Synthetic geomechanical logs and distributions for marcellus shale

Mohammad Omidvar Eshkalak
West Virginia University

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

Recommended Citation
Omidvar Eshkalak, Mohammad, "Synthetic geomechanical logs and distributions for marcellus shale" (2013). Graduate Theses, Dissertations, and Problem Reports. 585.
https://researchrepository.wvu.edu/etd/585

This Thesis 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 Thesis 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 Thesis 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.
SYNTHETIC GEOMECHANICAL LOGS AND DISTRIBUTIONS FOR MARCELLUS SHALE

By

Mohammad Omidvar Eshkalak

Thesis submitted to the
Benjamin M. Statler College of Engineering and Mineral Resources
At West Virginia University
In partial fulfillment of the requirements
For the degree of
Master of Science
In
Petroleum and Natural Gas Engineering
Approved by
Shahab D. Mohaghegh, PhD., Chair
Sam Ameri, Professor.
Razi Gaskari, PhD.

Department of Petroleum and Natural Gas Engineering

Morgantown, West Virginia
2013

Keywords: Marcellus Shale, Geomechanical Logs, Reservoir Stress Profiles, Artificial Intelligence, Data-Driven Modeling

Copyright 2013 Mohammad Omidvar Eshkalak
ABSTRACT

Synthetic Geomechanical Logs and Distributions for Marcellus Shale

Mohammad O. Eshkalak

The intent of this study is to generate synthetic Geomechanical Logs for a specific Marcellus Shale asset using Artificial Intelligence and Data mining Technology. Geomechanical Distributions (Map and Volume) for the entire Marcellus Shale asset was completed. In order to accomplish the objectives, conventional well logs such as Gamma Ray and Bulk Density are used to build Data-Driven models. The Data-Driven technique used in this study is applicable to other shale reservoirs.

Successful recovery of hydrocarbons from the reservoirs, notably shale, is attributed to realizing the key fundamentals of reservoir rock properties. Having adequate and sufficient information regarding the variable lithology and mineralogy is crucial in order to identify the “right” pay-zone intervals for shale gas production. In addition, contribution of mechanical properties (Principal stress profiles) of shale to hydraulic fracturing strategies is a well-understood concept. It may also contribute to better, more accurate simulation models of production from shale gas reservoirs.

In this study, synthetic Geomechanical logs (Including following properties: Poisson’s Ratio, Total Minimum Horizontal Stress, Bulk and Shear Modulus, etc.) are developed for more than 50 Marcellus Shale wells. Using Artificial Intelligence and Data Mining (AI&DM), data-driven models are developed that are capable of generating synthetic Geomechanical logs from conventional logs such as Gamma Ray and Density Porosity. The data-driven models are validated using wells with actual Geomechanical logs that have been removed from the database to serve as blind validation wells. In addition, having access to necessary data to building Geomechanical distributions (Map and Volume) model can assist in understanding the rock mechanical behavior and consequently creating effective hydraulic fractures that is considered an essential step in economically development of Shale assets.

Moreover, running Geomechanical logs on a subset of wells, but having the luxury of generating logs of similar quality for all the existing wells in a Shale asset can prove to be a sound reservoir management tool for better reservoir characterization, modeling and efficient production of Marcellus Shale reservoir.
Acknowledgments

I would like to thank Professor Shahab Mohaghegh for being a fantastic advisor & friend and sometimes-tolerant one, too! Working with Shahab was a very joyful experience. I learnt a lot from him in both academic and personal level.

My appreciation goes to Professor Sam Ameri and Dr. Razi Gaskari who generously accepted to be a member of my thesis committee. Thank them for their valuable comments and recommendations.

I have been fortunate to spend my time with PEARl members. I am thankful to my friends for making some great moments, to name some of the soon-be doctors, Masoud Kalantari, Soodabeh Esmaili, Shohre Amini, Vida Gholami and many others for providing a friendly environment and for making helpful scientific discussions, which sometimes ends with sharing stories of life that I gained invaluable life experience.

Above all, I would like to thank my beautiful wife Roya for being everything I could ask for and more in a partner and best friend. This dissertation is as much hers as it is mine.

I am and always will be thankful to Roya’s parents, Zohre and Kazem and my parents, Ameneh and Ahmad for encouraging me all the way to the end. My own success is a reflection of the love and support they have given me.
Dedication

To Roya,

Who stood by me, through Rain and Shine.
# Table of Contents

Acknowledgments ..................................................................................................................... vii  
ABSTRACT .................................................................................................................................. vi  
List of Figures .............................................................................................................................. xi  
Chapter 1 ................................................................................................................................... 14  
Introduction .............................................................................................................................. 14  
  1.1 Unconventional Gas Reservoirs ...................................................................................... 15  
  1.2 Geomechanical Logs and Properties ............................................................................ 18  
  1.3 Marcellus Shale Formation ......................................................................................... 24  
  1.4 Artificial Intelligence and Data-Driven Modeling ...................................................... 25  
  1.5 Synthetic Geomechanical Logs ............................................................................... 31  
  1.6 Geomechanical Models-Distributions .................................................................... 32  
Chapter 2 ................................................................................................................................... 33  
Methodology ............................................................................................................................. 33  
  2.1 Data Preparation ...................................................................................................... 34  
  2.2 Data-driven Model Development ............................................................................ 37  
  2.3 Validation of Data-driven Models ............................................................................ 41  
  2.4 Error Analysis ....................................................................................................... 42  
  2.5 Wells Selection Criteria for NN Training ................................................................. 43  
  2.6 Geomechancial Property Distribution .................................................................. 47  
Chapter 3 ................................................................................................................................... 50  
Results and Discussion ............................................................................................................. 50  
  3.1 KPI analysis: .................................................................................................................. 52  
  3.2 Scenario # 1- Training with 25 wells and output of 5 Blind wells ......................... 57  
    3.2.1 Error analysis of Scenario# 1 ................................................................. 69  
  3.3 Scenario # 2- Training with 5 wells and output of 25 Blind wells ....................... 72  
    3.3.1 Error analysis of Scenario# 2 ................................................................. 73  
  3.4 Scenario # 3- Training with 8 wells and output of 22 Blind wells ....................... 76  
    3.4.1 Error analysis of Scenario# 3 ................................................................. 77  
Chapter 4 ................................................................................................................................... 80
List of Figures

Figure 1. Major shale gas basins in the United States .............................................................. 17
Figure 2. Shear Modulus definition .......................................................................................... 20
Figure 3. Bulk Modulus definition ........................................................................................... 21
Figure 4. Young modulus Definition........................................................................................ 21
Figure 5. Poisson’s Ratio .......................................................................................................... 22
Figure 6. Total minimum Horizontal Stress Definition............................................................ 23
Figure 7. Neural Network Architecture .................................................................................... 26
Figure 8. Maximum, Average and Minimum of Sonic porosity vs. Bulk Density ................. 44
Figure 9. Maximum, Average and Minimum Sonic Porosity vs. Gamma Ray ....................... 45
Figure 10. Maximum, Average and Minimum of Bulk Density vs. Gamma Ray .................. 46
Figure 11. Marcellus Shale Gas Filed Under Study-80 wells ................................................... 50
Figure 12. 50 Wells with Conventional Logs in Marcellus Shale asset .................................. 51
Figure 13. 30 Wells with Geomechanical Logs in Marcellus Shale asset ............................ 51
Figure 14. KPI Analysis- Output is Bulk Density ................................................................. 52
Figure 15. KPI Analysis- Output is Sonic Density ................................................................. 53
Figure 16. KPI Analysis- Output is Young Modulus ............................................................... 53
Figure 17. KPI Analysis- Output is Bulk Modulus ................................................................. 54
Figure 18. KPI Analysis- Output is Shear Modulus ................................................................. 54
Figure 19. KPI Analysis- Output is Poissons Ratio ................................................................. 55
Figure 20. KPI Analysis-Output is Total Min. Hor. Stress ....................................................... 55
Figure 21. Scenario # 1 well distributions .............................................................................. 57
Figure 22. Well # 1 in Scenario 1- Actual Vs. NN generated well logs .................................. 58
Figure 23. Well # 2 in Scenario 1- Actual Vs. NN generated well logs .................................. 59
Figure 24. Well # 3 in Scenario 1- Actual Vs. NN generated well logs .................................. 60
Figure 25. Well # 4 in Scenario 1- Actual Vs. NN generated well logs ................................... 61
Figure 26. Well # 5 in Scenario 1- Actual Vs. NN generated well logs ................................... 62
Figure 27. Bulk Density and Gamma Ray for all wells Vs. Well # 5 ....................................... 64
Figure 28. Bulk Modulus, Poison’s Ratio and Shear modulus of all wells Vs. Well # 5 ....... 65
Figure 29. Young Modulus, TMHS and Sonic porosity of all wells Vs. Well # 5 .................... 66
Figure 30. Data-driven generated Sonic Porosity Vs. Actual Log (Wrong log) for Well # 5 .. 67
Figure 31. Well # 5 in Scenario 1- Actual Vs. NN generated well logs ................................... 68
Figure 32. Young Modulus Error histogram for scenario #1 .................................................... 69
Figure 33. Shear Modulus Error histogram for scenario #1 .................................................... 69
Figure 34. Bulk Modulus Error histogram for scenario #1 .................................................... 70
Figure 35. Poisson’s Ratio Error histogram for scenario #1 .................................................... 70
Figure 36. Total min. Hor. Stress Error histogram for scenario #1 ......................................... 71
Figure 37. Scenario # 2 well distributions .............................................................................. 72
Figure 38. Young Modulus Error histogram for scenario #2 .................................................... 73
Figure 39. Bulk Modulus Error histogram for scenario #2 .................................................... 73
Figure 40. Shear Modulus Error histogram for scenario #2 .................................................... 74
Figure 41. Total min. Hor. Stress Error histogram for scenario #2 .......................................... 74
Figure 42. Poisson’s Ratio Error histogram for scenario #2 ..................................................... 75
Figure 43. Young Modulus Error histogram for scenario #3 .................................................... 77
Figure 44. Bulk Modulus Error histogram for scenario #3 ....................................................... 78
Figure 45. Shear Modulus Error histogram for scenario #3 ..................................................... 78
Figure 46. Poisson’s Ratio Error histogram for scenario #3 ..................................................... 79
Figure 47. Total min. Hor. Stress Error histogram for scenario #3 .......................................... 79
Figure 48. Distribution of Young Modulus based on actual data (Top) and Distribution with full-field data. (Bottom) ........................................................................................................... 82
Figure 49. Distribution of Shear Modulus based on actual data (Top) and Distribution with full-field data. (Bottom) ........................................................................................................... 83
Figure 50. Distribution of Poisson’s Ratio based on actual data (Top) and Distribution with full-field data. (Bottom) ........................................................................................................... 84
Figure 51. Distribution of Total Min. Hor. Stress based on actual data (Top) and Distribution with full-field data. (Bottom) ........................................................................................................... 85
Figure 52. Distribution of Bulk Modulus based on actual data (Top) and Distribution with full-field data. (Bottom) ........................................................................................................... 86
Figure 53. Actual Vs. NN generated Geomechanical logs for Wells # 1 and 2 ......................... 87
Figure 54. Actual Vs. NN generated Geomechanical logs for Wells # 3 and 4 ......................... 88
Figure 55. Actual Vs. NN generated Geomechanical logs for Wells # 5 and 6 ......................... 89
Figure 56. Actual Vs. NN generated Geomechanical logs for Wells # 7 and 8 ......................... 90
Figure 57. Actual Vs. NN generated Geomechanical logs for Wells # 9 and 10 ....................... 91
Figure 58. Actual Vs. NN generated Geomechanical logs for Wells # 11 and 12 ..................... 92
Figure 59. Actual Vs. NN generated Geomechanical logs for Wells # 13 and 14 ..................... 93
Figure 60. Actual Vs. NN generated Geomechanical logs for Wells # 15 and 16 ..................... 94
Figure 61. Actual Vs. NN generated Geomechanical logs for Wells # 17 and 18 ..................... 95
Figure 62. Actual Vs. NN generated Geomechanical logs for Wells # 19 and 20 ..................... 96
Figure 63. Actual Vs. NN generated Geomechanical logs for Wells # 21 and 22 ..................... 97
Figure 64. Actual Vs. NN generated Geomechanical logs for Wells # 23 and 24 ..................... 98
Figure 65. Actual Vs. NN generated Geomechanical logs for Well # 25 ................................. 99
Figure 66. Actual Vs. NN generated Geomechanical logs for Wells # 1 and 2 ......................... 100
Figure 67. Actual Vs. NN generated Geomechanical logs for Wells # 3 and 4 ......................... 101
Figure 68. Actual Vs. NN generated Geomechanical logs for Wells # 5 and 6 ......................... 102
Figure 69. Actual Vs. NN generated Geomechanical logs for Wells # 7 and 8 ......................... 103
Figure 70. Actual Vs. NN generated Geomechanical logs for Wells # 11 and 12 ..................... 104
Figure 71. Actual Vs. NN generated Geomechanical logs for Wells # 13 and 14 ..................... 105
Figure 72. Actual Vs. NN generated Geomechanical logs for Wells # 15 and 16 ..................... 106
Figure 73. Actual Vs. NN generated Geomechanical logs for Wells # 17 and 18 ..................... 107
Figure 74. Actual Vs. NN generated Geomechanical logs for Wells # 19 and 20 ..................... 108
Figure 75. Actual Vs. NN generated Geomechanical logs for Wells # 21 and 22 ..................... 109
Figure 76. Actual Vs. NN generated Geomechanical logs for Wells # 23 and 24 ..................... 110
List of Tables

Table 1. Methodology............................................................................................................... 34
Table 2. Average value of the Input data for shale and non-Shale Formation ...................... 35
Table 3. Information used for data-sets for developing Data-driven models ......................... 57
Table 4. Information used for data-sets for developing Data-driven models ....................... 72
Table 5. Scenario # 2 well distributions ................................................................................... 76
Table 6. Information used for data-sets for developing Data-driven models ....................... 76
Chapter 1

Introduction

Well logging, also known as borehole logging is the practice of making a detailed record (a well log) of the geologic formations penetrated by a borehole. The oil and gas industry uses logging to obtain a continuous record of a formation's rock properties. Well logging has been used for almost one century in petroleum industry as an essential tool for determination and classification of potential production formations in oil and gas reservoirs. Log analysts interpret the data from the log in order to determine the petro physical parameters of the well. In this study, geomechanical well logs are generated with help of neural network and data mining. These geomechanical logs include Total Minimum Horizontal Stress, Poisson’s Ratio, Young, Bulk and Shear modulus.

For economic reasons, companies do not always run all the logs for all available wells that are required to analyze reservoir characteristics and behavior. This study presents a methodology that can help to solve the aforementioned problem by generating synthetic well logs for those locations (Wells) where the set of logs that are necessary to analyze the reservoir properties, are absent or are not complete. Synthetic geomechanical logs are generated with the same quality of the actual ones which is been validated with. The intention of the technique used here is not to eliminate well logging in a field but it meant to become a tool for reducing costs for companies whenever logging proves to be insufficient and/or difficult to obtain. The approach presented involves the use of artificial neural networks- Data Driven modeling, as the main tool, in conjunction with data obtained from conventional well logs such as Bulk Density, Gamma Ray and Sonic Porosity.

After all, in this work, one of the purposes is meant to propose a technique in order to reduce cost of running geo-mechanical logs for the rest of the field after just obtaining sufficient information from some part of the asset.
1.1 Unconventional Gas Reservoirs

Unconventional Gas reservoirs continue to be a fast growing sector within the petroleum industry. These reservoirs include Coal Bed Methane, Shale Gas, and Tight Gas Sands.

Outside of the United States, with a few exceptions, unconventional gas resources have largely been overlooked and understudied. In most of the world, the natural gas industry is focusing on producing gas from conventional reservoirs and has yet to turn its attention to unconventional gas reservoirs. These unconventional gas reservoirs represent a vast, long-term, global source of natural gas and have not been appraised in any systematic way.

Unconventional gas resources—including tight sands, coal bed methane, and gas shales—constitute some of the largest components of remaining natural gas resources in the United States. Research and development concerning the geologic controls and production technologies required to evaluate and produce these unconventional gas resources has provided many new technologies during the past several decades. These new technologies have enabled operators in the United States to unlock the vast potential of these challenging resources, boosting production levels to an estimated 30% of the natural gas production in the United States.

Around the world, unconventional gas resources are widespread, but with several exceptions, they have not received close attention from natural gas operators. This is due in part because geologic and engineering information on unconventional resources is scarce, and natural gas policies and market conditions have been unfavorable for development in many countries. In addition, there is a chronic shortage of expertise in the specific technologies needed to develop these resources successfully. As a result, only limited development has taken place to date outside of North America.

Interest is growing, however, and during the last decade development of unconventional gas reservoirs has occurred in Canada, Australia, Mexico, Venezuela, Argentina, Indonesia, China, Russia, Egypt, and Saudi Arabia. Many of those who have estimated the volumes of gas in place within unconventional gas reservoirs agree on one aspect: that it is a large resource. In Error! Reference source not found., Kawata and Fujita summarized the work of Rogner, who estimated the worldwide unconventional gas resource.1 Using the United States as an analogy, there is good reason to expect that unconventional gas production will increase significantly around the world in the coming decades for the following reasons:

- A significant number of geologic basins around the world contain unconventional gas reservoirs.
Rogner estimates that in the world there are around

- 9,000 Tcf of gas in place in coalbed methane,
- 16,000 Tcf of gas in place in shale gas, and
- 7,400 Tcf of gas in place in tight gas sands.

Any reasonable recovery efficiency leads one to the conclusion that there are many opportunities in the future to develop unconventional gas worldwide.

Tight gas sand development in the United States, critical to future U.S. gas supply, has to over four Tcf/year and is supported by ongoing technological development.

The technology developed in the United States over the past 3 to 4 decades will be available for application around the world.

New technology is rapidly becoming a worldwide commodity through efforts of major service companies.

The global need for energy, particularly natural gas, will continue to be an incentive for worldwide unconventional gas resource development.

Tight gas sands, gas shales, and coalbed methane are already critical to North America today and will be an important energy source worldwide during the 21st Century.

Currently Producing Gas Shale Reservoirs-The Barnett Shale in the Fort Worth basin of North Central Texas is by far the most active shale gas play in the United States. The reservoir ranges from 100 ft to more than 1,000 ft in gross thickness and holds from 50 × 109 ft³ (50 bcf) to 200 bcf of gas per square mile. The Gas Technology Institute estimates that organic shale reservoirs in the United States contain up to 780 trillion (780 × 1012) ft³ of gas. Equally, gas-rich organic shales almost certainly exist elsewhere around the world, but so far, the United States is the only country with a large shale gas industry. Figure 1 shows the main shale gas reservoirs in US.
Shale gas wells are not hard to drill, but they are difficult to complete. In almost every case, the rock around the wellbore must be hydraulically fractured before the well can produce significant amounts of gas. Fracturing involves isolating sections of the well in the producing zone, and then pumping fluids and proppant (grains of sand or other material used to hold the cracks open) down the wellbore through perforations in the casing and out into the shale. The pumped fluid, under pressures up to 8,000 psi, is enough to crack shale as much as 3,000 ft in each direction from the wellbore. In the deeper high-pressure shales, operators pump slickwater (a low-viscosity waterbased fluid) and proppant. Nitrogen-foamed fracturing fluids are commonly pumped on shallower shales and shales with low reservoir pressures.

Fractures are the key to good production. The more fractures in the shale around the wellbore, the faster the gas will be produced. Because of shale’s extremely low permeability, the best fracture treatments are those that expose as much of the shale as possible to the pressure drop that allows the gas to flow. The natural formation pressure of a large gas shale reservoir will decline only slightly over decades of production. Any pressure drop on individual wells is likely the result of fractures closing up, rather than depletion of the reservoir. The key to good shale gas production over time is having the proper distribution and placement of proppant to keep the fractures open.

Shale gas is completely different. Shale gas wells do not come on as strong as tight gas, but once the production stabilizes, they will produce consistently for 30 years or more. Suppose that new horizontal wells in a typical shale gas play produce 1 million ft3 per day (1 MMcf/d). If the operator puts 10 such wells on 1 square mile, that section will produce 10 MMcf/d. With an estimated 120 bcf of gas per square mile in the ground, these gas shale reservoirs will be producing gas for a very long time.
time. That realization, plus increasingly effective horizontal drilling tools, 3D seismic imaging, and advanced reservoir-modeling software, has many people looking at shale gas as an important new resource.

The price of gas is linked to oil and based on each fuel’s heating value. The ratio is about 6 to 1. In other words, 1 bbl of oil contains about 6 times more heat energy than 1,000 ft³ of gas. If a barrel of oil sells for USD 50.00, then 1,000 ft³ of gas is worth about USD 8.00. As long as oil prices remain high, there is no reason for natural gas prices to go down. Although gas is abundant in much of the world, it is expensive and potentially dangerous to transport internationally.

1.2 Geomechanical Logs and Properties

Geomechanical properties are important across the full reservoir lifecycle. Investigating geomechanical properties and stress profiles of Marcellus Shale is vital since natural gas is the second largest source of primary energy in the United States. Change in stresses will result in understanding of how flow properties evolve which is still a poorly understood concept when it comes to unconventional resources. The Marcellus Shale is a complex rock, which makes accurate analysis of its geological, geomechanical and transport properties an ongoing challenge.

Geomechanical properties are important since they are affecting the production in the Marcellus shale reservoir. Since hydraulic fracturing can significantly alter the stress profiles and geomechanical properties through the reservoir, it is important to understand and predict these parameters for the entire reservoir.

The Geomechanical properties of Marcellus Shale, which analyzed in this thesis, are listed in Table 1. Geomechanical properties for the study are coming from Well Logs.

<table>
<thead>
<tr>
<th>Geomechanical Properties</th>
<th>Average Value for Marcellus Shale</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Minimum Horizontal Stress</td>
<td>0.8 psi/ft</td>
</tr>
<tr>
<td>Poisson’s Ratio</td>
<td>0.15 (Unitless)</td>
</tr>
<tr>
<td>Bulk Modulus</td>
<td>4 Mpsi</td>
</tr>
<tr>
<td>Young Modulus</td>
<td>3.5 Mpsi</td>
</tr>
<tr>
<td>Shear Modulus</td>
<td>2.5 Mpsi</td>
</tr>
</tbody>
</table>
The log data for the computation of this parameters are mainly delta t shear and compressional. In the following paragraph, there is a detailed definition of these parameters that are measured while logging.

Delta shear is explained here. The values of this log represent the acoustic slowness of a shear wave (Vs.); reciprocal of Shear Velocity; a shear wave is a type of seismic wave in which the particle motion is perpendicular to the direction of propagation; also called S-wave or secondary wave. At a more general level these values are measures of the interval transit time (The travel time of a wave over a unit distance, hence the reciprocal of wave velocity).\textsuperscript{2}

For the compressional one, this log displays readings of the time it takes for a compressional wave to traverse a fixed distance; reciprocal of Compressional Velocity. Furthermore, this can be explained by looking at the interval transit time (The travel time of a wave over a unit distance, hence the reciprocal of wave velocity).\textsuperscript{3}

When this two time intervals are measured in a formation, properties such as Sonic Porosity and geomechanical properties such as Poisson’s Ratio, Total Minimum Horizontal Stress, Bulk, Young and Shear Modulus are calculated at the same time. Some other information such as pore pressures as well as density of rock and fluid is used in their computation. In this study, this geomechancial properties as such are calculated without having those time intervals of shear and compressional. In addition, the aim of this study is not investigating how these parameters calculated.

To have an overview of these geomechanical properties that are off interest in this thesis, a short and clear explanation as well as formulas relevant to each characteristic is added.

1. Shear Modulus-SM

The shear modulus is one of several quantities for measuring the stiffness of materials. The shear modulus is concerned with the deformation of a solid when it experiences a force parallel to one of its surfaces while its opposite face experiences an opposing force (such as friction). Figure 2 is showing the forces applied in definition of Shear Modulus.

In brief, Shear modulus is the stiffness to a shearing force. This is vaguely like collapsing a cardboard box and a wooden crate. The wooden crate would have a greater shear modulus, because it would be more difficult to collapse than the cardboard box. It should be noted that, as porosity increases Shear Modulus generally decreases.\textsuperscript{4}
Figure 2. Shear Modulus definition

\[ G = \frac{\tau_{xy}}{\gamma_{xy}} = \frac{F/A}{\Delta x/l} = \frac{Fl}{A\Delta x} \]

Where

\[ \tau_{xy} = \frac{F}{A} = \text{Shear Stress} \]

F is the force which acts
A is area on which force acts
\( \Delta x \) is transverse displacement
L is the initial length

Shear modulus' derived SI unit is the pascal (Pa), although it is usually expressed in gigapascals (GPa) or in thousands of pounds per square inch (kpsi).

2. Bulk Modulus- BM

The bulk modulus of a substance measures the substance's resistance to uniform compression. The bulk modulus can be formally defined by the equation. Its base unit is the Pascal. Figure 3 is showing bulk modulus definition schematically.

In brief, bulk modulus is exactly the inverse or reciprocal of bulk compressibility. A high compressibility corresponds to a low bulk modulus. It should be noted that as porosity increases Bulk modulus generally decreases.\(^5\)
Figure 3. Bulk Modulus definition

\[ K = -V \frac{dp}{dv} \] where the p is pressure and the V is volume.

3. Young Modulus- YM

Young's modulus, also known as the tensile modulus or elastic modulus, is a measure of the stiffness of an elastic material and is a quantity used to characterize materials. It is defined as the ratio of the stress along an axis over the strain along that axis in the range of stress in which Hooke's law holds. Figure 4 is showing the schematic explanation of young modulus.

Young's modulus is the ratio of stress, which has units of pressure, to strain, which is dimensionless; therefore, Young's modulus has units of pressure.²

\[ \frac{P}{\Delta L/\Delta L} \]

Figure 4. Young modulus Definition

\[ E = \frac{\text{Tensile Stress}}{\text{Tensile Strain}} = \frac{\sigma}{\varepsilon} = \frac{F/A_0}{\Delta L/L_0} \]

Where:

E is the Young's modulus (modulus of elasticity)

F is the force exerted on an object under tension;
A0 is the original cross-sectional area through which the force is applied; 

\( \Delta L \) is the amount by which the length of the object changes; 

L0 is the original length of the object.

4. Poisson’s Ratio - PR

Poisson's ratio (\( \nu \)), named after Siméon Poisson, is the negative ratio of transverse to axial strain. In fact, when a sample object is stretched (or squeezed), to an extension (or contraction) in the direction of the applied load, it corresponds a contraction (or extension) in a direction perpendicular to the applied load. The ratio between these two quantities is the Poisson's ratio.

In brief, Poisson’s Ratio represents the amount that the sides bulge out when the top is compressed. Water is alone extreme with the maximum bulging. A sponge would be at the other extreme with no bulging.\(^7\) Figure 5 is showing the forces in defining Poisson's Ratio.

![Figure 5. Poisson’s Ratio](image)

Assuming that the material is stretched or compressed along the axial direction (the x axis in the below diagram):

\[
\nu = -\frac{\varepsilon_{\text{trans}}}{\varepsilon_{\text{axial}}}
\]

Where:

\( \nu \) is the resulting Poisson's ratio,

\( \varepsilon_{\text{trans}} \) is transverse strain (negative for axial tension (stretching), positive for axial compression)

\( \varepsilon_{\text{axial}} \) is axial strain (positive for axial tension, negative for axial compression).
5. Total Minimum horizontal Stress-TMHS or Hmin

Underground formations are confined and under stress. The stresses can be divided into three principal stresses. These are vertical, maximum and minimum stresses. These stresses are normally compressive, anisotropic, and nonhomogeneous, which means that the compressive stresses on the rock are not equal and vary in magnitude on the basis of direction.\(^8\) The magnitude and direction of the principal stresses are important because they control the pressure required to create and propagate a fracture, the shape and vertical extent of the fracture, the direction of the fracture, and the stresses trying to crush and/or embed the propping agent during production. A hydraulic fracture will propagate perpendicular to the minimum principal stress.\(^9\) In figure 6, Total Minimum Horizontal Stress is shown schematically.

The minimum horizontal Stress can be calculated with:

\[
\sigma_{\text{min}} = \frac{v}{1-v} \left( \sigma_1 - \alpha p_p \right) + \alpha p_p + \sigma_{\text{ext}}
\]

Where:

- \(\sigma_{\text{min}}\) is minimum horizontal stress
- \(v\) Poisson’s ratio,
- \(\sigma_1\) Overburden stress,
- \(\alpha\) Biot’s constant,
- \(p_p\) Reservoir fluid pressure or pore pressure
- \(\sigma_{\text{ext}}\) Tectonic stress

![Figure 6. Total minimum Horizontal Stress Definition](image-url)
To perform a broader view of minimum horizontal stress, the following is explained to make a clear-cut conception in this regard. Great advances have been made in rock mechanics, particularly in computer modeling of rocks. These models have been used to help solve problems as borehole instability, hydraulic fracturing, and sand-production prediction. A key input parameter for these models is Minimum Horizontal Stress. Wells drilled in the direction of minimum horizontal stress are potentially more favorable from the perspective of reservoir development and optimal production. In such a situation, hydraulic fractures grow transverse to the wellbore axis allowing multiple fractures to be placed without the possibility of fracture overlapping.\textsuperscript{10}

1.3 Marcellus Shale Formation

The Marcellus Shale, deposited 389.5-388 million years ago, is an important gas reservoir containing 500 trillion cubic feet (Tcf) of natural gas, of which 10\% is technically recoverable. Marcellus Formation is a unit of marine sedimentary rock found in eastern North America. Named for a distinctive outcrop near the village of Marcellus, New York in the United States, it extends throughout much of the Appalachian Basin. The shale contains largely untapped natural gas reserves, and its proximity to the high-demand markets along the East Coast of the United States makes it an attractive target for energy development.\textsuperscript{11}

The Marcellus Formation is black shale that may contain limestone beds and concentrations of iron pyrite (FeS2) and siderite (FeCO3). Its sedimentary structure, or bedding, is moderately well developed. Like most Shales, it tends to split easily along the bedding plane, a property known as fissility. Lighter colored Shales in the upper portion of the formation tend to split into small thin-edged fragments after exposure. These fragments may have rust stains from exposure of pyrite to air, and tiny gypsum (CaSO4•2H2O) crystals from the reaction between pyrite and limestone particles. Fresh exposures of the pyriteiferous shale may develop the secondary mineralization of orange limonite (FeO(OH)•nH2O), and the pale yellow efflorescence or bloom of sulfur, associated with acid rock drainage.\textsuperscript{12}

Pyrite is especially abundant near the base, and the upper contacts of limestones, but frambooidal microcrystals and ephedra crystals of pyrite occur throughout the organic-rich deposits. The Marcellus also contains uranium, and the radioactive decay of the uranium-238 makes it a source rock for radioactive radon gas.\textsuperscript{13}

Measured total organic content of the Marcellus Formation ranges from less than 1\% in eastern New York, to over 11\% in the central part of the state and the shale may contain enough carbon to support combustion. The more organic-rich black Shales can be bituminous, but are too old to contain
bituminous coal formed from land plants. In petroleum geology, these black Shales are an important source rock that filled conventional petroleum reservoirs in overlying formations, are an unconventional shale gas reservoir, and are an impermeable seal that traps underlying conventional natural gas reservoirs. To the west, the formation may produce liquid petroleum; further east heating during deeper burial more than 240 million years ago cracked this oil into gas.\textsuperscript{14}

Outcrops of the Marcellus can contain very small beds that resemble coal. The New York outcrops, and others further south in Pennsylvania and New Jersey, were extensively excavated in the early 19th century, sometimes at great expense, in the false hope of finding minable coal seams. In Perry County, Pennsylvania along the Juniata River the false coal beds become up to .3 m (1 ft) thick, but they did not produce a valuable fuel, despite the considerable effort expended to mine it from the surrounding hills. Seaweed and marine plants probably formed the false coal. True coal is formed from terrestrial plants, which only began to appear in Marcellus and later fossils.\textsuperscript{15}

Maximum thickness of the Marcellus ranges from 890 ft in New Jersey, to 40 ft in Canada. In West Virginia, the Marcellus Formation is as much as 200 ft thick. In extreme eastern Pennsylvania, it is 790 ft thick, thinning to the west, becoming only 49 ft thick along the Ohio River, and only a few feet in Licking County, Ohio. The thinning, or stratigraphic convergence, from east to west is caused by decreasing grain size in the clastic deposits, which entered the basin from the east. The beds finally "pinch out" westward because deposition was limited by the Cincinnati Arch, the bulge that formed the west shore of the basin. Where the formation is relatively thick, it is divided into several members, and as the formation continues to thicken to the east, these members are further divided. Some workers chose to classify the Marcellus as a subgroup, and classify some of the members as separate formations.\textsuperscript{16}

The Marcellus Shale filed under study in this thesis is a part of an asset which is located in Southern Pennsylvania and Northern West Virginia.

1.4 Artificial Intelligence and Data-Driven Modeling

An artificial neural network is an information processing system that has certain performance characteristics in common with biological neural networks. Therefore it is appropriate to describe briefly a biological neural network before offering a detail definition of artificial neural networks. All living organisms are made up of cells. The basic building blocks of the nervous system are nerve cells, called neurons.
An artificial neural network is a collection of neurons that are arranged in specific formations. Neurons are grouped into layers. In a multi-layer network there are usually an input layer, one or more hidden layers and an output layer. The number of neurons in the input layer corresponds to the number of parameters that are being presented to the network as input. The same is true for the output layer. It should be noted that neural network analysis is not limited to a single output and that neural nets can be trained to build neuro-models with multiple outputs. The neurons in the hidden layer or layers are mainly responsible for feature extraction. They provide increased dimensionality and accommodate tasks such as classification and pattern recognition. Figure 7 is a schematic diagram of a fully connected three-layered neural network.

There are many kinds of neural networks. Neural network scientists and practitioners have provided different classifications for neural networks. One of the most popular classifications is based on the training methods. Neural nets can be divided into two major categories based on the training methods, namely supervised and unsupervised neural networks. Unsupervised neural networks, also known as self-organizing maps, are mainly clustering and classification algorithms. They have been used in oil and gas industry to interpret well logs and to identify lithology. They are called unsupervised simply because no feedback is provided to the network. The network is asked to classify the input vectors into groups and clusters. This requires a certain degree of redundancy in the input data and hence the notion that redundancy is knowledge.

Figure 7. Neural Network Architecture

Most of the neural network applications in the oil and gas industry are based on supervised training algorithms. During a supervised training process both input and output are presented to the network to permit learning on a feedback basis. A specific architecture, topology and training algorithm is selected and the network is trained until it converges. During the training process neural network tries...
to converge to an internal representation of the system behavior. Although by definition neural nets are model-free function approximators, some people choose to call the trained network a neuro-model.

In a typical neural data processing procedure, the database is divided into three separate portions called training, calibration and verification sets. The training set is used to develop the desired network. In this process (depending on the paradigm that is being used), the desired output in the training set is used to help the network adjust the weights between its neurons or processing elements. During the training process the question arises as when to stop the training. How many times should the network go through the data in the training set in order to learn the system behavior? When should the training stop? These are legitimate questions, since a network can be over trained. In the neural network related literature over-training is also referred to as memorization. Once the network memorizes a data set, it would be incapable of generalization. It will fit the training data set quite accurately, but suffers in generalization. Performance of an over-trained neural network is similar to a complex non-linear regression analysis.

Over-training does not apply to some neural network paradigms simply because they are not trained using an iterative process. Memorization and over-training is applicable to those networks that are historically among the most popular ones for engineering problem solving. These include back-propagation networks that use an iterative process during the training.

In order to avoid over training or memorization, it is a common practice to stop the training process every so often and apply the network to the calibration data set. Since the output of the calibration data set is not presented to the network, one can evaluate network's generalization capabilities by how well it predicts the calibration set's output. Once the training process is completed successfully, the network is applied to the Verification Data Set. Back propagation method is used in this study.\(^\text{17}\)

Backpropagation is an abbreviation for "backward propagation of errors", is a common method of training artificial neural networks. From a desired output, the network learns from many inputs, similar to the way a child learns to identify a dog from examples of dogs.

The procedure for developing a back propagation neural network, which has been used in this study, is explained in this section.\(^\text{18}\)

In this algorithm, all outputs are computed using sigmoid thresholding of the inner product of the corresponding weight and input vectors. Meanwhile all outputs at each stage \(n\) are connected to all
inputs at stage n+2. The errors propagate backwards by apportioning them to each unit according to the amount of this error the unit is responsible for.

We now derive the stochastic Back propagation algorithm for the general case.

\[ \tilde{x}_j = \text{input vector for unit } j \ (x_{ji} = \text{i th input to the j th unit}) \]

\[ \tilde{w}_j = \text{Weight vector for unit } j \ (W_{ji} = \text{weight on } x_{ji}) \]

\[ z_j = \tilde{w}_j \cdot \tilde{x}_j, \text{ the weighted sum of inputs for unit } j \]

\[ o_j = \text{Output of unit } j \ (o_j = \sigma(z_j)) \]

\[ t_j = \text{target for unit } j \]

Downstream (j) = set of units whose immediate inputs include the output of j

Outputs = set of output units in the final layer

Since we update after each training example, we can simplify the notation somewhat by imagining that the training set consists of exactly one example and so the error can simply be denoted by E.

We want to calculate \( \frac{\partial E}{\partial w_{ji}} \) for each input weight \( W_{ji} \) for each output unit j. Note first that since \( z_j \) is a function of \( W_{ji} \) regardless of where in the network unit j is located,

\[ \frac{\partial E}{\partial w_{ji}} = \frac{\partial E}{\partial z_j} \cdot \frac{\partial z_j}{\partial w_{ji}} = \frac{\partial E}{\partial z_j} \cdot x_{ji} \]

Furthermore, \( \frac{\partial E}{\partial z_j} \) is the same regardless of which input weight of unit j we are trying to update. So we denote this quantity by \( \delta_j \). Consider the case when \( j \in \text{Outputs} \). We know:
\[ E = \frac{1}{2} \sum_{k \in \text{Outputs}} (t_k - \sigma(z_k))^2 \]

Since the outputs of all units \( k \neq j \) are independent of \( W_{ji} \), we can drop the summation and consider just the contribution to \( E \) by \( j \).

\[
\delta_j = \frac{\partial E}{\partial z_j} = \frac{\partial}{\partial z_j} \frac{1}{2} (t_j - o_j)^2 = -(t_j - o_j) \frac{\partial o_j}{\partial z_j} = -(t_j - o_j) \sigma'(z_j) = -(t_j - o_j)(1 - \sigma(z_j)) \sigma(z_j) = -(t_j - o_j)(1 - o_j) o_j
\]

\[
\Delta w_{ji} = -\eta \frac{\partial E}{\partial w_{ji}} = \eta \delta_j x_{ji}
\]

Thus,

Now consider the case when \( j \) is a hidden unit. Like before, we make the following two important observations.

For each unit \( k \) downstream from \( j \), \( z_k \) is a function of \( z_j \).

The contribution of error by all units \( l \neq j \) in the same layer as \( j \) is independent of \( W_{ji} \).

We want to calculate \( \frac{\partial E}{\partial w_{ji}} \) for each input weight \( W_{ji} \) for each hidden unit \( j \). Note that \( W_{ji} \) influences just \( z_j \) which influences \( o_j \) which influences \( z_k \) for each \( k \in \text{Downstream}(j) \) each of which influences \( E \). So we can write:

\[
\frac{\partial E}{\partial w_{ji}} = \sum_{k \in \text{Downstream}(j)} \frac{\partial E}{\partial z_k} \frac{\partial z_k}{\partial o_j} \frac{\partial o_j}{\partial z_j} \frac{\partial z_j}{\partial w_{ji}} = \sum_{k \in \text{Downstream}(j)} \frac{\partial E}{\partial z_k} \frac{\partial z_k}{\partial o_j} \frac{\partial o_j}{\partial z_j} \cdot x_{ji}
\]

Again, note that all the terms except \( x_{ji} \) in the above product are the same regardless of which input weight of unit \( j \) we are trying to update. Like before, we denote this common quantity by \( \delta_j \). Also note that \( \frac{\partial E}{\partial z_k} = \delta_k \), \( \frac{\partial z_k}{\partial o_j} = w_{kj} \) and \( \frac{\partial o_j}{\partial z_j} = o_j(1 - o_j) \). Substituting,
\[ \delta_j = \sum_{k \in \text{Downstream}(j)} \frac{\partial E}{\partial z_k} \cdot \frac{\partial o_j}{\partial z_j} = \sum_{k \in \text{Downstream}(j)} \delta_k w_{kj} (1 - o_j) \]

Thus,
\[ \delta_k = o_j (1 - o_j) \sum_{k \in \text{Downstream}(j)} \delta_k w_{kj} \]

We are now in a position to state the Back-Propagation algorithm formally.

Formal statement of the algorithm:

Stochastic Back-Propagation (training examples, \( \eta, n_i, n_h, n_o \)).

Each training example is in the form of \( \langle \vec{x}, \vec{t} \rangle \) where \( \vec{x} \) is the input vector and \( \vec{t} \) is the target vector. \( \eta \) is the learning rate (e.g., 0.05). \( n_i, n_h \) and \( n_o \) are the number of input, hidden and output nodes respectively. Input from unit \( i \) to unit \( j \) is denoted \( x_{ji} \) and its weight is denoted by \( W_{ji} \).

Create a feed-forward network with \( n_i \) inputs, \( n_h \) hidden units and \( n_o \) output units.

Initialize all the weights to small random values (e.g., between -0.05 and 0.05)

Until termination condition is met, Do

For each training example \( \langle \vec{x}, \vec{t} \rangle \), Do

Input the instance \( \vec{x} \) and compute the output \( o_u \) of every unit

For each output unit \( k \), calculate \( \delta_k = o_k (1 - o_k) (t_k - o_k) \)

For each hidden unit \( h \), calculate \( \delta_h = o_h (1 - o_h) \sum_{k \in \text{Downstream}(h)} w_{kh} \delta_k \)

Update each network weight \( W_{ji} \) as follows:

\[ W_{ji} \leftarrow W_{ji} + \Delta w_{ji} \]

Where \( \Delta w_{ji} = \eta \delta_j x_{ji} \)
1.5 Synthetic Geomechanical Logs

Artificial Intelligent and Data Mining have been used with in recent years in reservoir modeling and characterization to perform analysis of formation evaluation. Studies indicates that ANN is a powerful tool for pattern recognition and system identification such as methodology developed by Mohaghegh et al. (1998) to generate synthetic Magnetic Resonance Imagin (MRI) logs using conventional logs such as SP, GR and Resistivity. In their study, the methodology incorporated an artificial neural network as its main tool. The synthetic magnetic resonance imaging logs were generated with a high degree of accuracy even when the model developed used data not employed during model development.  

Basheer, I. A. shows that ANN is suitable to predict and classify soil compaction and rocks characteristics as well as to determine some mechanical parameters such as Young’s modulus, Poisson’s Ratio. They mainly investigate Neural Network capability in solving geotechnical engineering problems and they provided a general view of some Neural Network applications in their field of research.  

In another attempt, synthetic Gamma Ray logs are generated using backpropagation method with a hash-base time function. In this research they have used the times and pressures in their training process. For reflecting the time feature in their models, hash function is used through the neural network modeling. In order to validate their model by time, they have presented two categories of single well and multi wells.  

In all this studies, the formations under study are not complex like Shale rocks. The methodology used in this study is based on back propagation algorithm but not only various designs are used but also more accurate and reliable results achieved. Geomechanical properties of Marcellus shale play significant role in gas production that is related to hydraulic fracturing issues. In analyzing the hydraulic fractures, the most substantial part is geomechanics of the rock and properties.
1.6 Geomechanical Models-Distributions

A geomechanical model is a numerical representation of the state of stress and rock mechanical properties for a field. These models are primarily driven from static elastic properties, rock strength, pore pressure, in-situ stress magnitudes and stress direction.\textsuperscript{24}

Geomechanics relates to almost all technical aspects of reservoir production. Some examples of areas where geomechanics is used in the oil and gas industry include pore pressure prediction, sand production, wellbore instability issues, casing design, safe mud weight window prediction, subsidence, compaction and fully coupled simulation.\textsuperscript{25} The most critical geomechanical issues are in weak rock systems, deep-water reservoirs and tight fractured formations.\textsuperscript{26}

Furthermore, a Geomechanical Model captures the initial equilibrium state that describes the stresses, pore pressure, and mechanical properties of the rock mass in which the reservoir resides. The creation of the geomechanical model is the starting point of any geomechanical analysis for which the model will be used to evaluate the changes in hydro-thermo-mechanical conditions due to hydrocarbon production or geological processes.\textsuperscript{27}
Chapter 2

Methodology

The goal of the study is to develop Data-Driven models to generate synthetic geomechanical well logs such as Total minimum horizontal Stress, Poisson’s Ratio, Bulk, Young and Shear Modulus for a specific Marcellus shale asset. The followings explain the concept, procedure and methodology used in this study.

Principal stress profiles in a field are related to rock geomechanical properties. Geomechanical properties of rock include Poisson’s Ratio, Total Minimum Horizontal Stress, Bulk, Young and Shear Modulus. These properties play significant role in developing Shale assets in terms of fluid flow characteristics and as well as hydraulic fracturing which results in more production of an unconventional reservoirs.

Having access to geomechanical data can assist engineers and geoscientists during geomechanical modeling, hydraulic fracture treatment design and in some cases during reservoir modeling in Shale Gas fields. A common source of securing such data is geomechanical well logs. Running geomechanical well logs (in all wells in a Shale asset) is not common practice among operators. This may be attributed to the cost associated with running such logs.

In this study, it is demonstrated that Artificial Intelligence and Neural Network technology is capable of developing data-driven models to generate rock geomechanical properties. The workflow includes development of synthetic geomechanical well logs from conventional logs such as Gamma Ray and Bulk Density that are commonly available for assets.

Generally, data-driven models used 6, 10 and 30 percent of data (coming from geo-mechanical logs) of the entire asset which were available to expand them for the rest of the filed with conventional logs but no geomechanical logs. Criteria in choosing what percentage of asset for training purposes are discussed later.

In order to validate and verify the accuracy of Data-Driven models, blind wells have been defined as follows. Blind wells are wells with actual data, which selected due to different locations in the asset to generate the synthetic logs for them to show the actual and generated log in the same plot. Data-driven models are validated with these blind wells.
Highly accurate results are achieved which demonstrate the capability of Artificial Intelligence and Data mining in predicting highly nonlinear relationships between parameters. Later in this study, a method to select the blind wells and a method to select minimum amount wells of an asset in order to train the neural network models are discussed. The Marcellus Shale field under study includes 80 lateral and horizontal gas wells. Most of the wells are producing and some of them are considered as monitoring wells.

Moreover, once all data-driven models are trained and the outcome is acceptable according to validation process, then, the entire asset now has the full data of geomechanical properties. This information generated from data-driven models is used to build integrated geomechanical distribution (Maps & Volumes) of rock geomechanical properties for the entire asset. In order to create these distributions, commercial software is used with different geostatistical methods to achieve the best results. Difference between distributions when the entire asset has available data compared to the time with missed data is clearly shown in the results and discussion chapter.

The methodology is used to accomplish the objectives of this study-includes several steps. These steps showed in Table 2 below.

Table 1. Methodology

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Data Preparation</td>
</tr>
<tr>
<td>2.2</td>
<td>Data-driven Model Development</td>
</tr>
<tr>
<td>2.3</td>
<td>Validation of Data-driven Models</td>
</tr>
<tr>
<td>2.4</td>
<td>Error Analysis</td>
</tr>
<tr>
<td>2.5</td>
<td>Well Selections for Training of Data-Driven Models</td>
</tr>
<tr>
<td>2.6</td>
<td>Geomechanical Property Distribution (Maps and Volumes)</td>
</tr>
</tbody>
</table>

In the following sections, all the above steps are explained in details.

### 2.1 Data Preparation

In order to prepare the data sets as the first step of the methodology, all wells logs of the Marcellus Shale asset is turned into data file which means all the well logs, pictures, were digitized and converted to ASCI files. Afterwards, all 80 wells and correspondent well logs such as conventional and geomechanical logs are categorized in two main groups.
First group of wells which are 30 wells have both actual conventional such as Gamma Ray, Sonic porosity and Bulk Density as well as geomechanical well logs with following attributes; total minimum horizontal Stress, Poisson’s Ratio, Bulk, Young and Shear modulus . Second group includes wells with Conventional well logs such as Bulk Density, Gamma Ray and Sonic Porosity.

Another sub set of the main data sets are also defined to complete the project and training purposes. Since for asset in this study, some wells had missing data for their Sonic Porosity and some of them for Bulk Density, there is also two other data sets prepared in order to train some models which is called Conventional Models to generate the missing data of those wells in the asset. These conventional models help the data sets to be completed before training the main models in order to generate the geomechanical well logs for the entire asset.

Once data sets are categorized and arranged in the order to train the model, zone of interest which production of Marcellus Shale asset is happening, shale formation, is selected. In order to do this step, well legs are analyzed to see the best way of choosing the pay-zone. Among all logs, Gamma Ray logs are showing the best difference between Shale and non-Shale formations.

Gamma ray is an indicator of shale formation and other well logs are analyzed as well to see the right pay-zone of production. Average Gamma Ray log of below 150 is been chosen for non-shale formation and anything above is considered Shale formation of interest. Table 3 below shows average well logs used in this study for Shale and non-shale formations.

<table>
<thead>
<tr>
<th>Average Gamma Ray</th>
<th>Average Bulk Density</th>
<th>Average Sonic Porosity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shale</td>
<td>Non Shale</td>
<td>Shale</td>
</tr>
<tr>
<td>150</td>
<td>&gt;150</td>
<td>&gt;2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>&lt;2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>&lt;0.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>&gt;0.2</td>
</tr>
</tbody>
</table>

Producing pay zones of Marcellus shale for each individual well has been selected according to the Gamma Ray well logs. Difference between Shale and non-shale formations is identified for each well. Vertical depth of the wells in the assets starts from 4000 for some wells and continues to 8000 ft for the longest wells.

Production pay zones of different wells are in different depths as mentioned. Marcellus Shale thickness is different in various locations. In overall, thickness of Shale pay zones for all the wells in the asset is from 50 ft to 250 ft. Wells with higher depth of pay zone will be producing more than the others.
After identifying the depth of producing zone intervals for each well, available correspondent data that are conventional data such as Gamma Ray, Sonic Porosity and Bulk Density as well as geomechanical log data such as Total minimum horizontal stress, Poisson’s Ratio, Bulk, Young and Shear Modulus, for each well is extracted from the main data sets. These data are available from well logs in most of the wells for each 1 ft. Also in minority of wells, there is data available for every 0.5 ft. Wrong data (bad data) is been deleted during this process. This wrong data includes Negative values or either Zeros values of properties that is coming from the logging tools calibration process. These wrong values are common in well logs that are due to logging process.

Number of well logs that data is used in this study is 390 well logs. Amount of 240 well logs are for those wells which they have both conventional and geomechanical well logs. Amount of 30 wells times eight different well logs, which are 240 well logs. The amount of 150 of them entire well logs is for those wells that just have conventional well logs. Amount of 50 wells times three, which is 150 well logs.

Furthermore, the important issue related to extracting data from well logs is in identifying the contrast between Shale and Non-shale formations. In order to accomplish this issue in a right way, 50 ft above and below the Marcellus Shale interval is added to all main data sets. This data mining procedure helps the models to understand the different characteristics of various formations. This also helps in having larger data and wider range data sets for every foot of data, which is helpful in Data-Driven modeling process. These 50 feet of data has very different characteristics in terms of all well logs.

The database contains the well name, the Depth, the well Coordinates, the values for Gamma Ray (GR), Bulk Density (BD), Sonic Porosity, Bulk Modulus (BM), Shear Modulus (SM), Young’s Modulus (YM), Poisson’s Ratio (PR) and Total Minimum Horizontal Stress (TMHS) for each well. It is important to mention that not all the wells include geomechanical well logs, thus geomechanical values are only recorded for wells that have such data, which is 30 wells out of 80 wells in Marcellus Shale asset under study.

Data cleansing is the final consideration of data preparation step. Data cleansing is the process of detecting and correcting bad or inaccurate records from a database. This part is performed with some criteria such as Accuracy, consistency and Uniformity. By analyzing data using the values of mean, standard deviation, ranges, these criteria have been satisfied for all well logs for each individual well. Since data are coming form well logs, availability of outlier data is inevitable. Each well has been monitored several times for their own well logs to eliminate outlier data, which is associated to errors while logging tools are running in the wellbore.
Data sets, Main and sub-Main, are categorized as mentioned. Amount of data for each well is different since the pay zone of production is different for each well. Average available data for each well is around 800 cells to 1000 cells. For the entire data sets, average data is around 80000 cells. Each cell indicates a value for each depth in our data sets.

Data sets analyzed with data cleansing methods. Average, mean, minimum, maximum of each well on the pay zone is been calculated and any type of unclear results in terms of average values for each formation is detected and removed from the data sets. Beside these method, each row of data, which represents all values for each depth, is been monitored for all data sets in order to find wrong data that must be removed in order to keep the data driven models out of confusion. In this regard, the output well logs would be more reliable. Once the anomalous data were deleted and cleaned, predictions improved noticeably.

2.2 Data-driven Model Development

In this step, the prepared database was processed using Back propagation Algorithm of Neural Network in two parts. Conventional data-driven models and Geomechanical data-driven models.

Development of the data-driven models is completed using data sets prepared in last step. Different Training, calibration and verification algorithms are attempted until the best results were achieved.

Part 1- Conventional Models:

In order to have the consistent conventional logs information for all wells, the first part is defined and the conventional log data were generated. Some well logs which are Bulk Density and Sonic Porosity were missing for some wells. To complete this missing data which results in complete data sets of conventional logs, data-driven models are trained, calibrated and verified. Outcome of this models are used as the input of part II that is the generation of geomechanical well logs. The bulk density and sonic porosity for 30 wells were produced by using two different data-driven models.

First model (Data-Driven model#1) used Gamma Ray, Depth, Location and well Coordinates as input to develop training, calibration and verification segments for generating the Bulk Density of around 30 wells in asset where bulk density and sonic data were missing. To validate the accuracy of predicted data, blind wells are used to plot both actual and generated one.

In part one also Sonic porosity were missing. Second model (Data-Driven model 2) used also Bulk Density as input beside inputs of first step to generate Sonic Porosity for the part of asset without this property (Around 30 wells also used in this part). Blind wells are used for the validation purposes.
At the end of this step, all wells in the asset have complete set of conventional logs such as Gamma Ray, Bulk Density and Sonic Porosity. In part 2 of the procedure, these data is considered as inputs. All geomechancial logs generated from data-driven models are coming from these conventional logs. Any type of error or wrong data in this data sets will make the task harder in the next step and as such these data are cleaned and have been thoroughly analyzed to make sure all aspects of data mining process is successfully respected.

Part 2- Geomechanical Models:

After having the conventional logs information for all wells, the neural network model was developed by creating five different data-driven models for five geomechanical characteristics which are Total Minimum Horizontal Stress, Poisson’s Ratio, Bulk, Young and Shear Modulus. Backpropagation algorithm is used with several different structures to achieve the best results.

First model is trained, calibrated and verified to generate the Young Modulus. The second and third data driven model is also used the first models output as the input to generate Bulk and Shear modulus respectfully. Once first three modules are generated, then the poison’s Ratio is generated with data driven model number four. In each model, the results of last model are used as the input. Finally model number 5 that is trained and validated for Predicting the Total Minimum Horizontal Stress.

In brief, for each of these neural networks, the same conventional information of 30 wells has been provided and one geomechanical property was generated at a time. Then each generated geomechanical property was used as an input for the next neural network model and the process continued until all five determined geomechanical properties were achieved.

During neural network designing process of these data-driven models, various neural network architecture is been tried and analyzed. Number of hidden layers and the partitioning process are changed through the entire modeling time. The optimized amount of hidden layers and partitioning sections of main data set is used which performed the best outcome. In brief, around 80% of data was used for training, and 20% for calibration and verification (10% for each).

The next two pages is showing the schematic workflow of the body of this study.
Work Flow for Models in part 1:

Model (1.1) Development Procedure:
- Train the model using wells with Gamma Ray and Bulk Density Logs (Output: Bulk Density)
- Validate the model using Blind Wells (Output: Bulk Density)
- Generate Bulk Density for wells that are missing Bulk Density

Model (1.2) Development Procedure:
- Train the model using wells with Gamma Ray, Bulk Density, & Sonic Logs (Output: Sonic)
- Validate the model using Blind Wells (Output: Sonic)
- Generate Sonic Logs for wells that are missing Sonic
Work Flow for models in part 2:

Model (2.1 to 2.5) Development Procedure for 5 Geomechanical Properties:
- Train the model using wells with Conventional Logs and Geomechancial Logs:
  Output: Geomechanical Properties
- Validate the model using Blind Wells:
  Output: Young, Bulk and Shear modulus, Poisson’s Ratio and Total Min. Hor. Stress
- Generate Geomech. logs for wells that are missing Geomechanical properties

To explain further about the process of neural network modeling, this information is helpful. In this study, a multilayer neural networks or multilayer perceptions are used which is mainly mentioned that these networks are most suitable for pattern recognition specially in non-linear problems Neural network have one hidden layers with different number of hidden neurons that are selected based on the number of data record available and the number of input parameters selected in each training process.

The Intelligent partitioning process allows the network to adapt to new data one it is being trained. The first segment, which includes the majority of the data, is used to train the model. In order to prevent the memorizing and overtraining effect in the neural network training process, a second segment of the data is taken for calibration that is blind to the neural network and at each step of training process, the network is tested for this set.
If the updated network given better predictions for the calibration set, it will replace the previous neural network; otherwise, the previous network is selected. Training will be continued once the error of predictions for both the calibration and training data set is satisfactory. This will be achieved only if the calibration and training partitions are showing similar statistical characteristics. Verification partition is the third and last segment used for the process that is kept out of training and calibration process and is used only to test the precision of the neural networks. Once the network is trained and calibrated, then the final model is applied to the verification set. If the results are satisfactory then the neural network is accepted as part of the entire prediction system.

2.3 Validation of Data-driven Models

Five different Data-driven models are trained and the appropriate results are achieved. The question that comes to mind is that how much reliable are these data-driven models. Trained models are used to generate Bulk, Young and Shear modulus as well as Total minimum horizontal Stress, Poisson’s Ratio respectively.

Models Validation and Verification is the most important part of the study since it presents how close the actual data are to data-driven model generated data. One of the simplest ways of showing the accuracy is to generate the data-driven generated logs for those wells that already have the actual well logs. Removing some wells which also called Blind wells from the main data set of Training, Verification and Calibration process and considering them as the output of neural network is a simple but in turn a best way of showing how much reliable are these data-driven models. The models have not seen the blind wells (Data sets) during the Training, Calibration and verification process.

To examine the model validity, the information of some wells (which have both conventional and geomechanical logs) was removed from the training dataset and it was attempted to re-generate the geomechanical logs. These removed wells are so called “blind wells”. Blind wells have been chosen from different location of Marcellus Shale asset. Data-driven models used this new data set to be trained to generate geomechanical properties for those blind wells.

All Data Driven models have been validated separately to generate geomechanical properties for blind wells. In each step, the generated property compared with the actual values that have been removed from main dataset. Validation results have been discussed in the Result and Discussion section of this thesis.
2.4 Error Analysis

In this section, histograms of errors is used to show the percentage error of data-driven models in generating five geomechanical properties of blind wells. Method of Analyzing Errors:

The mean absolute percentage error (MAPE), also known as mean absolute percentage deviation (MAPD), is a measure of accuracy of a method for constructing fitted time series values in statistics, specifically in trend estimation. It usually expresses accuracy as a percentage, and is defined by the formula:

\[ M = \frac{100\%}{n} \sum_{t=1}^{n} \left| \frac{A_t - F_t}{A_t} \right| \]

where At is the actual value and Ft is the forecast value.

The difference between At and Ft is divided by the actual value At again. The absolute value in this calculation is summed for every fitted or forecasted point in time and divided again by the number of fitted points n. multiplying by 100 makes it a percentage error. Error analysis for all three scenarios is analyzed in the next chapter.
2.5 Wells Selection Criteria for NN Training

There is a fact that the more data available in the training process the better result for the output of neural network. From the beginning of this project, for the training process, 25 wells is used. Then the five wells remained is considered to perform the reliability analysis and validation of data-driven models. However, the important issue is that how many wells with geomechanical logs is necessary to do the task (Training, Calibration and Verification of data-driven models) that is generating logs for the entire asset.

At first, Selection of the wells for Training was based on locations of available wells with geomechanical logs in the asset. However, when it comes to a new project and new asset, the question, which comes to picture is: How many wells are needed to be logged with geomechanical instruments? With answering this question, three scenarios explained on the next section. A procedure is designed here to explain how many wells and which wells in an asset is needed to accomplish the training of neural network models.

The criteria is related to the fact that neural network is finding the nonlinear relationship between inputs and outputs. The range of data as input is in our hand to select. Since different wells have different well logs data in terms of average, minimum and maximum, the plots of the average, minimum and maximum of each characteristics of well logs, which in this study is Bulk Density, Gamma Ray and Sonic porosity, is showing the wells needed to be in training. It is clearly shown that wells in the border of these plots and in the middle are candidate wells to put in training data sets.

For this study, the mentioned plots are as follows.
Figure 8. Maximum, Average and Minimum of Sonic porosity vs. Bulk Density
Figure 9. Maximum, Average and Minimum Sonic Porosity vs. Gamma Ray
Figure 10. Maximum, Average and Minimum of Bulk Density vs. Gamma Ray
For the training purposes, it is necessary for the data-driven models to cover the entire range of data. In this regard, the output would be much more acceptable since models are familiar with input data which will correlate them to the output parameters. Data analyst should consider a data set of Minimum data which covers the whole range of inputs.

When you plot the minimum, maximum and average data of each well vs. each other, it is easier to pick wells for the input data sets that are around the boundary of these plots. As you cover the entire range, more wells are added to data sets. It is necessary also to pick some wells with the middle of above plots which will result in better distribution of data in input of the training section.

In this study, it is tried to pick the least amount of wells that cover the whole range, which in turn will result in faster training in terms of time as well as achieving the better results. As the models trained and the blind validating process completed, it might be necessary to add more wells from these plots to the data sets to optimize the number of the wells needed to get the best match in terms of geomechanical property validation with actual data sets. Five red spots in these plots are those blind wells in the first scenario that served as the training data set this time and afterwards three other wells (Green spots) are selected to better cover the range of inputs for the training purposes. The result of all scenarios are discussed in the next chapter of this thesis.

### 2.6 Geomechanical Property Distribution

In this step, the first objective of this paper is accomplished; geo-mechanical properties are generated for all existing wells in the asset. Different geostatistic methods is analyzed to perform geomechanical property distribution.

Methodology for creating geomechanical distributions are as follows. The specific data used here came from well logs from both neural network generated logs and actual well logs. Methodology includes two steps, constructing the Structure Framework and Property Modeling.

1) **Structural Framework**

In order to construct a structural framework, the first step is to Import Well Header LAS-file which includes all your wells name, wells coordinates (x,y) Kelly bushing and TVD (Total Vertical Depth). Now all the wells are in the Model with different locations. Since most of the wells in this study are horizontal wells, trajectory LAS-file is needed to input in the model to show the exact pathway of drilled wells. For all the wells, this trajectory files are created and then imported to the model.
Afterwards, most important part of this process is to import all well logs such as Gamma Ray and all five geomechanical properties (Total Minimum Horizontal Stress, Poisson’s Ratio, Bulk, Young and Shear Modulus). For each well the LAS-files are generated in this step. All the logs are double-checked to remove any possible errors in generating them. Once you import these files to the model, by clicking on the name of each well, you can see the correspondent well logs that are imported.

Creating the cross sections is important in this step. According to the location of the well, five cross sections, which are selected wells in different direction, is selected to cover the entire reservoir. With not selecting appropriate cross sections will result in inappropriate distribution of geomechanical properties. In this step, the possible number of cross sections is unlimited so here several attempts are completed with selecting different cross sections to achieve the best results in terms of making final distributions.

In the well section display, you will see the well logs right next to each other for all the wells in each cross section you have chosen. In this step, well tops, which are tops of formations, should be picked up for each cross section. As it is explained before, the criteria of selecting this is Gamma Ray logs. Here you choose the Gamma Ray to be shown in the display windows. Going through this step, for each specific cross section all different layers of upper and lower Marcellus shale is been selected.

By selecting the well tops, surfaces are created for the asset. In order to make surfaces for upper and lower Marcellus shale, Minimum curvature method is used. Several different algorithms such as 1. Convergent 2. Kriging 3. Minimum Curvature 4. Moving average is tried to make the after mentioned surfaces. All this methods are interpolating the data in the model with different criterias and weights.28

Finally, minimum curvature method is been selected since other methods did not keep the vertical distance constant since the surfaces must not have any contact. Min. Curv. Method showed the best results for all four surfaces In brief, the Minimum Curvature Method smooths two straight-line segments of the Balanced Tangential Method by using a Ratio Factor (RF). Calculates node values near data and then uses minimum curvature method for extrapolating the trends smoothly between and away from the data. The available parameters have little impact on the resulting surface except when Minimum curvature is switched to Full tension in which case extreme spikes occur at the data locations. This method is most used algorithm in constructing surfaces in structural modeling.

2) Property Modeling
When the surfaces are created, average values of geomechanical properties such as Total Minimum Horizontal stress, Poisson’s Ratio, bulk, young and Shear Modulus are added to the model for both upper and lower Marcellus shale formations separately. The averages data is gathered one by one form the main data sets that the data-driven models are trained with. For each well in the data set, all values for both layers were averaged and gathered in a spreadsheet. Then, all values are imported to their correspondent well in the model.

A well log is created for each property for all wells in the model, with the input data, before up scaling the properties to the scale of each grid block. After up-scaling all the geomechanical properties, the petro-physical modeling which is considered the last part of geomechanical modeling is completed.

At the end, Sequential Gaussian simulation (SGS) used to create distribution according to well locations for the entire field. Two types of maps are created. First map is only incorporated with 30 wells, which already have actual geomechanical logs. The second map is related to entire field (50 wells with generated property and 30 wells with actual data). These two maps demonstrate differences between geomechanical property distribution with and without having full-field data. Ten maps which shows distribution of five rock geomechancial properties in the Marcellus shale asset is performed and showed in appendix. These distributions have been discussed later in this paper.

Basically Geostatistics is intimately related to interpolation methods, but extends far beyond simple interpolation problems. Geostatistical techniques rely on statistical model that is based on random function (or random variable) theory to model the uncertainty associated with spatial estimation and simulation.
Chapter 3

Results and Discussion

The Marcellus Shale filed under study is a part of an asset that is located in southern Pennsylvania and northern West Virginia. In the following plots all the wells in the asset is shown. To be clearer, in the first map you see all the 80 wells in the asset. Half of them are vertical and the rest are horizontal wells. As it goes ahead, wells, which are used in training that wells with actual geomechanical logs are, are shown separately with different colors. Also five wells out of 30 wells, which served as blind wells, are shown differently.

Figure 11. Marcellus Shale Gas Filed Under Study-80 wells
Figure 12. 50 Wells with Conventional Logs in Marcellus Shale asset

Figure 13. 30 Wells with Geomechanical Logs in Marcellus Shale asset
3.1 KPI analysis:

Once models are developed, Key Performance Indicator process (KPI) is performed in order to investigate the influence of each parameter in data-driven models. This process is analyzed for each step during model developments. KPI of both conventional and geomechanical data-driven models are measured and plotted.

These plots show the importance of inputs in managing and gaining the outputs by order. It also demonstrates the models capability of keeping in order all the inputs in each step. Analyzing the KPI is not a simple task since the range of data is really wide and at the time that there are more input parameters, it is harder to find the most influential parameter and rank it by their impact in output of data-driven models. Figures 14 to 20 illustrate the ranking of different inputs when Total Minimum Horizon Stress is used as an output.

For the figure 14, it is showing that in calculating the bulk Density, the Gamma ray has the most influence and most important parameter. Clustering which is general facial separation of data has the second rank in the model. Depth and location has around 20 and 5 percent impact in Bulk density generation.

![Bar Chart](Figure 14. KPI Analysis- Output is Bulk Density)
In order to rank the impact of parameters in Input data set on the Sonic porosity prediction, the KPI here is showing that Bulk Density has the most important role. Which Gamma Ray has the second rank with 80 percent influence on Sonic Porosity prediction by data-driven models. Data clustering has around 60 percent influence in Sonic Porosity prediction and so on.

As you follow, the data from last step is added to the next model. For the Young modulus as the first parameter of geomechanical properties, the Sonic Porosity has the most influence which conceptually it is acceptable since all of these parameters can be calculated with the Delta t Shear and Compression which is explained in last chapter.(In this study we do not have acces to this data). Buk Density and Gamma Ray has the same amount of impact in predicting the yound Modulus. The rest is shown also in the Figure 16.
Bulk Density is calculated with the inputs showed in the figure below. It is obvious that Young Modulus has the most influence since these parameter are conceptually related to the forces applied in different directions to the rock. Sonic porosity has the second rank since the times that waves passed through the rock and come back is the way that they are calculatable. Gamma Ray and Clustering has less influence and are at the last levels.

Figure 17. KPI Analysis- Output is Bulk Modulus

Once it comes to Shear modulus the trend of KPI is the same and most influences are coming from the last two parameters predicted. All other parameters are showing less impact on the output.

Figure 18. KPI Analysis- Output is Shear Modulus
The trend interestingly is followed for Poisson’s Ratio as well. Last geomechanical properties generated have shown most part of influential ranking of these characteristics as well. Even in this step, Gamma Ray and Sonic and the others are showing really less impact which is basically acceptable and understandable.

For the final data driven model, in generating the Total Minimum Horizontal Stress, all the geomechanical properties mostly have the same influence. Since they are all related to speed of waves going through the rock and the time corresponded, the results are acceptable and even the Gamma Ray and the rest has around 30 and less impact on this parameters to be predicted.
Additionally, R-squared calculated for training, calibration and verification segments, is also an indicator of accuracy of the data-driven models in performing results as output. The higher the R-squared, the closest the results to the actual values.

R-squared is calculated with the formula below:

\[
\text{R-squared} = 1 - \frac{\text{Sum of squared distances between the actual and predicted values}}{\text{Sum of squared distances between the actual and their mean values}}
\]

In our study the highest achieved R-squared is around 95 and the lower one in some cases around 85 percent, which in both situation the results presented are highly acceptable but the critical point is when the model gained very low R-squared, the results were so poor in generating a matched data. Higher level of R-squared reflects, in all three stages of training, calibration and verification, the reliable correlation between actual and generated data.

It is also important to mention that during the initial training of data sets; the results obtained were with low R-squared. Unsuccessful behavior of models was understood because of having some wells with log data for each 0.5 ft., which is in contrast with the rest of the wells with every 1 ft. log data available. Once data of 0.5 ft. turned to 1 ft., which considered as discrepancy that misleads the models, the results came out properly and the data-driven models showed rapid improvements.

In appendix, we present actual well logs and generated logs for 5 blind wells. To compare the results, both actual and generated properties are plotted in the same figure like an actual well log. Properties such as Bulk Modulus, Young Modulus, Poisson’s Ratio, Shear Modulus and Total Minimum Horizontal Stress are presented respectively.

Blue line shows the actual value and the red line is for generated values by data-driven models. For well # 1 to well # 4, there is perfect match between blue and red lines. These wells are in proximity of wells with actual geomechanical properties according to their locations and depths. As it was expected, results showed for these wells are accurate which demonstrate data-driven models capability in predicting geomechanical properties.
3.2 Scenario # 1- Training with 25 wells and output of 5 Blind wells

The study focuses on part of Marcellus Shale field. Figure 21 shows the distribution of existing wells in the Marcellus Shale asset that is used in this scenario. Table 4 shows the information and number of the wells that were used to develop data-driven models in this scenario.

![Figure 21.Scenario # 1 well distributions](image)

Table 3. Information used for data-sets for developing Data-driven models

<table>
<thead>
<tr>
<th>Well Identifier</th>
<th>Description</th>
<th>Conventional Well Logs</th>
<th>Geomechanical Well logs</th>
<th>Number of wells</th>
</tr>
</thead>
<tbody>
<tr>
<td>Green Circle</td>
<td>Blind Validation Wells</td>
<td>YES</td>
<td>YES</td>
<td>5</td>
</tr>
<tr>
<td>Red Square</td>
<td>Wells Used for Training, Calibration &amp; Validation</td>
<td>YES</td>
<td>YES</td>
<td>25</td>
</tr>
<tr>
<td>Blue Diamond</td>
<td>Wells with no Geo-mechanical logs</td>
<td>YES</td>
<td>NO</td>
<td>50</td>
</tr>
</tbody>
</table>
Figure 22. Well #1 in Scenario 1 - Actual Vs. NN generated well logs
Figure 23. Well #2 in Scenario 1 - Actual Vs. NN generated well logs
Figure 24. Well # 3 in Scenario 1- Actual Vs. NN generated well logs
Figure 25. Well #4 in Scenario 1- Actual Vs. NN generated well logs
For well #5, the generated data does not match with the actual well logs and it might be related to the fact that the data associated to this well is not perfect. This fact indicates that the models could not be
able to predict the behavior of outlier wells. Further study related to this well is done in the following part of this study.

Well number 5 is not showing acceptable results at all since the generated data is not matched with the actual one. Since the data-driven models are showing perfect result for most of the blind wells, the problem is not possibly coming from the design issues of neural network structure. To find an answer for this well, all the data sets are evaluated and analyzed with different methods. Finally, the problem related to this well in having very different behavior with was successfully found.

Since the neural network models are using the data sets provided from the beginning, there should be something wrong with them. In this regard, all the input and output well logs of all 80 wells in the asset is plotted to see a trend that results a different behavior for well number five. After doing this task, it is understood that the sonic well log of this well which is the one of the main inputs which affects the outputs, due to KPI analysis, has different trend compared to all other wells in the asset.

Sonic log of this well is look like to be wrong which results in confusion of data driven models, all five of them. To solve this issue, a creative way came to mind. Generating Sonic log for this specific well with other available logs such as Bulk Density and Gamma Ray. Another data-driven model is trained, calibrated and verified with the same architecture with optimum output. By doing so, a synthetic sonic well log is generated for well number five. Having the true values for this well, this time we used this information in the main data sets to run one last model to generate 5 geomechanical logs for this specific well. After all, the outcome of these models was fantastic which the results are in the next pages.

Figures 27, 28 and 29 are showing the Input data of our model such as Gamma Ray, Bulk Density, Sonic Porosity and five geomechanical properties of all wells versus Well # 5.
Figure 27. Bulk Density and Gamma Ray for all wells Vs. Well # 5
Figure 28. Bulk Modulus, Poison’s Ratio and Shear modulus of all wells Vs. Well # 5
Figure 29. Young Modulus, TMHS and Sonic porosity of all wells Vs. Well # 5
Figure 30. Data-driven generated Sonic Porosity Vs. Actual Log (Wrong log) for Well # 5
Figure 31. Well # 5 in Scenario 1 - Actual Vs. NN generated well logs
3.2.1 Error analysis of Scenario# 1

These error histograms are showing that for the first scenario. These models are trained with 25 wells. five blind wells are validated and the histograms are plotted. For four wells in the blind set, the error in all geomechanical characteristics generation with less than 5 percent. The only well that is showing errors around 10 percent and more is well # 5, which is explained, in next chapter.

![Young Modulus-Histogram](image1)

Figure 32. Young Modulus Error histogram for scenario #1

![Shear Modulus-Histogram](image2)

Figure 33. Shear Modulus Error histogram for scenario #1
Figure 34. Bulk Modulus Error histogram for scenario #1

Figure 35. Poisson’s Ratio Error histogram for scenario #1
Figure 36. Total min. Hor. Stress Error histogram for scenario #1
3.3 Scenario # 2- Training with 5 wells and output of 25 Blind wells

The results which shows the blind set of 25 wells is in the appendix B. Table below is showing the number of wells used in this scenario.

![Figure 37. Scenario # 2 well distributions](image)

<table>
<thead>
<tr>
<th>Well Identifier</th>
<th>Description</th>
<th>Conventional Well Logs</th>
<th>Geomechanical Well Logs</th>
<th>Number of wells</th>
</tr>
</thead>
<tbody>
<tr>
<td>Green Circle</td>
<td>Wells Used for Training, Calibration &amp; Validation</td>
<td>YES</td>
<td>YES</td>
<td>5</td>
</tr>
<tr>
<td>Red Square</td>
<td>Blind Validation Wells</td>
<td>YES</td>
<td>YES</td>
<td>25</td>
</tr>
<tr>
<td>Blue Diamond</td>
<td>Wells with no Geo-mechanical logs</td>
<td>YES</td>
<td>NO</td>
<td>50</td>
</tr>
</tbody>
</table>
3.3.1 Error analysis of Scenario #2

These models are trained with five wells which were blind set in scenario #1. 25 other wells which scenario #1 was trained with herein served as blind wells. Blind wells are validated and the histograms are plotted. Most of the wells (around 120 wells) have less than 10 percent error. One well has the error around 20 percent and the other 4 wells are showing 15 percent error.

Figure 38. Young Modulus Error histogram for scenario #2

Figure 39. Bulk Modulus Error histogram for scenario #2
Figure 40. Shear Modulus Error histogram for scenario #2

Figure 41. Total min. Hor. Stress Error histogram for scenario #2
Figure 42. Poisson’s Ratio Error histogram for scenario #2
### 3.4 Scenario # 3- Training with 8 wells and output of 22 Blind wells

The results which shows the blind set of 22 wells is in the appendix C. Table below is showing the number of wells used in this scenario.

<table>
<thead>
<tr>
<th>Well Identifier</th>
<th>Description</th>
<th>Conventional Well Logs</th>
<th>Geomechanical Well logs</th>
<th>Number of wells</th>
</tr>
</thead>
<tbody>
<tr>
<td>Green Circle</td>
<td>Wells Used for Training, Calibration &amp; Validation</td>
<td>YES</td>
<td>YES</td>
<td>8</td>
</tr>
<tr>
<td>Red Square</td>
<td>Blind Validation Wells</td>
<td>YES</td>
<td>YES</td>
<td>22</td>
</tr>
<tr>
<td>Blue Diamond</td>
<td>Wells with no Geo-mechanical logs</td>
<td>YES</td>
<td>NO</td>
<td>50</td>
</tr>
</tbody>
</table>
3.4.1 Error analysis of Scenario# 3

This is interesting to know that with 8 wells, the errors are less but the trends of their reduction is as same as last scenarios. These error histograms are showing Percent of Errors for the third scenario. These models are trained with eight wells. 22 blind wells are validated and the histograms are plotted. Most of the wells (around 18 wells) have less than 5 percent error. The rest of the wells have errors around 10 to 15 percent.

![Young Modulus Error histogram for scenario #3](image)

Figure 43. Young Modulus Error histogram for scenario #3
Figure 44. Bulk Modulus Error histogram for scenario #3

Figure 45. Shear Modulus Error histogram for scenario #3
Figure 46. Poisson's Ratio Error histogram for scenario #3

Figure 47. Total min. Hor. Stress Error histogram for scenario #3
Chapter 4

Concluding Remarks and Future Work

In this thesis, it is demonstrated that the AI&DM is a reliable tool in performing accurate results for generating synthetic geomechanical logs. In simple terms, we used conventional well logs to generate geomechanical properties and create distribution maps of geomechanical properties for the entire asset.

Five data-driven models have been designed to predict five geomechanical properties of Marcellus Shale rock in three different scenarios. All three scenarios have been completed successfully. Validation process conducted to illustrate data-driven models accuracy in predicting Young Modulus, Poisson Ratio, Bulk Modulus, Shear Modulus and Total Minimum Horizontal Stress. Data mining issues in some cases related to geomechanical properties are successfully managed to reach to a reliable prediction. It should be noted that before building the data-driven models, it is necessary to ensure the input/output data’s favorable quality. Errors related to the prediction of all three scenarios are analyzed which mostly are less than 10 percent in scenarios # 2 and less than 5 percent in scenario # 3.

Geomechanical property distribution model of the entire asset shows more reasonable between distributions when there are just a few available actual data rather than having access to the full-field data. A comparison between maps for each property shows that distribution that is more reasonable gained using more data for the asset. The Sequential Gaussian Simulation (SGS) algorithm was used in order to generate the maps. These distributions will help in understanding and determination of the hydraulic fracture growth in the reservoir rock.

Synthetic geomechanical logs and property distributions for Marcellus shale exhibits a great deal of assistant to better performing reservoir modeling and the optimizing of hydraulic fracturing issues related to Marcellus Shale development strategies. Authors expect these models will conclude also accurate results in other shale assets.
For the future work, there are some ideas that are interesting which author is planning to perform them as the future steps of his research.

1. Developing a workflow with this technology for Quality Control purpose of the well logs in the asset.
2. Finding the optimum number of wells needed to be logged with geomechanical log tools
3. Using different well logs such as Resistivity, Density Porosity etc. to generate synthetic geomechanical logs.
4. Finding the relationship of ISIP (Initial Shut-in pressure) and geomechanical properties to find a dynamic way of showing the change of these properties after each step of hydraulic fracturing.
5. Correlating the cumulative production to dynamic geomechanical properties model
6. Developing a robust method with AI&DM technology to edit and refine the main data sets (well logs) before starting data-driven modeling.
7. Using different neural network algorithms to compare the results and analyzing them.

Appendix A showing distribution models for five geomechanical rock properties. For each property, there are two distributions; one by using the actual data and the second one by using the information of both generated and actual data (Full-field data).

Appendix B and C are showing the well logs served as a blind wells in validation of scenarios 2 and 3.
Appendix A

Geomechanical properties for the asset under studies. Top distribution is with limited number of data and the bottom distribution is related to full-field data after generating them by data-driven models.

Figure 48. Distribution of Young Modulus based on actual data (Top) and Distribution with full-field data. (Bottom)
Figure 49. Distribution of Shear Modulus based on actual data (Top) and Distribution with full-field data. (Bottom)
Figure 50. Distribution of Poisson’s Ratio based on actual data (Top) and Distribution with full-field data. (Bottom)
Figure 51. Distribution of Total Min. Hor. Stress based on actual data (Top) and Distribution with full-field data. (Bottom)
Figure 52. Distribution of Bulk Modulus based on actual data (Top) and Distribution with full-field data. (Bottom)
Appendix B:

Results for Scenario #2: Blue Line is Actual Well Logs and the Red line is Neural Network Generated well log.

Figure 53. Actual Vs. NN generated Geomechanical logs for Wells # 1 and 2
Figure 54. Actual Vs. NN generated Geomechanical logs for Wells # 3 and 4
Figure 55. Actual Vs. NN generated Geomechanical logs for Wells # 5 and 6
Figure 56. Actual Vs. NN generated Geomechanical logs for Wells # 7 and 8
Figure 57. Actual Vs. NN generated Geomechanical logs for Wells # 9 and 10
Figure 58. Actual Vs. NN generated Geomechanical logs for Wells # 11 and 12
Figure 59. Actual Vs. NN generated Geomechanical logs for Wells # 13 and 14
Figure 60. Actual Vs. NN generated Geomechanical logs for Wells # 15 and 16
Figure 61. Actual Vs. NN generated Geomechanical logs for Wells # 17 and 18
Figure 62. Actual Vs. NN generated Geomechanical logs for Wells # 19 and 20
Figure 63. Actual Vs. NN generated Geomechanical logs for Wells # 21 and 22
Figure 64. Actual Vs. NN generated Geomechanical logs for Wells # 23 and 24
Figure 65. Actual Vs. NN generated Geomechanical logs for Well # 25
Appendix C:
Results for Scenario #3: Blue Line is Actual Well Logs and the Red line is Neural Network Generated well log.
Figure 67. Actual Vs. NN generated Geomechanical logs for Wells # 3 and 4
Figure 68. Actual Vs. NN generated Geomechanical logs for Wells # 5 and 6
Figure 69. Actual Vs. NN generated Geomechanical logs for Wells # 7 and 8
Figure 70. Actual Vs. NN generated Geomechanical logs for Wells # 11 and 12
Figure 71. Actual Vs. NN generated Geomechanical logs for Wells # 13 and 14
Figure 72. Actual Vs. NN generated Geomechanical logs for Wells # 15 and 16
Figure 73. Actual Vs. NN generated Geomechanical logs for Wells # 17 and 18
Figure 74. Actual Vs. NN generated Geomechanical logs for Wells # 19 and 20
Figure 75. Actual Vs. NN generated Geomechanical logs for Wells # 21 and 22
Figure 76. Actual Vs. NN generated Geomechanical logs for Wells # 23 and 24
References

2. Fundamental of Well-log Interpretation, O Serra, Elsevier Series, Developments of Petroleum Science 15 A
3. Fundamental of Well-log Interpretation, O Serra, Elsevier Series, Developments of Petroleum Science 15 A

17. VIRTUAL INTELLIGENCE AND ITS APPLICATIONS IN PETROLEUM ENGINEERING, 1. Artificial Neural Networks, Shahab Mohaghegh Intelligent Solutions, Inc. West Virginia University, JPT - Journal of Petroleum Technology, September 2000
23. An Effective Hash-based Method for Generating Synthetic Well Log, Yi Du , Wen-an Tan', Chuanqun Jiang', Detang Lu2, Daolun Li2, 'Department ofComputer and Information, Shanghai Second Polytechnic University, Shanghai, P.R. China, 2 University ofScience and Technology ofChina, Hefei, Anhui, P. R. China

111


ITASCA HOUSTON, INC.: EXPERTISE, GEOMECHANICAL MODEL, The necessary starting point for a quality geomechanical evaluation
