

2023

Comparative Analysis of Artificial Intelligence and Numerical Reservoir Simulation in Marcellus Shale Wells

Arya Maher Sattari

West Virginia University, asattari@mix.wvu.edu

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

 Part of the [Petroleum Engineering Commons](#)

Recommended Citation

Sattari, Arya Maher, "Comparative Analysis of Artificial Intelligence and Numerical Reservoir Simulation in Marcellus Shale Wells" (2023). *Graduate Theses, Dissertations, and Problem Reports*. 12256.
<https://researchrepository.wvu.edu/etd/12256>

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

Comparative Analysis of Artificial Intelligence and Numerical Reservoir Simulation in Marcellus Shale Wells

Arya M. Sattari

Dissertation 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

Doctor of Philosophy in
Petroleum and Natural Gas Engineering

Shahab D. Mohaghegh, Ph.D., Chair
Samuel Ameri, M.S., P.E.
Kashy Aminian, Ph.D.
Mohamed El Sgher, Ph.D.
Qingqing Huang, Ph.D.

Department of Petroleum and Natural Gas Engineering

Morgantown, West Virginia
2023

Keywords: Petroleum Data Analytics, Numerical Reservoir Simulation, Artificial Intelligence, Machine Learning, Data Science, Unconventional Reservoir, Production Optimization, Marcellus Shale, Shale Gas

©Copyright 2023 Arya M. Sattari

ABSTRACT

Comparative Analysis of Artificial Intelligence and Numerical Reservoir Simulation in Marcellus Shale Wells

Arya M. Sattari

This dissertation addresses the limitations of conventional numerical reservoir simulation techniques in the context of unconventional shale plays and proposes the use of data-driven artificial intelligence (AI) models as a promising alternative. Traditional methods, while providing valuable insights, often rely on simplifying assumptions and are constrained by time, resources, and data quality. The research leverages AI models to handle the complexities of shale behavior more effectively, facilitating accurate predictions and optimizations with less resource expenditure.

Two specific methodologies are investigated for this purpose: traditional numerical reservoir simulations using Computer Modelling Group's GEM reservoir simulation software, and an AI-based Shale Analytics approach using IMPROVE™ software from Intelligent Solutions, Inc. The investigation covers the impact of key parameters on production prediction, assumptions made, predictive accuracy, data requirements, workflow complexity, and time efficiency.

By comparing these methods, the research aims to offer guidelines for incorporating AI models into reservoir simulation and identify areas for increased efficiency and accuracy. The study concludes by presenting recommendations to advance the field of reservoir simulation and encourage the adoption of innovative methodologies in the energy industry. The results are anticipated to considerably enhance reservoir simulation processes and optimize production strategies for unconventional shale plays.

Acknowledgement

I would like to express my deepest gratitude to my supervisor, Dr. Shahab Mohaghegh, for his unwavering support and invaluable guidance throughout this journey. His expertise, patience, and motivation have been instrumental in my academic pursuits, and I will forever be grateful to have been a part of the LEADS team.

My sincere thanks also go to the members of my dissertation committee: Dr. Kashy Aminian, for his unwavering support through every challenge; Department Chair, Professor Samuel Ameri, for his constant encouragement and unending support, Dr. Mohamed El Sgher for his patience, especially when learning new software and model development, and Dr. Qingqing Huang, for her invaluable feedback and insight. Your collective support has been instrumental in my success, and I deeply appreciate the time and effort you have all invested in me.

I am also grateful to my colleagues and friends in the PNGE department and LEADS team for fostering an environment of enthusiasm and innovation. The time we've spent together, and the conversations we've shared hold a special place in my heart. I look forward to seeing where our futures lead, the successes you will all have, and how our friendships will grow as we pursue our individual goals.

Heartfelt thanks go to my friends and family, who have supported me and patiently listened to my academic musings, often enduring one-sided conversations.

To my parents, your unconditional love, support, and encouragement have been my cornerstone; I could not have achieved this without you.

And to R, you were there when this all began, and your memory continues to inspire me, I miss you.

This dissertation would not have been possible without the support and encouragement from each one of you. Thank you all for being an integral part of my academic journey.

Table of Contents

List of Tables	vi
List of Figures	vii
Chapter 1: Introduction	1
1.1 Problem Statement.....	2
1.2 Research Objective	2
Chapter 2: Literature Review	5
2.1 Overview	5
2.2 Numerical Reservoir Simulation	6
2.2.1 Numerical Reservoir Simulation: Unconventional Shale Reservoirs	8
2.2.2 Applications of Numerical reservoir Simulation in Unconventional Shale Reservoirs	13
2.3 Data-Driven AI models and Shale Analytics	16
Chapter 3: Objectives and Methodology	18
3.1 Objectives and Methodology Summary.....	18
3.2 Methodology: Numerical Reservoir Simulation (NRS).....	18
3.2.1 Data Collection: MSEEL Site and MIP-3H Well.....	19
3.2.2 Data Analysis	24
3.2.3 Model Development	26
3.2.4 Model Performance and Validation	32
3.3 Methodology: Data-Driven Artificial Intelligence Model.....	34
3.3.1 Data Collection and Preparation: AI Model Study Area.....	35
3.3.2 Data Exploration and Analysis: Marcellus Shale Well Dataset	38
3.3.2.1 Basic Statistical Analysis.....	39
3.3.2.2 Well Quality Analysis (WQA).....	43
3.3.2.3 Key Performance Indicators (KPI)	46
3.3.3 Predictive Analytics and AI Model Development.....	50
3.3.3.1 Feature Selection	51
3.3.3.2 Data Partitioning and Database Construction	52
3.3.3.3 Model Development: Predictive AI Model for 30 Day Rich Gas production	54
3.3.3.4 Model Validation and Predictive Analysis.....	58
Chapter 4: Results and Discussion	60
4.1 Numerical Reservoir Simulation Model Results.....	60
4.1.1 Scenario 1: Base Model.....	61

4.1.2 Scenario 2: Number of (Producing) Hydraulic Fracture Perforations & Fracture Half-Length ..	63
4.1.3 Scenario 3: Matrix Porosity and Fracture Conductivity	65
4.1.4 Scenario 4: Skin and Skin Effect	67
4.1.5 Summary of NRS Scenario Results	69
4.2 AI Data Driven Model results	70
4.2.1 Data Driven AI Model: Initial Model Validation	72
4.2.2 Data Driven AI Model: Tuned Model Validation & Discussion	73
4.2.3 AI Model: Blind Validation and Model Tuning Discussion	74
4.2.4 AI Reservoir modelling – Top-Down Model (TDM) Results	76
Chapter 6: Conclusion and Recommendations.....	79
6.1 Conclusion.....	79
6.2 Recommendations	80
References	81
Appendix	84
AI Dataset: Input Attribute Behavior (Fuzzy Pattern Recognition).....	84

List of Tables

Table 1: "Hard Data" and "Soft Data" in Hydraulic Fracturing (Mohaghegh, 2017).....	12
Table 2: Basic Model Parameters for MIP-6H and MIP-4H.....	14
Table 3: Completion and stimulation parameters for MIP-3H (Carr et al., 2015)	22
Table 4: NRS MIP-3H Base Model Parameters.....	27
Table 5: NRS MIP-3H Parametric Study: Number of (Producing) Hydraulic Fracture Perforations & Fracture Half Length	28
Table 6: NRS MIP-3H Parametric Study: Matrix Porosity and Fracture Conductivity.....	29
Table 7: NRS MIP-3H Parametric Study: Skin Effect	31
Table 8: Input Attribute Summary for AI Model Development	36
Table 9: Output Attribute Summary for AI Model Development	37
Table 10: Key Performance Indicators (Attributes Ranking and Degree of Influence).....	47
Table 11: NRS MIP-3H Base Model Parameters	61
Table 12: NRS MIP-3H Base Model Results.....	62
Table 13: NRS MIP-3H Parametric Study Scenario: Number of (Producing) Hydraulic Fracture Perforations & Fracture Half Length.....	63
Table 14: NRS MIP-3H Parametric Study Scenario Results: Number of (Producing) Hydraulic Fracture Perforations & Fracture Half Length.....	64
Table 15: NRS MIP-3H Parametric Study Scenario: Matrix Porosity and Fracture Conductivity	65
Table 16: NRS MIP-3H Parametric Study Scenario Results: Matrix Porosity and Fracture Conductivity ...	66
Table 17: NRS MIP-3H Parametric Study Scenario: Skin Effect.....	67
Table 18: NRS MIP-3H Parametric Study Scenario Results: Skin Effect	68
Table 19: Results Summary for NRS Scenarios	69
Table 20: Data Driven AI Model Preparation.....	70
Table 21: Backpropagation NN Preparation	70
Table 22: Data Driven AI Model: Initial Model Results	72
Table 23: Data Driven AI Model: Initial Model Blind Validation Results.....	72
Table 24: Data Driven AI Model: Tuned Model Results.....	73
Table 25: Data Driven AI Model: Tuned Model Blind Validation Results	73

List of Figures

Figure 1: Worldwide Shale Gas Resources (EIA, 2015)	5
Figure 2: "Sugar Cube Model" of Naturally Fractured Reservoir (Warren & Root, 1963)	8
Figure 3: Key elements needed for successful shale gas play (Kundert & Mullen, 2009)	9
Figure 4: "Real" and "Imaginary" sides in Simulation Modelling (Islam et al., 2016)	11
Figure 5: History Matching for MIP-6H with geo-mechanical impact for first two years (El Sgher, 2021)	14
Figure 6: Process Workflow - Numerical Reservoir Simulation Methodology.....	19
Figure 7: Well layout in the Marcellus Shale Energy and Environment Laboratory (MSEEL) MIP site	20
Figure 8: Logs from the Vertical Section of the MIP-3H Well (El Sgher, 2021)	23
Figure 9: Production History MIP-3H – Cumulative Gas (2015.12.12 - 2021.07.08)	23
Figure 10: Reservoir Parameter Analysis	24
Figure 11: Parametric Study Example – CMG™ CMOST™ Simulations.....	26
Figure 12: Data Driven AI Model Workflow	34
Figure 13: Data Collection and Preparation.....	35
Figure 14: Normalized Marcellus Shale Well Locations	38
Figure 15: Basic Statistical Analysis - Bubble Plot (30-Day Rich Gas MCF Production).....	40
Figure 16: Regression Analysis - Measured Depth vs 30-day Rich Gas Production	41
Figure 17: Regression Analysis - Stimulated Lateral Length vs 30-day Rich Gas Production.....	41
Figure 18: Regression Analysis - Max Proppant Concentration vs 30-day Rich Gas Production	42
Figure 19: Regression Analysis - Total Number of Stages vs 30-day Rich Gas Production	42
Figure 20: WQA Preparation - 3 Fuzzy Sets	43
Figure 21: WQA Preparation - 4 Fuzzy Sets	44
Figure 22: WQA - 3 Fuzzy Sets for PD-A1-H1 Well	44
Figure 23: WQA - 4 Fuzzy sets for PD-A1-H1 Well	45
Figure 24: WQA - Fuzzy Categories for All Wells (X, Y Plane)	46
Figure 25: Fuzzy Pattern Recognition (FPR) - Measured Depth (MD).....	49
Figure 26: Fuzzy Pattern Recognition (FPR) - Initial Water Saturation (Swi)	49
Figure 27: Neural Network Input Selection – IMPROVE™ Software.....	51
Figure 28: Input Selection for AI Model Development	52
Figure 29: Dataset Partitioning	53
Figure 30: Data Partitioning – IMPROVE™ Software	54
Figure 31: Neural Network Design – IMPROVE™ Software	56
Figure 32: Backpropagation Neural Network Training Module – IMPROVE™ Software	57
Figure 33: Model Validation Workflow	58
Figure 34: NRS MIP-3H Base Model Predicted Production Behavior and History Match.....	62
Figure 35: NRS MIP-3H Parametric Study Scenario: Number of (Producing) Hydraulic Fracture Perforations & Fracture Half Length – Predicted Production Behavior and History Match.....	64
Figure 36: NRS MIP-3H Parametric Study Scenario: Matrix Porosity and Fracture Conductivity – Predicted Production Behavior and History Match	66
Figure 37: NRS MIP-3H Parametric Study: Skin Effect – Predicted Production Behavior and History Match	68
Figure 38: Model Selection Workflow	71
Figure 39: Data Driven AI Model: Initial Model Predicted Blind Validation Results.....	72
Figure 40: Data Driven AI Model: Tuned Model Predicted Blind Validation Results.....	73

Figure 41: Data Driven AI Model: Initial Model Partition Visualization	74
Figure 42: Data Driven AI Model: Tuned Model Partition Visualization.....	75
Figure 43: TDM Model: Entire Field Gas Production Results.....	76
Figure 44: NRS Model Scenarios: Forecasted Production	77
Figure 45: TDM Model: Forecasted Production (Example Well 1)	78
Figure 46: TDM Model: Forecasted Production (Example Well 2)	78
Figure 47: Parameter Influence (Fuzzy Pattern Recognition) - Cluster Spacing	84
Figure 48: Parameter Influence (Fuzzy Pattern Recognition) - Completions Date	85
Figure 49: Parameter Influence (Fuzzy Pattern Recognition) - Shot Density.....	86
Figure 50: Parameter Influence (Fuzzy Pattern Recognition) - Stimulated Lateral Length	87
Figure 51: Parameter Influence (Fuzzy Pattern Recognition) - Total Clusters.....	88
Figure 52: Parameter Influence (Fuzzy Pattern Recognition) - Total No. Stages.....	89
Figure 53: Parameter Influence (Fuzzy Pattern Recognition) - Average ISIP.....	90
Figure 54: Parameter Influence (Fuzzy Pattern Recognition) - Average Fracture Gradient	91
Figure 55: Parameter Influence (Fuzzy Pattern Recognition) - BTU Area.....	92
Figure 56: Parameter Influence (Fuzzy Pattern Recognition) - Bulk Modulus.....	93
Figure 57: Parameter Influence (Fuzzy Pattern Recognition) - Minimum Horizontal Stress.....	94
Figure 58: Parameter Influence (Fuzzy Pattern Recognition) - Net Thickness	95
Figure 59: Parameter Influence (Fuzzy Pattern Recognition) - Poisson's Ratio.....	96
Figure 60: Parameter Influence (Fuzzy Pattern Recognition) - Porosity.....	97
Figure 61: Parameter Influence (Fuzzy Pattern Recognition) - Shear Modulus.....	98
Figure 62: Parameter Influence (Fuzzy Pattern Recognition) - Initial Water Saturation (Swi %)	99
Figure 63: Parameter Influence (Fuzzy Pattern Recognition) - Total Organic Content (TOC %).....	100
Figure 64: Parameter Influence (Fuzzy Pattern Recognition) - Young's Modulus	101
Figure 65: Parameter Influence (Fuzzy Pattern Recognition) - Average Breakdown Pressure	102
Figure 66: Parameter Influence (Fuzzy Pattern Recognition) - Average Breakdown Rate	103
Figure 67: Parameter Influence (Fuzzy Pattern Recognition) - Average Injection Pressure.....	104
Figure 68: Parameter Influence (Fuzzy Pattern Recognition) - Average Injection Rate	105
Figure 69: Parameter Influence (Fuzzy Pattern Recognition) - Average Max. Pressure.....	106
Figure 70: Parameter Influence (Fuzzy Pattern Recognition) - Average Max. Rate	107
Figure 71: Parameter Influence (Fuzzy Pattern Recognition) - Clean Volume	108
Figure 72: Parameter Influence (Fuzzy Pattern Recognition) - Max. Proppant Concentration.....	109
Figure 73: Parameter Influence (Fuzzy Pattern Recognition) - Proppant per Stage.....	110
Figure 74: Parameter Influence (Fuzzy Pattern Recognition) - Slurry Volume	111
Figure 75: Parameter Influence (Fuzzy Pattern Recognition) - Total Proppant Injected.....	112
Figure 76: Parameter Influence (Fuzzy Pattern Recognition) - Start Production Date.....	113
Figure 77: Parameter Influence (Fuzzy Pattern Recognition) - Wellhead Pressure WPH (30 Days).....	114
Figure 78: Parameter Influence (Fuzzy Pattern Recognition) - Azimuth	115
Figure 79: Parameter Influence (Fuzzy Pattern Recognition) - Deviation Type.....	116
Figure 80: Parameter Influence (Fuzzy Pattern Recognition) - Inclination.....	117
Figure 81: Parameter Influence (Fuzzy Pattern Recognition) - Measured Depth.....	118
Figure 82: Parameter Influence (Fuzzy Pattern Recognition) - True Vertical Depth TVD.....	119
Figure 83: Parameter Influence (Fuzzy Pattern Recognition) - Well X Location Coordinates	120
Figure 84: Parameter Influence (Fuzzy Pattern Recognition) - Well Y Location Coordinates.....	121

Chapter 1: Introduction

The shale gas revolution, beginning in the early 2000s, fundamentally transformed the global energy landscape by enabling profitable extraction of oil and natural gas from unconventional shale reservoirs. This success was made possible due to key technological advancements including hydraulic fracturing and horizontal drilling techniques, which together revolutionized extraction methods of previously inaccessible shale formations. Hydraulic fracturing, or fracking, involves the injection of high-pressure fluids to access shale formations, while horizontal drilling allows for access to a larger area of the formation compared to traditional vertical drilling. This increased contact with the shale formation results in high production rates, and greater overall hydrocarbon recovery (Hunt, 2012). Coupled with numerical reservoir simulation, a tool instrumental in understanding reservoir behavior and optimizing production techniques, these technologies have substantially influenced decisions regarding well placement, fracture design, and production forecasting, contributing to the unprecedented growth in natural gas production. However, as innovative and game-changing as this technology has been in bringing us to this point, as an industry we are still using old, conventional techniques to analyze, model, and optimize recovery from unconventional shale plays (Mohaghegh, 2017).

“When it comes to production from shale using long horizontal wells that are hydraulically fractured in multiple stages, these conventional technologies are too simplistic and are not capable of realistically modelling the physics (as much of it as we understand) of the problem. Therefore, they make unreasonable simplifying assumptions to a degree that make their use all but irrelevant. However, in the absence of any other widely accepted technology as an alternative for modeling the storage and transport phenomena in shale, these technologies flourished in the past several years.” (Mohaghegh, 2017)

Numerical reservoir simulation has long been a cornerstone of the petroleum industry, providing valuable insights and predictions for reservoir performance and optimization. However, with the rise of unconventional shale plays and the limitations of traditional simulation models, there is a pressing need to explore alternative approaches.

This study aims to investigate the use of artificial intelligence and petroleum data analytics for reservoir management and production operation of shale. This approach, which is also referred to as “Shale Analytics” within this study, demonstrates how the existing data from the development of shale assets can help in developing a better understanding of the nuances associated with the operation of shale wells

and by utilizing modern tools such as artificial intelligence (AI) models for unconventional shale plays, one can overcome the shortcomings of conventional numerical models and enhance overall efficiency and accuracy (Mohaghegh, 2017).

1.1 Problem Statement

Although numerical reservoir simulation has been effective for analyzing complex reservoir problems, the introduction of unconventional shale plays has uncovered several disadvantages. The common concern of reservoir simulation and modeling is accuracy, nonetheless conventional models often rely on unrealistic simplifications and assumptions. This can lead to inaccurate results incapable of realistically modeling the physics of the problem, particularly for long, multi-stage horizontal wells common in shale production. Furthermore, these models are often time-consuming, require extensive data-collection, and are limited by the quality of the data they are based on (Mohaghegh, 2017).

Artificial Intelligence (AI) data-driven models and Shale Analytics offer a promising alternative to traditional numerical simulations, particularly in the context of unconventional shale plays. AI models can significantly reduce the time and resources required for reservoir simulation while minimizing or eliminating the assumptions often present in conventional models. These methods and advanced algorithms can handle the complexities of shale behavior more effectively, leading to more accurate predictions and optimizations.

This research aims to compare the effectiveness of AI techniques against conventional numerical reservoir simulation in the context of unconventional shale plays. The analysis will focus on the accuracy of the predictions, the resources required for model development and simulation, and the ability of each technique to capture the complexities of shale behavior. By evaluating the relative strengths and weaknesses of each approach, the research will provide insights into the potential advantages of Shale Analytics, data-driven AI models, and their workflows for reservoir simulation, predictive analysis, and optimization.

1.2 Research Objective

The primary research objective of this study is to gain an in-depth understanding of how Artificial Intelligence models and Shale Analytics can be applied to reservoir simulation and optimization within the context of unconventional shale plays. By comparing the performance of AI models to traditional numerical simulations, this research aims to pinpoint areas where increased efficiency and accuracy can be achieved. The anticipated outcomes are a collection of guidelines for incorporating AI models into

reservoir simulation for shale plays. The study also suggests further advancements in AI-based modeling methods within the energy industry. This research has the potential to considerably enhance reservoir simulation processes and aid in optimizing production strategies for unconventional shale plays.

This will be achieved through the following objectives and methodologies:

1. Investigate traditional workflows and procedures followed in building numerical reservoir simulations. For this, Computer Modelling Group's (CMG) GEM reservoir simulation software was used. This analysis will:
 - a. Demonstrate the impact of modifying key parameters, such as porosity, permeability, fracture half-length, conductivity, and the number of producing fracture perforations, on the production prediction for a single well (MSEEL MIP-3H).
 - b. Discuss the attribute parameters, including their determination and the modified assumptions made, along with the justification for each assumption.
 - c. Present a detailed review of steps taken and results found from tuning the assumptions and the impact reflected in the production history match of the scenario models.
2. Investigate the application of a data-driven AI model approach for developing predictive models using a dataset of 400 unique wells within the Marcellus Shale. For this purpose, IMPROVE™ software provided by ISI (Intelligent Solutions, Inc.) was utilized. This analysis will:
 - a. Develop a predictive Data-Driven AI model.
 - b. Select input attributes for the AI model neural network (NN) based on domain expertise, patterns and trends observed in the data through analysis, and the level of influence on the desired output (Gas Production).
 - c. Demonstrate the ability of the model to make predictions quickly and accurately on production without assumptions, and validate these predictions through blind validation.
 - d. Explain the steps and results obtained and discuss potential improvements through parameter fine-tuning.
 - e. Discuss the wells and the available attribute data, their likely influence based on domain expertise and how the analysis for developing the model depicts them.

3. Perform a systematic comparison of the NRS and AI NN approaches with respect to predictive accuracy, data requirements, workflow complexity, and time efficiency. This will enable the identification of the most suitable approach for production prediction in reservoir simulations and facilitate the development of improved strategies for reservoir management and optimization.
4. Based on the comparative analysis, provide recommendations to enhance overall performance, reliability, and efficiency in production prediction in reservoir simulations. This will contribute to the advancement of the field and promote the adoption of innovative methodologies in the energy industry.

Chapter 2: Literature Review

2.1 Overview

Within the context of the global energy industry, accurate forecasting of oil and gas production is a critical endeavor to ensure efficient resource management, economic planning, and strategic decision making. Existing forecasting techniques range from empirical decline curve analysis to more complex numerical reservoir simulations, each offering unique insights into reservoir behavior and opportunities for production optimization. Despite these strategies, the rapid evolution of the industry necessitates ongoing innovation and research. The scope of this study is grounded in these ongoing advancements, primarily focusing on the emergence of data-driven Artificial Intelligence (AI) models as transformative tools for production forecasting.

Unconventional resources, such as shale gas, differ considerably from their conventional counterparts. Conventional resources are typically characterized by permeable rock formations that naturally facilitate the flow of oil and gas. In contrast, shale gas is trapped in low-permeability formations, making its extraction more challenging and introducing new dimensions of complexity to the process. The unique characteristics and low permeability of these unconventional reservoirs demand the use of advanced extraction techniques and complex analyses.

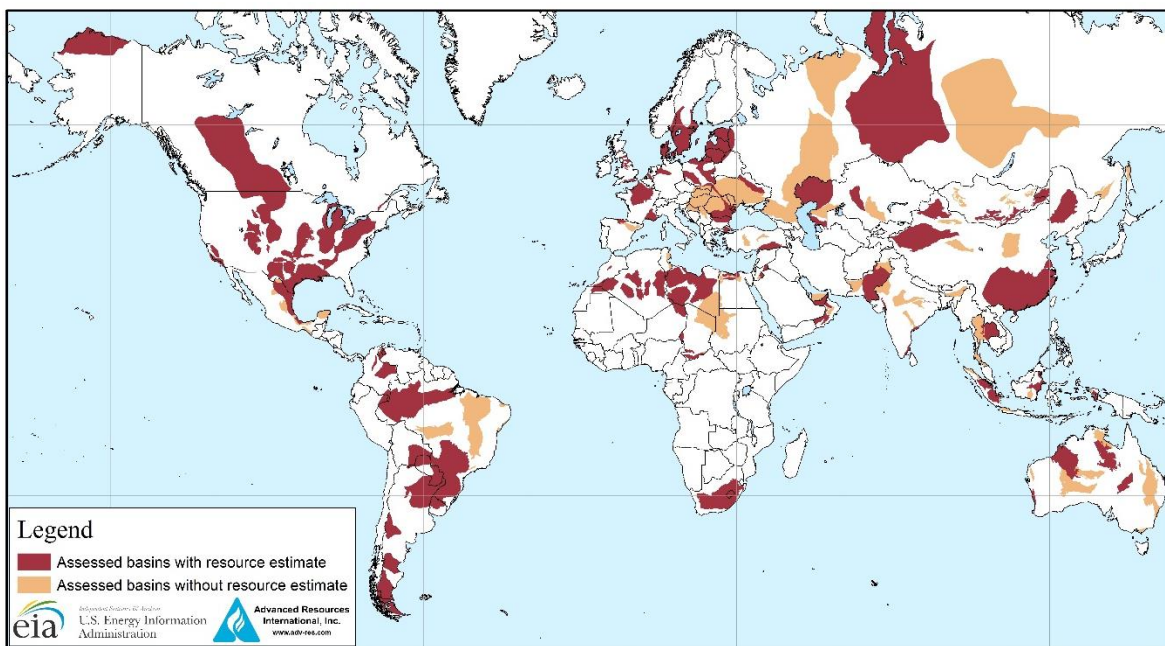


Figure 1: Worldwide Shale Gas Resources (EIA, 2015)

The term "Shale Revolution" refers to the innovative utilization of hydraulic fracturing and horizontal drilling techniques, which have led to a substantial increase in oil and gas production in the United States. These technological breakthroughs have made once-inaccessible shale gas resources exploitable, fundamentally reshaping our understanding of natural gas reserves and the methods of their extraction. As depicted in Figure 1 (EIA, 2015), shale gas reservoirs are globally abundant.

These unconventional shale reservoirs dramatically differ from conventional reservoirs in terms of extraction techniques, resource abundance, and their economic and environmental impacts. However, despite the technological advancements that have facilitated the extraction of these resources; fully understanding and optimizing shale gas production remains a considerable challenge due to the inherent complexities of these reservoirs.

This literature review examines the historical development, current utilization, and challenges of these technologies within the context of unconventional shale reservoirs. It then introduces AI models and Shale Analytics as innovative solutions to address these complex problems, with the potential to transform the future of oil and gas production.

2.2 Numerical Reservoir Simulation

Numerical reservoir simulation has proven itself instrumental as a tool in reservoir management and engineering to forecast the flow of fluids such as oil, water, and gas through porous media. With nearly all major modern reservoir development decisions based in some way on simulation results (Watts, 1997), this technology has been used to predict the productivity of a well, optimize the placement of wells, and to identify areas of the reservoir that are most likely to be productive.

The first reservoir simulation began in 1954, using electronic punch cards to perform numerical methods based on the radial gas flow equations of (Aronofsky & Jenkins, 1954). This work introduced one of the first practical methods for predicting and interpreting the flowing well pressure history in a developed natural gas reservoir. It was considered a breakthrough as it demonstrated that reservoir simulation could be used to optimize production strategies and to estimate ultimate recovery. In the years to come, the work done by notable researchers and major oil company laboratories led to the development of even more advanced reservoir simulators, eventually reaching a wider audience with D.R. McCord and Associates being the first to commercially market a reservoir simulator in 1966 (Watts, 1997).

Advancements in high-speed computers and technology over the next several decades led to reservoir simulation becoming widely adopted in the industry for use in conventional reservoir management and

development. Today's reservoir modeling and simulation practices reflect many of the foundational principles established during their early development.

Fundamentally, reservoir simulation operates on the principle of simultaneously solving the flow equations between adjacent blocks of rock. This takes place beginning within a reservoir model, constructed as an array of discrete cells to represent the physical space of a subsurface oil, gas, or water reservoir. The array of cells is typically three-dimensional, although 1D and 2D models are sometimes used.

To enhance the accuracy and reliability of the simulation, each cell in the model is characterized by associated reservoir properties such as porosity, permeability, and fluid saturation, using data collected from well logs, core samples, and production data. The model also integrates the reservoir's physical attributes including its size, shape, and depth. In addition to these static parameters, dynamic aspects are also incorporated into the model. These can include changes in pressure, temperature gradients, and the resulting fluid behaviors due to these variations.

From this point, a conceptual model of the reservoir system is developed and serves as a simplified representation of the real system. The conceptual model is then used to simulate how the system will behave under varying inputs and conditions. Additionally, it forms the groundwork for the development of a more detailed mathematical model.

The mathematical model represents a series of equations depicting the physical processes that occur within the reservoir system. This set of equations is employed to simulate the operation of the reservoir system, examining various input data to predict the system's response (Coats, 1969). The results of the simulation can then be used to make decisions about how to operate the reservoir system.

However, despite their effectiveness in supporting conventional reservoir management, it's essential to understand that reservoir simulation models are not perfect. There are limitations and potential sources of error in the calculated outcomes that must be understood when used for decision-making purposes.

One fundamental issue arises from the model's inherent approximations, as it involves certain assumptions which may only be partially valid or accurate to the actual reservoir conditions. Additionally, replacement of the model differential equations by difference equations can lead to truncation error. This happens when there is a small disparity between the exact solution of the difference equation from the solution to the original differential equation.

Lastly, and arguably of most importance, uncertainty in the reservoir description data, including parameters such as permeability and porosity distributions, significantly adds to the overall error of the simulation model. This is due to these parameters being difficult to accurately measure, as reservoirs are inherently heterogeneous, and the rock and fluid properties vary from point to point making it difficult to accurately represent in a numerical model (Coats, 1969). It is important to understand these limitations and how they can affect the accuracy of the model's output.

2.2.1 Numerical Reservoir Simulation: Unconventional Shale Reservoirs

Unconventional reservoirs are defined at present as those reservoirs whose porosity, permeability, fluid trapping mechanism, or other characteristics differ from conventional sandstone and carbonate reservoirs. Where conventional reservoirs, traditionally defined as a reservoir in which buoyant forces keep hydrocarbons in place below a sealing caprock, use the natural energy in the environment to flow oil and gas to the surface unaided, unconventional reservoirs require the introduction of energy to facilitate extraction and exploitation (SLB Energy Glossary, 2023b).

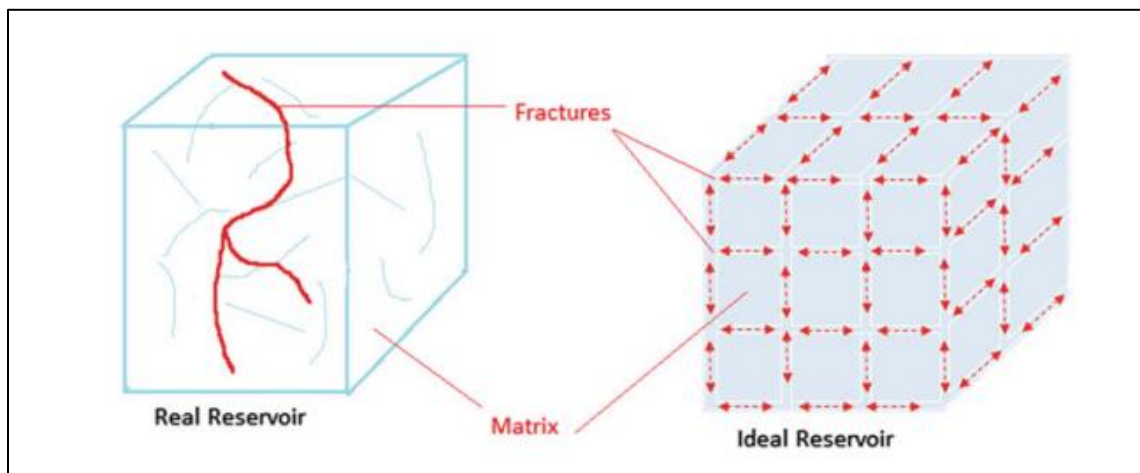


Figure 2: "Sugar Cube Model" of Naturally Fractured Reservoir (Warren & Root, 1963)

Naturally fractured reservoirs, such as shale reservoirs, are characterized by the presence of two distinct types of porous media: the porous matrix and the fracture network. Often termed as dual porosity systems, these reservoirs are unique in their coexistence of these two types of porous media (Barenblatt et al., 1960). In these systems, the porous matrix predominately acts as a fluid supplier to the fractures, which subsequently shape a well linked, continuous network. The dual porosity model, established by (Warren & Root, 1963), divides the reservoir into interconnected matrix and fracture subsystems, each exhibiting distinct characteristics of porosity, permeability, and either compressibility or connectivity.

They conceptualized this system as an orthogonal arrangement of intersecting fractures, supplemented by cubic matrix blocks, referred to as the Sugar Cube Model as shown in Figure 2. These models, based on Darcy's law and the intercommunication between the matrix and fracture, have transformed the understanding and analysis of naturally fractured reservoir and fluid flow, providing utility across various flow regimes.

Shale Gas reservoirs, as unconventional energy sources, are organic-rich formations which serve dually as the source rock and the reservoir. The gas is largely confined to the limited pore space available within the rock, while a substantial fraction of gas is adsorbed onto the organic material. These reservoirs exhibit extremely low permeability and thus require effective stimulation strategies to produce economically (Gholinezhad et al., 2018).

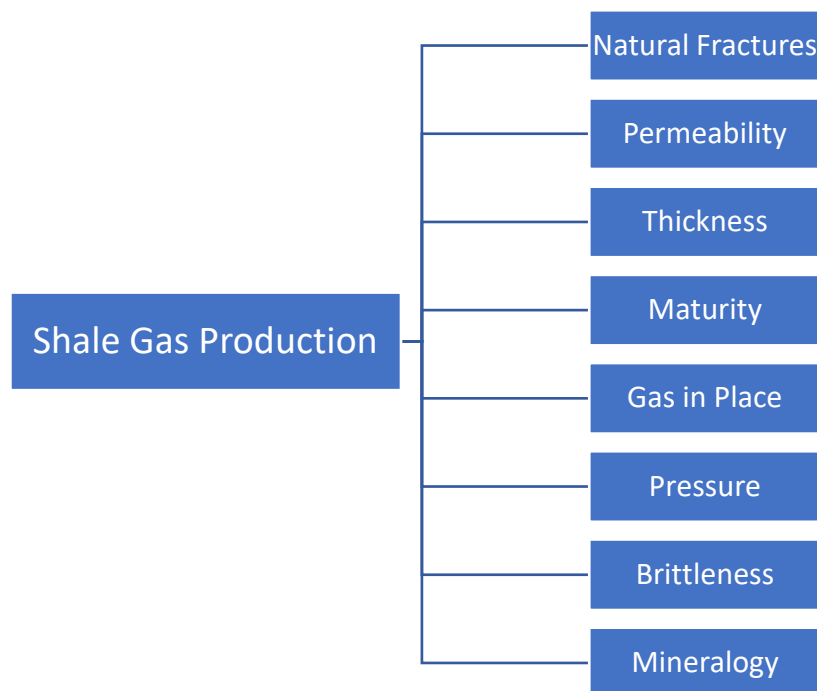


Figure 3: Key elements needed for successful shale gas play (Kundert & Mullen, 2009)

Figure 3 depicts the essential components for a commercially successful shale play, as identified by (Kundert & Mullen, 2009). These key components – maturity, free gas, total gas in place, thickness, natural fractures, and reservoir pressure – are found in varying proportions within a shale reservoir. Other favorable shale reservoir properties include organic richness, brittleness, and mineralogy.

Permeability is a critical parameter in understanding unconventional shale reservoirs. However, unlike conventional reservoirs, permeability is not the primary factor that dictates production from

unconventional shale reservoirs. Instead, production from unconventional shale reservoirs is limited by the ability to create fractures in the rock that allow fluids to flow more easily. Therefore, hydraulic fracturing, or fracking, is used to produce unconventional shale reservoirs. Fracking involves injecting a high-pressure fluid into the rock to create fractures. These fractures allow fluids to flow more easily, which increases production.

In comparison to conventional reservoirs, ultra-low permeability unconventional reservoir exploration is in its infancy, thus modeling conventional and unconventional reservoirs using numerical reservoir simulation models requires a fundamental understanding of techniques used to arrive at this point. In the case of conventional reservoirs, these models typically use a single-porosity approach, assuming the entire reservoir is made up of a single rock type. This approach has worked well as conventional reservoirs often boast high permeability, allowing the hydrocarbon to flow easily throughout the rock.

Unconventional reservoirs are more difficult to model using numerical reservoir simulation due to their inherent added complexities. These models typically require a dual-porosity approach, assuming the reservoir is made up of two types of rocks: a matrix and a fracture network. The matrix is the solid rock, and the fracture network is the network of highly interconnected natural fractures and fissures in the rock where the hydrocarbons are stored within the matrix, and flow to the wellbore through the fracture network. Small pore sizes and property heterogeneities at different scales can also dramatically change flow processes and related physical phenomena.

Several widely used commercial simulators for shale gas simulation include CMG/GEM™ developed by Computer Modelling Group (CMG), Eclipse 300™ created by SLB, and COMET3™ developed by Advanced Resources International Inc. (Andrade et al., 2011). Each simulation tool offers unique features, which can be applied more effectively based on the characteristics of the reservoir being modeled.

(Islam et al., 2016) discussed that reservoir simulation has a "real" side and an "imaginary" side. The real side refers to tangible data and outcomes, such as physical reservoir properties, initial and boundary conditions, and production data. Conversely, the imaginary side is composed of parameters, mathematical models, and outcomes not directly observable such as grid cells, fluid viscosity, and simulation results. These two sides interconnect, with the real side informing the imaginary side, and the imaginary side producing results that help manage and understand the real reservoir. A visual depiction of this interconnected workflow can be seen in Figure 4. However, this interdependence introduces a

degree of uncertainty in the output results, as minor fluctuations in the physical properties can potentially lead to significant changes in the predicted reservoir behavior.

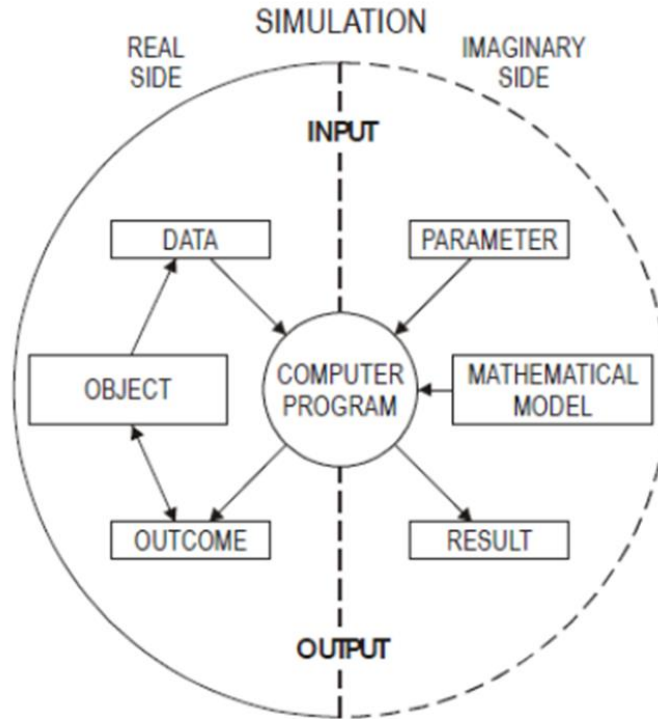


Figure 4: "Real" and "Imaginary" sides in Simulation Modelling (Islam et al., 2016)

The level of uncertainty in reservoir parameters, and the assumptions used, can directly influence the outcome of each model. In the creation of a reservoir model, the caliber of data used is vital, as it dictates the model's accuracy and ability to realistically represent the reservoir. Nonetheless, many of these parameters are not easy to measure, often only offering a glimpse of a small portion of the full reservoir. As such, certain assumptions and compromises are required to proportionately scale these parameters into a comprehensive representation of the entire reservoir.

For instance, direct porosity measurements are obtained by laboratory measurements of core samples brought to the surface during drilling operations. While these samples depict conditions near the wellbore, they are frequently integrated into models that suggest a nearly uniform reservoir area. However, this doesn't accurately represent shale reservoirs, which exhibit varying and heterogeneous porosity throughout. This scenario provides a segway into discussing what we refer to as "Hard Data" and "Soft Data," which will be demonstrated later in this study as data-driven techniques use "Hard Data" to

model production from shale, while many of the traditional technologies and numerical reservoir models rely on “Soft Data.”

The term “Hard Data” is used to describe field measurements, which are often directly gathered during operations and do not contain assumptions when applied to a model. For example, in hydraulic fracturing operations, parameters such as fluid type and amount, proppant type and amount, injection, breakdown and closure pressures and rates fall under “Hard Data.” Most shale assets offer a detailed account of such data recorded in reasonable detail.

Remaining within the scope of hydraulic fracturing, “Soft Data” refers to variables that require interpretation, estimation, or conjecture. For example, the hydraulic fracture half-length, height, width, and conductivity are unmeasurable directly. Software applications may be utilized to estimate these parameters, but the overly simplified assumptions made undermine the relevance of “Soft Data” in the design and optimization of frac jobs (Esmaili & Mohaghegh, 2016).

“... although “Soft Data” may help engineers and modelers during the history matching process, it fails to provide a means for truly analyzing the impact of what is being done. (Mohaghegh, 2017)”

Table 1 shows a variety of examples of both “Hard Data” collected from hydraulic fracturing and “Soft Data” commonly employed by reservoir engineers and modelers.

Table 1: “Hard Data” and “Soft Data” in Hydraulic Fracturing (Mohaghegh, 2017)

“Hard Data”		“Soft Data”
Fluid types	Proppant amounts (lbs)	Hydraulic fracture Half-Length
Fluid amounts (bbls)	Proppant concentration	Hydraulic fracture Width
Pad volume (bbls)	Injection rates	Hydraulic fracture Height
Slurry volume (bbls)	Injection pressures	Hydraulic fracture Conductivity
Proppant types		Stimulated Reservoir Volume

Understanding the workings of numerical reservoir simulation, its original use case, and the data types used in developing reservoir models helps us appreciate the present state of shale reservoir modeling technology. Despite their acknowledged limitations, conventional techniques originally intended for traditional reservoirs are commonly used with the consensus that these methods are the best currently available for numerical simulation of fluid flow through shale.

However, the utility of new data-driven AI models is emerging in this space. These models have the potential to not only expedite the process and lessen processing power required, but also mitigate many assumptions. This opens a promising path forward in shale reservoir modeling, redefining our use of “hard” data (directly measured and unchangeable) and “soft” data (not fully understood, assumed, and easy-to-change).

As we transition into an analysis of scholarly papers focused on numerical reservoir simulators for shale, it is crucial to keep these considerations in mind, particularly the distinction and significance of ‘hard’ and “soft” data. The exploration of these papers will further illuminate how these methodologies and technologies are applied and developed.

2.2.2 Applications of Numerical reservoir Simulation in Unconventional Shale Reservoirs

The research analysis on numerical reservoir simulation in shale has unsurprisingly uncovered numerous studies that successfully employ this tool to generate innovative discoveries, while also acknowledging the inherent assumptions and limitations. This review will showcase notable accomplishments in the field and explore the challenges faced by researchers, establishing the necessary groundwork for our own study in this intricate and evolving arena.

(El Sgher, 2021) conducted an examination of how propped fracture conductivity, influenced by net stress, impacts horizontal wells with multiple fractures in the Marcellus shale. To develop the base reservoir model for the investigation, they utilized a commercial reservoir simulation tool using data from published laboratory studies to predict the production performance. The results were then compared to the production history for evaluation and verification, with the model then used to perform several parametric studies.

Upon developing their base model, (El Sgher, 2021) performed history matching by altering the hydraulic fracture half-length, while keeping the remaining reservoir properties constant. These alterations, which aimed at achieving an optimal history match with the well’s known production, illustrated in Figure 5 highlight the presumptive variations required by reservoir model simulators. This is an attempt to formulate a model that best replicates the reservoir’s behaviors, even though it is understood that the model does not completely mirror the reality of the reservoir. They make note of the difficulty in accurately modeling the complex nature of shale reservoirs, due to challenges in ascertaining good quality reservoir information and the low permeability of the shale matrix. The base model developed for their work was constructed using model parameters seen in Table 2.

Table 2: Basic Model Parameters for MIP-6H and MIP-4H

Basic Reservoir Model Parameters for MIP-6H and MIP-4H		
Parameters	Value	Unit
Model Dimensions (MIP-6H)	4000 (Length) x 1000 (Width) x 90 (Height	ft.
Model Dimensions (MIP-4H)	4500 (Length) x 1500 (Width) x 90 (Height	ft.
Initial Pressure	4700	psia.
Fissure Porosity	0.0001	Fraction
Matrix Porosity	0.02	Fraction
Fissure Permeability i, j, k	0.0013, 0.0013, 0.00013	md
Matrix Permeability i, j, k	0.000124, 0.000124, 0.0000124	md
Gas Saturation	0.85	Fraction
Water Saturation	0.15	Fraction
Density	120	lb/ft ³
Langmuir Pressure	0.0023	psi ⁻¹
Langmuir Volume	0.12	gmol/lb
Fracture Spacing for (MIP-6H)	340	ft.
Fracture Spacing for (MIP-4H)	380	ft.

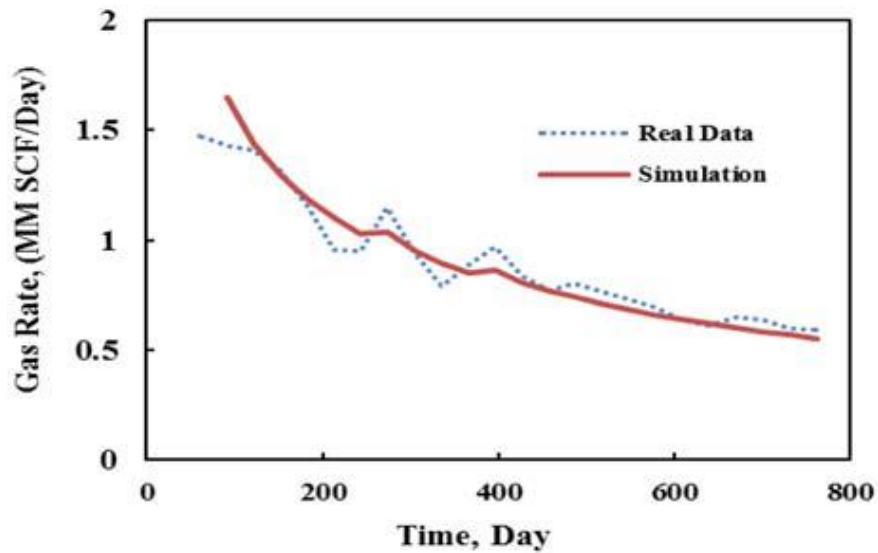


Figure 5: History Matching for MIP-6H with geo-mechanical impact for first two years (El Sgher, 2021)

(C. L. Cipolla et al., 2010) illustrated the impact of gas desorption on the production profile and ultimate gas recovery in shale reservoirs. Moreover, they provided an exploration of how fluctuating closure stress distribution within the fracture network impacts the well productivity and gas extraction. The reservoir simulations were conducted using a comprehensive numerical grid, described as meticulously replicating the intricate fracture network, primary fracture, and tight shale matrix. Nonetheless, their conclusions propose that, although the numerical reservoir simulation history match may not be unique in the complexity of shale gas reservoirs, the modelling can be effectively constrained given that matrix permeability estimates are available from core analyses. Their research underscores the need for a more comprehensive and refined method of data analysis in unconventional reservoirs.

(Liu, 2022) critically examined the growing importance, unique challenges, and inherent complexities of unconventional natural gas resources from ultra-low permeability reservoirs, including tight sandstone and shale. The study highlights the complex factors influencing gas production, such as nonlinear flow equations, multi-scale fracture networks, and heterogeneous, stress-sensitive rocks that drive gas production in these reservoirs. The author reviews mathematical numerical simulation methodologies as an approach to understand reservoir capacity and key influential parameters. Additionally, they delve into current challenges, including the absence of standard procedures for determining physical parameters and the need for more effective prediction models for future production profiles. While acknowledging advancements in simulation software, Liu cautions that the exploration of ultra-low permeability reservoirs is still in its infancy. He advocates for further research to enhance the reliability and accuracy of new models and methodologies, while also suggesting the need for new techniques for more reliable future production predictions in these unconventional reservoirs.

(Guo et al., 2023) examined the limitations of traditional numerical reservoir simulation models, and subsequently introduced a novel approach – the Reservoir Graph Network (RGNet). Their research effectively highlights the complications associated with reservoir models, primarily due to their inherent complexity and difficulties requiring regular updates. The advent of RGNet, described as a generalized data-driven approach that models a reservoir using a set of drainage volumes controlled by wells, promises a hybrid model that seamlessly merges reservoir physics and data-driven methods, thereby minimizing complexity and runtime. However, a more thorough discussion and validation of the RGNet's practical applications and potential limitations could further enhance the study's merits.

2.3 Data-Driven AI models and Shale Analytics

In recent years, there has been an incredible surge in the popularity of Artificial Intelligence (AI) and Machine Learning (ML) algorithms. The rapid advancements in technology have placed these algorithms into the spotlight, capturing the attention and interest of numerous industries. The exponential growth in the adoption of AI and ML signifies their transformative impact, as organizations across different sectors recognize the immense value and potential these algorithms hold.

These techniques have become of interest for petroleum engineers tackling problems involving big data analytics, offering a promising alternative to traditional numerical simulations. AI models can significantly reduce the time and resources required for reservoir simulation while minimizing or eliminating the assumptions often present in conventional models (Mohaghegh, 2017). These methods and advanced algorithms can handle the complexities of shale behavior more effectively, leading to more accurate predictions and optimizations.

Data driven AI models are created by using historical data and computational techniques to reveal trends, identify patterns, and make predictions. The vast amounts of existing data routinely collected throughout the development of shale assets can help in developing a better understanding of the nuances associated with the operation of shale wells. Utilizing modern tools such as data-driven AI models for unconventional shale plays, one can overcome the shortcomings of conventional numerical models and enhance overall efficiency and accuracy.

Data, rather than physics and geology, is the driving force behind this technology (Mohaghegh, 2017) and the energy industry's collection of vast amounts of data further fuels this shift. For example, AI models have been used to predict well performance, optimize production strategies, and identify new drilling opportunities. As a result, AI and ML are becoming increasingly powerful and valuable tools for petroleum engineers, offering viable alternatives to conventional solutions traditionally used in the industry.

(Mohaghegh, 2017) introduced the term Shale Analytics, defined as the application of Big Data Analytics in shale integrating data science, artificial intelligence, and machine learning, enabling a multi-pronged approach to reservoir management. Its primary tasks involve understanding patterns in collected data, identifying key parameters controlling production, ranking areas based on response to design implementations, and constructing predictive models. These models estimate well performance, employing variables such as reservoir characteristics, well spacing, and completion parameters.

Moreover, they are validated using blind wells, i.e., wells that were not part of the predictive model's development.

The success of Shale Analytics has been witnessed in various shale formations, including Marcellus, Utica, Eagle Ford, Bakken, Niobrara, with over 3000 wells having been evaluated. The predictive models provide a comprehensive toolset to quantify uncertainties related to well productivity, evaluate historical frac jobs, estimate reserves potentially missed due to sub-optimal practices, and gauge the effectiveness of previous completions and stimulation practices. Furthermore, these models can help identify optimized designs for new wells.

Chapter 3: Objectives and Methodology

3.1 Objectives and Methodology Summary

The objective of this research is to conduct a comparative analysis of Artificial Intelligence (AI) techniques and traditional numerical reservoir simulations (NRS) to improve the efficiency and effectiveness of natural gas production within the scope of unconventional Marcellus Shale plays. The focus will be:

1. To compare the prediction accuracy of each methodology (AI and NRS)
2. To compare the resources required for model creation and simulation
3. To compare the ability of each technique to accurately depict the intricate behavior of shale formations
4. To compare the accuracy of each technique in predicting and history matching production
5. To identify the strengths and weaknesses of each approach

Through a thorough evaluation of both the AI techniques and traditional numerical simulations, the study aspires to shed light on the benefits of employing data-driven AI models and their respective workflows for reservoir simulation, predictive analytics, and optimization procedures. This research intends to illustrate the potential advantages and explore how these AI models can revolutionize conventional practices by increasing efficiency and accuracy.

To achieve the aim of this study, the methodology section will provide a detailed overview of the methods used to collect and analyze data for each of the two methodologies.

3.2 Methodology: Numerical Reservoir Simulation (NRS)

The methodology for the development and analysis of the Numerical Reservoir Simulation model used in this study will be structured as follows:

Methodology for Numerical Reservoir Simulation (NRS)

1. Data collection and analysis
 - Study area: collect data on the geology, petrophysics, and production history of the Marcellus Shale Well (MSEEL MIP-3H)
 - Analyze data to identify key parameters affecting well's production
2. Model development
 - Develop a NRS model using CMG software
 - Calibrate model to well's production history

3. Parametric studies and Predictive Analysis

- Conduct parametric studies to investigate the impact of uncertainty in the “soft” data on the production history match within NRS model

A high-level overview of the Numerical Reservoir Simulation methodology workflow is seen in Figure 6.

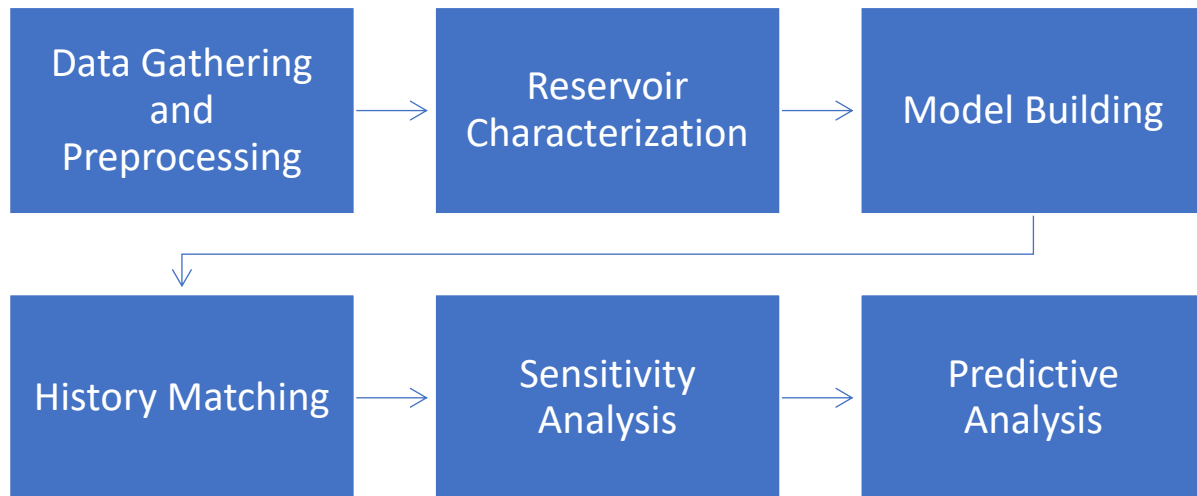


Figure 6: Process Workflow - Numerical Reservoir Simulation Methodology

3.2.1 Data Collection: MSEEL Site and MIP-3H Well

The Morgantown Industrial Park (MIP) site in West Virginia, U.S., hosts the Marcellus shale well that is the primary subject within this portion of the study. The MIP site falls under the purview of the Marcellus Shale Energy and Environment Laboratory (MSEEL), a multidisciplinary research project supported by the U.S. Department of Energy. The primary goal of MSEEL is to foster a comprehensive understanding of unconventional shale reservoirs through advanced reservoir characterization and monitoring techniques, with a focused effort to enhance recovery efficiency and mitigate environmental impacts of unconventional resource development.

The MIP site accommodates four horizontal Marcellus Shale wells (MIP-3H, MIP-4H, MIP-5H and MIP-6H), as well as a vertical scientific observation well (MIP-SW), illustrated in Figure 7. The Marcellus shale well under consideration for this study is the MIP-3H, a 28-stage horizontal well with a lateral length of 6,058 feet. Outfitted with several multi-scale and multi-sensor measurement tools, the available data includes parameters such as surface pressure, surface temperature, petrophysical logs, geo-mechanical logs, and production logs. The data, accessible to the public via the MSEEL website, was utilized for the purposes of this study (Carr et al., 2015).

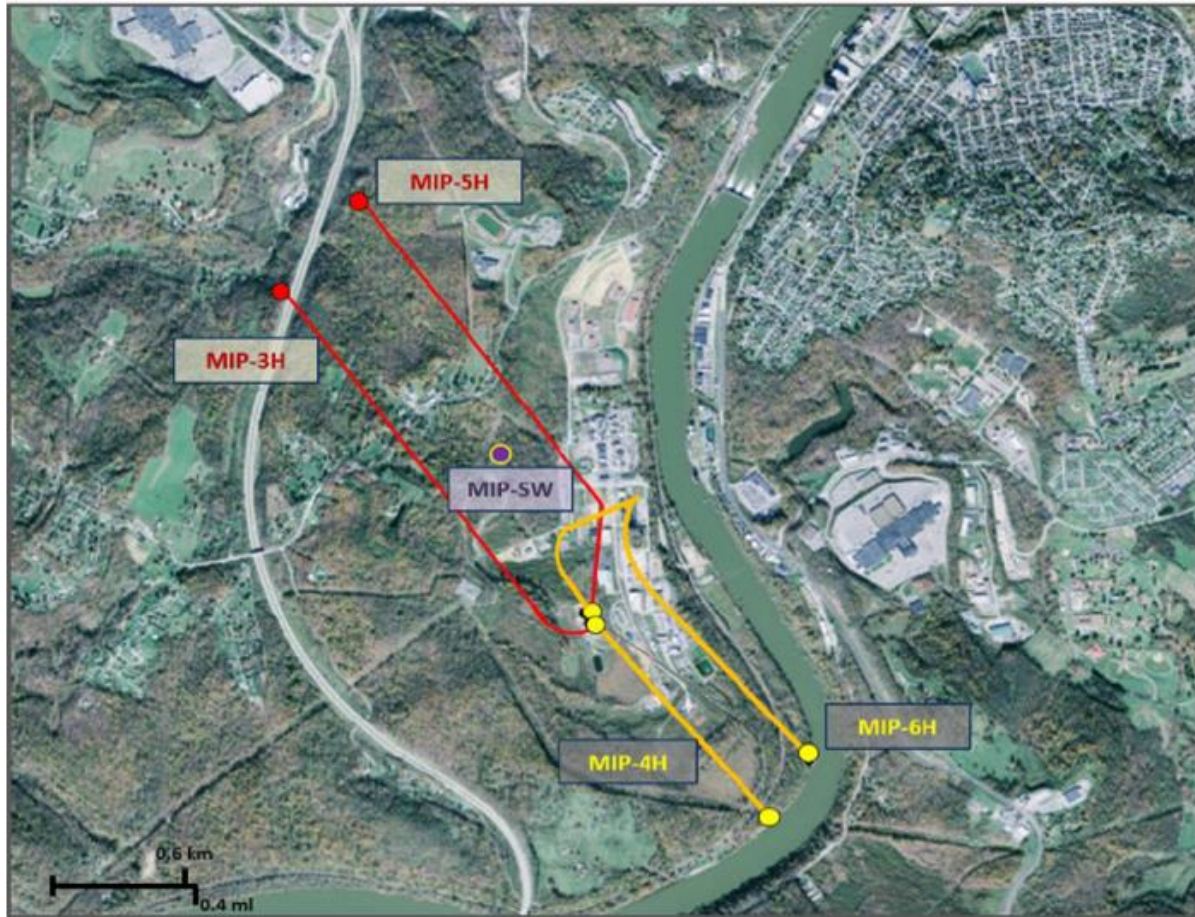


Figure 7: Well layout in the Marcellus Shale Energy and Environment Laboratory (MSEEL) MIP site

Having established the context of our study and introduced the MIP-3H well, we will next discuss the nuanced datasets and methodologies that are essential to our work. This data, which covers everything from geo-mechanical details to completion designs, will form the foundation of our ongoing analyses. We will then break down these findings and explain their significance in guiding our research.

A comprehensive set of geo-mechanical and petrophysical logs were acquired using the Schlumberger Sonic Scanner tool, as shown in Figure 8. The logs included gamma ray, resistivity, density, neutron, and acoustic logs, providing insights into the reservoir rock's mechanical properties and stress responses, such as Young's modulus, Poisson's ratio, minimum horizontal stress, pore pressure, and overburden pressure.

In conjunction with measurements derived from core samples collected during the drilling operations, from the observation well (MIP-SW), these logs provided complimentary data on the reservoir parameters, such as lithology, thickness, porosity, permeability, rock density, initial saturation, and initial pressure (Elsaig et al., 2016). The logs provided continuous data along the borehole, while the core

samples provided discrete data that could be analyzed in detail. Together, the data from the logs and core samples helped to establish the basic shale properties.

The completion design for the well consists of 133 perforation clusters across its 28 stages, which are sectioned into five parts from the toe to the heel. Each section was uniquely designed to incorporate different fracture treatment methodologies, as seen in Table 3, to observe increased efficiency and impact on production. Each stage is roughly 200 feet long with 4 or 5 perforation clusters, each 3-5 feet in length and consisting of 5-6 shots per foot. Two proppant types were implemented in the treatment design, 100 mesh sand and 40/70 mesh white sand, with proppant proportions varying across sections.

A production logging operation was performed using the Schlumberger Flow Scanner to acquire data on fluid production and movement in the wellbore. The data included measurements of mini-spinners, water holdup, and gas holdup. Due to obstructions, a depth limit of 13,530 ft was noted for the Flow Scanner, causing seven perforation clusters to be incompletely logged. However, rates exceeding this depth were collectively noted for these clusters. Production history was collected with data available for a period of 2036 days (2015-12-12 to 2021-07-08), seen plotted in Figure 9.

The extensive data collected from the MSEEL study area forms a robust foundation for our research and the subsequent development of the MIP-3H base reservoir model. The diverse datasets, ranging from geo-mechanical logs to completion design details, offer an in-depth understanding of the reservoir's physical and mechanical characteristics. This, coupled with the precise measurements from tools like the Schlumberger Sonic Scanner and Flow Scanner, ensures a comprehensive overview of the reservoir's behavior and properties. As we move forward, this information will be instrumental in calibrating and refining our models.

Table 3: Completion and stimulation parameters for MIP-3H (Carr et al., 2015)

Section		Stage	Cluster Count	Average Mid-perforation Depth (ft)	Total Shot Count	Shot Density (shot/ft)	Stage Length (ft)	Pump schedule
E	Best Practice Applied	28	4	7,828	40	6	191	A
		27	4	8,017	40	6	184	A
		26	5	8,220	40	6	225	A
		25	5	8,449	32	6	231	A
		24	5	8,678	30	6	222	A
		23	5	8,911	40	6	237	C
		22	5	9,138	40	6	220	C
D	Saphire VEF	21	5	9,363	40	5	218	D
		20	5	9,596	40	5	240	D
C	SLB Engineered Completion	19	4	9,802	32	6	180	C
		18	4	9,981	32	8	180	C
		17	4	10,115	32	6	181	C
		16	4	10,343	26	6	178	C
		15	4	10,526	26	6	186	C
		14	5	10,731	30	6	228	A
		13	5	10,956	30	6	230	A
B	NNE 75% 100-Mesh	12	5	11,186	50	5	231	B
		11	5	11,416	50	5	232	B
		10	5	11,648	50	5	227	B
		9	5	11,878	50	5	237	B
		8	5	12,103	50	5	222	B
		7	5	12,328	50	5	224	B
A	NNE Standard 35% 100-Mesh	6	5	12,557	50	5	245	A
		5	5	12,792	50	5	234	A
		4	5	13,025	50	5	230	A
		3	5	13,259	50	5	238	A
		2	5	13,489	50	5	223	A
		1	5	13,720	50	5	233	A

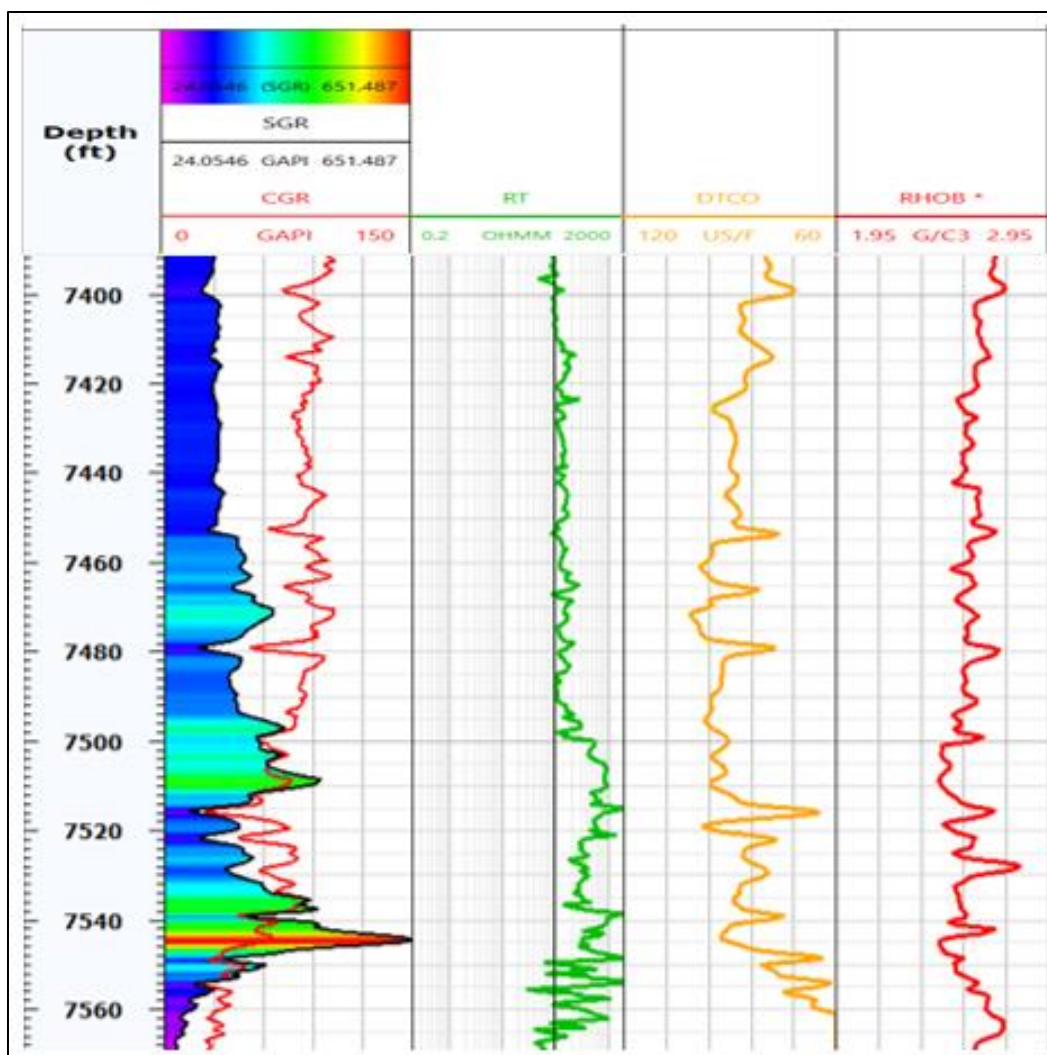


Figure 8: Logs from the Vertical Section of the MIP-3H Well (El Sgher, 2021)

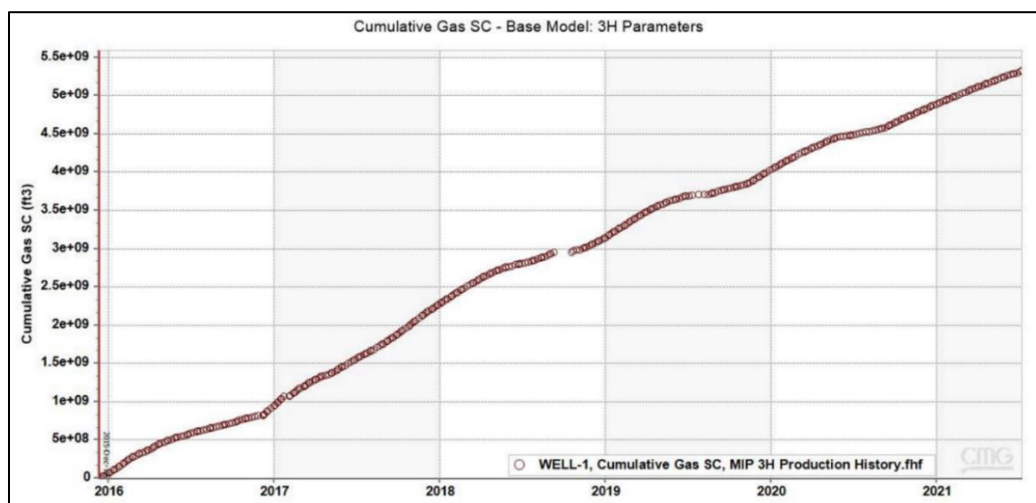


Figure 9: Production History MIP-3H – Cumulative Gas (2015.12.12 - 2021.07.08)

3.2.2 Data Analysis

The development of our NRS model for the MIP-3H well utilized a combination of the parameters and data collected and presented in the preceding section. The most important of these parameters, necessary for the construction of our base reservoir model within CMG software, are shown in Table 4.

Interestingly, many of the parameters can be categorized as “soft” data. As previously mentioned, “soft” data refers to parameters that are not often directly measured and are rather easily assumed or modifiable based on various factors. Figure 10 offers a clear distinction between

parameters, differentiating the “hard” data, obtained through direct measurements, from the “soft” data that are more flexible and susceptible to change.

Reservoir Parameters		Units	
Model Dimensions (MIP-3H)	ft.	Assumed - Easy to Change	
Well Length (Horizontal)	ft.	Measured - No Change	
Initial Reservoir Pressure	psia.	Measured - No Change	
Initial Fissure Porosity	percent	Assumed - Easy to Change	
Initial Matrix Porosity	percent	Assumed - Easy to Change	
Initial Fissure Permeability i, j, k	nd	Assumed - Easy to Change	
Initial Matrix Permeability i, j, k	nd	Assumed - Easy to Change	
Number of Hydraulic Fractures		Measured - No Change	
Fracture Half-Length, X_f	ft.	Assumed - Easy to Change	
Initial Fracture Conductivity, $k_f w_f$	md-ft	Assumed - Easy to Change	
Water Saturation	Fraction	Assumed - Easy to Change	
Rock Density	lb/ft ³	Assumed - Easy to Change	
Langmuir Pressure Constant	psi ⁻¹	Assumed - Easy to Change	
Langmuir Volume Constant	g-mol/lb	Assumed - Easy to Change	

Number of Hydraulic Fractures	No. of HF that Produce	Assumed – Easy to Change
	No. of HF that Do Not Produce	Assumed – Easy to Change

Figure 10: Reservoir Parameter Analysis

The alterable and inferable nature of these parameters pose a particular challenge when it comes to accurately depicting them in heterogeneous reservoirs, like those seen in the Marcellus Shale. In reservoir simulations, some of these parameters, due to their inherent variability and the challenges in measuring them, can introduce uncertainties that can substantially affect the outcome.

The parameters listed below were selected from the available data for their significant impact on reservoir performance, as observed in both real-world scenarios and numerical simulation modelling:

1. **Porosity:** The percentage of pore volume or void space within reservoir rock, indicating the proportion of the rock's volume that can contain fluids such as oil, gas, or water. Field measurements of porosity employ techniques such as core analysis, geophysical well logging, and seismic surveys. However, these methods can be inaccurate due to factors such as heterogeneity of the rock, borehole damage, and invasion of fluids
2. **Permeability:** The ability of a rock to allow fluids to flow through it, typically measured in millidarcies (md). Permeability can be measured in the field using core sample analysis, logging, and mini-frac tests. Like porosity, challenges with methods arise from rock heterogeneity, borehole damage, and the presence of natural fractures.
3. **Fracture half-length:** The distance from the wellbore to the midpoint of the fracture, measured in feet (ft). Fracture half-length can be estimated in the field using production logging, tracer tests, and pressure transient analysis. The precision of these methods can be affected by the complexity of the fracture geometry, the presence of proppant, and reservoir rock heterogeneity.
4. **Fracture conductivity:** A measure of how well a propped-fracture can carry fluids from a well, measured as the product of fracture permeability and fracture width (md-ft). Fracture conductivity can be measured in the field using methods such as production logging, tracer tests, and pressure transient analysis.
5. **Number of producing fracture perforations:** The number of perforations in the wellbore that are open and allow fluids to flow from the fracture. The number of producing fracture perforations can be estimated in the field using production logging. However, this method can be inaccurate due to factors such as the complexity of the fracture geometry and the presence of proppant (Yang et al., 2021)

In summary, the parameters mentioned are challenging to measure accurately in the field. This is because the rocks and formations that contain oil and gas are often heterogeneous and complex, and the processes that create and modify these parameters are not fully understood. As a result, there is always some uncertainty associated with the measurements of these parameters, which can introduce variability into reservoir simulation models.

3.2.3 Model Development

In this portion of the study, the objective was to develop a base model that could accurately predict the production of the MIP-3H well, matching it with known production data. The model parameters used to build this initial base model are summarized in Table 4. Notably, the base model for the MIP-3H aligns closely with similar models and studies carried out by other researchers using the same datasets and parameters collected through the MSEEL project.

After successfully developing a base model for the MIP-3H well that closely matched historical production data, our study will investigate the influence of “soft” data parameters on the model’s production outcomes. This base model will serve as our initial scenario, illustrating how inherent assumptions in “soft” data parameters can contribute to the development of an NRS model.

To assess the impact of changes to “soft” parameters on model predictions, and their alignment with historical production data, this study modified select parameters in subsequent scenarios. We aimed to show that reasonable assumptions can steer the models towards desired outcomes.

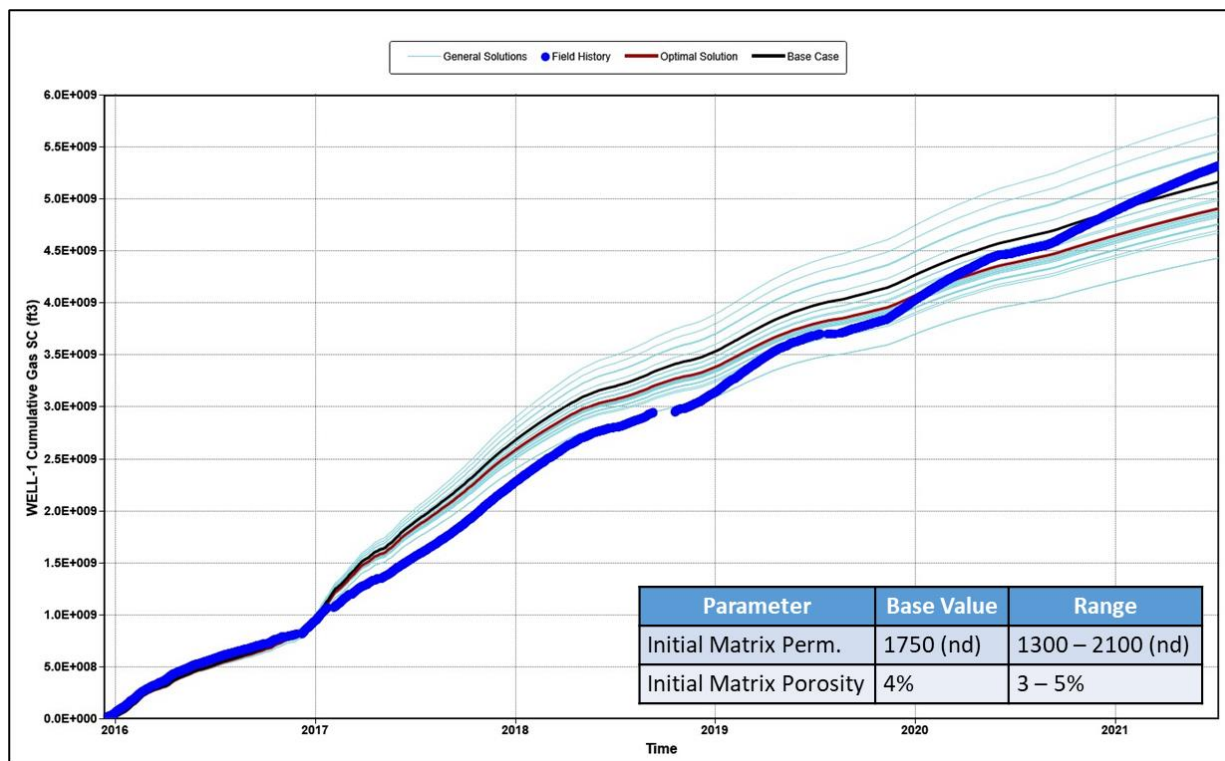


Figure 11: Parametric Study Example – CMG™ CMOST™ Simulations

For instance, in the initial phases of our research, we examined the effects of parameters values such as initial matrix permeability and initial matrix porosity. This was done using CMG’s CMOST tool (CMG Ltd.,

n.d.) to automate sensitivity analysis and history matching, as shown in Figure 11. This allowed us to develop comprehensive models that simulated varying conditions and scenarios, providing a deeper understanding of reservoir dynamics.

The automation of the sensitivity analysis process allowed us to identify, evaluate, and optimize the impact of specific reservoir parameters. This not only refined the accuracy of our scenario models, but also provided valuable insights into how subtle changes in parameters can have cascading effects on the overall production in the simulated reservoir systems.

Building on the MIP-3H well's base model, we explored the effects of different reservoir parameters. Using this foundational model, we conducted a series of parametric studies. These were designed to understand how specific tweaks could influence overall production. The subsequent section provides a breakdown of these studies, highlighting the significance of each parameter in shaping our predictions.

The model scenarios developed for parametric studies are summarized as follows:

1. Base Model

- i. A well-validated model developed using parameters consistent with MSEEL data and other published models for the same dataset.
- ii. Model parameters are shown in Table 4.

Table 4: NRS MIP-3H Base Model Parameters

MIP-3H: Base Model Reservoir Parameters		
Reservoir Parameters (MIP-3H)	Base Model Parameter Values	Units
Model Dimensions (MIP-3H)	7000 (Length)×1500(Width) ×90(Height)	ft.
Well Length (Horizontal)	6350	ft.
Initial Reservoir Pressure	4800	psia.
Initial Fissure Porosity	0.1	percent
Initial Matrix Porosity	4	percent
Initial Fissure Permeability i, j, k	7000, 7000, 700	nd
Initial Matrix Permeability i, j, k	1750, 1750, 175	nd
Number of Hydraulic Fractures	128	-
Fracture Half-Length, Xf	300	ft.
Initial Fracture Conductivity, kf wf	6	md-ft
Water Saturation	0.15	Fraction
Rock Density	120	lb/ft3
Langmuir Pressure Constant	0.00036	psi-1
Langmuir Volume Constant	0.05	g-mol/lb

2. Parametric Study: Number of (Producing) Hydraulic Fracture Perforations & Fracture Half Length

- i. Base model with dynamic modifications made to the parameters for producing hydraulic fracture perforations and fracture half-length.
- ii. Model parameters and modifications are shown in Table 5.

Table 5: NRS MIP-3H Parametric Study: Number of (Producing) Hydraulic Fracture Perforations & Fracture Half Length

Reservoir Parameters (MIP-3H)	Modified Scenario Parameter Values	Units
Model Dimensions (MIP-3H)	7000 (L)×1500(W) ×90(H)	ft.
Well Length (Horizontal)	6350	ft.
Initial Reservoir Pressure	4800	psia.
Initial Fissure Porosity	0.1	percent
Initial Matrix Porosity	4	percent
Initial Fissure Permeability i, j, k	7000, 7000, 700	nd
Initial Matrix Permeability i, j, k	1750, 1750, 175	nd
Number of Hydraulic Fractures	Parametric study	-
Fracture Half-Length, Xf	Parametric study	ft.
Initial Frac Conductivity, kf wf	6	md-ft
Water Saturation	0.15	Fraction
a Rock Density	120	lb/ft3
Langmuir Pressure Constant	0.00036	psi-1
Langmuir Volume Constant	0.05	g-mol/lb

For the parametric study on producing hydraulic fracture perforations, we grounded our modifications in both empirical observations and validated research. Conversations with industry professionals highlighted real-world challenges, including perforation blockages caused by debris, scale, and displaced proppant post-fracturing.

Several studies confirm these insights. (C. Cipolla et al., 2011) and (Miller et al., 2011) observed production variability along the horizontal length of wells, noting that only about 60-64% of perforation clusters significantly contribute to production. Further research by (Chorn et al., 2014), (Spain et al., 2015), (Slocombe et al., 2013), and (Ugueto C. et al., 2016) shows that many of these clusters either produce minimally or aren't adequately stimulated, underscoring the production variability in horizontal wells.

In the case of the MIP-3H well, findings from (Aboaba, 2022) and (Carr et al., 2015) reveal that approximately 44% of perforation clusters did not produce any gas. This combination of industry feedback and supportive studies informs our research approach and inclusion of this parameter.

Hydraulic fracture half-length is similarly difficult to accurately measure in a horizontal well due to the complex geology and challenges of obtaining reliable data (Chen et al., 2022). In shale formations, several factors contribute to this uncertainty, including the geo-mechanical heterogeneity of the formation, variability in the hydraulic fracturing process, and the subsequent interactions between the created fractures (Cook et al., 2014). Various methods, ranging from flow pattern evaluations to sophisticated hydraulic fracturing diagnostics like micro-seismic monitoring (Fisher et al., 2004), are employed to assess fracture geometry. However, these methods often come with their own sets of uncertainties and challenges, necessitating multiple sensitivity tests and engineering data analysis to make informed decisions and optimize strategies.

3. Parametric Study: Matrix Porosity and Fracture Conductivity

- i. Base model with modifications made to the parameters for matrix Porosity and fracture conductivity.
- ii. Model parameters and modifications are shown in Table 6.

Table 6: NRS MIP-3H Parametric Study: Matrix Porosity and Fracture Conductivity

Reservoir Parameters (MIP-3H)	Modified Scenario Parameter Values	Units
Model Dimensions (MIP-3H)	7000 (L)×1500(W) ×90(H)	ft.
Well Length (Horizontal)	6350	ft.
Initial Reservoir Pressure	4800	psia.
Initial Fissure Porosity	0.1	percent
Initial Matrix Porosity	Parametric Study	percent
Initial Fissure Permeability i, j, k	7000, 7000, 700	nd
Initial Matrix Permeability i, j, k	1750, 1750, 175	nd
Number of Hydraulic Fractures	128	-
Fracture Half-Length, Xf	300	ft.
Initial Frac Conductivity, kf wf	Parametric Study	md-ft
Water Saturation	0.15	Fraction
Rock Density	120	lb/ft ³
Langmuir Pressure Constant	0.00036	psi-1
Langmuir Volume Constant	0.05	g-mol/lb

Within this parametric study scenario, we focused our investigation on the effects of matrix porosity and fracture conductivity, both of which play a pivotal role in modelling shale reservoirs and forecasting their productivity.

The accurate estimation of matrix porosity is important to determine the potential volume of oil, gas, or water that can be extracted from a reservoir. While various methods such as acoustic logging, laboratory analysis, direct determination, nuclear magnetic resonance (NMR), and empirical formulas exist for this purpose (Kazatchenko et al., 2003), challenges persist. Shale reservoirs are inherently heterogeneous, making the representation of microscopic phenomena increasingly limited as observation scales increase. Laboratory measurements of core samples, for instance, may not be fully representative due to variations in porosity throughout the reservoir (Keelan, 1982). Factors such as the minute size of shale pores (often less than 1 micron), the presence of organic matter that could occupy these pores, and stress-induced pore closure further complicate accurate matrix porosity measurements (Qian et al., 2022).

Similarly, assessing fracture conductivity, essential in model development for determining the production rate from hydraulically fractured wells, presents its own set of challenges. Influenced by multiple factors, including the type of proppant used during fracking operations, fracture width, and closure stress (Davies & Kuiper, 1988), it's further complicated by the intricate fracture networks in shale reservoirs. Variations in conductivity over time, due to proppant settling and the migration of fines, add another layer of complexity (Zheng et al., 2020).

Given these intricacies and the impact of minor parameter variations on model accuracy, it was vital to include these in our study. Interpreting these results in the context of broader geological and reservoir features is essential to refine our model assumptions, as we aim to understand how subtle changes in matrix porosity and fracture conductivity can significantly alter model predictions.

4. Parametric Study: Skin Effect

- i. Base model with Skin dynamically applied.
- ii. Model parameters and modifications are shown in Table 7.

Table 7: NRS MIP-3H Parametric Study: Skin Effect

Reservoir Parameters (MIP-3H)	Modified Scenario Parameter Values	Units
Model Dimensions (MIP-3H)	7000 (L)×1500(W) ×90(H)	ft.
Well Length (Horizontal)	6350	ft.
Initial Reservoir Pressure	4800	psia.
Initial Fissure Porosity	0.1	percent
Initial Matrix Porosity	4	percent
Initial Fissure Permeability i, j, k	7000, 7000, 700	nd
Initial Matrix Permeability i, j, k	1750, 1750, 175	nd
Number of Hydraulic Fractures	128	-
Fracture Half-Length, Xf	300	ft.
Initial Fracture Conductivity, kf wf	6	md-ft
Water Saturation	0.15	Fraction
Rock Density	120	lb/ft3
Langmuir Pressure Constant	0.00036	psi-1
Langmuir Volume Constant	0.05	g-mol/lb
* Skin	Parametric Study	-

Skin is a dimensionless parameter used to measure a well's production efficiency by comparing its actual conditions to theoretical or ideal ones. The term "Skin Effect" refers to the deviation in the pressure drop from what's predicted with Darcy's law using the value of permeability thickness (kh), due to the skin's influence (SLB Energy Glossary, 2023a). Simply put, a positive skin value suggests some damage or influences that are impairing productivity, while a negative value indicates enhanced performance.

Our parametric study incorporated skin into numerical models to offer deeper insights. Over time, many phenomena near the wellbore can impact flow and productivity. These include the closure of fractures, the migration of particles through fractures, the buildup of scale or wax near the perforation zones, or the slow degradation of the formation's integrity around the wellbore (Raji et al., 2020). Given the complexities of these processes, they can be difficult to model analytically.

To address this complexity, modeling software like CMG™ and IHS Harmony™ utilize skin or "Time-Dependent Skin" to consolidate the effects of these phenomena into a single term at a discrete point,

independent of geometrical or fluid-flow complexities. IHS™ effectively summarizes this concept within their reference materials, stating that “skin should be considered a “tuning parameter” during the history-matching phase of numerical modeling, intended to be a “catch-all” for any fluid or reservoir property changes over time that are not accounted for by the analytical model.” (IHS Harmony, 2020)

Incorporating skin and skin effect into our parametric study highlighted how certain modeling parameters can serve as a 'catch-all' for representing complex phenomena in NRS models. This underscores the significant influence of model parameters that might not be grounded in concrete data.

3.2.4 Model Performance and Validation

To assess the performance of the numerical reservoir simulation model scenarios and measure the accuracy of the predicted outcomes, we utilized two key metrics: the Mean Absolute Percentage Error (MAPE) and the Mean Absolute Error (MAE). The MAE provided a direct measure of the average absolute differences between the predicted and actual values. The MAPE, on the other hand, provided a clear percentage-based understanding of prediction accuracy, measuring the average magnitude of error produced by a model.

Equation 1: Mean Absolute Percentage Error (MAPE)

$$MAPE = \frac{1}{n} \sum \left(\frac{|A_t - F_t|}{A_t} \right) \times 100$$

Equation 2: Mean Absolute Error (MAE)

$$MAE = \frac{1}{n} \sum |A_t - F_t|$$

Where:

- n is the number of data points
- A_t is the actual value, and
- F_t is the forecasted value

In simpler terms, MAPE is easy to understand and interpret, but can be sensitive to outliers. MAE is not as sensitive to outliers but is also not as sensitive to changes in trend.

These metrics show us, on average, how much our predictions deviated from the actual production values, regardless of direction (over or under-predicting). Both metrics were crucial as they offered different perspectives on the model's performance. Observations of a lower value in both MAPE and MAE indicated a closer match between our model's predictions and the actual historic production data for the MIP 3H well, implying a more reliable model.

Additional metrics used to compare the model performance included CPU specs, such as the number of processors used when running the simulation model, and the time for each model to run. Furthermore, an estimate of the resources required for model development was recorded, such as time to collect and process the MSEEL data, initial model preparation, and scenario investigations to identify "soft" parameter selection for scenario development were to compare the NRS process with the AI approach.

3.3 Methodology: Data-Driven Artificial Intelligence Model

The methodology for the development and analysis of the Data-Driven AI model in this study will be structured as follows:

Methodology for Artificial Intelligence (AI) techniques

1. Data Collection and Preparation
 - Study area: dataset encompassing 400 Marcellus Shale Wells
2. Data Exploration and Analysis
 - Basic Statistical Analysis
 - Well Quality Analysis (WQA)
 - Identify Key Performance Indicators
3. Model Development
 - Feature selection
 - Data partitioning and database construction
 - Develop predictive AI model for 30-day rich gas production
 - i. IMprove™ software – backpropagation neural network
 - ii. Model Training
4. Model Validation and Predictive Analysis
 - Perform blind validation of the predictive neural network
 - Input selection and hyperparameter tuning to improve model accuracy

A high-level illustration of the workflow is seen Figure 12.

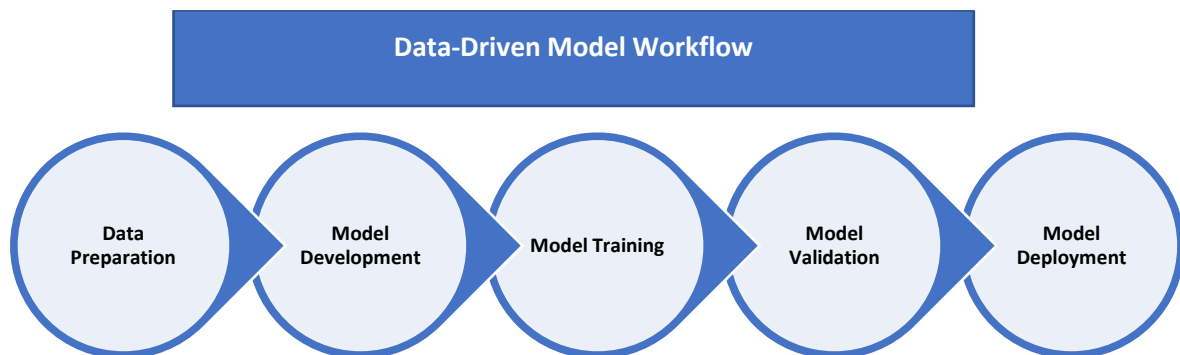


Figure 12: Data Driven AI Model Workflow

3.3.1 Data Collection and Preparation: AI Model Study Area

The AI model in our study utilized a comprehensive dataset derived from 400 Marcellus shale wells. This dataset comprises of field-measured parameters coupled with production records for 30-day, 12-month, and 24-month periods, encompassing both gas and water production outputs.

The dataset consists of 46 parameters: 40 are input variables (also known as features or predictors), which represent the characteristics or attributes of the data used for predictions or actions. The remaining 6 are output variables, signifying the desired outcomes or predictions (also known as targets) that the model is trained to produce.

It should be noted that while most attributes have a consistent record count of 400, there is variability in the output parameters. This variance is due to differences in data collection intervals and the individual production lifespans of the wells. Upon examining the output variables (production records), the 30-day production data offers a complete set of 400 records, whereas the 12-month and 24-month data contain 257 and 153 records, respectively.

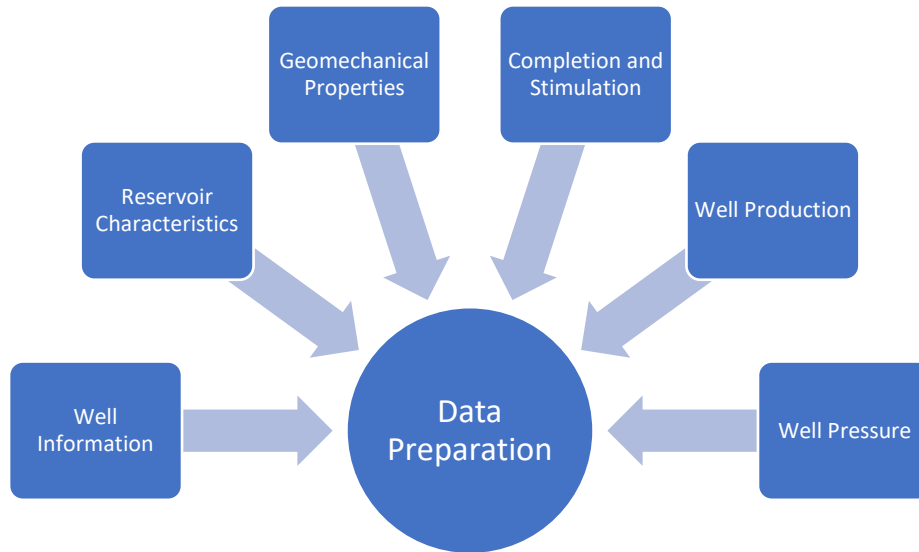


Figure 13: Data Collection and Preparation

To make the data analysis more intuitive, all attribute variables were categorized. The detailed classifications for the input variables are presented in Table 8, organized by category and number of available records. The output variables for this study are the production records previously discussed, shown in Table 9.

Table 8: Input Attribute Summary for AI Model Development

Category	Input Variable	# Records
Well	X	400
Well	Y	400
Well	Measured Depth	400
Well	True Vertical Depth	400
Well	Deviation Type	400
Well	Inclination	400
Well	Azimuth	400
Hydraulic Fracture	Avg. Inj. Pressure (psi)	400
Hydraulic Fracture	Avg. Max. Pressure	400
Hydraulic Fracture	Avg. Inj. Rate (bbl/min)	400
Hydraulic Fracture	Avg. Max. Rate	400
Hydraulic Fracture	Slurry Volume (bbl)	400
Hydraulic Fracture	Clean Volume (bbl)	400
Hydraulic Fracture	Max. Prop. Concentration (lb/gal)	400
Hydraulic Fracture	Prop./Stage(lb)	400
Hydraulic Fracture	Total Prop. Inj. (lb)	400
Hydraulic Fracture	Avg. Breakdown Pressure	400
Hydraulic Fracture	Avg. Breakdown Rate	400
Formation	BTU Area	400
Formation	Avg. Frac. Gradient	400
Formation	Porosity(%)	400
Formation	Net Thickness(ft)	400
Formation	Swi (%)	400
Formation	TOC (%)	400
Formation	Bulk Modulus	400
Formation	Shear Modulus	400
Formation	Youngs Modulus	400
Formation	Poisson's Ratio	400
Formation	Min Horizontal Stress	400
Formation	Avg. ISIP	400
Completion Design	Completions Date	400
Completion Design	Stimulated Lateral Length(ft)	400
Completion Design	Shot Density (Shots/ft)	400
Completion Design	Total Clusters	400
Completion Design	Total No. Stages	400
Completion Design	Cluster Spacing	400
Operational	Start Production Date	400
Operational	Wellhead Pressure - 30 days-PSI	400
Operational	Wellhead Pressure - 12 Months-PSI	257
Operational	Wellhead Pressure - 24 Months-PSI	153

Table 9: Output Attribute Summary for AI Model Development

Output Variable	# Records
Rich Gas-30 days-MCF	400
Rich Gas-12 Months-MCF	257
Rich Gas-24 Months-MCF	153
Water 30 days-BBL	400
Water 12 Months-BBL	257
Water 24 Months-BBL	153

To safeguard proprietary data, the exact names and locations of these wells have been anonymized, yet their inherent patterns and relationships remain intact. This allows for trend and pattern analysis without exposing sensitive information. Figure 14 shows the relative positions of the 400 wells within the dataset. All other dataset attributes are unchanged from their originally measured values.

A more robust dataset is less susceptible to the influence of outliers, which can distort the results of machine learning algorithms. Moreover, a larger dataset allows the model to better discern complex relationships between data points, which is critical as many real-world challenges involve intricate interconnections. Given these considerations, we will opt for the output variable with the highest record count (30-day production) to provide the machine learning algorithm with the most robust dataset possible.

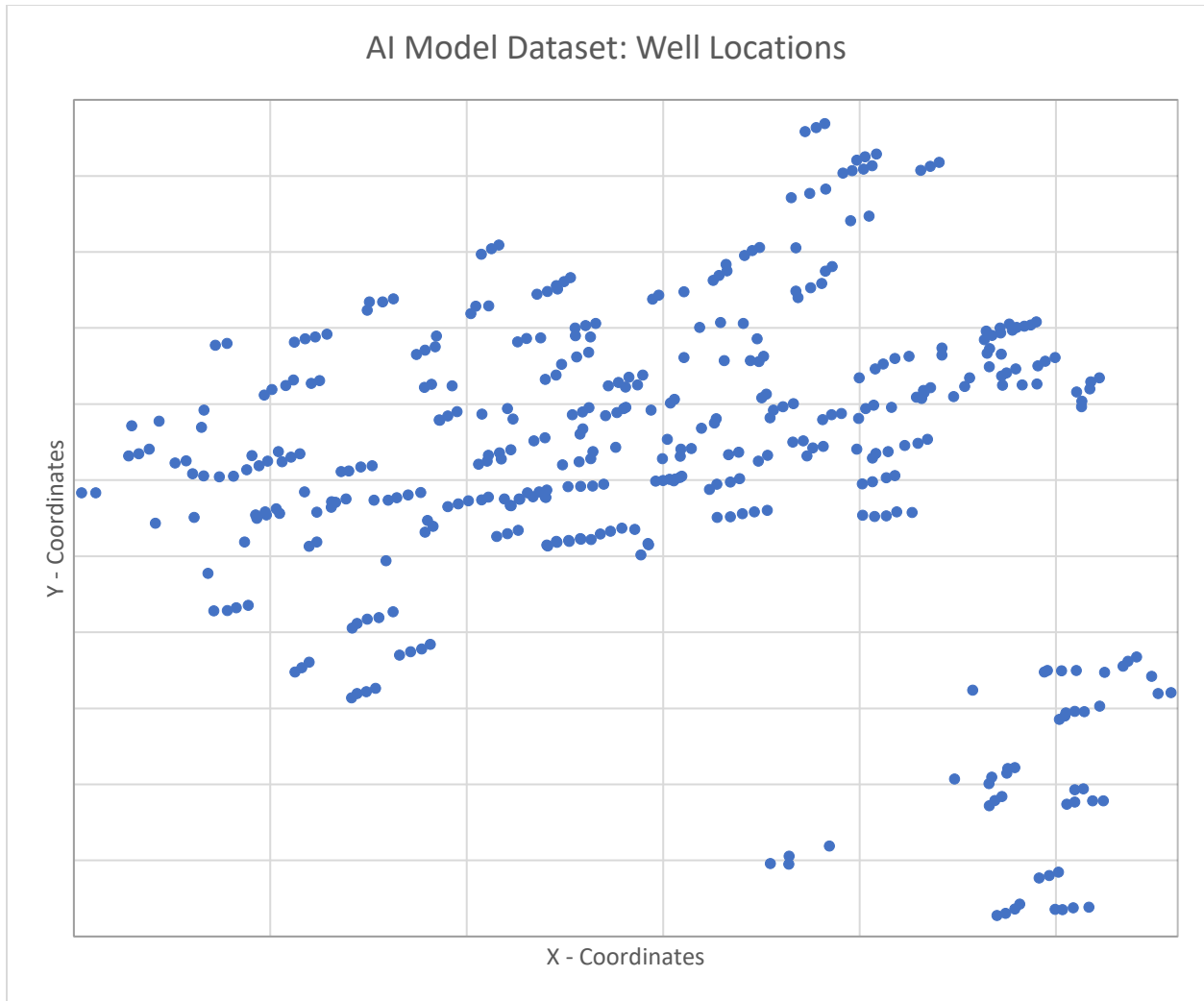


Figure 14: Normalized Marcellus Shale Well Locations

3.3.2 Data Exploration and Analysis: Marcellus Shale Well Dataset

Data exploration and analysis was performed using several techniques within IMPROVE™ software. Our objective in this step was to thoroughly explore the dataset and explore our understanding of the various attributes involved, given the objective of predicting production on a well-by-well basis using a data-driven AI model. The following descriptive analytics will serve to evaluate the data's quality while highlighting promising candidates and relevant information for subsequent analysis and model development.

3.3.2.1 Basic Statistical Analysis

We began our data exploration process by performing basic statistical analysis to investigate the relationship between the dataset's attributes. The primary goal of this analysis was to gain a deeper understanding of the data before transitioning into the development of our AI data-driven models.

This approach allows us to:

1. **Develop Initial Insights:** Visualization tools, such as the bubble plot depicted in Figure 15, enhance our understanding by making patterns and correlations more intuitive, highlighting behavior that might be missed when viewing the data in its raw format. In this plot, the size of each bubble represents the cumulative gas production value at their respective X and Y-coordinate for each well. This visual representation offers a clear view of how production behavior differs across various locations and their relative proximities, suggesting that the formation reservoir plays a significant role.
2. **Understand the distribution of the data:** By observing parameters like the mean, median, range, and standard deviation, we gained insights into the characteristics of each variable, identifying if it is normally distributed, skewed, or exhibiting other tendencies.
3. **Identify relationship between variables:** Through regression analysis, we can measure the relationship between one or more independent (predictor) variables and a dependent (outcome) variable. For our study, the main interest lies in understanding the connection between available input variables and the 30-day gas production, which we defined as our output variable. Different patterns and behaviors observed for select variable examples are seen in Figure 16 - Figure 19.
4. **Spot and Scrutinize Outliers:** Outliers and anomalies in the data can compromise the accuracy and reliability of subsequent models. By identifying data points that significantly deviate from the rest, or are not consistent with the expected behavior, we can assess their legitimacy and decide whether to remove them or treat them as special cases.
5. **Address Potential Data Problems:** Basic statistical analysis can highlight issues such as missing values or errors that might compromise the integrity of our data.

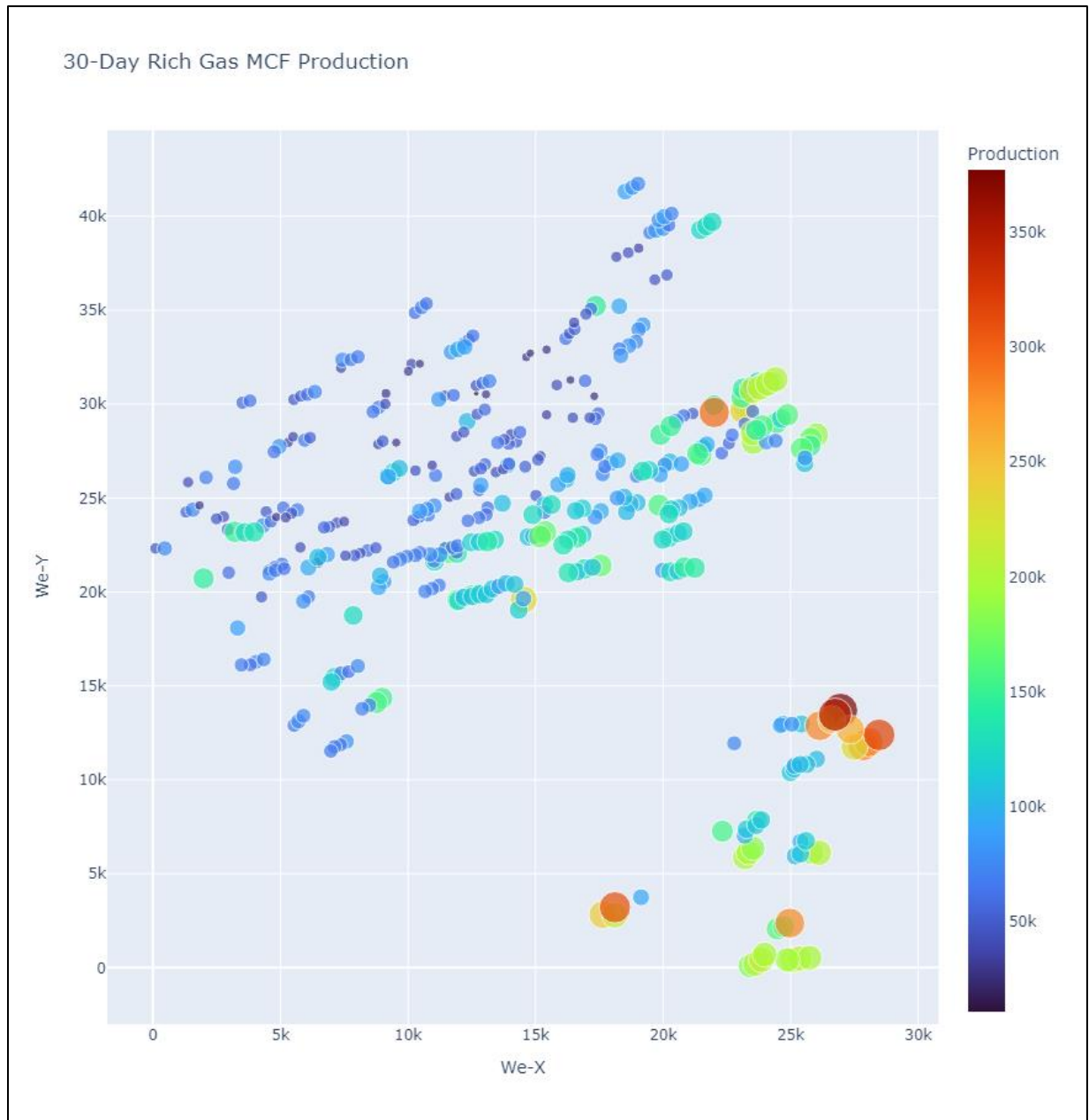


Figure 15: Basic Statistical Analysis - Bubble Plot (30-Day Rich Gas MCF Production)

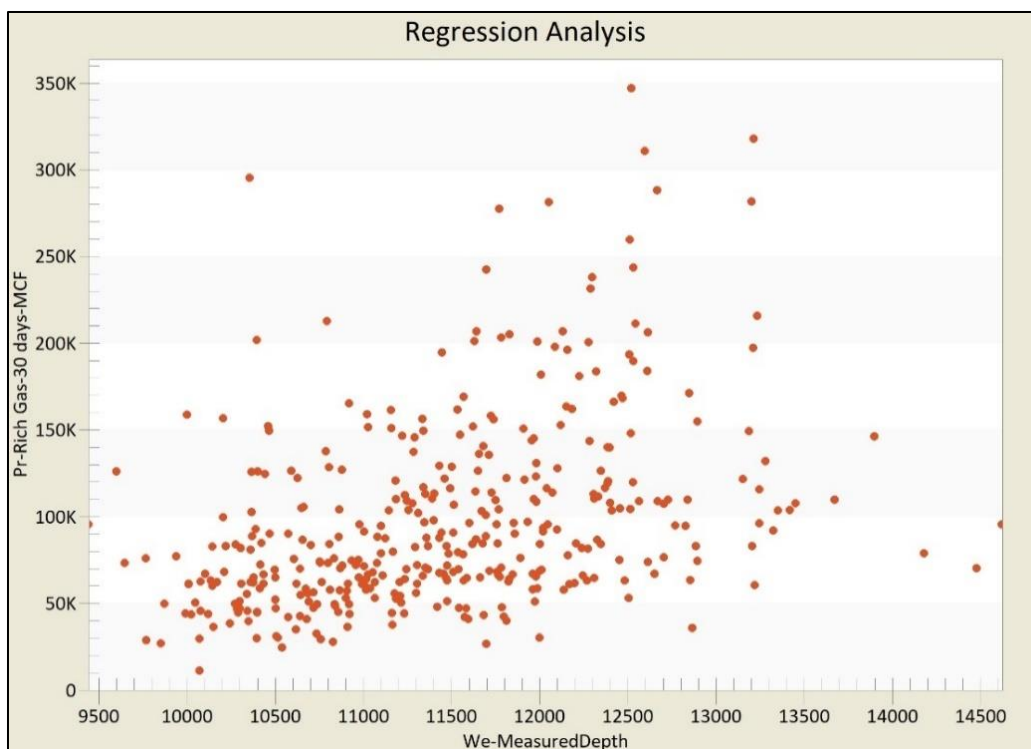


Figure 16: Regression Analysis - Measured Depth vs 30-day Rich Gas Production

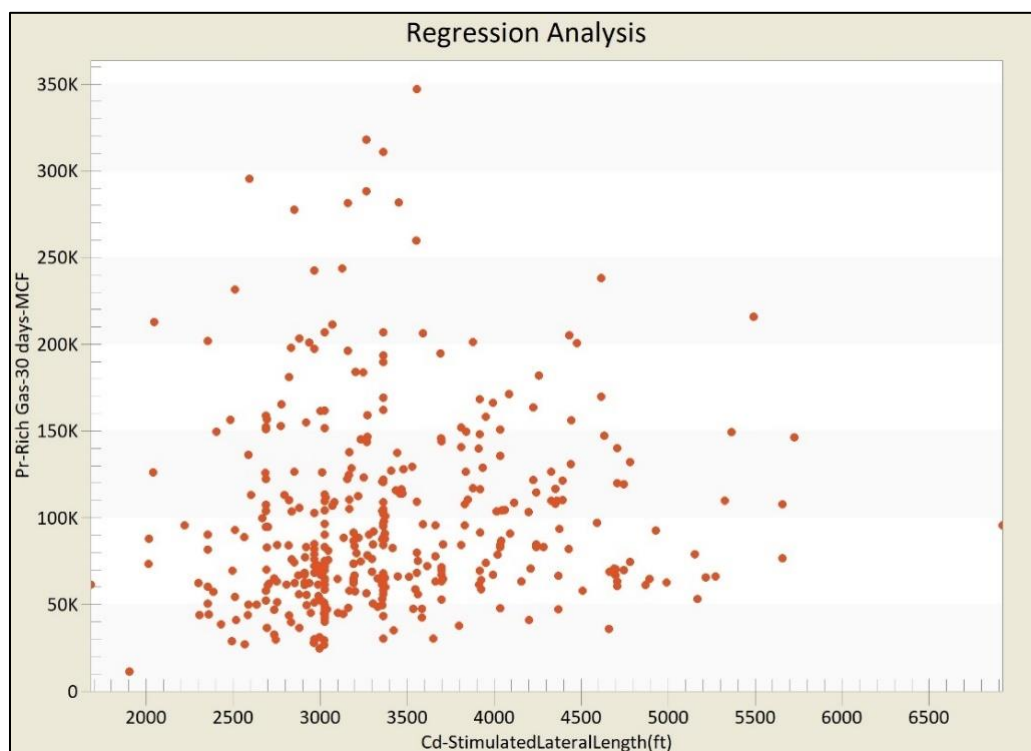


Figure 17: Regression Analysis - Stimulated Lateral Length vs 30-day Rich Gas Production

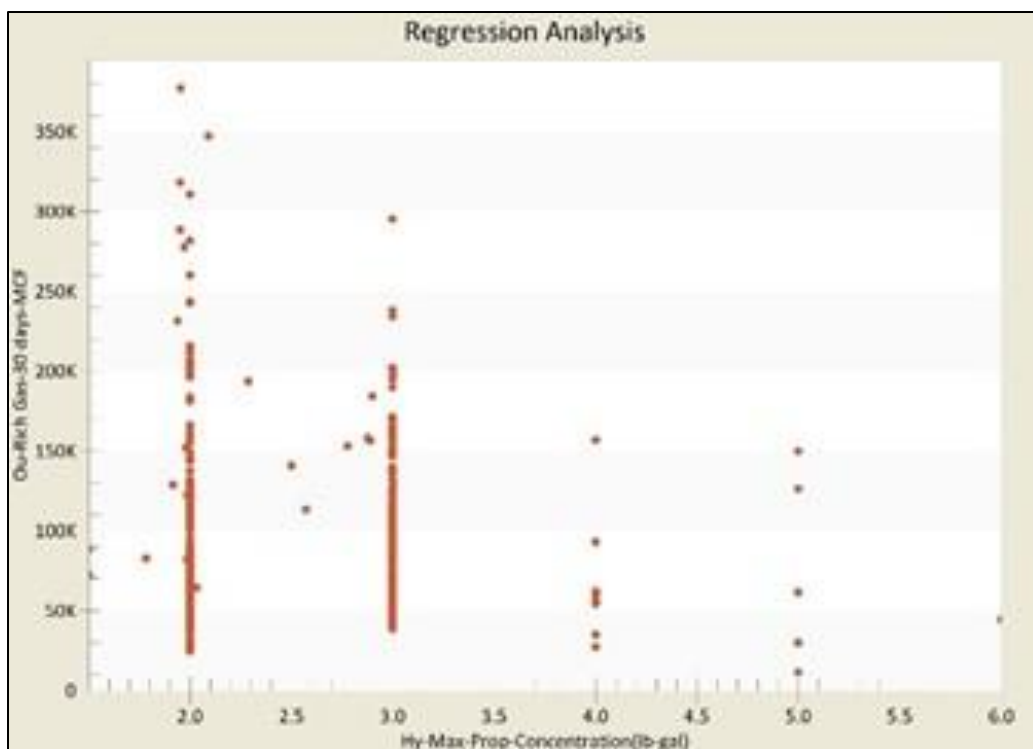


Figure 18: Regression Analysis - Max Proppant Concentration vs 30-day Rich Gas Production

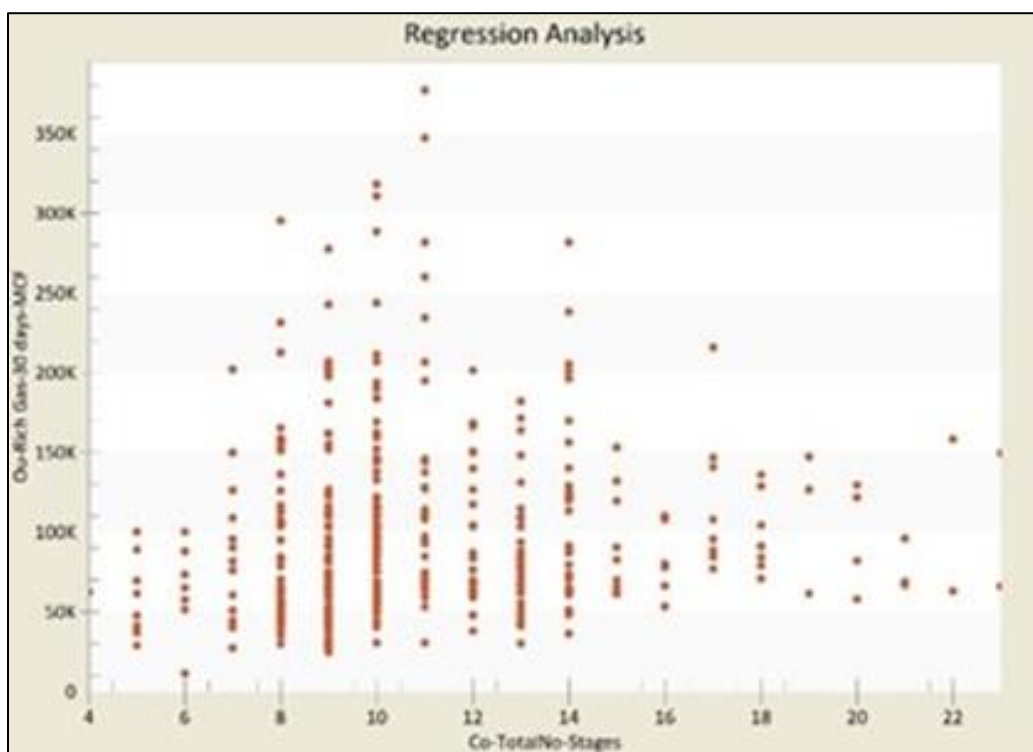


Figure 19: Regression Analysis - Total Number of Stages vs 30-day Rich Gas Production

3.3.2.2 Well Quality Analysis (WQA)

IMPROVE™ software was used to perform well quality analysis (WQA). This method of using fuzzy set analysis on the input data, based on the 30-day gas production output, allowed us to group the wells into categories by a soft criterion. The input data can then be classified by not just a hard category but also the value of membership to the category.

The WQA breaks the output variable into distinct categories of “Poor”, “Average”, and “Good” production for 3 sets, providing a more nuanced view of the data. Figure 20 and Figure 21 below depict the WQA preparation for multiple example scenarios of 3 and 4 fuzzy sets. The example seen in Figure 22 and Figure 23 for well *PD-A1-H1* demonstrates that a 3 fuzzy set scenario categorizes the well “Average” production, but when the 4 fuzzy set categories are applied, we can observe the well to have membership to both “Average” (36%) and “Good” (64%) categories.

Select Output :

Pr-Rich Gas-30 days-MCF

Number of Sets

3

Default Table

Data Range

Minimum

11044

Maximum

376993

Fuzzy Set	Rise	Top	Top/Fall	Fall
Poor	11044	11044	54234	127424
Average	54234	127424	200613	273803
Good	200613	273803	376993	376993

Pr-Rich Gas-30 days-MCF	Poor	Average	Good	Total Cases = 400
# Cases per Set	313	319	26	658
% Cases per Set	78.2	79.8	6.5	164.5

Figure 20: WQA Preparation - 3 Fuzzy Sets

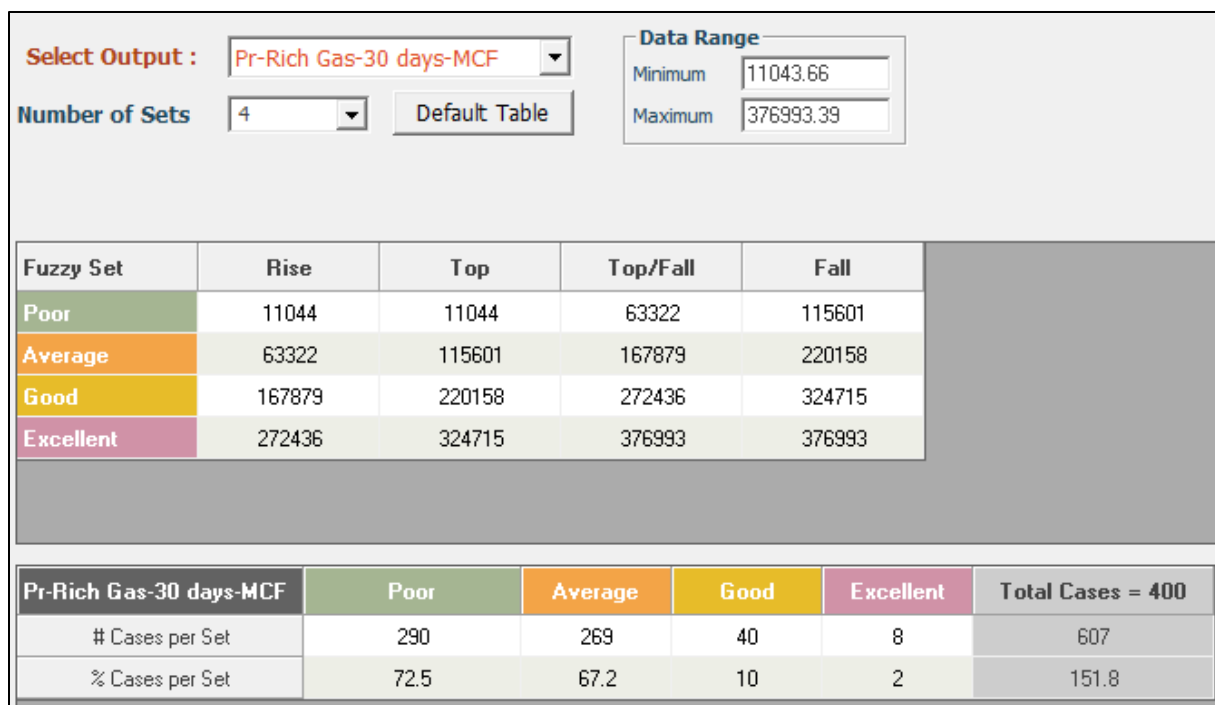


Figure 21: WQA Preparation - 4 Fuzzy Sets

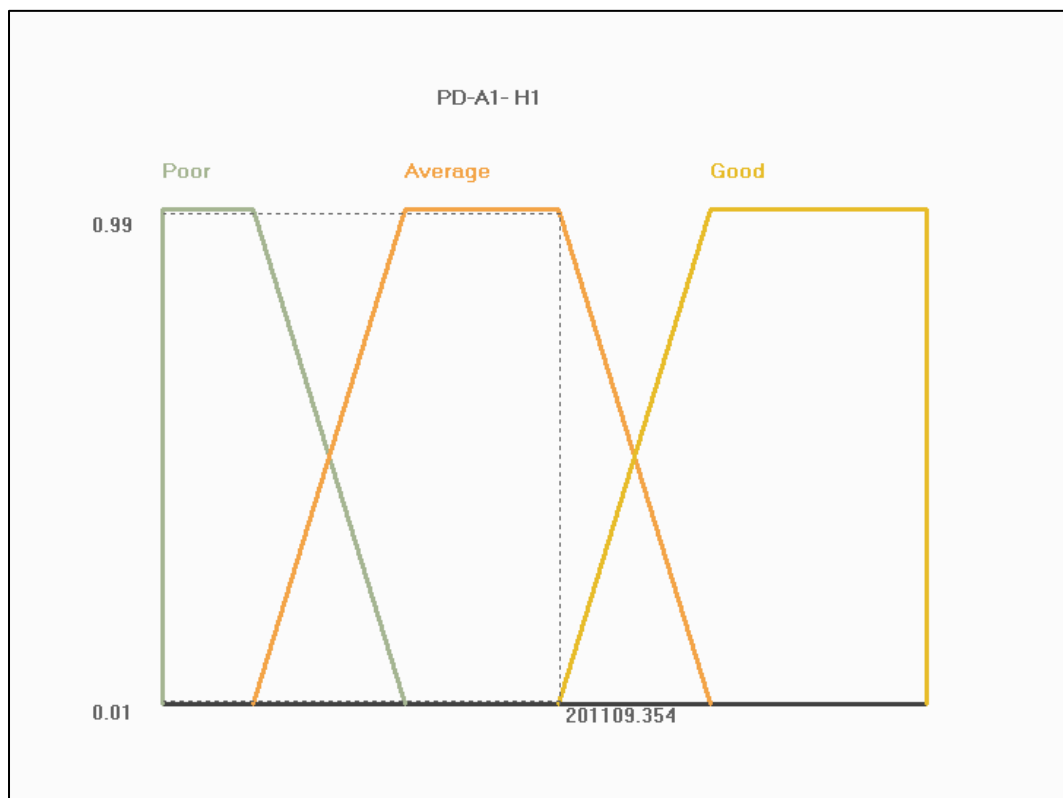


Figure 22: WQA - 3 Fuzzy Sets for PD-A1-H1 Well

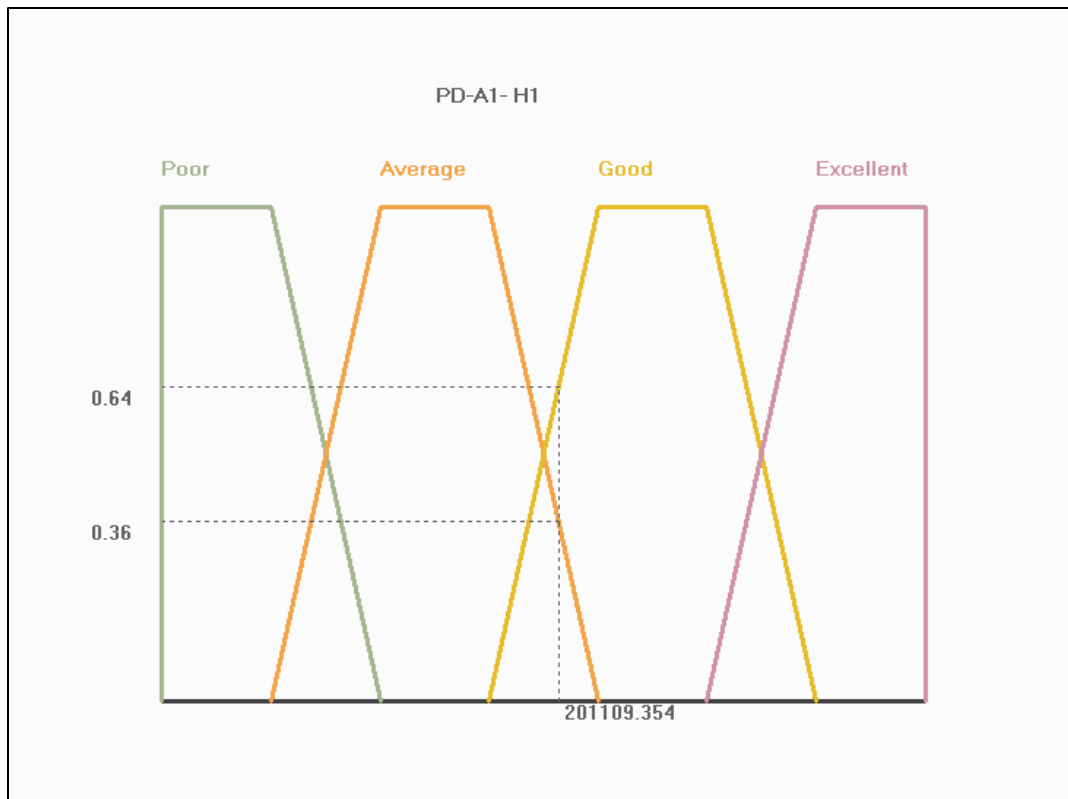


Figure 23: WQA - 4 Fuzzy sets for PD-A1-H1 Well

Plotting well attributes of X, Y coordinates classified by fuzzy sets visualizes the distribution of well quality across the dataset. As Figure 24 shows, gas production quality appears to trend from left (poor) to bottom right (good), suggesting that the formation reservoir plays a significant role. Despite this visible trend, inconsistencies in some production areas mean we can't depend entirely on these factors. Still, this analysis provides more context and a deeper understanding of the dataset's features, setting the stage for a more thorough investigation.

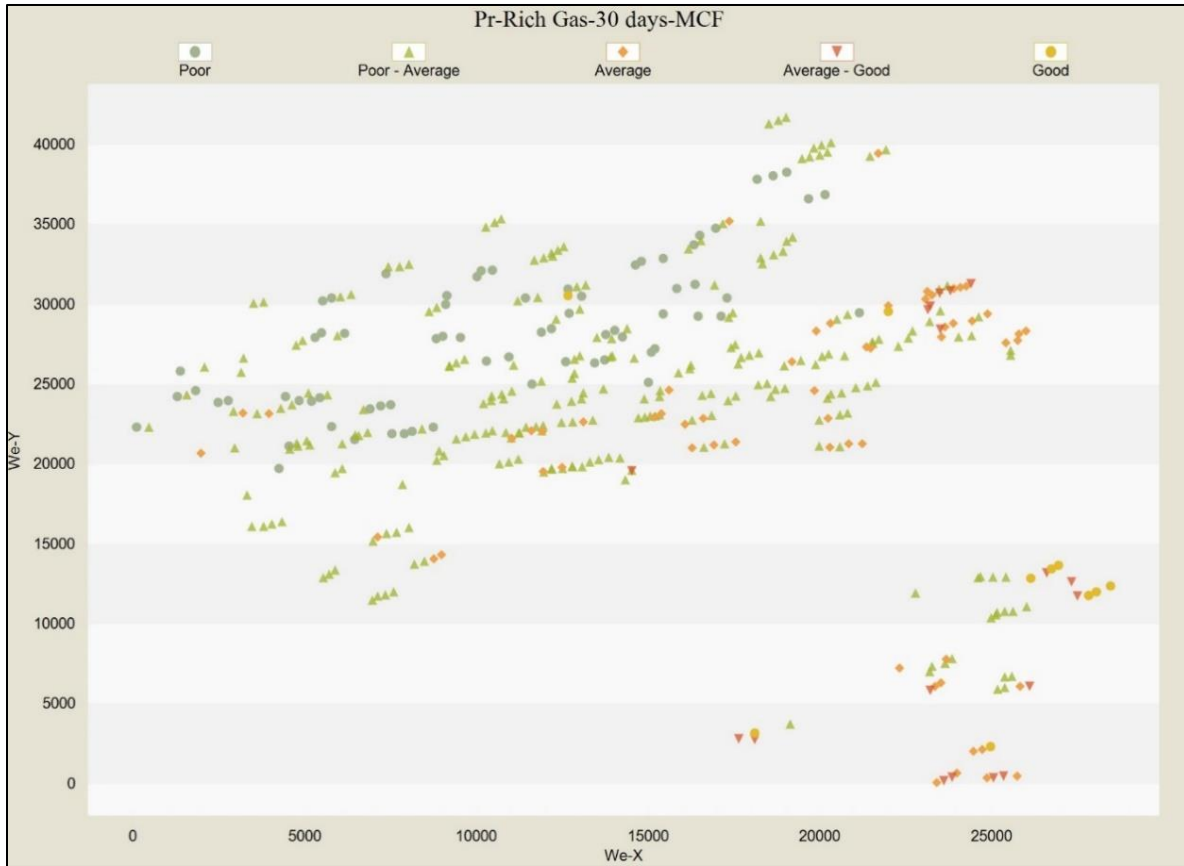


Figure 24: WQA - Fuzzy Categories for All Wells (X, Y Plane)

3.3.2.3 Key Performance Indicators (KPI)

Identification of key performance indicators (KPI) is an important task in understanding any system's behavior. Using IMPROVE™ software's propriety fuzzy pattern recognition technology (Intelligent Solutions, 2018), we identify the contribution of parameters to the process outcome.

Fuzzy Pattern Recognition uses the interconnection of parameters in complex, non-linear and dynamic processes. Each parameter affects and is affected by others, altering the system's behavior. For effective analysis, it's vital to examine these interactions collectively and identify key drivers in an integrated or combined approach.

A summary of this study's attribute KPI values and their degree of influence on the process outcome are shown in Table 10.

Table 10: Key Performance Indicators (Attributes Ranking and Degree of Influence)

Rank	Attribute	% Degree of Influence
1	Fr-NetThickness(ft)	100
2	Fr-BTU Area	99
3	Fr-Avg- ISIP	95
4	We-X	92
5	Fr-TOC(%)	88
6	Fr-Youngs Modulus	76
7	We-True Vertical Depth	72
8	Op-WPH-30 days-PSI	64
9	We-Y	57
10	Fr-Swi(%)	51
11	We-Measured Depth	44
12	Fr-Bulk Modulus	38
13	Fr-Avg-Frac-Gradient	36
14	Fr-Shear Modulus	32
15	Fr-Porosity(%)	30
16	Hf-Avg-Inj-Pressure(psi)	19
17	Cd-Completions Date	18
18	Op-Start Production Date	14
19	Hf-Avg-Max-Pressure	13
20	Fr-Min Horizontal Stress	10
21	Cd-Shot Density (Shots-ft)	9
22	We-Azimuth	7
23	Hf-Slurry Volume(bbl)	7
24	We-Inclination	7
25	Hf-Avg-Inj-Rate(bbl-min)	6
26	Hf-Avg-Max-Rate	6
27	Cd-TotalClusters	6
28	Cd-TotalNo-Stages	5
29	Fr-Poisson's Ratio	5
30	Cd-Cluster Spacing	5
31	Hf-Avg-Breakdown Rate	5
32	Cd-Stimulated Lateral Length(ft)	4
33	Hf-Avg-Breakdown Pressure	2
34	Hf-Total Prop- Inj- (lb)	2
35	We-Deviation Type	2
36	Hf-Max-Prop-Concentration(lb-gal)	2
37	Hf-Clean Volume(bbl)	1
38	Hf-Prop--Stage(lb)	1

For a deeper understanding of the KPIs behavior, the IMPROVE™ software utilizes Fuzzy Pattern Recognition (FPR) to visualize and inspect input parameter's individual impact on the output parameter. FPR is a technique used in pattern recognition and machine learning. It helps to identify key performance drivers in a process, and it provides more nuanced information by analyzing the effect of different ranges of each parameter on the output.

To understand the effect of each parameter on the output we consider the following points:

1. High FPR values:

- a. When the FPR of the output demonstrates high values within certain ranges, it indicates that the parameter has a significant positive effect on the output. Essentially, these ranges are where the parameter most positively influences the results.

2. Low FPR values:

- a. Conversely, when the FPR demonstrates low values in certain ranges, it indicates the parameter negatively influences the output. These are ranges that should typically be avoided, as the parameter in these ranges has a negative correlation effect on the output.

3. Slope in the FPR Plot:

- a. The slope of the graph – its steepness and direction – indicates the sensitivity of the output to changes in the parameter's value within those ranges. For example, a steep slope means a small change in the parameter leads to a significant change in the output, suggesting high sensitivity.
 - i. Positive Slope: when the graph's slope is positive it suggests an increase in the parameter's value will result in an increase in the output.
 - ii. Negative Slope: a negative slope means that an increase in the parameter's value results in a decrease in the output.

This analysis helps in fine-tuning parameters to get the most desirable output and gain insights about the behavior of the process under various conditions. The following Figure 25 and Figure 26 present examples of two parameters: Measured Depth (MD) shows a positive slope and indicates that FPR values increase as the depth increases. This relationship can be explained by the fact that a longer MD provides a larger area for gas production. Conversely, Initial Water Saturation (Swi) exhibits a negative slope, implying that lower Swi values correspond to higher FPR values. This observation aligns with the negative correlation between initial water saturation and gas production in a reservoir.

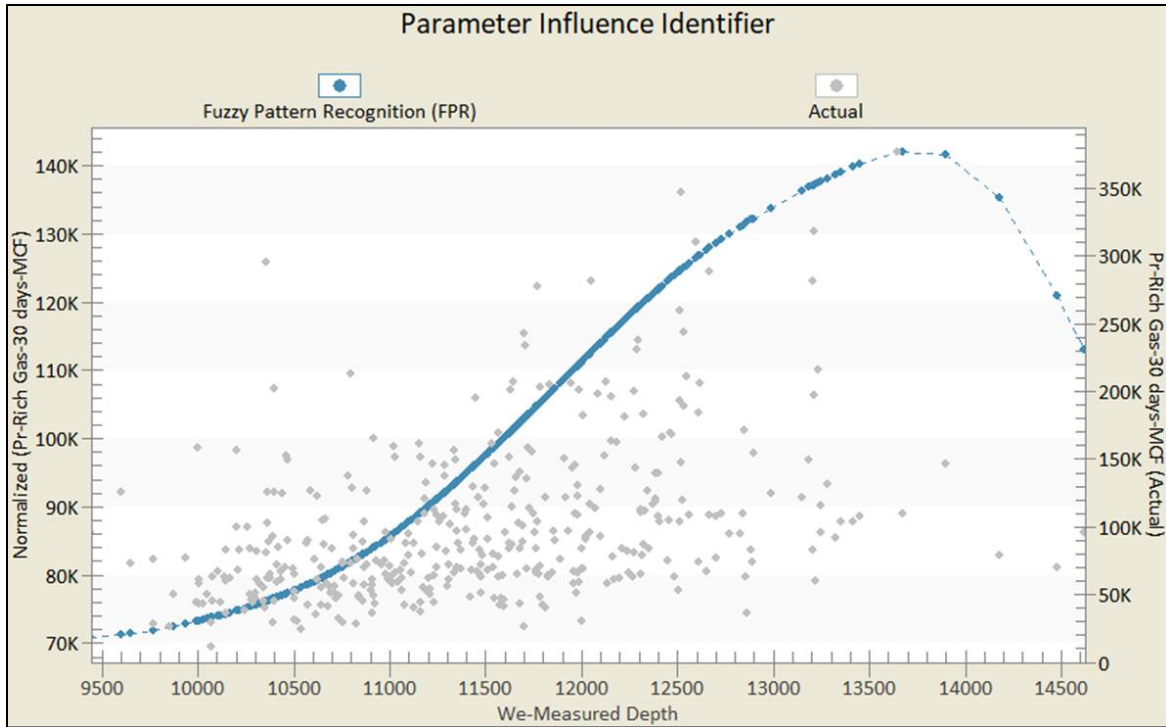


Figure 25: Fuzzy Pattern Recognition (FPR) - Measured Depth (MD)

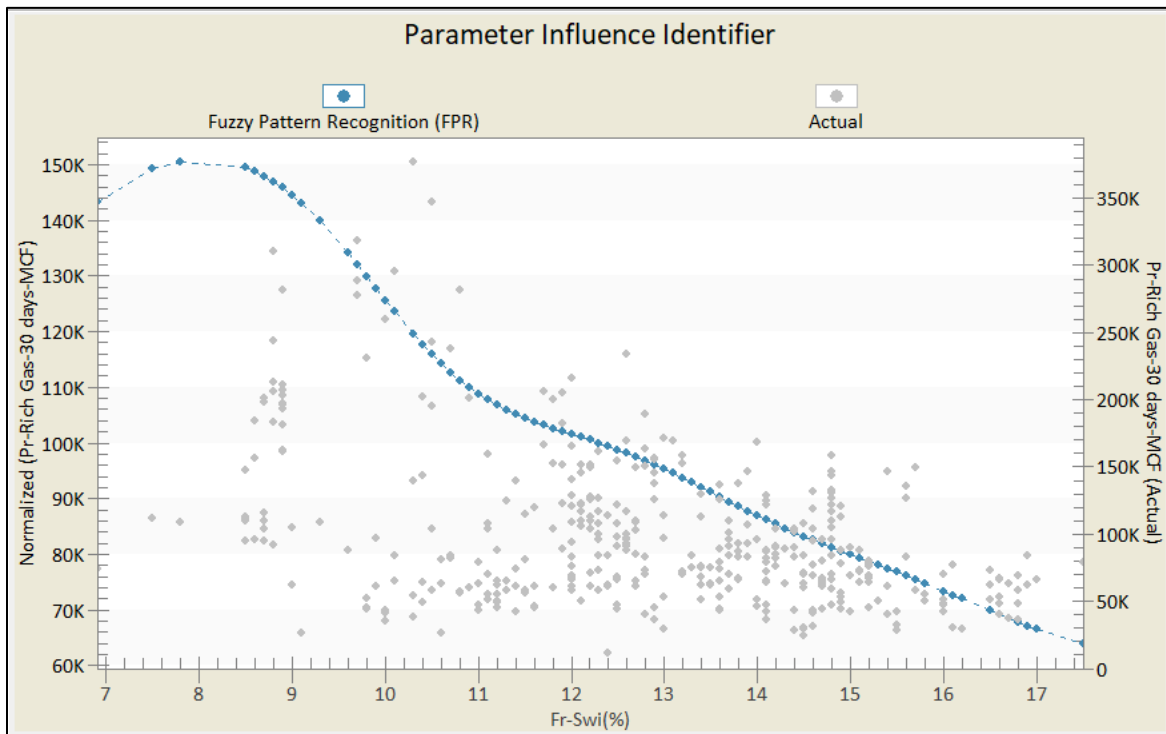


Figure 26: Fuzzy Pattern Recognition (FPR) - Initial Water Saturation (Swi)

3.3.3 Predictive Analytics and AI Model Development

The objective of this phase of the study is to create a predictive AI model using the insights gained during data exploration and analysis.

We will use a machine learning algorithm, specifically an artificial neural network, to study the measured well data. This approach goes beyond traditional statistical methods by allowing the model to learn from patterns in the data, resulting in more accurate predictions.

The Predictive Analytics workflow can be summarized as follows:

1. **Feature Selection:**

- a. Feature selection plays a pivotal role in predictive analytics by determining the input parameters that a neural network learns from. Various strategies, including processing and weighting these parameters, can optimize the learning process. This is often supplemented with domain expertise, both in the subject matter and the development of diverse neural network approaches. Initial input features are selected with careful examination of the impact of adding or removing specific features on the overall output.

2. **Model Development:**

- a. Model development includes tuning hyperparameters and partitioning datasets. Dataset partitioning follows a four-part breakdown: Training, calibration, verification, and blind validation. Different strategies in this step can greatly influence the neural network's performance. This project will use IMPROVE™ software for model development, negating the need for extensive mathematical background.

3. **Model Comparison:**

- a. This study will compare models based on the R2 score and Mean Absolute Error (MAE). This goal is to have a relatively even error score across all data partitions. This approach should ensure the model's robustness and ability to handle unseen data.

4. **Model Selection:**

- a. The final model will be selected based on the information gathered in the previous step. We will evaluate the performance of each model on a validation set of unseen "Blind" data. The model with the best performance will be selected.

3.3.3.1 Feature Selection

Feature selection is an important step in predictive analytics and model development. The input features represent the data that the network will learn from, so it is important to select the most relevant features.

In this research, we used a combination of methods to select the input features. As previously discussed, we first explored the data and performed descriptive analytics to identify the most important features. We then used our domain expertise to further refine the selection. Finally, we evaluated the impact of different feature combinations using a neural network approach.

An example of input selection within the IMPROVE™ software can be seen in Figure 27. The initial input features selected for our model, with associated KPI rank and degree of influence as determined during descriptive analytics, can be seen in Figure 28.

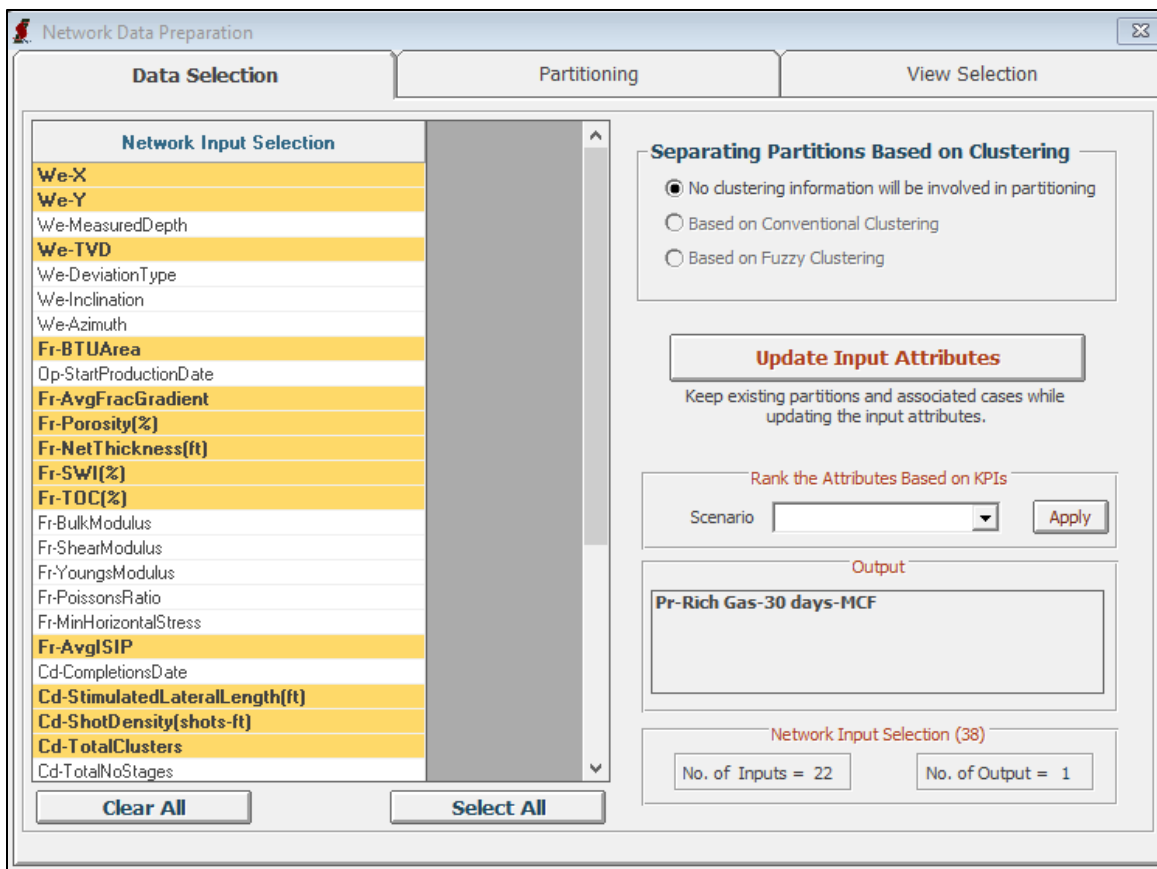


Figure 27: Neural Network Input Selection – IMPROVE™ Software

KPI Rank	Attribute	KPI % Degree of Influence
1	Fr-NetThickness(ft)	100
2	Fr-BTU Area	99
3	Fr-Avg- ISIP	95
4	We-X	92
5	Fr-TOC(%)	88
7	We-True Vertical Depth	72
8	Op-WPH-30 days-PSI	64
9	We-Y	57
10	Fr-Swi(%)	51
13	Fr-Avg-Frac-Gradient	36
15	Fr-Porosity(%)	30
16	Hf-Avg-Inj-Pressure(psi)	19
19	Hf-Avg-Max-Pressure	13
21	Cd-Shot Density (Shots-ft)	9
23	Hf-Slurry Volume(bbl)	7
25	Hf-Avg-Inj-Rate(bbl-min)	6
27	Cd-TotalClusters	6
30	Cd-Cluster Spacing	5
31	Hf-Avg-Breakdown Rate	5
32	Cd-Stimulated Lateral Length(ft)	4
33	Hf-Avg-Breakdown Pressure	2
34	Hf-Total Prop- Inj- (lb)	2

Figure 28: Input Selection for AI Model Development

3.3.3.2 Data Partitioning and Database Construction

After selecting our initial input features and defining 30-day gas production as our desired output, the next step in model development was to partition the data. The data was split into three partitions within the IMPROVE™ software: Training, Calibration, and Verification. This is a critical step, as the way the data is partitioned will often have a large impact on the success of the model.

The dataset partitions are summarized as follows:

- **Training dataset:** This is the most important part of the development process. The training set is part of the data shown to the ANNs during the training process, used to teach the machine learning algorithms patterns in the dataset by establishing correlations between the input and output features. This is where the model learns to make predictions.

- **Calibration dataset:** This data is not used directly during the training process or to adjust the outputs. The calibration data is used to monitor the performance of the neural network as it learns and to determine when to stop the training process.
- **Verification/Validation dataset:** This data is used to validate the performance of the trained model at the end of the training process. It is used to see if the trained neural network can perform reasonably well on out-of-sample data. This data has no bearing on the machine learning models training or calibration.

In addition to the training, calibration, and verification partitions, a distinct subset of wells was deliberately set aside from the primary model dataset. These wells are referred to as blind samples. They are used to further validate the model in a process called blind validation. In blind validation, the blind subset of data is not used or seen during the training process. This ensures that the model is not overfitting to the training data and that it can generalize to new data. The blind samples will be used to supplement the validation process and demonstrate the effectiveness of the model on data that it has not had access to during its development.

Figure 29 illustrates the partitioning of the dataset.

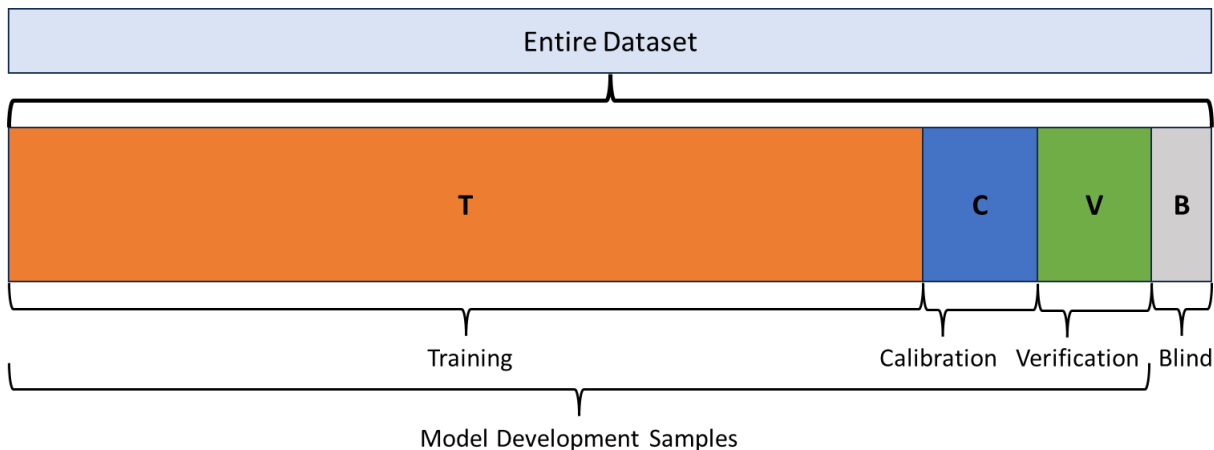


Figure 29: Dataset Partitioning

IMPROVE™ software offers various methods for partitioning the dataset, as shown in Figure 30. In this study, we used the Intelligent Partitioning technique, a proprietary method within the software that ensures all partitions are statistically representative of the dataset. We also investigated a manual partitioning approach to further direct the training process. These techniques will be discussed in more detail within the results chapter of this work.

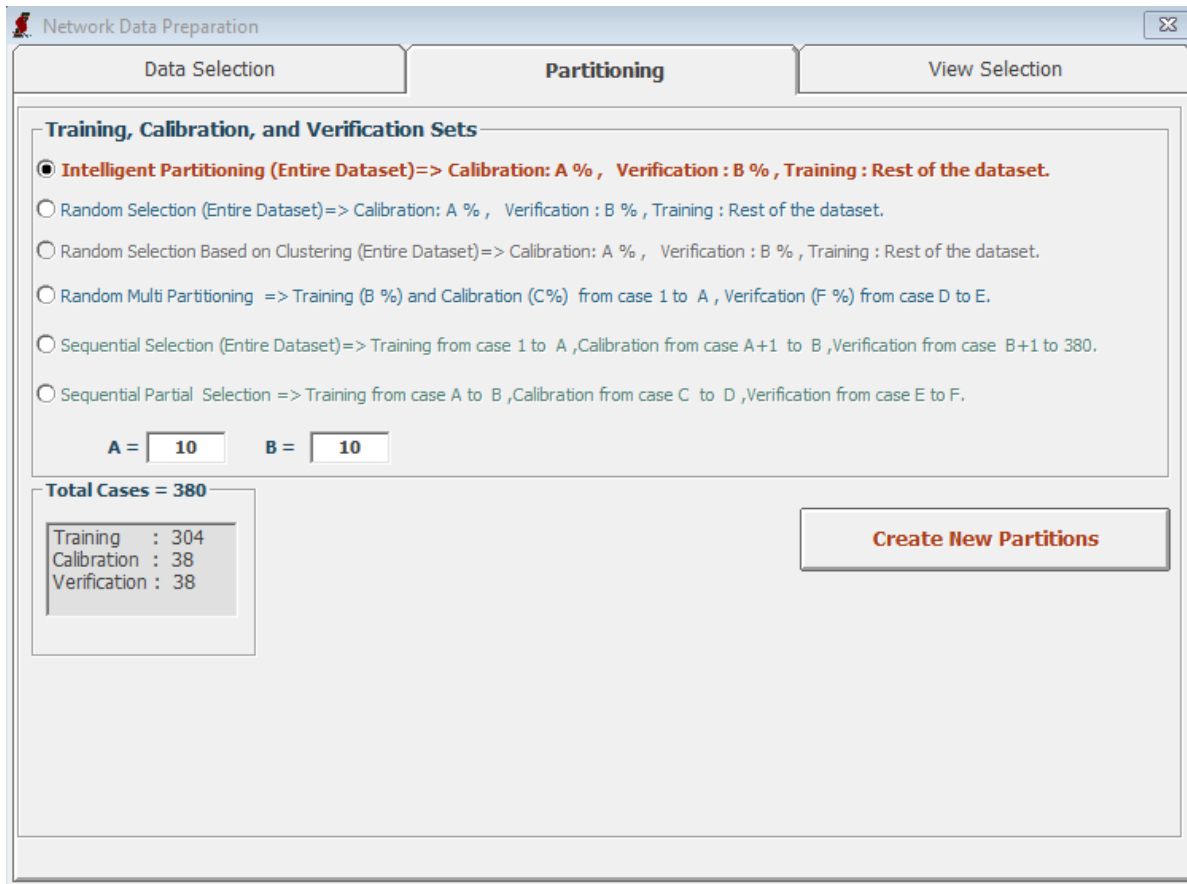


Figure 30: Data Partitioning – IMPROVE™ Software

3.3.3.3 Model Development: Predictive AI Model for 30 Day Rich Gas production

Once the partitions have been defined and the input-output features have been selected, the database is primed to train our initial models. The IMPROVE™ software offers various neural network training methods. In this research, we employed the backpropagation algorithm.

Backpropagation is a learning algorithm for training feedforward neural networks. Particularly effective for networks with a high number of parameters, it works by iteratively adjusting the weights of the network's connections, aiming to minimize the discrepancy between the predicted output and the target output (Fausett, 1993).

The backpropagation algorithm works in two phases:

1. **Forward Propagation:** The input data is fed into the network and the network calculates its output.

2. **Backward propagation:** The error between the network's predicted output and the target is calculated. This error is then propagated back through the network to update the weights of the connections.

The key to backpropagation's efficiency is the proper tuning of weights. The weights of a neural network are the parameters that control how the network learns. By adjusting the weights, we can fine-tune the network's performance and make it more accurate and reliable.

When the weights are properly tuned, the network can learn the patterns in the data more effectively. This leads to a reduction in the error rate, which means the network is more likely to make accurate predictions. Additionally, a well-tuned network is more likely to generalize well to new data, meaning that it can be used to make predictions on data that is not seen before.

The weights of the connections are updated using a gradient descent algorithm. Gradient descent is an iterative optimization algorithm that works by moving in the direction of the steepest descent of a function. In the case of backpropagation, the function is the error between the network's predicted output and the target output (Kriesel, 2007).

The following steps illustrate how the backpropagation algorithm works for a simple feedforward neural network with one hidden layer, two input neurons, one hidden neuron, and one output neuron. The weights of the connections are initialized randomly.

1. The input data is fed into the network.
2. The network calculates its output.
3. The error between the network's output and the desired output is calculated.
4. The error is propagated back through the network, layer by layer.
5. The weights of the connections are updated using gradient descent.
6. Steps 2-5 are repeated until the error is minimized.

An example backpropagation neural network as designed in this research can be seen in Figure 31. Within the IMPROVE™ software, many of the default hyperparameters were suggested to be kept at their default values, as these values represented several years of cumulative experience working with different neural network models. With these hyperparameters constant, the first iterations of model tuning allow us to observe the impact of input selection and data partitioning in our initial models before further hyperparameter tuning is performed.

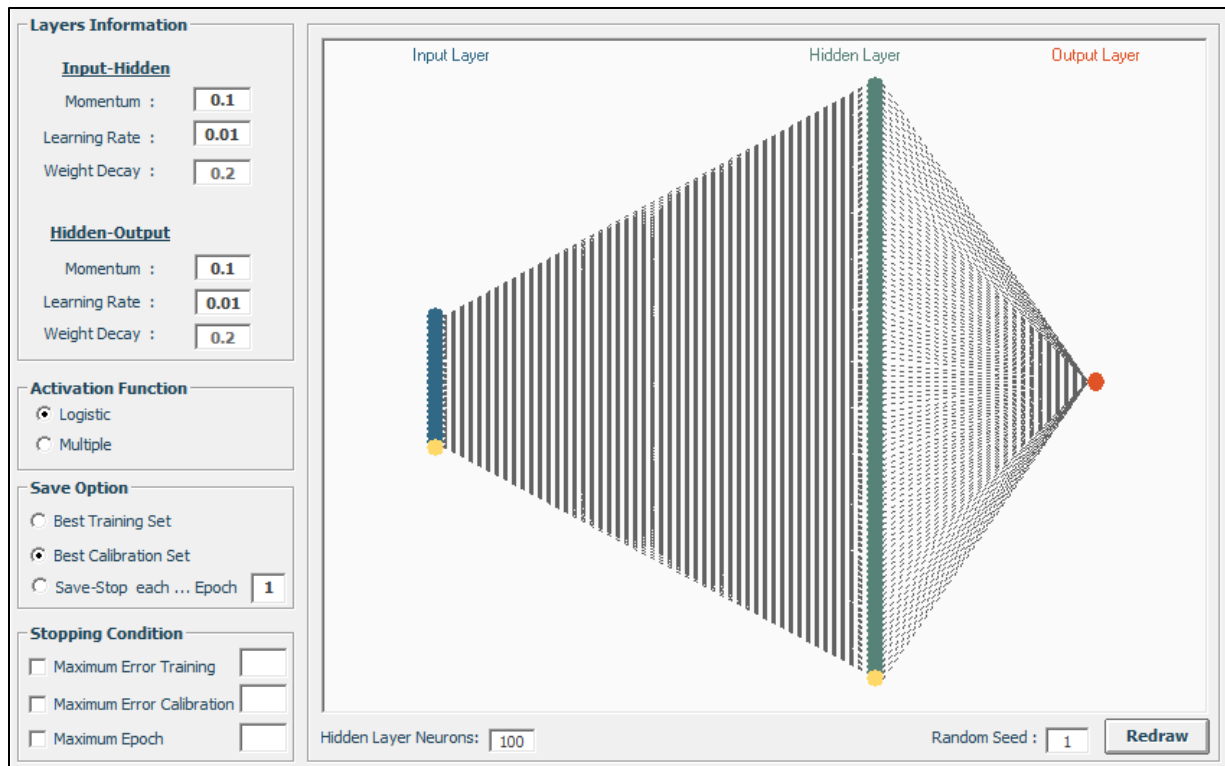


Figure 31: Neural Network Design – IMPROVE™ Software

A summary of the hyperparameter controls and their impact on the model are as follows (Intelligent Solutions, 2018):

- **Layers Information** – there are two sets of synaptic connections in the network. First are the synaptic connections between the input layer and the hidden layer, and the second set of connections are those between the hidden layer and the output layer. The “Layers Information” frame allows you to assign three parameters to each of these connection sets. These hyperparameters are summarized as follows:
 - **Momentum** – hyperparameter that helps the neural network to converge more quickly and smoothly.
 - **Learning Rate** – hyperparameter controlling how quickly the neural network learns.
 - **Weight Decay** – hyperparameter that penalizes large weights in a neural network, helping to prevent overfitting.
- **Activation Function** – for this study we used the Logistic Function, a hyperparameter that transforms the output of a neural network to a value between 0 and 1, which can be interpreted as a probability.

- **Save Option** – option for how and when to save the trained network.
- **Stopping Condition** – option for when to stop the training process, based on condition set.
- **Epoch** – one cycle of training the neural network with all training data.
- **Hidden Layer Neuron** – hyperparameter controlling the complexity of a neural network. A higher number will make the NN more complex but may also make it more difficult to train.
- **Random Seed Number** – hyperparameter controlling the randomness of the initial weights in the neural network.

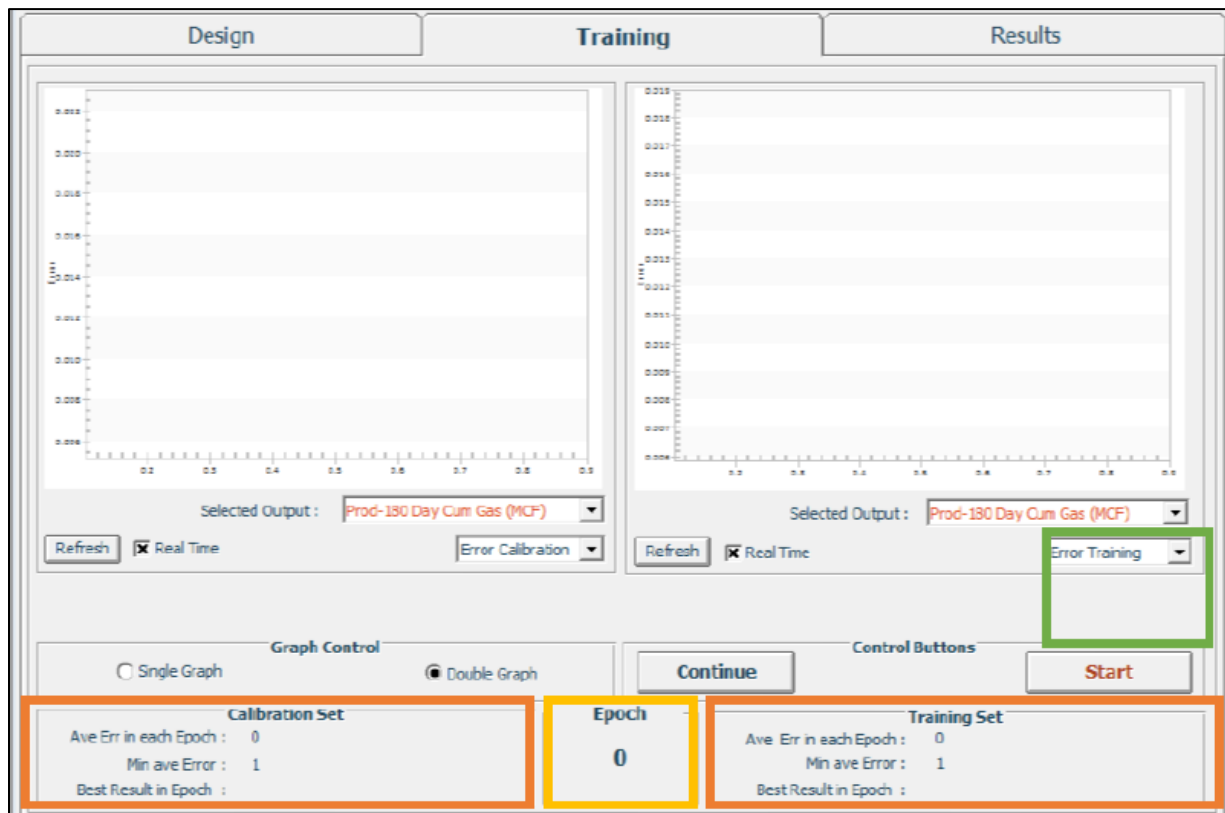


Figure 32: Backpropagation Neural Network Training Module – IMPROVE™ Software

Following the initial model design and preparation, we can begin training our neural network. Within the IMPROVE™ software we can observe this training in real-time, with the ability to monitor both the Training and Calibration sets, using an available double-graph option within the training module. Based on our model settings, and stopping conditions, the training process will continue for an unlimited number of epochs, saving the best model as it goes. An example of the training module is shown in Figure 32.

3.3.3.4 Model Validation and Predictive Analysis

After developing and training the data-driven model with promising accuracy seen on the verification data, the next step is to validate it using a blind dataset. Since the model has never encountered this dataset during training, it represents “new” data, allowing us to assess the model’s ability to generalize to new data.

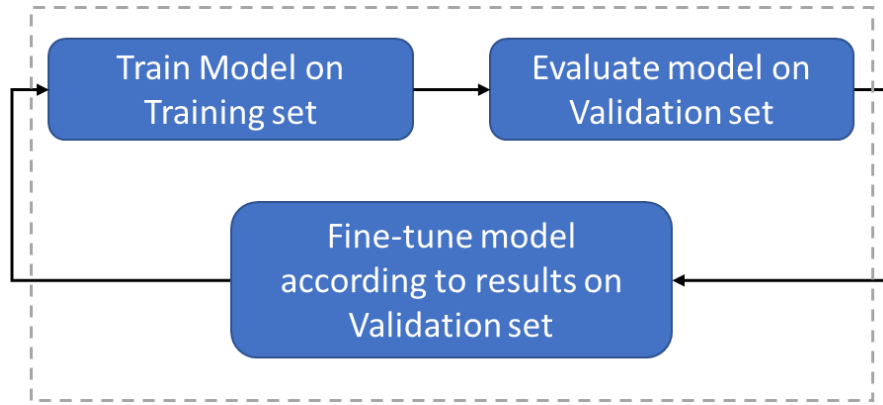


Figure 33: Model Validation Workflow

When importing the blind dataset, we must indicate the input attributes that are present both in the blind dataset and the trained model. Following this step, we can deploy our trained model on the blind data and examine predicted output values, sensitivity analysis, and model performance.

Based on the results throughout training, calibration, verification, and finally blind validation, we will track the model’s performance and error between actual and predicted output values. The key drivers in this analysis will be R-Squared (R^2) value, measuring the proportion of variation in two datapoints when applied to actual and predicted values, and the Correlation Coefficient, which measures the strength of the linear relationship between two variables.

Equation 3: R-Squared (R^2)

$$R^2 = 1 - \left(\frac{SS_{res}}{SS_{tot}} \right)$$

Where:

- SS_{res} is the sum of squared residuals (i.e., the difference between the actual values of the dependent variable and the predicted values from the model)
- SS_{tot} is the total sum of square (i.e., the total variation in the dependent variable)

Equation 4: Correlation Coefficient

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum((x_i - \bar{x})^2 \sum(y_i - \bar{y})^2)}}$$

Where:

- r = correlation coefficient
- x_i = values of the x-variable in a sample
- \bar{x} = mean of the values of the x-variable
- y_i = values of the x-variable in a sample
- \bar{y} = mean of the values of the y-variable

Additionally, the model performance will also be analyzed on the Mean Absolute Error (MAE) presented earlier, which can be interpreted as the average error the model's predictions have in comparison with their actual targets.

This analysis will evaluate the validity of our model, with the objective of selecting a model that generalizes well to new data. We aim to enhance our predictive accuracy through an iterative process of input and hyperparameter tuning, as depicted in Figure 33. Once the model demonstrates consistent performance, especially on the blind validation dataset, it can be considered ready to use.

However, it is important to note that data validation is an ongoing process. It should be performed routinely, especially when new data is added to a project. In our case, taking our model into the future would entail regular updates and validation as new wells are developed and as their related data becomes available.

Chapter 4: Results and Discussion

This chapter presents the findings from our comprehensive comparison between Artificial Intelligence (AI) techniques and traditional numerical reservoir simulations (NRS) within the context of natural gas extraction in unconventional Marcellus Shale plays.

Key highlights include:

1. Assessment of prediction accuracy of AI and NRS.
2. Analysis of resource allocation for model development and simulation.
3. Evaluation of the capability of both techniques to capture intricacies found in shale formations.
4. Comparison of both methodologies in forecasting and production history matching.
5. Advantages and limitations for each approach.

By comparing AI with traditional NRS, this chapter aims to provide insights into the potential advantages and challenges of both techniques in reservoir simulation, predictive analytics, and optimization.

The following sections will detail the findings, supported by empirical data, and examine their significance within the domain.

4.1 Numerical Reservoir Simulation Model Results

In this section, we present the results of the NRS model scenarios, developed to understand the impact of key parameters on the model's predictive performance. The values for these scenarios were determined after experimenting with a wide range of parameters. They were ultimately selected based on their ability to best history match the observed production data, while demonstrating their impact within the context of this research.

The results highlight the influence of assumptions and biases in parameter values on the NRS model's accuracy and reliability. The following sections will:

- Summarize the model scenarios applying the methodology described earlier.
- List the final reservoir parameters used in model development.
- Visualize the predicted production behavior alongside historical data.
- Quantify the model's accuracy by comparing its results with the known production history for 2036 days (December 12, 2015 – July 8, 2021).

4.1.1 Scenario 1: Base Model

The first scenario developed, referred to as our Base Model, was prepared using parameters that best matched the available data from the MSEEL project for the MIP-3H well. These parameter values were either directly measured or considered the best match based on their inclusion in other researcher's work using the same dataset.

The results of Scenario 1 are shared as follows, where:

- Table 11 displays the MIP-3H Base Model Parameters.
- Figure 34 illustrates the NRS MIP-3H Base Model's production behavior matched with the known MIP-3H production history.
- Table 12 presents the Base Model predictive results.

Table 11: NRS MIP-3H Base Model Parameters

MIP-3H: Base Model Reservoir Parameters		
Reservoir Parameters (MIP-3H)	Base Model Parameter Values	Units
Model Dimensions (MIP-3H)	7000 (Length)×1500(Width) ×90(Height)	ft.
Well Length (Horizontal)	6350	ft.
Initial Reservoir Pressure	4800	psia.
Initial Fissure Porosity	0.1	percent
Initial Matrix Porosity	4	percent
Initial Fissure Permeability i, j, k	7000, 7000, 700	nd
Initial Matrix Permeability i, j, k	1750, 1750, 175	nd
Number of Hydraulic Fractures	128	-
Fracture Half-Length, Xf	300	ft.
Initial Fracture Conductivity, kf wf	6	md-ft
Water Saturation	0.15	Fraction
Rock Density	120	lb/ft3
Langmuir Pressure Constant	0.00036	psi-1
Langmuir Volume Constant	0.05	g-mol/lb

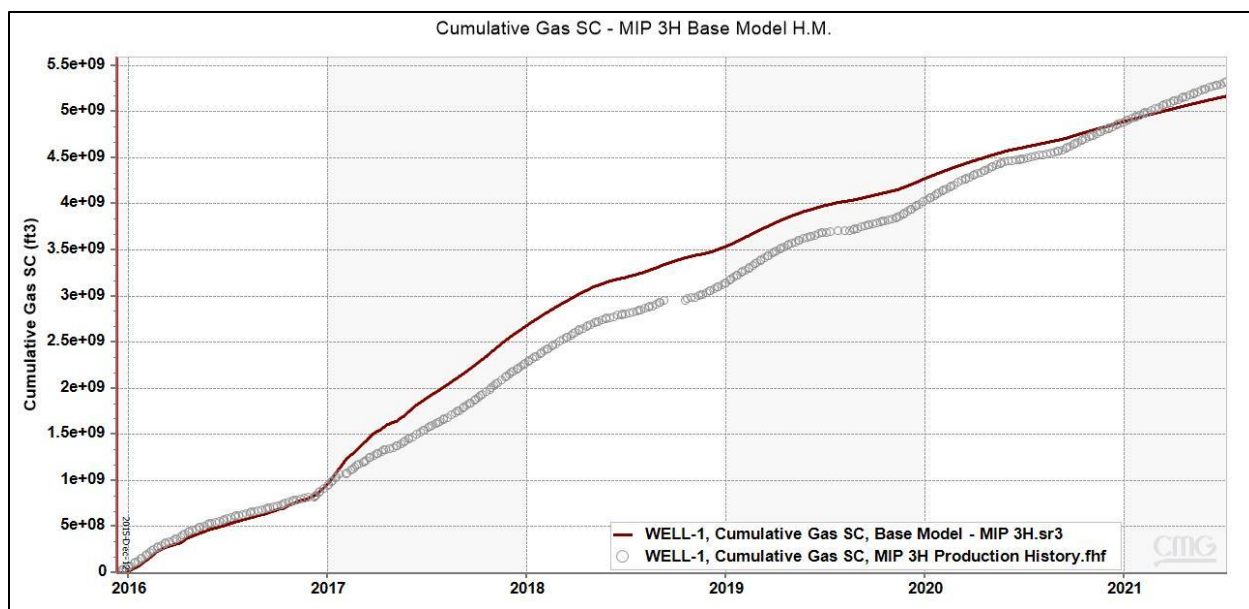


Figure 34: NRS MIP-3H Base Model Predicted Production Behavior and History Match

Table 12: NRS MIP-3H Base Model Results

MIP 3H Scenario – Base Model Results	
Mean Absolute Error (MAE)	215956762.9
Mean Absolute Percentage Error (MAPE)	11.44%
Time to Run Model	51:47 min.

The Base Model scenario results seen in Table 12, demonstrated the predictive capability of the NRS model when using the best available data. However, it's important to acknowledge that these values may not necessarily reflect the reality of the entire reservoir.

The NRS model is still limited by its inability to account for all factors that affect reservoir performance. This can be seen in the history match plot (Figure 34), which shows discrepancies between the simulated and observed data. These discrepancies suggest that there is still uncertainty in the parameter values, and that the model may not be able to accurately predict the reservoir performance under all conditions.

4.1.2 Scenario 2: Number of (Producing) Hydraulic Fracture Perforations & Fracture Half-Length

In our second scenario on the MIP-3H well, we studied the impact of producing hydraulic fractures and fracture half-length parameters, keeping all other parameter values consistent with the Base Model scenario above.

- Table 13 displays the MIP-3H Scenario 2 model parameters.
- Figure 35 illustrates the MIP-3H Scenario 2 model's production behavior matched with the known MIP-3H production history.
- Table 14 presents the predictive model results for Scenario 2.

Table 13: NRS MIP-3H Parametric Study Scenario: Number of (Producing) Hydraulic Fracture Perforations & Fracture Half Length

Scenario: Number of (Producing) Hydraulic Fracture Perforations & Fracture Half Length			
Reservoir Parameters (MIP-3H)	Base Model Parameter Values	Modified Scenario Parameter Values	Units
Model Dimensions (MIP-3H)	7000 (L)×1500(W) ×90(H)	7000 (L)×1500(W) ×90(H)	ft.
Well Length (Horizontal)	6350	6350	ft.
Initial Reservoir Pressure	4800	4800	psia.
Initial Fissure Porosity	0.1	0.1	percent
Initial Matrix Porosity	4	4	percent
Initial Fissure Permeability l, j, k	7000, 7000, 700	7000, 7000, 700	nd
Initial Matrix Permeability l, j, k	1750, 1750, 175	1750, 1750, 175	nd
Number of Hydraulic Fractures	128	98	-
Fracture Half-Length, Xf	300	450	ft.
Initial Frac Conductivity, kf wf	6	6	md-ft
Water Saturation	0.15	0.15	Fraction
a Rock Density	120	120	lb/ft3
Langmuir Pressure Constant	0.00036	0.00036	psi-1
Langmuir Volume Constant	0.05	0.05	g-mol/lb
Comments: Perforations dynamically produce (2 years)			

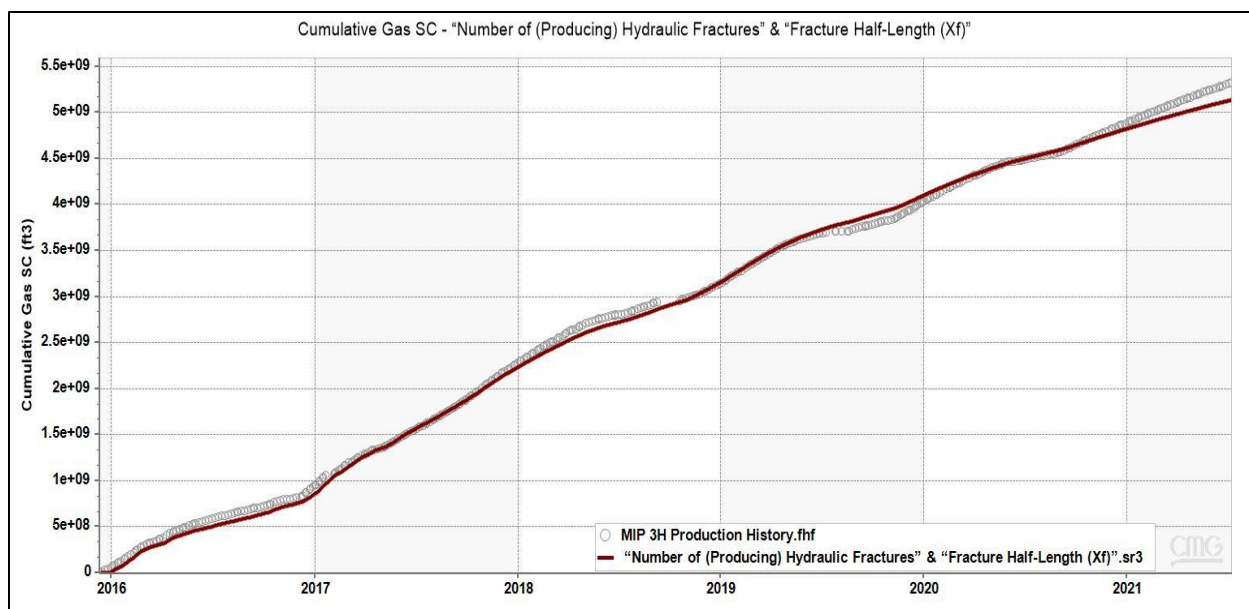


Figure 35: NRS MIP-3H Parametric Study Scenario: Number of (Producing) Hydraulic Fracture Perforations & Fracture Half Length – Predicted Production Behavior and History Match

Table 14: NRS MIP-3H Parametric Study Scenario Results: Number of (Producing) Hydraulic Fracture Perforations & Fracture Half Length

MIP 3H Scenario – Number of Hydraulic Fractures & Fracture Half Length Model Results	
Mean Absolute Error (MAE)	54075714.41
Mean Absolute Percentage Error (MAPE)	5.32%
Time to Run Model	98:55 min.

The results of Scenario 2 parametric study showed that the number of producing perforations and the fracture half-length parameters significantly influence the NRS model’s history matching output.

Table 14 and Figure 35 show that the MIP-3H model has a very good match to the known production. When comparing our results to the Base Model scenario, our MAPE saw an improvement from 11.44% down to 5.32%.

For the fracture half-length parameter, an initial estimated length of 300 feet was increased to 450 feet. This increase in half-length improved the model’s ability to match the well’s production history due to the model’s assumption of uniform, planar fractures, and their direct correlation with stimulated reservoir volume (SRV). These assumptions are common in NRS models to address complexity and improve modelling time.

Adjustments to the parameter for producing fracture perforations were more complex as they were applied dynamically following iterative changes done prior to selecting the model presented. Initially set at 128, the number of producing perforations dropped to 98 for a period of two years (from November 1, 2016 to November 1, 2018) before returning to 128. This change reflected perforations that were no longer contributing to production, directly influencing the model's predicted production.

Although the modifications in this scenario were exaggerated to best demonstrate the impact of these parameters, the theory behind these phenomena can be justified in real-world applications. This is supported by the detailed discussion provided earlier in this work.

4.1.3 Scenario 3: Matrix Porosity and Fracture Conductivity

In our third scenario on the MIP-3H well, we studied the impact of matrix porosity and fracture conductivity parameters, keeping all other parameter values consistent with the Base Model scenario above.

- Table 15 displays the MIP-3H Scenario 2 model parameters.
- Figure 36 illustrates the MIP-3H Scenario 2 model's production behavior matched with the known MIP-3H production history.
- Table 16 presents the predictive model results for Scenario 2.

Table 15: NRS MIP-3H Parametric Study Scenario: Matrix Porosity and Fracture Conductivity

Scenario: Matrix Porosity and Fracture Conductivity			
Reservoir Parameters (MIP-3H)	Base Model Parameter Values	Modified Scenario Parameter Values	Units
Model Dimensions (MIP-3H)	7000 (L)×1500(W) ×90(H)	7000 (L)×1500(W) ×90(H)	ft.
Well Length (Horizontal)	6350	6350	ft.
Initial Reservoir Pressure	4800	4800	psia.
Initial Fissure Porosity	0.1	0.1	percent
Initial Matrix Porosity	4	4.5	percent
Initial Fissure Permeability I, j, k	7000, 7000, 700	7000, 7000, 700	nd
Initial Matrix Permeability I, j, k	1750, 1750, 175	1750, 1750, 175	nd
Number of Hydraulic Fractures	128	128	-
Fracture Half-Length, Xf	300	300	ft.
Initial Frac Conductivity, kf wf	6	4.5	md-ft
Water Saturation	0.15	0.15	Fraction
Rock Density	120	120	lb/ft3
Langmuir Pressure Constant	0.00036	0.00036	psi-1
Langmuir Volume Constant	0.05	0.05	g-mol/lb

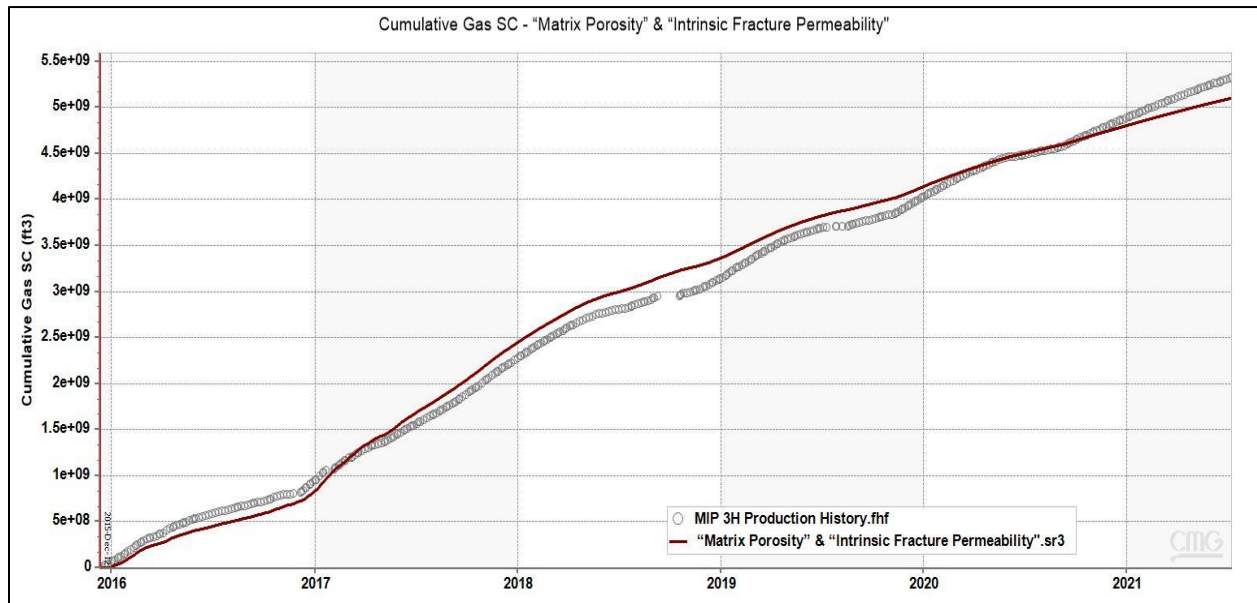


Figure 36: NRS MIP-3H Parametric Study Scenario: Matrix Porosity and Fracture Conductivity – Predicted Production Behavior and History Match

Table 16: NRS MIP-3H Parametric Study Scenario Results: Matrix Porosity and Fracture Conductivity

MIP 3H Scenario – Matrix Porosity and Intrinsic Fracture Permeability Model Results	
Mean Absolute Error (MAE)	122604974.9
Mean Absolute Percentage Error (MAPE)	9.27%
Time to Run Model	63:07 min.

The results of Scenario 3 parametric study illustrated the influence of the matrix porosity and fracture conductivity parameters on the NRS model’s history matching output. We specifically focus on small changes in the values of the parameters, as these can be explained by the difficulty of making accurate measurements and the different properties of heterogenous shale reservoirs. Greater detail and discussion of these parameters has been presented earlier in this research.

Table 16 and Figure 36 show that the MIP-3H model for Scenario 3 has a very good match to the known production. When comparing our results to the Base Model scenario, our MAPE saw an improvement from 11.44% down to 9.27%.

4.1.4 Scenario 4: Skin and Skin Effect

In our fourth scenario on the MIP-3H well, we studied the impact of the skin parameter, keeping all other parameter values consistent with the Base Model scenario above.

- Table 17 displays the MIP-3H Scenario 2 model parameters.
- Figure 37 illustrates the MIP-3H Scenario 2 model's production behavior matched with the known MIP-3H production history.
- Table 18 presents the predictive model results for Scenario 2.

Table 17: NRS MIP-3H Parametric Study Scenario: Skin Effect

Scenario: Skin Effect			
Reservoir Parameters (MIP-3H)	Base Model Parameter Values	Modified Scenario Parameter Values	Units
Model Dimensions (MIP-3H)	7000 (L)×1500(W) ×90(H)	7000 (L)×1500(W) ×90(H)	ft.
Well Length (Horizontal)	6350	6350	ft.
Initial Reservoir Pressure	4800	4800	psia.
Initial Fissure Porosity	0.1	0.1	percent
Initial Matrix Porosity	4	4	percent
Initial Fissure Permeability I, j, k	7000, 7000, 700	7000, 7000, 700	nd
Initial Matrix Permeability I, j, k	1750, 1750, 175	1750, 1750, 175	nd
Number of Hydraulic Fractures	128	128	-
Fracture Half-Length, Xf	300	300	ft.
Initial Fracture Conductivity, kf wf	6	6	md-ft
Water Saturation	0.15	0.15	Fraction
Rock Density	120	120	lb/ft3
Langmuir Pressure Constant	0.00036	0.00036	psi-1
Langmuir Volume Constant	0.05	0.05	g-mol/lb
* Skin	0	10	-
Comments: Skin dynamically modified (2 years) *Skin Factor included for parametric study.			

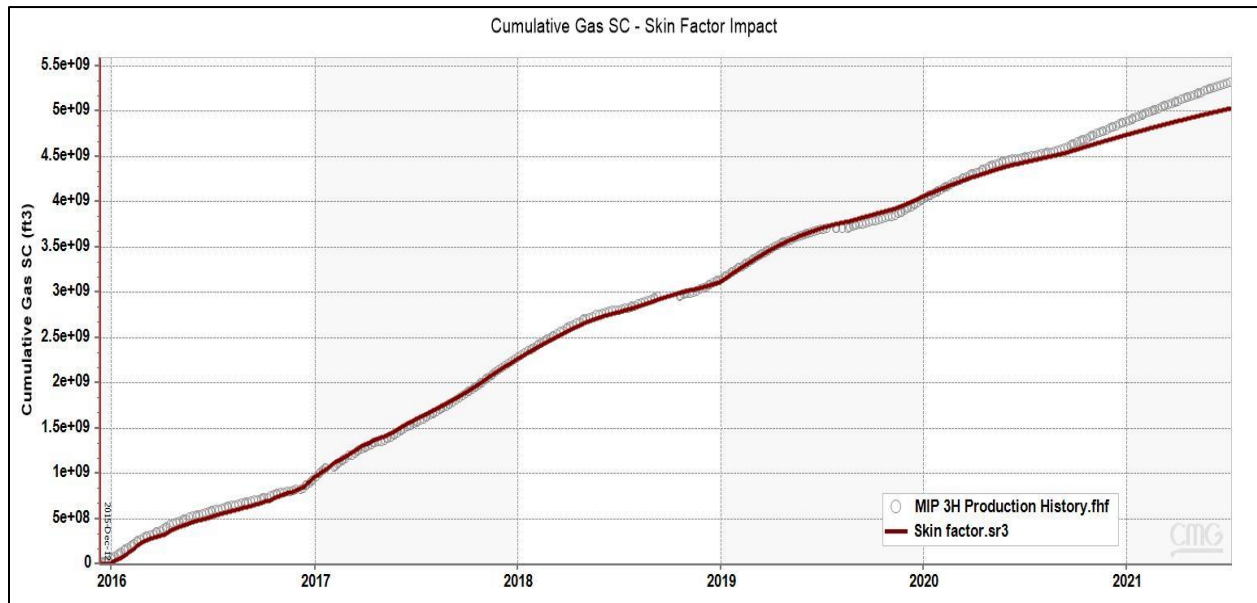


Figure 37: NRS MIP-3H Parametric Study: Skin Effect – Predicted Production Behavior and History Match

Table 18: NRS MIP-3H Parametric Study Scenario Results: Skin Effect

MIP 3H Scenario – Skin Factor Model	
Mean Absolute Error (MAE)	57372834.54
Mean Absolute Percentage Error (MAPE)	4.77%
Time to Run Model	69:52 min.

The results of Scenario 4 parametric study illustrate the impact of skin and skin effect on the NRS model. This model was unique in that used the same parameters as the MIP-3H Base Model from Scenario 1, with the only modification being the application of skin.

Table 18 and Figure 37 show that the MIP-3H model has a very good match to the known production. When comparing our results to the Base Model scenario, our MAPE saw an improvement from 11.44% down to 4.77%, which was the best history match of our 4 scenarios.

Like Scenario 2, skin was applied dynamically to the NRS model. The skin value was initially zero and was increased to 10 for a period of two years (from January 1, 2017 to January 1, 2019). After that, the skin was returned to zero. The skin value was estimated using limited data, so it is possible that the true skin

value was different from the value used in the study. However, this scenario reasonably demonstrates the theory presented, with additional details to justify the parameter modification presented earlier in this work.

Despite these limitations, the study provides valuable insights into the impact of skin on NRS models. The results of the study suggest that skin is an important parameter to consider when developing and calibrating NRS models, and its use-case in history matching a well can be easily applied to capture complexities that would otherwise be missed. This “catch-all” parameter aptly demonstrates the biases that can be brought into NRS models, and how the influence can be monumental given the problem one is trying to solve.

4.1.5 Summary of NRS Scenario Results

A summary of the presented results can be seen in Table 19 below.

Table 19: Results Summary for NRS Scenarios

Model	CPUs used	Time to run Model (elapsed seconds)	Convert to mins	MAE	MAPE
Scenario 1: Base Model	6	3107	51:47	215,956,762	11.44%
Scenario 2: Number of (Producing) Perforations & Fracture Half-Length	6	5935	98:55	54,075,714	5.32%
Scenario 3: Matrix porosity and fracture conductivity	6	3787	63:07	122,604,974	9.27%
Scenario 4: Skin and Skin Effect	6	4192	69:52	57,372,834	4.77%

Notably, the PC used for the modelling in this study had a 9th generation Intel Core i5-9500 processor with 6 cores and 6 threads. It had a base clock speed of 3 GHz and a max turbo frequency of 4.4 GHz. The PC also had 32GB of RAM running at 2666 MHz.

4.2 AI Data Driven Model results

In this section, we present the results of our investigation into the use of AI data-driven models to predict the production of wells within the Marcellus Shale region. As discussed within the methodology, a dataset containing 400 Marcellus Shale wells was used to develop and train the AI model using a back propagation neural network (NN) to predict the 30-Day Rich Gas production.

The development of a predictive NN model is an iterative process of continuous improvement. To illustrate our study's findings, we'll showcase our initial model's outcomes and discuss the fine-tuning processes that enhanced its predictive accuracy during both training and blind validation. This enhanced version is termed our 'tuned model'

The results demonstrate the ability of the NN model to learn from the provided dataset to predict quickly, accurately, and without assumptions. The predictions made for 30-day gas production are further validated based on the model's ability to generalize to the blind validation set of wells withheld throughout the training process.

Our presented model was developed with the following parameters shown in Table 20 and Table 21:

Table 20: Data Driven AI Model Preparation

AI Model Data Preparation	
Total number of Cases	400
Total cases used for Model Development	380
Number of Input Attributes used	25
Training % of Development cases	80%
Training Count	304
Calibration % of Development cases	10%
Calibration Count	38
Verification % of Development cases	10%
Verification Count	38
Total cases withheld for Blind Validation	20

Table 21: Backpropagation NN Preparation

Backpropagation NN Preparation	
Momentum	0.1
Learning Rate	0.01
Weight Decay	0.2
Activation Function	Logistic
Hidden Layer Neurons	100

Our iterative process is further illustrated in Figure 38. Through each iteration of model development, we compare the results of the developed models. Our aim is to continually move towards improved performance, ultimately selecting the model that both most accurately predicts and best generalizes to new data.

The following sections will present the results of our initial model and the improvements seen in our tuned model based on their predictive performance throughout development and blind validation.

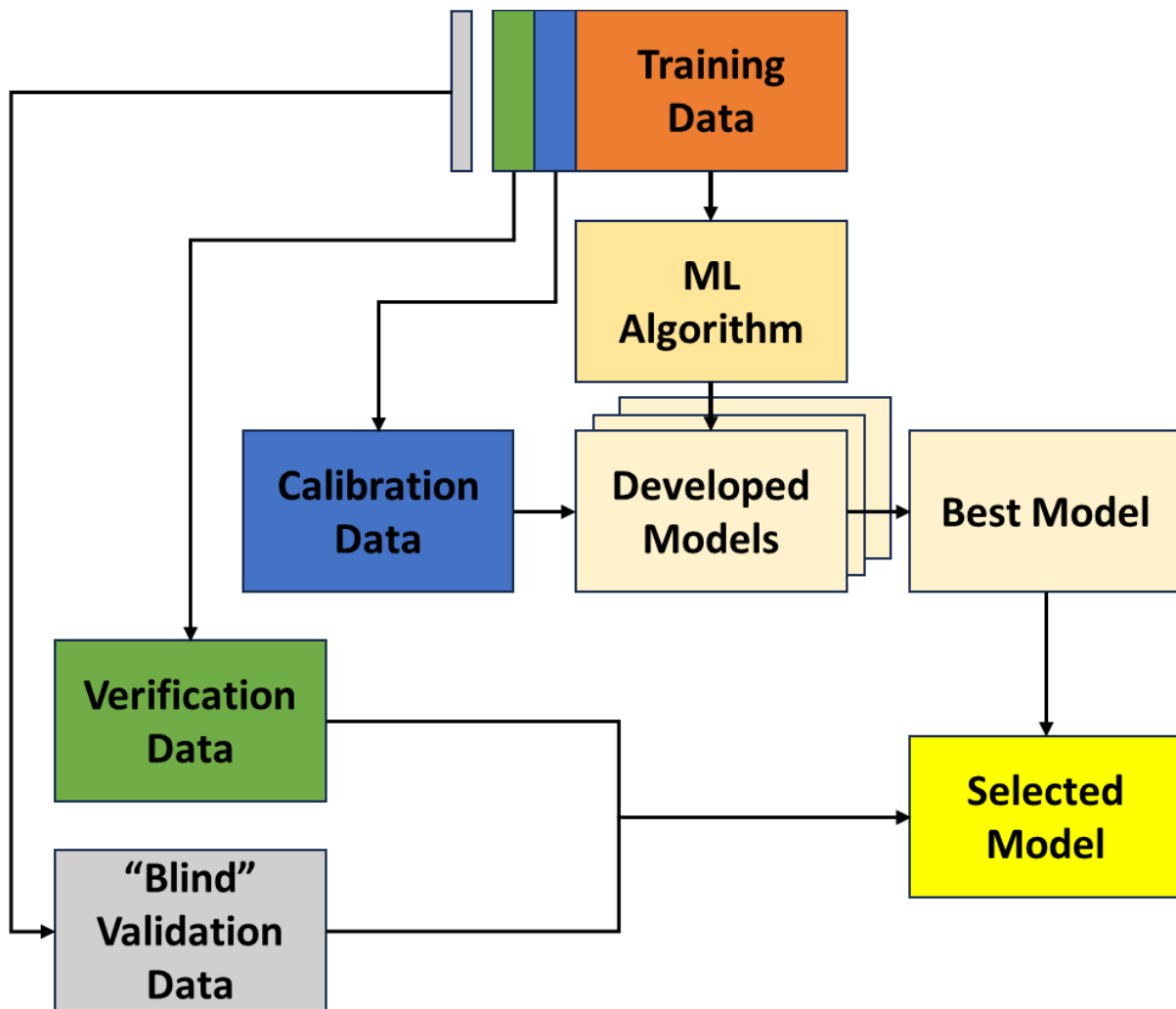


Figure 38: Model Selection Workflow

4.2.1 Data Driven AI Model: Initial Model Validation

Table 22: Data Driven AI Model: Initial Model Results

Initial Model Results	Train	Calibration	Verification
R²	83.19%	86.13%	71.62%
MAE	17012.48	17639.86	21297.99
MSE	391077166.55	49747835.93	90882109.16
Correlation Coefficient	0.9253	0.9366	0.8583

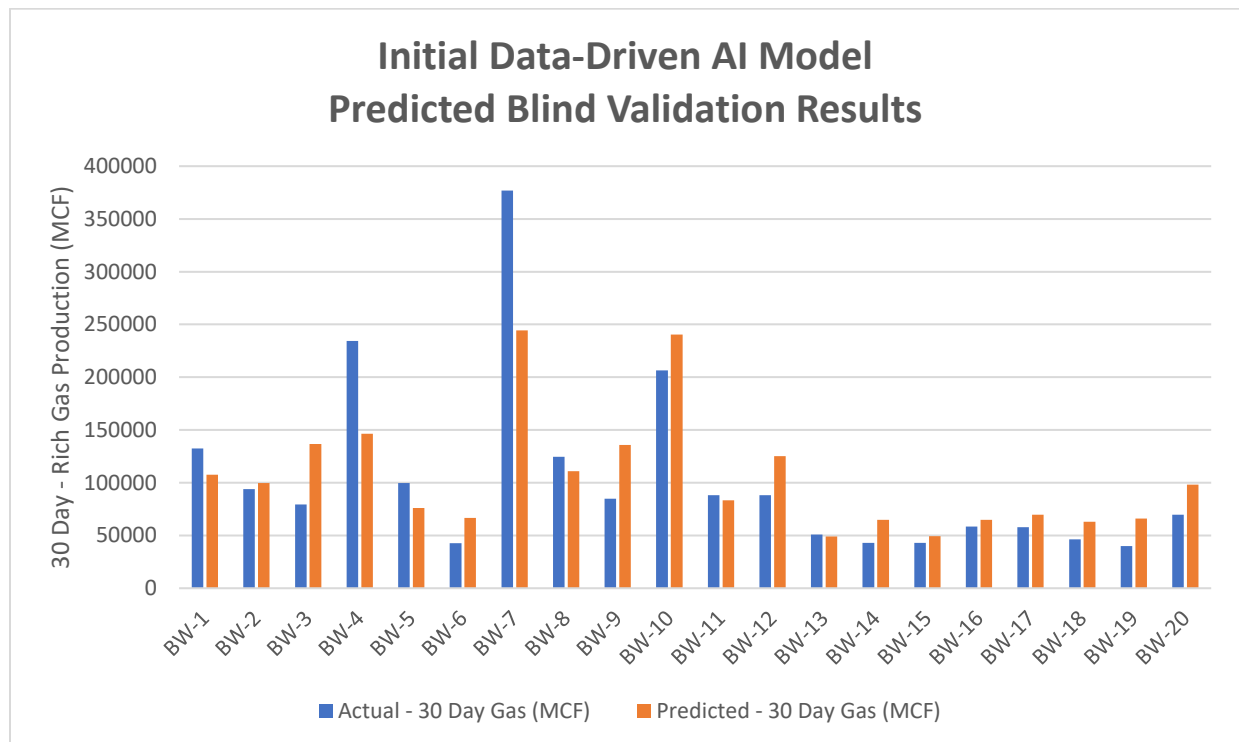


Figure 39: Data Driven AI Model: Initial Model Predicted Blind Validation Results

Table 23: Data Driven AI Model: Initial Model Blind Validation Results

Initial Model Results Blind Validation	
R²	74.98%
MAE	30794.799
MSE	1908473257.84
Correlation Coefficient	0.8659

4.2.2 Data Driven AI Model: Tuned Model Validation & Discussion

Table 24: Data Driven AI Model: Tuned Model Results

Tuned Model Results	Train	Calibration	Verification
R²	99.10%	65.19%	70.66%
MAE	3974.69	24013.66	22986.08
MSE	22889797.60	120649691.33	103827262.19
Correlation Coefficient	0.9956	0.8095	0.8853

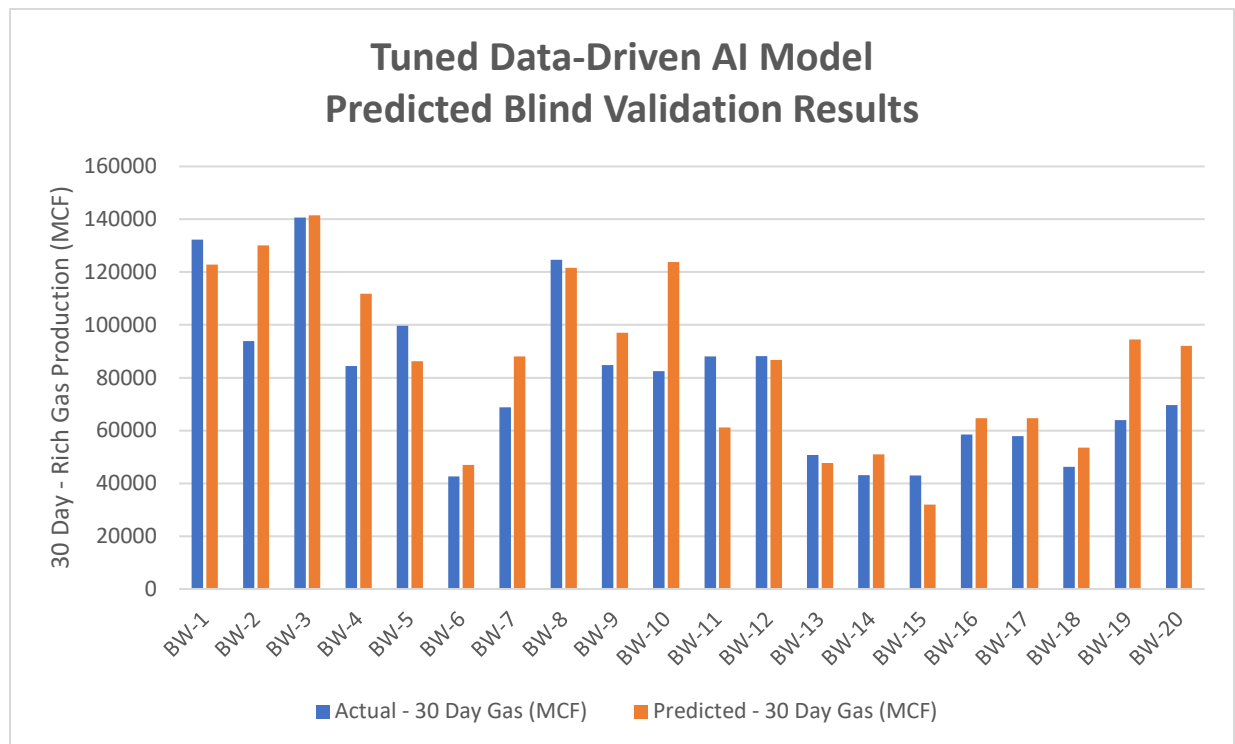


Figure 40: Data Driven AI Model: Tuned Model Predicted Blind Validation Results

Table 25: Data Driven AI Model: Tuned Model Blind Validation Results

Tuned Model Results Blind Validation	
R²	84.38%
MAE	17624.582
MSE	799524014.24
Correlation Coefficient	0.9186

4.2.3 AI Model: Blind Validation and Model Tuning Discussion

As discussed prior to training our initial model, we set aside blind validation samples from our dataset with the remaining samples used for model development and training. Post-training, our model achieved an overall R^2 score of 74.98%. However, its performance varied, particularly struggling with certain validation inputs.

Figure 41 illustrates the dataset partitions for the Initial model based on well locations to visualize the location of each well and their respective partition.

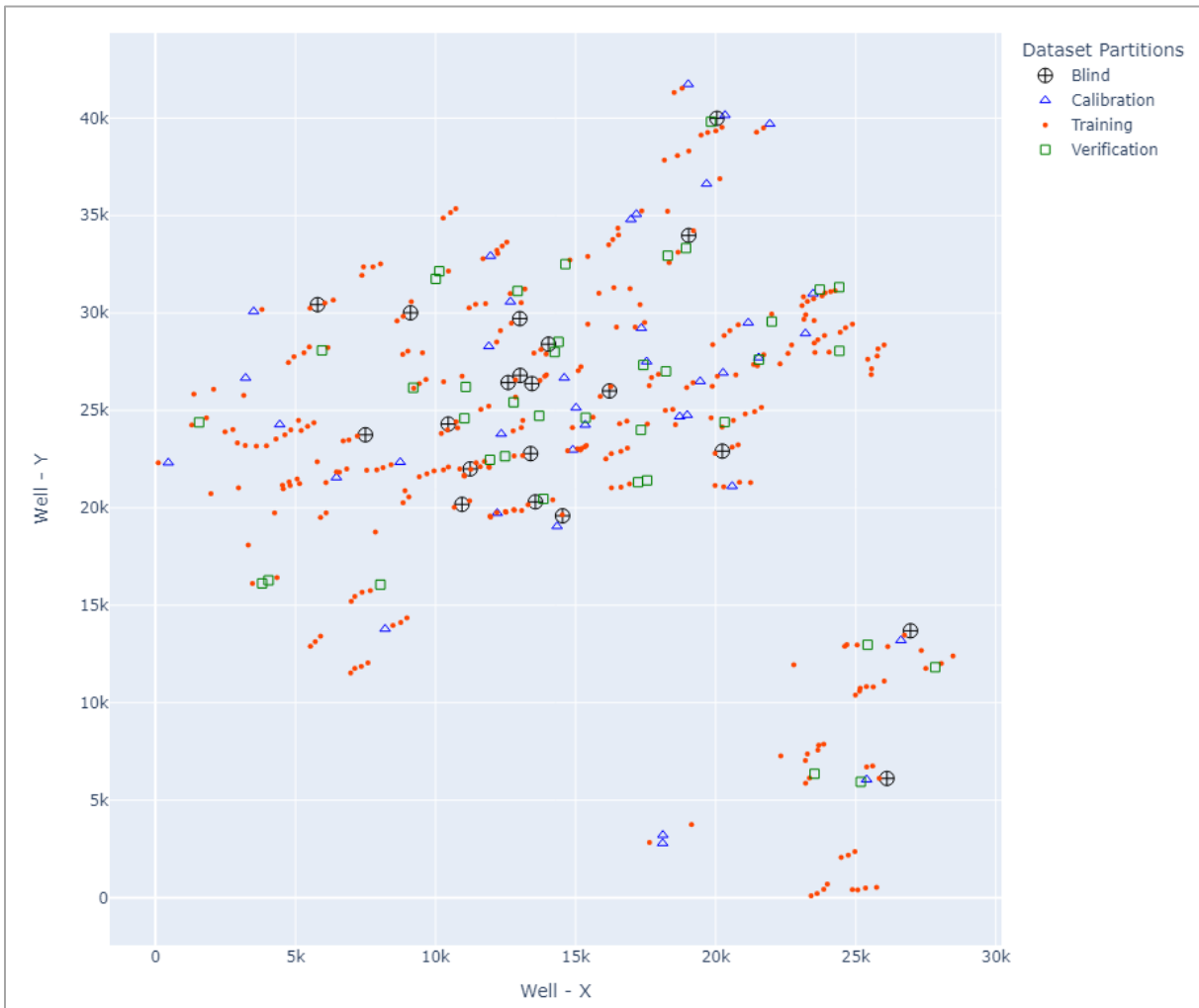


Figure 41: Data Driven AI Model: Initial Model Partition Visualization

To address this, we tuned our model and applied data augmentation, swapping challenging validation inputs with similar training datapoints. This aimed to provide the model with more representative data to enhance its learning capability. Figure 42 depicts the tuned model dataset following data augmentation.

After this adjustment, we retrained the neural network and the model's overall R^2 score improved to 84.38%. While data augmentation was effective for us, it might not be a one-size-fits-all solution. However, our results highlight the potential benefits of iterative training and the incorporation of augmentation techniques to improve predictive accuracy and continuous improvement when using NN models.

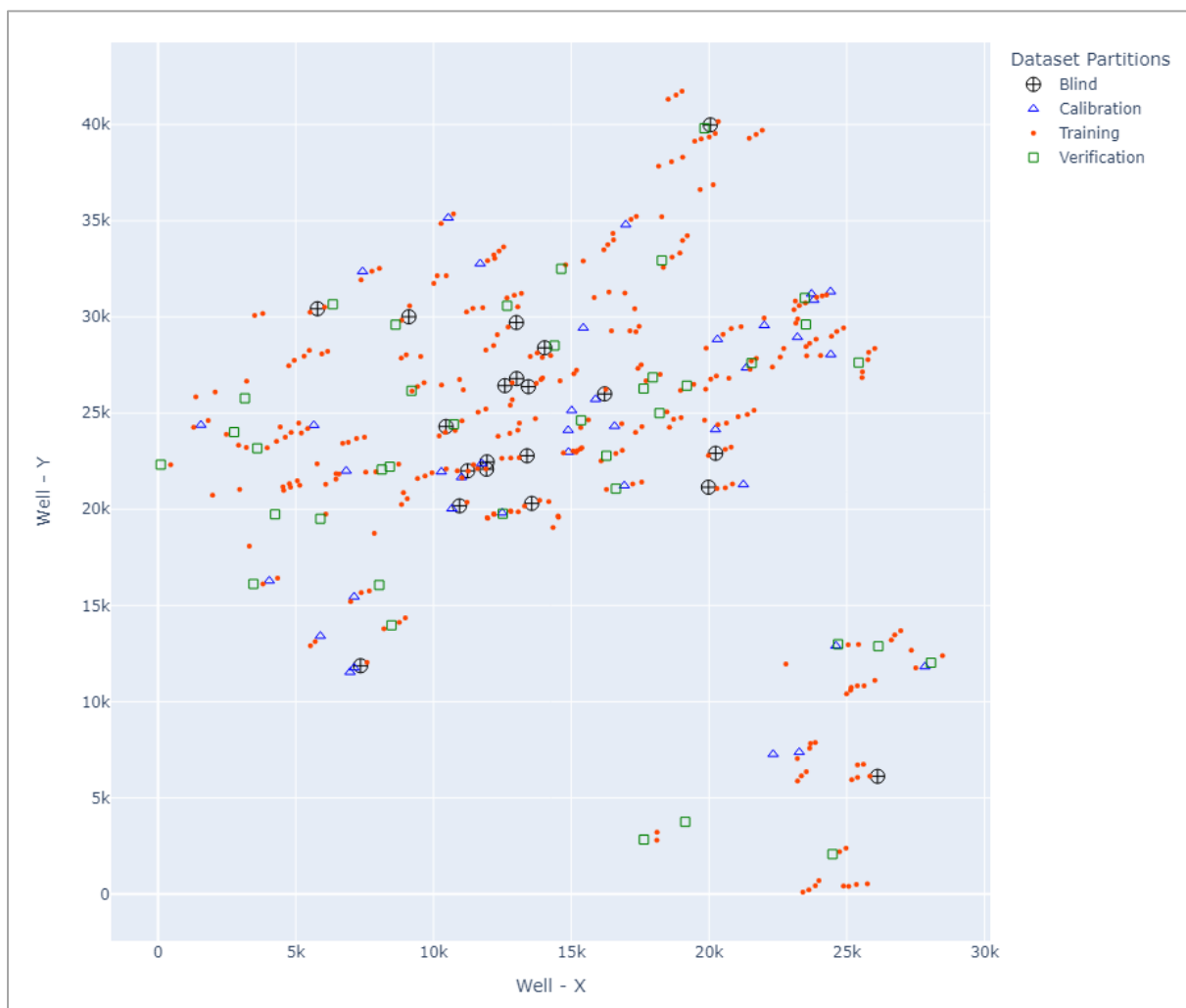


Figure 42: Data Driven AI Model: Tuned Model Partition Visualization

4.2.4 AI Reservoir modelling – Top-Down Model (TDM) Results

In addition to our backpropagation NN approach, we also explored AI Reservoir Modelling, a Top-Down Model (TDM) methodology using the same Marcellus Shale dataset. The TDM approach, unlike the gradient descent nature of backpropagation, starts with a high-level understanding, progressively refining its insights based on specific data patterns.

Essentially, while the backpropagation NN began with random weight initialization and adjusted based on error minimization, TDM began with overarching patterns in the Marcellus Shale dataset, diving deeper into details iteratively. This exploration was driven by the notion that TDM might offer a more intuitive, human-like breakdown of complex problems, providing an alternative perspective to the results gleaned from our backpropagation NN.

Interestingly, the TDM and Backpropagation NN models showcased complementary strengths in certain prediction scenarios, enhancing our comprehensive understanding of the Marcellus Shale region's well production. The model validation results for our TDM approach can be seen in the following Figure 43 for both the entire field of Marcellus shale wells as well as examples of individual wells within the dataset.

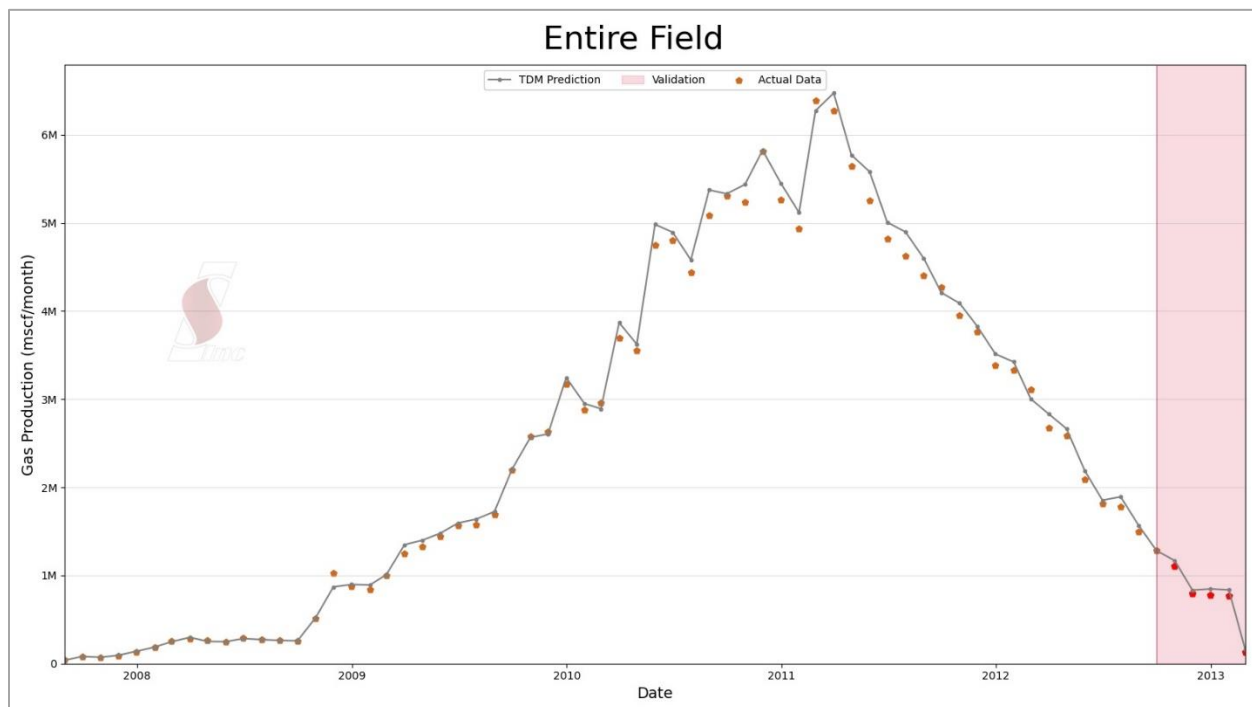


Figure 43: TDM Model: Entire Field Gas Production Results

Comparing our TDM Well results with the NRS model outcomes, we found consistent forecasts in the NRS model across all scenarios. This suggests that these NRS forecasts may not accurately reflect reality and need refinement. Conversely, TDM wells displayed unique and closely matched forecasting patterns, which were further supported through blind validation.

Though not as exhaustively analyzed as done throughout the history matching parametric study, the NRS scenarios consistently showed similar forecasting patterns, regardless of model development and parametric biases. These patterns, while requiring more validation, contrast with TDM outcomes which align more closely with known production data. This difference emphasizes TDM's success in data-driven predictions over assumptive parametric modeling.

The subsequent Figure 44 displays the forecasted production based on the NRS model scenarios discussed earlier in this study. In contrast, Figure 45 and Figure 46 showcase the TDM model's predictions and forecasts for select sample of wells from our 400-well Marcellus dataset, demonstrating the performance of TDM on a per well basis.

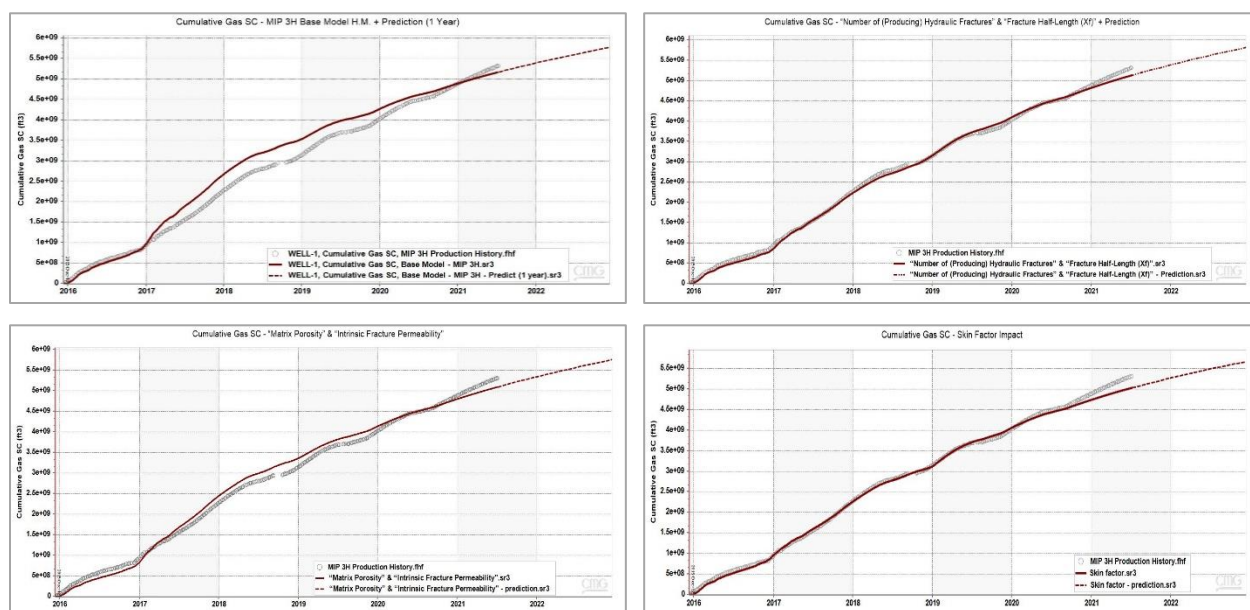


Figure 44: NRS Model Scenarios: Forecasted Production

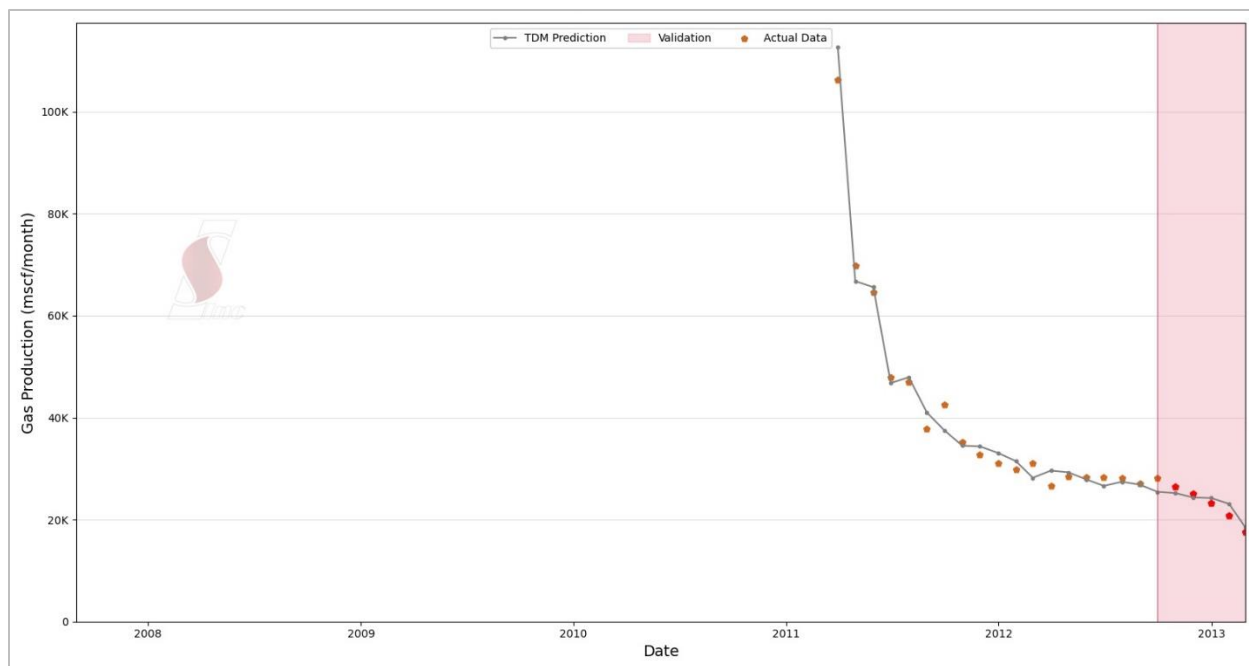


Figure 45: TDM Model: Forecasted Production (Example Well 1)

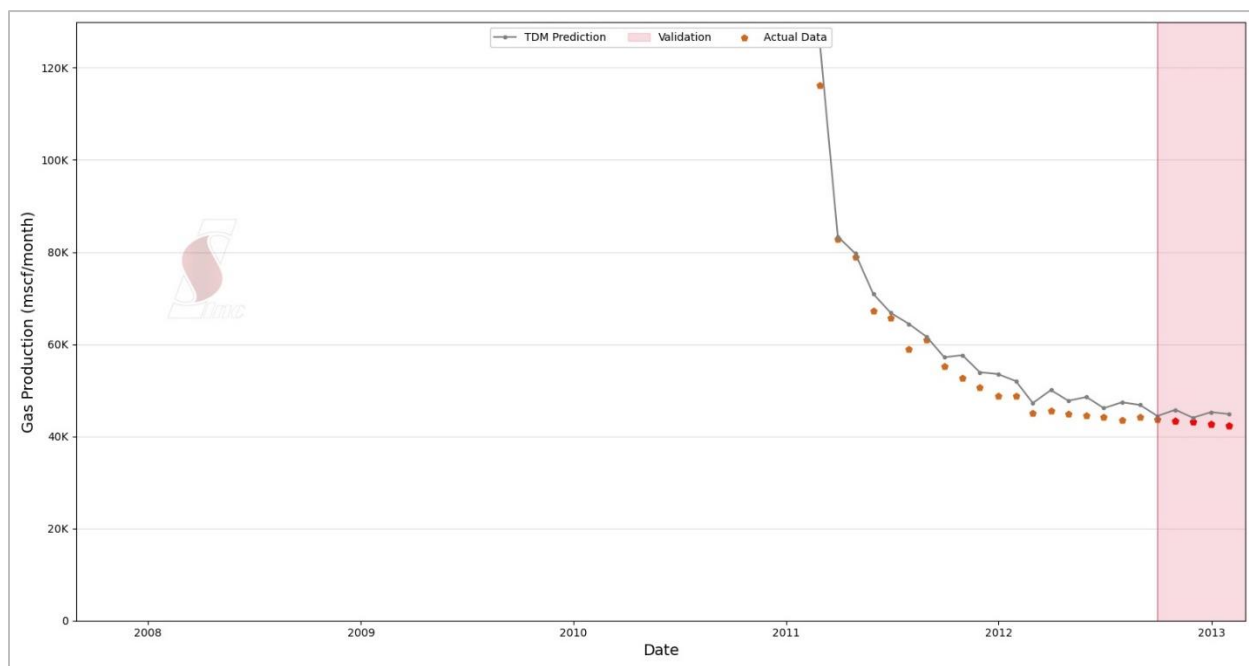


Figure 46: TDM Model: Forecasted Production (Example Well 2)

Chapter 6: Conclusion and Recommendations

6.1 Conclusion

Predictive analytics is essential for optimizing production and maximizing value in unconventional shale reservoirs. However, the complex nature of these reservoirs makes it challenging to develop accurate predictive models. This study has compared the modeling approaches and outcomes of two predictive analytics methods, numerical reservoir simulation (NRS) and data-driven artificial intelligence (AI) and machine learning (ML) techniques, in unconventional Marcellus Shale reservoirs.

NRS is a powerful tool for modeling complex reservoir dynamics, but its accuracy is extremely sensitive to parameter assumptions and data quality. These findings emphasize that even minor assumptions throughout NRS model development can significantly impact results, shedding light on the crucial role of parameter accuracy. The implications of these findings are that the NRS model's production prediction can be unreliable. The actual production rate could be higher or lower than the predicted rate, depending on the real-world values of the key parameters. Thus, stakeholders should carefully consider the uncertainty in NRS predictions when making decisions about future asset development and management.

AI models, on the other hand, demonstrate a strong ability to capture the intricacies of shale behavior, as the models can learn from large datasets of historical data. Our results suggest that AI models have the potential to be a valuable tool for predictive analytics in unconventional shale reservoirs, delivering predictions and optimizations that are not only more precise but also significantly more resource efficient.

However, acknowledging the limitations for both techniques is also important. The NRS model reveals discrepancies between simulated and actual data, illustrating the effects of parameter uncertainties. It does not capture all reservoir performance influencers, and assumptions, such as uniform, planar fractures, do not accurately represent actual reservoir conditions. The AI model, meanwhile, requires more data initially to begin training and may struggle to generalize if new data is significantly different from the data used to train and validate the initial model.

Looking forward, we can improve the reliability and confidence of our NRS model by limiting the use of assumptions and soft data. Furthermore, for the AI model, it is important to use a diverse and representative dataset during development, and to implement robust validation and testing protocols. This will help to ensure that the model is able to generalize well to new data and be more useful in real-world applications. Continuous iterative improvement, expanding the training dataset, and exploring different neural network architectures can further enhance predictive accuracy.

In summary, this research bridged traditional reservoir simulations with emerging AI methodologies, aiming to enhance the prediction accuracy of reservoir outputs. The significance of parameter selection in NRS models and the potential of iterative training in AI were notably emphasized. Despite identified limitations, this work paves the way for future studies, ensuring that the predictive capabilities in the reservoir domain continue to advance.

6.2 Recommendations

The findings of this study demonstrate the potential of artificial intelligence (AI) and machine learning (ML) for predictive analytics in unconventional shale reservoirs. However, further research and development are necessary to enhance the robustness, generalizability, and interpretability of AI models. The following recommendations are made to address these challenges and accelerate the adoption of AI-based predictive analytics solutions in the oil and gas industry:

- **Develop more robust and generalizable AI models.** This could be achieved by using more diverse and representative training datasets, developing more sophisticated neural network architectures, and using ensemble learning techniques.
- **Investigate the use of AI models to predict reservoir performance under different operating conditions.** This would allow operators to optimize their wells for different production targets and reservoir conditions.
- **Integrate Explainable AI (XAI) techniques to improve transparency and interpretability of employed AI and ML models.** By providing insights into the model's decision-making process, XAI can help stakeholders better understand and trust the predictions, improving the adoption of data-driven approaches.
- **Develop AI-powered decision support tools for operators.** These tools could help operators to make better decisions about well placement, completion design, and production management.
- **Collaborate with industry stakeholders to develop and implement AI-based predictive analytics solutions.** This would help to ensure that AI models are developed and used in a way that is beneficial to all stakeholders, with domain knowledge shared and implemented throughout the entire development and execution cycle of the asset.

In addition to these specific recommendations, it is also important to continue to invest in research on the fundamental principles of AI and machine learning. This research will help us to develop more powerful and efficient AI models that can be used to solve a wider range of problems in both the Marcellus Shale and the larger energy industry.

References

- Aboaba, A. L. (2022). Machine Learning Based Real-Time Quantification of Production from Individual Clusters in Shale Wells [Graduate Theses, Dissertations, and Problem Reports, West Virginia University Libraries]. In *Graduate Theses, Dissertations, and Problem Reports*.
<https://doi.org/https://doi.org/10.33915/etd.11269>
- Andrade, J., Civan, F., Devegowda, D., & Sigal, R. F. (2011). *SPE 144401 Design and Examination of Requirements for a Rigorous Shale-Gas Reservoir Simulator Compared to Current Shale-Gas Simulators*. <http://onepetro.org/SPEGTS/proceedings-pdf/11UGC/All-11UGC/SPE-144401-MS/1703509/spe-144401-ms.pdf/1>
- Barenblatt, G. I., Zheltov, I. P., & Kochina, I. N. (1960). Basic concepts in the theory of seepage of homogeneous liquids in fissured rocks [strata]. *Journal of Applied Mathematics and Mechanics*, 24(5), 1286–1303. [https://doi.org/10.1016/0021-8928\(60\)90107-6](https://doi.org/10.1016/0021-8928(60)90107-6)
- Carr, T., NETL, & DOE. (2015). *Marcellus Shale Energy and Environment Laboratory, MSEEL*.
<http://www.mseel.org/>
- Chen, Y., Wang, H., Li, F., & Zhou, T. (2022). Research Advance on the Diagnosis Technology of Hydraulic Fractures in Shale Reservoirs. *Frontiers in Energy Research*.
<https://doi.org/10.3389/FENRG.2022.919917/BIBTEX>
- Chorn, L., Stegent, N., & Yarus, J. (2014). Optimizing Lateral Lengths in Horizontal Wells for a Heterogeneous Shale Play. *Society of Petroleum Engineers - European Unconventional Resources Conference and Exhibition 2014: Unlocking European Potential, 1*, 124–135.
<https://doi.org/10.2118/167692-MS>
- Cipolla, C. L., Lolon, E. P., Erdle, J. C., & Rubin, B. (2010). *Reservoir Modeling in Shale-Gas Reservoirs*.
<http://onepetro.org/REE/article-pdf/13/04/638/2137977/spe-125530-pa.pdf/1>
- Cipolla, C., Weng, X., Onda, H., Nadaraja, T., Ganguly, U., & Malpani, R. (2011). New Algorithms and Integrated Workflow for Tight Gas and Shale Completions. *Proceedings - SPE Annual Technical Conference and Exhibition*. <https://doi.org/10.2118/146872-MS>
- CMG Ltd. (n.d.). *CMG Software | CMOST AI*. Retrieved August 25, 2023, from
<https://www.cmgl.ca/cmmost>
- Coats, K. H. (1969). *Use and Misuse of Reservoir Simulation Models - SPE-2367-pa*.
<http://onepetro.org/JPT/article-pdf/21/11/1391/2222685/spe-2367-pa.pdf/1>
- Cook, D., Downing, K., Bayer, S., Watkins, H., Sun Chee Fore, V., Stansberry, M., Saksena, S., & Peck, D. (2014). Unconventional Asset Development Work Flow in the Eagle Ford Shale. *SPE Unconventional Resources Conference*. <https://doi.org/10.2118/168973-MS>
- Davies, D. R., & Kuiper, T. O. H. (1988). Fracture Conductivity in Hydraulic Fracture Stimulation. *Journal of Petroleum Technology*, 40(05). <https://doi.org/10.2118/17655-PA>
- EIA. (2015). *World Shale Resource Assessments - Independent Statistics and Analysis*. U.S. Energy Information Administration - EIA. <https://www.eia.gov/analysis/studies/worldshalegas/>

- El Sgher, M. (2021). *The Impacts of the Net Stress and Stress Shadow on the Productivity of Marcellus Shale Horizontal Well Productivity of Marcellus Shale Horizontal*. <https://researchrepository.wvu.edu/etd/8321>
- Elsaig, M., Aminian, K., Ameri, S., & Zamirian, M. (2016). *Accurate Evaluation of Marcellus Shale Petrophysical Properties* (Vol. 2).
- Esmaili, S., & Mohaghegh, S. D. (2016). Full field reservoir modeling of shale assets using advanced data-driven analytics. *Geoscience Frontiers*, 7(1), 11–20. <https://doi.org/10.1016/j.gsf.2014.12.006>
- Fausett, L. (1993). *Fundamentals of Neural Networks: Architectures, Algorithms, and Applications*. Prentice Hall.
- Fisher, M. K., Heinze, J. R., Harris, C. D., Davidson, B. M., Wright, C. A., & Dunn, K. P. (2004). Optimizing Horizontal Completion Techniques in the Barnett Shale Using Microseismic Fracture Mapping. *SPE Annual Technical Conference Proceedings*. <https://doi.org/10.2118/90051-MS>
- Gholinezhad, J., Senam, J., Mohamed, F., & Hassan, G. (2018). *Challenges in Modelling and Simulation of Shale Gas Reservoirs*. <http://www.springer.com/series/15391>
- Guo, Z., Sankaran, S., & Sun, W. (2023). *Reservoir Modeling, History Matching, and Characterization with a Reservoir Graph Network Model*. <https://doi.org/10.2118/209337-PA/3065454/spe-209337-pa.pdf/1>
- IHS Harmony. (2020, February 24). *Time-Dependent Skin*. IHS Harmony Reference Materials. https://www.ihsenergy.ca/support/documentation_ca/Harmony/content/html_files/reference_material/general_concepts/time_dependent_skin.htm
- Intelligent Solutions, Inc. (2018). *IMprove Manual*. www.IntelligentSolutionsInc.com
- Islam, M. R., Hossain, M. E., Moussavizadegan, S. H., Mustafiz, S., & Abou-Kassem, J. H. (2016). Advanced Petroleum Reservoir Simulation: Towards Developing Reservoir Emulators: Second Edition. In *Advanced Petroleum Reservoir Simulation: Towards Developing Reservoir Emulators: Second Edition*. Wiley. <https://doi.org/10.1002/9781119038573>
- Kazatchenko, E., Markov, M., & Mousatov, A. (2003). Determination of Primary and Secondary Porosity in Carbonate Formations Using Acoustic Data. *Proceedings - SPE Annual Technical Conference and Exhibition*. <https://doi.org/10.2118/84209-MS>
- Keelan, D. K. (1982). Core Analysis for Aid in Reservoir Description. *Journal of Petroleum Technology*. <https://doi.org/10.2118/10011-PA>
- Kriesel, D. (2007). *A Brief Introduction to Neural Networks*. <http://www.dkriesel.com/>
- Kundert, D., & Mullen, M. (2009). *Proper Evaluation of Shale Gas Reservoirs Leads to a More Effective Hydraulic-Fracture Stimulation*. <http://onepetro.org/SPERMPTC/proceedings-pdf/09RMPC/All-09RMPC/SPE-123586-MS/2719974/spe-123586-ms.pdf/1>
- Liu, C. (2022). Numerical Simulation of Ultra-Low Permeability Reservoirs: Progress and Challenges. *Frontiers in Energy Research*, 10. <https://doi.org/10.3389/fenrg.2022.895135>

- Miller, C., Waters, G., & Rylander, E. (2011). Evaluation of Production Log Data from Horizontal Wells Drilled in Organic Shales. *Society of Petroleum Engineers - SPE Americas Unconventional Gas Conference 2011, UGC 2011*. <https://doi.org/10.2118/144326-MS>
- Mohaghegh, S. D. (2017). *Shale Analytics: Data-Driven Analytics in Unconventional Resources*.
- Qian, Y., Gao, P., Fang, X., Sun, F., Cai, Y., & Zhou, Y. (2022). Microstructure Characterization Techniques for Shale Reservoirs: A Review. *Frontiers in Earth Science*. <https://doi.org/10.3389/FEART.2022.930474/BIBTEX>
- Raji, A., Ogolo, O., & Anochie, V. (2020). Effect of Skin Factor on a Pressure Maintenance Project. *Society of Petroleum Engineers - SPE Nigeria Annual International Conference and Exhibition 2020, NAIC 2020*. <https://doi.org/10.2118/203743-MS>
- SLB Energy Glossary. (2023a). *Skin, Skin Factor, Skin Effect*. SLB Energy Glossary. <https://glossary.slb.com/en/terms/s/skin>
- SLB Energy Glossary. (2023b). *Unconventional Resource*. SLB Energy Glossary.
- Slocombe, R., Acock, A., Fisher, K., Viswanathan, A., Chadwick, C., Reischman, R., & Wigger, E. (2013). Eagle Ford Completion Optimization Using Horizontal Log Data. *Proceedings - SPE Annual Technical Conference and Exhibition*. <https://doi.org/10.2118/166242-MS>
- Spain, D. R., Gil, I., Sebastian, H., Smith, P. S., Wampler, J., Cadwallader, S., & Graff, M. (2015, January 26). Geo-Engineered Completion Optimization: An Integrated, Multi-Disciplinary Approach to Improve Stimulation Efficiency in Unconventional Shale Reservoirs. *SPE Middle East Unconventional Resources Conference and Exhibition*. <https://doi.org/10.2118/SPE-172921-MS>
- Ugueto C., G. A., Huckabee, P. T., Molenaar, M. M., Wyker, B., & Somanchi, K. (2016). Perforation Cluster Efficiency of Cemented Plug and Perf Limited Entry Completions; Insights from Fiber Optics Diagnostics. *Society of Petroleum Engineers - SPE Hydraulic Fracturing Technology Conference, HFTC 2016*. <https://doi.org/10.2118/179124-MS>
- Warren, J. E., & Root, P. J. (1963). The Behavior of Naturally Fractured Reservoirs. *Society of Petroleum Engineers Journal*, 3(03), 245–255. <https://doi.org/10.2118/426-PA>
- Yang, Q., Li, B., Shao, M., & Zhang, H. (2021). Production Performance of Perforation Clusters during Multistage Fracturing in Shale Gas Reservoirs. *ACS Omega*, 6(40), 26231–26238. <https://doi.org/10.1021/acsomega.1c03233>
- Zheng, S., Manchanda, R., & Sharma, M. M. (2020). Modeling Fracture Closure with Proppant Settling and Embedment during Shut-In and Production. *SPE Drilling & Completion*, 35(4). <https://doi.org/10.2118/201205-PA>

Appendix

AI Dataset: Input Attribute Behavior (Fuzzy Pattern Recognition)

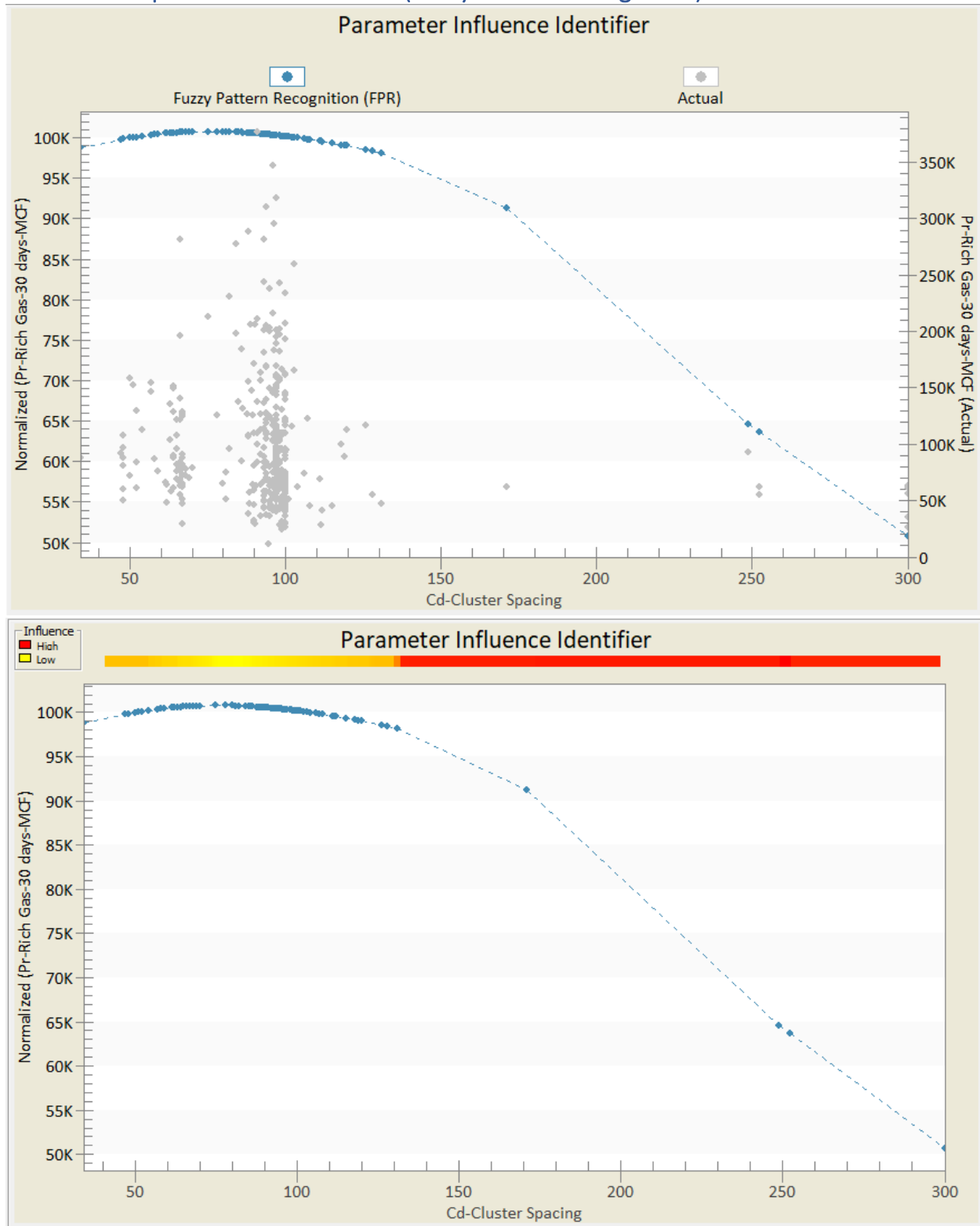


Figure 47: Parameter Influence (Fuzzy Pattern Recognition) - Cluster Spacing

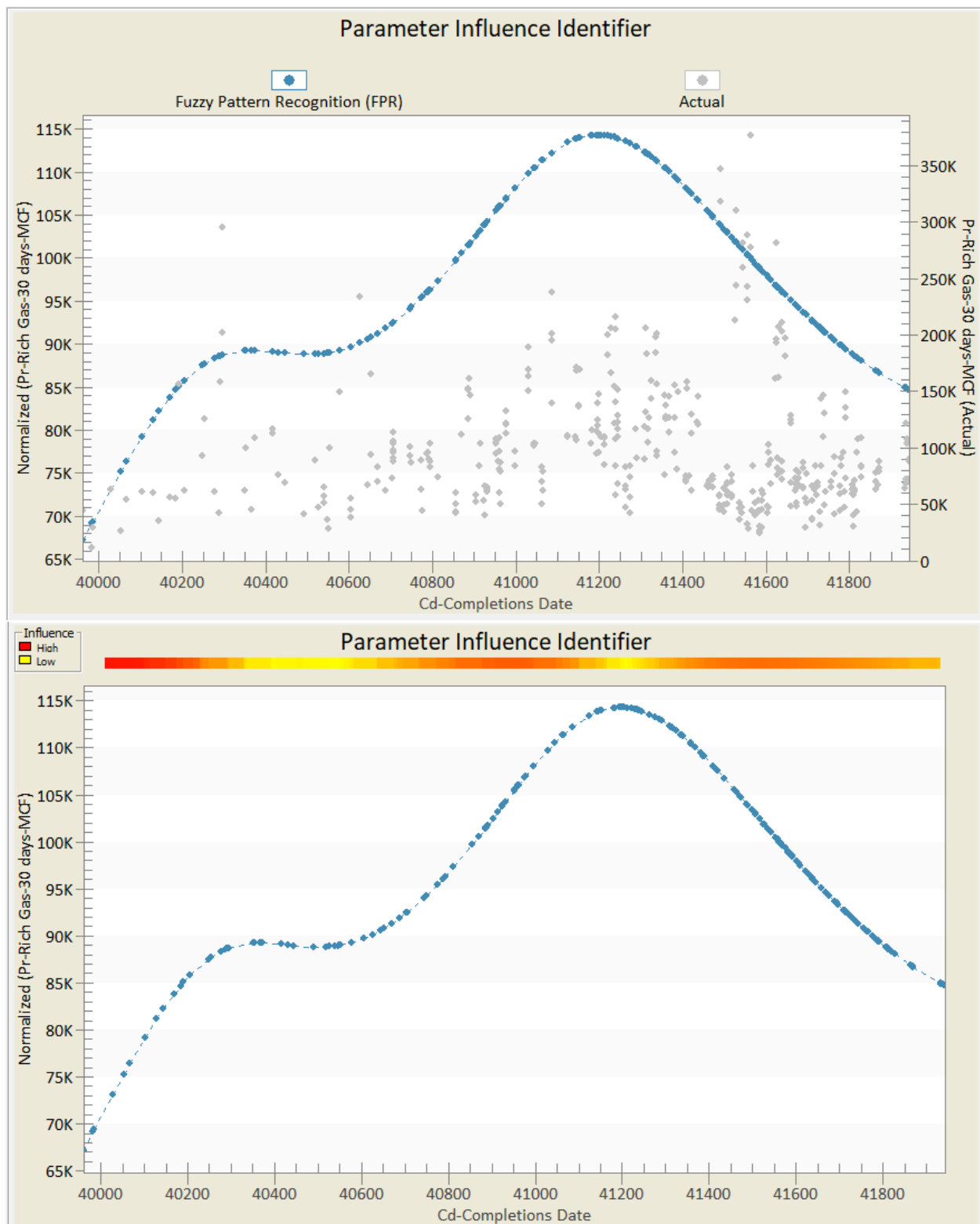


Figure 48: Parameter Influence (Fuzzy Pattern Recognition) - Completions Date

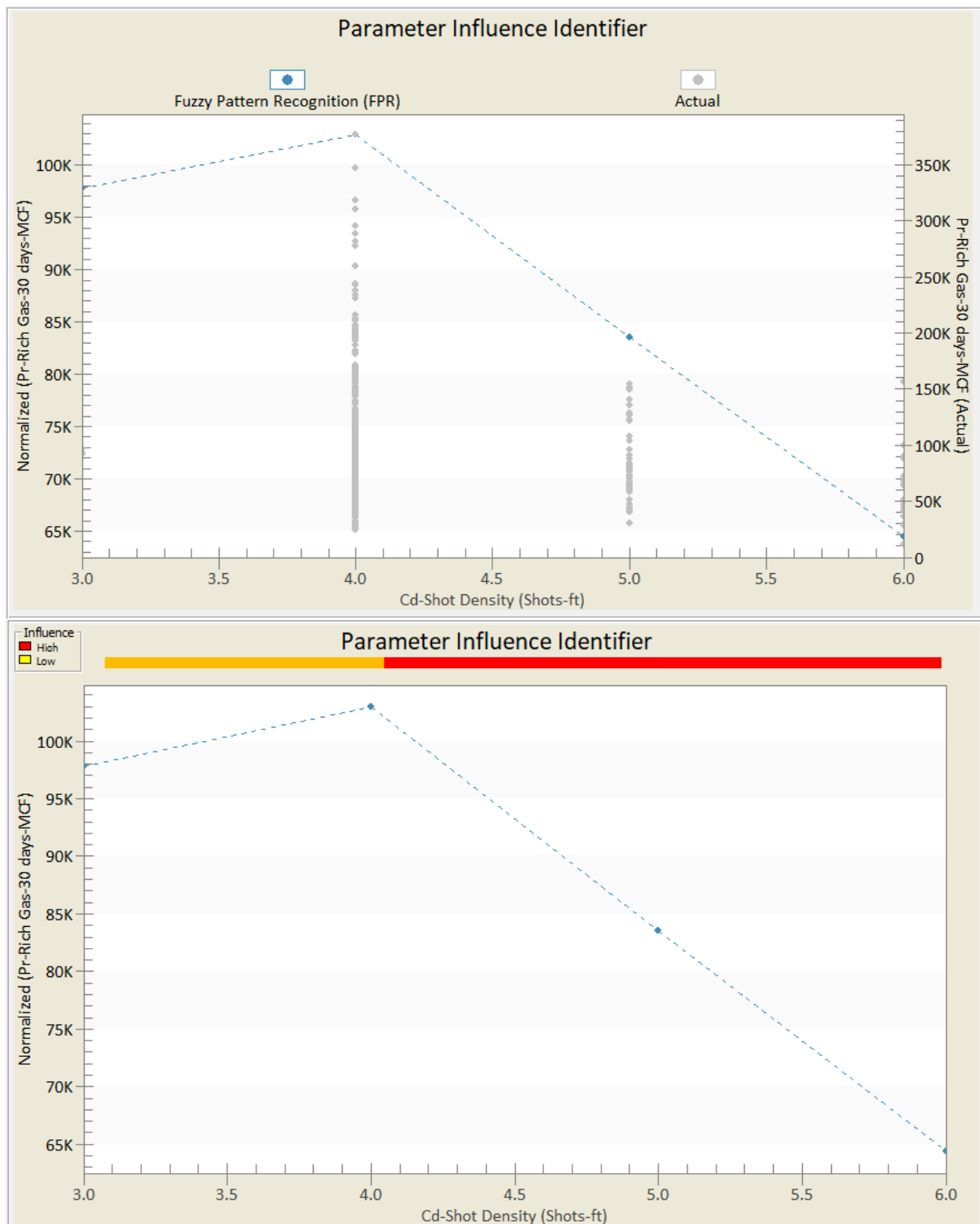


Figure 49: Parameter Influence (Fuzzy Pattern Recognition) - Shot Density

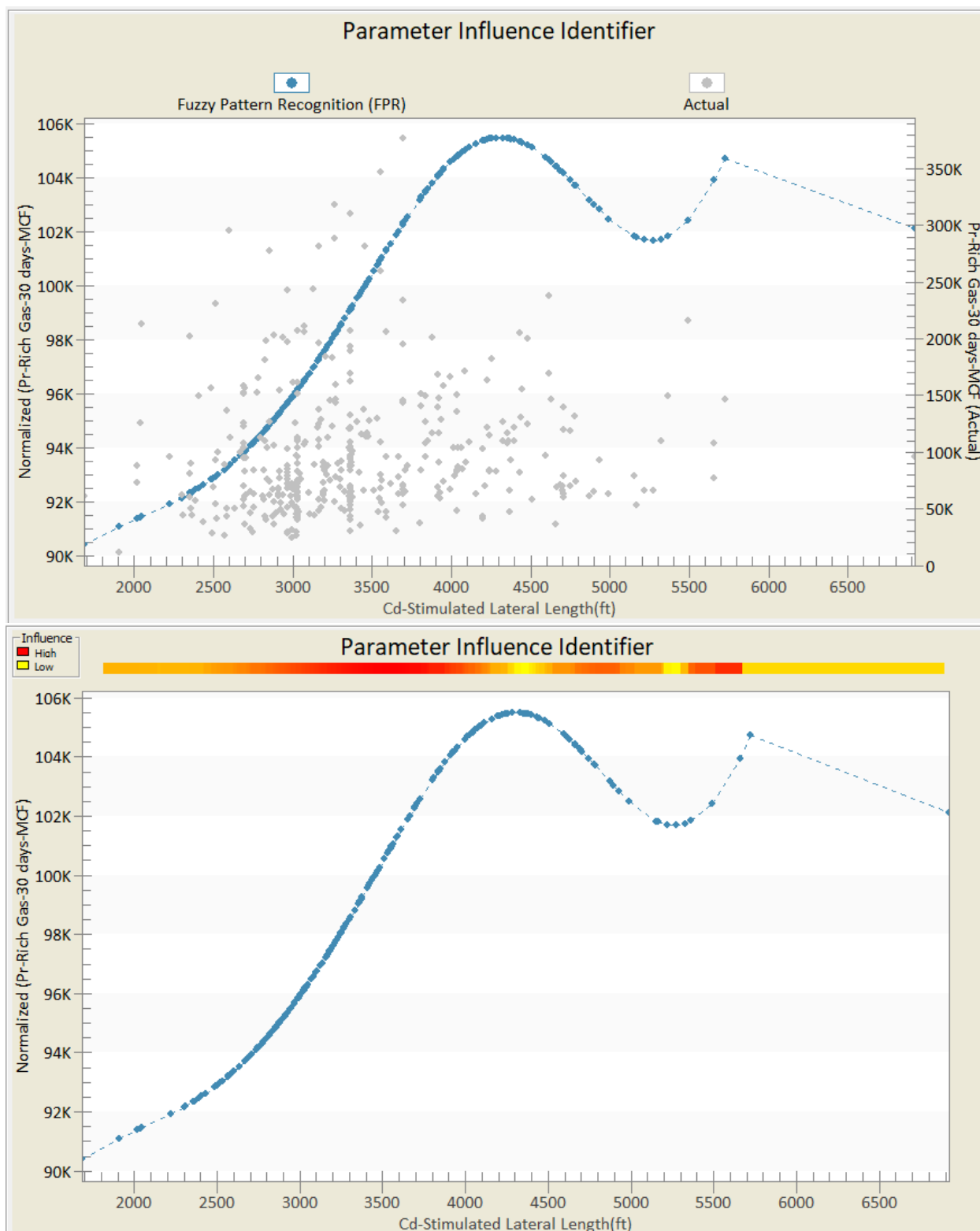


Figure 50: Parameter Influence (Fuzzy Pattern Recognition) - Stimulated Lateral Length

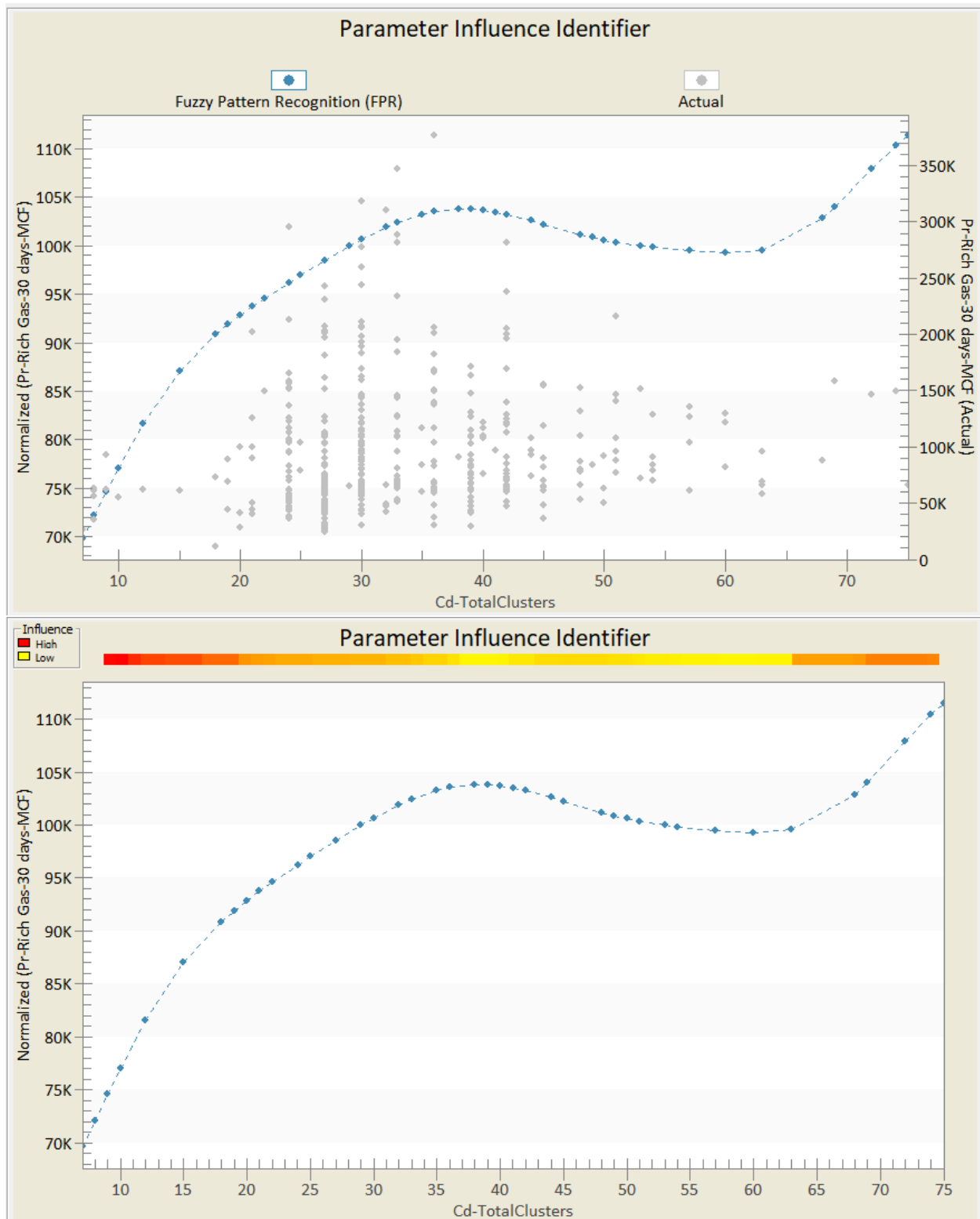


Figure 51: Parameter Influence (Fuzzy Pattern Recognition) - Total Clusters

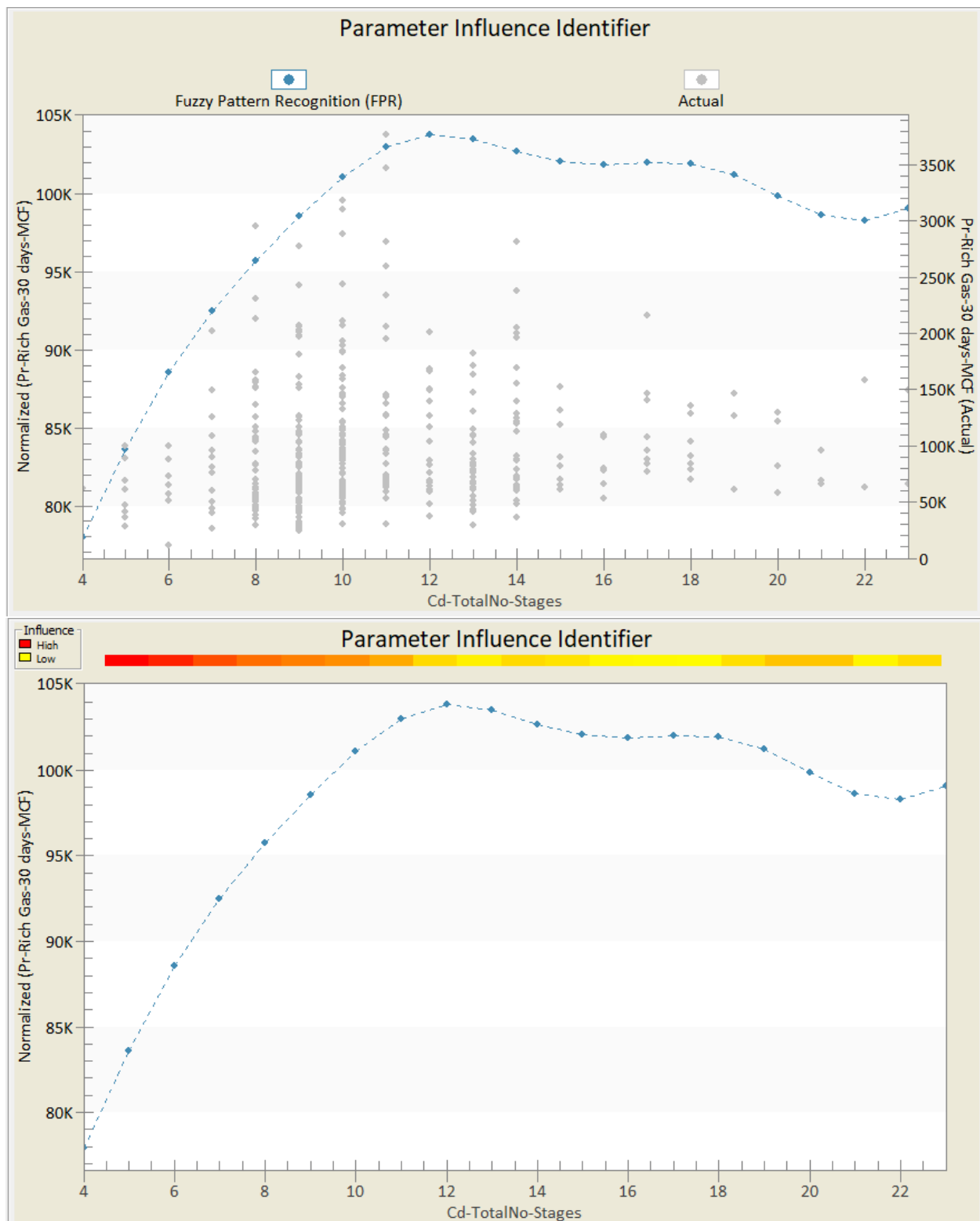


Figure 52: Parameter Influence (Fuzzy Pattern Recognition) - Total No. Stages

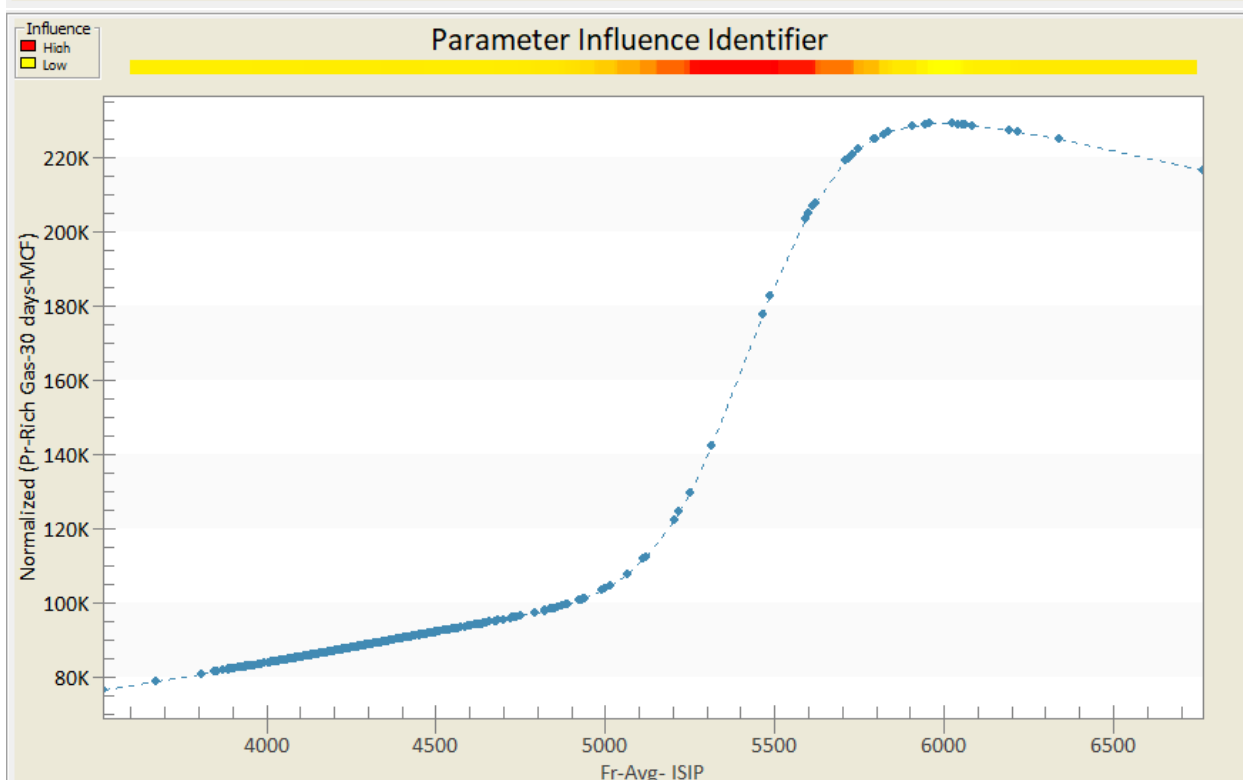
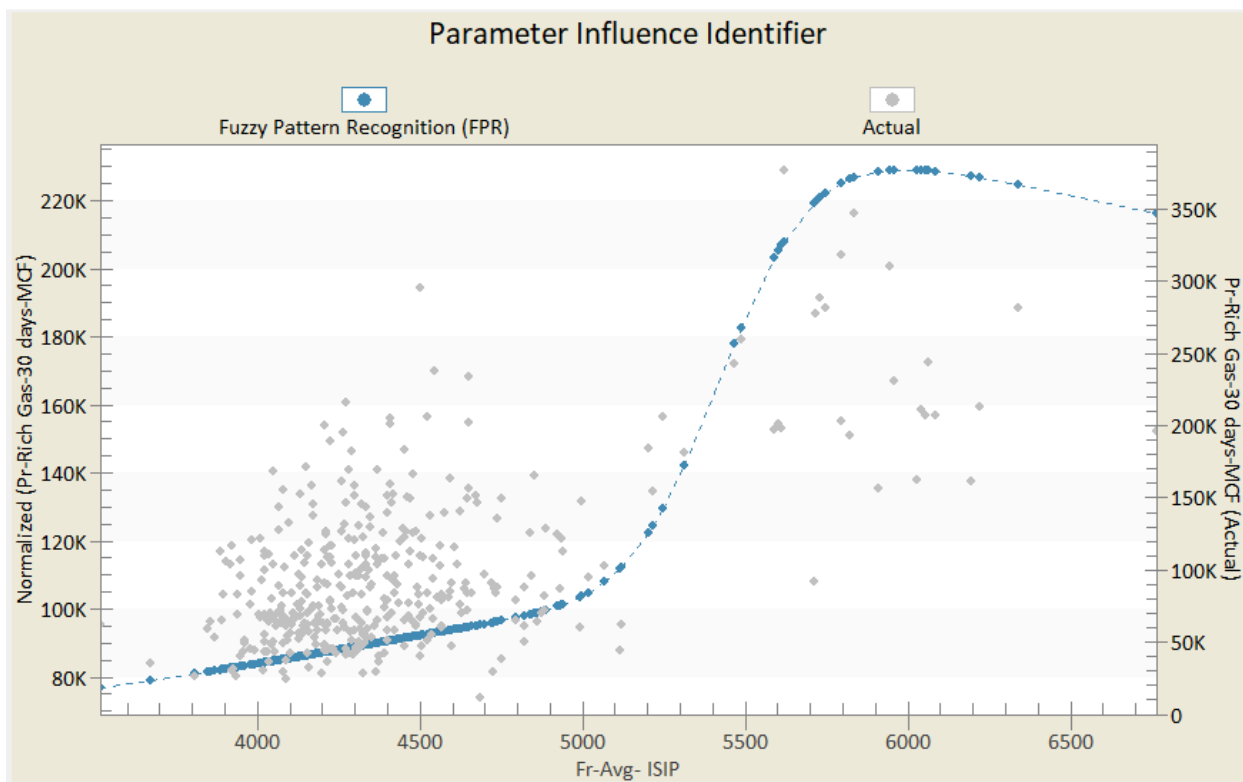


Figure 53: Parameter Influence (Fuzzy Pattern Recognition) - Average ISIP

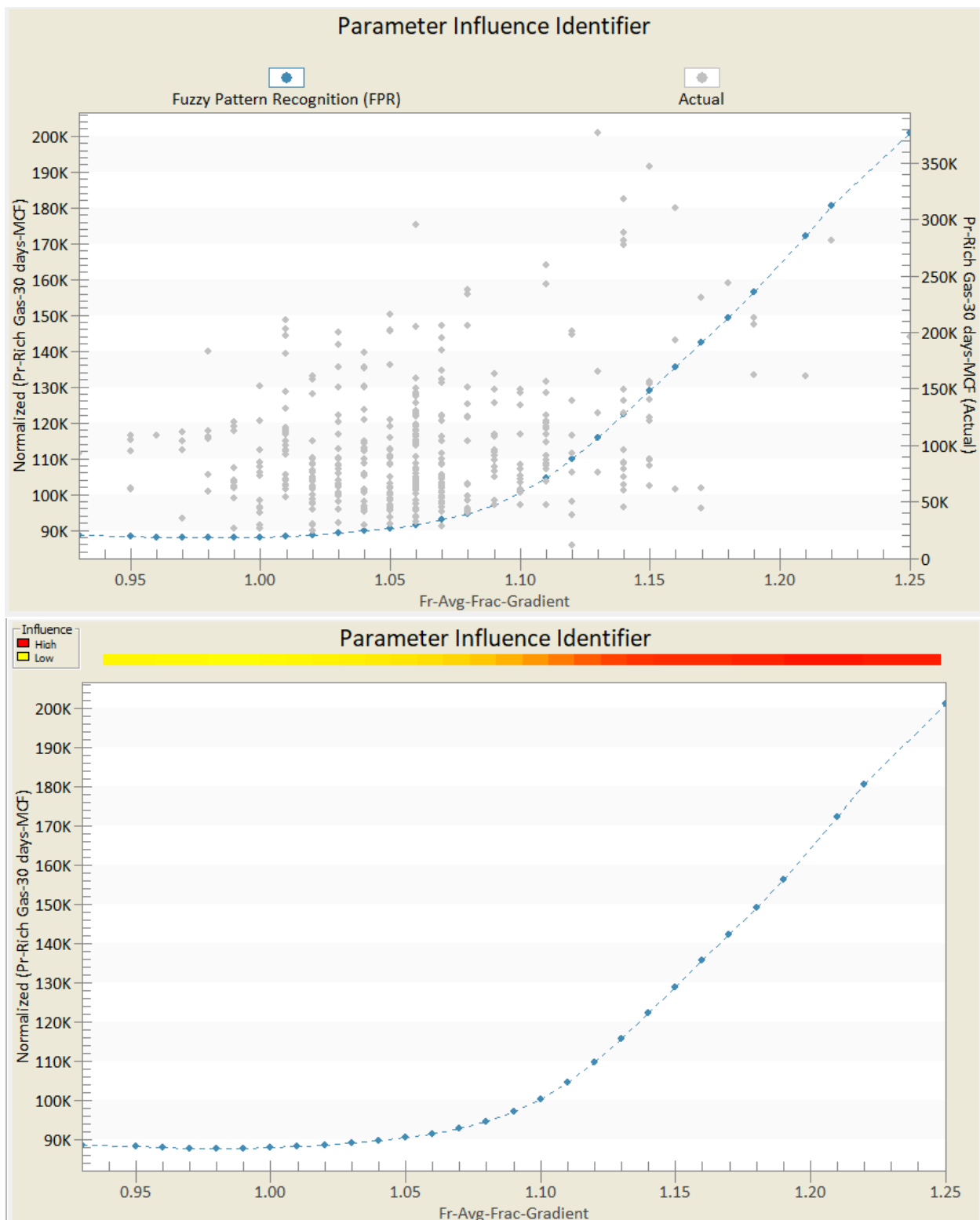


Figure 54: Parameter Influence (Fuzzy Pattern Recognition) - Average Fracture Gradient

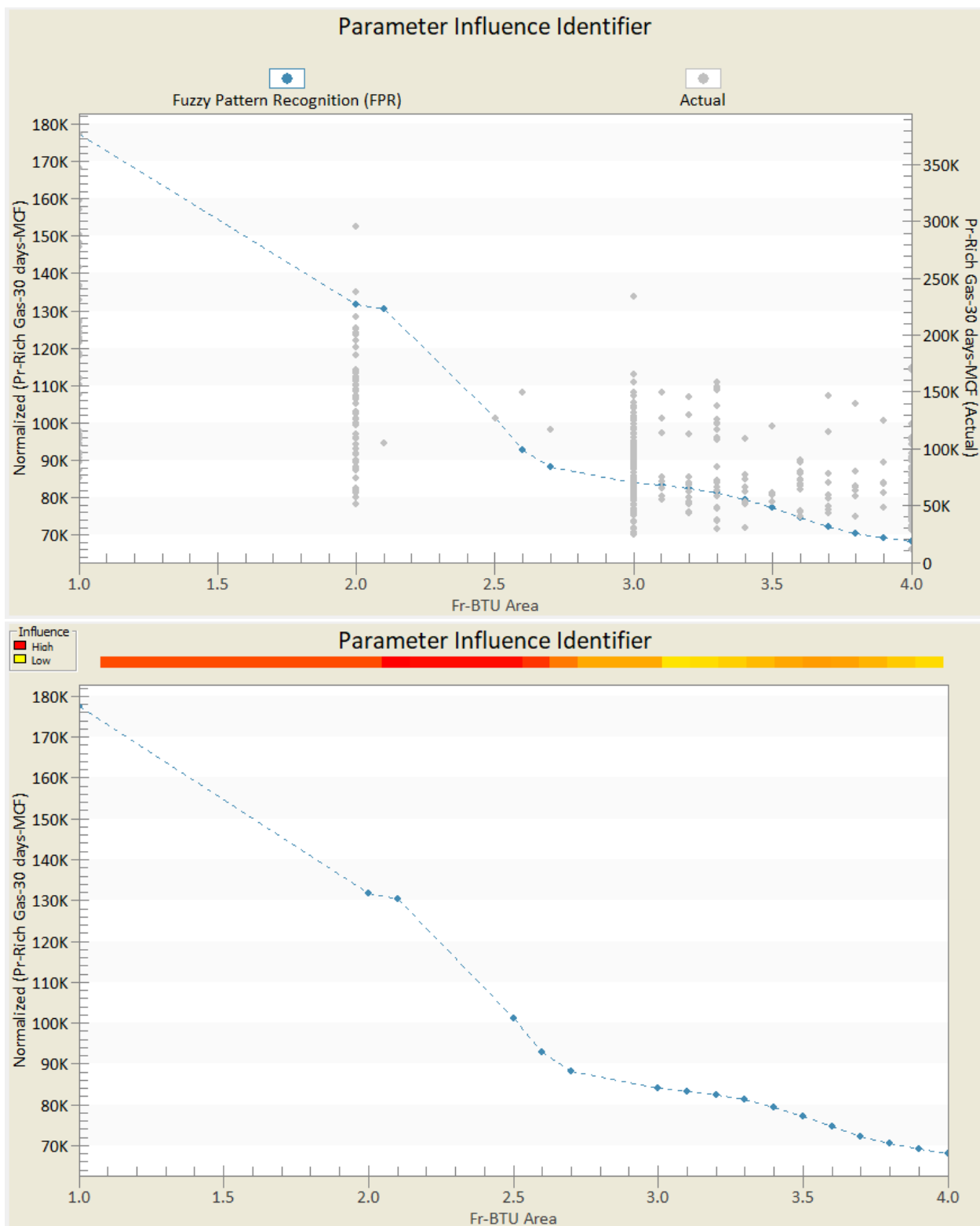


Figure 55: Parameter Influence (Fuzzy Pattern Recognition) - BTU Area

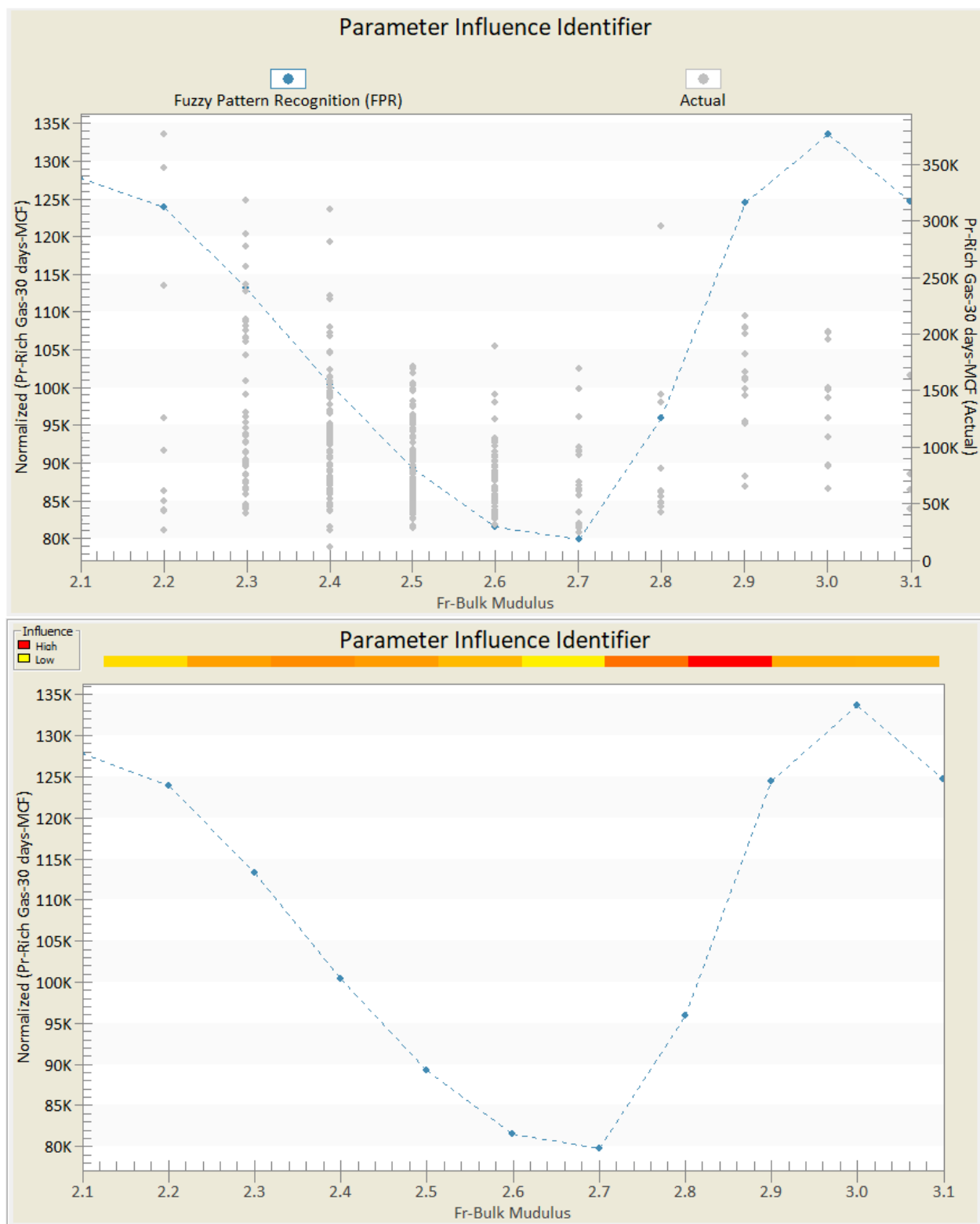


Figure 56: Parameter Influence (Fuzzy Pattern Recognition) - Bulk Modulus

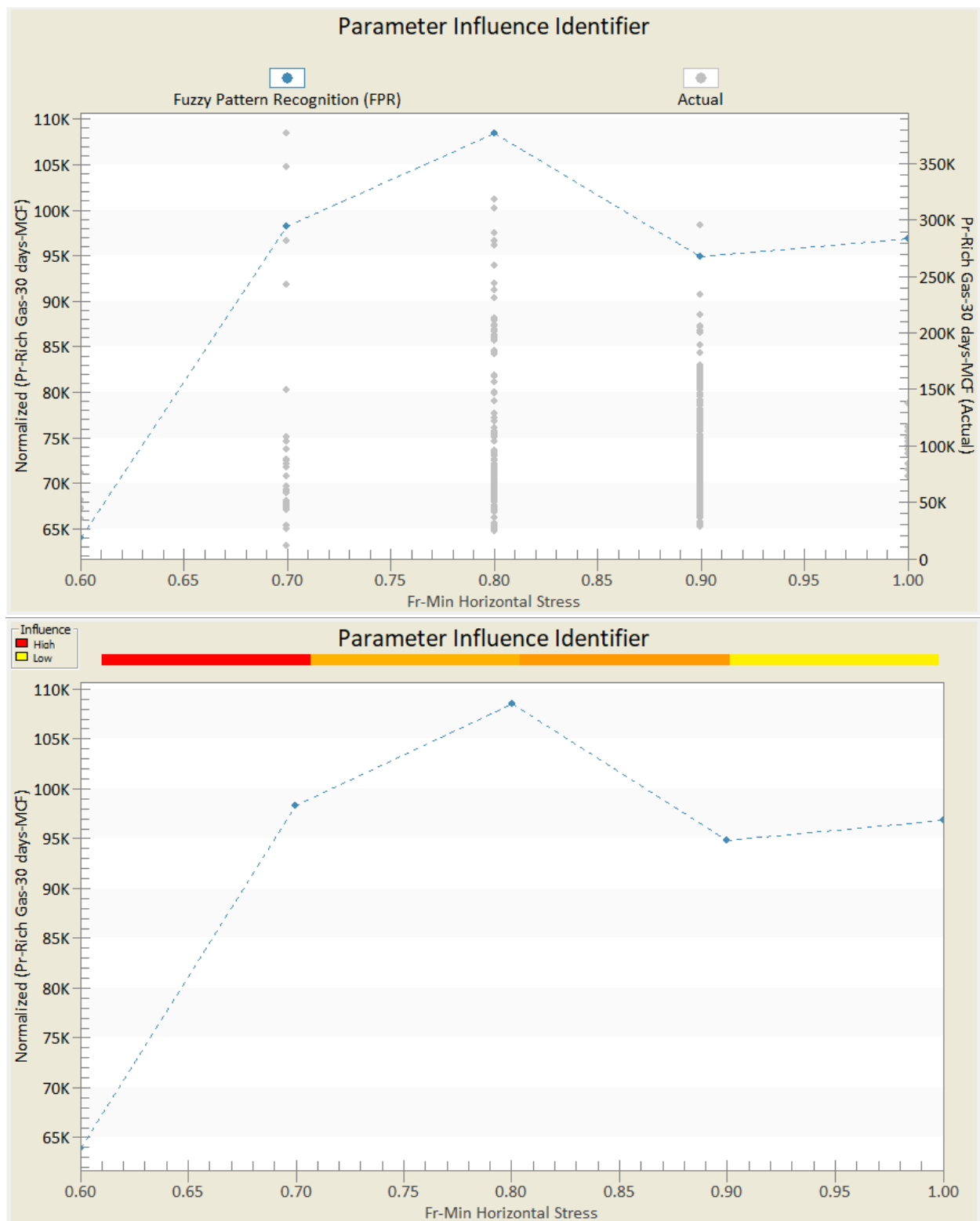


Figure 57: Parameter Influence (Fuzzy Pattern Recognition) - Minimum Horizontal Stress

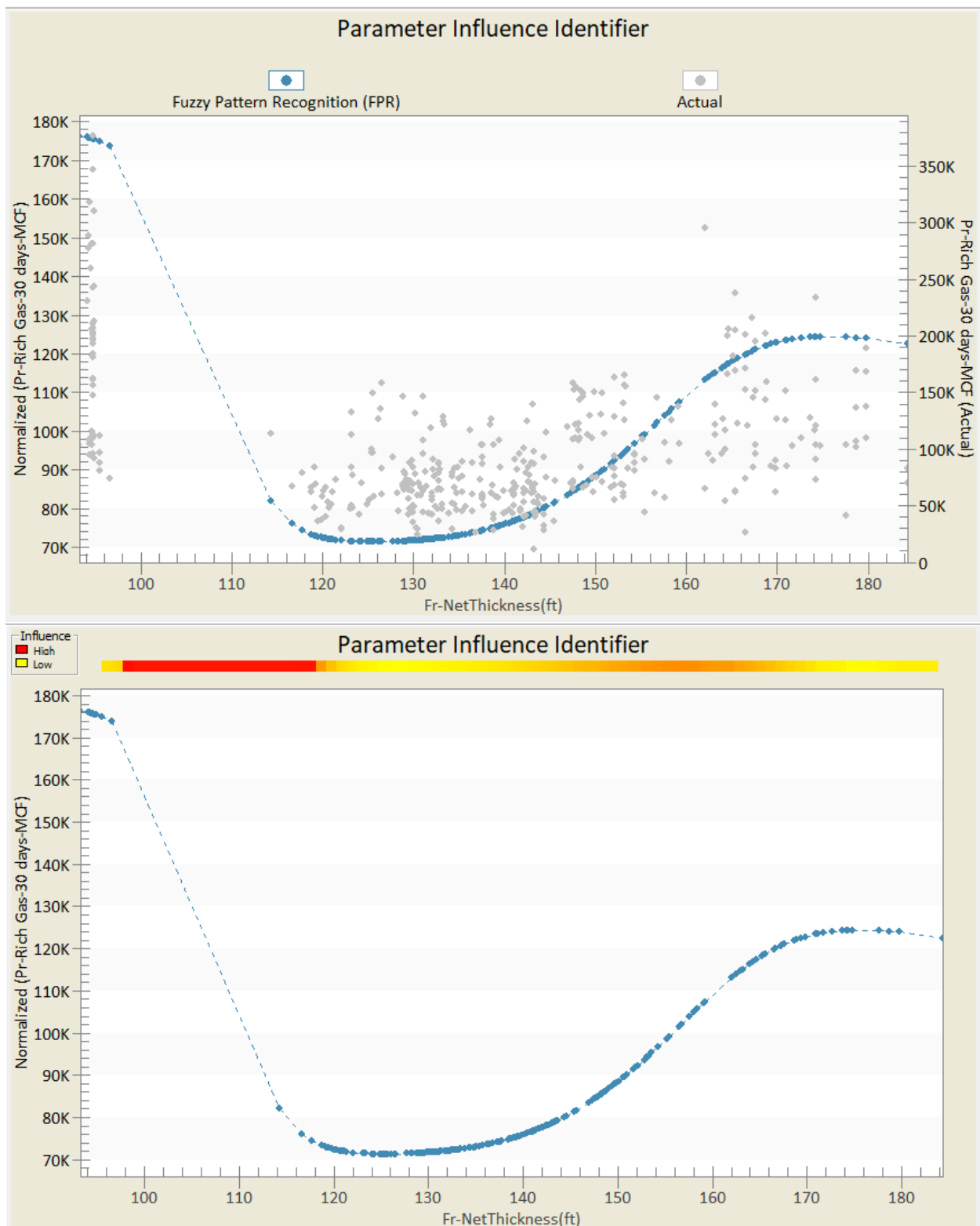


Figure 58: Parameter Influence (Fuzzy Pattern Recognition) - Net Thickness

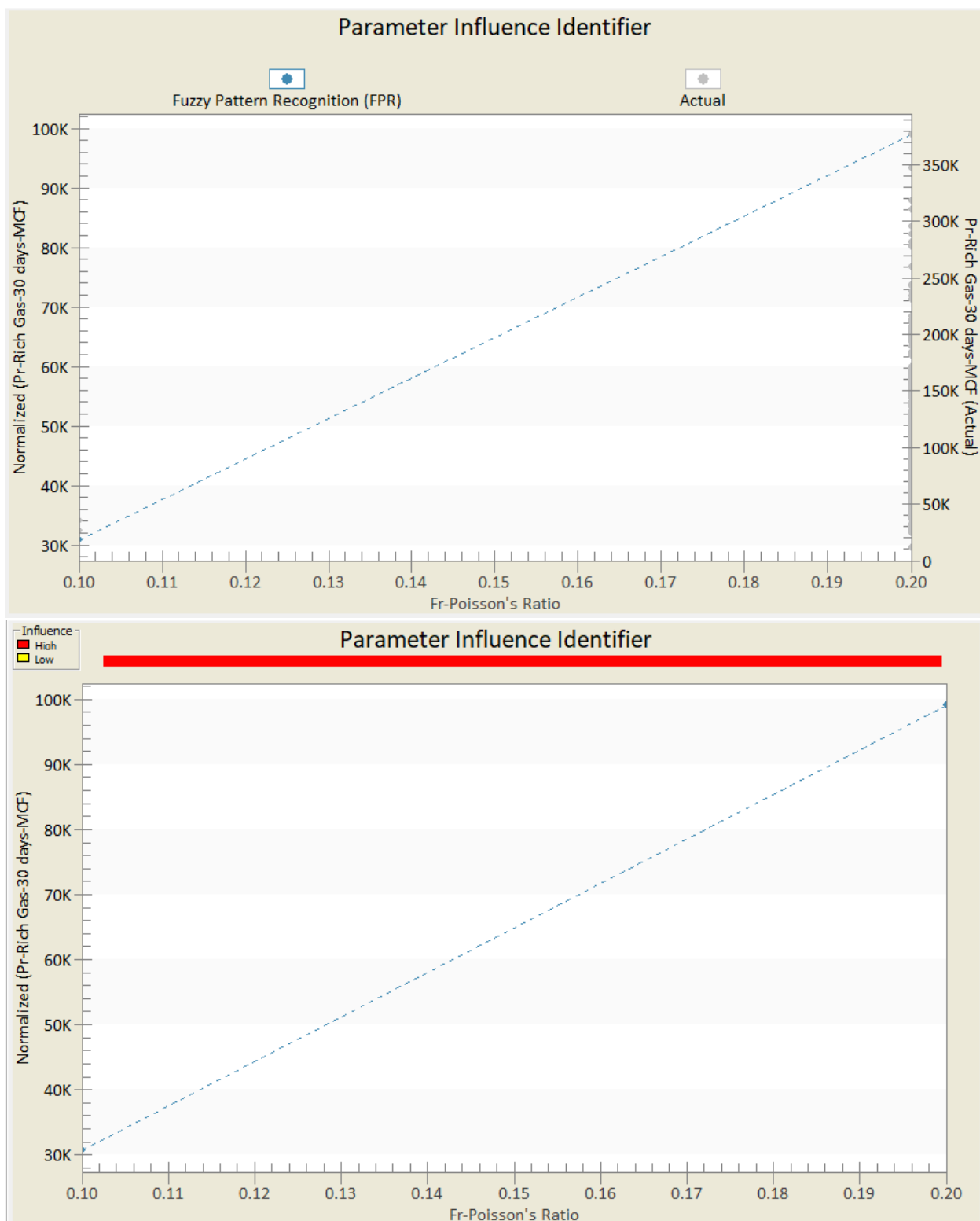


Figure 59: Parameter Influence (Fuzzy Pattern Recognition) - Poisson's Ratio

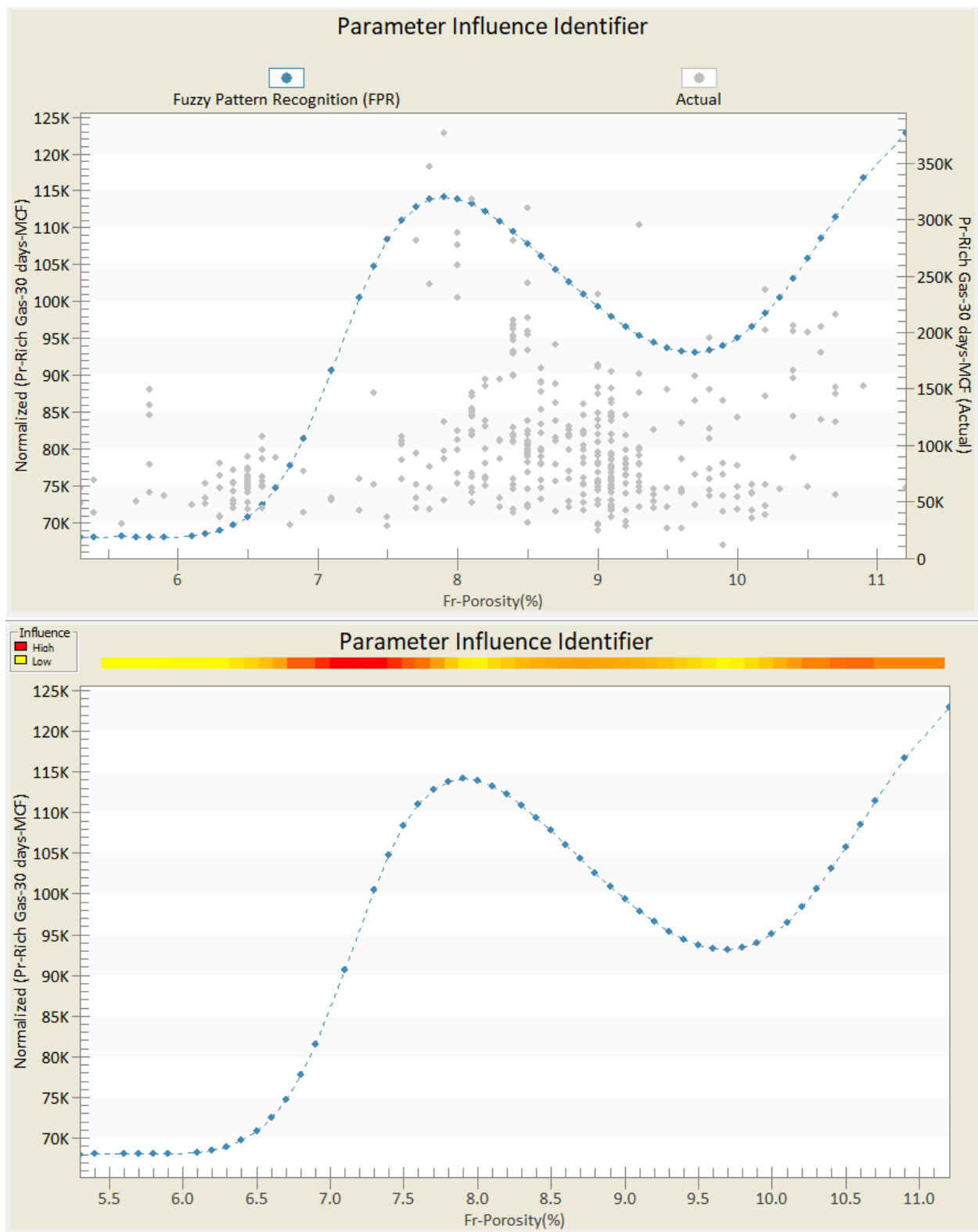


Figure 60: Parameter Influence (Fuzzy Pattern Recognition) - Porosity

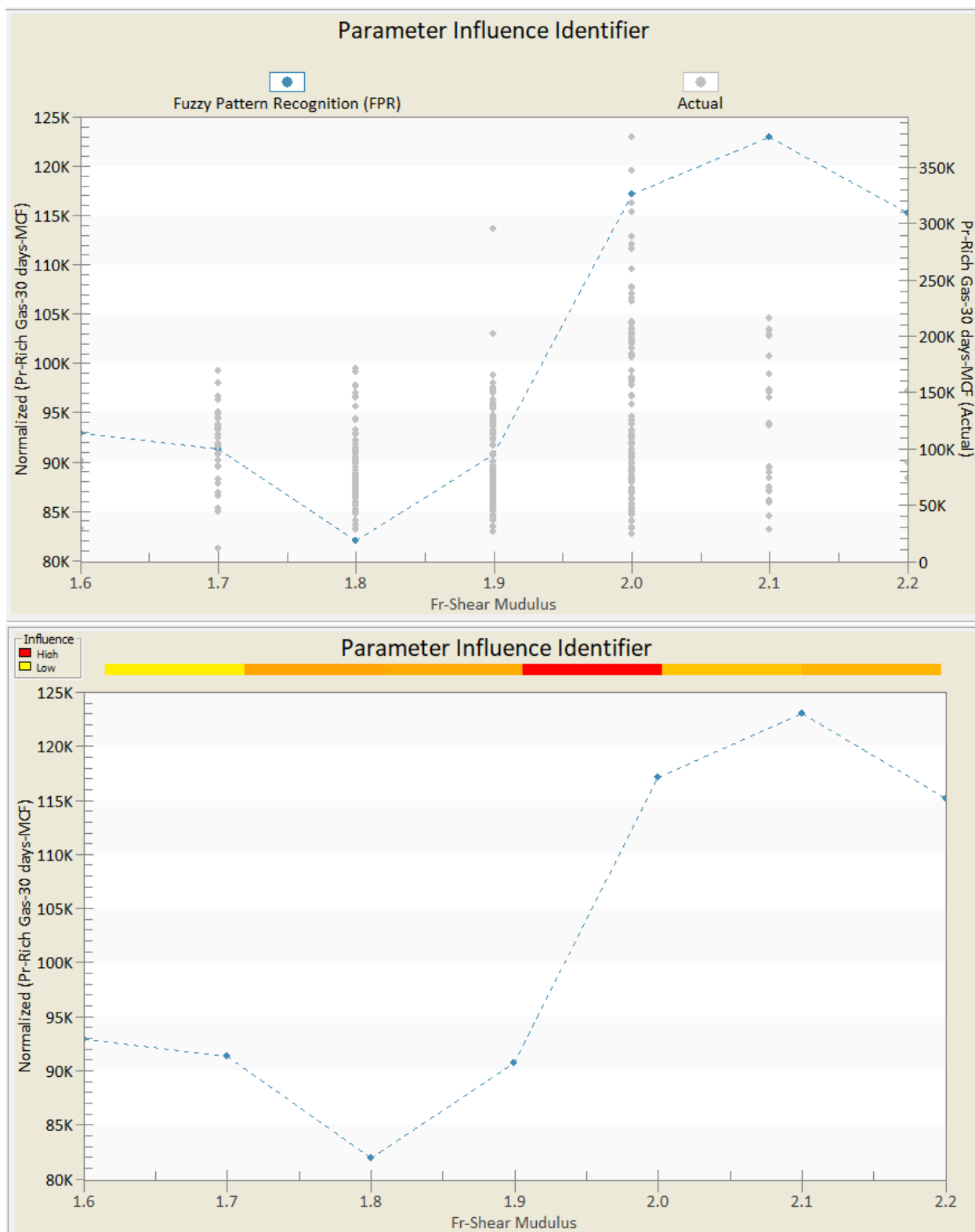


Figure 61: Parameter Influence (Fuzzy Pattern Recognition) - Shear Modulus

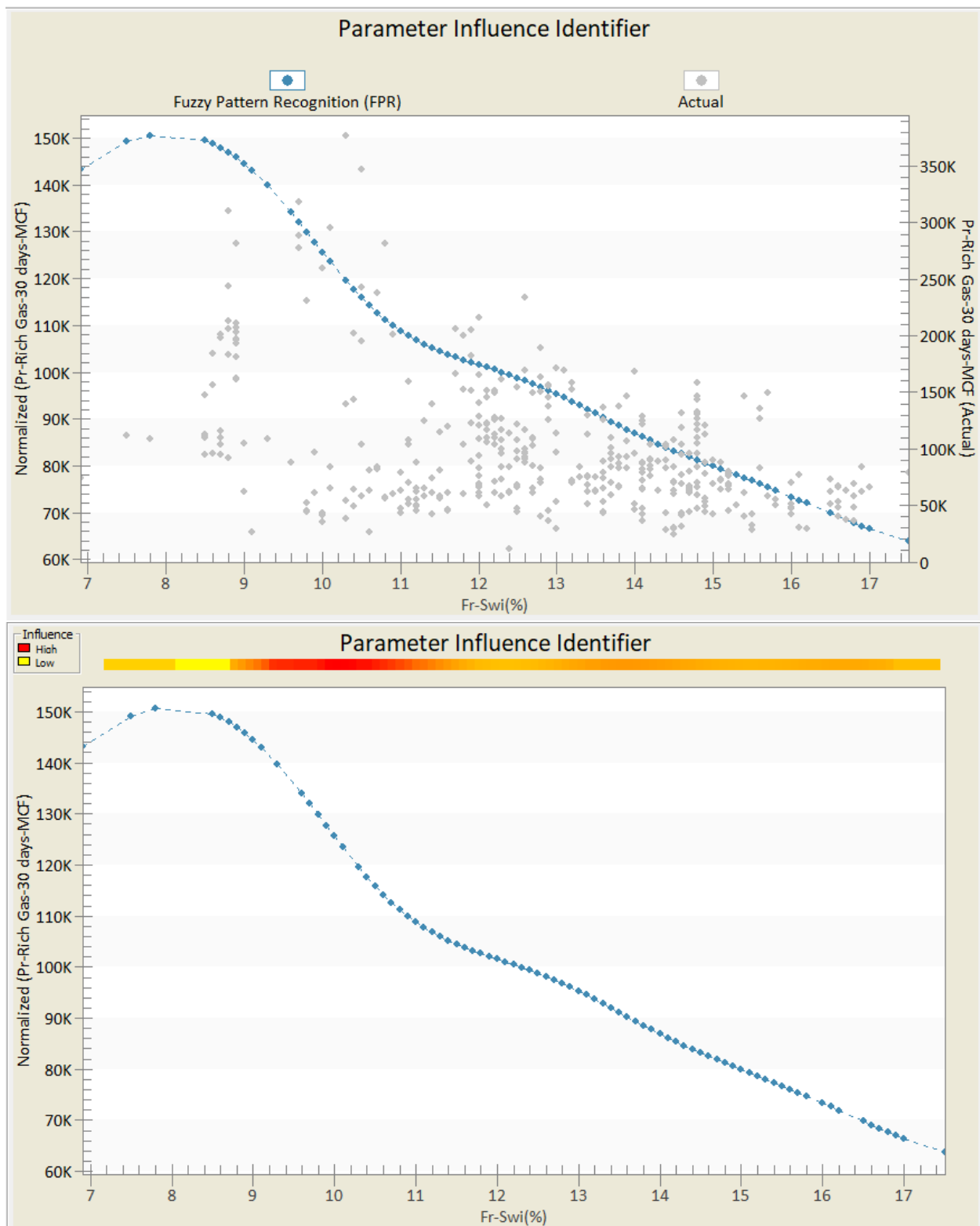


Figure 62: Parameter Influence (Fuzzy Pattern Recognition) - Initial Water Saturation (Swi %)

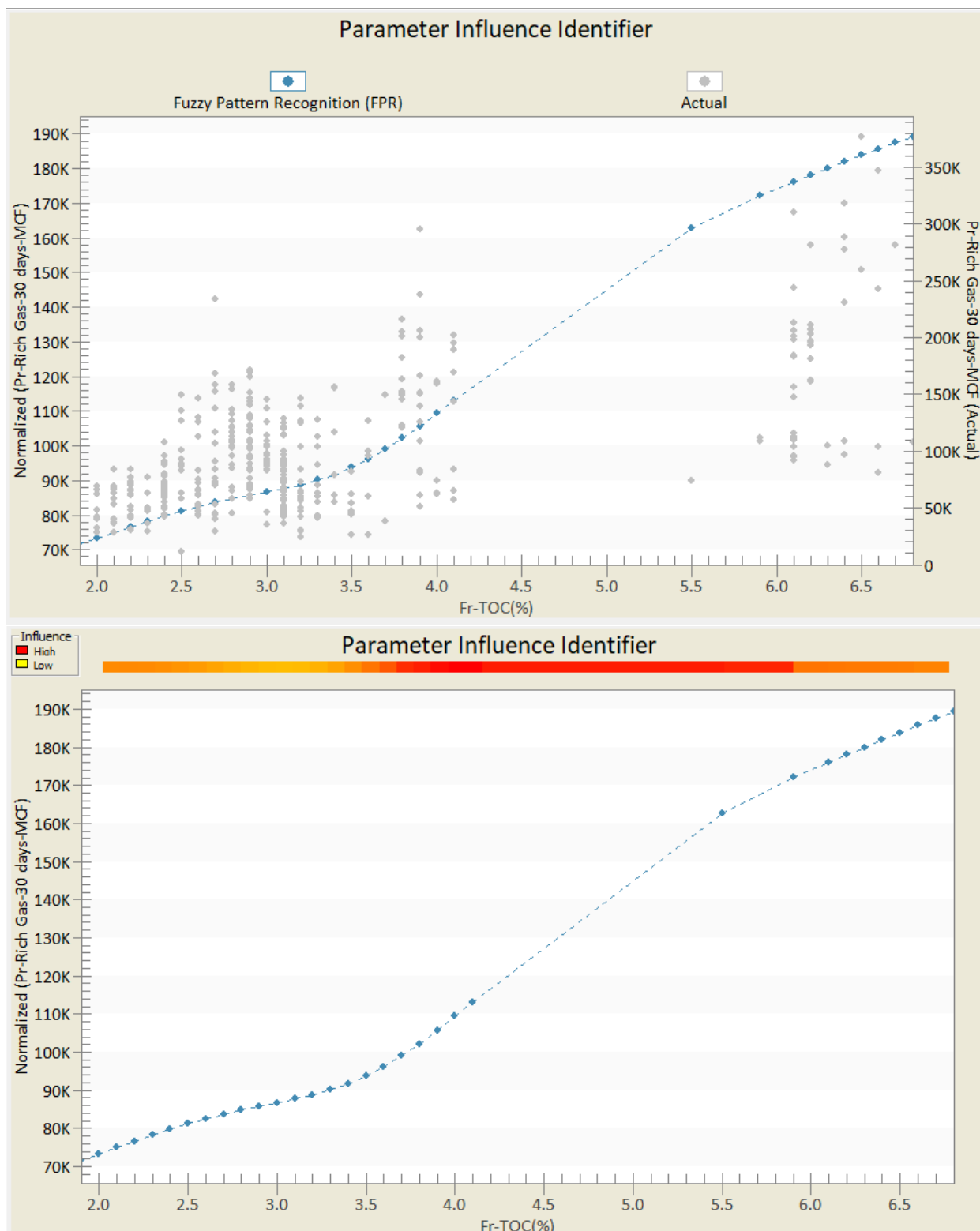


Figure 63: Parameter Influence (Fuzzy Pattern Recognition) - Total Organic Content (TOC %)

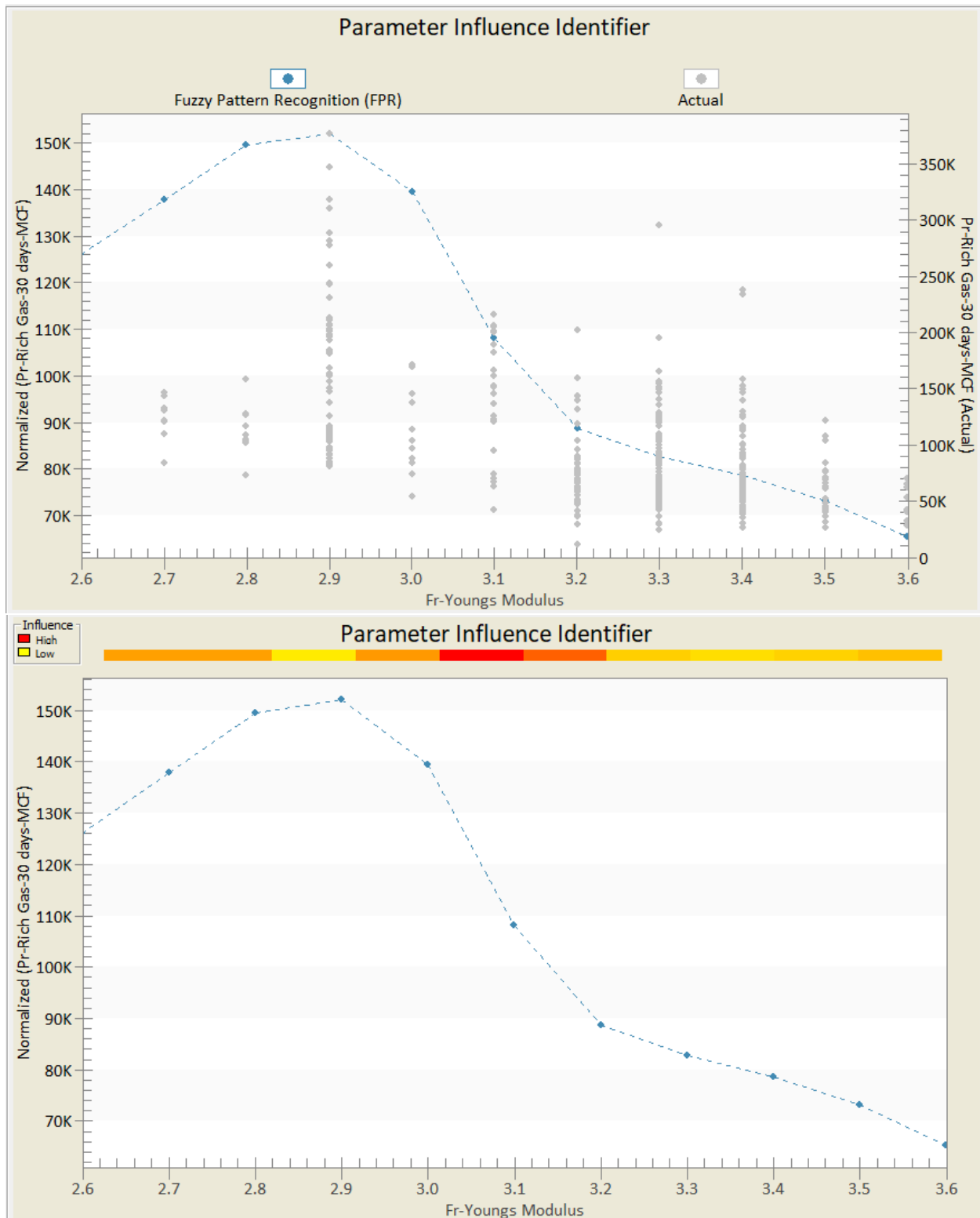


Figure 64: Parameter Influence (Fuzzy Pattern Recognition) - Young's Modulus

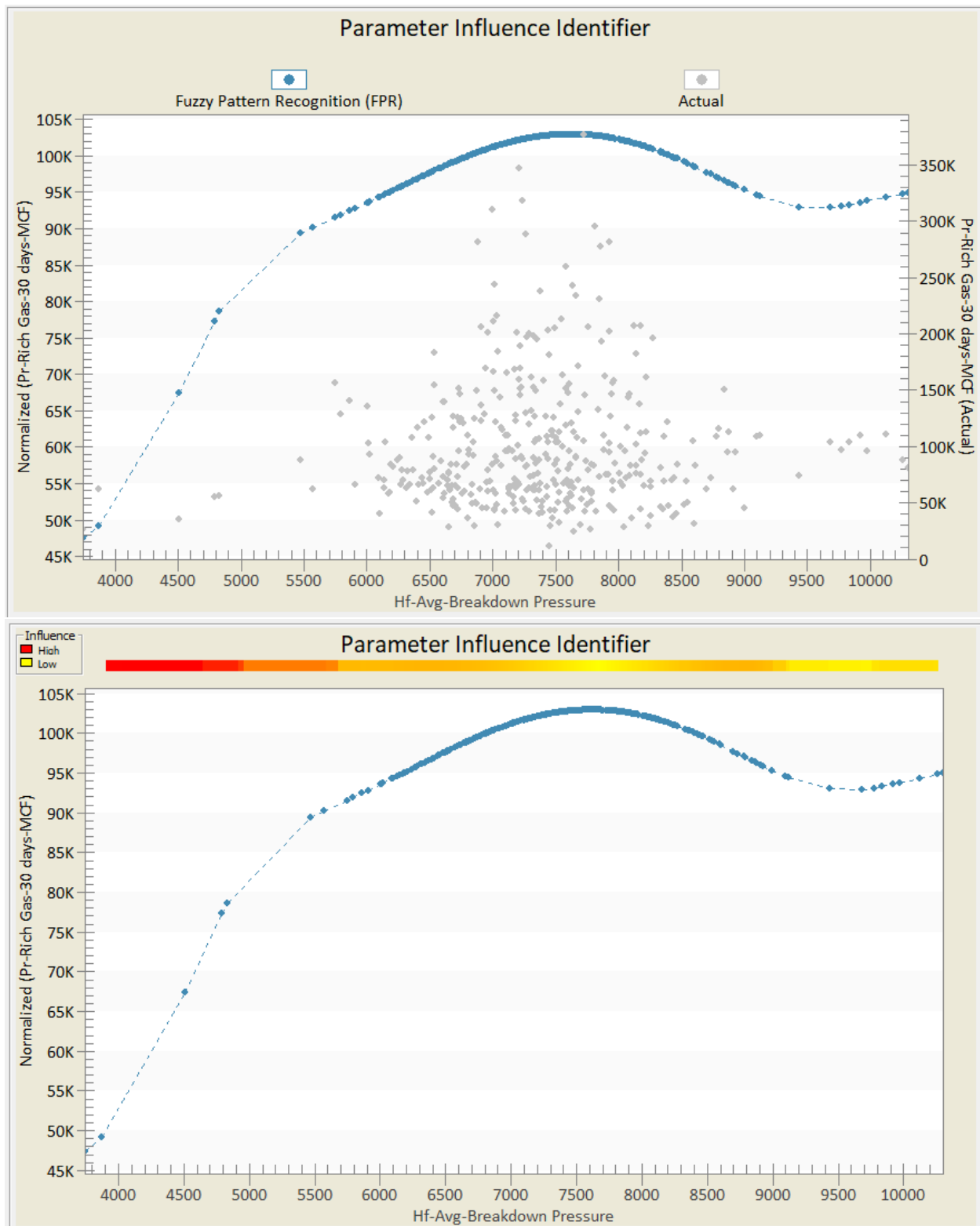


Figure 65: Parameter Influence (Fuzzy Pattern Recognition) - Average Breakdown Pressure

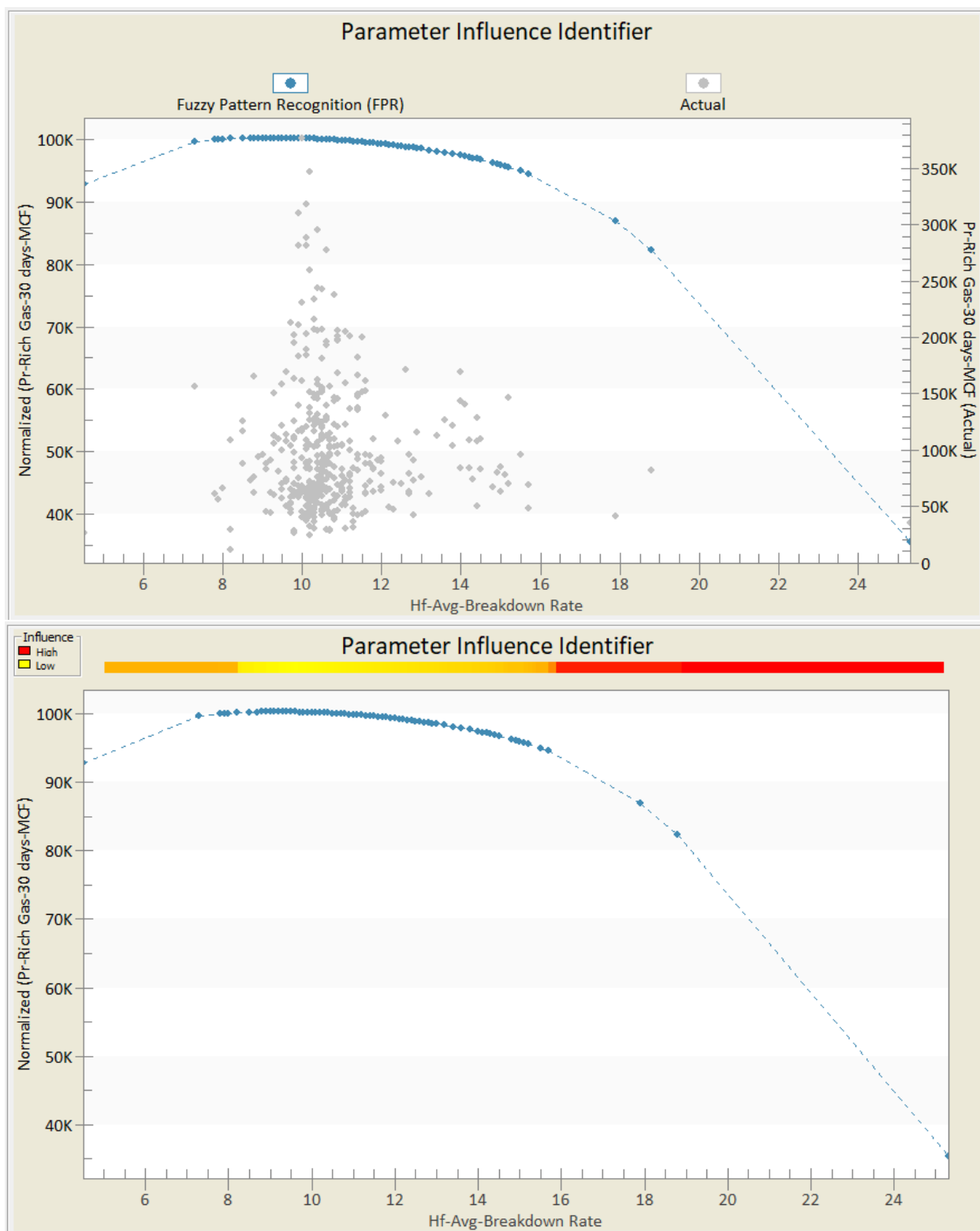


Figure 66: Parameter Influence (Fuzzy Pattern Recognition) - Average Breakdown Rate

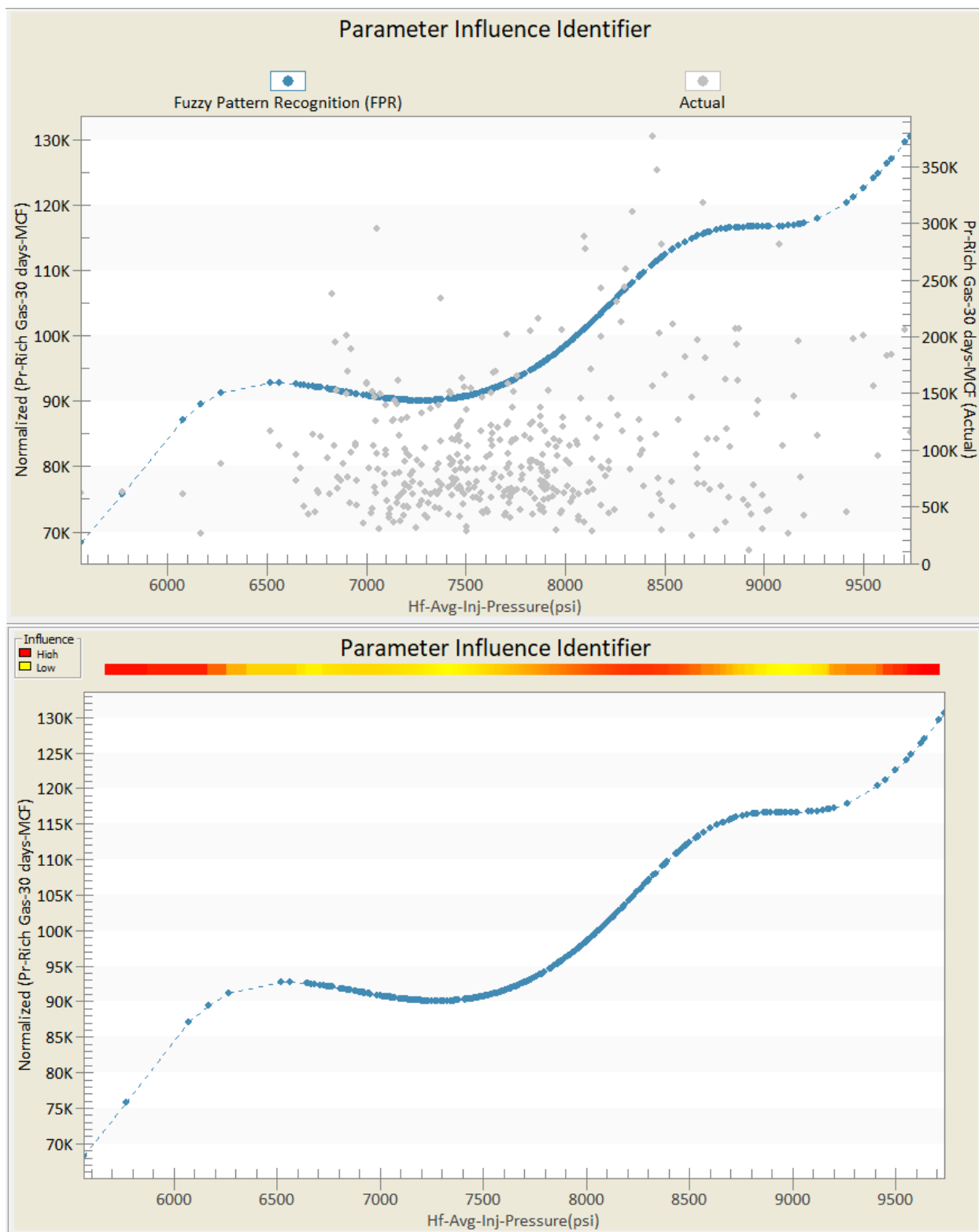


Figure 67: Parameter Influence (Fuzzy Pattern Recognition) - Average Injection Pressure

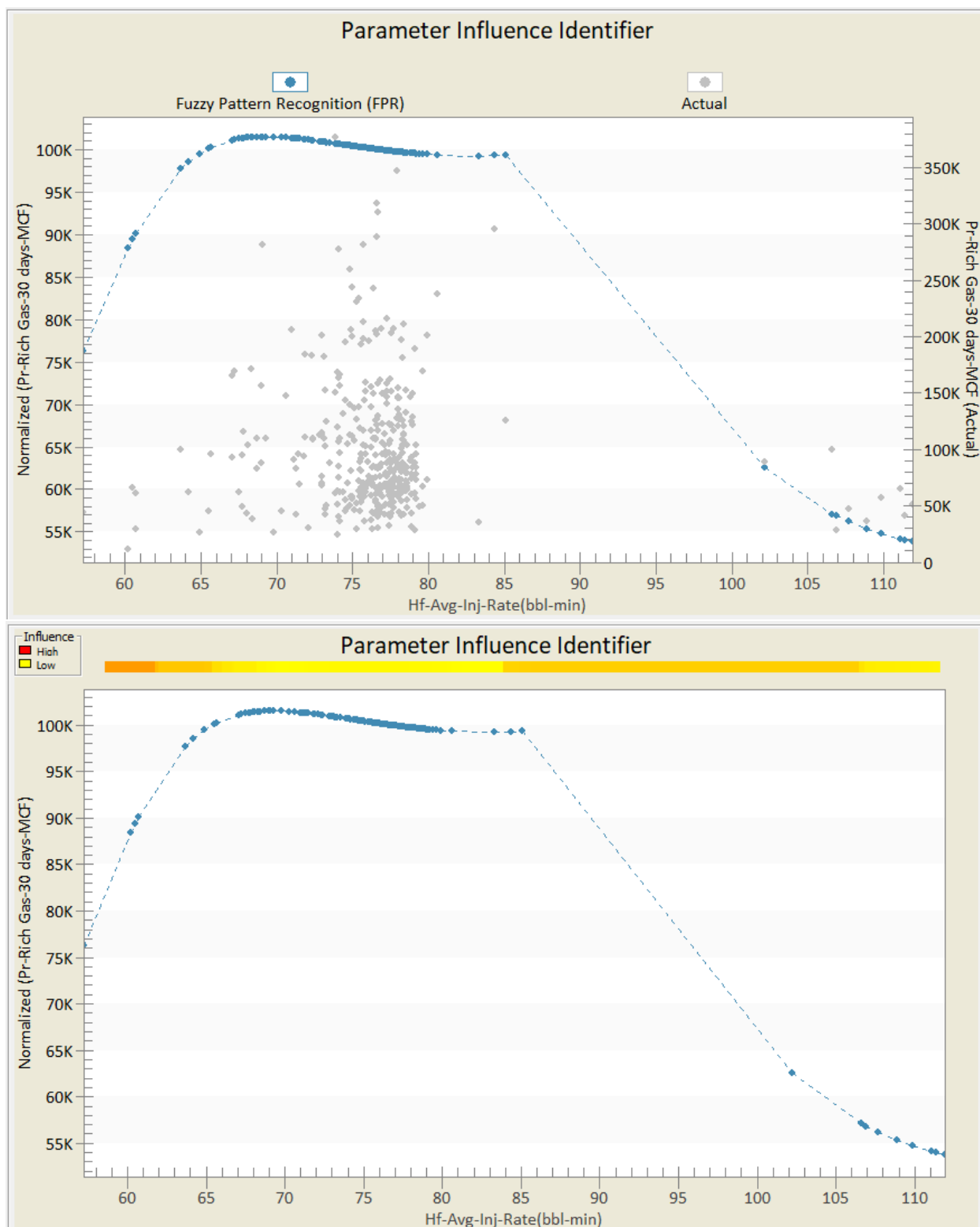


Figure 68: Parameter Influence (Fuzzy Pattern Recognition) - Average Injection Rate

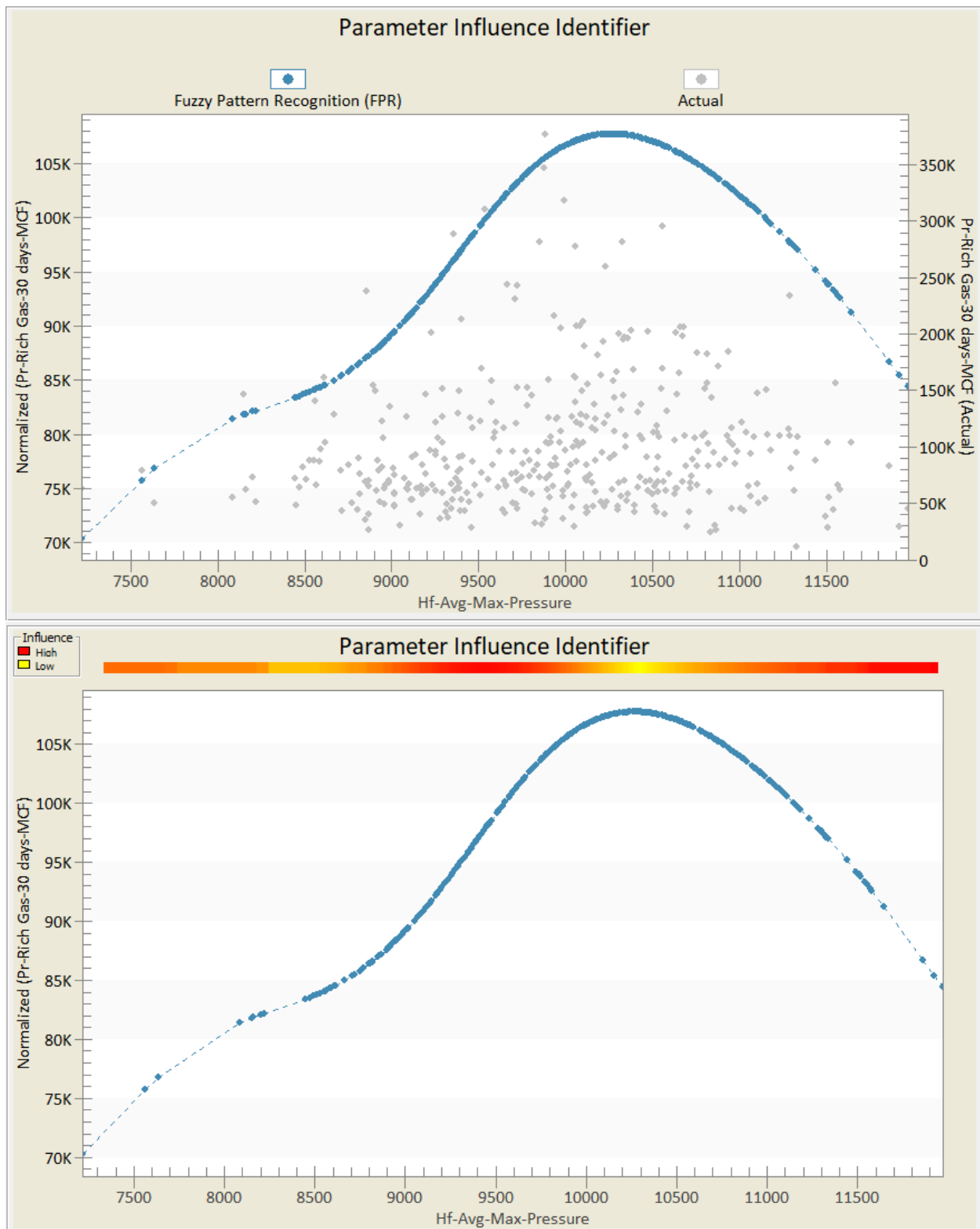


Figure 69: Parameter Influence (Fuzzy Pattern Recognition) - Average Max. Pressure

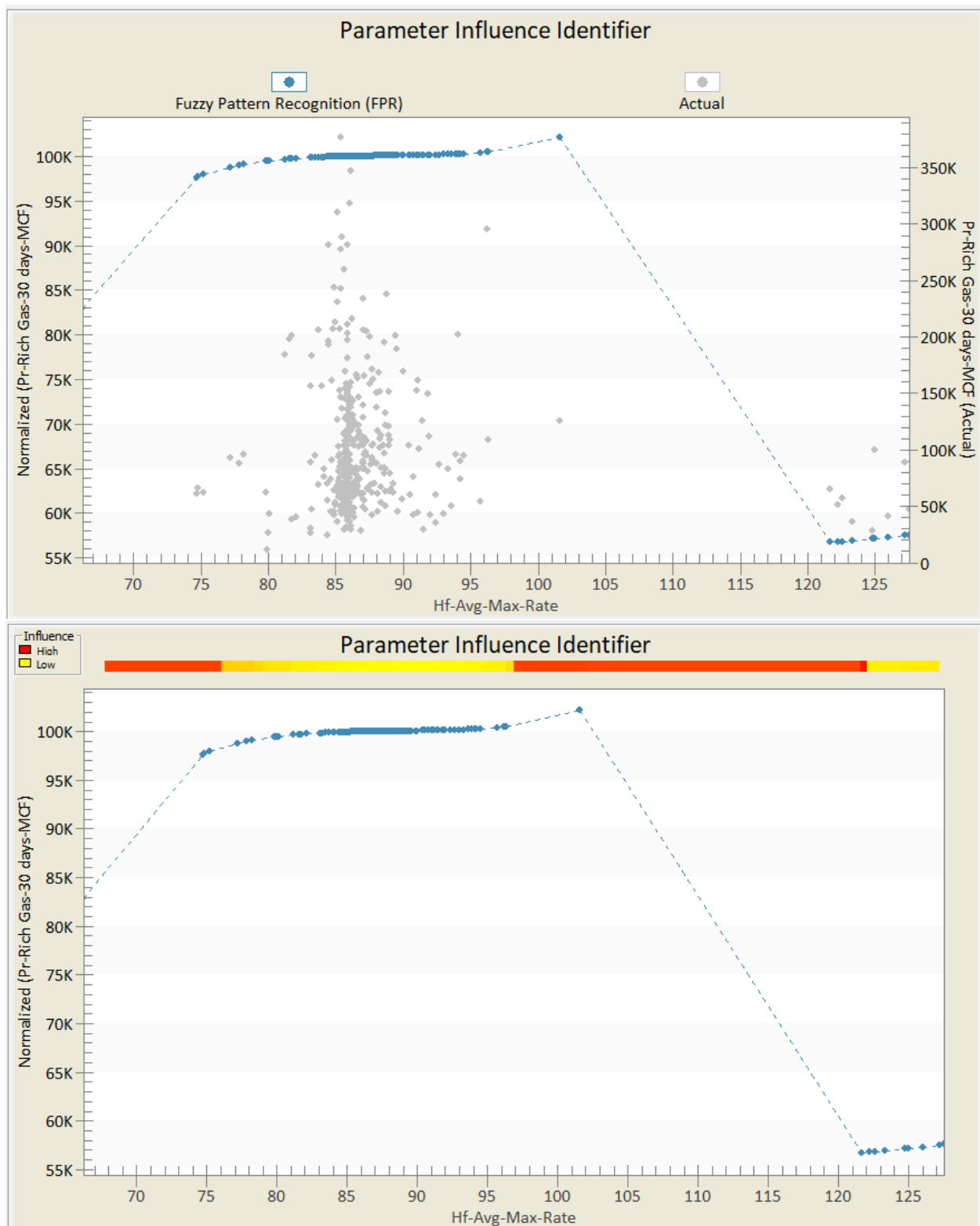


Figure 70: Parameter Influence (Fuzzy Pattern Recognition) - Average Max. Rate

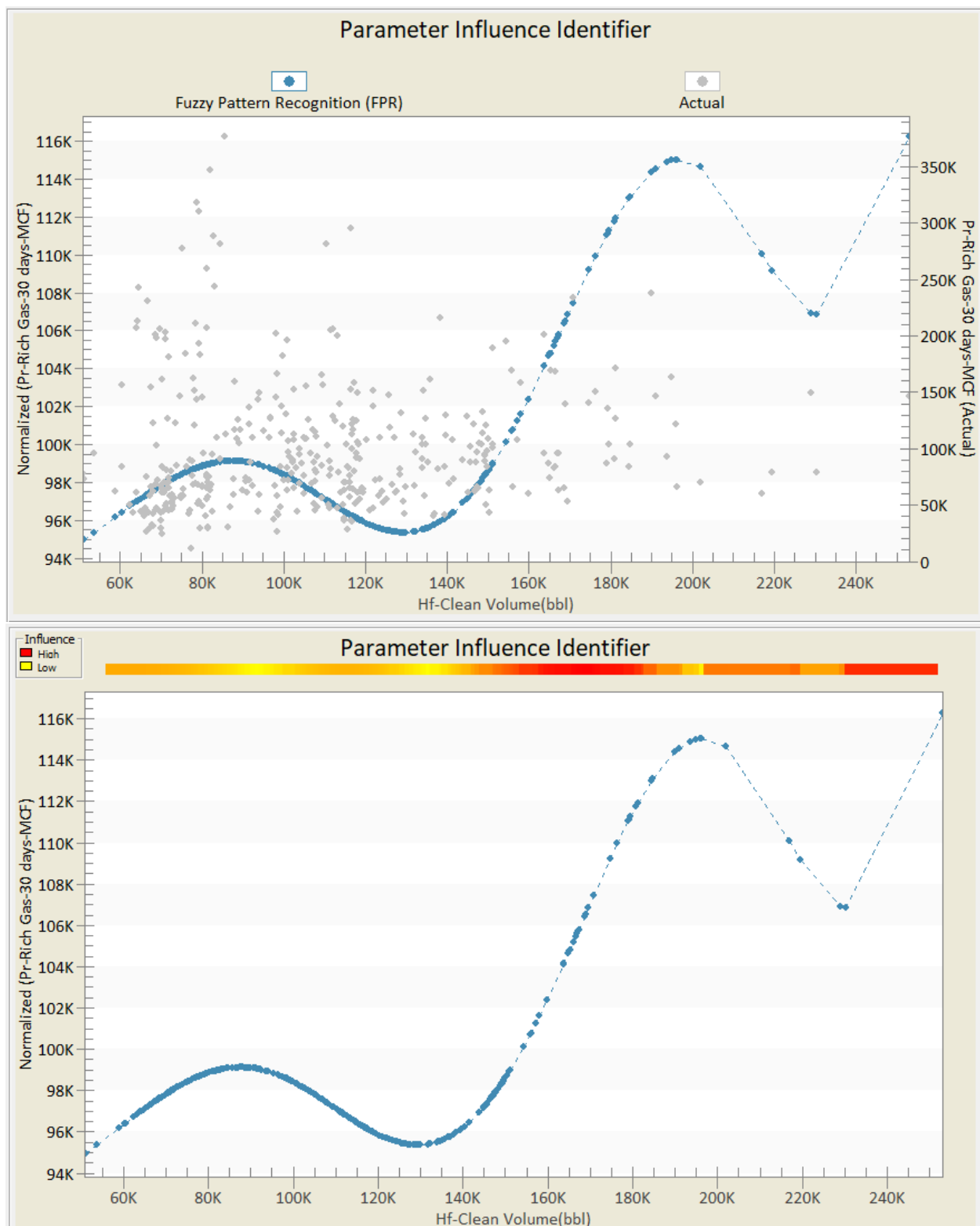


Figure 71: Parameter Influence (Fuzzy Pattern Recognition) - Clean Volume

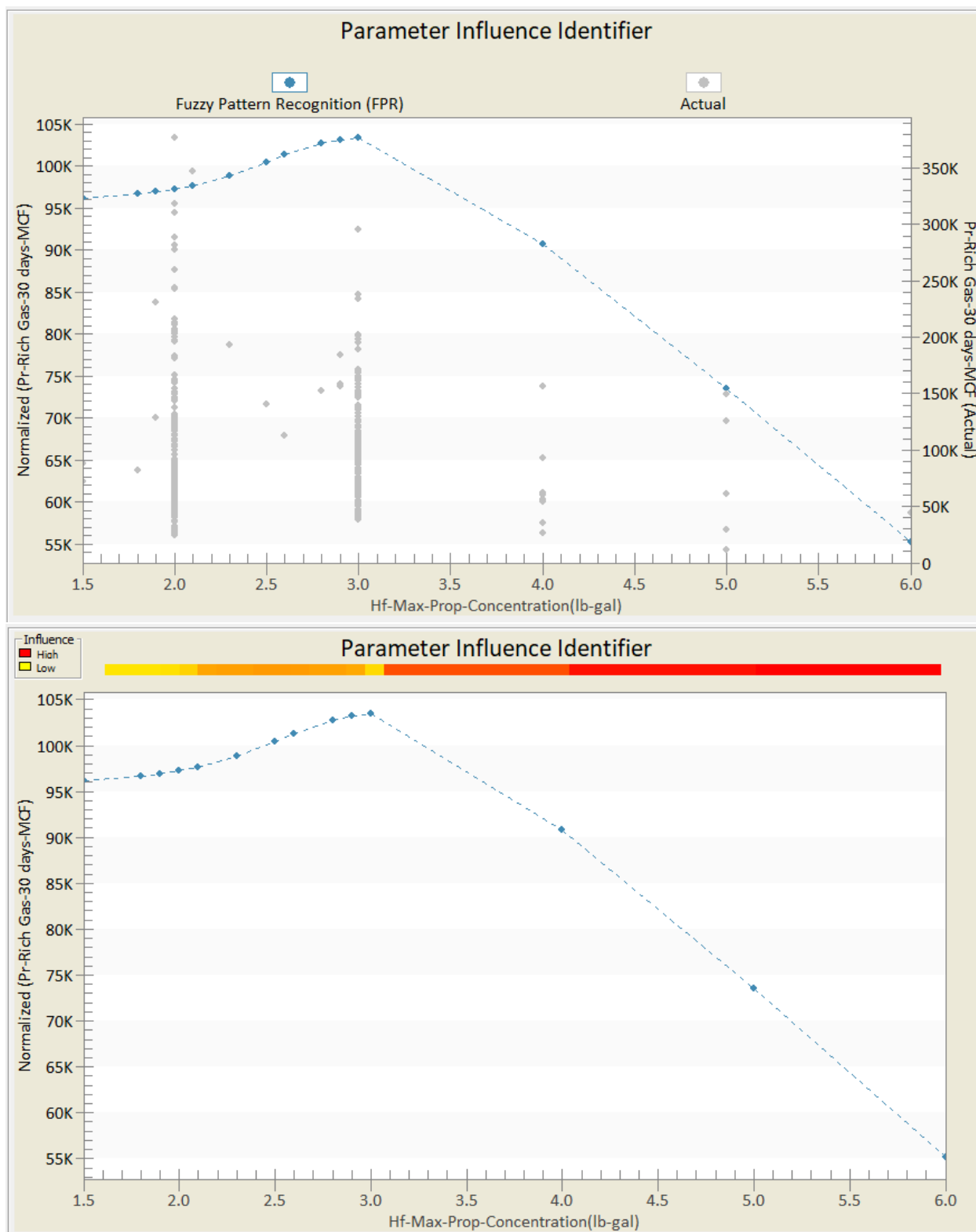


Figure 72: Parameter Influence (Fuzzy Pattern Recognition) - Max. Proppant Concentration

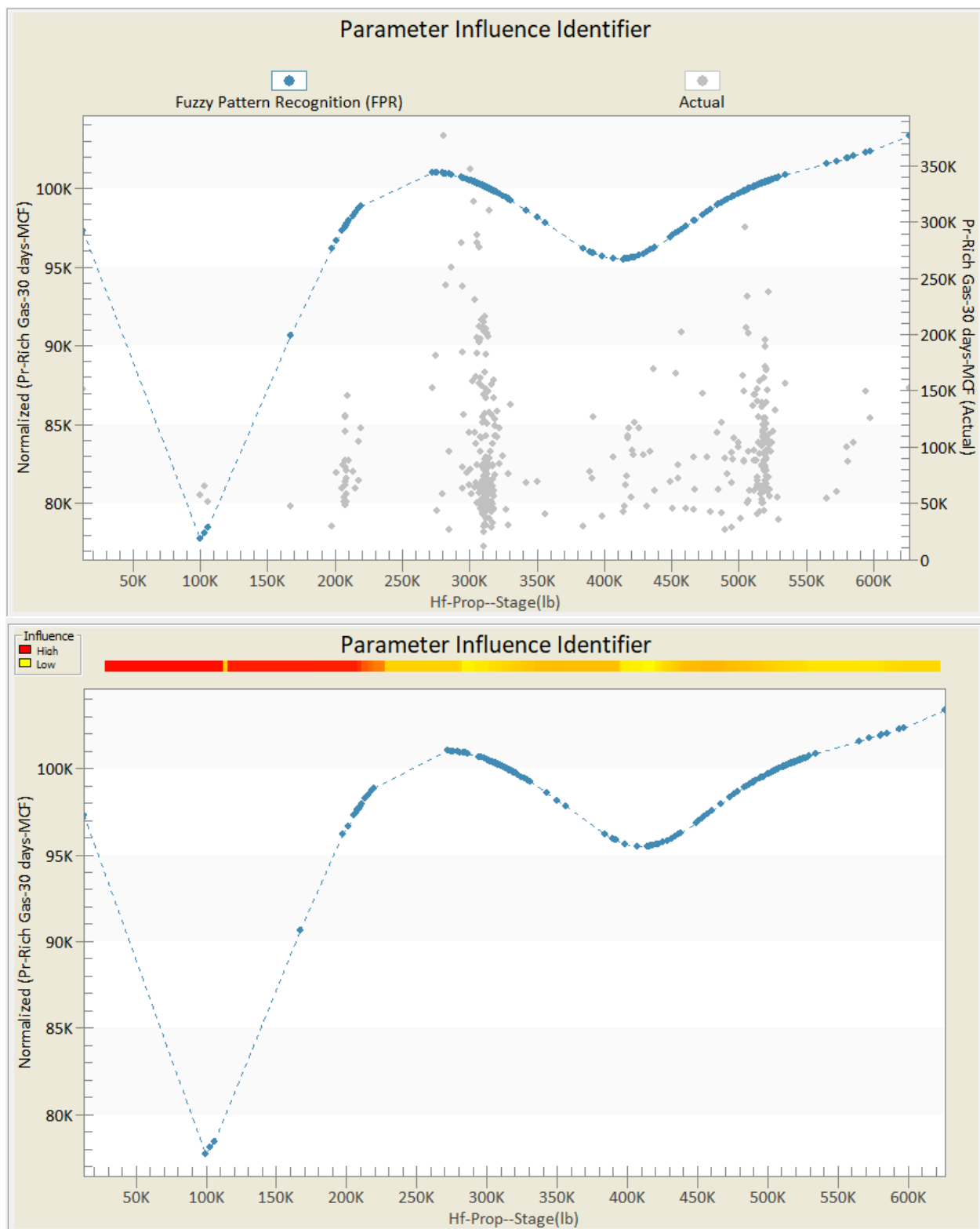


Figure 73: Parameter Influence (Fuzzy Pattern Recognition) - Proppant per Stage

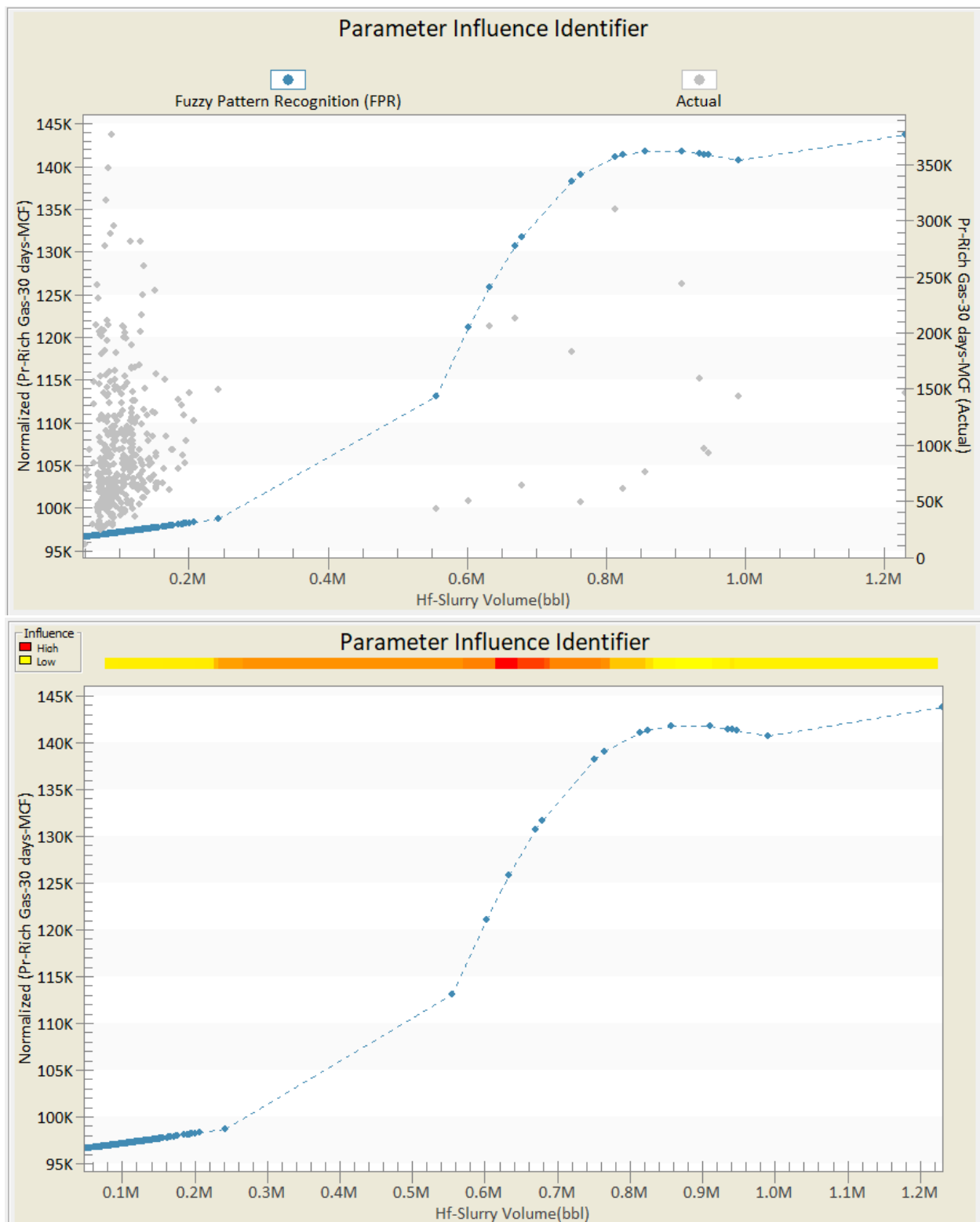


Figure 74: Parameter Influence (Fuzzy Pattern Recognition) - Slurry Volume

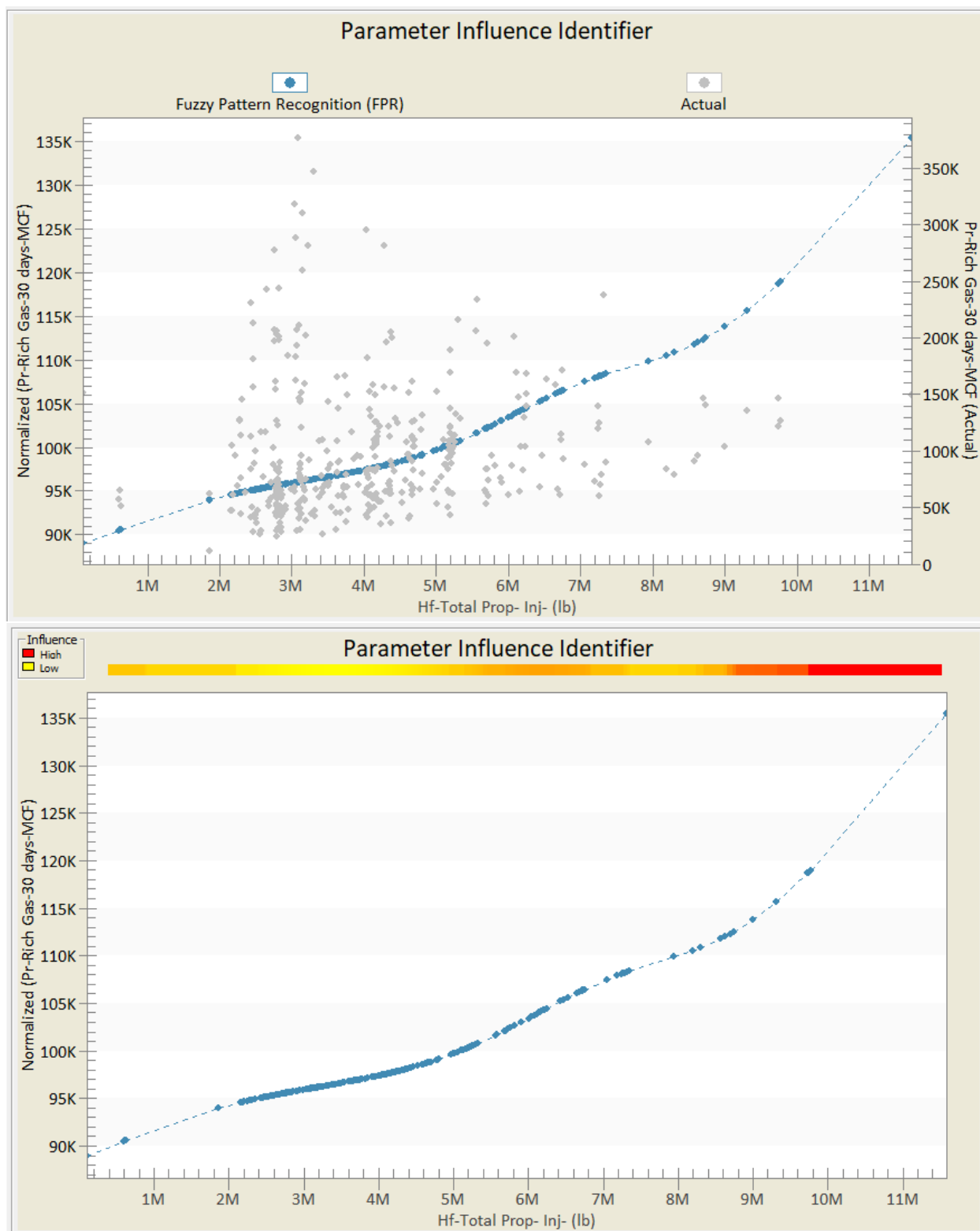


Figure 75: Parameter Influence (Fuzzy Pattern Recognition) - Total Proppant Injected

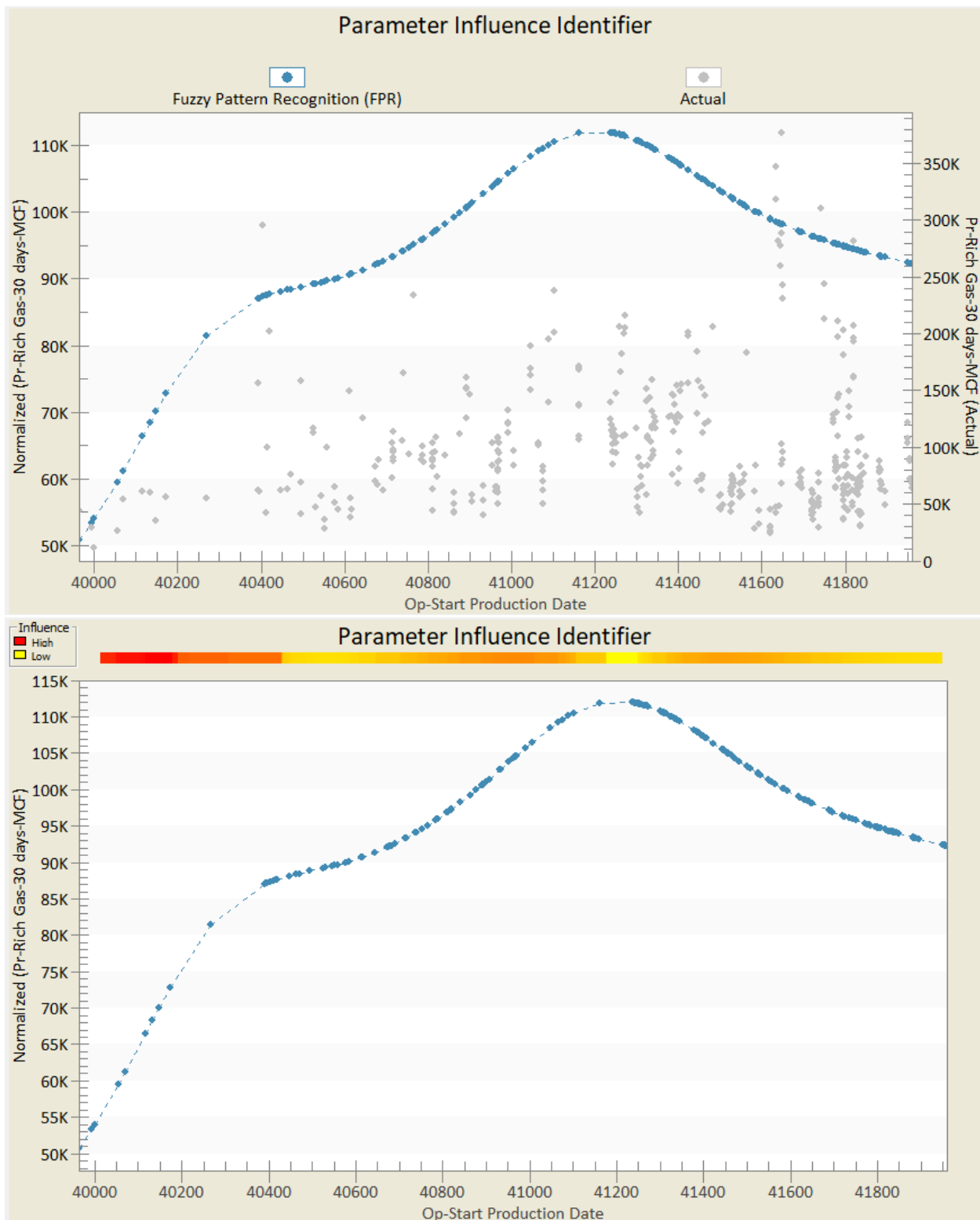


Figure 76: Parameter Influence (Fuzzy Pattern Recognition) - Start Production Date

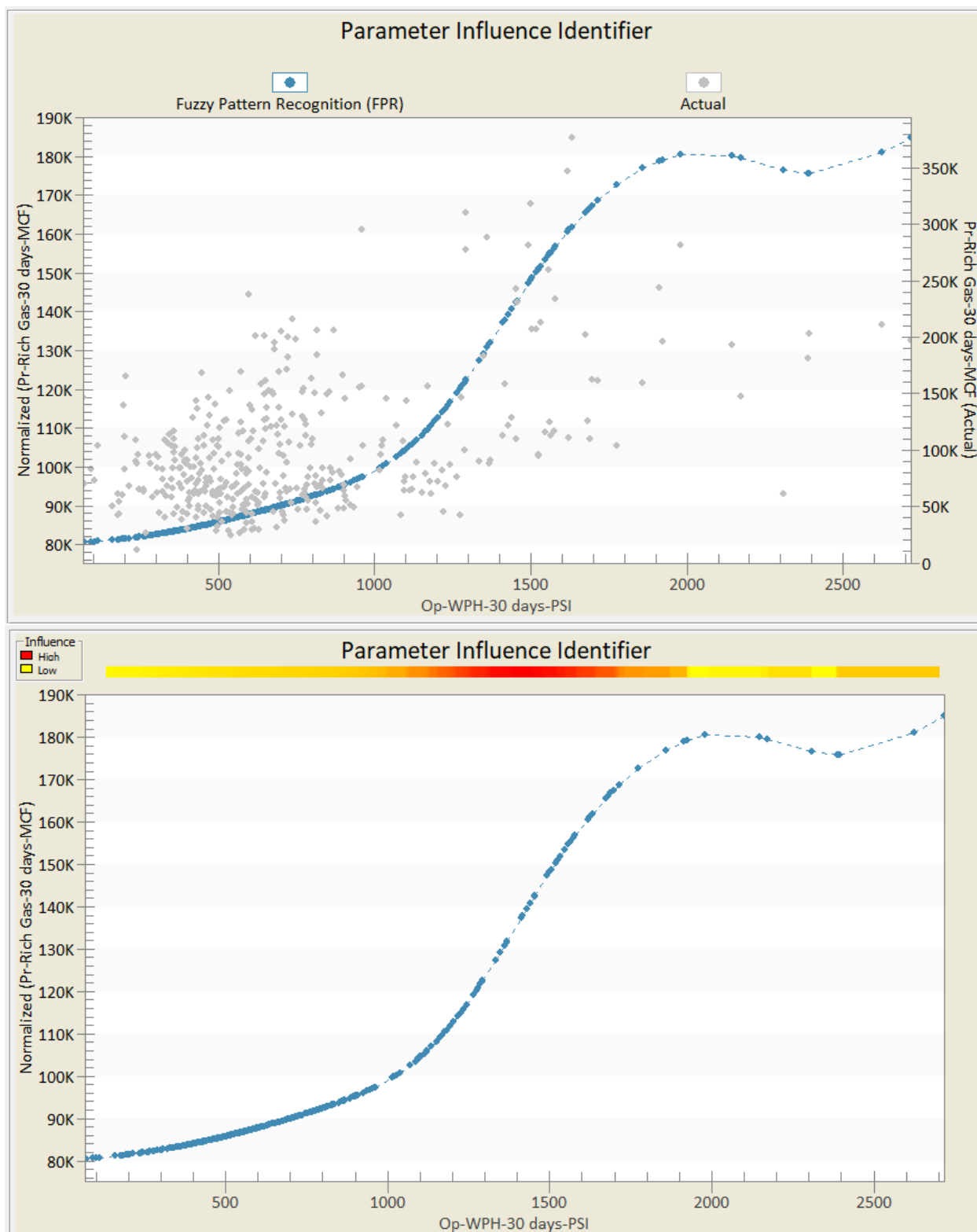


Figure 77: Parameter Influence (Fuzzy Pattern Recognition) - Wellhead Pressure WPH (30 Days)

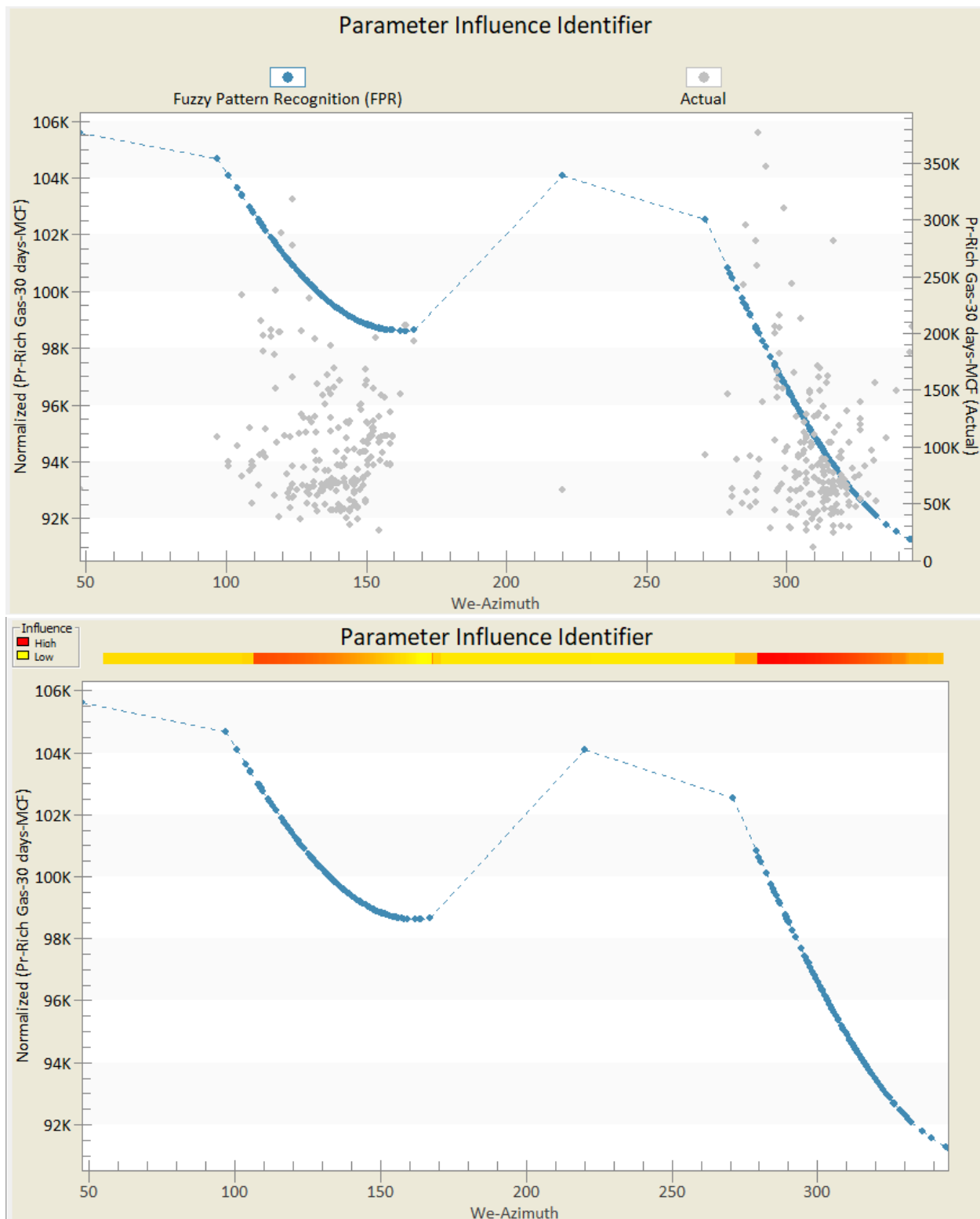


Figure 78: Parameter Influence (Fuzzy Pattern Recognition) - Azimuth

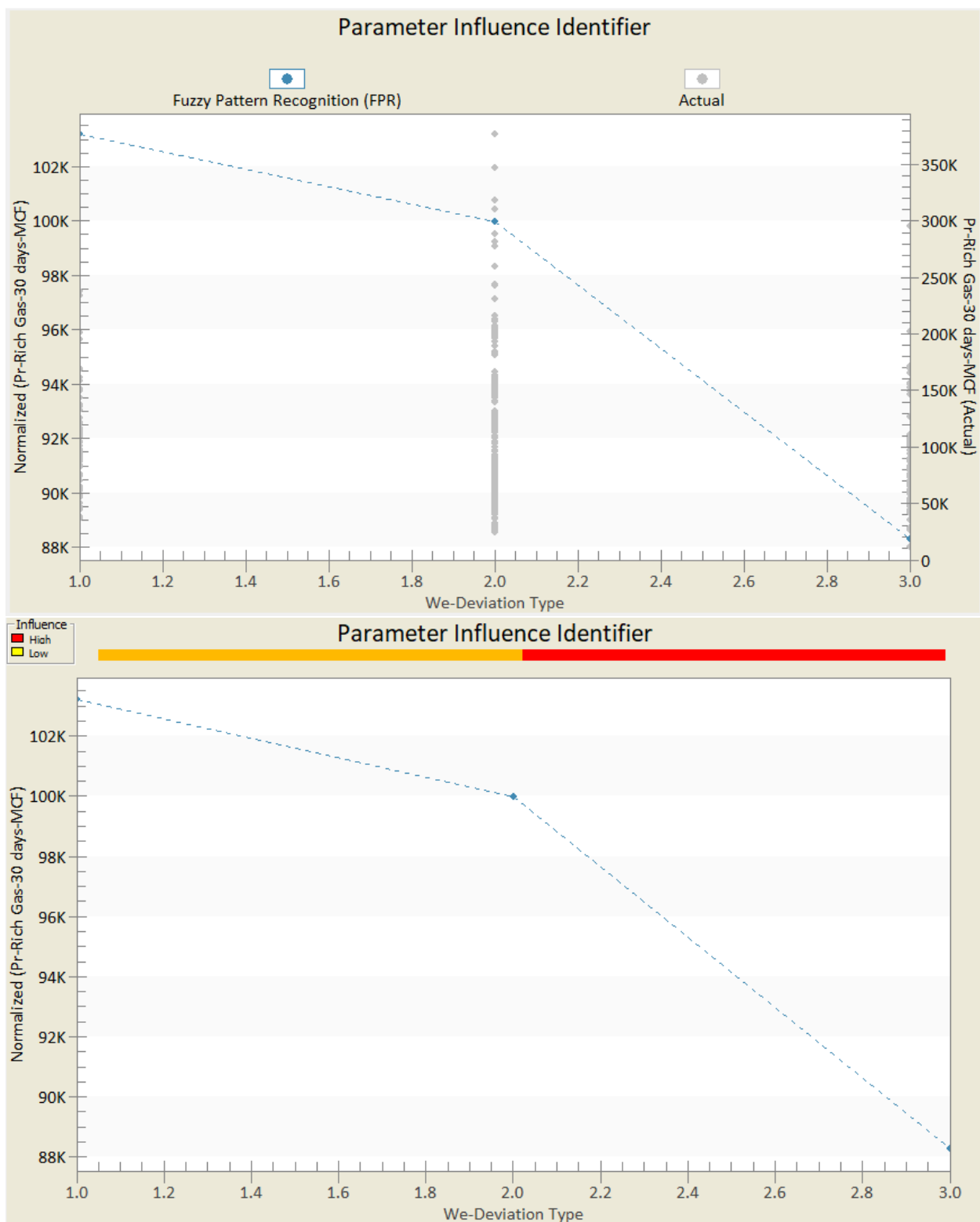


Figure 79: Parameter Influence (Fuzzy Pattern Recognition) - Deviation Type

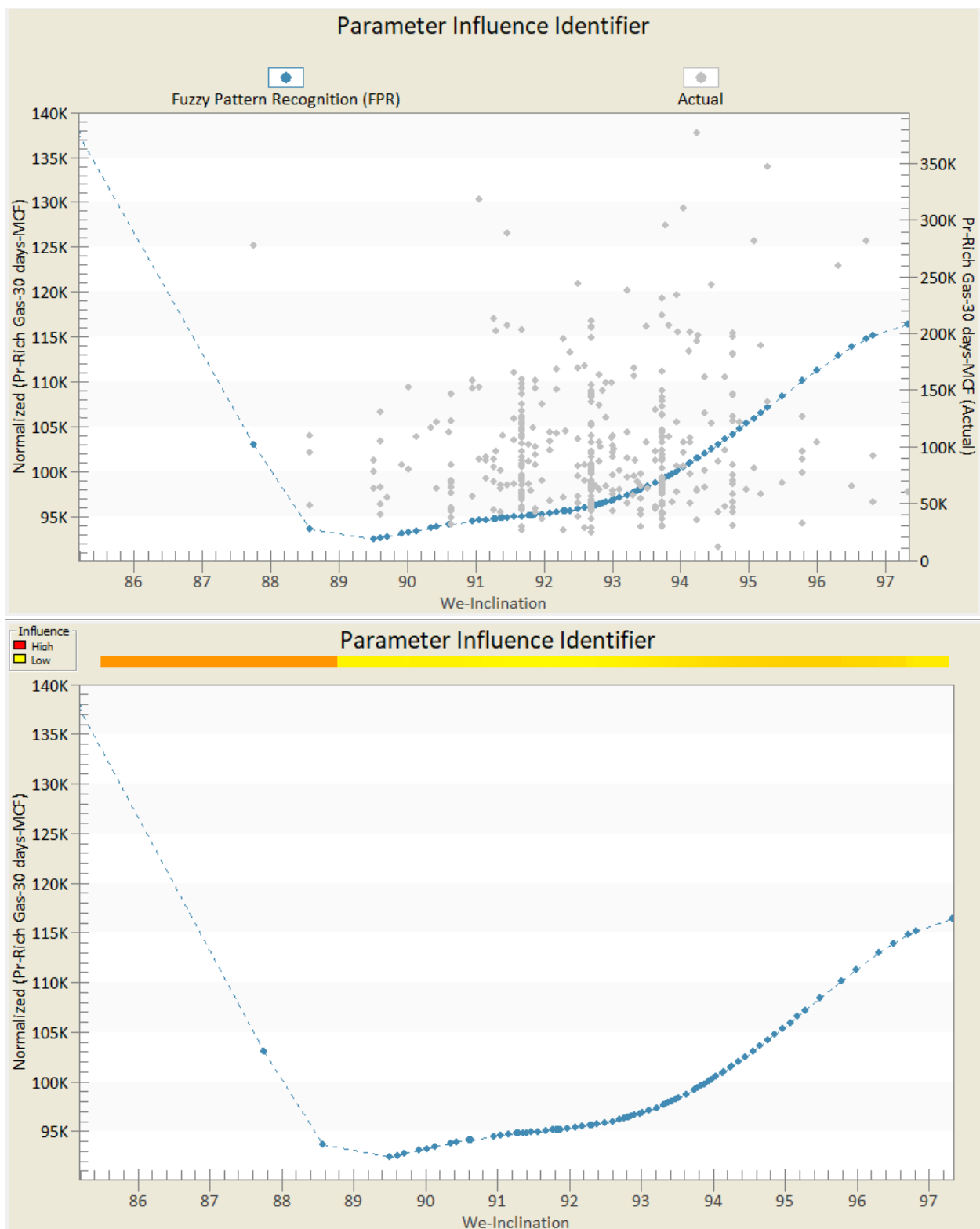


Figure 80: Parameter Influence (Fuzzy Pattern Recognition) - Inclination

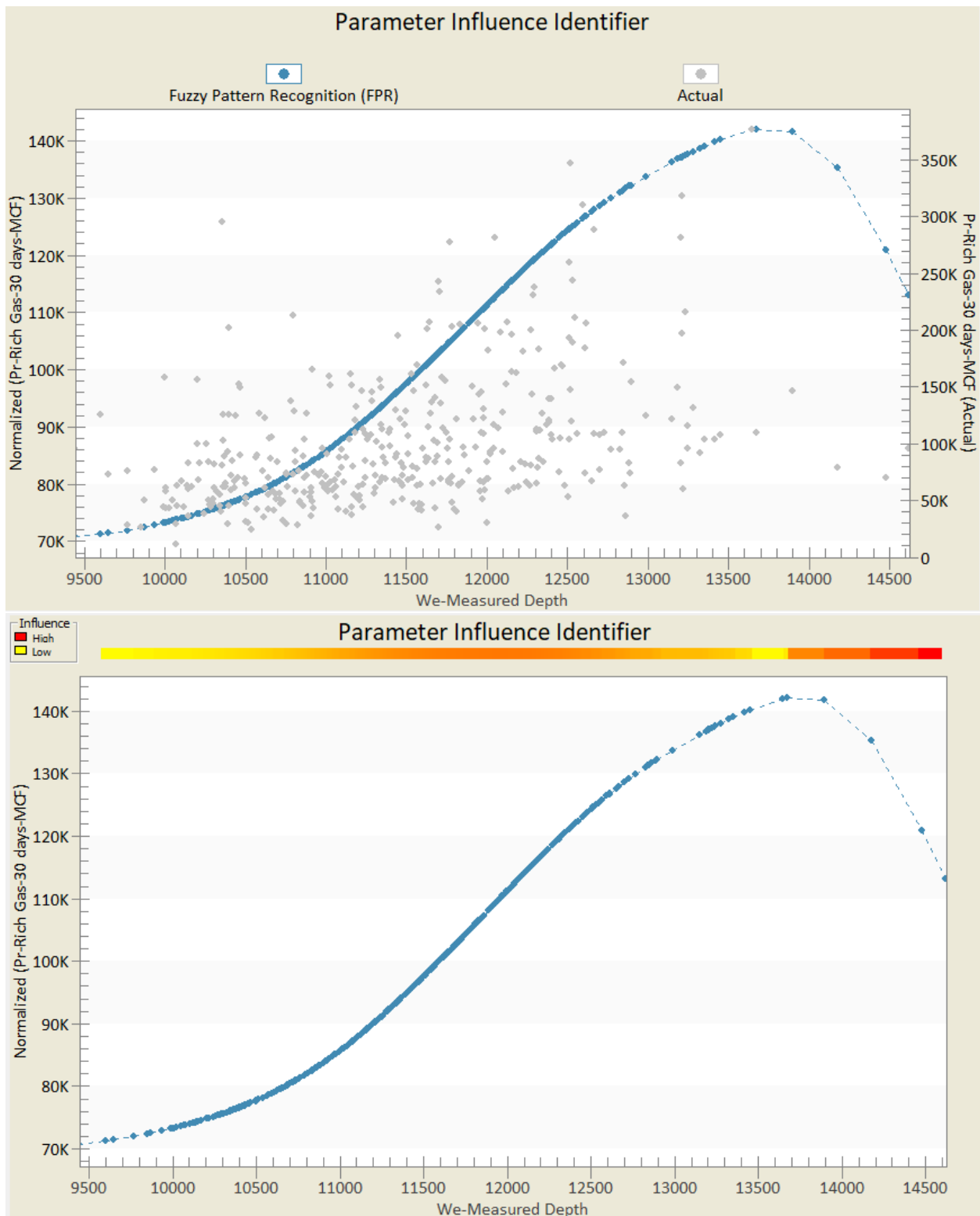


Figure 81: Parameter Influence (Fuzzy Pattern Recognition) - Measured Depth

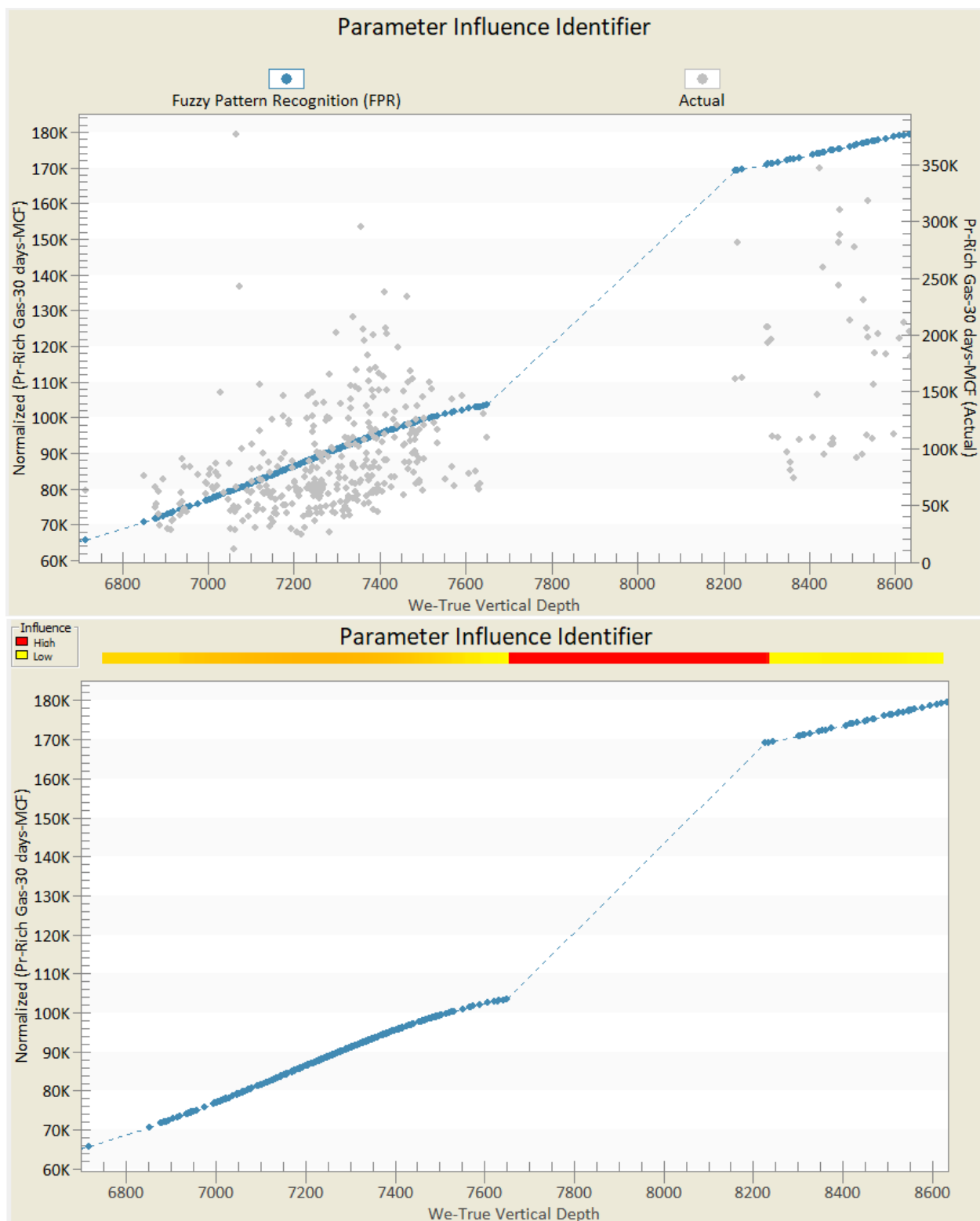


Figure 82: Parameter Influence (Fuzzy Pattern Recognition) - True Vertical Depth TVD

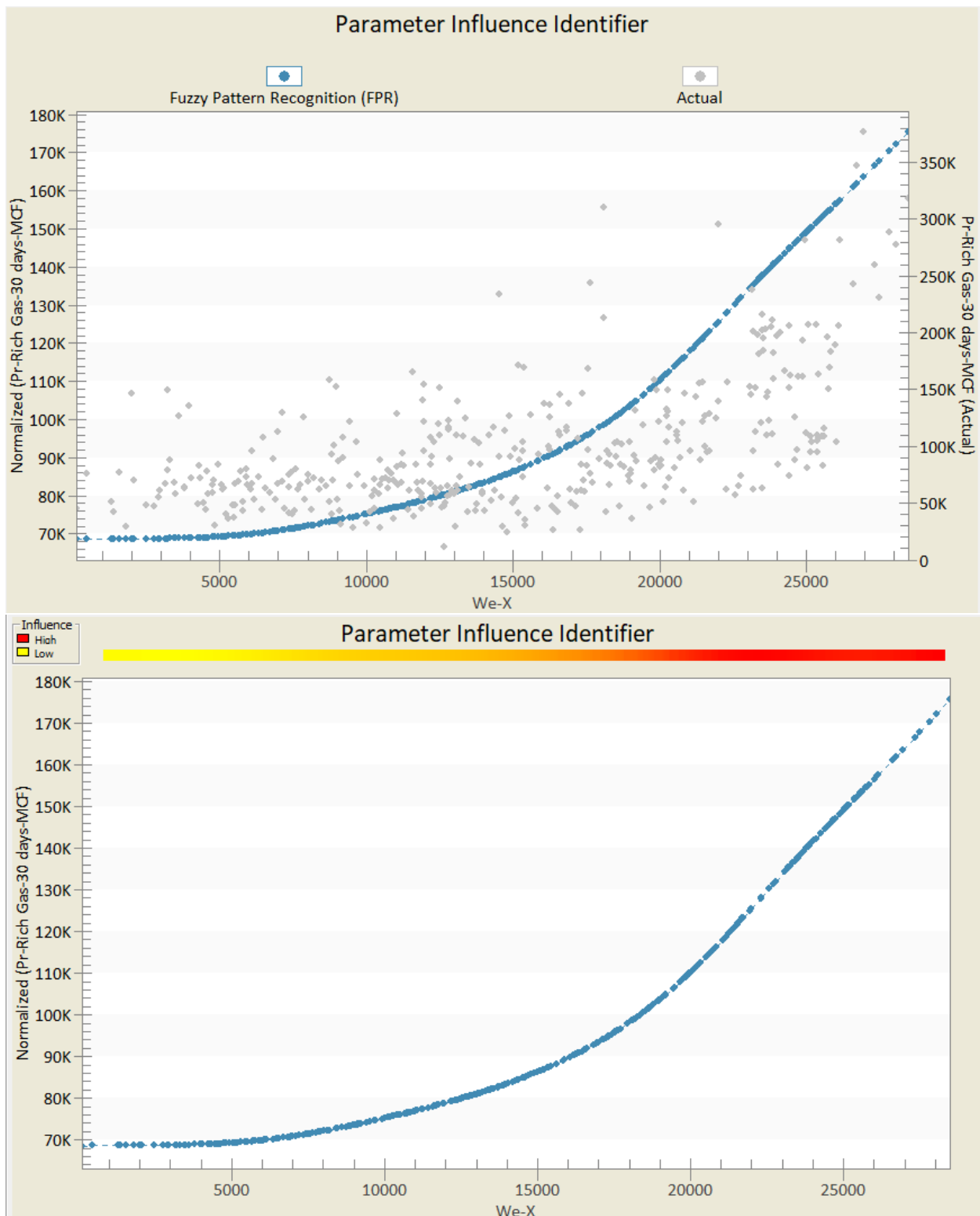


Figure 83: Parameter Influence (Fuzzy Pattern Recognition) - Well X Location Coordinates

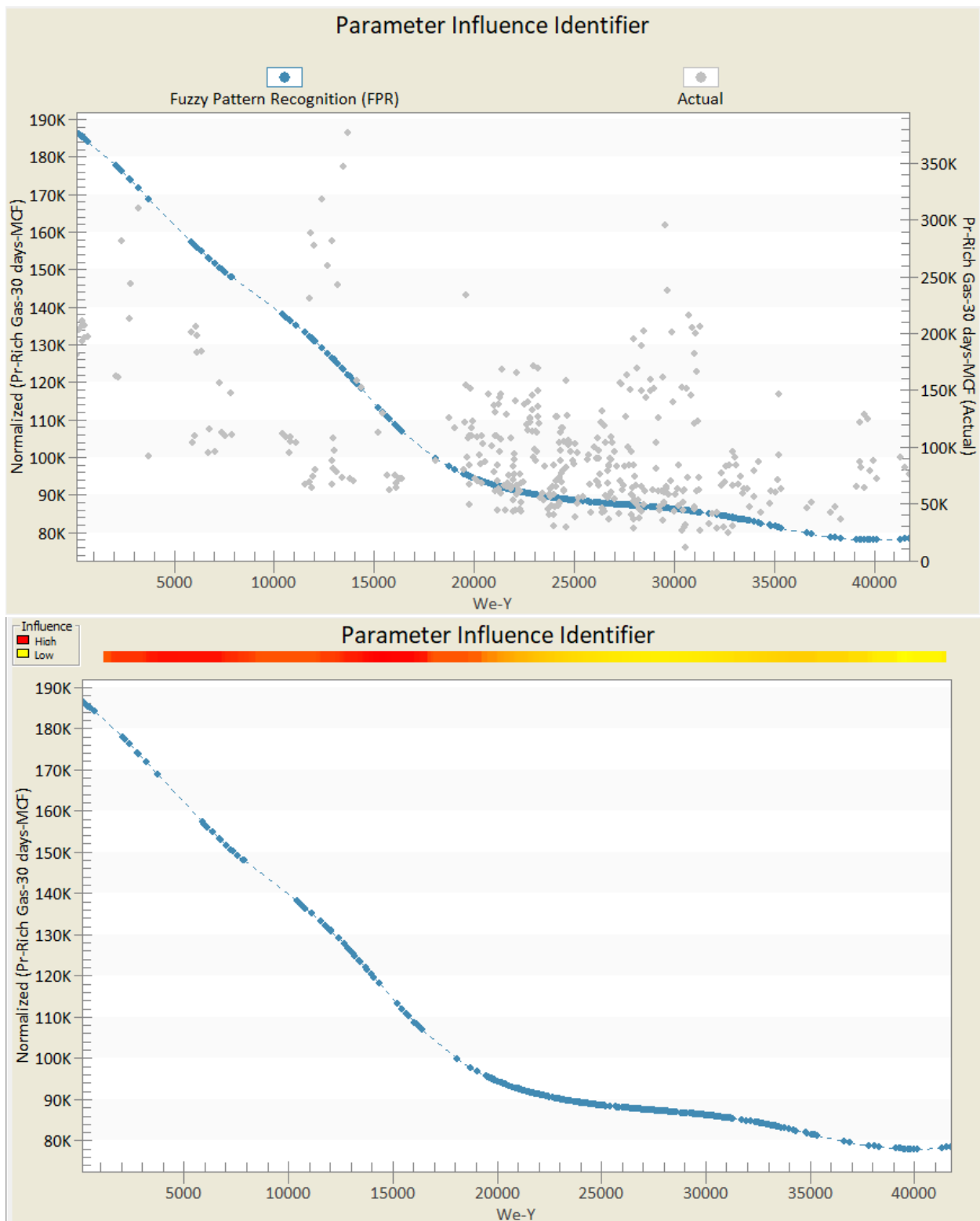


Figure 84: Parameter Influence (Fuzzy Pattern Recognition) - Well Y Location Coordinates