Petroleum Source Rocks and Stratigraphy of the Bakken Formation in North Dakota, *in* Woodward, J., Meissner, F.F., and Clayton, J.L., eds., Hydrocarbon Source Rocks of the Greater Rocky Mountain Region, Rocky Mountain Association of Geologists, 57–81.

CHAPTER 2

Comparison of Supervised and Unsupervised Approaches for Mudstone Lithofacies Classification: Case Studies from the Bakken and Mahantango-Marcellus Shale, USA

Shuvajit Bhattacharya¹, Timothy R. Carr¹, and Mahesh Pal²

¹Department of Geology and Geography

West Virginia University, Morgantown

WV- 26506, USA

(304)-777-9885

Email: subhattacharya@mix.wvu.edu; tim.carr@mail.wvu.edu

²Department of Civil Engineering

National Institute of Technology Kurukshetra

India- 136119

Abstract

Quantitative lithofacies modeling is important to understand the depositional and diagenetic history, and hydrocarbon potential of unconventional resources at a regional scale. The complex heterogeneous nature and large data dimensionality of unconventional mudstone reservoirs increase the challenge of lithofacies interpretation by conventional qualitative methods. Quantitative shale lithofacies, which are meaningful, mappable, and predictable at core, well log, and regional scales, can be defined based on mineralogy and Total Organic Carbon (TOC) derived from core analysis and advanced geochemical spectroscopy logs (e.g. Pulsed Neutron Spectroscopy, PNS). However, access to numerous and widespread core samples and geochemical log responses is typically limited by cost and time.

We apply different mathematical techniques to ubiquitous conventional well log suites calibrated to rock types, defined by the limited number of wells with high-quality core and geochemical logs. The documented interrelationships between lithofacies and conventional logs are propagated with a quantified degree of accuracy in wells without advanced log or core data. Our study addresses issues of different approaches of quantitative lithofacies classification and prediction techniques from well logs. Various data-driven supervised and unsupervised computational approaches, such as Support Vector Machine (SVM), Artificial Neural Network (ANN), Self-Organizing Map (SOM) and Multi Resolution Graph-based Clustering (MRGC), are applied and compared to reduce uncertainty of propagating single-well based lithofacies analysis, and efficiently understand geological trends.

Two different dataset from the Devonian Bakken and Mahantango-Marcellus Shale formations in North America are used, in order to undertake a comparative assessment of

computational techniques for lithofacies characterization. Original shale lithofacies, defined from geochemical logs and core data, are used to compare the results of selected supervised and unsupervised computational approaches. The results show that both Bakken and Mahantango-Marcellus shale members are vertically and laterally heterogeneous, but can be classified into at least five mudstone lithofacies, along with calcareous siltstone and limestone lithofacies. SVM works better than other techniques for lithofacies classification and prediction in reduced computational time, no iteration, and with highly repeatable results. Accuracy of lithofacies prediction increases if the algorithms are supervised with geological rules.

Keywords: Quantitative modeling, Organic mudstone, Lithofacies, Mineralogy, Well logs, Machine learning

2.1 Introduction

This paper discusses application of different computational techniques in mudstone lithofacies modeling. Lithofacies modeling or assigning a rock type to specific rock samples on the basis of petrography or measured physical properties, is fundamental to many subsurface investigations. Even though various clastic and carbonate facies have been studied in detail for depositional and diagenetic environment study, research in mudstone lithofacies is relatively rare. Mudstone is heterogeneous, in terms of variable mineralogy and organic matter content. Two approaches have been followed for mudstone facies modeling. The first approach deals with either single borehole study or descriptive core and outcrop sample analysis (Egenhoff and Fishman, 2013; Schieber, 1999; Singh, 2008), whereas the second approach, which is more recent, is pursued from a quantitative perspective (Bhattacharya et al., 2015; Qi and Carr, 2006; Wang, 2012). In this study we choose two world class Devonian mudstone formations in North America such as the Bakken and Mahantango-Marcellus formations. There have been extensive qualitative studies on mineralogical composition, facies description, and depositional history in both of the shale formations (Egenhoff and Fishman, 2013; Emmanuel and Sonnenberg, 2013; LeFever et al., 2011).

However, it is not always possible to have access to widespread and abundant core, advanced well logs, due to economic and logistic constraints, which poses significant challenges in correct lithofacies interpretation, facies identification, and reservoir characterization. Outcrop shale sample studies can prove to be futile, due to weathering and chance of kerogen degradation over time. Therefore, we emphasize quantitative approaches over qualitative approaches, which contribute to only visual idea about lithology. Quantitative approaches can be used to classify

lithofacies using conventional wireline logs at borehole scale, and propagated to multi-well studies to predict lithofacies, and measure accuracy of prediction with available original lithofacies model, derived from core and/or advanced well logs.

In recent years several mathematical techniques have been proposed for facies and petrophysical property modeling (Al-Anaji and Gates, 2010; Qi and Carr, 2006; Wang et al., 2014). There is also a significant bias to use conventional well log based cutoff values to build facies models quickly. Such deterministic and simplified cutoff values may not work well in complex mudstone formations (Schlanser et al., 2014). In addition, well log cutoff values can be misleading in regional studies, while using them without normalization. Hence, there is a need to compile mathematically powerful machine learning algorithms together, to assess their individual capacity for predictive lithofacies modeling, and define a protocol for using them in specific geological scenarios. We claim that a few quantitative techniques can learn complex subsurface geological pattern better than others.

This paper compares different mathematical methods to classify and predict mudstone lithofacies from ubiquitous conventional well logs, in terms of reducing uncertainty in exploration. Different supervised and unsupervised machine learning algorithms including Support Vector Machine (SVM), Artificial Neural Network (ANN), Self-Organizing Map (SOM), and Multi Resolution Graph-based Clustering (MRGC) are used and discussed in detail, with corresponding advantages and disadvantages, to design a suitable schematic workflow for intelligent mudstone lithofacies pattern recognition. We first introduce mathematical principles of all four algorithms, followed by actual lithofacies characteristics in the dataset, and then use them to classify and predict lithofacies. The result shows that both Bakken and Mahantango-

Marcellus shale members are heterogeneous, but can be classified into at least five mudstone lithofacies, along with calcareous siltstone and limestone lithofacies. Supervised approaches, especially SVM, work better than other techniques for lithofacies classification and prediction purposes in reduced computational time, no iteration, and with highly repeatable results.

2.2 Principles Of Mathematical Algorithms

2.2.1 Support Vector Machine (SVM)

SVM is one of the fastest growing approaches of computational intelligence, in terms of data classification and prediction in all scientific disciplines. It is based on solid mathematical background of statistical learning theory (Kordon, 2010; Vapnik, 1995).

The idea behind SVM is to map the original data from input space to a higher dimensional or even infinite-dimensional feature space so that distance between each data points gets increased, and classification among different variables into different clusters becomes simpler (Luts et al., 2010). The mapping is done by a suitable choice of kernel function.

The key feature of SVM is support vectors (Figure 2-1). Support vectors are the data points (i.e. samples), which lie on the boundaries of different classes (such as different lithofacies) during classification. There can be a large number of hyperplanes, which can discriminate between two classes, out of which SVM tries to find the optimal hyperplane, which is the farthest from both classes. For two-class problems, it assumes two planes that support each class, and maximizes the distance (called margin) between them. Optimization problem involves pushing these parallel planes apart until they collide with data points, representative of each class (called

support vectors). Therefore, SVM uses a significantly small portion of the training dataset (Pal and Foody, 2012).

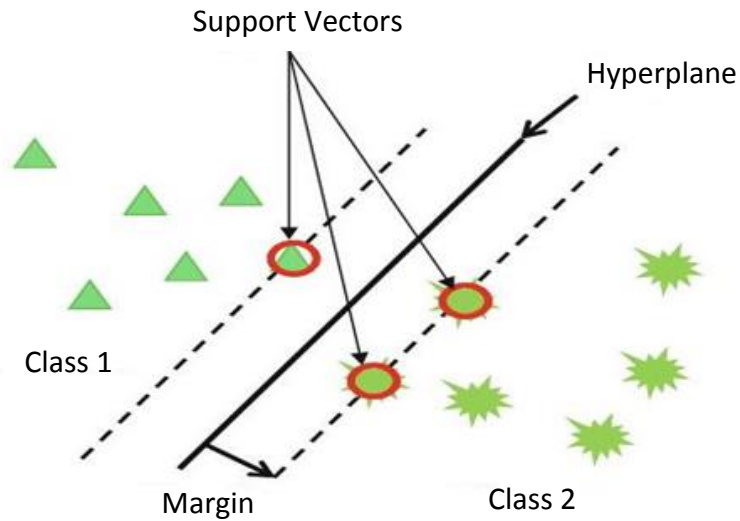


Figure 2-1. An optimal hyperplane separating two classes of data in SVM model (Kordon, 2010)

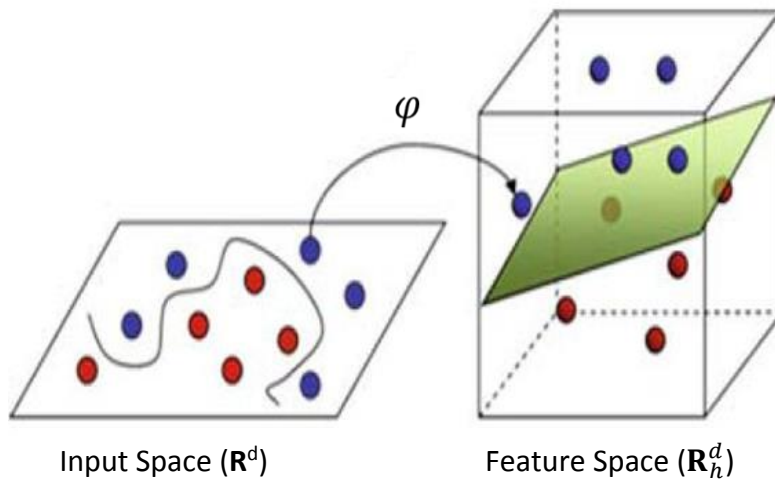


Figure 2-2. SVM kernel function maps data from input space to feature space (modified after Kordon, 2010)

Consider a training dataset $\{x_i, y_i\}_{i=1}^N$, with $x_i \in \mathbf{R}^d$ being the input vectors and $y_i \in \{-1, +1\}$ the class labels.

SVM maps the input vector x from the input space to d^h -dimensional feature space

using a function $\varphi(\cdot): \mathbf{R}^d \rightarrow \mathbf{R}_h^d$ (Figure 2-2). The separating hyperplane or classification function (indicator function) in the feature space is defined as

$$w^T \varphi(x) + b = 0 \quad (1)$$

where $b \in \mathbf{R}$. Parameters b and w determine the offset from origin, and the orientation of discriminating plane respectively. A data point x is assigned to the class 1 (or lithofacies 1) if $f(x) = \text{sign}(w^T \varphi(x) + b) \geq +1$ or to class 2 (or lithofacies 2) if $f(x) \leq -1$.

However, in most of the real cases, perfect linear separation is impossible due to overlapping classes, especially in geology, which is why a restricted number of misclassifications are tolerated around the margin. The resulting optimization problem for SVM is written as

$$\min_{w, b, \xi_1, \dots, \xi_k} \left[\frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \xi_i \right]$$

$$\text{such that } y_i(w^T \varphi(x) + b) \geq 1 - \xi_i, \quad i = 1, \dots, n \quad (2)$$

$\xi_i \geq 0, i=1, \dots, n$ where C is a positive regularization constant (or penalty parameter).

The first term in the above equation corresponds to minimization of the quadratic program to maximize the margin, and the second term relates to empirical error or penalty (Kecman, 2005; Luts et al., 2010). Larger C assigns a higher penalty to errors.

In case of non-linear data modeling, SVM maps data to a significantly large feature space to make them linearly separable and it uses a kernel functions to solve the optimization problem in that space. Four type of kernel functions are generally used: linear, polynomial, Radial Basis

Function (RBF), and multilayer perceptron (Cortes and Vapnik, 1995; Tan et al., 2015; Wang et al., 2014).

Although originally built as a binary classifier, SVM algorithm can be used in multi-class geological pattern recognition problems. SVM handles the multi-class problems efficiently by one-against-all and pairwise comparison (one-against-one) techniques (Hastie and Tibshiranu, 1998; Li et al., 2003). Pairwise comparison method leads to larger number of simpler binary classification, compared to one-against-all method (Wang et al., 2014).

2.2.2 Artificial Neural Network (ANN)

ANN is a popular machine learning algorithm, which is used extensively in data classification and prediction studies (Bishop, 1995). ANN attempts to mimic some of the basic information processing methods in human brain (Kordon, 2010; McCulloch and Pitts, 1943).

Generally, ANN consists of three layers: input layer, hidden layer and output layer, all of which are connected via artificial neurons (Figure 2-3). The input layer receives the incoming data, which is distributed to the hidden layer. Hidden layer is the key part of the ANN structure that learns data structure, in terms of patterns and interrelationship among input variables, and then distributes the learned data patterns (mathematically expressed as weight) to the output layer. The output of a node is controlled by an activation function, which is basically a function of total input parameters and a threshold, which determines the initiation of output (Doveton, 1994).

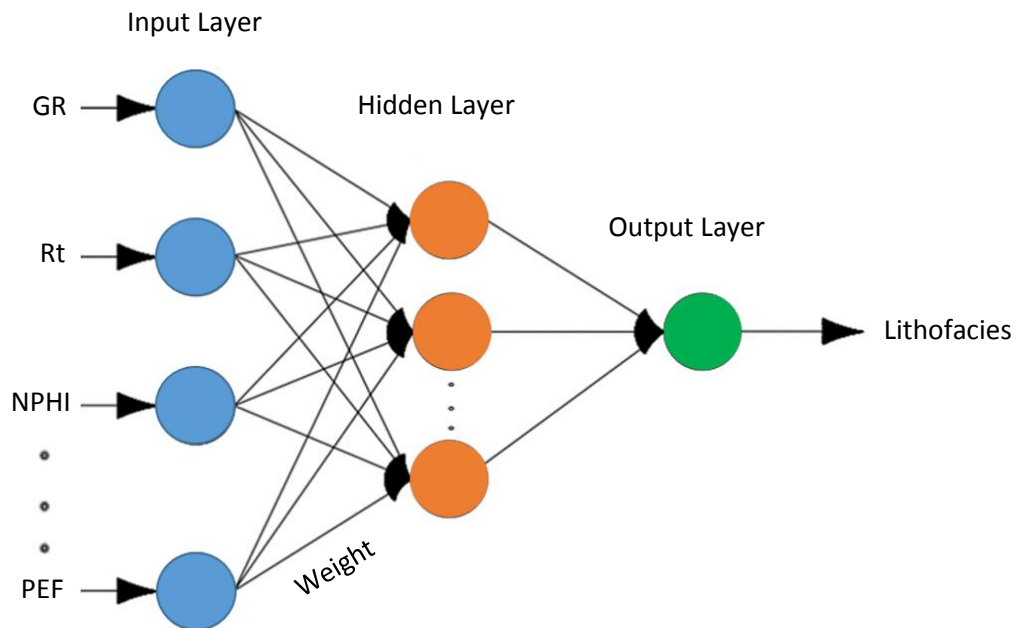


Figure 2-3. Schematic architecture of an ANN algorithm where input layer corresponds to different well logs and output refers to different lithofacies (modified after Manshad et al., 2015)

ANN modeling starts with randomly assigned weight coefficients. Then a set of data patterns is fed forward repeatedly and the weights of the neurons are modified until the output matches closely with the actual values (Doveton, 1994). Training is usually accomplished by “back-propagation technique” in iterative manner, while minimizing the error between ANN computed output and target output.

There are several parameters to control, while designing ANN such as number of hidden layers, number of hidden layer nodes, learning rate, damping coefficient or momentum, and number of iterations for better optimization.

2.2.3 Self-Organizing Map (SOM)

SOM was first introduced by Kohonen (1982, 2001), as a technique for clustering and reduction of data dimensionality. SOM clusters data in such a way that the statistical relationship

among multidimensional input dataset is converted to a lower dimensional SOM grid space without losing topological properties among data points (Roy, 2013). SOM is different from ANN, in the sense that the former preserves geometrical relationships of input data (Venkatesan and Mullai, 2014).

Consider a 2D SOM represented by PV (prototype vectors) m_i , $m_i = [m_{i1}, m_{i2}, \dots, m_{iN}]$, where N represents dimensions of prototype vectors. After arranging the PVs in a lower-dimensional 2D grid space, SOM model is trained and tested with an available dataset.

During SOM training process, an input vector (x) is randomly chosen from a set of input vectors, then the Euclidean distance between this vector and all other vectors (m_i , $i = 1, 2, \dots, n$) are computed. According to Kohonen (2001), the prototype vector (m_b), with minimum distance from the input vector, is called the Best Matching Unit (BMU), which can be expressed as:

$$\|x - m_b\| = \min\{\|x - m_i\|\} \quad (3)$$

Then, the BMU prototype vector m_b , and other input vectors within a neighborhood radius of σ are updated as a part of the training. The rule of updating the weight of prototype vectors inside and outside the neighborhood radius is given by (Kohonen, 2001; Roy, 2013)

$$\begin{aligned} m_i(t+1) &= m_i(t) + \alpha(t)h_{bi}(t)[x - m_i(t)] \text{ if } \|r_i - r_b\| \leq \sigma(t) \\ &= m_i(t) \text{ if } \|r_i - r_b\| > \sigma(t) \end{aligned} \quad (4)$$

where r_b and r_i are position vectors of m_b and m_i respectively. The neighborhood function is defined as $h_{bi}(t)$ where t represents the length of training. The neighborhood radius decreases with successive iteration and BMUs are brought closer to the input vectors, finally forming different clusters. Next, the trained prototype vectors are color coded for better visualization of various clusters.

2.2.4 Multi Resolution Graph-based Clustering (MRGC)

MRGC is an unsupervised data classification technique, which can solve the dimensionality problem and obtain valuable information about the input dataset (Ye and Rabiller, 2000). It is based on K-nearest neighbor and graph based presentation. Most of the other techniques require exact number of clusters prior to training, whereas MRGC automatically clusters the data, and yet allows the user to either lump or subdivide different clusters based on “ground truth” at the end.

MRGC describes data pattern by two indices: Neighboring Index (NI) and Kernel Representative Index (KRI). NI measures how close the data points are with respect to each other in the input domain. It is computed by mutual rank of neighboring data points, which can be expressed as

$$\begin{aligned} NI(x) &= \exp(-m/\alpha) \quad (5) \\ &= 1 \text{ if } m=0 \\ &\cong 0 \text{ otherwise} \end{aligned}$$

Here x is the m^{th} nearest neighbor of y , which is n^{th} nearest neighbor of x in the dataset, $m \leq N - 1$, and $\alpha > 0$.

However, it is important to recognize optimal number of clusters, otherwise infinite number of clusters can be constructed, including local irregularities, which is not correct. Hence, the other index KRI measures the probability of each data point to be a representative cluster kernel (Ye and Rabiller, 2000). KRI can be expressed as

KRI (x) = NI(x) M(x,y) D(x,y), where M(x,y) = M if y is the mth neighbor of x, and D(x,y) is the distance between x and y. Data points corresponding to the highest KRI values are assigned to form final clusters or lithofacies.

2.3 Bakken and Mahantango-Marcellus Shale Lithofacies

We chose the Devonian Bakken and Marcellus formations in North America to apply, and compare different quantitative techniques for mudstone lithofacies modeling. Both Bakken and Marcellus formations are big unconventional oil and gas plays in North America. The Bakken Formation in the Williston basin spans over portions of North Dakota, South Dakota, Montana, and extends in to Canada, whereas the Mahantango-Marcellus formations in the Appalachian basin is present in West Virginia, Pennsylvania, and New York etc. in the United States (Figure 2-4).

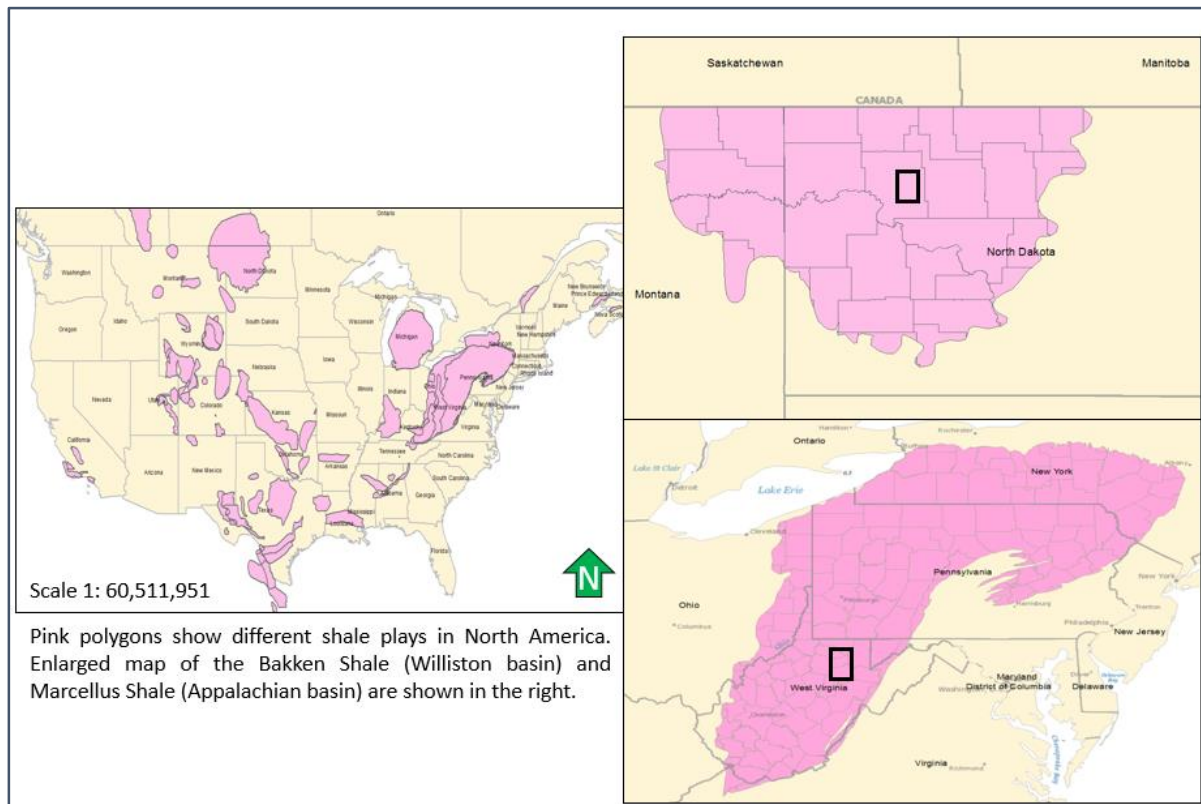


Figure 2-4. Study areas in the Williston and Appalachian basins marked by black rectangles

The Bakken Formation consists of two world class source rocks (upper Bakken and lower Bakken shale) that sandwich the middle Bakken member, which is composed of mixed lithologies, including sandstone, dolostone and limestone. The Mahantango-Marcellus interval is composed of gray shale and organic rich shale units with carbonate interlayers.

Original shale lithofacies models, based on core geochemical data (XRD and TOC) and advanced PNS logs, are available for the Bakken and Mahantango-Marcellus mudstone intervals (Bhattacharya et al., 2015; Wang, 2012). Both shale members are heterogeneous; they are rich in quartz and clay, compared to carbonate (Figure 2-5).

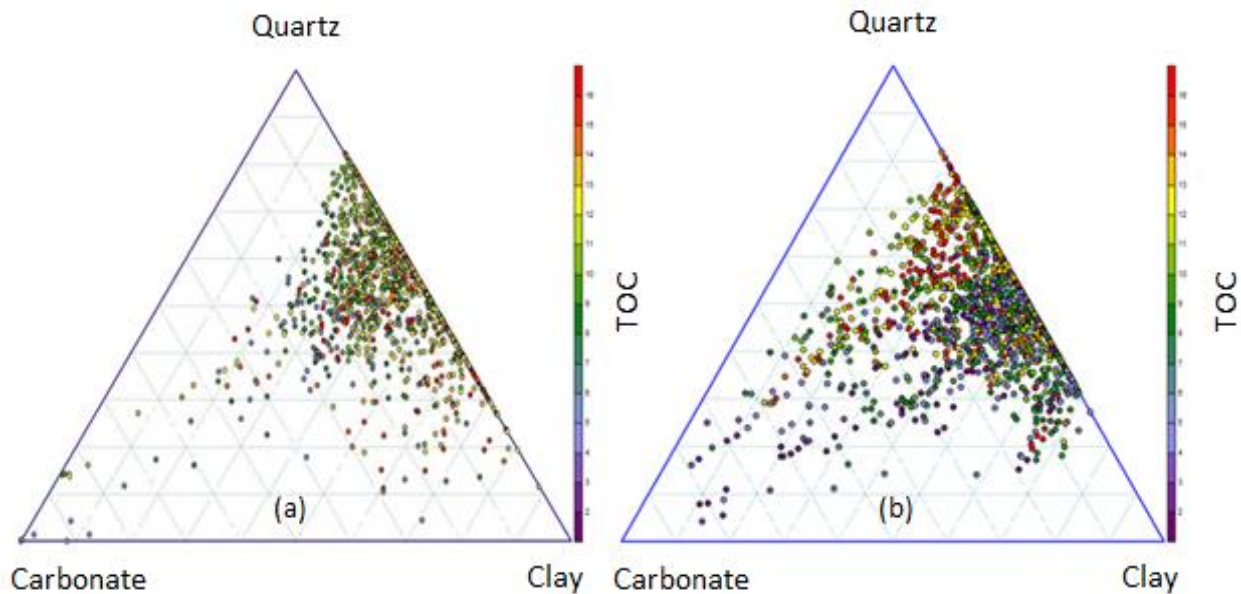


Figure 2-5. Ternary diagrams (a and b) showing variation in major mineralogy and TOC content (all in percentage) in the Bakken and Marcellus shale formations respectively. Both shale formations are siliceous mudstone (modified after Bhattacharya and Carr, 2016).

We apply a schematic methodology to classify the shale lithofacies in a quantitative manner (Bhattacharya and Carr, 2016). We use three criteria such as clay volume, TOC content, and

quartz-to-carbonate ratio to classify different lithofacies (Figure 2-6). We use clay volume of 30% in the Bakken dataset, whereas 40% cutoff is used in the Mahantango-Marcellus dataset, calibrated to core-based ground truth information. Both Bakken and Mahantango-Marcellus shale members are composed of at least five different lithofacies such as Organic Mudstone (OMD), Organic Siliceous Shale (OSS), Gray Mudstone (GMD), Gray Siliceous Shale (GSS), and Gray Mixed Shale (GMS), along with limestone and mixed lithology (calcareous siltstone). Organic Mixed Shale lithofacies is absent from the available dataset, however, the schematic methodology can be used to evaluate other mudstone formations. The non-shaly middle Bakken member is composed of mixed lithology (siltstone, dolostone, and limestone), which is considered as a single lithofacies unit for the scope of this study. OSS lithofacies has high silica and high organic matter content, which make it amenable to effective hydraulic stimulation and hydrocarbon production. Figure 2-7 and Figure 2-8 show depth displays of various conventional well logs, mineralogical analysis, and original lithofacies model in the Bakken and Mahantango-Marcellus intervals.

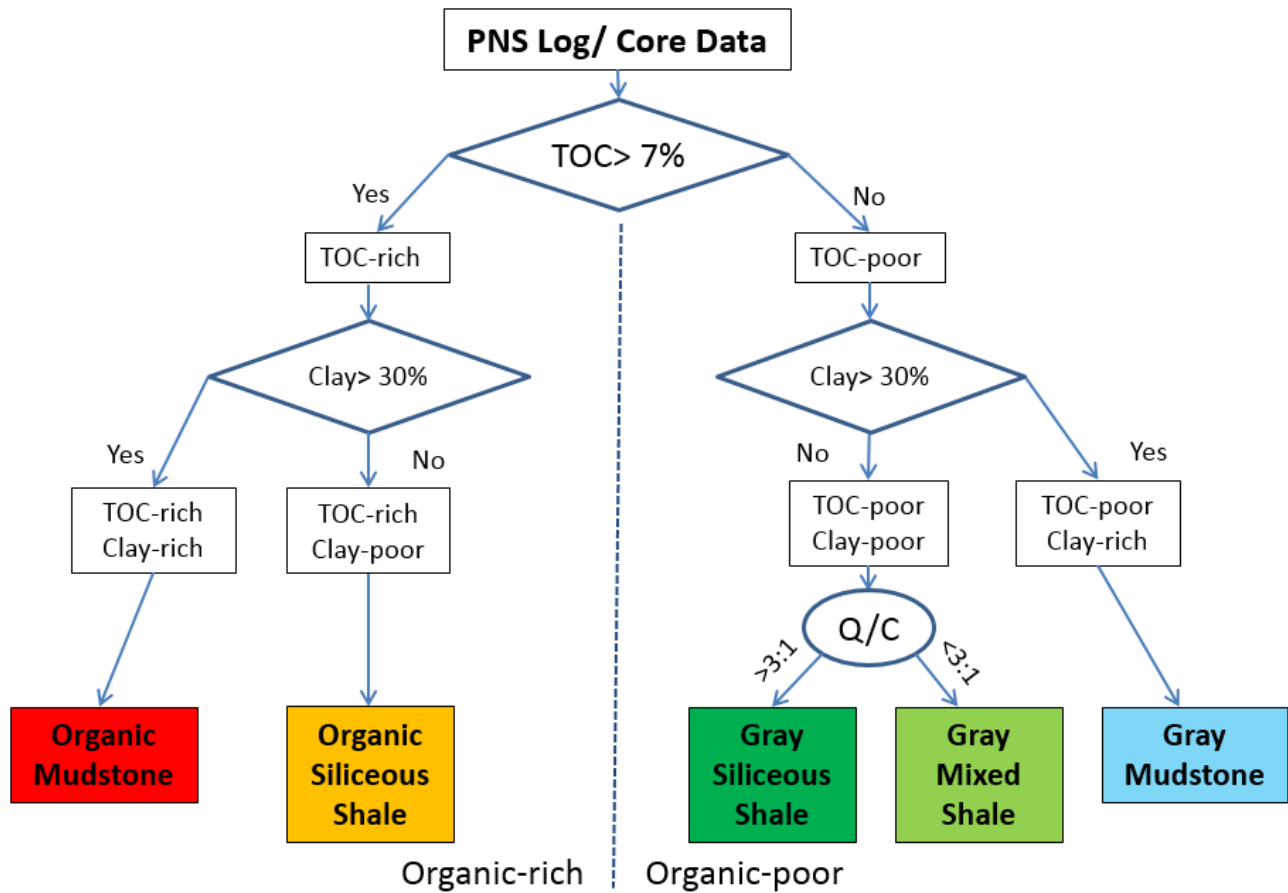


Figure 2-6. Schematic methodology to classify shale lithofacies. Three criteria such as clay volume, TOC content, and quartz-to-carbonate ratio (Q/C) are used to classify shale lithofacies quantitatively (modified after Bhattacharya and Carr, 2016).

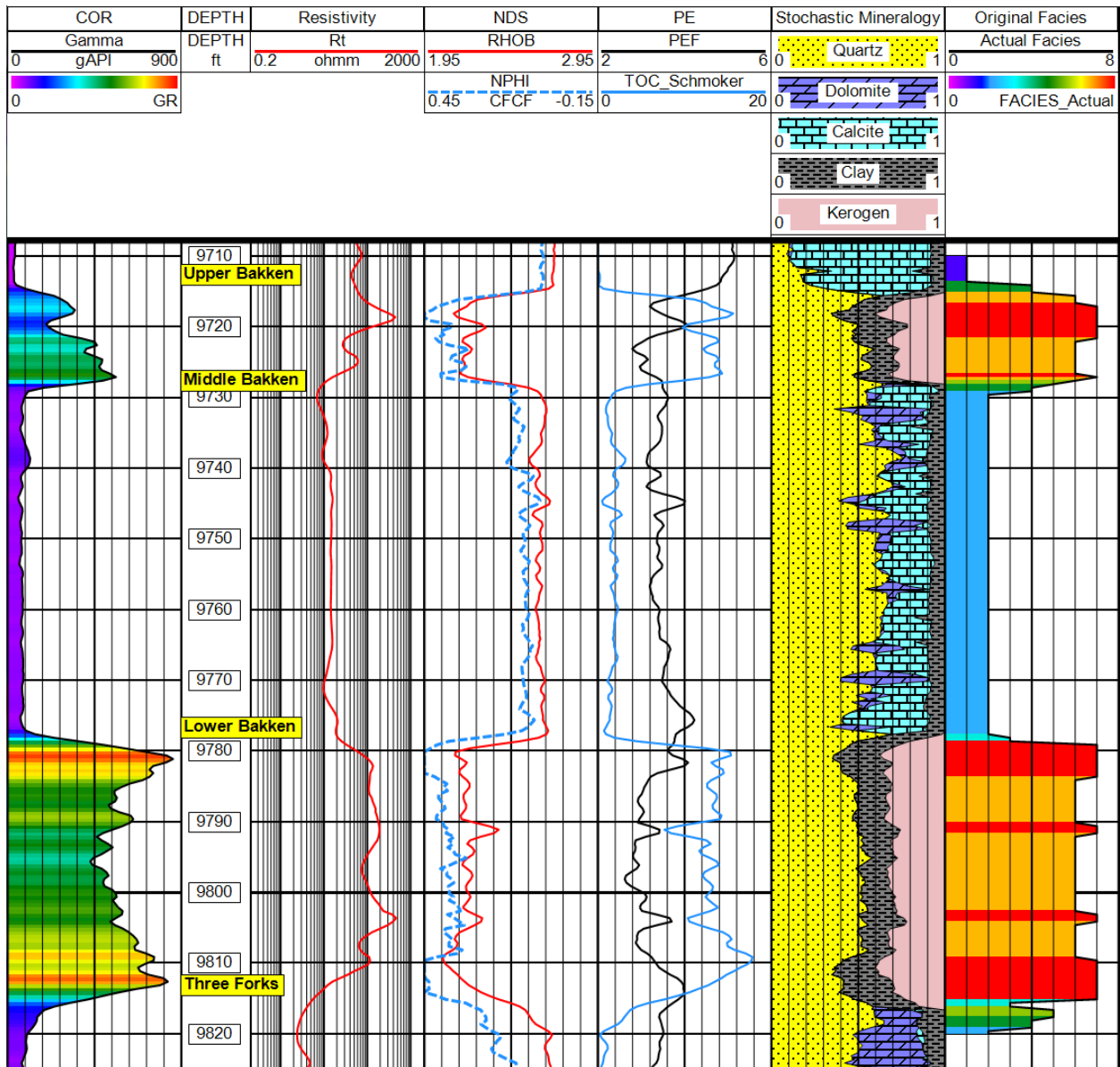


Figure 2-7. Integrated well log and original lithofacies model display in one of the training wells in the Bakken dataset. Upper Bakken and lower Bakken shale members are composed five different mudstone lithofacies, based on mineralogy and organic matter richness. Limestone, mixed lithology (calcareous siltstone), GMD, GMS, GSS, OSS, and OMD lithofacies are represented by numerical codes from 1 to 7 in a rainbow pattern.

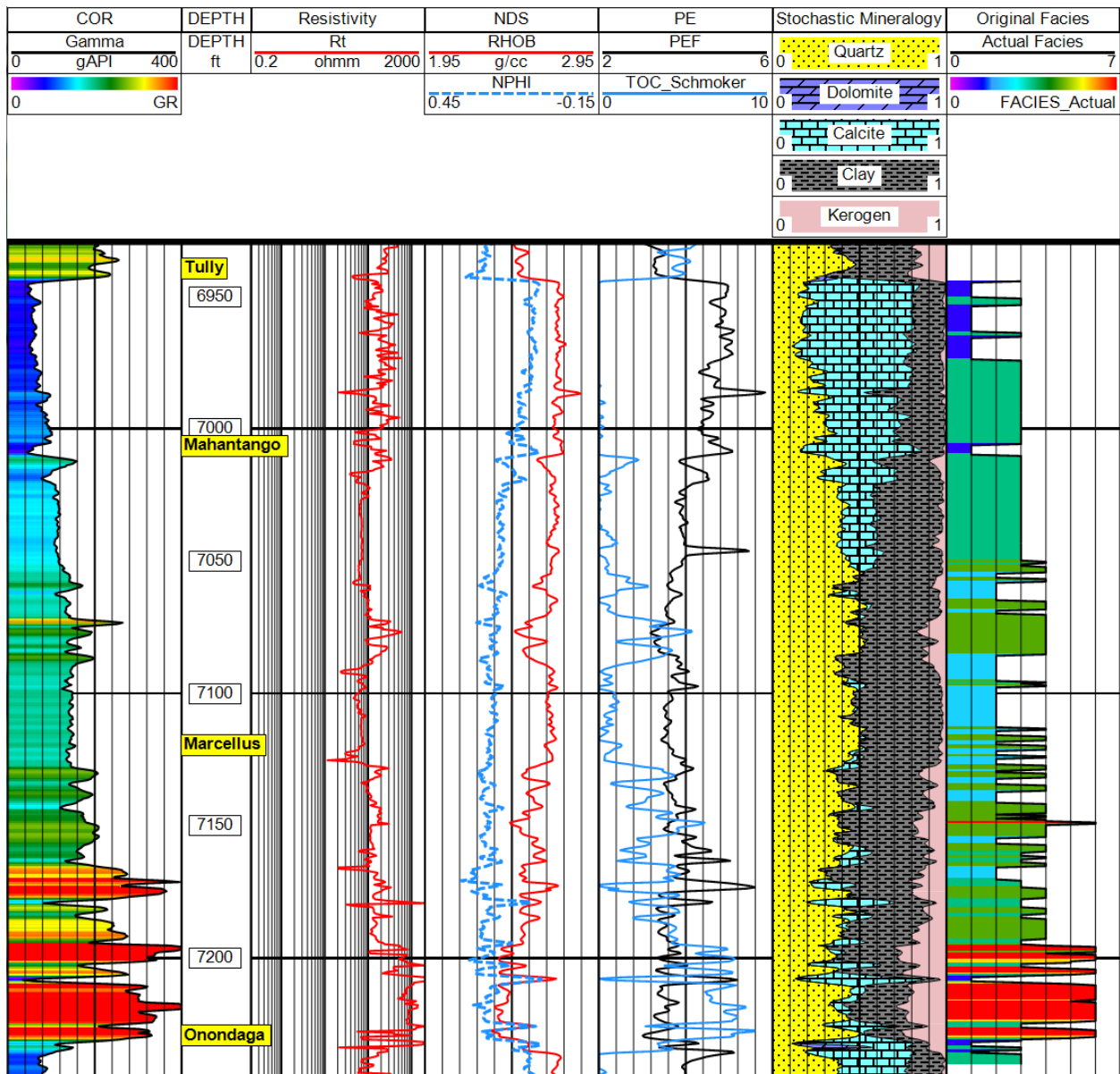


Figure 2-8. Integrated well log and original lithofacies model display in one of the training wells in the Mahantango-Marcellus interval. The Mahantango interval is mostly composed of lower order gray shale facies, whereas the Marcellus Formation is mostly composed of different types of organic shale lithofacies, along with carbonate interlayers.

2.4 Materials and Methods

Two different dataset containing conventional well log suites, such as gamma, resistivity, neutron porosity, density, and photoelectric factors, are compiled from the Devonian Bakken and

Mahantango-Marcellus shale formations, in order to undertake a comparative assessment of computational techniques for lithofacies characterization. The Bakken dataset consists of petrophysical logs from three wells, whereas the dataset from the Mahantango-Marcellus interval contains similar data from two wells. In general, the petrophysical log data are obtained from around 100 feet (30.48 meters) thick Bakken interval, whereas the data interval is of around 200 feet (60.96 meters) in the Mahantango-Marcellus formations. Each well log is recorded at half-foot (0.15 meters) increment.

We compare, and investigate the capabilities of all four quantitative algorithms (SVM, ANN, SOM, and MRGC) in shale lithofacies classification and prediction, using well logs from the Bakken and Mahantango-Marcellus Shale formations as examples. The emphasis of the assessment is on the potential for value creation in subsurface geologic analysis by reducing uncertainties in lithofacies pattern recognition. The whole methodology includes: (1) pre-processing input petrophysical dataset; and (2) training and testing the classifiers for lithofacies modeling.

2.4.1 Pre-processing Input Petrophysical Dataset

Pre-processing input petrophysical logs is an important step to construct a reliable lithofacies model. This process involves proper selection of well logs, which are facies-sensitive. Apart from that, well logs should be normalized and environmentally corrected based on regional geology. For example, resistivity logs should be corrected for hydrocarbon effects, sonic log should be free from overburden related compression effects near basin center, and neutron porosity logs should be free from clay bound-water effects, so that facies models generated with these logs will be independent of any fluid and normalization related issues.

We choose five conventional well logs such as gamma (GR), natural logarithm of deep resistivity (LnRt), neutron porosity (NPHI), bulk density (RHOB), and photoelectric factor (PEF), and five other derived parameters such as apparent matrix density (RHOMaa), photoelectric absorption index (Umaa), gamma/density (GR/RHOB), photoelectric/density (PEF/RHOB), and TOC as input for lithofacies modeling (Figure 2-9). Shale lithofacies modeling is a multi-class problem; it is better to use more input parameters to increase the average distance between data points, thereby forming easily separable lithofacies clusters. GR, LnRt, and RHOB logs are used as input petrophysical parameters, because organic shale has higher uranium concentration (associated with high GR), higher resistivity, and lower density, compared to organic-poor lithofacies. NPHI log is used as it can help to distinguish clay-rich intervals from clay-poor intervals, whereas PEF log is utilized to separate carbonate and non-carbonate layers. Other parameters such as RHOMaa, Umaa, GR/RHOB, and PEF/RHOB are utilized to enhance the evaluation of high silica content in the shale formations, detect carbonate interlayers, and remove parts of the noise if any. TOC parameter is used as a quantitative measure of kerogen richness in the mudrocks, which can distinguish organic-rich shale from organic-poor gray shale. Kerogen has low density (~1.1 g/cc). TOC is derived from RHOB log using Schmoker's method (1983).

$$\text{TOC} = (154.497/\text{RHOB}) - 57.261 \quad (6)$$

$$\text{RHOMaa} = (\text{RHOB} - \text{PHIA}) / (1 - \text{PHIA}) \quad (7)$$

$$\text{Umaa} = (\text{PEF} * \text{RHOB} - 0.5 * \text{PHIA}) / (1 - \text{PHIA}) \quad (8)$$

where PHIA corresponds to average porosity from NPHI and RHOB logs.

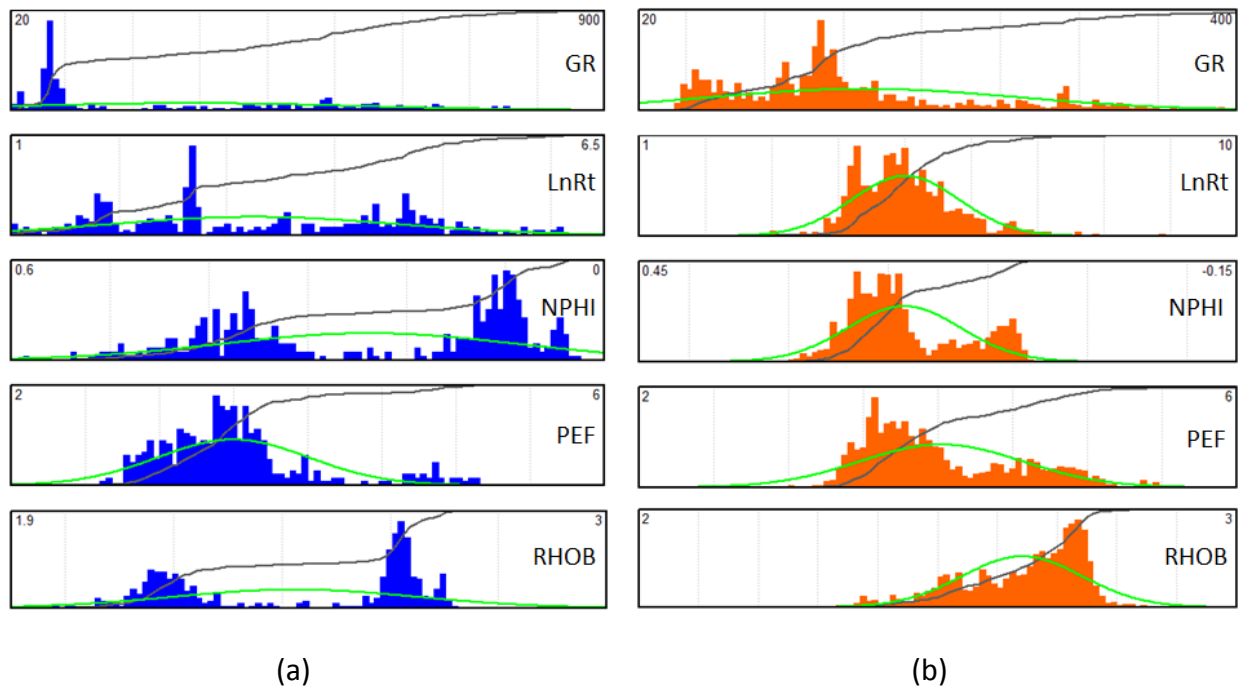


Figure 2-9. Histograms showing statistical feature of conventional well log suites (as model input) in the Bakken (a) and Mahantango-Marcellus (b) dataset. The green curve indicates Gaussian distribution and the black curve indicates cumulative distribution.

2.4.2 Training and Testing the Classifiers

All three machine learning algorithms (SVM, ANN, and SOM) need to be set up with optimum network parameters to enable them work efficiently for lithofacies training and testing. A large number of trials with various combination of user-defined parameters are performed to train the classifiers. Geolog™ and WEKA programs are used to train and test petrophysical dataset, and PowerLog™ to display well logs and lithofacies model. We used an Intel computer with 32 GB RAM and CPU@3.6 GHz for this study.

For SVM classifier, choice of kernel function is significantly important. Literature review in different research areas suggests that RBF function performs quite well in comparison to other kernels (Camps-Valls, 2009; Hsu et al., 2003), and hence used in this study. RBF kernel based SVM

classifier requires fine tuning of two user-defined parameters: gamma (kernel parameter), and penalty parameter (C).

We apply back propagation ANN algorithm with one hidden layer with both datasets. Optimal values of learning rate, momentum, and epoch (i.e. number of iterations) obtained after large number of trials is used for enhanced petrophysical data classification and testing. In case of SOM, we use 2D SOM method, with square grid of neurons (five in each side).

In case of Bakken dataset, SVM, ANN and SOM algorithms are trained and tested on dataset obtained from different wells for facies prediction. Whereas, a 10-fold cross-validation is used with the Marcellus dataset, due to smaller sample size. With unsupervised MRGC method, first a number of clusters are formed to generate a high-resolution lithofacies model, which is then lumped into a coarser model based on similarity of data statistics in different clusters, and the original lithofacies model. Results from SVM, ANN and SOM are represented in form of confusion matrices and overall classification accuracy.

2.5 Results and Discussions

Results from both Bakken and Mahantango-Marcellus dataset show that SVM method is the best lithofacies classifier in training and testing domains. This performance is attributed to the RBF kernel. Application of RBF function in the SVM classifier generated best possible result. Because, RBF kernel can efficiently map non-linear data to a higher-dimension feature space, and solve the problem of data classification.

In the Bakken dataset SVM accuracy of data prediction is 87.3%, compared to ANN (83.7%), SOM (84.03%), and MRGC (71.3%), shown in the confusion matrices (Tables 2-1, 2-2, 2-3, and 2-

4). The best values of gamma and C in the SVM model are found to be 1.2 and 25 respectively. The optimum values of learning rate and momentum in the ANN algorithm are 0.3 and 0.8 respectively. For 2D SOM a lower dimensional square grid (5x5) is chosen (Figure 2-10). In case of MRGC, first a high-resolution lithofacies model is built up with 16 clusters, which is then lumped into relatively coarser resolution lithofacies model of seven classes to match with the original lithofacies model.

Table 2-1. Confusion matrix illustrates accuracy of SVM algorithm in lithofacies prediction in the Bakken test dataset. Overall accuracy by SVM is 87.3%.

		Actual						
		Lithofacies	1	2	3	4	5	6
Predicted	1	25	3	0	0	0	0	0
	2	0	140	0	3	2	0	0
	3	0	0	6	0	2	1	1
	4	1	0	0	1	3	0	0
	5	0	0	0	0	1	0	0
	6	0	0	0	0	0	45	23
	7	0	0	0	0	0	0	50
	Absolute Accuracy	96.15%	97.90%	100%	25%	12.50%	97.82%	67.60%

Table 2-2. Confusion matrix illustrates accuracy of ANN algorithm in lithofacies prediction in the Bakken test dataset. Overall accuracy by ANN is 83.7%.

		Actual						
		Lithofacies	1	2	3	4	5	6
Predicted	1	23	0	0	0	0	0	0
	2	2	125	1	1	2	0	0
	3	1	13	2	1	1	0	0
	4	0	5	1	2	1	0	0
	5	0	0	2	0	4	1	0
	6	0	0	0	0	0	44	17
	7	0	0	0	0	0	1	57
Absolute Accuracy		88.46%	87.41%	33%	50%	50.00%	95.65%	77.03%

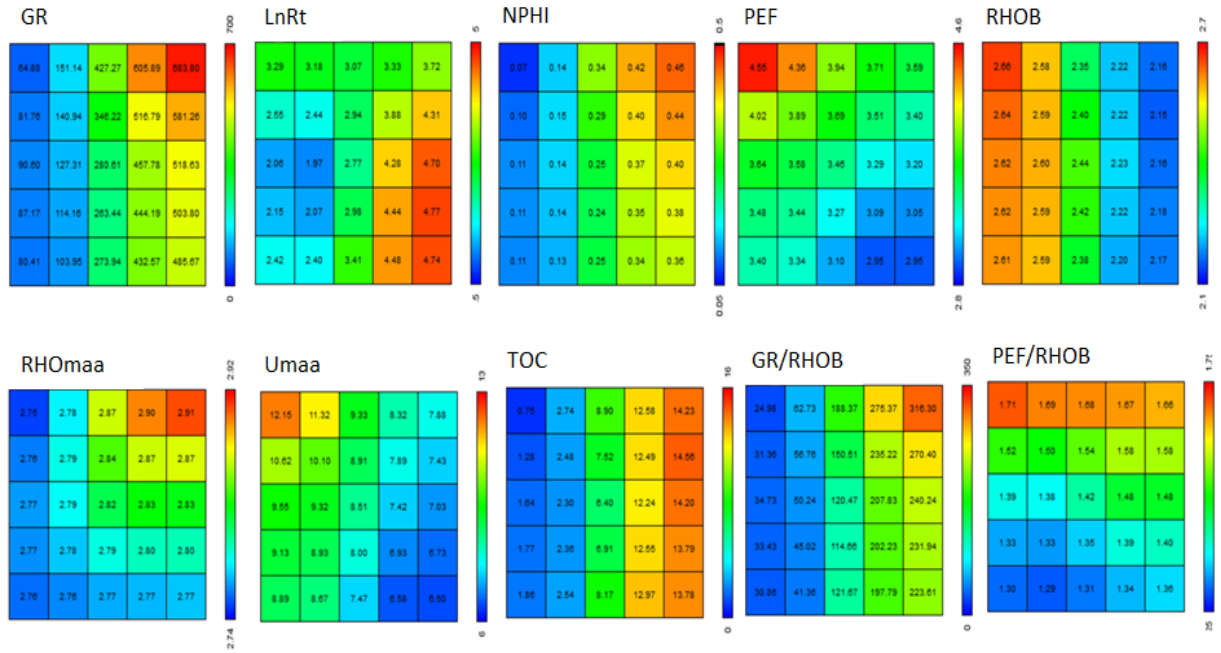
Table 2-3. Confusion matrix illustrates accuracy of 2D SOM algorithm in lithofacies prediction in the Bakken test dataset. Overall accuracy by SOM is 84.03%.

		Actual						
		Lithofacies	1	2	3	4	5	6
Predicted	1	26	5	0	0	1	0	0
	2	0	138	0	4	5	0	0
	3	0	0	0	0	1	0	0
	4	0	0	0	0	0	0	0
	5	0	0	6	0	1	1	6
	6	0	0	0	0	0	41	16
	7	0	0	0	0	0	4	52
Absolute Accuracy		100.00%	96.50%	0%	0%	0%	89.13%	70.27%

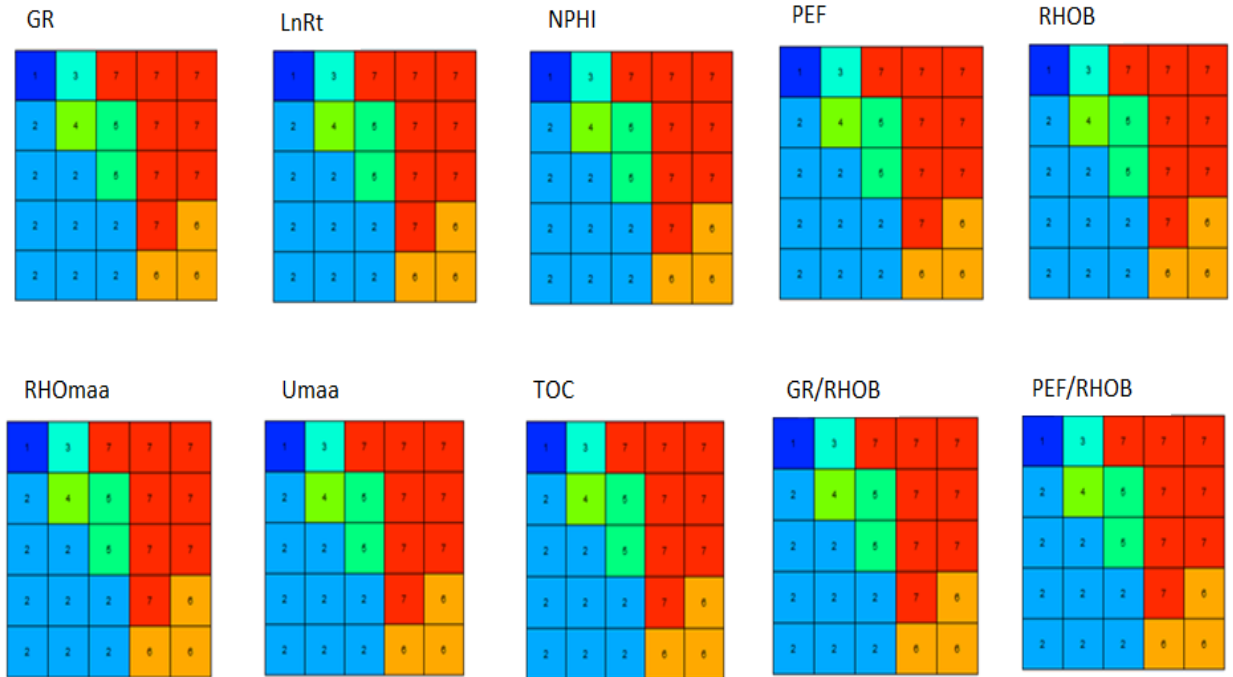
Table 2-4. Confusion matrix illustrates accuracy of MRGC algorithm in lithofacies prediction in the Bakken test dataset. Overall accuracy by MRGC is 71.3%.

		Actual						
		Lithofacies	1	2	3	4	5	6
Predicted	1	26	4	0	0	1	0	0
	2	0	98	0	0	3	0	0
	3	0	33	0	1	0	0	0
	4	0	8	0	0	0	0	0
	5	0	0	6	3	4	1	2
	6	0	0	0	0	0	38	19
	7	0	0	0	0	0	7	53
	Absolute Accuracy	100.00%	68.53%	0%	0%	50.00%	82.60%	71.62%

Figure 2-11 and Figure 2-12 show visual comparison of four different facies models, with respect to the original lithofacies model during training and testing phases. It appears that all four algorithms can be trained to understand the well log data pattern associated with lithofacies distribution significantly well, however, most of them fail to predict lithofacies accurately during testing.



(a)



(b)

Figure 2-10. An example of 2D SOM maps representing all ten input parameters in lower-dimensional grid space (five neurons in each side) by representative log values (a) and lithofacies (b) respectively in the Bakken dataset

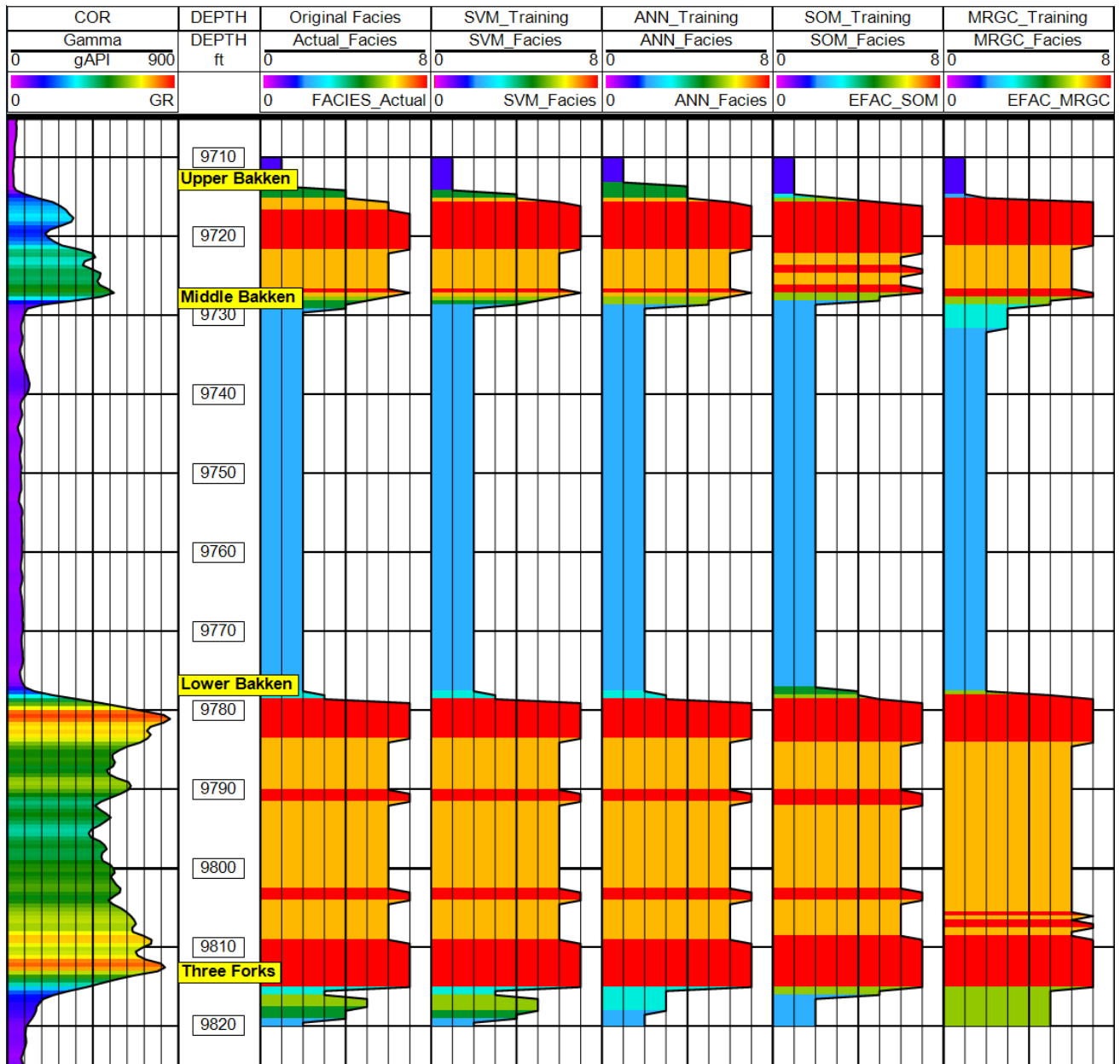


Figure 2-11. The Bakken interval lithofacies are classified in a training well. The second track shows original lithofacies, and third, fourth, fifth, and sixth track are showing training results from four different algorithms. Lithofacies models classified by all three supervised techniques (SVM, ANN, and SOM) look significantly similar to the original lithofacies model.

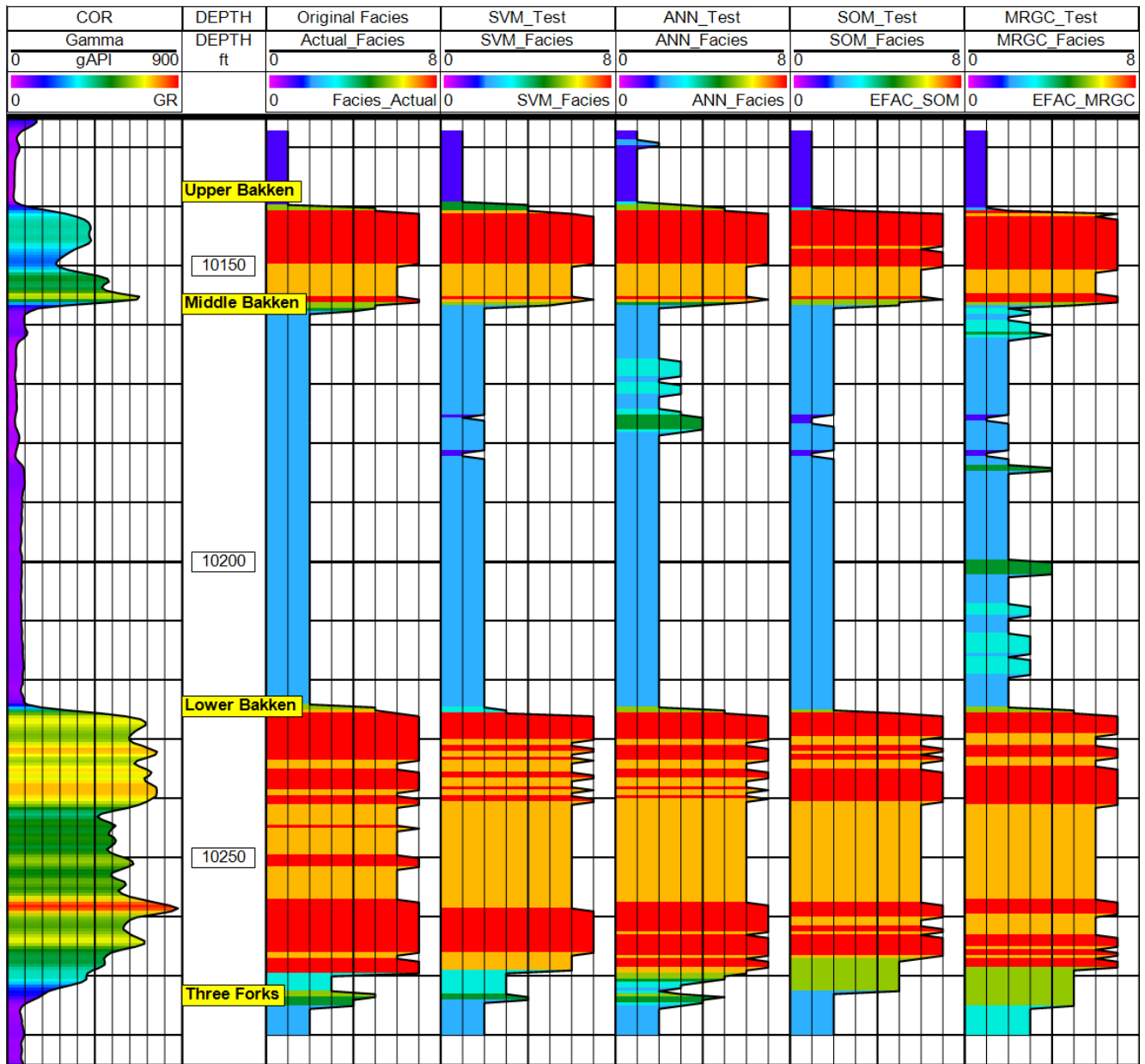


Figure 2-12. The Bakken interval lithofacies are classified in a test well. The second track shows original lithofacies, and four other tracks showing predicted lithofacies results from four different algorithms.

In the Marcellus dataset, SVM achieves accuracy of 82.46% for lithofacies classification, which is much higher than that of ANN (78.75%), SOM (69.65%), and MRGC (64.42%). The optimal values of gamma (i.e. 4.1) and C (i.e. 20) are used with SVM classifier with this dataset. The optimal parameters used with ANN and SOM are found to be the same as used in the Bakken

dataset. In case of MRGC, first a high-resolution lithofacies model is built up with 18 clusters, which is then merged to six classes, matching with the original lithofacies model. A comparative assessment of all four quantitative techniques is shown in Figure 2-13 and confusion matrices (Tables 2-5, 2-6, 2-7, and 2-8). Figure 2-14 summarizes lithofacies prediction results by four algorithms in both dataset.

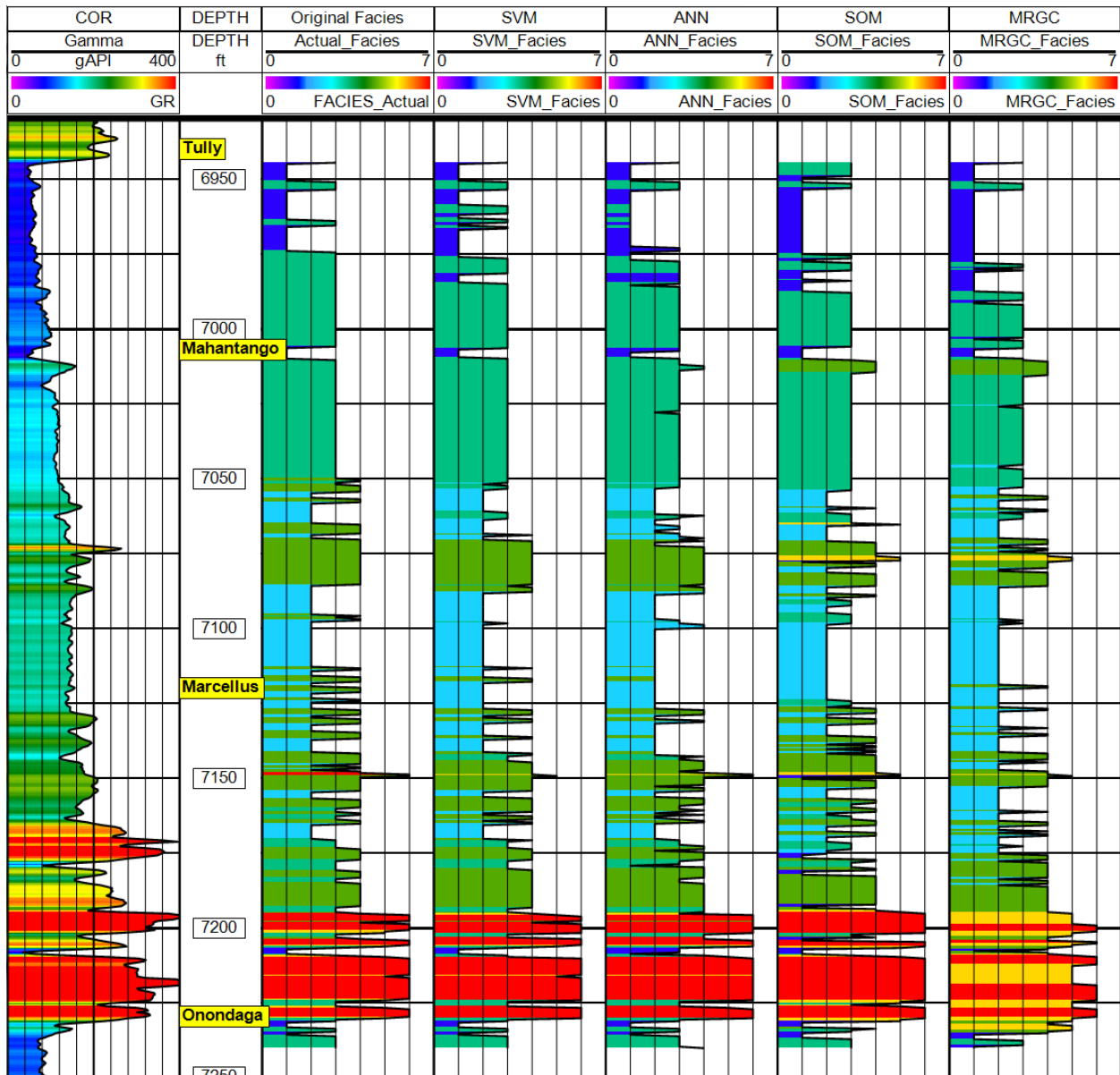


Figure 2-13. The Mahantango-Marcellus interval lithofacies are predicted in a test well. The second track shows original lithofacies, and third, fourth, fifth, and sixth tracks show test results from four different algorithms.

Table 2-5. Confusion matrix illustrates accuracy of SVM algorithm in lithofacies prediction in the Mahantango-Marcellus dataset. Overall accuracy by SVM is 82.46%.

		Actual					
		Lithofacies	1	2	3	4	5
Predicted	1	56	0	12	0	0	0
	2	0	106	2	32	0	0
	3	11	7	181	9	1	0
	4	0	8	12	99	0	1
	5	0	0	0	0	1	2
	6	0	0	0	1	6	46
	Absolute Accuracy	83.58%	87.60%	87.44%	70.21%	12.50%	93.88%

Table 2-6. Confusion matrix illustrates accuracy of ANN algorithm in lithofacies prediction in the Mahantango-Marcellus dataset. Overall accuracy by ANN is 78.75%.

		Actual					
		Lithofacies	1	2	3	4	5
Predicted	1	57	0	12	0	0	0
	2	0	101	4	36	0	0
	3	10	7	173	18	0	0
	4	0	13	18	85	0	0
	5	0	0	0	0	2	0
	6	0	0	0	2	6	49
	Absolute Accuracy	85.07%	83.47%	83.57%	60.28%	25.00%	100.00%

Table 2-7. Confusion matrix illustrates accuracy of 2D SOM algorithm in lithofacies prediction in the Mahantango-Marcellus dataset. Overall accuracy by SOM is 69.64%.

		Actual					
		Lithofacies	1	2	3	4	5
Predicted	1	58	0	26	11	0	1
	2	0	86	2	39	0	0
	3	9	17	151	15	0	0
	4	0	18	21	71	0	0
	5	0	0	7	3	1	2
	6	0	0	0	2	7	46
	Absolute Accuracy	86.56%	71.07%	72.95%	50.35%	12.50%	93.88%

Table 2-8. Confusion matrix illustrates accuracy of MRGC algorithm in lithofacies prediction in the Mahantango-Marcellus dataset. Overall accuracy by MRGC is 64.42%.

		Actual					
		Lithofacies	1	2	3	4	5
Predicted	1	58	0	36	0	0	0
	2	0	104	16	61	0	0
	3	1	5	118	5	0	0
	4	6	12	26	70	0	2
	5	2	0	11	5	5	20
	6	0	0	0	0	3	27
	Absolute Accuracy	86.56%	85.95%	57.00%	49.65%	62.50%	55.10%

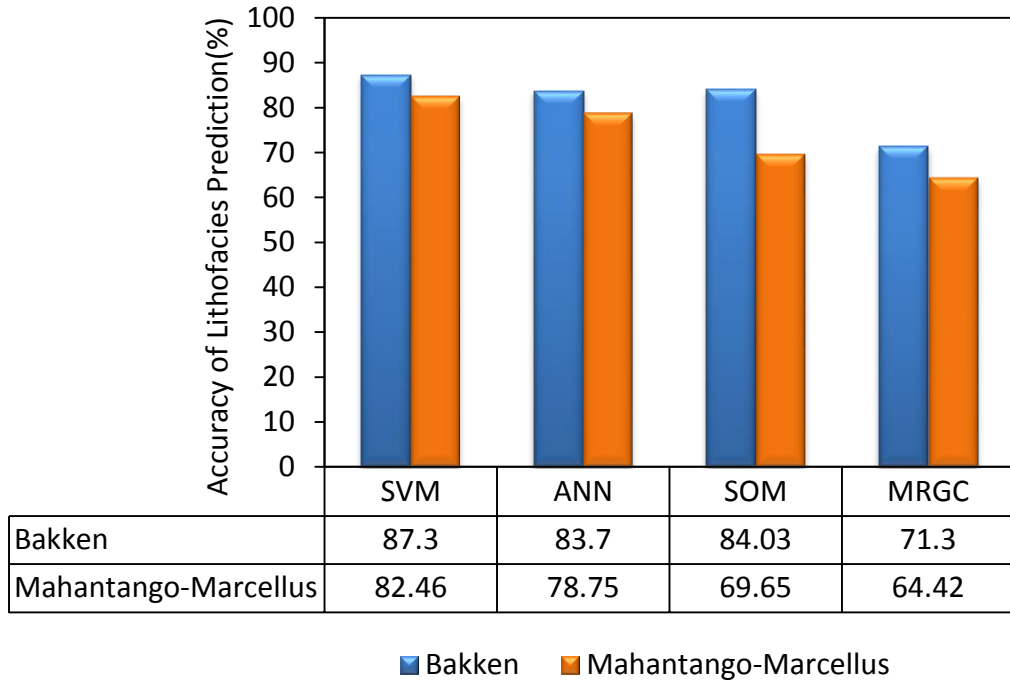


Figure 2-14. Summary of lithofacies prediction by SVM, ANN, SOM and MRGC methods in the Bakken and Mahantango-Marcellus dataset. SVM appears to be the best method.

Both Bakken and Mahantango-Marcellus shale members are vertically and laterally heterogeneous, but can be classified into five different mudstone lithofacies, along with calcareous siltstone and limestone units. All supervised algorithms can train and test shale lithofacies patterns. However, the overall uncertainty of each method mostly stems from relatively poor sample proportion of the gray shale lithofacies (GMD, GMS, and GSS). Results from both dataset indicate that SVM works far better than other techniques for lithofacies classification and prediction purposes in reduced computational time (0.11 sec and 0.18 sec with the Bakken and Mahantango-Marcellus dataset, in comparison to 5.58 sec and 2.38 sec respectively by ANN algorithm), requires no iteration, and with highly repeatable results with the same dataset, because it does not require setting of randomized interconnected weights as

opposed to ANN algorithm. SVM also uses a significantly small portion of training data represented by support vectors compared to all other approaches, which require full training dataset during training. With the Bakken dataset, SVM uses 40 support vectors, whereas it needs only 26 support vectors to represent the full petrophysical data pattern associated with lithofacies in the Marcellus dataset. Further, SVM needs only two parameters (gamma and C) to optimize during training and testing. Figure 2-15 shows that accuracy of lithofacies classification approaches 100% by increasing values of gamma and C during training. In fact, SVM training models start mimicking the original lithofacies model, once gamma and C values are over 100. But the models fail to work successfully outside their training domain, with such high gamma and C parameters, due to the loss of its generalization property. Relatively low values of gamma and C help to balance the model during training and testing.

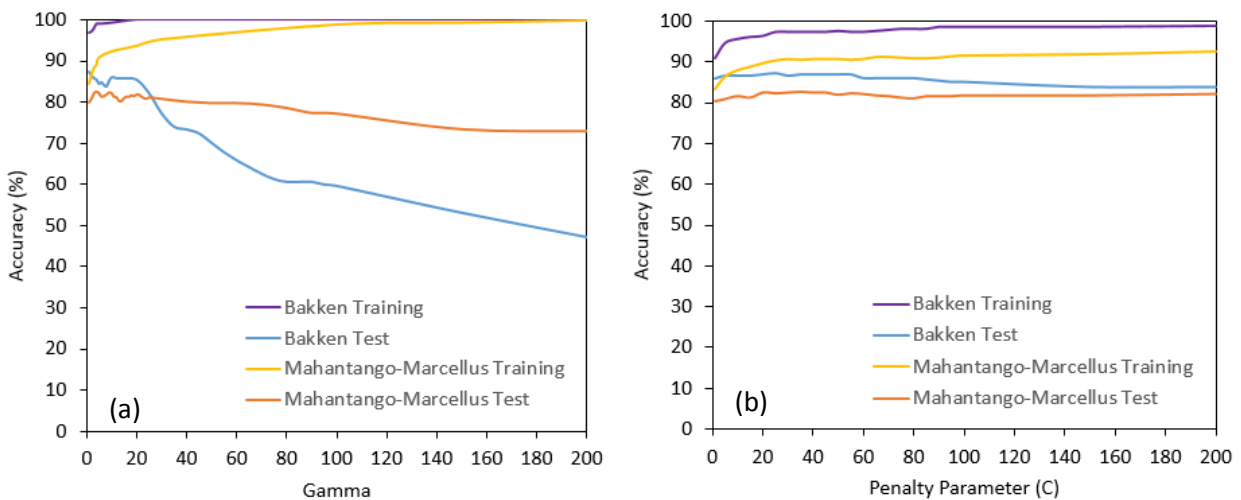


Figure 2-15. SVM Network parameter optimization by checking accuracy level at every gamma and C values (a and b).

In terms of computational cost used by SOM and MRGC algorithms (both of them take more than 120 sec per well to generate lithofacies models), their precise values of the computational costs cannot be exactly compared with SVM and ANN algorithms, because all algorithms were

implemented using different programming languages and software. Nevertheless, a comparison of computational cost suggests that the SVM is computationally less demanding than the ANN, SOM, and MRGC.

Figure 2-16 shows changes in accuracy of SVM based lithofacies classification with increasing size of the training dataset. For the Bakken interval, 4 training dataset of varying size (i.e. 157, 235, 313, and 391 samples for all seven lithofacies) were generated, using stratified random sampling. Stratified random sampling was used, because the original dataset is highly imbalanced, which implies proportions of all seven lithofacies are not the same. Bakken test dataset consisting of 307 samples was used with different training dataset. Results from Fig. 15 suggest that the performance of SVM classifier depends on the size of the training samples for lithofacies classification, which confirms the findings of Sebtosheikh and Salehi (2015), suggesting increase in classification accuracy with increasing training sample size. These results are also similar to other studies (Pal and Foody, 2010; Pal and Mather, 2005).

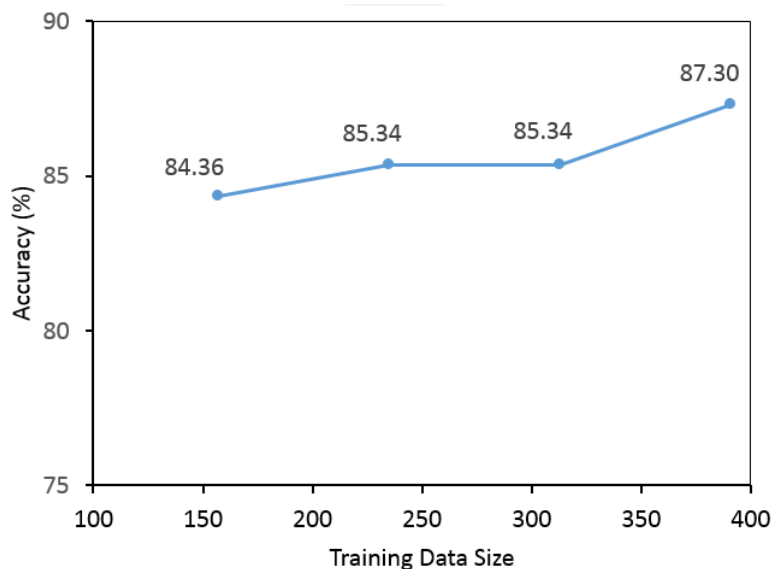


Figure 2-16. Variation in lithofacies classification accuracies with change in data size

Although, ANN and SOM methods produced good results, they work as black-box model. The lower accuracy of these supervised algorithms, compared to SVM, can be attributed to local minima problem. Number of iterations for these algorithms in the both dataset is set at the range of 1000 to 2000 to derive a global solution, rather than a solution stuck in local minima error. However, this process is not cost-effective (time). Number of iterations also depends on the complexity and size of dataset. Moreover, there are several user defined data dependent parameters required to control during ANN model building process; such as number of hidden layers, learning rate, and momentum, which are data dependent. MRGC is a powerful tool, which can reveal natural data structure, without interpreter's bias. The lower rate of accurate prediction by MRGC with these datasets may result from high error rates resulting from the small sample size of relatively rare lithofacies such as inorganic gray shale lithofacies or even suboptimum input well log conditioning. The probability of predicting relatively rare lithofacies in the sample population increases, if the data is intelligently partitioned rather than random splitting for cross-validation, so at least a few samples of each lithofacies are present during training, testing, and validation phases.

We recommend using MRGC as an exploratory technique if the geology is unknown, and then use the initially generated lithofacies model as a reference for SVM based lithofacies classification. MRGC helps to understand high-resolution natural data structure, which can be used to understand the complex lithofacies pattern. Once, the lithofacies pattern is established and/or calibrated to "ground truth" information at various scales of analysis, SVM algorithm can be used efficiently for lithofacies modeling in other test wells.

2.6 Conclusions

The study shows that heterogeneous and complex mudstone lithofacies can be modeled by pattern recognition tools. These kinds of machine learning tools are useful to classify and predict lithofacies in wells, with limited a priori geological knowledge. Once lithofacies model is defined in each well in the study area, they can be populated in grids to create 3D geocellular facies models, providing insight to depositional and diagenetic processes. The major conclusions of this study are as follows:

1. Both Bakken and Mahantango-Marcellus shale members are heterogeneous, but can be classified into at least five mudstone lithofacies, along with calcareous siltstone and limestone lithofacies.
2. Apart from fundamental working principles, the uncertainty of each method (SVM, ANN, SOM, and MRGC) also depends on the size, complexity of the dataset, and suitable choice of input parameters for lithofacies classification.
3. SVM is the best method for lithofacies classification and prediction purposes in reduced computational resources, no iteration, and with highly repeatable results.
4. Accuracy of lithofacies prediction increases with geological rules being superimposed.
5. The probability of predicting relatively rare facies in the sample population is increased if the data is intelligently partitioned, rather than random splitting, so at least a few samples from each lithofacies are present during training and testing processes.

Acknowledgements

We would like to thank Julie LeFever of North Dakota Geological Survey for providing access to well log database. We would also like to express our thanks to CGG Jason (PowerLog™), IHS

(Petra™), and Paradigm (Geolog™) for providing necessary software facilities at West Virginia University. The authors wish to thank three reviewers, whose comments help to strengthen the manuscript.

References

Al-Anazi, A., Gates, I.D., 2010, On the capability of Support Vector Machines to Classify Lithology from Well Logs, *Natural Resources Research*, **19** (2), 125-139.

Al-Anazi, A., Gates, I.D., 2010, A support vector machine algorithm to classify lithofacies and model permeability in heterogeneous reservoirs, *Engineering Geology*, **114**, 267-277.

Bishop C.M., 1995, *Neural Networks for Pattern Recognition*, New York, Oxford University Press.

Bhattacharya, S., Carr, T., Wang, G., 2015, Shale Lithofacies Classification and Modeling: Case Studies from the Bakken and Marcellus Formations, North America, presented at AAPG Annual Convention and Exhibition, Search and Discovery #90216, June 1.

Bhattacharya, S., Carr, T.R., 2016, Integrated Petrofacies Characterization and Interpretation of Depositional Environment of the Bakken Shale in the Williston Basin, North America, *Petrophysics*.

Camps-Valls, Gustavo, and Lorenzo Bruzzone, 2009, *Kernel methods for Remote Sensing Data Analysis*, John Wiley & Sons.

Christianini, C., Shawe-Taylor, J., 2000, *An introduction to support vector machines and other kernel-based learning methods*, Cambridge, Cambridge University Press.

Cortes, C., Vapnik, V., 1995, Support vector networks, *Mach. Learn.*, **20**, 273-297.

Doveton, J.H., 1994, Geologic Log Analysis Using Computer Methods, Tulsa, AAPG Computer Publications in Geology, 2.

Egenhoff, S.O., Fishman, N.S., 2013, Traces in the dark—sedimentary processes and facies gradients in the Upper Shale Member of the Upper Devonian–Lower Mississippian Bakken Formation, Williston basin, North Dakota, U.S.A, Journal of Sedimentary Research, **83**, 803-824.

Emmanuel, O.O., Sonnenberg, S.A., 2013, Concepts and Methods for the Recognition of Cyclicity in the Marcellus Shale of the Appalachian Basin, NE USA, presented at the Unconventional Resources Technology Conference, URTeC 1563723, August 12-14.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H., 2009, The WEKA Data Mining Software: An Update; SIGKDD Explorations, **11** (1), 10-18.

Hastie, T., Tibshirani, R., 1998, Classification by pairwise coupling. Ann. Stat, **26**(2), 451-471.

Hsu, C.W., Chih-Chung C., Chih-Jen L., 2003, A Practical Guide to Support Vector Classification, Technical report, Department of Computer Science, National Taiwan University.

Kecman, V., 2005, Support Vector Machines –an introduction. In: Wang, L., (Ed.), Support Vector Machines: Theory and Applications, Chap 1. Springer-Verlag, Berlin Heidelberg, 1-47.

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

Kohonen, T., 2001, Self-organizing Maps, 3rd ed., Berlin Heidelberg, Springer-Verlag.

Kordon, A.K., 2010, *Applying Computational Intelligence: How to Create Value*, Berlin Heidelberg, Springer-Verlag.

LeFever, J. A., LeFever, R.D., Nordeng, S.H., 2011, Revised nomenclature for the Bakken Formation (Mississippian- Devonian), North Dakota, In: Robinson, J.W., LeFever, J.A., Gaswirth, S.B., eds., *The Bakken-Three Forks Petroleum System in the Williston Basin: Denver, Colo., Rocky Mountain Association of Geologists*, 11–26.

Li, T., Zhu, S., Ogihara, M., 2003, Using discriminant analysis for multi-class classification: An experimental investigation,. In: *Proceedings of the Third IEEE International Conference on Data Mining*, 589-592.

Luts, J., Ojeda, F., Plas, R.V., Moor, B.D., Huffel, S.V., Suykens, J.A.K., 2010, A tutorial on support vector machine-based methods for classification problems on chemometrics, *Analytica Chimica Acta*, **665**, 129-145.

Manshad, A.K., Rostami, H., Rezaei, H., Hosseini, S.M., 2015, Application of Artificial Neural Network- Particle Swarm Optimization Algorithm for Prediction of Asphaltene Precipitation during Gas Injection Process and Comparison with Gaussian Process Algorithm, *Journal of Energy Resources Technology*, **137**, 1 – 5.

McCulloch, W., Pitts, W., 1943, A logical calculus of the ideas immanent in nervous activity, *Bulletin of Mathematical Biophysics*, **7**, 115 – 133.

Pal, M., Mathur, P.M., 2006, Some issues in the classification of DAIS hyperspectral data, *International Journal of Remote Sensing*, **27** (14), 2895-2916.

Pal, M., Foody G.M., 2010, Feature selection for classification of hyperspectral data by SVM, IEEE Transactions on Geoscience and Remote Sensing, **48** (5), 2297-2307.

Pal, M., Foody, G.M., 2012, Evaluation of SVM, RVM and SMLR for accurate image classification with limited ground data, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, **5** (5), 1344-1355.

Qi, L.S., Carr, T.R., 2006, Neural network prediction of carbonate lithofacies from well logs, Big Bow and Sand Arroyo Creek fields, Southwest Kansas, Computers & Geosciences, **32**, 947–964.

Roy, A., 2013. Latent Space Classification of Seismic Facies, Ph. D Dissertation, University of Oklahoma, Norman, Oklahoma.

Schieber, J., 1999, Distribution and deposition of mudstone facies in the Upper Devonian Sonyea Group of New York: Journal of Sedimentary Research, **69**, 909-925.

Schlanser, K., Grana, D., Campbell-Stone, E., 2014, Petro-elastic facies classification in the Marcellus Shale by applying Expectation Maximization to measured well logs, presented at SEG Annual Meeting (DOI <http://dx.doi.org/10.1190/segam2014-0939.1>).

Schmoker, J.W., Hester, T.C., 1983, Organic carbon in Bakken Formation, United States portion of Williston Basin: American Association of Petroleum Geologists, Bulletin, **67**, 2165–2174.

Sebtosheikh, M. A., Salehi, A., 2015, Lithology prediction by support vector classifiers using inverted seismic attributes data and petrophysical logs as a new approach and investigation of

training data set size effect on its performance in a heterogeneous carbonate reservoir, *Journal of Petroleum Science and Engineering*, **134**, 143-149.

Singh, P., 2008, Lithofacies and sequence stratigraphic framework of the Barnett Shale, northeast Texas, Ph. D Dissertation, University of Oklahoma, Norman, Oklahoma.

Tan, M.J., Liu, Q., Zhang, S.Y., 2013, A dynamic adaptive radial basis function approach for total organic carbon content prediction in organic shale, *Geophysics*, **78** (6), 445-459.

Wang, G., 2012, Black shale lithofacies prediction and distribution pattern analysis of Middle Devonian Marcellus shale in The Appalachian basin, northeastern U.S.A., Ph.D. Dissertation, West Virginia University, Morgantown, West Virginia.

Wang G., Carr, T.R., Ju, Y., Li, C., 2014, Identifying organic-rich Marcellus Shale Lithofacies by support vector machine classifier in the Appalachian basin, *Computers & Geosciences*, **64**, 52-60.

Vapnik, V., 1995, *The nature of Statistical Learning Theory*, New York, Springer.

Venkatesan, P., Mullai, M., 2014, Visualization of Breast Cancer Data by SOM Component Planes, *International Journal of Science and Technology*, **3**(2), 127-131.

Ye, Shin-Ju, Rabiller, P., 2000, A new tool for electro-facies analysis: multi-resolution graph-based clustering, presented at SPWLA 41st Annual Logging Symposium, June 4- 7.

CHAPTER 3

Integrated 3D Shale Lithofacies Modeling of the Bakken Shale in the Williston Basin, North Dakota

Shuvajit Bhattacharya¹ and Timothy R. Carr¹

¹Department of Geology and Geography

West Virginia University, Morgantown

WV- 26506, USA

(304)-777-9885

Email: subhattacharya@mix.wvu.edu; tim.carr@mail.wvu.edu

Abstract

Integrated 3D shale lithofacies modeling is important to visualize distribution pattern of different lithofacies, interpret depositional, diagenetic environments, and understand hydrocarbon potential of mudstone formations. A 3D shale lithofacies model is constructed for the shale members in the Bakken Formation of the Williston basin in North Dakota, United States. The principal objective of this multi-scale study is to identify different shale lithofacies in the Bakken Formation to better understand their depositional and diagenetic controls on mineral composition and organic-matter content, plus provide a quantitative geological framework at core, well, and regional scales. Shale lithofacies can be defined using quantitative mineralogy, Total Organic Carbon (TOC) content, and various petrophysical properties derived from core analysis, advanced geochemical spectroscopy logs, and conventional well logs at core and well scales. Machine learning algorithms such as Support Vector Machine (SVM) and Artificial Neural Network (ANN) are used and compared to recognize the pattern of different shale lithofacies, associated with basic petrophysical parameters from ubiquitous conventional well log suites from a large number of drilled wells. A limited set of ten petrophysical parameters are used as input to the mathematical algorithms, which can directly output shale lithofacies with high accuracy in wells, with or without core data and advanced geochemical log based constraints. Geostatistical algorithms, such as Sequential Indicator Simulation, are used to populate all lithofacies in a 3D grid, covering the portion of the Williston basin in North Dakota. The results show that the Bakken shale members are vertically and laterally heterogeneous, but can be classified into five different lithofacies. Proportion of organic-rich shale lithofacies outweigh the proportion of organic-poor (“gray”) shale lithofacies. Distribution pattern of organic-rich siliceous

shale in the Williston basin is hypothesized to be related to presence of biogenic and detrital components, whereas organic mudstone is believed to be associated with organic productivity.

3.1 Introduction

This paper discusses the application of geologic pattern recognition and 3D lithofacies modeling of the Bakken shale members, integrating core and well log data. Shale lithofacies classification and prediction is a complex and multi-class problem at various scales of analysis. Research in organic-rich shale lithofacies is relatively rare, compared to clastic and carbonate rocks. Most of the works on shale lithofacies are based on either outcrop, single well study or core based descriptive analysis, using petrography, fossil content, and texture etc. (Bhattacharya and Carr, 2016; Egenhoff and Fishman, 2013; Hickey and Henk, 2007; Schieber, 1999). However, core samples are limited due to economic and logistic constraints, which poses a significant challenge to accurately identify facies variations and correlate shale lithofacies at the regional scale. In addition, outcrop shale formations may not be always analogous to the subsurface, due to weathering processes and potential of kerogen degradation over time. Advanced geochemical logs such as Pulsed Neutron Spectroscopy (PNS) can be used to develop an understanding of continuous variation of mineralogy and organic-matter content. However, data acquisition and processing with PNS tools are expensive that is why the limited number of wells with such advanced logs is insufficient for regional scale studies. All these aspects complicate the problem of shale lithofacies modeling, because shale lithofacies are heterogeneous at various scales. Shale lithofacies should be geologically meaningful, mappable, and consistent at various scales of analysis, to have a broader implication on depositional, diagenetic environment studies, and integrated reservoir characterization (Wang, 2012). In this study, we attempt to classify and integrate shale lithofacies at core, well, and regional scales in a quantitative manner, calibrated to “ground truth”. Such multi-scale quantitative lithofacies modeling can be significantly useful

to analyze sedimentary formations, to interpret depositional and diagenetic environments, to predict variations in rock properties, and to characterize the petroleum system. The shale members of the Bakken Formation in the Williston basin of North Dakota are chosen for this study.

The Bakken Formation consists of two world class source rocks (upper Bakken and lower Bakken shale) that sandwich the reservoir middle Bakken, which is composed of mixed lithologies including sandstone, limestone, and dolostone. Most of the geological studies in the Bakken Formation have been focused on depositional, diagenetic history, facies analysis, and reservoir characterization of the middle Bakken member and the underlying Three Forks Formation (LeFever et al., 2011; Sesack, 2011; Simenson, 2011; Warner, 2011). Recently, a handful of studies on the Bakken shale members, ranging from source rock potential and chemostratigraphy, using limited core based X-ray Diffraction (XRD), X-ray Fluorescence (XRF), and Total Organic Carbon (TOC) data have been performed towards better understanding of geological processes controlling transportation, deposition, preservation, and diagenesis of organic matter and mud (Kocman, 2014; Maldonado, 2012; Nandy et al., 2014; Rowe et al., 2012). In this study, we will investigate vertical and lateral variations of different Bakken shale lithofacies at multiple scales, and attempt to associate the shale lithofacies with corresponding geologic factors.

This study starts with shale lithofacies classification at core scale, and then propagates the facies classification scheme at well scale, and finally upscales the integrated geologic data at regional scale, to generate 3D shale lithofacies models for the upper and lower Bakken shale members. We classify the Bakken shale lithofacies at core scale, using a schematic methodology

based on core-derived mineralogy from XRD and TOC data (Bhattacharya and Carr, 2016). Next, core data is calibrated to conventional and advanced PNS logs responses. Once, shale lithofacies is established at well scale, we use predictive algorithms such as SVM and ANN to train and test the model for all the available wells with ubiquitous wells with conventional well logs, no core data and/or advanced well log based controls. Such data-intensive machine learning algorithms are efficient for lithofacies modeling at regional scale, compared to conventional method of “manually” assigning lithofacies on a well-by-well basis. Next, all classified and predicted shale lithofacies are compiled and mapped at the regional scale, using geostatistical algorithms. In this study, we use Sequential Indicator Simulation (SIS) algorithm for 3D stochastic geocellular modeling. The results show that both upper and lower Bakken shale members are vertically and laterally heterogeneous, but can be classified into five different lithofacies. Organic-rich shale lithofacies are more dominant than organic-poor shale lithofacies in the Bakken Formation.

3.2 Geologic Setting

The Late Devonian-Mississippian Bakken Formation in the Williston basin (Figure 3-1) spans over portions of North Dakota, South Dakota, Montana in the United States, and Saskatchewan and Manitoba in Canada (~200,000 mi², ~500,000 Km²). The Williston basin is a large intracratonic sedimentary basin (Pitman et al., 2001). The Bakken Formation lies unconformably over the Three Forks Formation, and it is overlain by the Lodgepole Formation. The Bakken system is composed of three members- upper, middle, and lower units. Middle Bakken member, which is composed of siltstone, limestone, and dolostone (mixed lithology), is bound by two shale members at the top and bottom (Smith and Bustin, 2000; Webster, 1984) (Figure 3-2). Both

Bakken shale members are heterogeneous and complex mixture of quartz, smectite, illite, carbonate, pyrite, and kerogen in varying proportions (Bhattacharya and Carr, 2016). Upper and lower shale members in the Bakken Formation are considered to be deposited during the basin wide transgression and/or basin subsidence at the end of the deposition of the Three Forks Formation, whereas the middle member (composed of mixed lithology) has been interpreted to be deposited as a result of dramatic sea-level drop at the close of the Devonian period (LeFever et al., 1991; Sesack, 2011; Sonnenberg and Pramudito, 2009; Steptoe, 2012). Nesson Anticline, a north-south trending structural feature is present near the center of the basin, where all the Bakken members are thickest (>50 feet, 15m) in the basin.

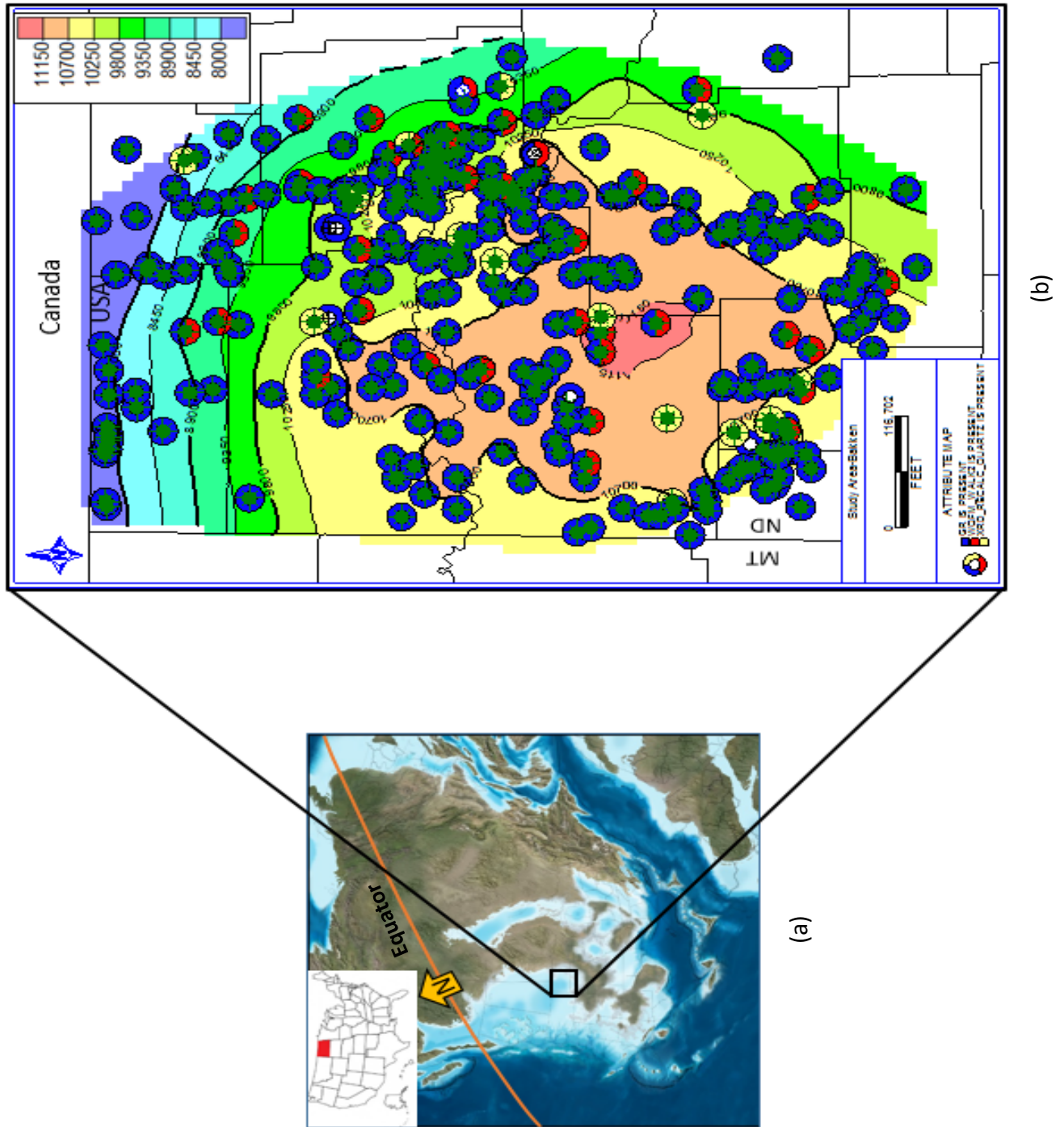


Figure 3-1. Paleogeographic location of the Williston basin in North America during the Late Devonian period (a) and study area in North Dakota, with available core and well log data (b). The circles on the study area map shows the type of data available (blue color indicates conventional well logs, red indicates PNS logs, whereas yellow represents core data.). Figure 3-1b shows the present day upper Bakken structure map (modified after Blakey, 2011).

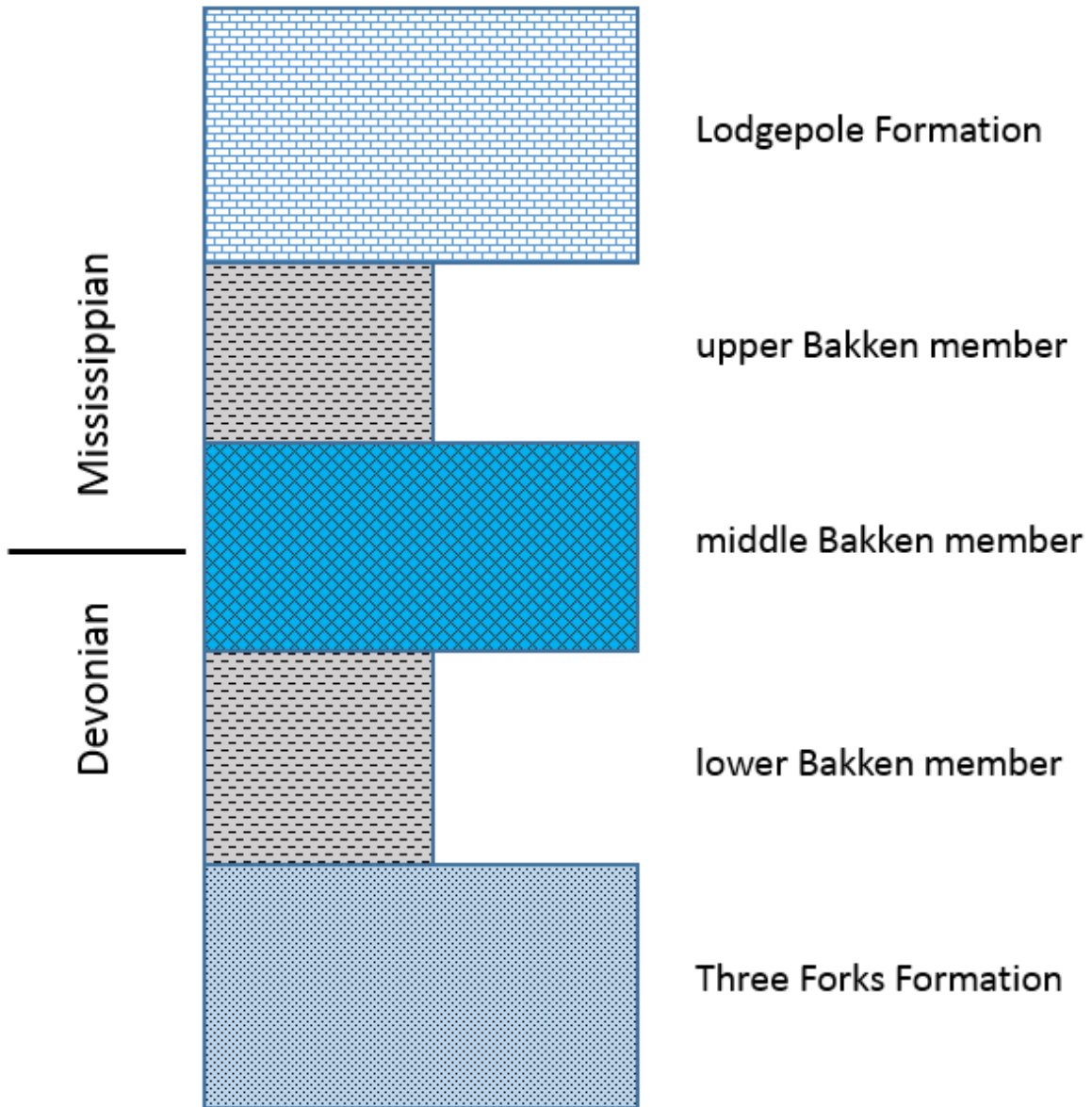


Figure 3-2. Lithostratigraphy of the Bakken Formation. The Bakken Formation is composed of three members- upper shale, middle mixed lithology, and lower shale (modified after Egenhoff et al., 2011).

As per paleogeography, the Williston basin was near the equator (within 5° to 10° south) during the Devonian period, when the Bakken Formation was deposited. This fact may attribute to eolian activities due to trade winds, which may influence lithofacies distribution pattern.

3.3 Dataset and Methods

The dataset used in this study consists of 417 wells containing conventional well log suites (such as gamma, resistivity, neutron porosity, density, and photoelectric factors), 37 wells with advanced PNS logs, and 17 wells with core data (XRD and TOC). Each well log is recorded at half-foot (0.15 meters) interval. Cores are sampled at North Dakota Geological survey in such a way, so that they cover broad geographical range to include as much geological variation as possible and none of the cores are more than 5 to 6 years old, which implies reduced chance of kerogen degradation over time.

The Bakken shale lithofacies is classified and integrated at core, well, and regional scales. First, mineralogy and organic matter richness data from XRD and TOC measurements were plotted in a ternary diagram to understand heterogeneity of the Bakken shale units at core scale (Figure 3-3) (Bhattacharya and Carr, 2016). A schematic methodology is used to classify the Bakken shale lithofacies, using the criteria, described by Bhattacharya and Carr (2016) (Figure 3-4). Three criteria are used to classify the Bakken shale lithofacies: TOC (cutoff 7%), clay volume (cutoff 30%) and quartz-to-carbonate ratio (cutoffs 3 and 1/3).

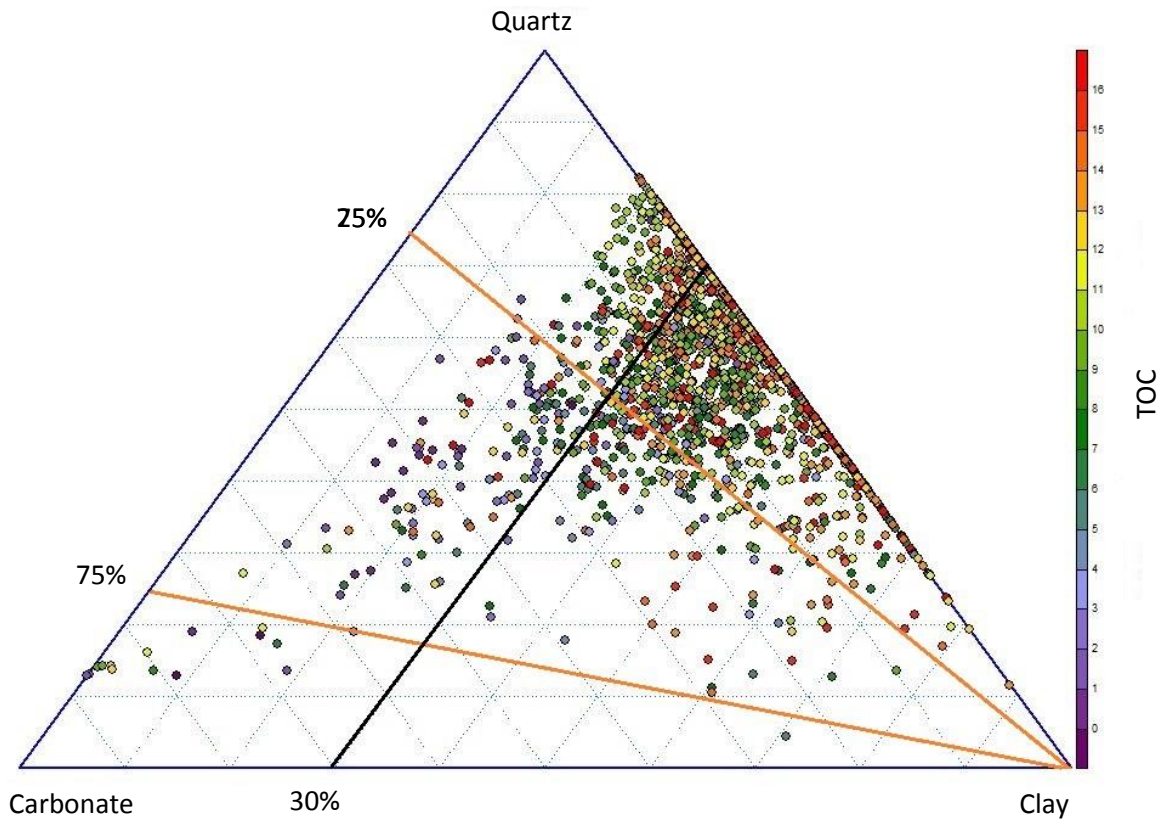


Figure 3-3. Ternary diagram showing variation of mineralogy and TOC content of the Bakken shale members based on core and PNS log data. The Bakken shale members are rich in quartz and clay, compared to carbonate. Most of the samples have high TOC. Three lines (quartz-to-carbonate ratio 3 and 1/3) and clay (30%) are drawn over the triangle to identify and distinguish different shale lithofacies (Bhattacharya and Carr, 2016).

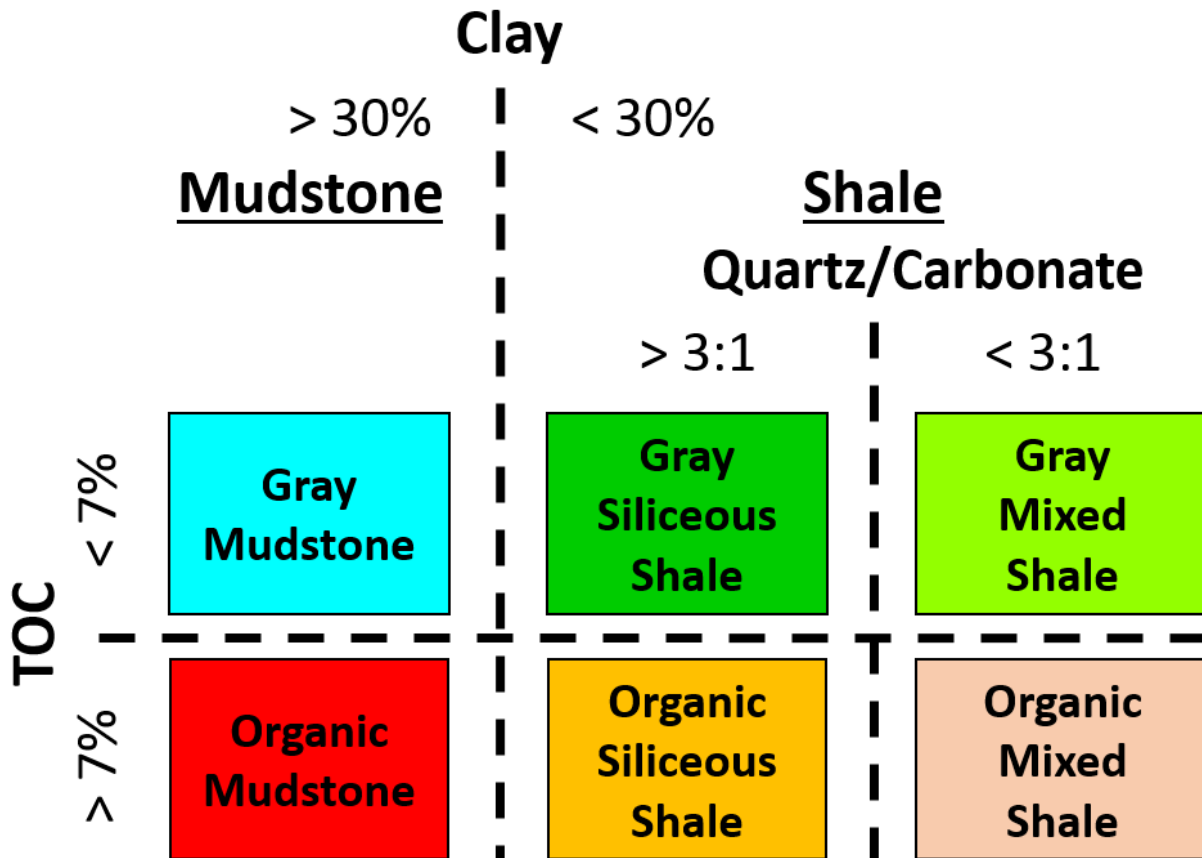


Figure 3-4. A general workflow to classify shale lithofacies using core and PNS log data. Six different shale lithofacies could be classified based on mineralogy and TOC data, however, organic mixed shale lithofacies is absent in the Bakken Formation.

Due to relative scarcity of core points, continuous mineralogy from advanced PNS logs, TOC logs, and stochastic mineralogical solutions were also used in wells, without required core data to enrich the dataset to better understand and evaluate shale heterogeneity. TOC is estimated from bulk density log, using Schmoker’s technique (1983). The derived TOC can be considered as a quantitative estimate of organic matter richness.

$$\text{TOC}_{\text{Schmoker}} = (154.497 / \text{Bulk Density}) - 57.261$$

Stochastic mineralogical solutions are obtained for many wells, using linear inversion technique (Kulyapin and Sokolova, 2014; Mitchell and Nelson, 1988; Moss and Harrison, 1985; Savre, 1963). Gamma, neutron porosity, bulk density, and Umaa parameters are chosen as input for stochastic mineralogical modeling using PowerLog™. Umaa is the product of photoelectric and bulk density log corrected by apparent total porosity. The final stochastic mineralogical solution is composed of quartz, clay, calcite, dolomite, kerogen and pyrite (Figure 3-5). The mineralogical solution is constrained, either by available core data and/or PNS logs. The kerogen estimation could also include any free oil and water present in the Bakken system, however, study of different types of fluid is out of the scope of this study. The calculated proportions of mineralogy and TOC data are used to classify different shale lithofacies in the Bakken Formation based on the tree-based facies classification scheme, described earlier (Figure 3-4). However, deriving geologically meaningful stochastic mineralogical solutions for all wells at regional scale, with this approach can take significantly high amount of time. Figure 3-6 shows the lithofacies distribution and depositional conditions (paleo-redox) based on trace element geochemical data and TOC content.

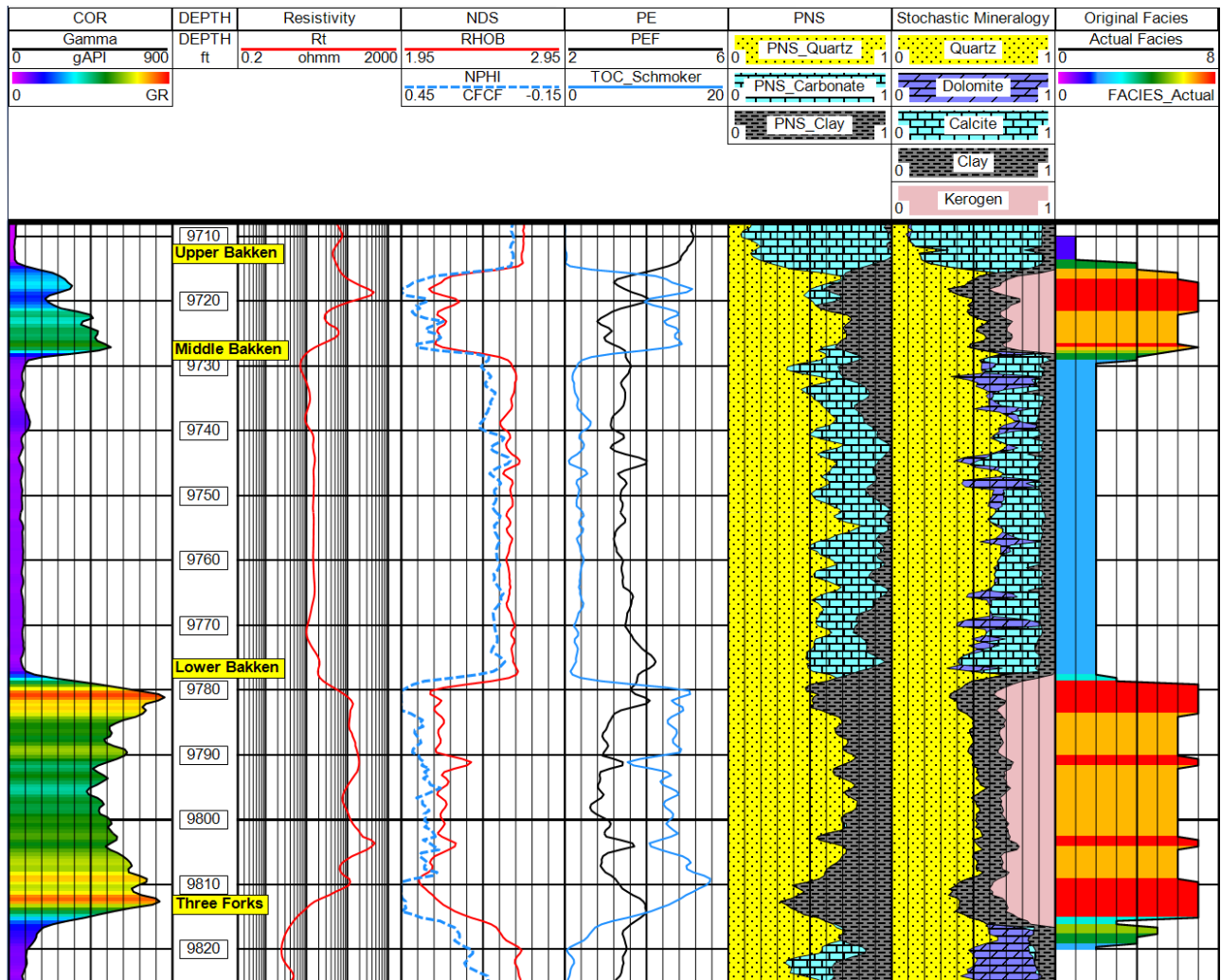


Figure 3-5. Integrated well log and original lithofacies model display in one of the training wells in the Bakken dataset. Upper Bakken and lower Bakken shale members are composed of five different shale lithofacies. Limestone, mixed lithology (calcareous siltstone), GMD, GMS, GSS, OSS, and OMD lithofacies are represented by numerical codes from 1 to 7 in a rainbow pattern.

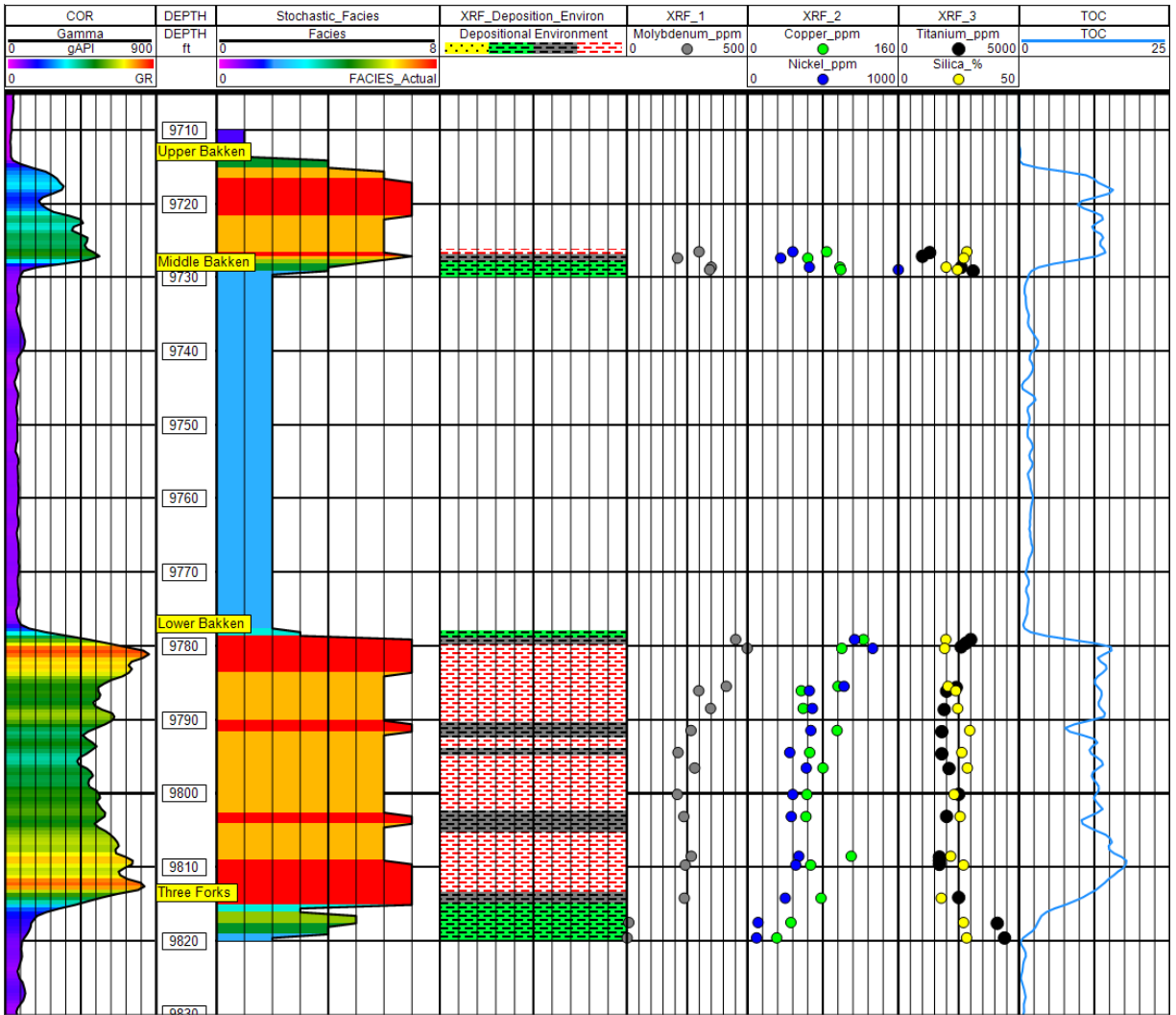


Figure 3-6. Integrated display of lithofacies model, depositional conditions (paleo-redox) and geochemical data. A gap in interpreting depositional environment in the upper Bakken member resulted due to unavailable core data.

We use a second approach here, to classify shale lithofacies, using ubiquitous conventional well logs from a large number of wells. We built predictive and cost-effective (time) mathematical models that can directly output the shale lithofacies at well scale with high accuracy, using conventional well log derived petrophysical parameters as input. SVM and ANN algorithms are utilized to build the mathematical models. Both of these machine learning algorithms are flexible in their design to deal with a complex, non-linear data pattern recognition problem, and have been applied successfully in the past for lithofacies classification and prediction purposes in complex reservoirs (Qi and Carr, 2006; Wang et al., 2014; Zhao et al., 2015). We choose five conventional well logs; gamma (GR), natural logarithm of deep resistivity (LnRt), neutron porosity (NPHI), bulk density (RHOB), and photoelectric factor (PEF), and five other derived parameters including apparent matrix density (RHOMaa), photoelectric absorption index (Umaa), gamma/density (GR/RHOB), photoelectric/density (PEF/RHOB), and TOC as input parameters for lithofacies modeling (Figure 3-7). All of the input parameters chosen are sensitive to variations in shale lithofacies. For example, organic-rich mudstone typically shows high GR, high LnRt, high NPHI, and low RHOB signatures, associated with high kerogen and high clay. Kerogen has a uranium component in the increased GR response, and high resistivity and low density (~1.1 g/cc). High NPHI response of organic mudstone is attributed to the bound water associated with high clay volume in it.

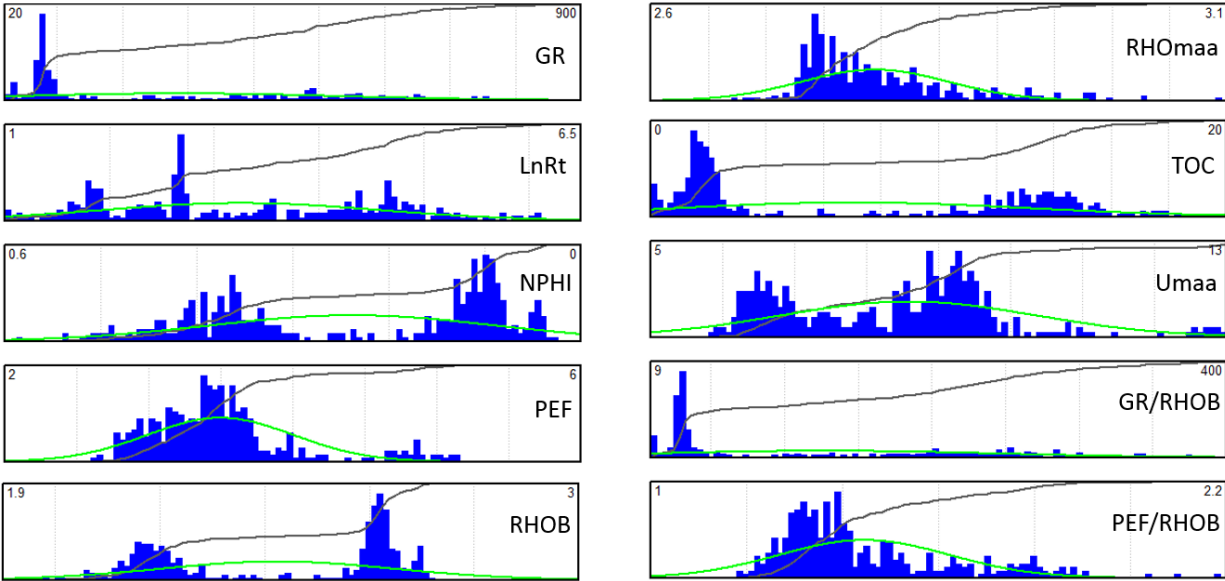


Figure 3-7. Histograms showing statistical feature of conventional well log suites in the Bakken dataset, used as input parameters for SVM and ANN model. Five basic well logs and five derived parameters are used as input shale lithofacies pattern recognition. The green curve indicates Gaussian distribution, whereas the black curve indicates cumulative distribution of well log data.

Based on numerous experiments, SVM and ANN models are set up with optimum network parameters to enable them work efficiently in both training and testing domains. We use Radial Basis Function (RBF) kernel for SVM classifier, which requires fine tuning of two user-defined parameters: gamma (kernel parameter), and penalty parameter (C). For SVM, the values of gamma and C are 4.1 and 20 respectively. We apply back propagation ANN algorithm with one hidden layer. Optimal values of learning rate (i.e. 0.3), momentum (i.e. 0.9), and number of iterations (i.e. 1000), obtained after large number of trials, are used for shale lithofacies classification. In the next step, the models are supervised with available shale lithofacies in a few wells for training, and then they are tested to check accuracy of lithofacies prediction outside

their training domain. SVM algorithm is chosen for further application as it proved to be better than ANN technique for lithofacies pattern recognition (Bhattacharya and Carr, in press).

The Bakken shale lithofacies classified and predicted in all available wells are used to generate a 3D geocellular lithofacies model at regional scale in an area of approximately 12,824 square mile (~33,213 square km). The unit boundaries, zones, layers in each zones, and corresponding structural maps of the upper Bakken, middle Bakken, lower Bakken, and the Three Forks Formation are utilized to create a 3D grid. Next, the Bakken lithofacies logs (from SVM model) are upscaled in such a way, so that vertical thickness of the 3D cells match with well log sampling rate to preserve formation details. Geostatistical techniques, such as SIS algorithm, are used to construct the 3D shale lithofacies models. SIS algorithm is used in this study, because we do not have sufficient knowledge about transportation, depositional mechanism of fine-grained rocks and associated shale lithofacies distribution pattern. Application of spatial statistics is useful to visualize vertical and lateral distribution of different lithofacies in three dimensional grid, which can aid in interpreting depositional and diagenetic environments, and drive decisions for petroleum system management. In this study, WEKA program is used for SVM and ANN based lithofacies classification and prediction, PowerLog™ to display well logs and vertical lithofacies distribution, Petra™ to generate maps, and Petrel™ to construct 3D geological models.

3.4 Results and Discussions

The ternary diagram shows that Bakken shale members are heterogeneous, in terms of mineralogy and organic matter content (Figure 3-3). Both Bakken shale members are siliceous mudstone- rich in quartz and clay, compared to carbonate. Most of the samples have high TOC

(>7 wt%). Five different shale lithofacies are defined, using the lithofacies classification scheme based on core data, PNS logs, and stochastic mineralogical solutions (Figure 3-4). Out of five different shale lithofacies two are organic-rich (Organic Mudstone (OMD) and Organic Siliceous Shale (OSS)) and the remaining three are organic-poor lithofacies (Gray Siliceous Shale (GSS), Gray Mixed Shale (GMS), and Gray Mudstone (GMD)).

Mineralogical solutions depict that Bakken shale members are vertically heterogeneous, in terms of different lithofacies at well scale, however, organic-rich lithofacies (OSS and OMD) assemblages outweigh the proportion of relatively organic-poor “gray” shale lithofacies members (GSS, GMS, and GMD) (Figure 3-5). Application of supervised SVM and ANN shows that the Bakken shale lithofacies can be predicted in a large number of test wells, using a few wells for training the lithofacies pattern associated with conventional well log signatures (Figure 3-8). SVM and ANN show accuracies of 87.3% and 83.7% respectively for lithofacies prediction (Tables 3-1 and 3-2). ANN algorithm could not predict the distribution of GSS and GMS, compared to SVM technique. This may be due to significantly small proportion of these two shale lithofacies, compared to other dominant lithofacies. This problem could be solved, either by training and testing the samples with a large number of iterations (more than 1000) to get rid of local minima problem, which is not cost-effective process or using intelligent data partitioning technique, so that at least a few samples from each shale lithofacies are present during training and testing processes (Bhattacharya and Carr, in press). Given the fact that, there are significantly small proportions of these two gray shale lithofacies in the Bakken dataset, intelligent data partitioning was not pursued.

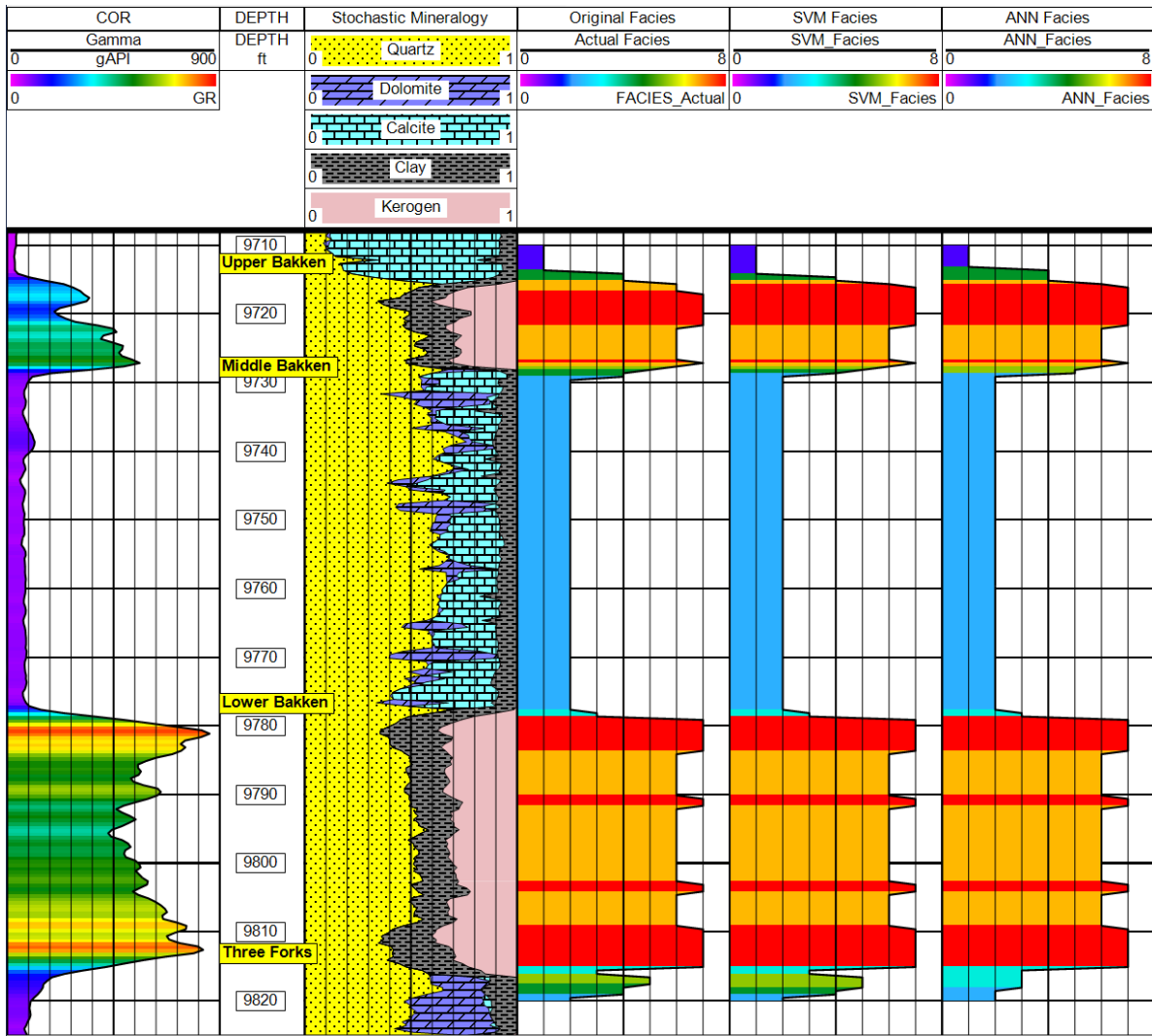


Figure 3-8. Lithofacies in the Bakken interval are classified in a well. The third track shows original lithofacies, whereas fourth, and fifth track are showing training results from Support Vector Machine (SVM) and Artificial Neural Network (ANN) algorithms. Lithofacies model generated by SVM looks significantly similar to the original lithofacies model, compared to the ANN model.

Table 3-1. Confusion matrix illustrating accuracy of Support Vector Machine algorithm in lithofacies prediction in the Bakken test dataset. Overall accuracy by SVM algorithm is ~87.3%.

Lithofacies		Actual Lithofacies						
		1	2	3	4	5	6	7
Predicted Lithofacies	1	25	3	0	0	0	0	0
	2	0	140	0	3	2	0	0
	3	0	0	6	0	2	1	1
	4	1	0	0	1	3	0	0
	5	0	0	0	0	1	0	0
	6	0	0	0	0	0	45	23
	7	0	0	0	0	0	0	50
	Absolute Accuracy	96.15%	97.90%	100%	25%	12.50%	97.82%	67.60%

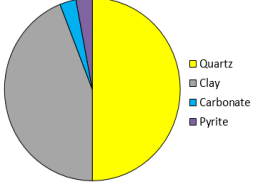
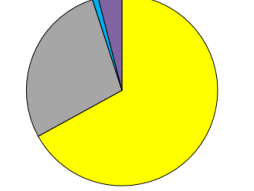
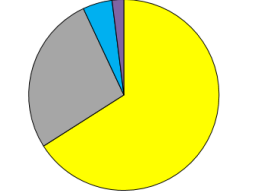
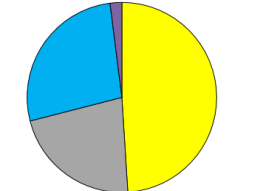
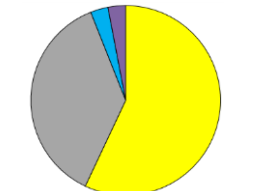
Table 3-2. Confusion matrix illustrating accuracy of Artificial Neural Network algorithm in lithofacies prediction in the Bakken test dataset. Overall accuracy by ANN technique is ~83.7%.

Lithofacies		Actual Lithofacies						
		1	2	3	4	5	6	7
Predicted Lithofacies	1	23	0	0	0	0	0	0
	2	2	125	1	1	2	0	0
	3	1	13	2	1	1	0	0
	4	0	5	1	2	1	0	0
	5	0	0	2	0	4	1	0
	6	0	0	0	0	0	44	17
	7	0	0	0	0	0	1	57
	Absolute Accuracy	88.46%	87.41%	33%	50%	50.00%	95.65%	77.03%

Table 3-3 summarizes features of different shale lithofacies in the Bakken Formation, in terms of mineralogy, TOC, and conventional well log responses. Figure 3-9 shows the vertical

lithofacies proportions in the upper and lower Bakken shale, and Figure 3-10 shows the 3D shale lithofacies model of the Bakken Formation at the regional scale. The figures depict the vertical and lateral variation of different shale lithofacies in the Bakken interval. However, the areal extent of the shale members are significantly larger than their vertical thickness (maximum 100 feet), which makes it difficult for enhanced visualization and interpretation. Lithofacies logs are upscaled separately for the upper Bakken and the remaining part of the Bakken Formation (middle and lower member), prior to application of SIS algorithm (Figure 3-11). The newly generated 3D models (for upper and lower Bakken shale members) are presented in the form of a “simbox” (in Petrel™, Schlumberger, 2012) for better visualization of lithofacies distribution pattern (Figure 3-12). In the “simbox” view, the topography of the model is flattened (Schlumberger, 2012). The total number of cells in the 3D geocellular models is around 1.72 million, with a dimension of each of cell of 101x143x1 feet.

Table 3-3. Summary of the mineralogical features, TOC content, and conventional well log responses of the five shale lithofacies defined from core and advanced well logs in the Bakken shale of the Williston basin.

Lithofacies	Pie Plot	Average Proportion (wt %)		Gamma (API)	Bulk Density (g/cc)	Neutron Ø (%)	Ln(Deep Resistivity) (ohm-m)	Umaa (barns/cc)
		Quartz:	TOC:					
Organic Mudstone (OMD)		Quartz: 52 Carb: 3 Clay: 45 TOC: 12.5		<u>337-873</u> 605	<u>2.05-2.3</u> 2.18	<u>30-48</u> 39	<u>2.1-6.56</u> 4.33	<u>6-9.6</u> 7.8
Organic Siliceous Shale (OSS)		Quartz: 70 Carb: 1 Clay: 29 TOC: 12.4		<u>303-750</u> 527	<u>2.06-2.38</u> 2.22	<u>22-40</u> 31	<u>2.4-5.8</u> 4.1	<u>5.5-9.5</u> 7.5
Gray Siliceous Shale (GSS)		Quartz: 67 Carb: 5 Clay: 28 TOC: 4.2		<u>150-347</u> 249	<u>2.44-2.6</u> 2.52	<u>18-26</u> 22	<u>1.03-2.3</u> 1.67	<u>8.3-10.6</u> 9.45
Gray Mixed Shale (GMS)		Quartz: 51 Carb: 27 Clay: 22 TOC: 2.08		<u>101-202</u> 152	<u>2.55-2.68</u> 2.62	<u>15-27</u> 21	<u>1.1-3.8</u> 2.45	<u>8.8-12.6</u> 10.7
Gray Mudstone (GMD)		Quartz: 60 Carb: 3 Clay: 37 TOC: 3.86		<u>238-410</u> 324	<u>2.43-2.5</u> 2.47	<u>26-36</u> 31	<u>1.3-3.2</u> 2.25	<u>8.8-9.84</u> 9.32

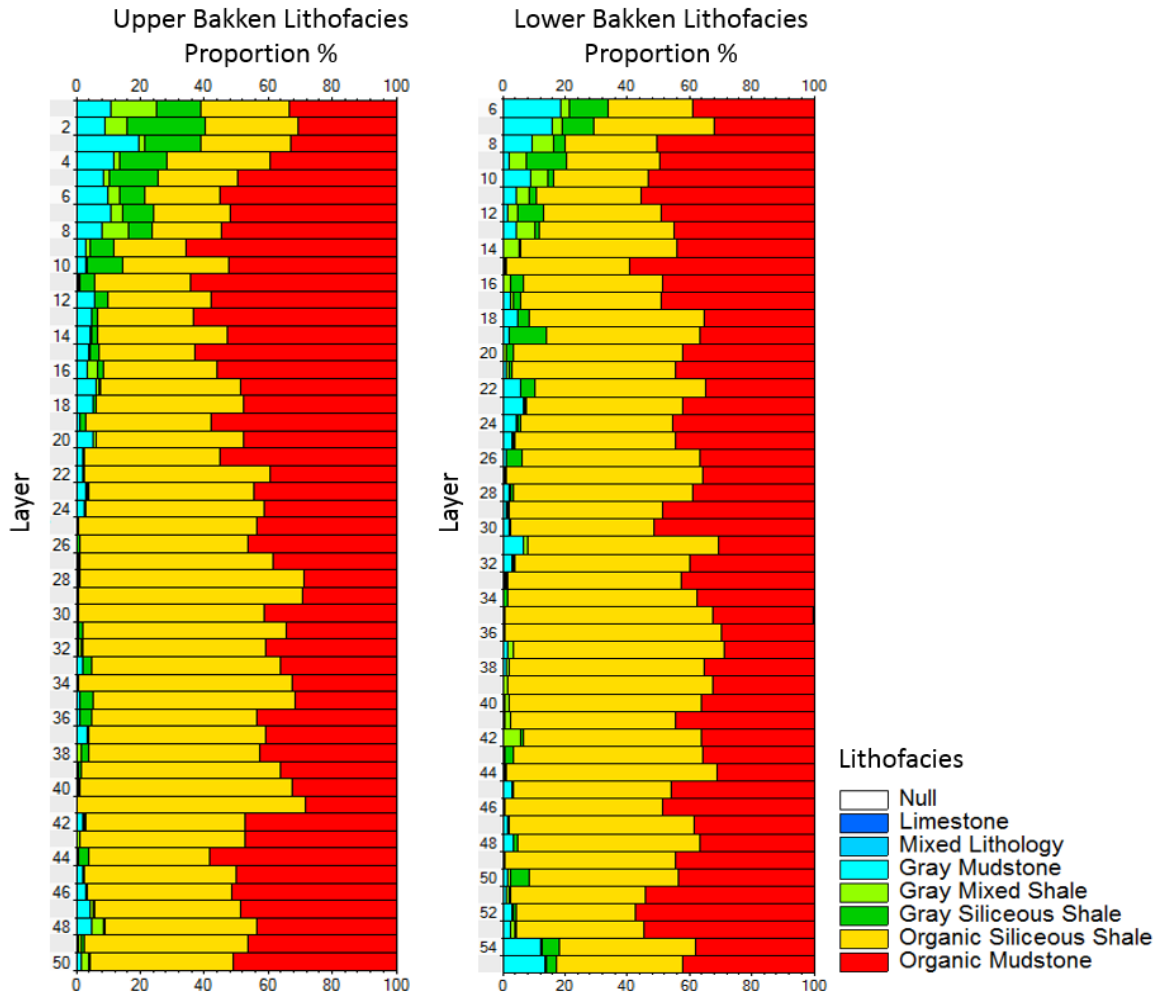


Figure 3-9. Lithofacies proportion in the upper and lower Bakken shale members. Both upper and lower shale members are classified into 50 different layers, to preserve the details of lithofacies distribution information during 3D modeling.

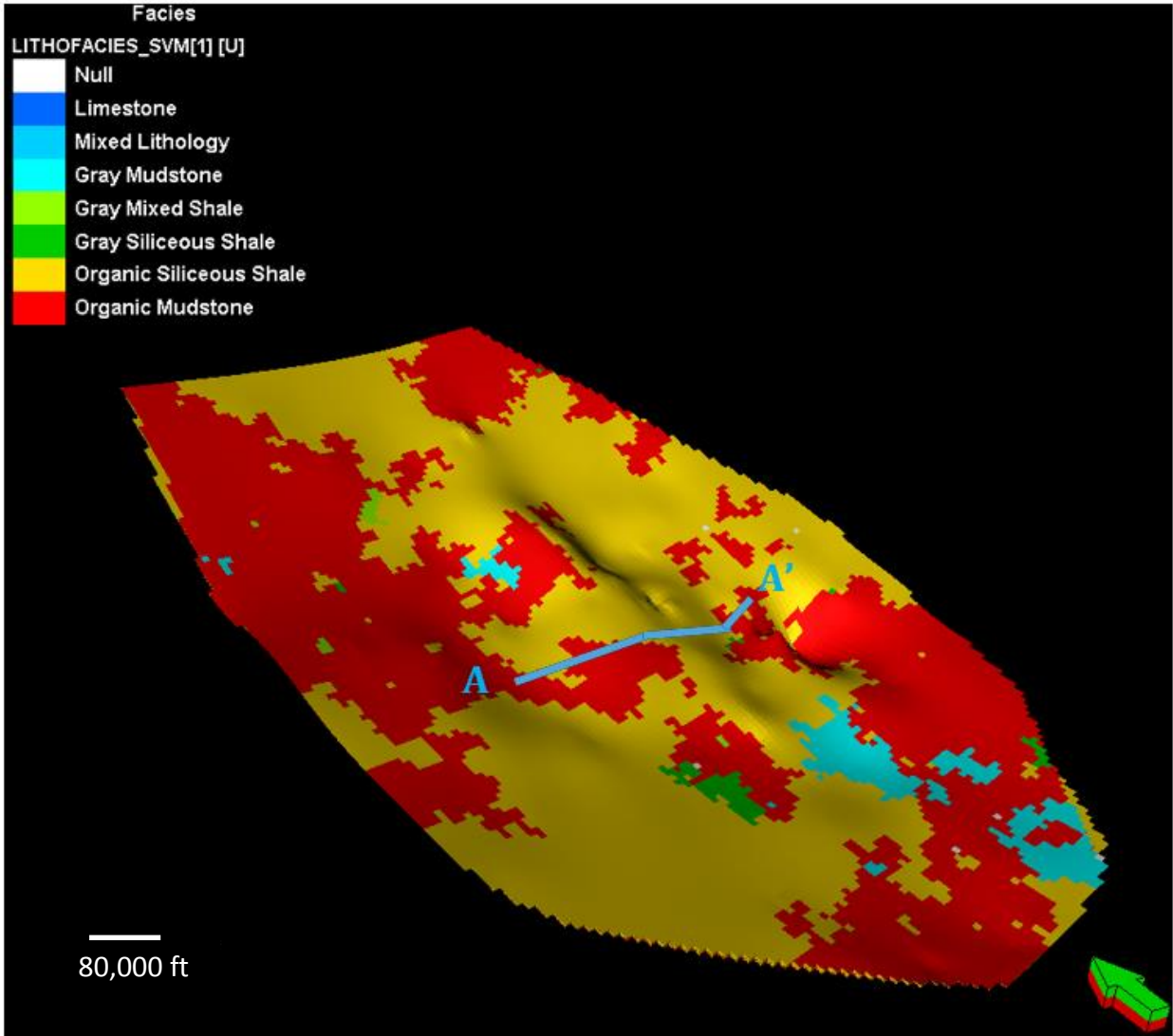


Figure 3-10. 3D shale lithofacies model of the Bakken Formation. AA' line represents the well cross-section in Figure 3-11.

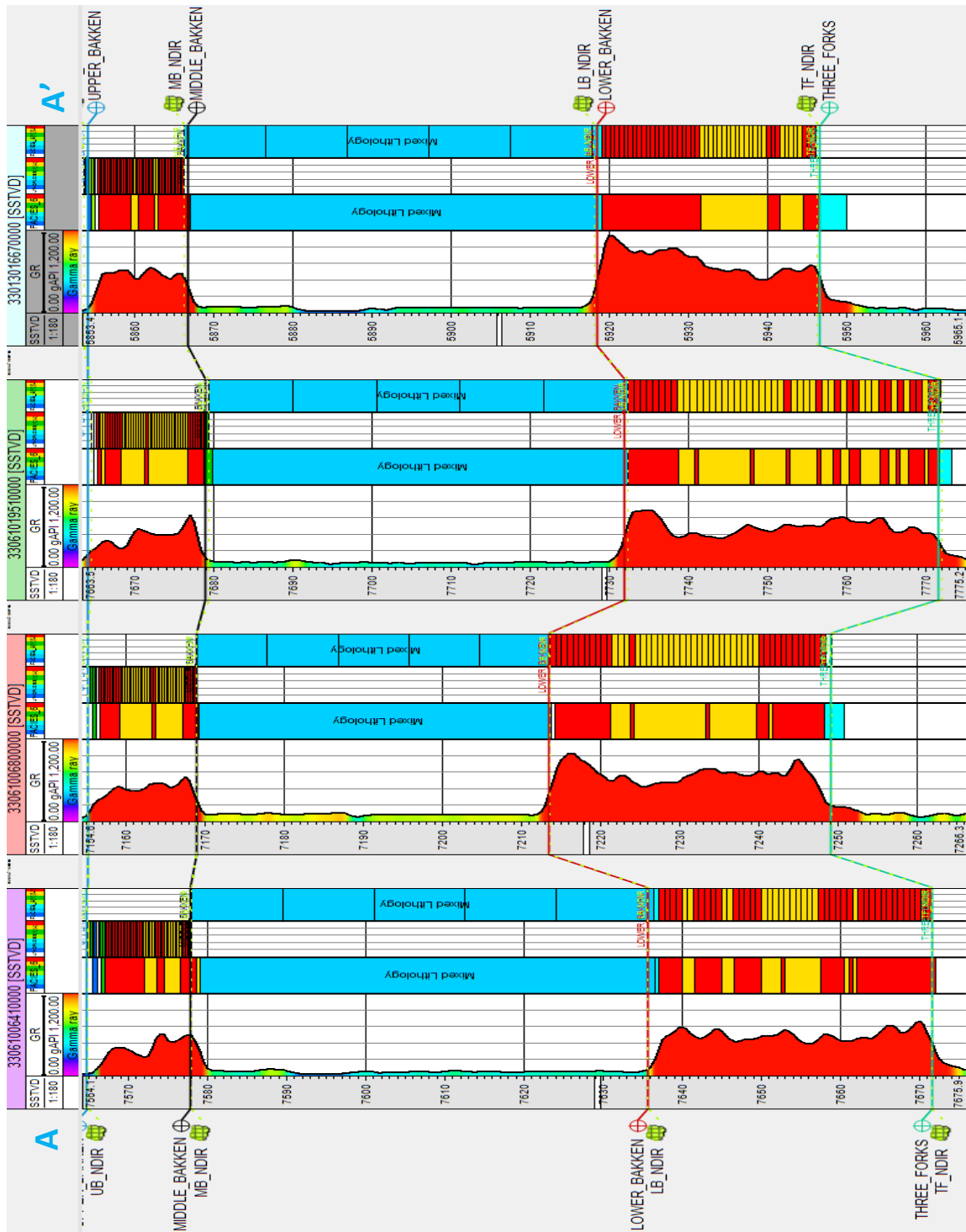
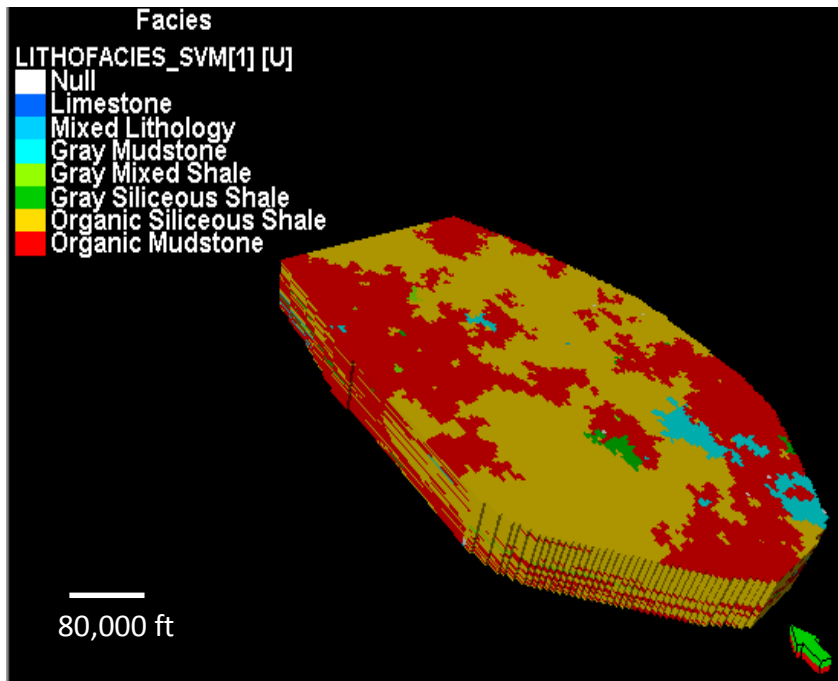
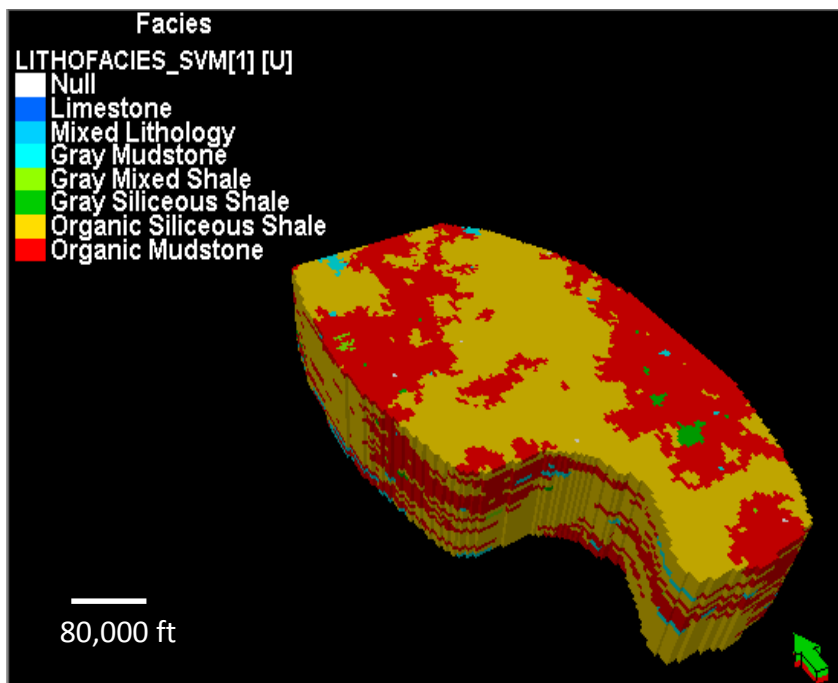


Figure 3-11. An example of lithofacies upscaling in the Bakken Formation. Lithofacies logs are upscaled separately for the upper Bakken, and the remaining part of the Bakken (middle plus lower members) separately. The first and second track show the GR log and original lithofacies, whereas third and fourth track show upscaled lithofacies logs. The well cross-section is flattened on the upper Bakken top and it is oriented along east-west (AA') (Figure 3-10).



(a)



(b)

Figure 3-12. 3D shale lithofacies models of the upper shale (a) and lower shale (b) members in “simbox” view

SIS algorithm used here for 3D modeling, is different from other available geostatistical techniques (such as Indicator Kriging and Gaussian Simulation) in a way that it can provide a non-parametric statistical framework for quantitative geological studies, without assuming any normal data distribution (Hansen, 1993). SIS algorithm can be used when the lithofacies lack a strong geometric pattern or a clear ordering in three different dimensions (Wang and Carr, 2013).

Figures 3-13a and 3-13b show the 3D shale lithofacies models of the upper and lower Bakken shale members. To test the robustness of the 3D shale lithofacies models, a limited number of wells (90% of total) were chosen for upscaling and generating the 3D lithofacies models again, while keeping all other parameters same. The results (Figures 3-13c and 3-13d) show that the 3D models are robust and they can be used for geological interpretation. OMD lithofacies can be observed surrounding the OSS lithofacies in both shale members near the central portion of the basin, which coincides with the depocenter. The major proportion of the OSS lithofacies are found near the basin center. Most of the gray shale lithofacies such as GMD, GMS, and GSS are found near the basin margin. Figure 3-14 shows lithofacies cross-section along east-west. Lithofacies packages are thicker near the Nesson Anticline, which could be the depocenter of the Williston basin.

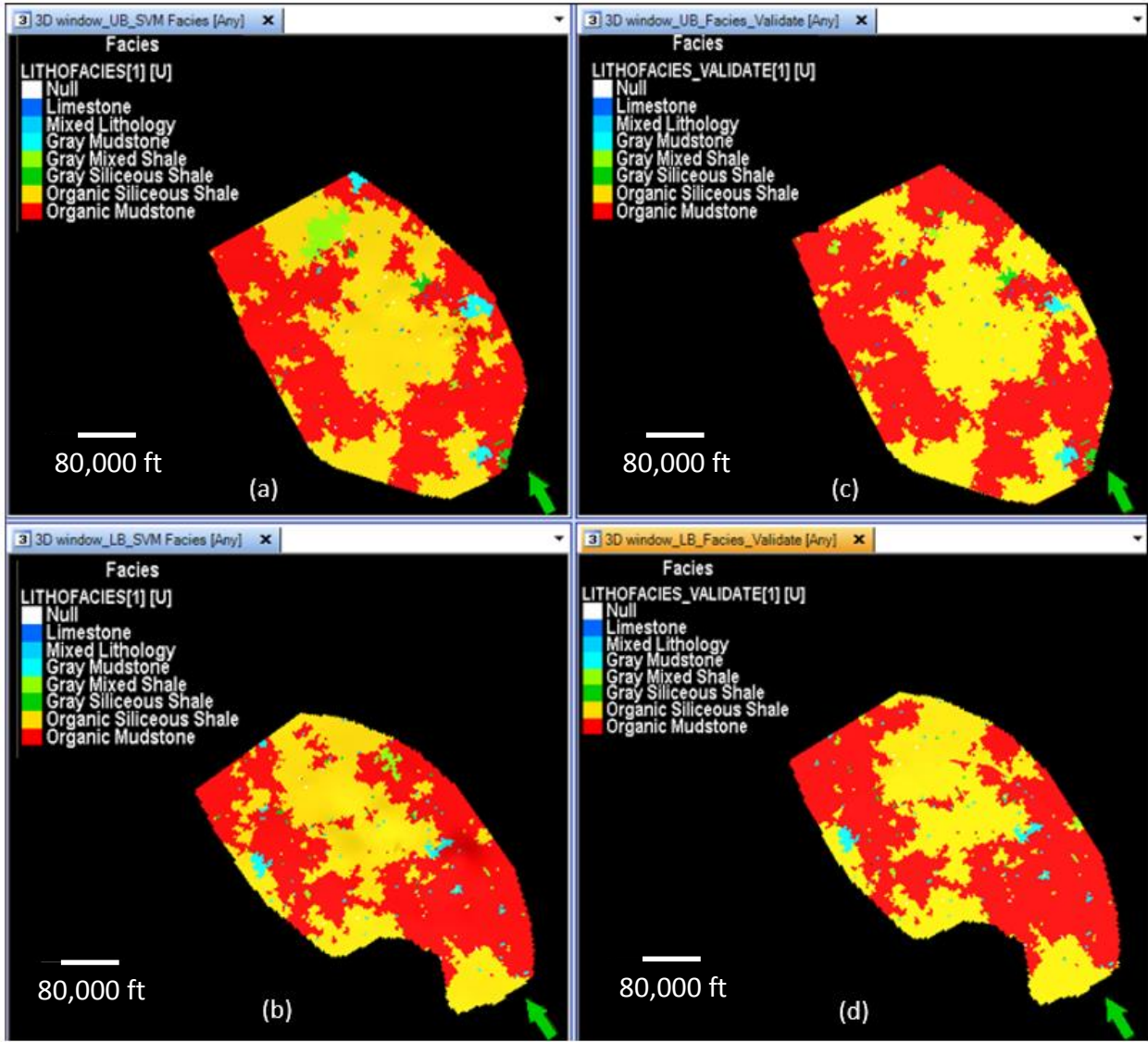


Figure 3-13. 3D shale lithofacies model validated using 90% of total number of wells. The validation of the 3D models is done for both upper and lower Bakken members (Figures 3-13c and 3-13d).

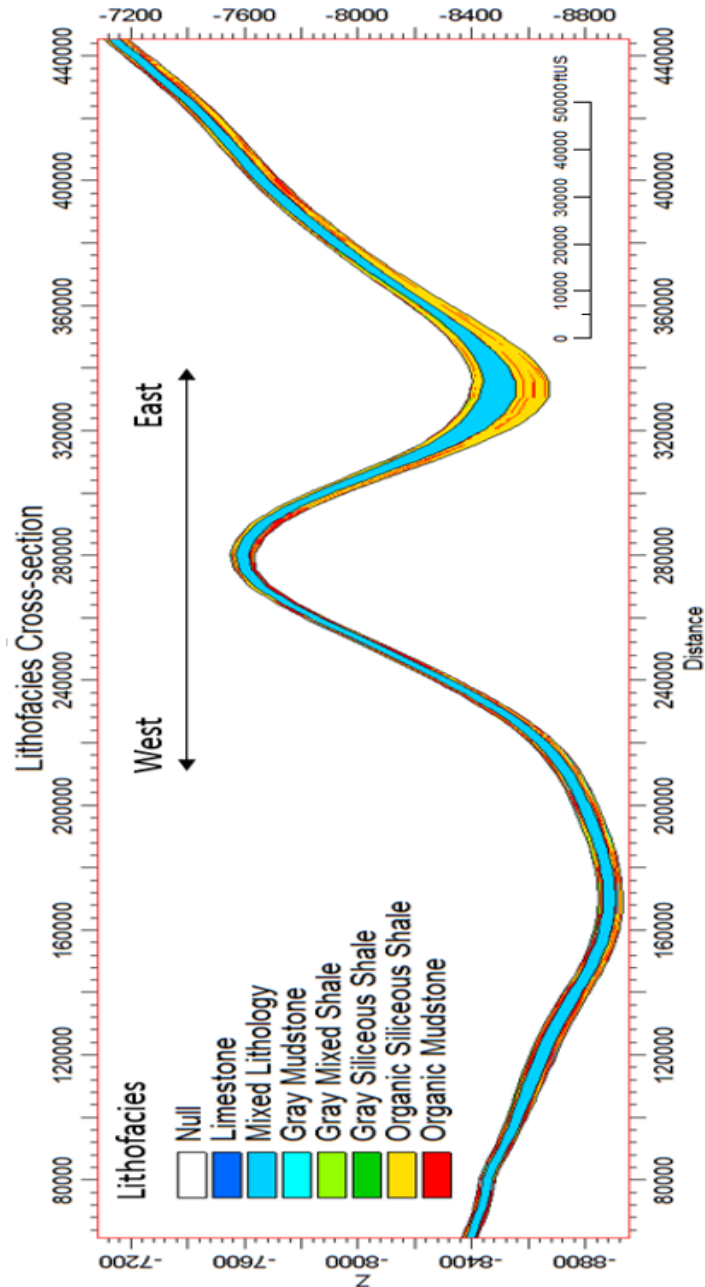


Figure 3-14. Lithofacies cross-section along east-west direction (along AA'). Nesson Anticline can be observed near the central portion of the basin. Middle Bakken member (colored blue) is assumed to be composed of one lithofacies- mixed lithology.

It appears several factors contributed to the Bakken shale lithofacies distribution pattern. The depositional pattern of different lithofacies in shale can be controlled by the interplay of source of minerals (terrigenous versus marine), organic matter productivity, and preservation

etc. (Arthur and Sageman, 2005; Carr et al., 2011; Sageman et al., 2003; Wang, 2012). Figure 3-15 shows the net thickness of OSS lithofacies in both upper and lower Bakken shale members.

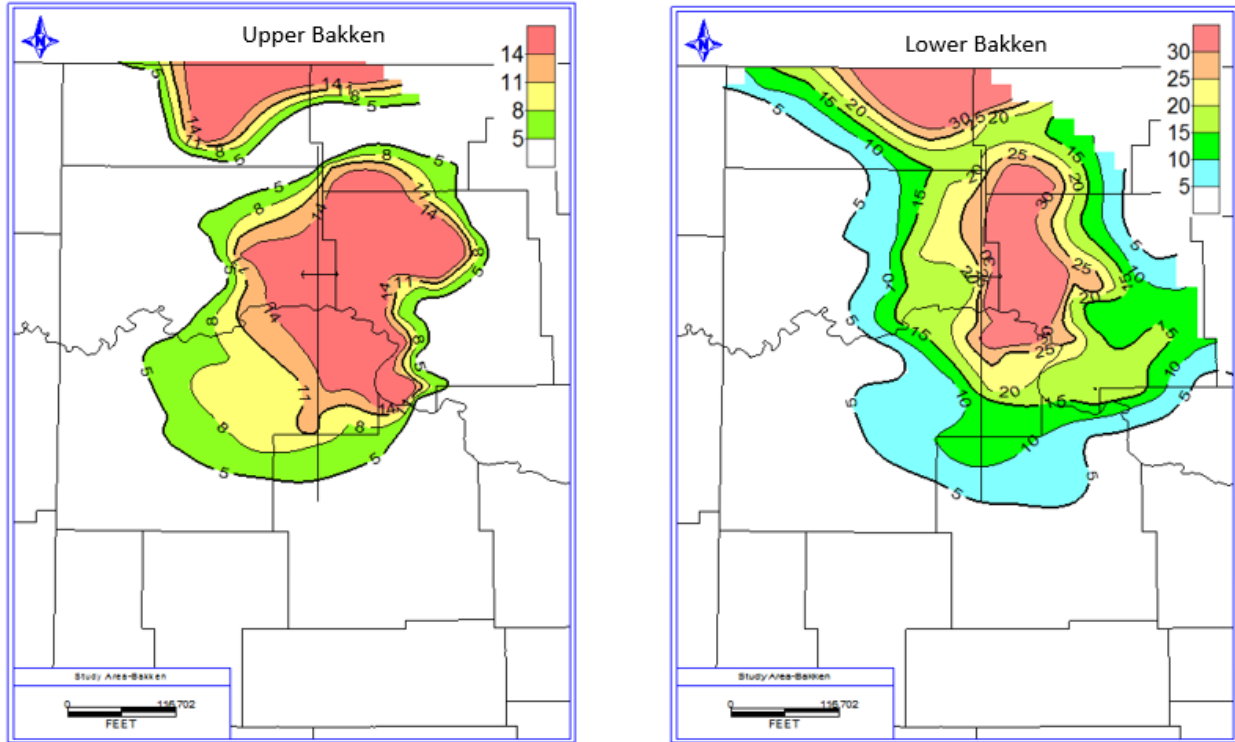


Figure 3-15. Net Organic Siliceous Shale (OSS) lithofacies thickness map of the upper and lower Bakken shale members. A major portion of thick OSS lithofacies is present near the central portion of the basin, along with the northern margin of the study area. Areas with OSS lithofacies thickness less than 5 feet are colored white, to better understand the subtle depositional signal of shale lithofacies.

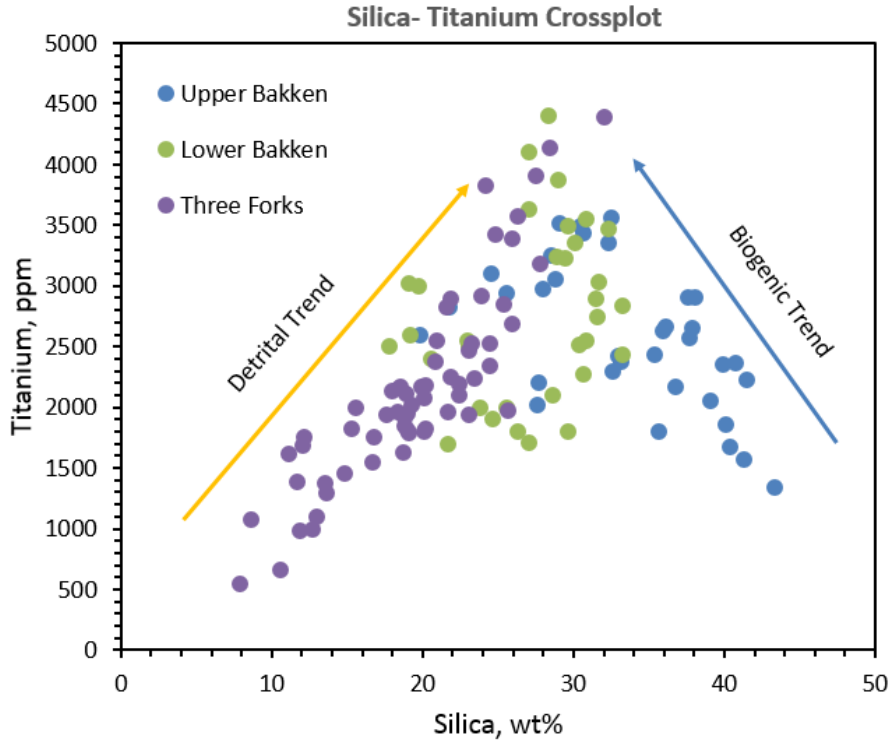


Figure 3-16. Silica versus titanium crossplot showing variation of silica content with titanium in the upper, lower Bakken members, and Three Forks Formation.

Silica in the shale lithofacies can come from detrital fractions (terrigenous) and/or biogenic components (e.g. radiolaria). Relative proportion of marine biogenic silica increases towards the distal part of the basin, which in this case, is near the central portion of the study area (near the Nesson Anticline). Silica in the OSS lithofacies near the basin center could be related to biogenic silica, whereas OSS lithofacies near the basin margin could be associated with eolian action. Cross-plot between XRF-derived titanium and silica shows that silica content increases steadily with titanium in the basal part of the Bakken or the Three Forks Formation, whereas it shows an opposite trend for significant portions of the upper and lower Bakken members (Figure 3-16). The positive relation between titanium and silica can be interpreted as detrital trend, and the negative relationship can be attributed to biogenic trend, because

titanium can be used as a proxy for detrital environment (Bertrand et al., 1996; Sageman et al., 2003). However, distinguishing biogenic and detrital silica, using well logs and “semi-quantitative” analytical geochemical data is hard, unless aided by thin section studies. Egenhoff and Fishman (2013) and Albert (2014) found evidences of radiolaria in the Bakken mudstone units, using petrography. The eolian transportation of terrigenous silica is possible, because of the prevailing trade-wind off the Canadian Shield, blowing near the equatorial position, where the Williston basin was located during the deposition of the Bakken Formation. Such silica can mix with algal matter near the basin margin, be lithified together, and classified as OSS lithofacies, based on core and well log data-driven lithofacies modeling scheme.

Organic matter in the OMD lithofacies could be attributed to the presence of algal matter near the basin margin. XRD data suggests that the clay type is mostly illite in the Bakken shale members. GMD, GMS, and GSS lithofacies are mostly found near the transition between the upper and lower Bakken shale members, with corresponding overlying and underlying carbonate units along the vertical direction, whereas they are mostly found near the basin margin along the lateral direction. Organic matter content is low in these lithofacies, which could be interpreted due to higher wave energy in surface water near the basin boundary, higher oxygen level, and effect of dilution by terrigenous matter, reducing the chance of kerogen preservation in sediment.

Using core derived XRF data, Bhattacharya and Carr (2016) showed that a significant proportion of the organic-rich shale lithofacies are deposited in euxinic and anoxic redox conditions, whereas gray shale lithofacies are deposited in mostly dysoxic conditions. Low

concentration of trace elements in gray shale lithofacies is associated with dysoxic and relatively oxic environments, because these trace elements are soluble in dysoxic and relatively oxic environments, and therefore they are not detected by measurements (Adams and Weaver, 1958; Algeo and Maynard, 2004; Bhattacharya and Carr, 2016; Nandy, 2014; Tribovillard et al., 2006). Presence of gray shale lithofacies at the boundary of organic shale lithofacies (OSS and OMD) makes geological sense, which can be explained either by increased supply of oxygen and/or gradual decrease in sea level for the deposition of calcareous middle Bakken member and the Lodgepole Formation.

Productivity of organic matter is hypothesized to be another major driver for the deposition of the dominant OSS and OMD lithofacies in the upper and lower shale members. The depositional pattern of OMD lithofacies suggests significant organic productivity, which can be related to upwelling. High productivity of organic matter is related to high supply of nutrients, which implies more organic carbon reaches the bottom water, and depletes oxygen level (Rabalais, 2010). Such a condition can be favorable for anoxia/euxinia. Euxinia in the Bakken shale members has been suggested by Nandy et al., (2015), based on $\delta^{34}\text{S}$ isotope data. On the other hand, Bakken shale members are significantly rich in trace elements. Some of these trace elements, such as, copper and nickel can be used as minor micronutrients by planktons (Algeo and Maynard, 2004). The significant source of nutrients is thought to be wind-blown.

The 3D geological models show no significant impact of the Nesson Anticline on the depositional pattern of different Bakken shale lithofacies. This may be due to great water depth (~100 feet) during deposition of the upper and lower Bakken shale members, which is why the

Nesson Anticline did not influence the depositional lithofacies. Apart from all these factors controlling Bakken shale lithofacies distribution pattern, we cannot account for the possible effects of paleo-salinity and paleo-acidity on diagenesis of different shale lithofacies, due to lack of relevant data.

3.5 Suggestions for Further Study

The present study can be extended to portions of Montana and Canada to understand the Bakken shale lithofacies variation at a larger basin scale. We may expect to find more gray shale lithofacies, as the basin is shallower towards the Canadian portion, and there could be significantly stronger effect of trade-winds off the Canadian Shield. This may affect hydrocarbon production. Although it is important to understand lithofacies distribution of the Bakken source rocks, this study can be integrated with lithofacies model of the middle Bakken member and the Three Forks Formation (reservoirs), and various engineering parameters such as lateral length, number of hydraulic stimulation stages, and proppant types etc. Such an integrated study could enable better understanding of the interplay of different geologic and non-geologic factors controlling hydrocarbon production from the non-shaly reservoir components. In addition, we anticipate applying the proposed shale lithofacies modeling scheme to other mudstone systems worldwide to check its robustness.

3.6 Conclusions

In this study we integrated core derived geochemical data with conventional well logs, and advanced PNS logs to build 3D shale lithofacies models of the Bakken Formation. The key points from the study can be enumerated as following:

1. Both upper and lower Bakken shale members are vertically and laterally heterogeneous, in terms of mineralogy and organic matter richness, which are classified into five different shale lithofacies (OMD, OSS, GSS, GMS, and GMD). The presence of all these lithofacies at core, well, and regional scales indicate that the classified shale lithofacies are geologically meaningful, mappable, and consistent at multiple scales.
2. Bakken shale members are primarily siliceous mudstone. Both upper and lower Bakken shale members have high silica and clay content. Silica in the OSS lithofacies can be derived from biogenic components and/or detrital fraction (eolian action). It appears that source of sediments, organic productivity, and anoxia control the distribution pattern of the Bakken shale lithofacies at regional scale significantly.
3. Quantitative methods such as SVM and ANN can be applied efficiently to understand relation of different shale lithofacies to various well log signatures (i.e. physical properties). These kinds of pattern recognition tools are useful to classify and predict lithofacies in wells, with limited a priori geological knowledge. Application of such powerful machine learning algorithms and quantitative geological modeling with soft data are significantly important, not only to better understand geology of the study area, but also help in reservoir management decisions. Usually there are minimal core data, scattered advanced well logs, and insufficient 3D seismic volumes in a petroliferous basin. In this study we showed that we can use ubiquitous conventional logs calibrated to limited core and advanced well log suites

to identify and characterize different shale lithofacies of the Bakken Formation to better understand their depositional and diagenetic controls on mineralogy and TOC, plus predict their distribution at the regional scale.

4. As geologic conditions for different shale formations are unique in their nature, this study and the corresponding lithofacies modeling techniques discussed here, can provide comprehensive methodology that can be used to better understand the depositional history and hydrocarbon exploration in the Williston basin, which perhaps can be applied to facilitate further research on mudrock systems in other basins.

Acknowledgments

We would like to thank Julie LeFever of North Dakota Geological Survey, PDC Energy, and Newfield Exploration for providing access to well log and core data. We thank Dr. Mahesh Pal for sharing his ideas about automated pattern recognition techniques. We thank Shared Research Facilities at West Virginia University, National Energy Technology Laboratory in Morgantown and Actlabs for analytical measurements. We would also like to express our thanks to Schlumberger (Petrel™), Jason (PowerLog™), and IHS (Petra™) for providing necessary computing facilities at West Virginia University.

References

Adams, J.A.S., Weaver, C.E., 1958, Thorium to Uranium Ratios as Indications of Sedimentary Processes- example of Concept of Geochemical Facies, AAPG Bulletin, **42**, 387-430.

Al-Anazi, A., Gates, I.D., 2010, A support vector machine algorithm to classify lithofacies and model permeability in heterogeneous reservoirs, *Engineering Geology*, **114**, 267-277.

Albert, Stacie, 2014, *Sedimentology, Facies Architecture and Sequence Stratigraphy of the Lower Bakken Shale Member in the Williston Basin, North Dakota, United States of America*, Master's Thesis, Colorado State University.

Algeo, T.J., Maynard, J.B., 2004, Trace-element Behavior and Redox Facies in Core Shales of Upper Pennsylvanian Kansas-type Cyclothems, *Chemical Geology*, **206**, 289-318.

Arthur, M. A., Sageman, B.B., 2005, Sea-level control on source-rock development: perspectives from the Holocene Black Sea, the Mid-Cretaceous western interior basin of North America, and the Late Devonian Appalachian basin, in Harris, N.B., Pradier, B., eds., *The Deposition of Organic Carbon-rich Sediments: Models, Mechanisms and Consequences*, SEPM Special Publication No. 82, 35-59.

Bertrand, P., Shimmield, G., Martinez, P., Grousset, F., Jorissen, F., Paterne, M., Pujol, C., Bouloubassi, I., Buat Menard, P., Peypouquet, J.P., Beaufort, L., Sicre, M.A., Lallier-Verges, E., Foster, J.M., Ternois, Y., 1996, The Glacial Ocean Productivity Hypothesis: the Importance of Regional Temporal and Spatial Studies, *Marine Geology*, **130**, 1-9.

Bhattacharya, S., Carr, T., 2016, Integrated Petrofacies Characterization and Interpretation of Depositional Environment of the Bakken Shale in the Williston Basin, North America, *Petrophysics*, **57** (2), 95-110.

Bhattacharya, S., Carr, T., Pal, M., in press, Comparison of Supervised and Unsupervised Approaches for Mudstone Lithofacies Classification: Case Studies from the Bakken and Mahantango-Marcellus Shale, USA, *Journal of Natural Gas Science & Engineering*.

Bishop C.M., 1995, *Neural Networks for Pattern Recognition*, New York, Oxford University Press.

Carr, T. R., Wang, W., Boyce, M.L., Yanni, A., 2011, Understanding controls on deposition of organic content in the Middle Devonian organic-rich shale intervals of West Virginia and western Pennsylvania, presented at Northeast (46th Annual) and North-Central (45th Annual) Joint Meeting, March 20–22, 2011, **43** (1), 50.

Dahl, T.W., Ruhl, M., Hammarlund, E.U., Canfield, D.E., Rosing, M.T., Bjerrum, C.J., 2013, Tracing Euxinia by Molybdenum Concentrations in Sediments using X-ray Fluorescence Spectroscopy (HHXRF), *Chemical Geology*, **360-361**, 241-251.

Egenhoff, S.O., Fishman, N.S., 2013, Traces in the dark—sedimentary processes and facies gradients in the Upper Shale Member of the Upper Devonian–Lower Mississippian Bakken Formation, Williston basin, North Dakota, U.S.A, *Journal of Sedimentary Research*, **83**, 803-824.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H., 2009, The WEKA Data Mining Software: An Update; *SIGKDD Explorations*, **11** (1), 10-18.

Kocman, K.B., 2014, Interpreting Depositional and Diagenetic Trends in the Bakken formation Based on Handheld X-ray Fluorescence Analysis, McLean, Dunn, and Mountrail Counties, North Dakota, Master's Thesis, Colorado School of Mines.

Kordon, A.K., 2010, *Applying Computational Intelligence: How to Create Value*, Berlin Heidelberg, Springer-Verlag.

LeFever, J.A., 1991, History of Oil Production from the Bakken Formation, North Dakota, *in* Hanson, W.B., ed., 1991 *Guidebook to Geology and Horizontal Drilling of the Bakken Formation*: Billings, Montana Geological Society, 3–17.

LeFever, J. A., LeFever, R.D., Nordeng, S.H., 2011, Revised nomenclature for the Bakken Formation (Mississippian- Devonian), North Dakota, *In*: Robinson, J.W., LeFever, J.A., Gaswirth, S.B., eds., *The Bakken-Three Forks Petroleum System in the Williston Basin*: Denver, Colo., Rocky Mountain Association of Geologists, 11–26.

Maldonado, D.N., 2012, *Chemostratigraphy and Geochemical Constraints on the Deposition of the Bakken formation, Williston Basin, eastern Montana and western North Dakota*, Master's Thesis, University of Texas at Arlington.

Meissner, F.F., 1978, *Petroleum Geology of the Bakken Formation, Williston Basin, North Dakota and Montana*, *in* Estelle, D., and Miller, R., eds., *The Economic Geology of the Williston Basin, 1978 Williston Basin Symposium*, Billings, Montana: Montana Geological Society, 207–230.

Mitchell, W.K., Nelson, R.J., 1988, *A Practical Approach to Statistical Log Analysis*, Paper S. Transactions, presented at the SPWLA 29th Annual Logging Symposium.

Nandy, D., Sonnenberg, S., Humphrey, J.D., 2014, *Application of Inorganic Geochemical Studies for Characterization of Bakken Shales, Williston Basin, North Dakota and Montana*, Paper URTEC-

1922974, presented at the Unconventional Resources Technology Conference, DOI 10.15530/urtec-2014-1922974.

Nandy, D., Sonnenberg, S.A., Humphrey, J.D., 2015, Factors Controlling Organic-richness in Upper and Lower Bakken Shale, Williston Basin: An Application Inorganic Geochemistry, presented at AAPG Annual Convention & Exhibition,

http://www.searchanddiscovery.com/pdfz/documents/2015/10775nandy/ndx_nandy.pdf.html

(accessed March 22nd, 2016).

Newton, R., 2012, Water Column Redox Proxies, presented at Life and Planet-NERC.

Pitman, J.K., Price, L.C., LeFever, J.A., 2001, Diagenesis and Fracture Development in the Bakken Formation, Williston Basin: Implications for Reservoir Quality in the Middle Member, U.S. Geological Survey Professional Paper 1653, U.S. Department of the Interior/U.S. Geological Survey, 19.

Pollastro R.M., Roberts, N.R., Cook, T.A., 2013, Geologic Assessment of Technically Recoverable Oil in the Devonian and Mississippian Bakken Formation, Chapter 5,

http://pubs.usgs.gov/dds/dds-069/dds-069_w/contents/REPORTS/69_W_CH_5.pdf (accessed

April 3rd, 2014).

Qi, L.S., Carr, T.R., 2006, Neural network prediction of carbonate lithofacies from well logs, Big Bow and Sand Arroyo Creek fields, Southwest Kansas, Computers & Geosciences, **32**, 947–964.

Rabalais, N.N., 2010, Hypoxia in Shelf Seas: Global Perspective

<http://www.bonusportal.org/files/646/Rabalais.pdf> (accessed March 22, 2016).

Rowe, H.D., Hughes, N., Robinson, K., 2012, The Quantification and Application of Handheld Energy-dispersive X-ray Fluorescence (ED-XRF) in Mudrock chemostratigraphy and Geochemistry, *Chemical Geology*, **324-325**, 122-131.

Sageman, B.B., Murphy, A.E., Werne, J.P., Ver Straeten, C.A., Hollander, D.J., Lyons, T.W., 2003, A Tale of Shales: the Relative Roles of Production, decomposition, and Dilution in the Accumulation of Organic-rich Strata, Middle-Upper Devonian, Appalachian Basin, *Chemical Geology*, **195**, 229-273.

Schieber, J., Zimmerle, W., Sethi, P.V., 1998a, Shales and Mudstones, Volume 1: Basin Studies, Sedimentology, and Paleontology, E. Schweizerbart'sche Verlagsbuchhandlung, Stuttgart, Germany.

Schieber, J., Zimmerle, W., Sethi, P.V., 1998b, Shales and Mudstones, Volume 2: Petrography, Petrophysics, Geochemistry, and Economic Geology, E. Schweizerbart'sche Verlagsbuchhandlung, Stuttgart, Germany.

Schieber, J., 1999, Distribution and deposition of mudstone facies in the Upper Devonian Sonyea Group of New York: *Journal of Sedimentary Research*, **69**, 909-925.

Schieber, J., Krinsley, D., Riciputi, L., 2000, Diagenetic Origin of Quartz Silt in Mudstones and Implications for Silica Cycling, *Nature*, **406**, 981-985

Schlumberger, 2012, Facies Modeling, Petrel Manual

Schmoker, J.W., Hester, T.C., 1983, Organic carbon in Bakken Formation, United States portion of Williston Basin: American Association of Petroleum Geologists, Bulletin, **67**, 2165–2174.

Sesack, S.A., 2011, Sequence Stratigraphy, Depositional Environments, and Regional Mapping of the Late Devonian Interval, Upper Three Forks Formation, Sanish Member, and Lower Bakken Shale, U.S. Portion of the Williston basin, MS Thesis, West Virginia University.

Simenson, A., 2011, Depositional Facies and Petrophysical Analysis of the Bakken Formation, Parshall field, Mountrail County, North Dakota, MS Thesis, Colorado School of Mines.

Singh, P., 2008, Lithofacies and sequence stratigraphic framework of the Barnett Shale, northeast Texas, Ph. D Dissertation, University of Oklahoma.

Smith, M.G., Bustin, R.M., 2000, Late Devonian and Early Mississippian Bakken and Exshaw Black Shale Source Rocks, Western Canada Sedimentary Basin: A sequence stratigraphic interpretation, AAPG Bulletin, **84**, 940-960.

Sonnenberg, S.A., Pramudito, A., 2009, Petroleum Geology of the Giant Elm Coulee Field, Williston basin, AAPG Bulletin, **93**, 1127–1153.

Steptoe, A., 2012, Petrofacies and Depositional Systems of the Bakken Formation in the Williston Basin, North Dakota, MS Thesis, West Virginia University.

Wang, G., 2012, Black shale lithofacies prediction and distribution pattern analysis of Middle Devonian Marcellus shale in The Appalachian basin, northeastern U.S.A., Ph.D. Dissertation, West Virginia University.

Wang G., Carr, T.R., Ju, Y., Li, C., 2014, Identifying organic-rich Marcellus Shale Lithofacies by support vector machine classifier in the Appalachian basin, *Computers & Geosciences*, **64**, 52-60.

Warner, T., 2011, Subsurface Horizontal Microfracture Propagation within the Middle Member of the Bakken Formation, Williston basin, North Dakota: Evidence and Implications, MS Thesis, West Virginia University.

Webster, R.W., 1984, Petroleum Source Rocks and Stratigraphy of the Bakken Formation in North Dakota, *in* Woodward, J., Meissner, F.F., and Clayton, J.L., eds., *Hydrocarbon Source Rocks of the Greater Rocky Mountain Region*, Rocky Mountain Association of Geologists, 57–81.

Vapnik, V., 1995, *The nature of Statistical Learning Theory*, New York, Springer.

Zhao, T., Jayaram, V., Roy, A., Marfurt K.J., 2015, A Comparison of Classification Techniques for Seismic Facies Recognition, *Interpretation*, 29-58.

SUMMARY

The upper and lower shale members in the Bakken Formation were studied, using a variety of geological data at core, well, and regional scales. The results of this study can be summarized as following:

- The Bakken shale lithofacies can be classified, using core data (XRD and TOC), PNS logs, and stochastic petrophysical solutions, in terms of mineralogy and organic matter content. Three criteria such as clay proportion, quartz-to-carbonate ratio, and TOC content can be directly used for shale lithofacies classification.
- Conventional well log suites can be used to understand complex shale lithofacies distribution. Cross-plots of various petrophysical parameters from well logs can reveal mathematical relations among them, and the underlying geological principles. A strong correlation exists between spectral gamma components (such as uranium) and TOC, because organic matter is generally associated with uranium in the reducing condition. Spectral gamma plots showed that the clay type in the Bakken shale members is primarily illite. Quantitative estimate of TOC content can be derived from bulk density log, as kerogen has low density.
- Petrophysical parameters from conventional well log suites can be trained and tested to recognize lithofacies pattern efficiently. Ten different petrophysical parameters were chosen as input parameters to machine learning algorithms. Supervised algorithms such as SVM is the best method for lithofacies classification and prediction in reduced time, no iteration, and highly repeatable results. Accuracy of lithofacies prediction can be increased by supervising the classification problem with the “known” geological rules.

- Optimal parameters should be used, while working with machine learning algorithms. Accuracy of lithofacies training can reach upto 100%, however, there is a significant chance that the mathematical models may fail outside the training domain. Algorithm-specific parameters (such as gamma, and penalty, in case of SVM) should be fine-tuned after a large number of experiments. Relatively low values of gamma and penalty parameter helped to balance the model during training and testing with the Bakken dataset.
- Sequential Indicator Simulation (SIS) performed well for populating all classified and predicted lithofacies in the 3D grid, covering the portion of the Williston basin in North Dakota. Bakken shale members are vertically and laterally heterogeneous at core, well, and regional scales. Both upper and lower shale members are composed of five different lithofacies such as GMD, GMS, GSS, OSS, and OMD. Organic-rich shale lithofacies dominate over the the orgnaic-poor gray shale lithofacies proportions.
- Cyclic depositional pattern exists between the dominant organic-rich OSS and OMD lithofacies. On the other hand, gray shale lithofacies such as GMD, GMS, and GSS are rare, and they are mostly found at the transitions between organic-rich shale lithofacies, with corresponding overlying and underlying non-shaly layers (such as the Lodgepole, Three Forks formations, and the middle Bakken member). Based on trace element geochemical data, OSS and OMD lithofacies are found to be deposited mostly in euxinic and anoxic conditions, whereas GMD, GMS, and GSS lithofacies are considered to be deposited mostly in the dysoxic-redox conditions.
- The distribution pattern of the Bakken shale lithofacies shows significant controls of source of sediment, transportation mechanism, paleo-redox conditions, organic matter productivity,

and preservation etc. This pattern is exhibited by the 3D stochastic lithofacies models and net OSS lithofacies thickness maps for the upper and lower Bakken shale members in the Williston basin.

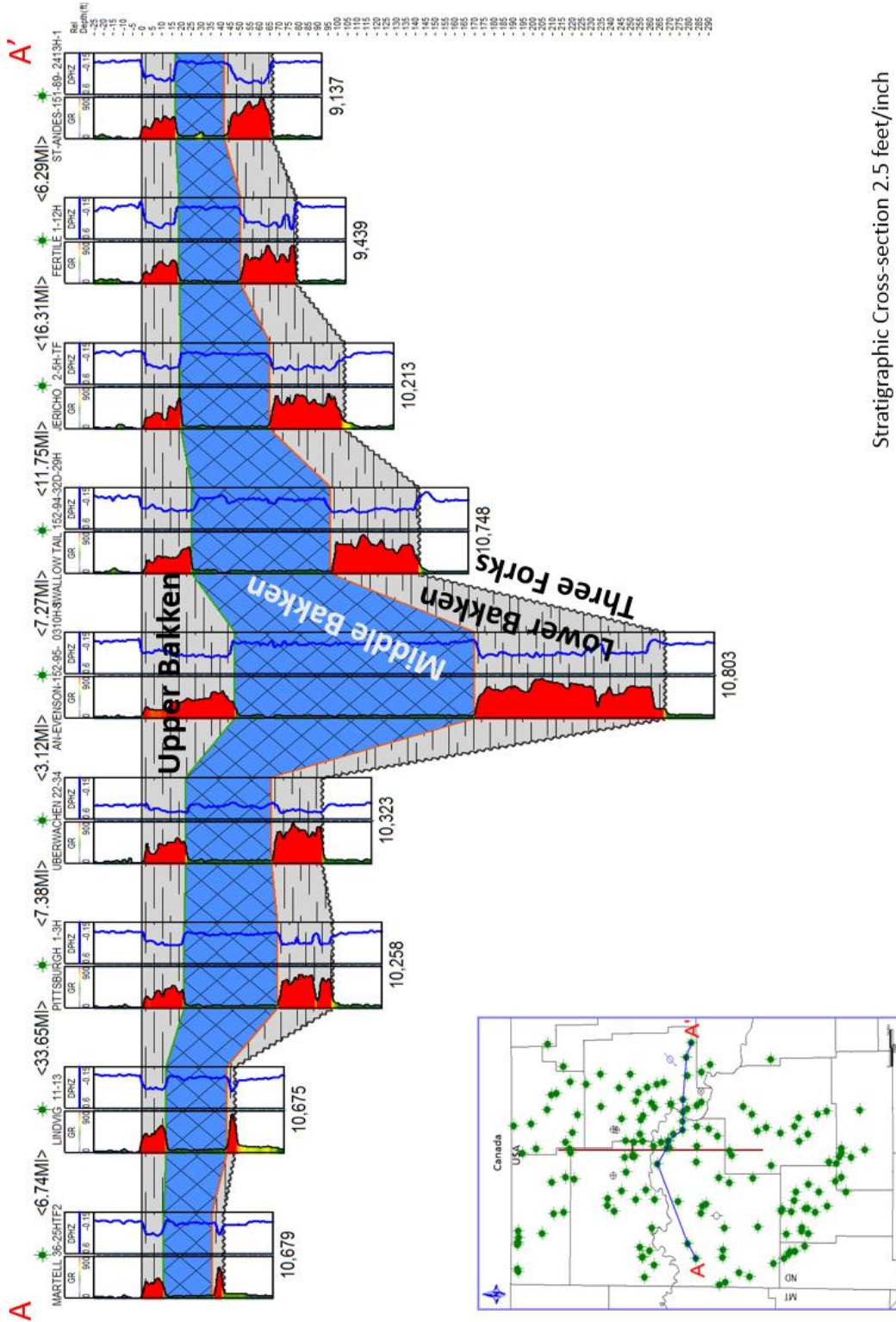
APPENDICES

Appendix A-1: Total Organic Carbon (TOC) content and Trace Element Geochemical data for the Sikes State Well. TOC data were measured at NETL laboratory, Morgantown, whereas trace element data were measured by Actlabs, Canada.

Depth (feet)	TOC (wt%)	Ti (wt%)	Mo (ppm)	U (ppm)	V (ppm)	Cu (ppm)	Ni (ppm)
9726.38	11.81	0.13	299	48.9	265	84	300
9727.25	10.26	0.1	210	40.5	213	64	220
9728.48	13.58	0.26	350	56.2	182	98	410
9728.83	12.23	0.31	343	68.3	203	99	1110
9779	20.13	0.3	451	68	1190	123	710
9780.2	21.47	0.26	499	80.8	1580	100	830
9785.35	18.76	0.24	412	101	1150	96	640
9785.95	14.56	0.2	298	67.7	869	57	410
9788.35	13.56	0.19	347	49.7	924	59	430
9791.34	14.25	0.18	266	62.6	709	95	420
9794.33	13.09	0.18	212	45.5	510	66	280
9796.43	14.63	0.21	281	55.6	728	80	390
9800.02	16.16	0.25	209	68.9	498	63	300
9803.02	14.46	0.2	236	56.5	337	62	290
9808.4	22.44	0.17	267	298	322	110	340

9809.6	17.4	0.17	242	71	309	67	320
9814.09	19.15	0.25	238	84.7	211	78	250
9817.39	3.92	0.41	8	9.1	125	46	70
9819.48	2.16	0.44	1	3.8	125	31	60

Appendix A-2: A well log-based cross-section along AA', showing different Bakken members



Stratigraphic Cross-section 2.5 feet/inch