The Pennsylvania State University

The Graduate School

John and Willie Leone Family Department of Energy and Mineral Engineering

ANALYSIS OF NATURAL GAS PRODUCTION AND FRACTURE-FLUIDS
FLOWBACK IN MARCELLUS SHALE USING DATA MINING APPROACHES

A Dissertation in

Energy and Mineral Engineering

by

Qiumei Zhou

©2017 Qiumei Zhou

Submitted in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

August 2017
The dissertation of Qiumei Zhou was reviewed and approved* by the following:

John Yilin Wang  
Associated Professor of Petroleum and Natural Gas Engineering  
Dissertation Co-Advisor  
Chair of Committee

Andrew Kleit  
Professor of Energy and Environmental Economics  
Dissertation Co-Advisor

Turgay Ertekin  
Head, Department of Energy and Mineral Engineering  
Professor of Petroleum and Natural Gas Engineering  
George E. Trimble Chair in Earth and Mineral Sciences

Terry Engelder  
Professor of Geosciences  
Department of Geosciences

Yitian Xiao  
Senior advisor of ExxonMobil Exploration Company

Luis Ayala  
William A. Fustos Family Professor  
Professor of Petroleum and Natural Gas Engineering  
Associate Department Head for Graduate Education  
FCMG Chair in Fluid Behavior and Rock Interactions

*Signatures are on file in the Graduate School
ABSTRACT

Marcellus has been development for more than a decade with the application of multi-staged hydraulically fractured horizontal well technology. The technology requires pumping large amount of fracture-fluids and proppant into the target formation at high pressure. The fracture-fluid will then be recovered as aqueous phase during the flowback periods after well shut-in, which can be treated and reused. Sweet spot identification and efficient fracture-fluid flowback management are keys requirement for sustainable and economic development of Marcellus Shale, which can be benefited greatly by optimizing drilling and completion practices, including accurate fracture-fluids flowback prediction.

In this work, a systematic study of the geology and engineering factors that influence fracture-fluids flowback, water production, and gas recovery was developed. The complex correlations between gas production and fracture-fluids flowback and produced water provide more understanding about flow mechanism in shale gas. The results suggest that the numbers of hydraulic fracturing stages and well lateral length have significant influence on gas production. The shut-in time and injected proppant volume have the most influence on fracture-fluids flowback. The correlations between gas and fracture-fluid flowback and produced water were different under certain geological conditions and time periods. These knowledges from previous results were used to develop economic analysis regional scale.

This work not only will provide the new insights about shale gas well production and fracture-fluid flowback, but also provide a new idea for how to effectively analyze limited field recorded data and to identify the ‘true story’ behind data.
# TABLE OF CONTENTS

List of Figures ........................................................................................................... vi
List of Tables ............................................................................................................... x
Acknowledgements ................................................................................................... xi
CHAPTER 1 Introduction .............................................................................................. 1
CHAPTER 2 Literature Review ..................................................................................... 5
  2.1 Gas Shale Reservoir and Marcellus Shale ......................................................... 5
  2.2 Key Factors Controlling Gas Production ........................................................ 6
  2.3 Shale Gas Production Prediction ..................................................................... 8
  2.4 Key Factors Controlling Fracture-fluid Flowback Volume ......................... 10
  2.5 Fracture-fluid Flowback Volume Prediction ................................................ 10
  2.6 Section References ......................................................................................... 11
CHAPTER 3 Problem Statement ................................................................................... 18
CHAPTER 4 Evaluating Gas Production Performances in Marcellus Using Data Mining Technologies ........................................................................................................ 19
  Abstract .................................................................................................................... 19
  4.1 Introduction ......................................................................................................... 20
  4.2 Methods ............................................................................................................... 29
    4.2.1 Data Gathering, Screening and Grouping ................................................ 29
    4.2.2 Principle Component Analysis ................................................................ 34
    4.2.3 K-Mean Clustering Analysis ................................................................... 35
    4.2.4 Regression Analysis ............................................................................... 35
  4.3 Results and Discussion ...................................................................................... 38
    4.3.1 Single Parameter Study ........................................................................... 38
    4.3.2 Multiple Parameter Study ....................................................................... 42
  4.4 Section Conclusions ......................................................................................... 52
  4.5 Section References ............................................................................................ 54
CHAPTER 5 Evaluating Fracture-fluid Flowback in Marcellus Using Data Mining Technologies ............................................................................................................. 56
  Abstract .................................................................................................................... 56
  5.1 Introduction ......................................................................................................... 57
  5.2 Methods ............................................................................................................... 61
    5.2.1 Data Gathering, Screening and Grouping .............................................. 61
    5.2.2 Principle Component Analysis ............................................................... 66
    5.2.3 K-Mean Clustering Analysis ................................................................. 67
    5.2.4 Regression Analysis .............................................................................. 67
  5.3 Results and Discussion ...................................................................................... 69
    5.3.1 Single Parameter Study ........................................................................... 69
LIST OF FIGURES

Figure 1-1 Major shale gas play in US ........................................................................................................... 1
Figure 1-2 Processes for fracture-fluid flowback to surface ........................................................................ 3
Figure 4-1 Isopach map of brittle lithofacies (Carr et al. 2013) ................................................................. 22
Figure 4-2 Isopach map of organic-rich lithofacies (Carr et al. 2013) ......................................................... 22
Figure 4-3 Formation thickness distribution of study area (NETL 2010) ..................................................... 23
Figure 4-4 Thermal maturity distribution of study area (NETL 2010) .......................................................... 23
Figure 4-5 Thermal maturity distribution of study area (Pool 2013) ............................................................ 24
Figure 4-6 Marcellus gas production distribution in 2012 (Earl et al. 2013) ................................................. 26
Figure 4-7 Marcellus liquid production distribution in 2012 (Earl et al. 2013) ............................................. 26
Figure 4-8 Group A, B distribution on the map of West Virginia ................................................................. 32
Figure 4-9 Group A1, A2, B1, B2 distribution on the map of West Virginia .................................................. 33
Figure 4-10 One year cum gas/oil production between Group A and Group B .............................................. 39
Figure 4-11 One year cum gas production between Group B1 and Group B2 ............................................. 39
Figure 4-12 Single parameters analysis between engineer parameter and one year gas production .................... 41
Figure 4-13 Variable importance of Group A1 ............................................................................................. 45
Figure 4-14 Variable importance of Group A2 ............................................................................................. 45
Figure 4-15 Variable importance of Group B1 ............................................................................................. 46
Figure 4-16 Variable importance of Group B2 ............................................................................................. 46
Figure 4-17 Clustering results for Group B ................................................................................................. 49
Figure 4-18 Compared results between Group B1, B2 ............................................................................... 50
Figure 4-19 Clustering results of Group A .................................................................................................. 51
Figure 4-20 Compared results between Group A1, A2 ............................................................................... 52
Figure 5-1 Group A1, A2, B1, B2 distribution on the map of West Virginia .................................................. 64
Figure 5-2 Workflow of this paper ............................................................................................................. 65
Figure 5-3 Three-week flowback recovery between Group A (52 cases) and Group B (34 cases); the flowback recovery increased as $Ro$ increased ................................................................. 70

Figure 5-4 Three-week flowback recovery between Group B1 (17 cases) and Group B2 (17 cases); the flowback recovery decreased as $h$ increased ......................................................... 70

Figure 5-5 Analysis of single engineering parameter on 3-week flowback recovery, ............ 72

Figure 5-6 Key factors and corresponding variable importance value of each geological group ........................................................................................................................................ 76

Figure 5-7 Proppant concentration distribution for Group A1 (28 cases). Around 86% wells in ........................................................................................................................................ 77

Figure 5-8 Proppant concentration distribution for Group A2 (24 cases). Around 17% wells in ........................................................................................................................................ 77

Figure 5-9 Proppant concentration distribution for Group B1 (17 cases). Around 72% wells in ........................................................................................................................................ 78

Figure 5-10 Proppant concentration distribution for Group B2 (17 cases). Around 82% wells in ........................................................................................................................................ 78

Figure 5-11 Compared results between Group A1, A2 .......................................................................................................................... 81

Figure 5-12 Compared results between Group B1, B2 .......................................................................................................................... 81

Figure 6-1 Map of gas production in West Virginia wells. Each square points represent one well with different one-month gas production as shown in different color. The blue circle contains wells in wet gas region which have relative high gas production. The red circle contains wells in dry gas region which have relative low gas production. The black dashed line is the potential boundary between wet gas region and dry gas region ............................................................................................................... 95

Figure 6-2 Map of fracture-fluid flowback in West Virginia wells. Each square points represent one well with different one-month fracture-fluid in bbl as shown in different color. The blue circle contains wells in wet gas region which have relative high gas production. The red circle contains wells in dry gas region which have relative low gas production. The black dashed line is the potential boundary between wet gas region and dry gas region ............................................................................................................... 96

Figure 6-3 Workflow of non-linear model to identify the correlation between gas and water .......................................................................................................................... 102

Figure 6-4 Correlation between one-month gas and fracture-fluid flowback in wet gas region .......................................................................................................................... 105

Figure 6-5 Correlation of wells between one-month gas and fracture-fluid flowback in dry gas region .......................................................................................................................... 105
Figure 6-6 Proposed gas flow mechanism at pore size scale: The black region indicates organic matter, while the orange part represents gas-bearing pores inside organic matter. The dark yellow region represents the inorganic matter, consisting of siliciclastic grains of quartz, feldspar and micas. The light yellow points are gas-bearing pores. The light yellow points with blue circle are gas pores with bonded water. The blue points are pores filled with water. The red dashed line is the possible flow path for gas. No scale is indicated because this model may be applied across a wide range of scales. (Figure is modified based on Milliken’s SEM observations and Zhou’s observations from AAPG Lear Blog, http://www.aapg.org/publications/blogs/learn/article.)

Figure 6-7 Relationship for 1st-year gas production vs produced water

Figure 6-8 Relationship for 2nd-year gas production vs produced water

Figure 6-9 Relationships for 3rd-year gas production vs produced water

Figure 6-10 Changes in water-gas ratio across 1st (blue dots), 2nd (green dots), and 3rd year production (red dots). The cyan colored line represents the mean value. Each of the dots represent data from individual well.

Figure 6-11 Water vapor in methane under reservoir conditions. The different colored dots represent different temperature conditions. (Sage and Lacey 1979)

Figure 7-1 Flowchart of one geological condition

Figure 7-2 Comprehensive numerical simulation layout and area (Seals 2015)

Figure 7-3 Cumulative gas production from numerical simulation model

Figure 7-4 Gas production rate before five years (left) and after five years (right)

Figure 7-5 Cumulative gas production for the case with difference < 5%

Figure 7-6 Cumulative gas production for the case with difference >5%

Figure 7-7 Processes of spatial interpolation

Figure 7-8 Well location of two subsets, blue points are wells used to simulate and red points are wells used to test

Figure 7-9 Estimated Herry Hub Natural Gas Price to 2040

Figure 7-10 Five sub-plays distribution on the map (EIA, 2016)

Figure 7-11 Life-cycle cumulative gas production for wells in Region-1

Figure 7-12 Life-cycle cumulative water production for wells in Region-1

Figure 7-13 Life-cycle cumulative gas production for dry gas wells in Region-2
Figure 7-14 Life-cycle cumulative water production for dry gas wells in Region-2 .......... 157

Figure 7-15 Estimation from 1 to 25 years cumulative gas production in region-1 .......... 159

Figure 7-16 Estimation from 1 to 25 years cumulative water production in region-1 ........ 160

Figure 7-17 Estimation from 1 to 25 years cumulative gas production in region-2 ........ 162

Figure 7-18 Estimation from 1 to 25 years cumulative water production in region-2 ........ 163

Figure 7-19 Net present value in different producing period with price increasing rate as 3% ........................................................................................................... 165

Figure 7-20 Net present value in different producing period with price increasing rate as 5% ........................................................................................................... 166

Figure 7-21 Net present value in different producing period with price increasing rate as 10% ........................................................................................................... 166

Figure 7-22 Net present value in map review at price increasing with 3% yearly in region-1 ........................................................................................................... 167

Figure 7-23 Net present value in map review at price increasing with 5% yearly in region-1 ........................................................................................................... 168

Figure 7-24 Net present value in map review at price increasing with 10% yearly in region-1 ........................................................................................................... 168

Figure 7-25 Net present value in map review in Region-2 ............................................. 169
# LIST OF TABLES

Table 4-1 Number of wells considered in this study, by county.............................................29  
Table 4-2 Summary of mean and range for stimulation completion parameters and production .......................................................................................................................... 30  
Table 4-3 Correlation coefficient and P-value between completion parameter and gas production .......................................................................................................................... 41  
Table 4-4 Principle component loading vectors ........................................................................ 48  
Table 5-1 Summary of mean and range for stimulation completion parameters and production .......................................................................................................................... 63  
Table 5-2 Thermal maturity (Ro) and formation thickness (h) ranges in four groups............63  
Table 5-3 Correlation coefficient and P-value between completion parameters and flowback recovery through three weeks post stimulation................................................. 71  
Table 5-4 Coefficients of multiple regression results for each group......................................... 74  
Table 6-1 Water gas ratio of wells in Pennsylvania ..................................................................97  
Table 6-2 Mean and range for completion design parameters used...........................................97  
Table 6-3 Mean and range for production data used .................................................................99  
Table 7-1 Cumulative and decline rate gas from numerical simulation model (Seals 2015) .....129  
Table 7-2 Error comparison from original GS and modified GS model................................. 142  
Table 7-3 Validation results of sum 4\textsuperscript{th} and 5\textsuperscript{th} gas rate in region-1 ................. 146  
Table 7-4 Validation results of sum 4\textsuperscript{th} and 5\textsuperscript{th} gas rate in region-2 ......................... 147  
Table 7-5 Validation results of 1\textsuperscript{st} year water production in region-1 ............................ 148  
Table 7-6 Validation results of 2\textsuperscript{nd} year water production in region-1 ............................ 149  
Table 7-7 Validation results of 3\textsuperscript{rd} year water production in region-1 ............................ 150  
Table 7-8 Parameters assumed to calculate net present value .............................................. 164  
Table 7-9 Gas price assumed to calculated net present value ............................................... 164  
Table 7-10 NPV in MM$ after 25 years by using history gas price ........................................... 170
ACKNOWLEDGEMENTS

Firstly, I would like to express my sincerest and deepest appreciation to my advisor Dr. John Yilin Wang, for his constant encouragement, efforts, and patience during my PhD journey, my co-advisor Dr. Andrew Kleit for his advice in shaping my manuscripts for publication, and my committee members Dr. Turgay Ertekin, Dr. Terry Engelder, and Dr. Yitian Xiao, for sharing time from their busy schedule to provide me advice and help and expand my perspective. Secondly, I would like to thank my friends and office mate whom I have great time and memories with at Penn State. Finally, I would like to thank my parents and husband for their unconditional love and support.
CHAPTER 1

Introduction

Multi-stage hydraulically fractured horizontal well (MFHW) is an effective way to unlock the hydrocarbon reserves from gas shale formations. It enlarges the borehole surface contact area and provides the pathway for hydrocarbons flow from matrix to wellbore by pumping the fracture-fluid and proppant into the target formation with a high pressure. However, the production of shale gas with multi-fractured horizontal wells is complex, which related to the reservoir properties, fracture properties and so on. Some of those properties are hardly got and changes with time. Significant uncertainty about production performance in these shale reservoirs represents a key risk for whether resource development will lead to favorable production performance.

Figure 1-1 Major shale gas play in US
On the other hand, drilling a multi-fractured horizontal well usually requires enormous amount of water to be injected into the formation during treatment. The water recovered after the stimulation could be injected into the disposal well or reused for new well hydraulic fracturing. The recovered water usually contain higher than the concentration of injection fluid, including dissolved solids, sodium, calcium, chloride, potassium and so on. Poor management of the recovered water from flowback can have serve impacts on the environment. Moreover, the injected fracture-fluid will not be fully recovered during the flowback period. The rest of the water could be trapped in the formation matrix by capillary pressure or mineral wettability. The volume of flowback can be used as the index to predict short and long term gas production as well as diagnoses of stimulation treatment effectiveness.

Therefore, it is desired to know how much gas and fracture-fluid could be recovered with a development plan before making a decision. In other words, it is important for the industry to know: 1) what factors control the production and flowback of a shale well? 2) How can data from existing wells be used to predict the performance and optimize the design of a new well?

Many studies have been worked along the line with factors affecting the shale well production and prediction model development, but not many researches have been done for flowback fluid as production did. Those conclusions are different and conflicted sometimes. In general, the impacts on shale gas production and flowback fluid and the relationship between each other still needs more explorations.
To bridge the gap, this research aims to develop an innovative method through data mining techniques to identify the factors and corresponding correlations for shale gas production and fracture-fluid flowback. Single well and regional gas and fracture-fluid prediction model will be developed based on those finds and correlations. Finally an economic analysis through net present value calculation will be used to give the guidelines for optimal engineering design under different geological region and time period. The whole work not only will provide the new insights and understandings about shale gas well production and fracture-fluid flowback, but also provide a new idea for how to effectively analyze limited field recorded data and to identify the ‘true story’ behind data.

Figure 1-2 Processes for fracture-fluid flowback to surface
This dissertation consists eight chapters, the core of which is composed of chapters 4, 5, 6 and 7. Chapter 4 presented a study of effects from geological, completion and stimulation attributions on gas production. Chapter 5 described and identified the key engineering factor effect on fracture-fluid flowback. Chapter 6 explored the relationship between gas production and fracture-fluid flowback and discussed the possible flow mechanisms of water being carried out under different production periods. Chapter 7 developed production prediction model and regional economic analysis based on the results of previous three chapters. Finally, summary, contribution of this research and recommendations for future work are made. All wells used in this dissertation are slickwater-type multistage hydraulic fracturing horizontal wells.

Chapter 4 and Chapter 5 share same methodology and workflow, but different for Chapter 6 and Chapter 7. In each Chapter, the structure includes: abstract, introduction, methods, results, conclusion, and reference. Chapters 4 and 5 already published in open journals and Chapter 6 and 7 were ready to be published.
CHAPTER 2

Literature Review

2.1 Gas Shale Reservoir and Marcellus Shale

Shale is a fine-grained rock that consists primarily of clay, quartz and calcite (Blatt and Tracy 1996). It represents extremely low-permeability (in the micro-darcy to nano-darcy range), clay-bearing rocks, which used to be considered as the source rock or cap rock for conventional reservoirs (Steiger 1992). Compared with conventional reservoir, there is no trap formation in shale. The hydrocarbon directly stored in source rock (shale) without migration. Shale contains hydratable clays that make them water-sensitive to take up water and swell. Gas in the shale exists as absorbed gas on the organic matter and to some degree mineral matrix surface, and free gas in the micro pores in the matrix and fractures (Shtepani et al. 2010; Kang et al. 2011; Santos and Akkutlu 2013). Around 85% gas was stored as absorbed on organic matters and clay in shale gas reservoir (Lane 1989). The volume of absorbed gas is function of matrix pressure according to Langmuir theory (Langmuir 1916). Shale is often naturally fractured (Larese and Heald 1977; Misak et al. 1978; Shumaker 1979; Evans 1980), and the hydraulic fracturing would open existing natural fractures (Ortega and Aguilera 2014; Zhang et al. 2014).

The Marcellus Formation is Devonian, organic-rich black shale located in the Appalachian Basin of the north-eastern part of the U.S., which covers an area of
approximately 95,000 square miles at an average thickness of 50 ft to 200 ft. The estimated depth ranges from 4,000 ft to 8,500 ft. It has been estimated that 1,500 Tcf of gas is in place in the whole play area and that approximate 489 Tcf of gas is recoverable (Bruner and Smosna 2010). In most areas the Marcellus is underlain by the Onondaga Limestone formation and is bounded at the top by the Hamilton shale member of the Hamilton Group. Marcellus is laminated and composed primarily of quartz, clay, calcite, and organic material. Other common minor constituents include pyrite and chlorite (Carr 2013). The depositional setting is variable throughout the whole basin. Porosity of the shale matrix is in the range of 0.5-5\% with a highest level of 9\% in West Virginia. However, the matrix pore spaces in Marcellus are poorly connected; thus the successful extraction of natural gas depends largely on the fracture porosity of the shale. Matrix permeability is in the order of $10^{-4}$ md and very strongly stress dependent (Soeder 1988). Doubling the net stress will decrease the permeability by nearly 70\%. Organic content ranges from 2 to 14\% but averages 3 to 6\% in most areas. It is primarily composed of type II algal marine kerogen (Ward 2010), which is formed from marine animal and plant during evolution. Thermal maturity increased from west part, where the reservoir is oily, to east part, where are gas window and no oil production (Bruner and Smosna 2010).

### 2.2 Key Factors Controlling Gas Production

The factors that control shale well production can be divided into two categories: geological and engineering. The geological factors include porosity, permeability, mineral type and percentage, formation thickness, total organic matter, etc. The
engineering factors include well vertical depth, lateral length, lateral azimuth, proppant mass and size, fracture-fluid volume, hydraulic fracture azimuth, perforation cluster spacing, average treatment/injection rate, number of stages, number of perforation clusters, maximum/average treating pressure, etc. For different shale reservoirs, different factors dominate. For example, the dominant factor for Eagle Ford is suggested to be proppant mass (Centurion 2011; Centurion et al. 2012; Centurion et al. 2013) and the dominant factor for Marcellus Shale is lateral length (Cunningham et al. 2012; Esmaili et al. 2012a). For the same shale reservoir, the conclusion about the dominant factors can also be different, perhaps as a consequence of different methods and parameters used. For example, for the Barnett Shale, the controlling factor is suggested to be the average treatment rate (Sharon et al. 2013), or vertical depth (LaFollette and Holcomb 2011; LaFollette et al. 2012), or proppant mass (Nejad et al. 2013; Shelly 2008). The methods involved include multi-variables linear regression, regression tree and artificial neural network. The controlling factor analysis by multi-variables linear regression is based on the calculation of deduction of residual sum of squares (RSS), which could be represented as the R-squared value. For regression tree, the conclusion is obtained by measuring the total RSS reduction by adding one predictor. By artificial neural network, the conclusion is obtained by sensitivity test, which varies only one parameter’s range and fixes the other one to calculate the output changes. However, it is hard to support any of those conclusions without geology or physical explanations. Even all of those researches can identify which factors are important, but none of them can describe how the single controlling factor affects the well production performance. Moreover, the
correlation between parameters is not constant in space. The correlation in completion and stimulation parameters may differ across spatial scales.

### 2.3 Shale Gas Production Prediction

The predictive models for shale gas production can be classified into three categories: Analytical, Numerical and Data-Driven. Analytical methods including Decline Curve Analysis and Rate Transient Analysis are the most common approach (Bello and Wattenbarger 2010; Ambrose et al. 2011; Boulis et al. 2012; Ikewun and Ahmadi 2012). However, they cannot capture the characteristics such as the complexity of fracture geometry; the effects of gas desorption; the long transient behavior in the matrix blocks; and the interaction between natural fracture and hydraulic fracture.

Numerical model can overcome part of the shortcomings of analytical models due to their ability to simulate complex physics, such as the complexity of the fracture network (Mayerhofer et al 2006; Cipolla et al 2008), the location of proppant within the complex fracture network (Cipolla et al. 2009), gas diffusion in kerogen, slip flow, Knudsen diffusion, Langmuir desorption and water imbibition (Cipolla et al 2010; Thompson et al. 2011; Shabro et al. 2012; Dong and Hoffman 2013; Cohen et al. 2013). However, all of those simulation models are developed for single well prediction because of limitation on computational expense, which result in the lack of ability to capture the impact of interference between wells.

Since analytical and numerical modeling of “full field” shale assets is either impractical or leaves much to be desired, data mining approach provides an alternative solution to petroleum industry. Data mining can be applied to extract implicit and
interesting information from large databases, and reveal hidden patterns in those data (Mohaghegh 2003; Tan 2006). The multilayer feed-forward artificial neural network with back-propagation algorithm is the method that has been applied to predict the shale gas production. It mimics the behavior of human neurons to reveal the structure of dataset by self-learning and self-programing processes and gives an exact prediction for a given observation. This method was applied in several researches (Esmaili et al. 2012b; Khazaeni and Mohaghegh 2011; Awoleke and Lane 2011). Another is the regression tree method (Bhattacharya 2013), which is less applied in industry compared with neuron network. It segments the predictor space into a number of simple regions and uses the mean of mode of the training observations in the region to which it belongs as the prediction for a given observation. However, all of those researches have the same shortcoming. The spatial dependencies were not considered. Spatial dependence is “the propensity for nearby locations to influence each other and to possess similar attributes” (Goodchild 1992). Because of the variation in the production, in some part, due to non-uniformities in the physical characteristics of the rock, it seems reasonable to suspect that there might be spatial correlation in the production. That is the production at two wells will partially depend on the distance separating those wells. In other words, the new well production will affect the production performance of nearby wells. Therefore, those old wells will not perform in accordance with its previous data. Both artificial neural network and regression tree model can’t reflect the spatial correlation within wells.
2.4 Key Factors Controlling Fracture-fluid Flowback Volume

The controlling factors for fracture-fluid flowback volume change have not been clearly identified by previous published research. There is no data-analysis type research that studied the correlations between fracture-fluid flowback with geology, completion and stimulation parameters. However, it is still significant to identify the key factors controlling fracture-fluid flowback volume. Because the fracture-fluid flowback volume can help to estimate the effective fracture half length (Crafton 1998). In later times, people realized that the flowback data can also help to estimate the effective fracture volume (Alkouh and Wattenbarger 2013a), hydraulic fracture permeability (Abbasi et al. 2012; Williams-Kovacs and Clarkson 2013) and gas production data analysis (Crafton 2008; Clarkson 2012; Alkouh et al. 2013b). On the other hand, some researches show that the salinity of flowback fluid is higher than the concentration of injection fluid (Blauch and Myers 2009; Haluszczak et al. 2013; Barbot et al. 2013; Carter et al. 2013). The poor management of such high salinity fracture-fluid flowback, such as improper treatment and disposal, will cause serve impacts on environment.

2.5 Fracture-fluid Flowback Volume Prediction

Though fracture-fluid flowback is important for post-fracture evaluation and disposal management, work on flowback water characterization and predictive model development is very limited. There is just one research in the last decade that developed a model to simulate the fracture-fluid flowback volume in shale reservoir (Clarkson and Williams-Kovacs 2013). However, the model is limited as in two dimensional, single bi-
wing fracture, inability to capture the interaction between natural fracture and hydraulic fracture and proppant transportation in fracture network.

2.6 Section References


presented at the SPE Asia Pacific Oil and Gas Conference and Exhibition, Perth, Australia, 10/22/2008.


Techniques. Paper SPE 149258 presented at the SPE Eastern Regional Meeting, Columbus, Ohio, 08/17/2011.


33. Ikewun, P., Ahmadi, M. 2012. Production Optimization and Forecasting of Shale Gas wells Using Simulation Models and Decline Curve Analysis. Paper SPE 153914 presented at the SPE Western Regional Meeting, Bakerfiled, California, 03/19/2012.


I.G.C.P. project meeting at Virginia Polytechnic Institute and State University

51. Shabro, V., Torres-Verdin, C., Sepehrnoori, K. 2012. Forecasting gas production
in organic shale with the combined numerical simulation of gas diffusion in
kerogen, langmuir desorption from kerogen surfaces, and advection in nanopores.
Paper SPE 159250 presented at the SPE Annual Technical Conference and
Exhibition, San Antonio, Texas. 10/08/2012.

52. Sharon, Y.W., Luo, X.L., Hurt, R.S. 2013. What We Learned from a Study of Re-
Fracturing in Barnett Shale: An Investigation of Completion/Fracturing, and
Production of Re-fractured Wells. Paper IPTC 17081 presented at the
International Petroleum Technology Conference, Beijing, China. 03/26/2013.

Mechanical Properties of Shales. Society of Petroleum Engineers.
doi:10.2118/18024-PA

54. Thompson, J. W., Kannegant, K., Fan, L., Grant, D., Lindsay, G. J., Martin, R.
2011. An overview of horizontal-well completions in the Haynesville


presented at the SPE Unconventional Gas Conference, Pittsburgh, Pennsylvania,
USA, 02/23/2010.

flowback of multi-fractured horizontal wells to estimate hydraulic fracture
properties and forecast production. Paper SPE 164550 presented at the SPE
Unconventional Resources Conference-USA. Woodlands, Texas. 04/10/2013.

Hydraulic-Fracture Conductivities in the Barnett Shale. Journal of SPE
Production and Operation. 14(03), 1-12.
CHAPTER 3

Problem Statement

The objectives of this research are as follows:

(1). To identify controlling factors for gas production and fracture-fluid flowback volume

(2). To identify the correlation between gas and fracture-fluid/produced water

(3). To develop models to predict well and whole field gas production and fracture-fluid flowback volume under different geological settings and time period with spatial dependency analysis.

The data is collected from public sources and confidential sources both. The geological conditions (formation thickness, thermal maturity, well location) are applied to partition the data into several geological groups. In each geological group, the key engineering factors are identified by multiple regression and variable importance calculation. The gas and water relationship is identified and discussed. The prediction model is developed under different geological setting and production period. After model validation, the simple economic analysis will be implemented to provide suggestions for further development.
CHAPTER 4
Evaluating Gas Production Performances in Marcellus Using Data Mining Technologies

Abstract

Shale gas development - enabled by the advent of advanced horizontal drilling and hydraulic fracturing technology - has become, over the past several years, a very important energy resource. The estimated ultimate recovery of natural gas from the Marcellus Shale in West Virginia alone has been estimated to be between 98 and 150 trillion cubic feet (Tcf). In 2008, 25 billion cubic feet (Bcf) of natural gas was produced from 41 horizontal wells in West Virginia. By 2012, that gas production reached 301.7 Bcf from 631 horizontal wells. However, the hydraulic fracture stimulation of horizontal wells with multiple stages mechanism, by which that natural gas is produced from shale, is complex. Significant uncertainty about production performance in these unconventional reservoirs represents significant risk for whether resource development will lead to favorable technical and economic performance.

The objective of this chapter is to use post-hoc analysis techniques to identify correlations between gas production performance of a well and attributes of its completion and geological setting, and to identify those factors most important to predicting gas recovery performance. To accomplish this, the geological attributes of Marcellus Shale in West Virginia were characterized through literature review. Then, the set of 631 wells was down selected to a representative subset of 187 wells for which
complete data are available, including well location, completion data, hydraulic fracture treatment data and production data. The wells were classified into four groups based on geological setting. For each geological group, engineering and statistical analyses were applied to study the correlation between well performance and well completion attributes through traditional regression methods. Important factors considered to affect gas production include number of hydraulic fracture stages, lateral length, vertical depth, proppant volume, and fracture-fluid volume and treatment rate. The numbers of hydraulic fracture stages and lateral length have relative large influence on well performance. With these analysis results, it was possible to estimate well-scale ultimate natural gas production performance as a function of known geological conditions and completion parameters.

The results lead to a better understanding of the trends in Marcellus Formation well performance. These approaches could, in the future, help to optimize stimulation treatments and well completions and improve resource recovery in the Marcellus, and other unconventional hydrocarbon formations.

4.1 Introduction

The Marcellus Formation is the Devonian, organic-rich black shale located in the Appalachian Basin of the north-eastern part of the U.S., which covers an area of approximately 95,000 square miles at an average thickness ranging from 50 to 200 ft. The depth of the Marcellus Formation ranges from 4,000 to 8,500 ft. It has been estimated that 1, 500 Tcf of natural gas is in place in the whole play area, and that
approximately 489 Tcf of that gas is recoverable (NETL 2010). In most areas, the Marcellus is overlain by the Hamilton shale member of the Hamilton Group and underlain by the Onondaga Limestone formation (Sweeney et al. 1986). Marcellus shale matrix consists primarily of quartz, clay, calcite, and organic material, with common minor constituents including pyrite and chlorite (Cunningham et al. 2012). The composition of minerals determines the brittleness of the formation, with higher quartz content and lower clay content yielding a higher brittleness quotient. That brittleness corresponds to the ability to hydraulically fracture shale matrix (Enderlin et al. 2011). Fig. 4-1 shows the isopach map of the brittle lithology of Marcellus in West Virginia.

The depositional setting is variable throughout the whole basin. Porosity of the shale matrix is most typically in the range of 0.5-5% with a highest-level of 9% in West Virginia (Lee et al. 2010). However, the matrix pore spaces in Marcellus are poorly connected; thus the successful extraction of natural gas depends largely on the fracture porosity of the shale. Matrix permeability is on the order of $10^{-4}$ md and very strongly stress dependent (Soeder 1988); the permeability decreases by nearly 70% by doubling the net stress (Soeder 1988). Organic content ranges from 2 to 14% but averages 3 to 6% in most areas (Cunningham et al. 2012). Fig. 4-2 shows the isopach map of the organic-rich lithofacies. The organic matter is primarily composed of type II algal marine kerogen (Ward 2010), which is formed from marine animal and plant during thermal evolution.
Figure 4-1 Isopach map of brittle lithofacies (Carr et al. 2013)

Figure 4-2 Isopach map of organic-rich lithofacies (Carr et al. 2013)
Figure 4-3 Formation thickness distribution of study area (NETL 2010)

Figure 4-4 Thermal maturity distribution of study area (NETL 2010)
The thickness, thermal evolution and maturity of organic content, and formation pressure gradients vary significantly though all wells in the study area (NETL 2010, Pool 2013). Generally speaking, the northeast central spatial subset is thick and the southwest part is thin (Fig. 4-3). Thermal maturity generally increases from west (where the reservoir is oily) to the east (where the reservoir produces natural gas with no oil). The thermal maturity, as estimated using vitrinite reflectance (Ro), is low in the south-western part of West Virginia and high in the north-eastern part (Fig. 4-4).

Figure 4-5 Thermal maturity distribution of study area (Pool 2013)

The Marcellus is under pressured to the southwest and it has been postulated to be normal to potentially over pressured to the northeast with a transitional area in between (Fig. 4-5). As we know, all of these are important parameters for reservoir
evaluation and prediction of production performance. Formation thickness is an important parameter for evaluation of the original hydrocarbon in place. Also, thermal maturity impacts pore-network characteristics and matrix flow mechanisms, which significantly affect production.

Fig. 4-6 and Fig. 4-7 show the geographical distribution of Marcellus gas and liquid production in West Virginia in 2012. The gas producing wells are concentrated in certain countries, while liquid producing wells are concentrated in other countries. As of the end of 2012, 2,089 wells with production data were reported in West Virginia, including 1,458 vertical wells and 631 horizontal wells, with cumulative gas production reaching 301.7 Bcf, and liquid production reaching 715.6 kbbl (Earl et al. 2013). Shale gas production could increase to around 12 Tcf annually in 2035; production in the entire Marcellus could reach 6 Tcf per year in the same year (Mason 2012).

Well production performance is determined based on the amount of original hydrocarbon in place and the fraction of the hydrocarbon that can be technically and economically extracted. Both of these factors can be explored by analysis of reservoir characteristics and well development methods. Industry experience provides practical evidence that drilling multi-fractured horizontal wells (MFHW) is a very effective method to unlock the hydrocarbon reserves of shale formations. In the Marcellus Play, for example, the current practice is the stimulation treatment with application of large volumes of fracturing fluid and mass of proppant per hydraulically-fractured stage. However, it can be difficult to determine which factors are most effectively predicting recovery performance.
Figure 4-6 Marcellus gas production distribution in 2012 (Earl et al. 2013)

Figure 4-7 Marcellus liquid production distribution in 2012 (Earl et al. 2013)
Data mining is one approach that can be applied to extract implicit and interesting information from data in large databases, and reveal patterns in those data. Data mining has, in recent years, been widely employed in the oil and gas industry for analysis and optimization of controlling factors in complicated reservoirs. Common approaches used for data mining include artificial intelligence and neural network, fuzzy logic, classification and regression trees (CART), and data clustering.

Artificial Intelligence or neural network is the most commonly used approach to analyze non-linear relationships between input and output data. In neural network analysis, inputs and outputs are connected by neurons, which represent the activation function for transferring the signal from input to output. Input data are weighted as initial condition, and that weighting value will evolve during the self-learning processes of neural networks. Once the neural network is adequately trained, the controlling factor can be identified through sensitivity study according to the weighting value and the structure of the neural network (Esmaili et al. 2012).

In CART analysis, a model is created to predict the value of a target output variable (such as gas production in this study) based on several input variables. A decision tree can be built by splitting the source dataset into subsets based on attribute value tests. After the tree is created, variable importance can be estimated from the tree's structure, with importance estimated based on the node hierarchical level and the number of cases influenced (Bhattacharya et al. 2013).

Fuzzy logic is a problem-solving methodology that provides a means of arriving at a conclusion or prediction based on imprecise or missing input information. Compared to traditional logic (where decision variables may be taken as true or false), fuzzy logic
variables may have truth values represented by continuous variables with values ranging between 0 and 1. In fuzzy logic, there is a membership function to judge the magnitude of the variable's participation to the event (such as gas production in this study). According to the value calculated by that function, the influence of each input variable on the output can be determined (Alimonti and Falcone 2002). Statistical and multiple linear regressions maybe applied to analyze variable importance through multivariate regression techniques including: R-squared, Beta weight, and zero-order correlation (Cunningham et al. 2012).
4.2 Methods

4.2.1 Data Gathering, Screening and Grouping

The database generated and used in this study contains data from completion and production reports from public resources. These wells are located in northern West Virginia, distributed over 11 counties (Table 4-1).

<table>
<thead>
<tr>
<th>County</th>
<th>Number of Wells</th>
<th>County</th>
<th>Well</th>
</tr>
</thead>
<tbody>
<tr>
<td>Barbour</td>
<td>2</td>
<td>Preston</td>
<td>2</td>
</tr>
<tr>
<td>Doddridge</td>
<td>45</td>
<td>Taylor</td>
<td>3</td>
</tr>
<tr>
<td>Harrison</td>
<td>25</td>
<td>Tyler</td>
<td>2</td>
</tr>
<tr>
<td>Marion</td>
<td>8</td>
<td>Upshur</td>
<td>20</td>
</tr>
<tr>
<td>Marshalls</td>
<td>31</td>
<td>Wetzel</td>
<td>45</td>
</tr>
<tr>
<td>Monongalia</td>
<td>4</td>
<td>Total</td>
<td>187</td>
</tr>
</tbody>
</table>

Completion parameters detailed in the database include proppant type, proppant mass, fracture-fluid type, fracture-fluid volume, vertical length, lateral length, number of stages, number of perforation shots, perforation intervals, treatment rate, average treatment pressure, maximum treatment pressure, ISIP, and acid volume. Production parameters include: initial gas/oil production, and cumulative gas/oil/water production. Because of data insufficiency of some wells, only the one-year cumulative gas production, fracture-fluid volume, proppant mass, vertical depth, lateral length, treatment rate, and number of stages as variables were considered for this study.
Table 4-2 Summary of mean and range for stimulation completion parameters and production

<table>
<thead>
<tr>
<th>Name of parameters</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fracture-fluid (million gallon)</td>
<td>Mean 4.97</td>
</tr>
<tr>
<td></td>
<td>Range 0.03-10.9</td>
</tr>
<tr>
<td>Proppant (million lbm)</td>
<td>Mean 4.46</td>
</tr>
<tr>
<td></td>
<td>Range 0.02-10.5</td>
</tr>
<tr>
<td>Number of Stage</td>
<td>Mean 11.58</td>
</tr>
<tr>
<td></td>
<td>Range 5-29</td>
</tr>
<tr>
<td>Treatment Rate (bpm)</td>
<td>Mean 87.38</td>
</tr>
<tr>
<td></td>
<td>Range 61.5-101.13</td>
</tr>
<tr>
<td>Vertical Depth (ft)</td>
<td>Mean 7,081</td>
</tr>
<tr>
<td></td>
<td>Range 4,483-8,732</td>
</tr>
<tr>
<td>Lateral Length (ft)</td>
<td>Mean 4,690</td>
</tr>
<tr>
<td></td>
<td>Range 1,096-10,616</td>
</tr>
<tr>
<td>One-year Cum Gas Production (bcf)</td>
<td>Mean 0.74</td>
</tr>
<tr>
<td></td>
<td>Range 0.09-2.09</td>
</tr>
</tbody>
</table>

Table 4-2 shows the mean and ranges for each of those parameters. It can be argued that one year cumulative gas production might not be a strong indicator of well productivity since the decline in gas production rate in the first year is large, and the well drainage boundary could not be reached in that amount of time. It is noted, however, that one year cumulative gas production account for the large amount of gas reserves of a well, and it can be considered an important indicator of well performance. This observation, and the practical consideration that data form multi-year cumulative production are limited, warrant its consideration for assessment of well performance.
A subset of 187 horizontal wells that satisfied the requirement for statistical analysis was chosen from the complete set of information. As such, the results from this data analysis are limited to the geological region in which the selected wells are located. As more well data are collected in the future, this analysis can be expanded to represent a broader geologic domain and improve reliability and representativeness of analysis results.

Before proceeding with data analysis, the outliers were detected by Mahalanobis distance, which is defined as the distance between a data point and a multi-dimension space's centroid. The centroid is determined as a function of the mean values of all variables (4.1). The x represents the observation with N features, and μ is the centroid consisted of N features’ mean, and S is the covariance distance. The points with the Mahalanobis distance larger than a certain cutoff were treated as outliers. We assumed that those outliers could have significant influence on the analysis results, and removed them from the dataset. After this screening, the data set decreases from 187 to 173 observations. All of the following analyses are based on that set of 173 well samples.

\[
D_M(x) = \sqrt{(x - \mu)^T(S)^{-1}(x - \mu)} \tag{4.1}
\]

\[
x = (x_1, x_2, \ldots, x_N)^T
\]

\[
\mu = (\mu_1, \mu_2, \ldots, \mu_N)^T
\]

\[
S = \begin{bmatrix}
E((x_1 - \mu_1)(x_1 - \mu_1)) & E((x_1 - \mu_1)(x_2 - \mu_2)) & \cdots & E((x_1 - \mu_1)(x_n - \mu_n)) \\
E((x_2 - \mu_2)(x_1 - \mu_1)) & E((x_2 - \mu_2)(x_2 - \mu_2)) & \cdots & E((x_2 - \mu_2)(x_n - \mu_n)) \\
\vdots & \vdots & \ddots & \vdots \\
E((x_n - \mu_n)(x_1 - \mu_1)) & E((x_n - \mu_n)(x_2 - \mu_2)) & \cdots & E((x_n - \mu_n)(x_n - \mu_n))
\end{bmatrix}
\]
Since the geological parameters are not available as single well scale, the data are grouped by formation thickness and thermal maturity distribution from Bruner and Smosna’s report (2010) as shown in Fig. 4-8 and Fig. 4-9. This helps to eliminate the effect from the correlations changes cross the spatial space. The hypothesis is that in each geological group, the wells share similar geological settings.

![Figure 4-8 Group A, B distribution on the map of West Virginia](image)

As shown in Fig. 4-8, Group A and Group B include 114 and 59 wells respectively. The green region, Group A, represents the region in which vitrinite reflectance (Ro) falls between 1.0 and 1.6 (gas and oil two phase region), and the production should contain natural gas and oil together. The yellow region, Group B,
represents the region in which Ro ranges between 1.6 and 2.5 (dry gas region) and production is expected to contain natural gas only. The white area is not considered in this study.

![Map of West Virginia showing Group A1, A2, B1, B2 distribution](image)

**Figure 4-9** Group A1, A2, B1, B2 distribution on the map of West Virginia

In addition to thermal maturity, formation thickness is also an important parameter for original hydrocarbon in place and production performance evaluation. According to formation thickness distribution of Marcellus in West Virginia, we subdivided the Group B into two subgroups – labeled Group B1, B2 with thickness ranges as 60-80 ft, 80-100 ft respectively. The number of wells in each group are 33 (B1) and 26 (B2) respectively (Fig. 4-9).
4.2.2 Principle Component Analysis

To understand how geological factors affect the gas and flowback recovery, two methods are used. One is K-Mean Clustering and another one is Principle Component Analysis.

The principle components are relatively small numbers of representative variables to achieve dimension reduction for visualization, such as 2-D visualization. In this study, the engineering parameters are transferred as two principle components for clustering analysis.

The first principal component direction of the data is that along which the observations vary the most. The principal component of a set of features \(X_1, X_2, \ldots, X_p\) is the normalized linear combination of the features (4.2), where \(\phi\) is the loadings of principle component.

\[
Z_1 = \phi_{11}X_1 + \phi_{21}X_2 + \cdots + \phi_{p1}X_p
\]  

In this research, the first and second principle components are shown as below:

\[
pca\text{Comp1} = \varphi_{11} \times \text{Fracture fluid Volume} + \varphi_{21} \times \text{Proppant} + \varphi_{31} \times \text{Depth} \\
+ \varphi_{41} \times \text{Lateral Length} + \varphi_{51} \times \text{Number of Stages} + \varphi_{61} \times \text{Treatment Rate}
\]

\[
pca\text{Comp2} = \varphi_{12} \times \text{Fracture fluid Volume} + \varphi_{22} \times \text{Proppant} + \varphi_{32} \times \text{Depth} \\
+ \varphi_{42} \times \text{Lateral Length} + \varphi_{52} \times \text{Number of Stages} + \varphi_{62} \times \text{Treatment Rate}
\]
4.2.3 K-Mean Clustering Analysis

K-mean clustering method is chosen for partitioning the dataset into subgroups by engineering parameters because of its simplicity and efficiency. Each cluster shares similar engineering characteristics but different geological conditions. One cluster will be picked to study how gas or flowback recovery differs under different geological settings.

In K-means clustering, the dataset is randomly divided into K clusters at the beginning, each with a centroid. These centroids represent the mean of all variables in each cluster. The process of K-mean is to calculate the distance between each observation and the centroid, and add them up. The final clustering result is achieved when the sum of the distances between observed value and centroid in each cluster is at minimum after several iterations. This optimization process can be summarized as following equation (4.3), where Ck means the number of clusters, \( (x_{ij} - x_{i'j}) \) is the Euclidean distance between every two observations in k cluster.

\[
\begin{align*}
\text{minimize}_{C_1 \cdots C_k} \left\{ \sum_{k=1}^{K} \frac{1}{C_k} \sum_{i \in C_k \cap l \neq k} \sum_{j=1}^{P} (x_{ij} - x_{i'j})^2 \right\} 
\end{align*}
\]  

(4.3)

4.2.4 Regression Analysis

The multiple variables analysis is to study the correlation between one engineering parameter and gas/flowback recovery considering other engineering parameters effects. It is achieved by constructing a multiple linear regression function
without considering the prediction accuracy test. Then the coefficient of determination is obtained between each engineering parameter and gas/flowback recovery.

The goal of regression is to analyze correlation between multiple variables. The $R^2$ Squared is the ratio of sum of regression error over the sum of total error \((4.4)\), which represents the proportion of the behavior of the dependent variable that is explained by independent variables. SSTO (total sum of squares) measures the total variation, which is calculated as the sum of squared of the difference between observation value ($Y_i$) and its corresponding mean value ($\bar{Y}_i$). SSR stands for regression sum of squares, which is calculated the sum of squared of the difference between the mean value ($\bar{Y}_i$) and fitted value ($\hat{Y}_i$).

$$R^2 = \frac{SSR}{SSTO} = \frac{\sum (\bar{Y}_i - \bar{Y})^2}{\sum (Y_i - \bar{Y})^2} \quad (4.4)$$

The method proposed by Lindeman, Merenda, and Gold (1980) (LMG) is used to quantify the variable importance in cases where they are correlated. This approach is chosen because of its easy application and the limitation on public data availability. In this research, it is applied on the coefficient of determination obtained from multiple variables correlation analysis.

It is the averaging sequential sums of squares over all orderings of predictors. The term “sequential” means that the predictors are entered into the model in the order in which they are listed. Division of sequential sums of squares by the total response sum of squares yields sequential $R$-Squared contributions. This method takes into account the
dependence on orderings by averaging over orderings, using simple un-weighted averages. Equation (4.5) shows the corresponding function, where \( \text{lmg}(x_k) \) is the variable importance of \( x_k \) calculated by LMG method, \( \text{seq}R^2 \) is the sequential sums of square, \( n \) is the data size, \( p \) is the number of predictors, \( S \) is the dataset excluding \( x_k \).

\[
\text{lmg}(x_k) = \frac{1}{p!} \sum_{S \subseteq (x_1, \ldots, x_p) \setminus \{x_k\}} n(S)! (p - n(S) - 1)! \text{seq}R^2 ([x_k]|S) \tag{4.5}
\]
4.3 Results and Discussion

4.3.1 Single Parameter Study

Single parameter studies were implemented to examine the correlation of each parameter with the one year cumulative gas production to identify statistically-significant correlations between the gas production and the geologic, stimulation and completion features.

4.3.1.1 Geologic Parameters

The geological differences could have significant influence on the production performance. Thermal maturity, as one significant indicator of hydrocarbon composition and phases in a reservoir, is important for the production performance evaluation.

Fig. 4-10 shows a comparison of the results of one year cumulative gas and oil production in Group A and B. In Group A, the average one year cumulative oil production is 3361 bbl/year, while no oil production was reported in Group B wells. This observation supports the utility of the Ro 1.6 threshold. In two phase production region, Group A, the average one year cumulative gas production is significantly lower than that of Group B. Group A is partitioned into Group A1 and A2 by the Ro value equal to 1.3, above which large amount of gas are known to be generated.
Fig. 4-10 shows the average one year cumulative oil and gas production within each subgroup. This result suggests that as the thickness increases, the gas production
decreases, but this observation is counter to expected performance based on physical mechanisms. The observed higher production in B1 may cause by stimulation design, such as large stimulation treatment with longer horizontal wells. To further understand the relationship between formation thicknesses and the gas production, the stimulation and completion parameters need to be taken into account.

### 4.3.1.2 Engineering Parameter

To further evaluate the relationship between engineering parameters and production performance, the dataset is explored to understand how each single engineering parameter influences production performance, and to describe the correlation characteristics. Here, gas production is considered because most wells from which data were collected have no oil production. Fig. 4-12 shows results of single parameter correlation to one year cumulative gas production; the coefficients of correlation (r) between single engineering parameter (total fracture-fluid volume, proppant mass, vertical depth, lateral length, number of stages, treatment rate) and one-year cumulative gas production are presented in Table 4-3. The absolute value describes the degree of correlation, and sign in front of the value accounts for the type of correlation. The number of stages has the largest coefficient of correlation as compared to other parameters, followed by lateral length, with r value as 0.4110 and 0.4080 respectively. Except vertical depth and treatment rate, the other parameters, including fracture-fluid volume, proppant mass, lateral length, and number of stages have positive relationship with the gas production. In other words, the gas production increases as those parameter increases. Some parameters do not have significant relationships with gas production,
such as vertical depth and treatment rate, which have P-value larger than 0.05 (representing weak correlation), while others are statistical significant. These results show general trends between single stimulation or completion parameters with the one year cumulative gas production of Marcellus shale in North-West West Virginia.

Table 4-3 Correlation coefficient and P-value between completion parameter and gas production

<table>
<thead>
<tr>
<th>Fracture-fluid</th>
<th>Proppant</th>
<th>Vertical Depth</th>
<th>Lateral Length</th>
<th>Number stage</th>
<th>Treatment Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>r</td>
<td>0.2250</td>
<td>0.2290</td>
<td>-0.0004</td>
<td>0.4080</td>
<td>-0.1450</td>
</tr>
<tr>
<td>P-value</td>
<td>0.0007</td>
<td>0.0025</td>
<td>0.9958</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

(T-critical=1.96 (two-tail), 95% confidence interval, the correlation is statistically significant if t-test>1.96, or t-test<-1.96, or P-value <0.05)

![Figure 4-12 Single parameters analysis between engineer parameter and one year gas production](image)
4.3.2 Multiple Parameter Study

Single variable analysis does not provide strong insights into reservoir and stimulation characteristics that influence well productivity. The failure may stem from the lack of normalization for geologic and engineering parameters, and from the collinearity among those variables. It is also possible that there were inconsistencies within the data that could not be identified. In an attempt to overcome these issues, multiple regression, principle component analysis (PCA) and clustering methods were used in combination in an attempt to identify patterns in the dataset.

4.3.2.1 Same Geological Condition with Different Stimulation and Completion Design

To identify key engineer factors that affect well production we first factor out the influence of geological properties by implementing multiple regression analysis between production and engineering parameters within each geological group. It is assumed that each group has similar geological characteristics and the wells are developed using the same operational strategy. Continuing the single parameter analysis, the Group A and B were subdivided into two groups. Group A is partitioned by the Ro value equal to 1.3, which is the threshold value above which large amounts of gas are observed to be generated. Therefore, Group A1 represents the geological condition with Ro ranges from 1.0-1.3, and formation thickness ranges from 40-60 ft. Group A2 represent the geological condition with Ro ranges from 1.3-1.6, and formation thickness ranges from 40-60 ft. Group B is subdivided by the formation thickness. Group B1 represent the geological
condition with Ro ranges from 1.6-2.5, formation thickness from 60-80 ft. Group B2 represent the geological condition with Ro ranges from 1.6-2.5, formation thickness from 80 - 100 ft. Since, the engineering parameter units are different and vary significantly a scaling process is needed before multiple regression analysis can be implemented. The relative importance of each parameter was reflected as the variable importance. It refers to the quantification of an individual regressor’s contribution (stimulation and completion parameter) to account for the variance of the dependent variable (one year cumulative gas production). If the individual regressor’s contribution can account for relatively large variance of the dependent variable, we say this regressor is important for the dependent variable. Fig. 4-13 to 4-16 shows the order of variable importance for different geological groups.

As illustrated in Fig. 4-13, the most important engineering factor for Group A1 is the number of fractured stages, followed by fracture-fluid volume and proppant mass applied (The wells in Group A1 are located in the western-most portion of the study region). According to the brittle lithofacies distribution, the thickness of brittle lithofacies decreases through the study area from North-East to North-West. Therefore, Group A1 is the sub region with the least amount of brittle lithology (containing more clay) as compared with other three groups. This may suggest that larger size stimulation treatment is needed to create the hydraulic fractures, corresponding to fracture geometries that tend to be wider fracture aperture and shorter lengths. Usually, wide and short hydraulic fractures have a relative good conductivity near wellbore, which has significant positive influence on initial production.
Fig. 4-14 to 4-16 shows a similar trend within rest of the three groups. The lateral length and number of stages are the most important factors for these geological groups. Start from Group A2, the gas generation becomes more and more. To Group B, the observations are all dry gas wells. For these groups, compared with Group A1, they contain more brittle lithofacies. Therefore, good fracturing results may depend less on the large stimulation treatment. In those regions, hydraulic fracture geometry will tend to longer lengths and narrow fracture aperture. In this case, the initial production may depend more on the hydraulic fracturing spacing, which is controlled by lateral length and number of stages.

In conclusion, the different trends observed in Group A1 and the other three Groups may be explained to be a function of the mineral type change. As the matrix lithotype becomes brittle (follow the trend from West to East), the rock become easier to fracture. Matrix which is more difficult to fracture may require larger stimulation treatment. The fracture geometry for those ductile regions may tend to be wide and short, with good conductivity near the wellbore. Rock matrix in such brittle regions, will not require large stimulation treatment as compared to that in ductile regions. These stimulations may generate relatively longer and narrow fracture geometry with relatively poor conductivity and production can be expected to rely more heavily on fracture spacing than stimulation size (fracture-fluid volume and proppant mass). In such cases, number of stages and lateral length become the most important factors for gas production (one year cumulative gas production).
**Figure 4-13** Variable importance of Group A1

**Figure 4-14** Variable importance of Group A2
Figure 4-15 Variable importance of Group B1

Figure 4-16 Variable importance of Group B2
4.3.2.2 Same Stimulation and Completion Design with Different Geological Condition

To examine the effect of geologic condition on one year gas production, PCA and K-means methods were implemented on the dataset to eliminated the effects from stimulation and completion parameters, and generate different clusters. In one cluster, the stimulation and completion parameters are similar, while there is a relative difference between clusters.

The PCA method is implemented here to replace the six variables (fracture-fluid volume, proppant mass, vertical depth, lateral length, number of stages, and treatment rate) as two principle components, pcaComp1 and pcaComp 2. Each component is the linear combination with certain coefficients (principle component loading vector) from the six variables. The coefficients for pcaComp 1 and pcaComp 2 are shown in Table 4-4.

First, the influence from formation thickness is examined. According to previous analysis, the Group B1 and B2 have the same Ro range and different thickness ranges. Group B is subdivided by the fracture-fluid volume, proppant mass, lateral length, vertical depth, number of stages and treatment rate, total six engineering parameters, into different clusters as shown in Fig. 4-17. The best choice for the number of clusters is made depending on the shape of the distribution of points, the Elbow Method analysis result, and the Silhouette analysis result. The Elbow Method is to compute the percentage of the variance explained by certain number of clusters. The Silhouette method is to measure how closely for data within its cluster and loosely for data of different clusters. Comparing the clustering results, the case in which the number of clusters equals three is
the best. The one year gas production from Group B1 and B2 was compared in one of the clusters with equal sample size; production comparison result is shown in Fig. 4-18.

<table>
<thead>
<tr>
<th>Group</th>
<th>pcaComp1</th>
<th>pcaComp2</th>
<th>Fracture-fluid</th>
<th>Proppant t</th>
<th>Vertical Depth</th>
<th>Lateral Length</th>
<th>Number of Stage</th>
<th>Treatment Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0.53</td>
<td>-0.18</td>
<td>0.50</td>
<td>0.11</td>
<td>0.50</td>
<td>0.50</td>
<td>0.37</td>
<td>-0.28</td>
</tr>
<tr>
<td></td>
<td>9.9e-01</td>
<td>-1.5e-02</td>
<td>-1.5e-02</td>
<td>-6.5e-05</td>
<td>-4.1e-04</td>
<td>1.9e-02</td>
<td>4.5e-05</td>
<td>-5.9e-05</td>
</tr>
</tbody>
</table>

There are 11 points selected from Group B1 and 10 points selected from Group B2. Those wells share similar thermal maturity and engineering parameters (Fracture-fluid, proppant, vertical depth, lateral length, number of stages, treatment rate). The one year gas production in Group B2 is higher than the production in Group B1 – a difference which may be attributed to difference in formation thickness.

Another interesting preliminary insight resulting from this analysis relates to the effect of thermal maturity changing. In single parameter study, the difference in production in the gas-oil region and dry gas region was noted. Those results showed that, the gas production from dry gas regions is a slightly larger than the one from gas-oil regions. However, the two regions have different formation thickness and engineering parameters. Therefore, it is not correct to conclude that the production difference is
caused by the thermal maturity difference without normalizing for those other attributes. PCA and clustering methods were applied on Group A wells, which share the same formation thickness range but different reservoir types. The clustering results are shown in Fig. 4-19. Compare the clustering results, when number of cluster equal to two is the best case.

![Figure 4-17 Clustering results for Group B](image-url)
Similar to the results of the previous analysis, the gas oil ratio (GOR) from Group A1 and A2 was compared in one of the clusters with equal sample size. The production comparison result is shown in Fig. 4-20.

![Figure 4-18 Compared results between Group B1, B2](image)

There are 23 points picked from Group A1 and 24 points picked from Group A2. Those wells share similar formation thickness and engineering parameters (fracture-fluid, proppant, vertical depth, lateral length, number of stages, treatment rate), but different Ro ranges. The Ro of Group A1 is smaller than Group A2 - Group A2 has a relatively high thermal evolution. As the result, the GOR of Group A1 is smaller than the Group A2. In other words, the Group A1 produces more gas than Group A2 at the same oil production.
Figure 4-19 Clustering results of Group A
4.4 Section Conclusions

In this study, we first reviewed available information on geology of Marcellus Formation and collected data from 187 Marcellus wells in northern West Virginia. We then classified wells into two groups and four sub-groups based on formation thickness and thermal maturity. Wells in Group A produce gas and oil (or condensate), while wells in Group B yields dry gas. The production of Group A1 contains a significantly higher fraction of liquid hydrocarbon than Group A2.

- Data screening was carried out by Mahalanobis distance analysis to identify outliers in whole dataset. Simple variable analysis results show the general trend between gas production with other geology, stimulation and completion parameters. The gas production has positive relationship with fracture-fluid volume, proppant mass, lateral length and number of stages, while no trends with
vertical depth and treatment rate.

- According to t-test results and P-value, correlations between single completion/stimulation parameter and production may not exist based on 95% confidence interval.

- Based on the multiple regression analysis and individual variable importance measurement in each geological group, the number of hydraulic fracture stages is found to be the most significant parameter among all factors studied for the horizontal wells in West Virginia.

- Based on variable importance, we ranked pertinent parameters. In Group A1, the fracture-fluid volume and proppant mass are relatively important compared to other parameters. This difference may cause by the brittleness lithology in this area. Compared with other three groups, the region of Group A1 are less brittle and may require larger stimulation treatment. In groups A2 and B, the number of stage and lateral length are important for initial gas production (one-year cumulative gas production), because it creates larger contact with the formation.

Advanced data mining and simulation techniques are needed for future studies. The authors also understand that with the limitation of the data available for analysis, the conclusions may not be applicable to the whole Marcellus Formation. Still, the study may provide a few new insights and understandings of the correlation of well completion and stimulation with well performances, which may help operators optimize field operations that leads to lower costs and higher gas production and recovery in the Marcellus Formation, and contribute to the development of data mining approaches to explore such relationships in other unconventional fossil resource plays.
4.5 Section References


CHAPTER 5

Evaluating Fracture-fluid Flowback in Marcellus Using Data Mining Technologies

Abstract

Natural gas recovery from low permeability unconventional reservoirs – enabled by the advanced horizontal drilling and multi-stage hydraulic fracture treatments - has become very important energy resource in past decade. While evaluating early gas production data in order to assess likely rate decline and ultimate gas recovery has been reported in literature, flowback water recovery is rarely considered. Fracture-fluid flowback is defined herein as aqueous phase produced within three weeks following a fracture treatment (exclusive of well shut-in time). Field data indicated about 2-26% of the fracture-fluid is recovered during flowback. However, stimulation of gas shale is a complex engineering process, and the factors that control the volumetric flowback performance are not well understood.

The objective of this paper is to use post-hoc analysis to identify correlations between fracture-fluid flowback water and attributes of a wells completion and geological setting, and to identify those factors that are most important in predicting flowback performances. To accomplish this objective we selected a representative subset of 187 wells for which complete data are available (from a full set of 631 wells), including well location, completion data, hydraulic fracture treatment data and production data. The wells were classified into four groups based on geological settings. For each geological group, engineering and statistical analyses were applied to study the
correlation between flowback data and well completion through traditional regression methods. Important factors considered to affect flowback water recovery efficiency include number of hydraulic fracture stages, lateral length, vertical depth, proppant mass applied, proppant size, fracture-fluid volume applied, treatment rate, and shut-in time. The total proppant mass, proppant size and shut-in time have relatively large influence on volumetric flowback performance.

The new results enable one to estimate flowback volume in a spatial domain, based on known geological conditions and completion parameters, and lead to a better understanding of flowback behaviors in Marcellus Shale. This also helps industry manage flowback water and optimize production operations.

5.1 Introduction

The Marcellus Formation – a Devonian-age, organic-rich black shale covering an area of approximately 95,000 square miles in the Appalachian Basin of the north-eastern part of the U.S. Marcellus formation has proved to be a major resource for economically-recoverable natural gas and liquid hydrocarbon condensate, with approximately one third of gas in place estimated to be recoverable (Zamnerilli 2010). The Marcellus ranges in thickness from an average thickness of 50 to 200 ft; the depth of the Marcellus ranges from 4,000 to 8,500 ft. In most areas, the Marcellus is overlain by the Hamilton shale member of the Hamilton Group and underlain by the Onondaga Limestone formation (Sweeney et al. 1986; Carter 2007). Marcellus shale matrix is composed primarily of quartz, clay, calcite, and organic material, with common minor constituents including
pyrite and chlorite (Cunningham et al. 2012). Porosity of the shale matrix is most typically in the range of 0.5-5\% with a highest-level of 9\% in West Virginia (Lee et al. 2010). However, the matrix pore spaces in Marcellus are poorly connected; thus the successful extraction of natural gas depends largely on the fractures. Matrix permeability is on the order of $10^{-4}$ md and strongly stress dependent (Soeder 1988); the permeability decreases by nearly 70\% by doubling the net stress (Soeder 1988). Because net stress will increase as pore pressure decreases during gas production, effective matrix permeability can be expected to decrease through the period of hydrocarbon recovery. Organic content ranges from 2 to 14\% but averages 3 to 6\% in most areas (Cunningham et al. 2012). The organic matter is primarily composed of type II algal marine kerogen (Ward 2010). Thickness, thermal evolution and maturity of organic content, and formation pressure gradients vary significantly. Thermal maturity generally increases from west (where the reservoir is oily) to the east (where the reservoir produces natural gas with no oil) (Zammerilli 2010; Bruner and Smosna 2010). The thermal maturity, as estimated using vitrinite reflectance ($Ro$), is low in the south-western part of West Virginia and high in the north-eastern part. Generally speaking, the northeast central spatial subset is thick and the southwest part is thin. The Marcellus is under pressured to the southwest and it has been postulated to be normal to potentially over pressured to the northeast with a transitional area in between.

Horizontal well with multi-stage fracture treatments is an established and economically viable method to enable economic gas flow rate from gas shale formations. These massive hydraulic fracture treatments are implemented by injecting million gallons of water per well into the formation though multiple stages, where each stage has 3 to 5
perforation clusters to create a complex hydraulic fracture network. However, the fracture-fluid injected into the formation will not be recovered fully during flowback periods. For Marcellus Shale wells in Northeastern West Virginia, the range of the recovered fracture-fluid within three weeks is from 2 to 26%. The hydraulic fracture-fluid may be trapped or imbibed into the formation or natural fractures due to high capillary pressure. The water recovered to surface have high salt concentration (Blauch and Myers 2009; Haluszczak et al. 2013; Barbot et al. 2013; Carter et al. 2013), which have caused concerns including formation damage, fracture damage, and environment. Therefore, it is desired to know what factor controls the flowback recovery and how much fracture-fluid could be recovered with a development plan before making a decision.

Though fracture-fluid flowback is important for post-fracture evaluation and disposal management, work on flowback water characterization and predictive model development has been relatively limited. Some researches have simulated fracture-fluid volumetric recovery during flowback periods in shale gas reservoir (Clarkson and Williams-Kovacs 2013; Alkouh et al. 2014; Bertoncello et al. 2014) employing various methods to predict, based on stimulation details and properties of two phase (water and gas) system the hydraulic fracture half-length and fracture permeability, which are, in turn, used to predict water and gas production from shale gas reservoirs. Those models are developed on different physical assumptions for fracture geometry and flow regimes. However, limitations of such modeling approaches to capture the full complexity of these systems, and challenges in adequately representing the uncertainty and variability in these systems suggests that complementing modeling with data driven approaches may aid in developing important insights.
Data mining approaches can avoid issues of physical assumptions and such as those made from above simulation works and extract the implicit and interesting information from real production data in large database. Data mining approaches were first applied in shale reservoirs for gas/oil production analysis (Centurion 2011; Centurion et al. 2012; Centurion et al. 2013; Cunningham et al. 2012; Esmaili et al. 2012a, 2012b; LaFollette and Holcomb 2011; LaFollette et al. 2012; Shelly 2008), but no such studies were found in the literature for fracture-fluid flowback analysis. Lan et al. (2014) performed experimental analysis for fracture-fluid flowback in shale reservoirs based on samples from limited wells from shale members of the Horn River Basin and Montney tight gas formation. The controlling factors for fracture-fluid flowback volume change have not been clearly identified by previous published research, and correlations between fracture-fluid flowback and geology, completion and stimulation parameters have not been adequately investigated.

This paper is focused on assessing trends in flowback recovery efficiency through data mining techniques. The analysis includes two parts: univariate engineering analysis used to identify correlations of completion and stimulation parameters that impact flowback recovery efficiency and a multiple variable linear regression model with variable importance calculation used to identify the controlling engineering factor that impacts flowback recovery efficiency. In the second part, the effects from geological conditions were examined by principal component analysis and clustering analysis. The analysis procedures and methods used in this paper are discussed in following part.
5.2 Methods

5.2.1 Data Gathering, Screening and Grouping

Details of the dataset development and screening criteria are provided in Zhou et al. (2014). Completion parameters detailed in the database include proppant type, proppant mass, fracture-fluid type, fracture-fluid volume, vertical depth, lateral length, number of stages, number of perforation shots, perforation intervals, treatment rate, average treatment pressure, maximum treatment pressure, instantaneous shut-in pressure (ISIP), acid volume and shut-in time. Because of insufficient data in some wells, only the fracture-fluid volume, proppant mass, vertical depth, lateral length, treatment rate, number of stages and shut-in time were considered as variables for this study. Table 5-1 shows the mean and ranges for each of those parameters over the entire spatial domain and in each geological group. The shut-in time was the number of days between end of treatment and beginning of flowback. Some wells with long calculated shut-in times may be attributed to a delay in well connection to production line. Volumetric flowback recovery efficiency was calculated as three weeks recovered fracture-fluid volume divided by total injected fracture-fluid volume. The three weeks cutoff was referred to Barbot’s work (2013). In Barbot’s paper (cited 77 times), flowback is defined as 20 days that may align with the practice in Marcellus. In addition, the three weeks cutoff was also supported by simulation results in Marcellus Shale (Seales et al. 2016). Max developed and validated a coupled fluid transport and flowback water ionic compositional model for modeling fracture-fluid clean up processes and simulating flowback water ionic composition by dual porosity paradigm with gas slippage and desorption effects.
According to Max’s model, around 80% of fracture-fluid recovered within first three weeks; around 100% increasing of chloride and potassium concentration happened in first three weeks. Finally, it should be noted that all the wells in this paper are multi-staged horizontal wells stimulated by slickwater fracturing.

Before proceeding with data analysis, the outliers were detected using the Mahalanobis distance approach (Mahalanobis 1963), which is defined as the distance between a data point and a multi-dimension space's centroid, with the centroid determined as a function of the mean values of all variables. Points with the Mahalanobis distance larger than a certain cutoff were treated as outliers. The outliers are assumed to bias the analysis result and were removed from the dataset. After this screening, the data set decreased from 187 to 173 observations. However, the 173 wells retained after this screening were further scrutinized to identify the set of wells most appropriate for use in this analysis. Some wells calculated shut-in periods of negative duration, those observations were removed. Additionally, to make the comparison more clearly, wells with flowback periods within three weeks were chosen as referred to Barbot’s paper (2013) mentioned above. As a result, the total well count is reduced to 86; all the following analyses are based on those 86 wells.

According to the thermal maturity distribution and the formation thickness distribution (Bruner and Smosna 2010), we divided the study domain into four groups (Fig. 5-1). The detailed partition rules and principles are discussed in Zhou et al. (2014). Table 5-2 listed the number of wells and the thermal maturity range and thickness range in each geological group.
Table 5-1 Summary of mean and range for stimulation completion parameters and production

<table>
<thead>
<tr>
<th>Name of parameters</th>
<th>Whole area</th>
<th>A1</th>
<th>A2</th>
<th>B1</th>
<th>B2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fracture-fluid Applied (million gallon)</td>
<td>Mean 5.2</td>
<td>6.0</td>
<td>5.1</td>
<td>5.1</td>
<td>4.5</td>
</tr>
<tr>
<td></td>
<td>Range 2.8-9.5</td>
<td>3.9-9.5</td>
<td>2.9-7.8</td>
<td>3.2-8.4</td>
<td>2.8-7.2</td>
</tr>
<tr>
<td></td>
<td>Mean 4.7</td>
<td>5.7</td>
<td>4.0</td>
<td>4.7</td>
<td>4.5</td>
</tr>
<tr>
<td>Proppant (million lbm)</td>
<td>Range 0.9-7.9</td>
<td>3.8-7.2</td>
<td>0.9-7.9</td>
<td>3.1-7.7</td>
<td>2.4-7.3</td>
</tr>
<tr>
<td></td>
<td>Mean 12</td>
<td>12</td>
<td>13</td>
<td>12</td>
<td>12</td>
</tr>
<tr>
<td>Number of Stages</td>
<td>Range 7-24</td>
<td>8-16</td>
<td>7-21</td>
<td>7-21</td>
<td>8-24</td>
</tr>
<tr>
<td></td>
<td>Mean 87</td>
<td>83</td>
<td>95</td>
<td>87</td>
<td>82</td>
</tr>
<tr>
<td>Treatment Rate (bpm)</td>
<td>Range 70-101</td>
<td>75-90</td>
<td>84-101</td>
<td>76-94</td>
<td>70-94</td>
</tr>
<tr>
<td></td>
<td>Mean 7157</td>
<td>7,149</td>
<td>6,650</td>
<td>7,308</td>
<td>7,521</td>
</tr>
<tr>
<td>Vertical Depth (ft)</td>
<td>Range 6388-8080</td>
<td>6,875-7,346</td>
<td>6,388-7,435</td>
<td>6,836-7,720</td>
<td>7,158-8,080</td>
</tr>
<tr>
<td></td>
<td>Mean 4628</td>
<td>5,248</td>
<td>4,050</td>
<td>5,167</td>
<td>4,047</td>
</tr>
<tr>
<td>Lateral Length (ft)</td>
<td>Range 2060-6763</td>
<td>3,673-6,500</td>
<td>2,060-6,627</td>
<td>2,846-6763</td>
<td>2,762-5,868</td>
</tr>
<tr>
<td></td>
<td>Mean 19</td>
<td>29</td>
<td>20</td>
<td>9</td>
<td>16</td>
</tr>
<tr>
<td>Shut-in Time (days)</td>
<td>Range 0-144</td>
<td>0-144</td>
<td>0-48</td>
<td>0-46</td>
<td>1-65</td>
</tr>
<tr>
<td>Flowback Recovery (vol %)*</td>
<td>Mean 6.6</td>
<td>5.7</td>
<td>3.1</td>
<td>9.7</td>
<td>7.8</td>
</tr>
<tr>
<td></td>
<td>Range 0.2-25.9</td>
<td>0.2-13.8</td>
<td>1.0-5.6</td>
<td>4.6-25.9</td>
<td>3.1-14.5</td>
</tr>
</tbody>
</table>

*Flowback recovery is defined as the volume fraction of injected fracture-fluid that is recovered within three weeks after a treatment, exclusive of shut-in-time; Vertical Depth is true vertical depth; Lateral length is the perforation interval

Table 5-2 Thermal maturity (Ro) and formation thickness (h) ranges in four groups.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>h (ft)</td>
<td>40-60</td>
<td>40-60</td>
<td>60-80</td>
<td>80-100</td>
</tr>
<tr>
<td>Ro (%)</td>
<td>1.0-1.3</td>
<td>1.3-1.6</td>
<td>1.6-2.5</td>
<td>1.6-2.5</td>
</tr>
</tbody>
</table>
Figure 5-1 Group A1, A2, B1, B2 distribution on the map of West Virginia

Fig. 5-2 shows a simplified schematic illustrating the analysis procedures of this paper. The workflow consists of four major parts as marked using different colors. First, the dataset was constructed by collecting the data from both public and confidential sources, including eight parameters, and then partitioned into four geological groups as discussed in the Data Gathering, Screening and Grouping section. In a third step, bivariate analysis was performed on the data; a fourth step was a more complex and comprehensive multiple parameter analysis. Results of each part of this workflow are discussed for both geological and engineering parameters.
Figure 5-2 Workflow of this paper
5.2.2 Principle Component Analysis

To understand how geological factors affect the gas and flowback recovery, two methods are used. One is K-Mean Clustering and another one is Principle Component Analysis.

The principle components are relatively small numbers of representative variables to achieve dimension reduction for visualization, such as 2-D visualization. In this study, the engineering parameters are transferred as two principle components for clustering analysis.

The first principal component direction of the data is that along which the observations vary the most. The principal component of a set of features $X_1, X_2, \ldots, X_p$ is the normalized linear combination of the features as shown in equation 2, where $\phi$ is the loadings of principle component.

$$Z_1 = \phi_{11}X_1 + \phi_{21}X_2 + \cdots + \phi_{p1}X_p \quad (5.1)$$

In this research, the first and second principle components are shown as below:

$pcaComp1 = \phi_{11} \times \text{Fracture fluid Volume} + \phi_{21} \times \text{Proppant} + \phi_{31} \times \text{Depth}$
$+ \phi_{41} \times \text{Lateral Length} + \phi_{51} \times \text{Number of Stages} + \phi_{61} \times \text{Treatment Rate}$

$pcaComp2 = \phi_{12} \times \text{Fracture fluid Volume} + \phi_{22} \times \text{Proppant} + \phi_{32} \times \text{Depth}$
$+ \phi_{42} \times \text{Lateral Length} + \phi_{52} \times \text{Number of Stages} + \phi_{62} \times \text{Treatment Rate}$
5.2.3 K-Mean Clustering Analysis

In this study, the K-mean clustering method is chosen for partitioning the dataset into subgroups by engineering parameters because of its simplicity and efficiency. Each cluster shares similar engineering characteristics but different geological conditions. One cluster will be picked to study how gas or flowback recovery differs under different geological settings.

In K-means clustering, the dataset is randomly divided into K clusters at the beginning, each with a centroid. These centroids represent the mean of all variables in each cluster. The process of K-mean is to calculate the distance between each observation and the centroid, and add them up. The final clustering result is achieved when the sum of the distances between observed value and centroid in each cluster is at minimum after several iterations. This optimization process can be summarized as following equation, where Ck means the number of clusters, \((x_{ij} - x_{i\eta})\) is the Euclidean distance between every two observations in k cluster.

\[
\minimize_{C_1, \ldots , C_k} \left\{ \sum_{k=1}^{C_k} \frac{1}{C_k} \sum_{i, \eta \in C_k} \sum_{j=1}^{p} (x_{ij} - x_{i\eta})^2 \right\}
\] (5.2)

5.2.4 Regression Analysis

The multiple variables analysis is to study the correlation between one engineering parameter and gas/flowback recovery considering other engineering parameters effects. It is achieved by constructing a multiple linear regression function
without considering the prediction accuracy test. Then the coefficient of determination is obtained between each engineering parameter and gas/flowback recovery.

The goal of regression is to analyze correlation between multiple variables. The R Squared is the ratio of sum of regression error over the sum of total error (5.4), which represents the proportion of the behavior of the dependent variable that is explained by independent variables. SSTO (total sum of squares) measures the total variation, which is calculated as the sum of squared of the difference between observation value ($Y_i$) and its corresponding mean value ($\bar{Y}_i$). SSR stands for regression sum of squares, which is calculated the sum of squared of the difference between the mean value ($\bar{Y}_i$) and fitted value ($\hat{Y}_i$).

$$R^2 = \frac{SSR}{SSTO} = \frac{\sum(Y_i - \bar{Y}_i)^2}{\sum(Y_i - \bar{Y}_i)^2} \quad (5.4)$$

The method proposed by Lindeman, Merenda, and Gold (1980) (LMG) is used to quantify the variable importance in cases where they are correlated. This approach is chosen because of its easy application and the limitation on public data availability. In this research, it is applied on the coefficient of determination obtained from multiple variables correlation analysis.

It is the averaging sequential sums of squares over all orderings of predictors. The term “sequential” means that the predictors are entered into the model in the order in which they are listed. Division of sequential sums of squares by the total response sum of squares yields sequential R-Squared contributions. This method takes into account the dependence on orderings by averaging over orderings, using simple un-weighted
averages. Equation (5.5) shows the corresponding function, where \( \text{lmg}(x_k) \) is the variable importance of \( x_k \) calculated by LMG method, seqR^2 is the sequential sums of square, \( n \) is the data size, \( p \) is the number of predictors, \( S \) is the dataset excluding \( x_k \).

\[
\text{lmg}(x_k) = \frac{1}{p!} \sum_{S \subseteq \{x_1, \ldots, x_p\} \setminus \{x_k\}} n(S)! (p - n(S) - 1)! \text{seqR}^2 (\{x_k\}|S) \quad (5.5)
\]

5.3 Results and Discussion

5.3.1 Single Parameter Study

A set of simple univariate analyses were performed to examine the correlation of each parameter with the cumulative water recovery within 3-week flowback to find potentially statistically significant correlations between the 3-week fracture fluid flowback and well geologic, stimulation, and completion features.

5.3.1.1 Geologic Parameters

The difference in geological setting could have significant effects on flowback performance. Thermal maturity, used as a practical index for the hydrocarbon composition and hydrocarbon phases in the reservoir, is important for the 3-week flowback evaluation. In addition to thermal maturity, formation thickness also significantly impacts 3-week flowback performance.
Figure 5-3 Three-week flowback recovery between Group A (52 cases) and Group B (34 cases); the flowback recovery increased as $R_o$ increased.

Figure 5-4 Three-week flowback recovery between Group B1 (17 cases) and Group B2 (17 cases); the flowback recovery decreased as $h$ increased.

**Fig. 5-3** shows a comparison of the results on 3-week flowback recovery in Groups A and B. In Group A, the average 3-week flowback recovery is 4.46%, while is 8.75% in Group B. This relatively large difference between Group A and Group B supports the decision to define this geological boundary. **Fig. 5-4** shows the average 3-week flowback recovery within group B. It seems that as the thickness increase, the flowback recovery decreases. To further understand the relationship between formation thicknesses with the flowback recovery, the stimulation and completion parameters need to be considered as well.
5.3.1.2 Engineering Parameter

Stimulation operational choices were also analyzed to explore which engineering parameters have significant influence on the flowback recovery (magnitude of correlation as absolute value), and the type of correlation (positive or negative). **Fig. 5-5** shows results of that univariate correlation analysis; the coefficients of correlations (r) between single engineering parameters (total fracture fluid volume, proppant mass, vertical depth, lateral length, number of stages, treatment rate, shut-in time) and 3-week flowback recovery are presented in Table 5-3.

<table>
<thead>
<tr>
<th>Fracture Fluid</th>
<th>Proppant Mass</th>
<th>Vertical Depth</th>
<th>Lateral Length</th>
<th>Number Stage</th>
<th>Treatment Rate</th>
<th>Shut-in Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>r</td>
<td>-0.1637</td>
<td>0.0749</td>
<td>0.0628</td>
<td>0.3713</td>
<td>-0.1416</td>
<td>-0.3157</td>
</tr>
<tr>
<td>P-value</td>
<td>0.1275</td>
<td>0.4879</td>
<td>0.5610</td>
<td>3.6852E-04</td>
<td>0.1880</td>
<td>0.0027</td>
</tr>
</tbody>
</table>

(T-critical=1.96 (two-tail), 95% confidence interval, the correlation is statistically significant if t-test>1.96, or t-test<-1.96, or P-value <0.05)

Those results show that the shut-in time has the largest coefficient of correlation compared with other parameters, followed by lateral length, with r value as -0.3828 and 0.3713, respectively. The vertical depth, proppant mass applied, and lateral length exhibit positive correlation with the 3-week flowback recovery- *i.e.*, the flowback recovery increases as those parameters increase. Other parameters, such as fracture-fluid volume, number of fracture stages, treatment rate and shut-in time have negative relationship with the 3-week flowback recovery – *i.e.*, the flowback recovery decreases as these parameters increase. Fracture-fluid volume, proppant mass, vertical depth and number of stages,
have P-value larger than 0.05, and therefore are determined to not have statistically significant relationship with the flowback recovery. All of those correlations are weak (even the ones with P value larger than 0.05). Still, this analysis provides some insight into the general trend between each stimulation or completion parameter and the 3-week flowback recovery of Marcellus shale in Northeastern West Virginia.

Figure 5-5 Analysis of single engineering parameter on 3-week flowback recovery, and shut-in time has a negative correlation with flowback recovery (86 cases).
5.3.2 Multiple Parameter Study

The lack of strong correlation between the reservoir and stimulation characteristics and well flowback recovery from univariate correlation analysis may stem from a lack of data normalization for geologic and engineering parameters and the colinearity among those variables. In an attempt to overcome these issues and identify meaningful patterns in the data, multi-variable regression, LMG, principal component analysis and clustering methods are considered in combination.

5.3.2.1 Engineering Parameter Ranking

To identify key engineering factors affecting the well flowback performance, multiple regression analysis between 3-week flowback and engineering parameters within each geological group is implemented. In this analysis, it is assumed that within each geological group, the geological characteristics are similar and the wells are operated using the same strategy. Since, the engineering parameters units vary significantly between wells; a scaling process is applied before multiple regression analysis is performed. To make the comparison more meaningful, the total fracture-fluid volume, total proppant mass and lateral length is divided by number of stages. Table 5-4 shows the correlation between each factor with the water production performance by the sign and value of coefficient. The multiple regression equation for each group is shown below. In order to overcome the range problem of flowback recovery, the logistic transformation was applied (Kutner et al. 2004). Equation (5.6) could be applied to estimate 3-week flowback with confidence level as shown in Table 5-4.
\[
\log \left( \frac{\text{recovery}}{1 - \frac{\text{recovery}}{100}} \right) = b_0 + b_1 \text{Fracture Fluid/Stage} + b_2 \text{Proppant Mass/Stage} \\
+ b_3 \text{Lateral Length/Stage} + b_4 \text{Vertical Depth} + b_5 \text{Average rate} + b_6 \text{Shut-in time}
\]

(5.6)

Table 5-4 Coefficients of multiple regression results for each group

<table>
<thead>
<tr>
<th>Group</th>
<th>Intercept</th>
<th>Fracture Fluid/Stage</th>
<th>Proppant Mass/Stage</th>
<th>Lateral Length/Stage</th>
<th>Vertical Depth</th>
<th>Average Rate</th>
<th>Shut-in Time</th>
<th>( R^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group A1</td>
<td>1.4663</td>
<td>-0.3510</td>
<td>0.0689</td>
<td>-0.1605</td>
<td>-0.0005</td>
<td>0.0356</td>
<td>-0.8348</td>
<td>0.69</td>
</tr>
<tr>
<td>Group A2</td>
<td>1.0436</td>
<td>-0.0590</td>
<td>0.3108</td>
<td>-0.1409</td>
<td>-0.1114</td>
<td>-0.2784</td>
<td>-0.2085</td>
<td>0.65</td>
</tr>
<tr>
<td>Group B1</td>
<td>2.2632</td>
<td>-0.3051</td>
<td>-0.0507</td>
<td>-0.0849</td>
<td>-0.1473</td>
<td>-0.1196</td>
<td>-0.1190</td>
<td>0.38</td>
</tr>
<tr>
<td>Group B2</td>
<td>2.0397</td>
<td>-0.5421</td>
<td>0.6277</td>
<td>0.0814</td>
<td>0.0475</td>
<td>-0.1654</td>
<td>-0.2112</td>
<td>0.70</td>
</tr>
</tbody>
</table>

For each geological group, the variable importance was calculated as shown in Appendix A. The factor of greatest importance for each of the four defined geological subgroups, and the corresponding variable importance value are shown in Fig. 5-6. The most important engineering factor for Group A1 and Group B2 is the shut-in time; for Group A2 and Group B2 the key factors are proppant mass/stage and fracture-fluid volume/stage, respectively.

To further understand what causes the different dominated engineering factor results, the proppant size and proppant concentration of each geological group were studied. The proppant size data based on single well scale is insufficient to do data
mining analysis; in each group, the number of wells with available proppant size information is 15, 0, 25, and 12. However, different proppant size in each group may explain why different engineering factors dominate the flowback recovery. For Group A1 and Group B2, the proppant size is 100, 40/70 and 30/50. For Group B1, the proppant size is 100, 40/70, 30/50 and 20/40. There is no information available for Group A2. In other words, Groups A1 and B2 share the same proppant size, and Group B1 contains larger proppant size as 20/40. Although not directly considered in the multiple regression analysis, proppant concentration is another factor significantly impacting fracture conductivity. **Fig. 5-7 to 5-10** shows the proppant concentration in each geological group. The proppant concentration is similar among Group A1, Group B1 and Group B2.

Therefore, the only difference among Groups A1, B1 and B2 is proppant size. According to the proppant size, Group A1 and B2 show characteristics of typical slickwater fracturing to carry proppant of small size. Slick-water fractures typically generate longer hydraulic fractures with larger matrix-fracture contact area and lower fracture conductivity. For those wells, flowback is most sensitive to shut-in time. Longer shut-in time may results in more fracture-fluid imbibition in matrix or natural fracture, and returning less to surface.

Larger size proppant, such as 20/40 mesh proppant (420 - 840 µm) applied in Group B1, may result in smaller proppant area (i.e., smaller matrix-fracture contact area) and higher fracture conductivity near wellbore (Ahn et al. 2014). Good fracture conductivity is important for fracture-fluid flowback to surface. The average shut-in time in Group B1 is shortest compared to Group A1 Group A2 and Group B2, and the average fracture-fluid flowback recovery is highest in Group B1, as shown in **Table 5-1**. The
shut-in time may be varied less in Group B1 and the flowback recovery may be more directly affected by the fracture-fluid volume/stage as the fracture with good conductivity.

In Group A2, the concentration distribution of applied proppant is smaller than that in the other three geological groups. In other words, with the same fracture-fluid volume, the total proppant mass applied would be less in Group A2 as compared with the other three geological groups. Because proppant is required to hold open stimulated fractures, insufficient proppant loading may result in increased fracture closure during flowback period and reduction in water recovery efficiency. Relatively low proppant mass loading is, therefore, proposed to be the dominated factor influencing flowback performance in Group A2.

![Figure 5-6](image-url) **Figure 5-6** Key factors and corresponding variable importance value of each geological group
Figure 5-7 Proppant concentration distribution for Group A1 (28 cases). Around 86% wells in this group with proppant concentration ranges from 1.2 to 1.47 lbm/gal.

Figure 5-8 Proppant concentration distribution for Group A2 (24 cases). Around 17% wells in this group with small proppant concentration as 0.4 lbm/gal, which is not observed in other three groups.
Figure 5-9 Proppant concentration distribution for Group B1 (17 cases). Around 72\% wells in this group with proppant concentration ranges from 1.2 to 1.47 lbm/gal.

Figure 5-10 Proppant concentration distribution for Group B2 (17 cases). Around 82\% wells in this group with proppant concentration ranges from 1.2 to 1.47.
5.3.2.2 Geological Parameter

The effect of geological properties on 3-week flowback recovery was also examined using PCA and K-means clustering methods to normalize the effects from stimulation and completion parameters, and generate different clusters of similar stimulation and completion parameters values.

The effect of thermal maturity was first examined. PCA and clustering methods were applied on GroupA wells, which share the same formation thickness range but have different thermal maturity. The PCA and clustering results are shown in the Appendix B. The 3-week flowback recovery from Groups A1 and A2 was compared in one of the clusters. The flowback comparison result shown in Fig. 5-11. There are 8 wells picked from Group A1 and 5 wells picked from Group A2. Those wells share similar formation thickness and engineering parameters (fracture-fluid, proppant, vertical depth, lateral length, number of stages, treatment rate, shut-in time), but different Ro ranges. The Ro of Group A1 is smaller than Group A2 – i.e., Group A2 has a relatively high thermal evolution. As a result, the 3-week flowback recovery of Group A1 is slightly smaller than that in Group A2. The difference may be result from the multi-phase flow. In Group A1, more condensate will be produced in the formation and may inhibit the fracture-fluid flowback as the effect of capillary pressure and relative permeability (Alfi et al. 2014). The clay content may be another cause of this phenomenon. The clay content in A1 is higher than A2, and it may be expected that, in this case, more water will be absorbed on clay surface and result in lower flowback recovery in A1 than A2.
The effect of formation thickness was then considered. Based on geological grouping, the Groups B1 and B2 have the same Ro and different thickness. In order to eliminate the effect of engineering attributes on wells in B1 and B2, we now pick wells in B1 and B2 that share similar engineering characteristics by using PCA and K-Mean. The PCA is transferring the seven engineering attributes, including fracture-fluid volume, proppant mass, lateral length, vertical depth, number of stages, shut-in time and treatment rate, into two principal components as shown in the Table 5-4 in Appendix B. The K-Mean method is applied on the two principal components and identified the wells in B1 and B2 which have similar engineering characteristics. As Fig. 5-12 shows, there are 11 wells selected from Group B1 and 10 wells selected from Group B2. Those wells share similar thermal maturity and engineering parameters. The 3-week flowback recovery in Group B2 is lower than the production in Group B1 – a difference which may be attributed to difference in formation thickness. In region B, the clay content is much less than in region A and the shale matrix should, therefore, be more brittle. The reason for different flowback recovery may be due to different fracture surface area. The thicker formation may have larger fracture surface area than the thin formation, and with that larger interfacial surface area more fracture-fluid will imbibe into the matrix by capillary pressure.
Figure 5-11 Compared results between Group A1, A2

Figure 5-12 Compared results between Group B1, B2
5.4 Section Conclusions

In this study, geological and engineering data of 187 Marcellus wells in Northeastern West Virginia were collected and classified into two groups and four subgroups based on formation thickness and thermal maturity. Wells in Group A produce gas and oil (or condensate), while wells in Group B yield dry gas. The production of Group A1 contains a significantly higher fraction of liquid hydrocarbon than Group A2. Based on our evaluation of 3-week fracture-fluid flowback, conclusions are as follows:

- Data screening was carried out by Mahalanobis distance analysis to identify outliers in whole dataset. Simple variable analysis results show the general trend between 3-week flowback recovery with other geological, stimulation and completion parameters. The 3-week flowback recovery exhibits positive relationship with proppant mass, lateral length and vertical depth, and negative relationships with fracture-fluid volume, number of stage, treatment rate and shut-in time. Correlations were developed for estimating 3-week flowback from wells in each of the four geological groups with different confidence intervals.

- Based on multiple regression analysis and individual variable importance measurement in each geological group, the shut-in time, proppant size and proppant mass are found to be the significant parameters for predicting volumes of water produced within 3-week after a fracture treatment.

- In Group A1 and Group B2, shut-in time is relatively important compared to other parameters. Groups A1 and B2 are typical slick-water fracturing to carry proppant of small size and generate larger matrix-fracture contact area and lower fracture
conductivity. For those wells, flowback is most sensitive to shut-in time. Longer shut-in time may result in more fracture-fluid imbibition to matrix.

- In Group B1, the key engineering factor is fracture-fluid volume/stage. This difference may be caused by the different proppant size applied in this area. Compared to Groups A1 and B2, larger proppants are used in Group B1, and may result in smaller proppant area and higher fracture conductivity. Therefore, the flowback recovery is more sensitive to fracture-fluid volume/stage. Some other factors, such as pressure difference between pore pressure and in-situ stress which is not available in this study may be reason that causes the low flowback recovery. Further study is needed.

- In Group A2, the key factor is proppant mass/stage besides treatment rate. The proppant concentration is smaller compared to other three groups, which may result in fracture closure during flowback period. Therefore, higher proppant mass ensures open paths for fluid flow.

- Based on principal component analysis and k-mean clustering in database, the effects from thermal maturity and formation thickness on 3-week fracture-fluid flowback were examined after eliminating the contribution from engineering parameters.

- The 3-week fracture-fluid flowback recovery increases as the thermal maturity increases, which may be explained as the effect of multi-phase flow. More condensate produced may inhibit the fracture-fluid flowback as affect of capillary pressure and relative permeability.
The 3-week fracture-fluid flowback recovery decreases as the formation thickness increases, which may be explained as the fracture surface area. The thicker formation may have larger fracture surface area than the thin formation, which will imbibe more fracture-fluid into the matrix by capillary pressure.
5.5 Section References


Meeting, Lexington, Kentucky, 3-5 October. SPE-161184-MS. http://dx.doi.org/10.2118/161184-MS.


http://dx.doi.org/10.1016/j.jngse.2014.06.014.
CHAPTER 6

Spatial and Temporal Relationships between Produced Gas and Water in Marcellus through Data Mining Analysis

Abstract

The relationship between produced gas and flowback/produced water is important for evaluating shale gas well performance; however, it is not fully understood yet due to complex flow mechanisms and interactions / feedback among various geoscience and engineering controls. Further investigation would provide valuable insight to adjust development plans to achieve optimal well/regional economic production.

In this study, an auto-updated nonlinear model method was applied to evaluate the relationship between water and gas in different spatial and temporal domains and to understand the micro-scale flow mechanisms from macro-scale data. Fracture-fluid flowback data in the dataset are water produced within one month, following a fracture treatment (exclusive of well shut-in time), and the produced water were 1 to 3 years. 114 wells from the Marcellus Formation in northwestern West Virginia were selected to investigate the relationship between fracture-fluid flowback and one month gas production in different spatial domains (wet and dry gas regions). 67 Marcellus wells in Lycoming County, Pennsylvania were selected to study the relationship between produced water and gas production across different time periods ranging from one to three years. The results indicate that the relationship between gas and fracture-fluid flowback in the wet gas region is positive while negative in the dry gas region. WGR (water gas ratio) is high (> 9 bbl/mmcf) during the 1st-year which indicated water be
carried out through displacement and leveled off at 3 bbl/mmcf after the 1st-year, indicating evaporation is the primary mechanism for water production. This study analyzed the relationship between gas and water production under different geological conditions and time periods and offers new insights on gas and fracture-fluid/produced water flow mechanisms in shale gas reservoirs.

6.1 Introduction

Fracture-fluid flowback is one of the most important indicators of shale gas well performance. However, it is uncertain how fracture-fluid flowback affects gas production and also unclear how produced water affects gas production after flowback periods under specific conditions. The primary problems in this study can be stated as:

- What’s the relationship between fracture-fluid flowback and early short-term gas production?
- What’s the relationship between produced water and long-term gas production?

The previous publications related to the above two problems could be classified as field analysis, experimental investigation, and simulation analysis. Two major characteristics of fracture-fluid flowback from field observation are: low fracture-fluid recovery and high salt concentration (King 2015). The relationship between fracture-fluid flowback with short-term gas production was observed in different trend. The field
analysis by Leonard et al. (2007) compared 6 wells in 3 areas of Barnett Shale indicated that higher fracture-fluid flowback (10 – 30 days) wells had lower gas production (second month production). The observation found from Curry et al. (2010) suggested the opposite situation. They compared 2 wells with similar completion design and hydraulic fracturing treatment in the Marcellus Shale indicating higher fracture-fluid flowback (around 12%) would result in higher early gas production (within two weeks, >5 mmcf/d). The well with lower fracture-fluid flowback (around 2%) has lower early gas production (<1.5 mmcf/d). Observations from long-term production data (Zhang and Ehlig-Economides 2014) show wells with higher recovery of produced water have higher gas production (6 out of 15 wells in Barnett shale). The rate of increase for salt concentration flattens later in the production period after fracture-fluid flowback procedures (Zhang and Ehlig-Economides 2014). Because of limited field data in previous studies, the relationship for both short and long-term could not be determined.

In order to understand the physics behind these field observations, more and more experimental studies were performed to understand the transport mechanisms on paths from hydraulic fracture to micro-fracture network/natural fracture to matrix. According to core and field observations, the pore or throat size of shale is extremely small and salt concentration in fracture-fluid flowback is higher compared to that in injected slick water. Therefore, the principal driving forces were interpreted to be capillary pressure and osmotic pressure. Most of the laboratory studies confirmed extremely high capillary pressure existed in shale matrix, which caused the spontaneous imbibition, drawing fracture-fluid deeper into the formation and expelling hydrocarbons to the wellbore. The imbibition processes could be happened from hydraulic fracture to natural fractures
(micro fracture network) (Roychaudhuri, Tsotsis, and Jessen 2011; Parmar et al. 2012, 2013) or natural fracture to matrix (Takahashi and Kovscek 2009; Kelly 2013; Kelly et al. 2015, Li et al. 2016). The imbibition process is faster early on but slows down later (Roychaudhuri, Tsotsis, Jessen 2011), and it is significantly affected by pore size and interfacial tension (wettability) (Takahashi and Kovscek 2009). Except for high capillary pressure, the effects of other forces, such as gravity, intermolecular force, and osmotic pressure, were investigated as well. The laboratory results indicated that water recovery from propped fractures is very low when the flow direction is against gravity (Parmar et al. 2012, 2013). The intermolecular force was observed at nano scale pores, which could be the counterforce to decrease imbibition and resist gas flow (Kelly 2013; Kelly et al. 2015). The osmotic pressure caused by the chemical concentration difference between injected fracture-fluid and initial brine may induce drainage processes after imbibition (Li et al. 2016).

Simulation studies, which were developed at larger scales, do not completely agree with the conclusions of the previously mentioned experimental studies. The simulation studies agree that capillary pressure is a significant force which caused spontaneous imbibition in shale formations, however, they disagree on how water imbibition affected gas flow. Some studies stated that imbibed water (which could be residual fracture-fluid or leak off fracture-fluid) has the ability to extrude hydrocarbon (gas) flow from matrix to hydraulic fractures, and in turn, improve early gas production (Bertoncello et al. 2014). Those studies believed well shut-in periods improve early gas production and reduce water production. Others works stated that the residual, or leak off fracture-fluid, could cause damage to formations at the intersection of fractures and
matrix as a result of water blockage, which may cause a reduction in relative gas permeability (Fakcharoenphol et al. 2013), or it may cause clay swelling in deeper formations, which may cause irreversible reductions in absolute permeability (surface mobilization and subsequent pore-throat plugging) due to water salinity gradient (Civan 2014).

Previous studies suggest that fracture-fluid flowback is a complex process controlled by fracture and matrix relative permeability, fracture conductivity, capillary pressure, wettability, pore size, and pore size distribution. Because of data limitations and narrow approaches, previously filed data analyses show conflicting results between each other. Information obtained from either experimental or simulation studies still cannot provide a clear understanding of the mechanism of flow pattern in gas shale. Data mining method seems to be a good alternative approach to investigate the multiple variable controls. To get a better understanding about how fracture-fluid flowback impacts gas production, field data from hundreds of wells (114 horizontal multi-stage hydraulic fractured wells in West Virginia and 67 horizontal multi-stage hydraulic fractured wells in Pennsylvania) was investigated in this report. All wells in this study were treated by slickwater hydraulic fracturing.

Production Data from 114 wells from West Virginia and 67 wells from Pennsylvania were collected for this study. The 114 wells from West Virginia contain one-month production data and were used to study the relationship between fracture-fluid flowback and short-term gas production. Fig. 6-1 and Fig. 6-2 show the spatial distribution of one-month gas and fracture-fluid values separately. A boundary could be observed on the two figures. It is assumed that the flow mechanisms on the two sides are
different and will result in different relationships. The boundary was defined in previous works (Zhou et al. 2014, 2016) and was used to divide 114 wells into two subsets: one in the wet gas region, and the other in the dry gas region.

Figure 6-1 Map of gas production in West Virginia wells. Each square points represent one well with different one-month gas production as shown in different color. The blue circle contains wells in wet gas region which have relative high gas production. The red circle contains wells in dry gas region which have relative low gas production. The black dashed line is the potential boundary between wet gas region and dry gas region.
Figure 6-2 Map of fracture-fluid flowback in West Virginia wells. Each square points represent one well with different one-month fracture-fluid in bbl as shown in different color. The blue circle contains wells in wet gas region which have relative high gas production. The red circle contains wells in dry gas region which have relative low gas production. The black dashed line is the potential boundary between wet gas region and dry gas region.

67 wells from Pennsylvania were analyzed to study the relationship between produced water and gas production data from 1<sup>st</sup>- to 3<sup>rd</sup>-year. According to the results of previous studies (Zhang and Ehlig-Economides 2014), differences in the declining trend of water gas ratio (WGR) may indicate different flow mechanisms of water such as displacement or vaporization, and the shift in the declining trend of WGR occurred after 1<sup>st</sup>-year. The WGR experienced a sharply declining trend in the 1<sup>st</sup>-year, which then began to flatten.
Table 6-1 Water gas ration of wells in Pennsylvania

<table>
<thead>
<tr>
<th>WGR (bbl/mmcf)</th>
<th>1st Year</th>
<th>2nd Year</th>
<th>3rd Year</th>
<th>1st Year</th>
<th>2nd Year</th>
<th>3rd Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>2nd Year with WGR &lt; 6</td>
<td>9.1</td>
<td>3.2</td>
<td>3.0</td>
<td>23.6</td>
<td>13.6</td>
<td>9.5</td>
</tr>
<tr>
<td>2nd Year with WGR &gt; 6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 6-2 Mean and range for completion design parameters used

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Injected Fluid</td>
<td>mean</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>106*gal</td>
</tr>
<tr>
<td>Injected Proppant</td>
<td>mean</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>4</td>
<td>106*lbm</td>
</tr>
<tr>
<td>Pump Rate</td>
<td>mean</td>
<td>87</td>
<td>82</td>
<td>70</td>
<td>70</td>
<td>bbl/min</td>
</tr>
<tr>
<td>Stage Number</td>
<td>mean</td>
<td>12</td>
<td>12</td>
<td>12</td>
<td>11</td>
<td>-</td>
</tr>
<tr>
<td>Vertical Length</td>
<td>mean</td>
<td>6,949</td>
<td>7,326</td>
<td>9115</td>
<td>9009</td>
<td>Ft</td>
</tr>
<tr>
<td>Horizontal Length</td>
<td>mean</td>
<td>4,779</td>
<td>4,440</td>
<td>3337</td>
<td>3049</td>
<td>Ft</td>
</tr>
<tr>
<td>Shut-in Time</td>
<td>mean</td>
<td>31</td>
<td>11</td>
<td>-</td>
<td>-</td>
<td>Day</td>
</tr>
</tbody>
</table>
This observation was used as a hypothesis to drive the investigation of WGR in each year of all 67 wells. As shown in Table 6-1, the wells with 2\textsuperscript{nd}-year average WGR of less than 6 were grouped in one subset, and the average WGR of those well after 2\textsuperscript{nd}-year leveled off. The rest were grouped in another subset, which continued to decline until 3\textsuperscript{rd}-year. Only the wells where WGR leveled off were used to identify the relationship between yearly gas production and yearly produced water. In the following tables (Table 6-2 & 6-3) there are four groups of wells: Group A contained 58 wells in wet gas region; B contained 56 wells in the dry gas region; C contained 36 wells with leveled off WGR; and D contained 31 wells with continued decline of WGR.

Table 6-2 summarizes the well engineering parameters used in each group. The engineering parameters include fracture-fluid volume, proppant mass, vertical depth, lateral length, treatment rate, number of stages and shut-in time. The fracture-fluid volume, proppant mass and horizontal length was normalized by stage number. Table 6-3 is the summary of production data. The first month cumulative gas production was used to study the relationship with fracture-fluid flowback. The first-year (1\textsuperscript{st}-12\textsuperscript{th} month), second-year (13\textsuperscript{th}-24\textsuperscript{th} month), and third-year (25\textsuperscript{th}-36\textsuperscript{th} month) gas production were used to study the relationship with first-year, second-year and third-year produced water.
### Table 6-3 Mean and range for production data used

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-month gas</td>
<td>mean</td>
<td>56</td>
<td>41</td>
<td>-</td>
<td>-</td>
<td>Mmcf</td>
</tr>
<tr>
<td>1-month flowback</td>
<td>mean</td>
<td>5,776</td>
<td>7,560</td>
<td>-</td>
<td>-</td>
<td>Bbl</td>
</tr>
<tr>
<td>1-month flowback</td>
<td>mean</td>
<td>5</td>
<td>8</td>
<td>-</td>
<td>-</td>
<td>%</td>
</tr>
<tr>
<td>1st-year gas</td>
<td>mean</td>
<td>-</td>
<td>-</td>
<td>1,136</td>
<td>804</td>
<td>Mmcf</td>
</tr>
<tr>
<td>1st-year water</td>
<td>mean</td>
<td>-</td>
<td>-</td>
<td>9,701</td>
<td>13,405</td>
<td>Bbl</td>
</tr>
<tr>
<td>1st-year water</td>
<td>mean</td>
<td>-</td>
<td>-</td>
<td>3</td>
<td>5</td>
<td>%</td>
</tr>
<tr>
<td>2nd-year gas</td>
<td>mean</td>
<td>-</td>
<td>-</td>
<td>629</td>
<td>484</td>
<td>Mmcf</td>
</tr>
<tr>
<td>2nd-year water</td>
<td>mean</td>
<td>-</td>
<td>-</td>
<td>1,936</td>
<td>5,416</td>
<td>Bbl</td>
</tr>
<tr>
<td>2nd-year water</td>
<td>mean</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>2</td>
<td>%</td>
</tr>
<tr>
<td>3rd-year gas</td>
<td>mean</td>
<td>-</td>
<td>-</td>
<td>430</td>
<td>325</td>
<td>Mmcf</td>
</tr>
<tr>
<td>3rd-year water</td>
<td>mean</td>
<td>-</td>
<td>-</td>
<td>1,196</td>
<td>2,280</td>
<td>Bbl</td>
</tr>
<tr>
<td>3rd-year water</td>
<td>mean</td>
<td>-</td>
<td>-</td>
<td>0.5</td>
<td>1</td>
<td>%</td>
</tr>
</tbody>
</table>
6.2 Methods

The methodology used in this study is a nonparametric nonlinear regression model. As mentioned in the data description section, the wells in West Virginia and Pennsylvania were partitioned into two subsets for each location. In each group, several mathematic models were developed to identify the relationship between gas and water. In each model, six engineering parameters (list in Table 6-1) were combined with one gas production value and one corresponding water production value. The six engineering parameters are: (1) injected fracture-fluid volume per stage; (2) injected proppant mass per stage; (3) pump rate; (4) vertical length; (5) horizontal length per stage; and (6) shut-in time, with the two production parameters (water and gas production) normalized by their mean and maximum values minus minimum values into dimensionless units. Each mathematic model can be represented as:

\[ Y = g(s(X)) + \varepsilon \quad (6.1) \]

where \( Y \) is the gas vector, \( X \) is the matrix contained water and the completion design parameters, and \( \varepsilon \) is the error term usually assumed to have expectation zero and constant variance. \( s \) is the function to convert the distribution of \( x \) into another domain to reach the minimum difference between estimated gas \( \hat{y} \) and true gas value \( y \). The form of \( s \) was shown in following equations:

\[ s(x) = E(g^{-1}(Y)|X = x) \quad (6.2) \]
if the \( g(.) \) is known s is the conditional expectation of \( g^{-1}(Y) \) when \( X = x \). The expectation could be written as:

\[
E(g^{-1}(Y) | X = x) = \int_{-\infty}^{\infty} g^{-1}(y)f_{g^{-1}(y)|X=x}(g^{-1}(y))dg^{-1}(y)
\]  \hspace{1cm} (6.3)

where \( f_{g^{-1}(y)|X=x} \) is the conditional probability density function of \( g^{-1}(y) \) given \( X = x \), could be estimated by kernel density function as shown:

\[
\hat{f}_h(x) = \frac{1}{nh} \sum_{i=1}^{n} K(\frac{x-x_i}{h})
\]  \hspace{1cm} (6.4)

where \( K \) is the non-negative kernel function and \( h \) is the smoothing parameter called bandwidth, \( n \) is the number of data sets, and \( x \) represents the column vector of \( X \) matrix to be converted.

The \( g(.) \) is initially guessed as:

\[
\hat{g}(Y) = (Y - EY)/[\text{var}(Y)]^{1/2}
\]  \hspace{1cm} (6.5)

Once the \( \hat{s}(x) \) is updated iteratively until the squared error between \( \hat{g}(Y) \) and \( \hat{s}(x) \) reaches minimum, in other words the \( R^2 \) is not changed and is written as:
According to the above procedures, the stimulation design factors and water data were converted into different probability domains. The predicted value was generated under probability distribution after converting.

The correlation between gas and water were extracted by fixing the completion design parameters as expectation value identified from converted probability domain. The whole algorithm and procedures of this work is shown in Fig. 6-3.

Figure 6-3 Workflow of non-linear model to identify the correlation between gas and water
6.3 Results and Discussion

6.3.1 Gas and Flowback Water Relationship

The one-month fracture-fluid flowback and corresponding gas production for wells in the wet gas region (group A) was shown in Fig. 6-4. Each point on the figure represents one well. The x-axes represent the one-month fracture-fluid flowback after shut-in period in bbl or percentage. The percentage is calculated as one-month fracture-fluid flowback volume over total injected fracture-fluid volume. The y-axes are the corresponding one-month gas production in mmcf computed from the above analytical model. The red points represent one-month gas production from the model and the green points represent actual one-month gas production for each well. The relationship between one-month fracture-fluid flowback and gas production is positive. In other words, if more fracture-fluid flowback is recovered from wells in the wet-gas region, then the well is likely to produce more gas in the first month. On the contrary, the relationship between one-month fracture-fluid flowback and gas production is negative (Fig. 6-5) in dry gas regions (group B), where a lower recovery of fracture-fluid flowback may indicate the potential for higher short-term gas production.

To explain the relationships between one-month fracture-fluid flowback and gas production in wet and dry gas regions identified from the analytical model, a physical model was proposed based on the following studies on shale gas pore and wettability characteristics. Without information on hydraulic fracture and natural fracture geometry, it was assumed that the damage of fracture in both regions is similar and could be overcome by enough pressure drawdown during production. Therefore, it was assumed
that different flow patterns were the primary control factor behind the different gas-water relationships in Fig. 6-4 and 6-5.

The Marcellus shale was recognized to contain both organic and inorganic pores (Curtis et al. 2010). The organic pores could be further classified into kerogen hosted (sponge-like smaller pores) and mineral-associated complex pores (larger voids left by alternation of organic particle or result of diagenesis) (Milner et al. 2010; Milliken et al. 2012). In the wet gas region (%Ro around 1.0), the complex organic pores were observed more often than sponge-like organic pores (Milner et al. 2010). Until thermal maturity reached a certain cutoff value, the sponge-like organic pores tended to increase with thermal maturity (Milner et al. 2010). Initially the organic pores were considered as hydrocarbon-wet and inorganic pores were considered as water-wet (Passey et al. 2010). However, some laboratory studies showed evidence of water content in kerogen, and as thermal maturity increased from immature to over-mature, the detected moisture content increased from 0.5% to 15% (Chalmers and Bustin 2010; Ruppert et al. 2013), which indicated the wettability of organic pores may change from hydrocarbon-wet to neutral-wet. Both the organic and inorganic pores are connected and could provide effective pore networks as observed from 3-D reconstructions built by Focused Ion Beam Scanning Electron Microscopy (FIB-SEM) data (Passey et al. 2010; Goral et al. 2015). However, lab observations using tracer migration (Hu et al. 2015) showed a limited accessibility (several millimeters from sample edge) and connectivity of nanopore spaces in shale with spatial wettability. Therefore, it could be possible that the fracture-fluid was imbibed in an area of the fracture network with narrow diameters and matrix area close to that of the fracture network.
Figure 6-4 Correlation between one-month gas and fracture-fluid flowback in wet gas region.

Figure 6-5 Correlation of wells between one-month gas and fracture-fluid flowback in dry gas region.
**Figure 6-6** Proposed gas flow mechanism at pore size scale: The black region indicates organic matter, while the orange part represents gas-bearing pores inside organic matter. The dark yellow region represents the inorganic matter, consisting of siliciclastic grains of quartz, feldspar and micas. The light yellow points are gas-bearing pores. The light yellow points with blue circle are gas pores with bonded water. The blue points are pores filled with water. The red dashed line is the possible flow path for gas. No scale is indicated because this model may be applied across a wide range of scales. (Figure is modified based on Milliken’s SEM observations and Zhou’s observations from AAPG Lear Blog, http://www.aapg.org/publications/blogs/learn/article.)

**Fig. 6-6** showed the proposed flow processes around the fractures in shale based on previous literature. The green region indicates areas which underwent hydraulic fracturing, while the orange part represents the matrix area around hydraulic fracture. The black part represents the organic matter in matrix. The yellow circles are gas pores, blue circles are water pores, and the yellow circles with blue borders indicate gas pores with bonded water. The gas was originally generated from organic matter and stored in organic pores. During processes of thermal evolution and burial history, some of the gas may have been expelled to the inorganic matrix and stored in inorganic pores.
During shut-in periods in the wet gas region, the gas in inorganic pores may be expelled as a result of spontaneous water imbibition or swelling, as some of the fracture-fluid may be drawn into the inorganic matrix around fractures. However, gas in organic pores may not be expelled as those pores are hydrocarbon-wet. The fracture-fluid imbibed in inorganic matter may form the block for gas flow from organic pores to the fracture network. If the pressure drawdown was large enough to get rid of the ‘water block’, then the gas in organic pores around hydraulic fractures could flow successfully. Therefore, as more fracture-fluid is recovered, more gas may be recovered.

In the dry gas region, as thermal maturity increases, the wettability of organic pores may change from hydrocarbon-wet to neutral-wet. During the shut-in period, the fracture-fluid may not only be drawn into inorganic matter, but also organic pores, and help to expel the gas from organic pores to hydraulic fractures. Therefore, as less fracture-fluid is recovered and more water was drawn into formation as result of high capillary pressure, more gas will be expelled into the wellbore through hydraulic and natural fracture network.

6.3.2 Gas and Produced Water Relationship

The relationships between 1\textsuperscript{st}-year gas and water production, 2\textsuperscript{nd}-year gas and water production, and 3\textsuperscript{rd}-year gas and water production were analyzed to determine how the relationship between gas and water production changed over time. The results were shown in Fig. 6-7 to Fig. 6-9. The red points represent the calculated gas production from mathematic model of each well, while the green points are the actual gas production.
Generally speaking, the total recovered water production decreases from 1\textsuperscript{st}-year to 3\textsuperscript{rd}-year, similar to the yearly decline in gas production. The average 1\textsuperscript{st}-year water production is 9,701 bbl, ranging from 2,000 to 30,000 bbl; the average 2\textsuperscript{nd}-year water production is 1,936 bbl, and ranges from 500 to 5,500 bbl; and average 3\textsuperscript{rd}-year water production is 1,196 bbl, ranging from 300 to 2,500 bbl.

However, the relationship between water and gas production shifts in each 12 month period. In the 1\textsuperscript{st}-year (Fig. 6-7), the relationship between gas and water is negative, indicating more gas will be recovered with lower water recovery. In the 2\textsuperscript{nd}- and 3\textsuperscript{rd}-year (Fig. 6-8 and Fig. 6-9), the relationship changed to a positive trend, where increased gas production correlates to increased water production. An obvious positive trend in the real gas and water production data (green points) can be observed, especially for 2\textsuperscript{nd}-year production (Fig. 6-8). The results indicated that water recovery may be controlled by different flow mechanisms in the 1\textsuperscript{st}-year, 2\textsuperscript{nd}-year and 3\textsuperscript{rd}-year production in the dry gas region (northeastern) of the Marcellus.
Figure 6-7 Relationship for 1st-year gas production vs produced water

Figure 6-8 Relationship for 2nd-year gas production vs produced water
The flow mechanisms could be displacement or evaporation (flow through-drying), according to previous experimental and simulation studies (Allerton, Brownell, and Katz 1949; Kamath and Laroche 2003; Mahadevan and Sharma 2005; Mahadevan, Sharma and Yortsos 2007). The results show that the water is primarily carried out by displacement at the beginning. When drawdown pressure decreases and closes to capillary entry pressure, water may be carried out though evaporation primarily as result of gas expansion. The flow of either saturated or unsaturated gas in a water saturated rock could induce evaporation. This is a result of a closed volume system, where water capacity increases as pressure decreases. Most of the liquid was recovered rapidly by displacement and the remaining traces are removed more slowly by evaporation.
processes. The evaporation processes take place at the matrix surface and may require more time than displacement. The drying rate is quick initially and becomes slower later on. According to the simulation results (Seals 2015), the effect from gas desorption was only take place in after 5 years production and potentially account for 4% increase in cumulative gas recovered in Marcellus Shale after 27 years. The effect from gas slippage was not obvious in first three years and can potentially increase the cumulative gas recovered after 27 years by 11%. Therefore, for our case, which within the production periods in first three years, the effects from both desorption and slippage was not considered in our analysis.

In order to understand whether different relationships between gas and water identified in 1st- to 3rd-year were caused by different flow mechanisms as discussed above, the average WGR at standard conditions in each year was converted in unit as mole fraction to compare with the maximum water vapor content in gas phase (Yw in mole fraction) at reservoir conditions. If the WGR is larger than Yw, it is believed that water is primarily carried out as liquid phase in the form of immiscible displacement. If WGR is less than Yw, it is believed that water is primarily carried out as vapor phase in the form of evaporation.

The maximum water vapor in methane under specific pressure and temperature were referred from experimental data (Sage and Lacey 1979). According to Raoul’s law, the mole fraction of water vapor in the gas phase could be calculated as:

$$y_w = \frac{p_s}{p_g}$$  \hspace{1cm} (6.7)
where $P_s$ is saturated vapor pressure. The saturated vapor pressure of water at standard conditions were estimated though the Antoine equation. The Antoine equation is a class of semi-empirical relationships describing the relation between vapor pressure and temperature for pure components. It is written as:

$$\log_{10} P_s = A - \frac{B}{C+T} \quad (6.8)$$

where $T$ is temperature and $A$, $B$, $C$ are component specified constant in S.I. units.

**Figure 6-10** Changes in water-gas ratio across 1$^{st}$ (blue dots), 2$^{nd}$ (green dots), and 3$^{rd}$ year production (red dots). The cyan colored line represents the mean value. Each of the dots represent data from individual well.
The WGR at standard condition (14.7 psi, 60°F) was calculated by total yearly water production over gas production in unit bbl/mmcf and mole/mole (Fig. 6-10). The average WGR drops dramatically by comparing 1st- year WGR and 2nd- year WGR, and then begins to flatten during the 2nd- year to 3rd- year. The average WGR in the 1st- year is 0.07 mole/mole (9.1 bbl/mmcf), the 2nd- year average WGR is 0.02 mole/mole (3.2 bbl/mmcf), and the 3rd- year average WGR is 0.02 mole/mole (3.0 bbl/mmcf).

The colorful diamond points in Fig. 6-11 show the saturated water vapor in methane at different pressure and temperature conditions. Each of the wells discussed here are dry gas wells, which contain over 95% to 97% methane. The estimated saturated water vapor in methane in Fig. 6-11 was assumed to be close to the saturated water vapor in gas phase of our dry gas wells under reservoir conditions. The reservoir temperature is around 140 °C (Marcon et al. 2016); the initial reservoir pressure ranges from 4,588 to 5,959 psi by assuming average reservoir pressure gradient to be 0.60 psi/ft (Milici 2005; Wrightstone 2008) based on the vertical depth of wells, which ranges from 7,647 to 9,931 ft.

It is assumed that the pressure of Marcellus wells in the dry gas region declined by over 50% (Chaudhary and Lee 2016) in 1st-year. By comparing the two figures (Fig. 6-10 and Fig. 6-11), if the reservoir temperature of wells studied in this region is around 140 °C (413 K), and the average initial reservoir pressure is around 5,274 psi (364 bar), then the blue points in Fig. 6-11 were used to get the corresponding Yw. Assuming at the end of 1st-year production, the pressure declined 50% of the initial pressure at around 2,637 psi (182 bar), then the average pressure from the start to the end of the 1st-year is 3,955 psi (273 bar), the corresponding Yw is around 0.03, which is smaller than the
average 1\textsuperscript{st}–year WGR of 0.07. It is indicated that the water may be recovered in both liquid and vapor phase.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{water_vapor_methane.png}
\caption{Water vapor in methane under reservoir conditions. The different colored dots represent different temperature conditions. (Sage and Lacey 1979)}
\end{figure}

After 1\textsuperscript{st}-year production, it was estimated that the pressure declined at a rate of 10\% (Chaudhary and Lee 2016) in the 2\textsuperscript{nd} and 3\textsuperscript{rd} year, then at the end of 2\textsuperscript{nd} and 3\textsuperscript{rd} year, the formation pressure averages around 2,373 psi (164 bar) and 2,136 psi (147 bar).
The average formation pressure in the 2\textsuperscript{nd} and 3\textsuperscript{rd}-year is 2,505 psi (173 bar) and 2,255 (155 bar). The corresponding Y\textsubscript{w} of the 2\textsuperscript{nd} and 3\textsuperscript{rd}-year is 0.04 and 0.045, which is larger than the average WGR of 2\textsuperscript{nd} and 3\textsuperscript{rd}-year as 0.02. These results indicate that water may be recovered in the vapor phase only.

6.3 Section Conclusions

The relationship between gas and water production (or fracture-fluid flowback) in shale formations was investigated based on data from 180 wells using data mining methods. The major findings include:

- The well data was partitioned into different subgroups based on the hydrocarbon composition and water-gas ratio decline trend. Mathematic models using nonparametric estimation methods were developed to identify the relationships between gas and water in different spatial and temporal domains.

- In the wet gas region, higher initial recovery of fracture-fluid may result in higher first month gas production. The water in wet gas regions was drawn into the inorganic mineral portion around the hydraulic fracture network, which may cause a ‘micro-water block’ for gas flow from organic matter to wellbore. While in dry gas region, lower initial recovery of fracture-fluid may result in higher gas production in the first month. The wettability of organic pores may change from hydrocarbon wet to neutral wet from mature to over-mature. Water was not only
drawn into the inorganic mineral portion but also may be drawn into organic pores around fracture network area as a result of spontaneous imbibition by high capillary pressure, which may result in gas expulsion from organic pores into effective flow path to the wellbore to improve gas production.

- The relationship between gas and water in 1st-year production is negative, and then shifts to a positive relationship in the 2nd- and 3rd-year. By comparing the standard condition water-gas ratio (WGR) and maximum water vapor in gas phase (Yw) under the reservoir pressure and temperature in each year, the changes in the trend of gas and water relationships from the 1st- to 2nd-year may be the result of different flow mechanisms. In the 1st-year, the average WGR is larger than Yw, which may indicate water is primarily carried out by immiscible displacement in the liquid phase. However, in the 2nd- and 3rd-year, the average WGR is smaller than Yw, which may indicate water is predominantly carried out by evaporation in the vapor phase.
6.4 Section References


http://dx.doi.org/10.1016/j.apgeochem.2016.11.005 0883-2927/.


Presented at the SPE Western Regional Meeting, San Jose, California, 24-26 March. SPE-121354-MS. http://dx.doi.org/10.2118/121354-MS.


CHAPTER 7

Comprehensive Economic Analysis of Marcellus Shale in Different Time and Space Scenario by Data-Driven Method

Abstract

According to previous chapters, number of fracture stages and lateral length were found to be the key factors for increased gas production. However, optimal number of stage and lateral length (or fracture spacing) to achieve maximum gas production may not be the same as highest commerciality in different time and space scenarios. So it is important to understand economics of shale gas development at regional levels.

In this chapter, a regional economic analysis was conducted based on previous data analysis results (Chapter 4 - 6). The lateral length per stage with the optimal value range was considered to get the pseudo-maximum gas production of each well. Long-term gas prediction model for different geological regions were developed based on our multiphase and multi-mechanics simulation results using decline curve analysis and mathematic non-linear regression methods integration. Corresponding fracture-fluid flowback/produced water was also estimated for all wells in the dataset. After the prediction of the gas and water production for individual wells in different geological regions and production periods, the prediction was extended to a region level through spatial statistics through modified Gaussian Simulation. Net present values were then calculated at three different history prices and three different predicted gas prices. The payback period and breakeven price with optimal completion designs under different
geological conditions and production periods were provided based on the calculated net presented value.

The results in this Chapter 7 includes: single well gas and water production prediction curve, map-view of gas and water prediction, natural gas price prediction curve, map-view of net present value and suggested payback period and breakeven price.

7.1 Introduction

Economic analysis is important part for both asset evaluation and development optimization. Previous economic studies could be summarized into two major categories. First category of economic studies were integrating fracture geometry model, production prediction model and economic model, to obtain estimated ultimate recovery (EUR) and optimize completion parameters by net present value (Agrawal et al. 2012; Jabbari and Zeng 2012; Yalavarthi et al. 2013; Liu et al 2013; Jabbari and Benson 2013; Cook et al. 2014; Ross et al. 2014; Wei and Kamy 2013, 2014; Barree et al. 2015; Bhattacharya and Nikolaou 2016). The major difference between those economic studies is how to develop fracture and production prediction model. Net present value was calculated where cost was considered from drilling, completion and stimulation, and the revenue was calculated from EUR times gas prices. However, there is no study considered the cost from fracture-fluid flowback treatment, as the prediction was based on single phase models in previous works. Moreover, the optimized economical completion or stimulation parameters identified from previous works may not be appropriate for new wells if the associated geological conditions are significant different.
Second category of economic analysis included only production prediction model and economic model, but usually contains thousands of wells with production data (Browning et al. 2013, 2015; Gulen et al. 2015; Williger et al. 2016; Darugar et al. 2016). The analysis does not optimize completion and stimulation design parameters, but estimated ultimate recovery and asset value based on number of wells and drilling space. The estimated ultimate recovery was obtained by different modified Arps decline curve analysis. The spatial dependency was considered when estimated the regional ultimate recovery and net present value to find the optimal number of drilling and drilling space. In reality, shale gas wells may not produce continuously as predicted by decline curves because of technique or non-technique issues and unexpected low oil or gas prices. Therefore, it may more important to get an idea about the payback period of a well or a region, besides ultimate profit.

The objective of this study is to provide regional level NPV under different geological conditions and production periods through comprehensive economic analysis based on hundreds of Marcellus wells across Pennsylvania and West Virginia with optimized engineering parameters. Two datasets were used in this Chapter: one includes wells in Lycoming County in PA (Region-1), which includes 1 to 3 years gas and water production; another one contains wells in Upshur County in WV (Region-2), which only includes 1 to 3 years gas production. Both datasets include same type of engineering parameters as lateral length, number of stage, injected fracture-fluid volume and proppant mass, pump rate. Following is the summary of the processes of whole study.

The key engineering factors for shale gas production and fracture-fluid flowback were found in previous studies (Zhou et al. 2014, 2016), where the number of stage and
lateral length are most important factors for first-year gas production. The optimal lateral length per stage were defined which could result in maximum first-year gas production and were estimated by methods as non-linear model in Chapter 6. For wells in region-1, the optimal lateral length per stage is 290 ft/stage. The wells with lateral length ranges from 290 ft-5% to 290 ft +5% were selected as subset used to predict well and regional production. There are 48 wells out of 71 wells were selected. Wells in dry gas region of WV are sparely distributed and cross more than one county. The wells in Upshur County (region-2) and with lateral length within the range from optimal lateral length 350 ft to 450 ft (the optimal lateral length per stage is 400 ft/stage for wells in our dataset) were selected as subset used to get the close relationship and corresponding covariance matrix and to predict well and regional production. In total, 21 wells were selected to generate regional prediction. Using correlation between gas and water identified previously, corresponding water/fracture-fluid flowback could be determined. However, those values were limited into certain time period and single well range. To overcome the limitation, the prediction model for single well and region were developed to find maximum production and corresponding water recovery from optimal engineering design under different production time period.

Prediction model was developed based on comprehensive simulation results. The comprehensive multiphase and multidimensional simulation model included flow mechanisms such as non-Darcy flow, diffusion and so on. The different production declining trends were observed in two production periods and were used to develop mathematic model to predict single well gas production. The associated produced water or waste was calculated based on the non-linear model and correlation found in previous
chapter (Chapter 6) and used to calculate the cost for water treatment for final net present value. The regional gas production and water recovery were estimated through modified Gaussian Simulation model. The modified Gaussian Simulation model is not only counted well production but also including new well to make prediction.

Once the gas and water production of region was obtained, the net present value was calculated at regional level as total revenue minus total cost, including the cost for waste treatment, which could take 10% of overall drilling and completion costs (Kohshour et al. 2016). The cost was consisted of drilling & completion, water treatment, and operation, which drilling & completion costs were assumed to be fixed and affected from geological location and water treatment and operation will be varied depends on gas and water production (Ikonnikova et al. 2016). The revenue was calculated at three different proposed gas price increasing rate. Following Fig. 7-1 shows the flowchart of this chapter.

This study provides a comprehensive and optimal analysis about the field economic value evaluation with available well data, which could give reasonable guidelines for field and asset strategy in further development and adjustment. Not only could the results guide the field development in studied area, but the workflow and methods could help others working in this area.
Figure 7-1 Flowchart of one geological condition
7.2 Methods

7.2.1 Production Estimation

7.2.2.1 Single Well Production Estimation

It is complicated to model hydraulic fracturing treatment in shale gas reservoirs because of the nature of hydraulic fracture growth, a poor understanding of the treatment processes, and lack of good-quality reservoir information. Reservoir simulation is the preferred method to predict and evaluate well performance in shale gas reservoirs. Integrated decline curve and mathematic regression methods based on numerical simulations and field data were used for our performance predictions.

![Figure 7-2 Comprehensive numerical simulation layout and area (Seals 2015)](image)

The life-cycle Marcellus shale gas production results from a fully implicit, 3-dimensional, dual-porosity, 2-phase fluid transport and flowback simulation model with integrated flow mechanisms, including gas slippage, gas desorption, rock compaction, proppant crushing, and proppant diagenesis, were used for life-cycle Marcellus shale gas
production prediction. The simulated area was shown in Fig. 7-2. Because the model only simulated half of a reservoir, the result was doubled to the match the whole simulated area. The original cumulative gas production from the numerical simulation model and gas decline rate by subtracting previous cumulative production of one horizontal well were shown in Table 7-1. The second row of table does not show the gas yearly but the sum of difference between periods. For example, the value on fourth raw third column is the sum of second and third year gas production. The curves on Cartesian plots for cumulative production and decline rate were shown in Fig. 7-3 and Fig. 7-4.

<table>
<thead>
<tr>
<th>Year</th>
<th>1</th>
<th>3</th>
<th>5</th>
<th>10</th>
<th>15</th>
<th>20</th>
<th>25</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cum/mmcf</td>
<td>787.76</td>
<td>1111.84</td>
<td>1322.40</td>
<td>1735.02</td>
<td>2058.92</td>
<td>2331.96</td>
<td>2570.80</td>
</tr>
<tr>
<td>Time Period</td>
<td>1</td>
<td>2\textsuperscript{nd} &amp; 3\textsuperscript{rd}</td>
<td>4\textsuperscript{th} &amp; 5\textsuperscript{th}</td>
<td>6\textsuperscript{th}–10\textsuperscript{th}</td>
<td>11\textsuperscript{th}–15\textsuperscript{th}</td>
<td>16\textsuperscript{th}–20\textsuperscript{th}</td>
<td>21\textsuperscript{st}–25\textsuperscript{th}</td>
</tr>
<tr>
<td>Rate/mmcf</td>
<td>787.76</td>
<td>324.08</td>
<td>210.56</td>
<td>412.62</td>
<td>323.90</td>
<td>273.04</td>
<td>238.84</td>
</tr>
</tbody>
</table>

According to Fig. 7-4, different decline trend could be observed by comparing the decline rate of gas production before first five years and after five years. Left-hand is for the period from 1\textsuperscript{st} year to 5\textsuperscript{th} year, and right-hand is for the period from 6\textsuperscript{th} year to 25\textsuperscript{th} year. The left-hand curve, from 1\textsuperscript{st} to 5\textsuperscript{th} year, has a sharp decline from 787.76 mmcf to 210.56 mmcf. The right-hand curve, from 6\textsuperscript{th} to 25\textsuperscript{th} year, has a relative gently decline from 412.62 mmcf to 238.84 mmcf. The two decline curves could be fitted by two power equations. The power of right-hand curve is one-thirds of power of left-hand curve.
Figure 7-3 Cumulative gas production from numerical simulation model

Figure 7-4 Gas production rate before five years (left) and after five years (right)

Based on the two power equations, there are two assumptions were made for shale gas prediction estimation: First, all of the Marcellus shale gas wells in the database
follow the same decline trend; second, the power from second period is 1/3 of first period. The mathematic functions were shown as follow:

\[ q_{g \text{- first period}} = q_{i \text{- first period}} t^b \]  \hspace{1cm} (7.1)

\[ q_{g \text{- second period}} = q_{i \text{- second period}} t^{b/3} \]  \hspace{1cm} (7.2)

Above are two power regression functions, \( q \) is gas production rate in mmcf, \( q_i \) is the initial gas rate for the two periods, and \( t \) is time period. For the case of numerical simulation results, the values of \( q_i \) for two periods are 787.76 mmcf and 412.62 mmcf separately. The power \( b \) is estimated by nonlinear least square method. The objective is to minimize the following equation:

\[ Q = \sum_{i=1}^{m} [y_i - f(x_i, b)]^2 \]  \hspace{1cm} (7.3)

With respect \( m \) is size of dataset, \( y_i \) is real gas rate for \( i^{th} \) case and \( f(x_i, b) \) is the predict gas rate for \( i^{th} \) case. Above is also called as minimize sum of square error. The minimum value is achieved when the partial derivative of \( Q \) close to zero. In nonlinear system, the derivatives are function of both the variable and parameter, so the gradient equation does not have a closed solution. Instead, initial value of \( b \) was guessed and Gauss-Newton and Taylor Series expansion were used to iteratively find value of \( b \), as shown in following equations:
\[ f(x_i, b) \approx f(x_i, g^{(0)}) + \sum_{j=0}^{p-1} \left[ \frac{\partial f(x_i, b)}{\partial b_j} \right]_{b=g^{(0)}} (b_j - g^{(0)}) \]  

(7.4)

where \( g^{(0)} \) is the initial guess of \( b \) and \( p \) is number of \( x \) in function \( f \). Since we only have one \( x \), above function could be simplified as:

\[ f(x_i, b) \approx f(x_i, g^{(0)}) + \left[ \frac{\partial f(x_i, b)}{\partial b} \right]_{b=g^{(0)}} (b - g^{(0)}) \]  

(7.5)

move \( f(x_i, g^{(0)}) \) from right-hand side to left-hand side and simplified notation as:

\[ y_i^{(0)} = D_i^{(0)} \beta^{(0)} \]  

(7.6)

\[ y_i^{(0)} = f(x_i, b) - f(x_i, g^{(0)}) \]  

(7.7)

\[ D_i^{(0)} = \left[ \frac{\partial f(x_i, b)}{\partial b} \right]_{b=g^{(0)}} \]  

(7.8)

\[ \beta^{(0)} = (b - g^{(0)}) \]  

(7.9)

Where \( y_i^{(0)} \) is the residuals which are the deviations of observation around the nonlinear regression function with the parameters replaced by initial estimates. \( D_i^{(0)} \) is the partial derivatives of the mean response evaluated with the parameter replaced by the
starting estimates. $\beta^{(0)}$ is the difference between true regression parameter and the initial estimate of parameter. According to eq.7.6 to eq.7.9, the $\beta^{(0)}$ is estimated and used to adjust initial estimation of $b$, which is to update $g^{(0)}$ to $g^{(1)}$, until meet the following converge criteria:

$$\left| \frac{SSE^{(k)} - SSE^{(k+1)}}{SSE^{(k)}} \right| < 0.0001 \quad (7.10)$$

where SSE is the sum of square error.

The longest production rate of Marcellus wells in the dataset is three year and the shortest production rate is one month. The three-year cumulative gas production was used to compare with the result from comprehensive numerical simulation model. If the difference is less than 5%, then the future gas production were estimated by simulation results add the difference from three-year comparison. Otherwise, if the difference is more than 5%, then the future gas production was estimated by the equations list above. Following two figures Fig. 7-5 and Fig. 7-6 with one of case for difference less than 5% and difference large than 5%. The associated produced water production was estimated by the correlations found in Chapter 6.
Figure 7-5 Cumulative gas production for the case with difference < 5%

Figure 7-6 Cumulative gas production for the case with difference > 5%
7.2.2.2 Regional Production Estimation through Gaussian Simulation

In order to give the production and economic analysis in the region scale, an interpolation method needs to be chosen and applied to predict the new well production at new location in same region.

The interpolation method chosen in here is modified Gaussian Simulation, which could give optimal interpolation, based on regression against observed values of surrounding data points through the spatial covariance. The assumption is: there is spatial covariance or spatial correlation among the single well production. The spatial covariance is caused by geological conditions. If two wells are close with each other, it is believed that the two wells share similar geological conditions, such as porosity, permeability, thickness and etc. Otherwise, if two wells are far away with each other, they may have different geological conditions. The well production is controlled by geological conditions and engineering parameters and the engineering parameters are decided based on geological conditions. In other words, the assumption could be translated as: wells close to each other have similar production but different when apart.

The method used here assumed that the mean is known in local neighborhood of each estimation point, written as:

\[
Z^*(u) = m(u) + \sum_{\alpha=1}^{n(u)} \lambda_{\alpha}(u) [Z(u_{\alpha}) - m(u)]
\]

\[
= \sum_{\alpha=1}^{n(u)} \lambda_{\alpha}(u) Z(u_{\alpha}) + \left[1 - \sum_{\alpha=1}^{n(u)} \lambda_{\alpha}(u)\right] m(u)
\]

(7.11)
Where $u$ is the location vector of estimation point and $u_\alpha$ is one of neighborhood used for estimate of $Z^*(u)$, $n(u)$ is the number of data points in local neighborhood used for estimation, $m(u)$ is the expected value of $Z^*(u)$, and $\lambda_\alpha$ is the weight assigned to datum $Z(u_\alpha)$, same datum will receive different weight for different estimation location, and is calculated from semi-variogram model.

Variogram is a statistic method to characterize spatial correlation and has two types: experimental variogram and variogram model. The experimental variogram is used to characterize the spatial correlation on the observed data, and is expressed as:

$$\gamma(h) = \frac{1}{2N(h)} \sum_{\alpha=1}^{N(h)} [Z(u_\alpha + h) - Z(u_\alpha)]^2$$  \hspace{1cm} (7.12)

Where $h$ is the lag vector, representing separation between two spatial locations, $N(h)$ is the number of lag vector, and $Z(u_\alpha + h)$ is the lagged version of variable under consideration. The lag vector has numerical value and direction, which could also reflect the anisotropic characteristics of reservoir. The number of pairs in dataset was found by:

$$C_n^2 = \frac{n!}{2!(n-2)!}$$  \hspace{1cm} (7.13)

Where $n$ is the number of data points. The lag array was estimated in 10 categories with largest value is one of third of maximum distances. Within each category, the mean value of $\gamma(h)$ was used to represent the experimental variogram within specified distance.
The variogram model is the mathematic equations to simulate the results of experimental variogram, and the results from variogram model are used to as $\lambda_a$ in estimation. There are nine typical types variogram model used in this chapter, which show in following equations:

Nugget: $\gamma(h) = \begin{cases} 0, & \text{if } h = 0 \\ c, & \text{if } h > 0 \end{cases}$ \hspace{1cm} (7.14)

Spherical: $\gamma(h) = \begin{cases} c(1.5 \left( \frac{h}{a} \right) - 0.5 \left( \frac{h}{a} \right)^3), & \text{if } 0 \leq h \leq a \\ c, & \text{if } h > a \end{cases}$ \hspace{1cm} (7.15)

Exponential: $\gamma(h) = c \left( 1 - exp \left( \frac{-3h}{a} \right) \right)$ \hspace{1cm} (7.16)

Gaussian: $\gamma(h) = c \left( 1 - exp \left( \frac{-3h^2}{a^2} \right) \right)$ \hspace{1cm} (7.17)

Power: $\gamma(h) = ch^\omega$ \hspace{0.5cm} with $0 < \omega < 2$ \hspace{1cm} (7.18)

Circular: $\gamma(h) = \begin{cases} c \left( \frac{2h}{\pi a} \sqrt{1 - \left( \frac{h}{a} \right)^2} + \frac{2}{\pi} arcsin \left( \frac{h}{a} \right) \right), & \text{if } 0 \leq h \leq a \\ c, & \text{if } h > a \end{cases}$ \hspace{1cm} (7.19)
Pentaspherical: \(\gamma(h) = \begin{cases} 
\left(c \frac{15h}{8a} - \frac{5}{4} \left(\frac{h}{a}\right)^3 + \frac{3}{8} \left(\frac{h}{a}\right)^5\right), & \text{if } 0 \leq h \leq a \\
c, & \text{if } h > a 
\end{cases} \) \tag{7.20}

Logarithmic: \(\gamma(h) = \begin{cases} 
0, & \text{if } h = 0 \\
c \left(\log(h + a)\right), & \text{if } h > 0 
\end{cases} \) \tag{7.21}

Where \(c(\text{sill})\) is the value as semivariogram leveling off, and \(a(\text{range})\) is the corresponding lag value. The initial sill is chosen as global variance calculated from observed data points and initial range is corresponding lag value. The range is updated through weighted nonlinear least squared method, which is to minimize:

\[
\min \sum_{j=1}^{p} w_j \left(\gamma(h) - \hat{\gamma}(h)\right)^2
\] \tag{7.22}

\[w_j = \frac{N_j}{h_j^2}\] \tag{7.23}

Where \(\gamma(h)\) is from experimental model, \(\hat{\gamma}(h)\) is parametric model from updated range and sill, \(w_j\) is the weighting factor calculated from specific \(j^{th}\) lag \(h_j\) and number of data points \(N_j\) under that lag. The nugget effect was considered when average experimental semivariogram value not equal to zero and lag distance smaller than 5m.

The variogram model could be a model listed above or combination of more than one model. To find the optimal model which has the best match with experimental model
value, following criteria is used. The $\gamma_e$ from experiment variogram and $\gamma_m$ from variogram model were compared through following equations:

$$\varepsilon = \sum_{\alpha=1}^{N(h)} (\gamma_e^{\alpha} - \gamma_m^{\alpha})^2$$

(7.24)

Where $\gamma_e^{\alpha}$ is the variogram value from experimental data for $\alpha$ lag and $\gamma_m^{\alpha}$ is the one from variogram model. The model was picked to reach minimum sum of squared error $\varepsilon$.

Once the best variogram was chosen, the spatial estimation was made based on the covariance matrix generated from variogram model and the weighting coefficients related with the spatial distance between pair points. The domain was divided into $m$ by $m$ gridded blocks, which was defined by hydraulic fracture length from fracture propagation model (Chong et al. 2016) based on average treatment size and injection rate. In each gridded block, the value was estimated by following equation:

$$Z^*(u) = \frac{1}{n(u)} \sum_{\alpha=1}^{n(u)} Z(u_{\alpha}) + \sum_{\alpha=1}^{n(u)} \lambda_{\alpha}^{*}(u) \left[ Z(u_{\alpha}) - \frac{1}{n(u)} \sum_{\alpha=1}^{n(u)} Z(u_{\alpha}) \right]$$

(7.25)

where $Z^*(u)$ is the value to be estimated at location $u$, $Z(u_{\alpha})$ are values at multiple locations around $u$, represented as $u_{\alpha}$, total $n$ points. The $n$ points were selected based on the distances from the estimated location $u$ smaller than range value (a), since once the distance large than range (a), there is no correlation between points. $\lambda_{\alpha}^{*}(u)$ is the
weighting coefficients calculated by inverse of covariance matrix between know points times covariance vector between know points and estimated point, shown as:

\[
\lambda^* = \begin{bmatrix}
\gamma(u_1 - u_1) & \gamma(u_1 - u_2) & \cdots & \gamma(u_1 - u_{j-1}) & \gamma(u_1 - u_j) \\
\gamma(u_2 - u_1) & \gamma(u_2 - u_2) & \cdots & \gamma(u_2 - u_{j-1}) & \gamma(u_2 - u_j) \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
\gamma(u_{i-1} - u_1) & \gamma(u_{i-1} - u_2) & \cdots & \gamma(u_{i-1} - u_{j-1}) & \gamma(u_{i-1} - u_j) \\
\gamma(u_i - u_1) & \gamma(u_i - u_2) & \cdots & \gamma(u_i - u_{j-1}) & \gamma(u_i - u_j)
\end{bmatrix}^{-1} \begin{bmatrix}
\gamma(u_1 - u_e) \\
\gamma(u_2 - u_e) \\
\vdots \\
\gamma(u_{n-1} - u_e) \\
\gamma(u_n - u_e)
\end{bmatrix}
\]

(7.26)

When one point was estimated by above equations, it was assumed followed normal distribution with mean and variance calculated from above equations and the value was random picked from the generated distribution. The new point estimated was added in every loop and iteration to calculate the value on next location. The estimation processes was summarized in following figure.
This method was better compared with traditional Gaussian Simulation for our case as limited data points available. Table 7-2 shows the prediction accuracy from traditional Gaussian Simulation (GS) and our modified model. The wells in region-1 (Lycoming County) were used to do the comparison study. The wells were divided into two subsets with one contains 40 wells within spatial range both easting and northing smaller than 4000 m, while another one contains 8 wells in range from over 4000 easting which were used as test wells. The spatial location of two subsets was shown in Fig. 7-8.

Figure 7-8 Well location of two subsets, blue points are wells used to simulate and red points are wells used to test
By comparing 8 test wells, the error from modified model is generally smaller than the error from traditional model. The average error from modified model is 14% and the average error from traditional model is 18%. The modified model could improve 4% in interpolation accuracy in our case.
7.2.2 Economic Estimation

The commerciality of a project depends on oil and gas prices, reserves per well and average well cost. Gas price is hard to predict compared with cost and production. Therefore, in this part, the gas price was assumed to increase at three different rates and cost was assumed fixed.

The Herry Hub Natural Gas price was used for price predictions as shown in Fig. 7-7. The red points are the real recorded Herry Hub Natural Gas price from U.S. Energy Information Administration (EIA) from Jan 1997 to Sep 2016. The blue points are the future gas price estimated at 10% increase each year. The orange points are the future gas price estimated at 5% increase each year. The green points are the future gas price estimated as 3% increase each year. If the increase rate for each year is 10%, to 2040, the gas price is estimated as 23.1 $/MMBtu; If the increase rate for each year is 5%, to 2040, the gas price is estimated as 7.6 $/MMBtu; If the increase rate for each year is 3%, to 2040, the Herry Hub Price is estimated as 4.8 $/MMBtu.

The highest monthly gas price in the past is 13.42 $/MMBtu, minimum gas price is 1.72 $/MMBtu, and the average is 4.5 $/MMBtu. These three history gas prices were used as well to calculate NPV in different region.

The major costs, such as drilling and completion, were referred to technical report (Trends in U.S. Oil and Natural Gas Upstream Costs, EIA 2016). The whole Marcellus area were divided into five regions, as super core shown in yellow points, northeast core shown in light blue points, southwest core shown in red points, liquid rich shown in dark blue points, and periphery shown in green points, based on high performance variation...
and depths. The average drilling and completion cost were estimated on those five sub-plays.

**Figure 7-9** Estimated Herry Hub Natural Gas Price to 2040

**Figure 7-10** Five sub-plays distribution on the map (EIA, 2016)
The average drilling and completion cost for five sub-plays in Marcellus were referred from EIA report (EIA, 2016). For Northeast core area, the average drilling cost is 1.9 MM$, completion cost is 2.9MM$, and facilities cost is 0.2 MM$. For Periphery area, the average drilling cost is 2.0 MM$, the completion cost is 4.4 MM$, and facilities cost is 0.2 MM$. The variation is from different completion and drilling design in each sub-play, which is majorly controlled by geological condition in each sub-play. For example, the proppant used in Northeast core, a highly prolific area, is about 50% more than other areas, where the stress and rock brittleness is higher than other sub-plays.

Except drilling and completion, other cost, such as operation and flowback fluid disposal were considered as well in this research, and estimated based on industry experiences. The operation cost was considered as 0.5 $/MMcf, and the water treatment cost was as 6 $/bbl. The average discount rate (IRR) was assumed to be 15%. The average net present value (NPV) was calculated as following equation:

\[
NPV_j = \sum_{t=0}^{N}(Price_g q_g - C_{oper} q_g - C_{treat} q_w)/(1 + IRR)^t - C_{dril&comp}
\]  

(7.27)
7.3 Model Validation

7.3.1 Gas Production Validation

To validate the gas production prediction model, 9 out of 134 active wells in Region-1 with full 5 years production data (‘full’ means produced over 1825 days) and 22 out of 53 active wells in Region-2 with full 5 years production data were picked. The first three years productions were used in gas production prediction model to generate the sum of 5th and 4th year production which is used to compare with recorded value. The percentage of error was calculated as the difference between recorded value and model value over recorded value. The results of comparison were listed in Table 7-3 and Table 7-4.

\[
\text{diff\%} = \frac{\text{abs}(G_{\text{real}} - G_{\text{model}})}{G_{\text{real}}} \times 100\% 
\]

Table 7-3 Validation results of sum 4th and 5th gas rate in region-1

<table>
<thead>
<tr>
<th>Well</th>
<th>Recorded (mmcf)</th>
<th>Predicted (mmcf)</th>
<th>Difference (mmcf)</th>
<th>Difference (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>L-1</td>
<td>1028</td>
<td>1145</td>
<td>-117</td>
<td>11</td>
</tr>
<tr>
<td>L-2</td>
<td>400</td>
<td>423</td>
<td>-23</td>
<td>6</td>
</tr>
<tr>
<td>L-3</td>
<td>537</td>
<td>571</td>
<td>-34</td>
<td>6</td>
</tr>
<tr>
<td>L-4</td>
<td>408</td>
<td>423</td>
<td>-15</td>
<td>4</td>
</tr>
<tr>
<td>L-5</td>
<td>402</td>
<td>426</td>
<td>-24</td>
<td>6</td>
</tr>
<tr>
<td>L-6</td>
<td>1034</td>
<td>1081</td>
<td>-47</td>
<td>5</td>
</tr>
<tr>
<td>L-7</td>
<td>611</td>
<td>629</td>
<td>-18</td>
<td>3</td>
</tr>
<tr>
<td>L-8</td>
<td>342</td>
<td>353</td>
<td>-11</td>
<td>3</td>
</tr>
<tr>
<td>L-9</td>
<td>324</td>
<td>349</td>
<td>-25</td>
<td>8</td>
</tr>
</tbody>
</table>
Table 7-3 shows the maximum error is 11%, minimum error is 3%, the average error is 6%. Table 7-4 shows the maximum error is 22%, minimum error is 0%, the average error is 11%. There are two wells with error larger than 20%. By comparing the proppant concentration of two wells with those with error below 10%, the proppant concentration is either lower than one as 0.6 lbm/gal of U-12 or larger than one as 1.5 lbm/gal of U-18.

Table 7-4 Validation results of sum 4th and 5th gas rate in region-2

<table>
<thead>
<tr>
<th>Well</th>
<th>Recorded (mmcf)</th>
<th>Predicted (mmcf)</th>
<th>Difference (mmcf)</th>
<th>Difference (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>U-1</td>
<td>387</td>
<td>449</td>
<td>-62</td>
<td>16</td>
</tr>
<tr>
<td>U-2</td>
<td>465</td>
<td>547</td>
<td>-82</td>
<td>18</td>
</tr>
<tr>
<td>U-3</td>
<td>611</td>
<td>686</td>
<td>-75</td>
<td>12</td>
</tr>
<tr>
<td>U-4</td>
<td>245</td>
<td>259</td>
<td>-14</td>
<td>6</td>
</tr>
<tr>
<td>U-5</td>
<td>376</td>
<td>366</td>
<td>10</td>
<td>3</td>
</tr>
<tr>
<td>U-6</td>
<td>262</td>
<td>273</td>
<td>-11</td>
<td>4</td>
</tr>
<tr>
<td>U-7</td>
<td>370</td>
<td>350</td>
<td>20</td>
<td>5</td>
</tr>
<tr>
<td>U-8</td>
<td>553</td>
<td>480</td>
<td>73</td>
<td>13</td>
</tr>
<tr>
<td>U-9</td>
<td>553</td>
<td>480</td>
<td>73</td>
<td>13</td>
</tr>
<tr>
<td>U-10</td>
<td>506</td>
<td>530</td>
<td>-24</td>
<td>5</td>
</tr>
<tr>
<td>U-11</td>
<td>499</td>
<td>560</td>
<td>-61</td>
<td>12</td>
</tr>
<tr>
<td>U-12</td>
<td>346</td>
<td>420</td>
<td>-74</td>
<td>22</td>
</tr>
<tr>
<td>U-13</td>
<td>371</td>
<td>426</td>
<td>-55</td>
<td>15</td>
</tr>
<tr>
<td>U-14</td>
<td>422</td>
<td>423</td>
<td>-1</td>
<td>0</td>
</tr>
<tr>
<td>U-15</td>
<td>483</td>
<td>504</td>
<td>-21</td>
<td>4</td>
</tr>
<tr>
<td>U-16</td>
<td>434</td>
<td>507</td>
<td>-73</td>
<td>17</td>
</tr>
<tr>
<td>U-17</td>
<td>438</td>
<td>474</td>
<td>-36</td>
<td>8</td>
</tr>
<tr>
<td>U-18</td>
<td>383</td>
<td>464</td>
<td>-81</td>
<td>21</td>
</tr>
<tr>
<td>U-19</td>
<td>407</td>
<td>485</td>
<td>-78</td>
<td>19</td>
</tr>
<tr>
<td>U-20</td>
<td>488</td>
<td>542</td>
<td>-54</td>
<td>11</td>
</tr>
<tr>
<td>U-21</td>
<td>430</td>
<td>477</td>
<td>-47</td>
<td>11</td>
</tr>
<tr>
<td>U-22</td>
<td>517</td>
<td>581</td>
<td>-64</td>
<td>12</td>
</tr>
</tbody>
</table>
7.3.2 Water Production Validation

The wells in region-2 only include gas production. The model developed in region-1 was modified to estimated single well water production through gas production and engineering parameters. The model was validated by compare the predicted water production and recorded water production from 1\textsuperscript{st} to 3\textsuperscript{rd} year for the original 36 wells in Chapter 6, which does not include the 9 wells in region-1 as mentioned above. The results were shown in following Table 7-5 to Table 7-7.

### Table 7-5 Validation results of 1\textsuperscript{st} year water production in region-1

<table>
<thead>
<tr>
<th>Well</th>
<th>Recorded (bbl)</th>
<th>Predicted (bbl)</th>
<th>Difference (bbl)</th>
<th>Difference (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>L-1</td>
<td>8174</td>
<td>9128</td>
<td>-954</td>
<td>12%</td>
</tr>
<tr>
<td>L-2</td>
<td>12156</td>
<td>12420</td>
<td>-264</td>
<td>2%</td>
</tr>
<tr>
<td>L-3</td>
<td>6598</td>
<td>7327</td>
<td>-729</td>
<td>11%</td>
</tr>
<tr>
<td>L-4</td>
<td>8143</td>
<td>8169</td>
<td>-26</td>
<td>0%</td>
</tr>
<tr>
<td>L-5</td>
<td>7825</td>
<td>9249</td>
<td>-1424</td>
<td>18%</td>
</tr>
<tr>
<td>L-6</td>
<td>8226</td>
<td>6765</td>
<td>1461</td>
<td>18%</td>
</tr>
<tr>
<td>L-7</td>
<td>28896</td>
<td>24789</td>
<td>4107</td>
<td>14%</td>
</tr>
<tr>
<td>L-8</td>
<td>21890</td>
<td>15644</td>
<td>6246</td>
<td>29%</td>
</tr>
<tr>
<td>L-9</td>
<td>21356</td>
<td>21413</td>
<td>-57</td>
<td>0%</td>
</tr>
<tr>
<td>L-10</td>
<td>9250</td>
<td>10964</td>
<td>-1714</td>
<td>19%</td>
</tr>
<tr>
<td>L-11</td>
<td>12407</td>
<td>14460</td>
<td>-2053</td>
<td>17%</td>
</tr>
<tr>
<td>L-12</td>
<td>8543</td>
<td>9532</td>
<td>-989</td>
<td>12%</td>
</tr>
<tr>
<td>L-13</td>
<td>7823</td>
<td>8297</td>
<td>-474</td>
<td>6%</td>
</tr>
<tr>
<td>L-14</td>
<td>9074</td>
<td>9277</td>
<td>-203</td>
<td>2%</td>
</tr>
<tr>
<td>L-15</td>
<td>11257</td>
<td>10922</td>
<td>335</td>
<td>3%</td>
</tr>
<tr>
<td>L-16</td>
<td>16554</td>
<td>16802</td>
<td>-248</td>
<td>1%</td>
</tr>
<tr>
<td>L-17</td>
<td>8307</td>
<td>8120</td>
<td>187</td>
<td>2%</td>
</tr>
<tr>
<td>L-18</td>
<td>11466</td>
<td>8926</td>
<td>2540</td>
<td>22%</td>
</tr>
<tr>
<td>L-19</td>
<td>7461</td>
<td>8182</td>
<td>-721</td>
<td>10%</td>
</tr>
<tr>
<td>L-20</td>
<td>15678</td>
<td>15039</td>
<td>639</td>
<td>4%</td>
</tr>
<tr>
<td>L-21</td>
<td>6826</td>
<td>8158</td>
<td>-1332</td>
<td>20%</td>
</tr>
<tr>
<td>L-22</td>
<td>9569</td>
<td>9785</td>
<td>-216</td>
<td>2%</td>
</tr>
<tr>
<td>L-23</td>
<td>10495</td>
<td>10835</td>
<td>-340</td>
<td>3%</td>
</tr>
<tr>
<td>L-24</td>
<td>12529</td>
<td>9879</td>
<td>2650</td>
<td>21%</td>
</tr>
<tr>
<td>L-25</td>
<td>17339</td>
<td>11032</td>
<td>6307</td>
<td>36%</td>
</tr>
</tbody>
</table>
Table 7-6 Validation results of 2nd year water production in region-1

<table>
<thead>
<tr>
<th>Well</th>
<th>Recorded (bbl)</th>
<th>Predicted (bbl)</th>
<th>Difference (bbl)</th>
<th>Difference (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>L-1</td>
<td>2411</td>
<td>2422</td>
<td>-11</td>
<td>0%</td>
</tr>
<tr>
<td>L-2</td>
<td>3031</td>
<td>2531</td>
<td>500</td>
<td>16%</td>
</tr>
<tr>
<td>L-3</td>
<td>1702</td>
<td>2019</td>
<td>-317</td>
<td>19%</td>
</tr>
<tr>
<td>L-4</td>
<td>1669</td>
<td>1928</td>
<td>-259</td>
<td>16%</td>
</tr>
<tr>
<td>L-5</td>
<td>1723</td>
<td>1996</td>
<td>-273</td>
<td>16%</td>
</tr>
<tr>
<td>L-6</td>
<td>2124</td>
<td>2018</td>
<td>106</td>
<td>5%</td>
</tr>
<tr>
<td>L-7</td>
<td>5175</td>
<td>4167</td>
<td>1008</td>
<td>19%</td>
</tr>
<tr>
<td>L-8</td>
<td>3252</td>
<td>2600</td>
<td>652</td>
<td>20%</td>
</tr>
<tr>
<td>L-9</td>
<td>4046</td>
<td>3400</td>
<td>646</td>
<td>16%</td>
</tr>
<tr>
<td>L-10</td>
<td>2768</td>
<td>3128</td>
<td>-360</td>
<td>13%</td>
</tr>
<tr>
<td>L-11</td>
<td>2469</td>
<td>2799</td>
<td>-330</td>
<td>13%</td>
</tr>
<tr>
<td>L-12</td>
<td>1946</td>
<td>2211</td>
<td>-265</td>
<td>14%</td>
</tr>
<tr>
<td>L-13</td>
<td>1889</td>
<td>2148</td>
<td>-259</td>
<td>14%</td>
</tr>
<tr>
<td>L-14</td>
<td>2200</td>
<td>2363</td>
<td>-163</td>
<td>7%</td>
</tr>
<tr>
<td>L-15</td>
<td>2320</td>
<td>1905</td>
<td>415</td>
<td>18%</td>
</tr>
<tr>
<td>L-16</td>
<td>2793</td>
<td>2770</td>
<td>23</td>
<td>1%</td>
</tr>
<tr>
<td>L-17</td>
<td>2097</td>
<td>2123</td>
<td>-26</td>
<td>1%</td>
</tr>
<tr>
<td>L-18</td>
<td>1799</td>
<td>764</td>
<td>1035</td>
<td>58%</td>
</tr>
<tr>
<td>L-19</td>
<td>1435</td>
<td>1562</td>
<td>-127</td>
<td>9%</td>
</tr>
<tr>
<td>L-20</td>
<td>4361</td>
<td>4339</td>
<td>22</td>
<td>0%</td>
</tr>
<tr>
<td>L-21</td>
<td>1404</td>
<td>1834</td>
<td>-430</td>
<td>31%</td>
</tr>
<tr>
<td>L-22</td>
<td>433</td>
<td>452</td>
<td>-19</td>
<td>4%</td>
</tr>
<tr>
<td>L-23</td>
<td>334</td>
<td>312</td>
<td>22</td>
<td>7%</td>
</tr>
<tr>
<td>L-24</td>
<td>2350</td>
<td>1697</td>
<td>653</td>
<td>28%</td>
</tr>
<tr>
<td>L-25</td>
<td>2816</td>
<td>2101</td>
<td>715</td>
<td>25%</td>
</tr>
<tr>
<td>L-26</td>
<td>2351</td>
<td>2090</td>
<td>261</td>
<td>11%</td>
</tr>
<tr>
<td>Well</td>
<td>Recorded (bbl)</td>
<td>Predicted (bbl)</td>
<td>Difference (bbl)</td>
<td>Difference (%)</td>
</tr>
<tr>
<td>------</td>
<td>----------------</td>
<td>----------------</td>
<td>------------------</td>
<td>----------------</td>
</tr>
<tr>
<td>L-1</td>
<td>1861</td>
<td>1775</td>
<td>86</td>
<td>5%</td>
</tr>
<tr>
<td>L-2</td>
<td>1741</td>
<td>1505</td>
<td>236</td>
<td>14%</td>
</tr>
<tr>
<td>L-3</td>
<td>1083</td>
<td>1073</td>
<td>10</td>
<td>1%</td>
</tr>
<tr>
<td>L-4</td>
<td>987</td>
<td>1028</td>
<td>41</td>
<td>4%</td>
</tr>
<tr>
<td>L-5</td>
<td>1088</td>
<td>1074</td>
<td>14</td>
<td>1%</td>
</tr>
<tr>
<td>L-6</td>
<td>1314</td>
<td>1087</td>
<td>227</td>
<td>17%</td>
</tr>
<tr>
<td>L-7</td>
<td>2540</td>
<td>1843</td>
<td>697</td>
<td>27%</td>
</tr>
<tr>
<td>L-8</td>
<td>1741</td>
<td>1526</td>
<td>215</td>
<td>12%</td>
</tr>
<tr>
<td>L-9</td>
<td>2223</td>
<td>1859</td>
<td>364</td>
<td>16%</td>
</tr>
<tr>
<td>L-10</td>
<td>1800</td>
<td>1951</td>
<td>-151</td>
<td>8%</td>
</tr>
<tr>
<td>L-11</td>
<td>1253</td>
<td>1426</td>
<td>-173</td>
<td>14%</td>
</tr>
<tr>
<td>L-12</td>
<td>1421</td>
<td>1526</td>
<td>-105</td>
<td>7%</td>
</tr>
<tr>
<td>L-13</td>
<td>1280</td>
<td>1485</td>
<td>-205</td>
<td>16%</td>
</tr>
<tr>
<td>L-14</td>
<td>1147</td>
<td>1526</td>
<td>-379</td>
<td>33%</td>
</tr>
<tr>
<td>L-15</td>
<td>1028</td>
<td>1073</td>
<td>-45</td>
<td>4%</td>
</tr>
<tr>
<td>L-16</td>
<td>1028</td>
<td>1218</td>
<td>-190</td>
<td>18%</td>
</tr>
<tr>
<td>L-17</td>
<td>1028</td>
<td>1028</td>
<td>0</td>
<td>0%</td>
</tr>
<tr>
<td>L-18</td>
<td>1028</td>
<td>688</td>
<td>340</td>
<td>33%</td>
</tr>
<tr>
<td>L-19</td>
<td>1028</td>
<td>1028</td>
<td>0</td>
<td>0%</td>
</tr>
<tr>
<td>L-20</td>
<td>2509</td>
<td>2540</td>
<td>-31</td>
<td>1%</td>
</tr>
<tr>
<td>L-21</td>
<td>1073</td>
<td>1131</td>
<td>-58</td>
<td>5%</td>
</tr>
<tr>
<td>L-22</td>
<td>540</td>
<td>540</td>
<td>0</td>
<td>0%</td>
</tr>
<tr>
<td>L-23</td>
<td>540</td>
<td>688</td>
<td>-148</td>
<td>27%</td>
</tr>
<tr>
<td>L-24</td>
<td>1526</td>
<td>1272</td>
<td>254</td>
<td>17%</td>
</tr>
<tr>
<td>L-25</td>
<td>1526</td>
<td>1083</td>
<td>443</td>
<td>29%</td>
</tr>
<tr>
<td>L-26</td>
<td>1526</td>
<td>1294</td>
<td>232</td>
<td>15%</td>
</tr>
<tr>
<td>L-27</td>
<td>1526</td>
<td>1311</td>
<td>215</td>
<td>14%</td>
</tr>
<tr>
<td></td>
<td>L-28</td>
<td>L-29</td>
<td>L-30</td>
<td>L-31</td>
</tr>
<tr>
<td>-----</td>
<td>------</td>
<td>------</td>
<td>------</td>
<td>------</td>
</tr>
<tr>
<td>688</td>
<td>688</td>
<td>640</td>
<td>688</td>
<td>688</td>
</tr>
<tr>
<td>0</td>
<td>48</td>
<td>48</td>
<td>0</td>
<td>-340</td>
</tr>
<tr>
<td>0%</td>
<td>7%</td>
<td>0%</td>
<td>0%</td>
<td>49%</td>
</tr>
</tbody>
</table>

Above three validation results indicated around 80% of wells could have error less than 20%. The wells with error larger than 20% may cause by different engineering parameters and gas production but similar recorded water production.

7.4 Results and Discussion

7.4.1 Production Estimation of Single Observed Wells

Fig.7-11 and Fig.7-12 show the predicted cumulative gas production and cumulative water production for wells within production periods from 1, 3, 5, 10, 15, 20 to 25 years, which represent 1<sup>st</sup> year, 2<sup>nd</sup> plus 3<sup>rd</sup> year, 4<sup>th</sup> plus 5<sup>th</sup> year, 6<sup>th</sup> to 10<sup>th</sup> year, 11<sup>th</sup> to 15<sup>th</sup> year, 16<sup>th</sup> to 20<sup>th</sup> year, 21<sup>st</sup> to 25<sup>th</sup> year separately. Red scatter plots are gas production, and blue scatter plots are water production. The production value was represented through the circle size and color bar. For both gas and water, the lighter and small circle represents smaller production, but darker and large circle represent larger production. Since some wells were close with each other, the corresponding circles were overlapped or covered by other points. From 1<sup>st</sup> year to 3<sup>rd</sup> year, productions are actual field data and start from 5<sup>th</sup> year production is predicted results from model.
The different color bar not only shows the gas and water production for each well but also tell the percentage of wells within certain production range under different production periods. For gas production, the 1\textsuperscript{st}-year cumulative production ranges from 300 to 2000 mmcf. During the production period from 2\textsuperscript{nd} year to 5\textsuperscript{th} year, small part of wells located on northeastern and southwestern produced cumulative gas more than 3000 mmcf. Over 10 years production, some wells located on northeastern corner produced gas more than 4000 mmcf. At end of 15 years production period, half of the wells had gas production over 3000 mmcf. From 20\textsuperscript{th} to 25\textsuperscript{th} year, one third of wells had gas production over 4000 mmcf. The maximum cumulative gas production is around 7000 mmcf and minimum is around 1000 mmcf after 25 years.

For water production, the change of production was not obvious as well as gas production after 3 years. In 1\textsuperscript{st} year, few wells recovered fracture-fluid over $20 \times 10^3$ bbl, some even over $30 \times 10^3$ bbl. After 3 years cumulative production, the number of wells with water production over $30 \times 10^3$ bbl increased and few wells with water production over $40 \times 10^3$ bbl. From 5 to 25 years, the change of water production distribution is not obvious. In other words, the water majorly was recovered in first 3 years. The maximum cumulative water production after 25 years is around $50 \times 10^3$ bbl and minimum is around $10 \times 10^3$ bbl.
Figure 7-11 Life-cycle cumulative gas production for wells in Region-1
Figure 7-12 Life-cycle cumulative water production for wells in Region-1
Fig. 7-13 and Fig. 7-14 show the cumulative gas production of dry gas wells in region-2 at the end of 1, 3, 5, 10, 15, 20 and 25 years. The water-gas relationship of wells in Pennsylvania was applied to wells of West Virginia in dry gas region to predict the 1 to 25 years cumulative water production. Compared with Fig. 7-11 and Fig. 7-12, the area of region-2 is twice larger than region-1 but the production of region-2 is half of wells in region-1, which is also match the result of EIA report (2016) that one is in Northeastern core area with relative high production and one is in Periphery with relative low production. For gas production, initially the production is low and all below 1000 mmcf, which production increases gradually before 10 years. After 15 years, small part of wells had gas production over than 2000 mmcf but still low. After 25 years, half of wells had gas production over than 2000 mmcf. For water production, initially some wells with water production as high as 150 kbbl. The water production was not changed much after 3 years, and more than half of wells produced water more than 130 kbbl after 25 years. In summary, both gas and water production in region-2 is quite low compared with those in region-1.
Figure 7-13 Life-cycle cumulative gas production for dry gas wells in Region-2
Figure 7-14 Life-cycle cumulative water production for dry gas wells in Region-2
7.4.2 Production Estimation of Region

Regional prediction of gas and water production was implemented based on the subset wells in 7.3.1. According to average treatment size, which is 200 gal per stage in region-1 and 400 gal per stage in region-2, and fracture propagation model (Chong 2016) the well spacing was assumed to be 500 ft in region-1 and 1000 ft in region-2. The domain of region-1 was divided into 40 x 40 (a total of 1600) grid blocks and domain of region-2 was divided into 50 x 50 (a total of 2500) grid blocks. The estimation represents the production of the point located in the center of each block.

Fig.7-15 shows maps of cumulative gas production in region-1. The yellow part represents the region with high cumulative gas production, and the dark blue part represents the region with low cumulative gas production, and the green part is in medial range. The black filled points are actual well locations. According to the maps of gas production, operators could refine their develop plans and choose drilling locations, while other stakeholders could calculate the economic impacts.

Fig.7-16 shows the results of cumulative water production for same region. Compared with the cumulative gas production, the low cumulative gas region corresponds to the high cumulative water region, but high cumulative gas region may not mean the lowest water production. After 25 years, the highest gas production is over 6000 mmcf per well and the lowest gas production is lower than 3500 mmcf; the highest water production is larger than 350 kbbl per well and lowest water production is smaller than 200 kbbl per well.
Figure 7-15 Estimation from 1 to 25 years cumulative gas production in region-1
Figure 7.16 Estimation from 1 to 25 years cumulative water production in region-1
**Fig.7-17** and **Fig.7-18** show estimations from 1 to 25 years cumulative gas and water production of wells in West Virginia. Compared with previous wells in PA, the wells in this region have longer fracturing spacing and longer well spacing. The fracture network assumed to be larger than those in PA. However, the average gas and water production in this region are relative low compared with previous cases, which may be caused by unknown geo conditions and engineering effects. Since the number of wells in this region is relative small and wells distributed sparsely, both the gas and water production is not varied much as a result of similar color cover large region of map. For example, as 1st year gas production, most wells with gas production from 500 to 700 mmcf, only small part of wells located in the northeastern corner produced gas below 350 mmcf. Some trend was found in 3 years, 10 years, 15 years, and 20 years scenarios. In 5 years and 25 years scenarios, the gas production distribution is more concentrated in middle value.

Same as the relationship in PA case, higher gas production corresponded lower water production. Unlike gas production, the water production also not change after 3 years as compared the percentage of color bar. After 25 years, the northern part has relative higher water production while in southern part relative low.
Figure 7-17 Estimation from 1 to 25 years cumulative gas production in region-2
Figure 7-18 Estimation from 1 to 25 years cumulative water production in region-2
7.4.3 Net Present Value Estimated at Different Increasing Rate

Table 7.8 shows the assumed parameters for calculate the present value (PV) in each production periods (1-year, 3-years, 5-years, 10-years, 15-years, 20-years, 25-years) and net present value (NPV), which is the sum of PV of different geological region. Table 7.9 shows the assumed gas price for 1, 3, 5, 10, 15, 20, 25 years periods at different increasing rate.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Region 1</th>
<th>Region 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Well Fixed Cost</td>
<td>5.0 MM$</td>
<td>6.6 MM$</td>
</tr>
<tr>
<td>Operation Cost</td>
<td>0.5 $/MMCF</td>
<td>0.5 $/MMCF</td>
</tr>
<tr>
<td>Water Treatment Cost</td>
<td>6 $/BBL</td>
<td>6 $/BBL</td>
</tr>
<tr>
<td>IRR</td>
<td>15%</td>
<td>15%</td>
</tr>
</tbody>
</table>

Table 7.9 Gas price assumed to calculated net present value

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>3</th>
<th>5</th>
<th>10</th>
<th>15</th>
<th>20</th>
<th>25</th>
</tr>
</thead>
<tbody>
<tr>
<td>$@3%</td>
<td>2.3</td>
<td>2.5</td>
<td>2.7</td>
<td>3.0</td>
<td>3.4</td>
<td>4.2</td>
<td>4.7</td>
</tr>
<tr>
<td>$@5%</td>
<td>2.3</td>
<td>2.6</td>
<td>2.8</td>
<td>3.6</td>
<td>4.6</td>
<td>5.9</td>
<td>7.5</td>
</tr>
<tr>
<td>$@10%</td>
<td>2.3</td>
<td>2.8</td>
<td>3.4</td>
<td>5.5</td>
<td>8.9</td>
<td>14.3</td>
<td>23</td>
</tr>
</tbody>
</table>

Fig.7.19 to Fig. 7.21 shows the net present value at different gas price increasing rate and different producing periods for the PA region. For each grid block, it shares
same cost. So the variation in net present value only caused by production and natural gas price. Each group points at specific production period show the individual NPV value of each grid block, and the red line is where NPV value equal to zero. Since in real scenario, wells could be stopped continually producing for some reason and may not all the way produce as long as 25 years. The purpose of these three plots is show at when and what gas price conditions, the well could be payoff. **Fig. 7-19** is the case for gas price increasing at 3%. **Fig. 7-20** is the case for gas price increasing at 5%. **Fig. 7-21** is the case for gas price increasing at 10%. For all three assumed price conditions, the wells could not be payoff during first year. After three years production, small part of wells may get NPV over than zero, which means those wells may have larger production. If the gas price increasing at 10%, after 5 years, more than half of wells in this region have NPV over than zero, and after 25 years, almost whole region have NPV over than zero.

![Figure 7-19](image)

**Figure 7-19** Net present value in different producing period with price increasing rate as 3%
Figure 7-20 Net present value in different producing period with price increasing rate as 5%

Figure 7-21 Net present value in different producing period with price increasing rate as 10%
Fig. 7-22 to Fig. 7-24 show the NPV after 25 years for whole PA region at 3%, 5% and 10% gas price increasing rate. If the gas price increasing at 3% yearly, after 25 years, the maximum NPV, which located at the north-eastern corner of region, is around 3 MM$, and the minimum NPV, which located at the south-western corner of region, is around -1.5 MMS. If the gas price increasing at 5% yearly, after 25 years, the maximum NPV is around 3 MM$ and minimum is around -1 MMS. If the gas price increasing at 10%, after 25 years, the maximum NPV is around 4 MM$, and the minimum is around -0.5 MM$. It looks like there is not much difference between 3% case with 5% and 10%. Before 5 years, the gas price is similar at different increasing rate. After 5 years, even the price is different, but the production is quite small compared with previous production.

**Figure 7-22** Net present value in map review at price increasing with 3% yearly in region-1
Figure 7-23 Net present value in map review at price increasing with 5% yearly in region-1

Figure 7-24 Net present value in map review at price increasing with 10% yearly in region-1
**Fig.7-25** shows the net present value of wells after 25 years in West Virginia. The green surface represents the NPV at price increasing at 10%, the blue one is 5% and the red one is 3%. Since the gas production in this region is significantly low, and the completion and drilling cost is 1.5 MM$ than those in PA region, the NPV is below zero at three scenarios.

*Figure 7-25* Net present value in map review in Region-2
Table 7-10 show the net present value of wells after 25 years in region-1 and region-2 calculated by using the highest history gas price 13.42 $/MMBtu, lowest history gas price 1.72 $/MMBtu and average history gas price 4.5 $/MMBtu. Average gas and water cumulative production after 25 years in region-1 and region-2 was used. The NPV in region-1 is almost double as the NPV in region-2 by using the average history gas price.

Table 7-10 NPV in MM$ after 25 years by using history gas price

<table>
<thead>
<tr>
<th>Region</th>
<th>Highest Price</th>
<th>Lowest Price</th>
<th>Average Price</th>
</tr>
</thead>
<tbody>
<tr>
<td>Region-1</td>
<td>1.6 MM$</td>
<td>0.07 MM$</td>
<td>0.4 MM$</td>
</tr>
<tr>
<td>Region-2</td>
<td>0.9 MM$</td>
<td>-0.02 MM$</td>
<td>0.2 MM$</td>
</tr>
</tbody>
</table>
7.5 Section Conclusions

In this study, economic analysis was performed based on predicted production. Several findings could be summaries as following:

- The gas production could be described as two different declining trends before and after 5 years; first declining within 5 years is sharp while second declining after 5 years is slowly. The declining rate of second period is one of third of first period. The validation results of well production prediction model show that the average error is 6% for region-1 and 11% for region-2. The 80% of wells show error of water production below 20%.

- A modified Gaussian Simulation model was developed and validated by comparing the validation results with traditional Gaussian Simulation model. The spatial interpolation accuracy could be improved as 4% by our modified model.

- The average spatial production in region-1 with the optimal lateral length per stage is twice than region-2 after 25 years. The water spatial distribution in both regions is almost not changed after 3 years. The wells with larger gas production and lower water production located in northeastern corner of region-1 while the wells with larger gas production and lower water production located in southeastern corner of region-2.

- By calculated the regional NPV, some wells in region-1 may payoff after 3 years production, and almost all wells have NPV over zero after 25 years production if gas price increasing as 10% yearly. For wells in region-2, since the production is small and cost of drilling and completion is high, NPV is below zero.
7.6 Section References


CHAPTER 8

Summary

8.1 Summary

This study provides a detailed discussion about applying multiple data mining analysis approaches to identify hidden information behind the massive production data, which gives a new idea and workflow for analyzing production data efficiently over the assumptions for conventional experimental and numerical analysis.

8.1.1 Findings on Controlling Factors Affect Gas and Fracture-Fluid Flowback

- The number of hydraulic fracture stages is found to be the most significant parameter among all factors studied affecting one-year cumulative gas production in West Virginia.

- The shut-in time, proppant size, and proppant mass are found to be the significant parameters affecting the fracture-fluid flowback within 3-week after a fracture treatment.

8.1.2 Findings on Correlations between Gas and Fracture-Fluid Flowback/Produced Water

- In the wet gas region, higher initial recovery of fracture-fluid may result in higher first month gas production. While in the dry gas region, lower recovery of
fracture-fluid may result in higher first month gas production. It may be result of different wettability condition and flow mechanisms.

- The relationship between gas and water in 1\textsuperscript{st}-year production is negative, and then shifts to a positive relationship in the 2\textsuperscript{nd} and 3\textsuperscript{rd}-year. It may be result of different flow mechanisms as displacement or evaporation, and the proposed analysis was further validated by comparing the standard condition water-gas ratio (WGR) and maximum water vapor in gas phase (Yw) under the reservoir pressure and temperature in each year.

### 8.1.3 Findings on Economic Analysis

- The gas production could be successfully described as two different declining trends before and after 5 years. The declining rate of second period is one of third of first period. The spatial interpolation accuracy could be improved as 4% by our modified Gaussian Simulation model.

- The average spatial production in region-1 with the optimal lateral length per stage is twice than region-2 after 25 years. The water spatial distribution in both regions is almost not changed after 3 years. The wells with larger gas production and lower water production located in northeastern corner of region-1 while the wells with larger gas production and lower water production located in southeastern corner of region-2.

- By calculated the regional NPV, the commercial value of region-1, which may get payoff after 3 years production, is much higher than region-2. 8.2
VITA

Qiumei Zhou

Qiumei Zhou was born and grew up in Chengdu, China, the hometown of Panda and also the heart of western China. She finished her B.S in Petroleum Geology in Chengdu University of Technology. After graduating from Chengdu University of Technology, she attended Penn State University for Master and Ph.D. study in Petroleum and Natural Gas Engineering. She has been appointed as teaching assistant and research assistant during 4 years Ph.D. studies.