

© Copyright by Aaditya Khanal 2016
All Rights Reserved

**Numerical Study and Production Forecasting of Unconventional
Liquid-Rich Shale Reservoirs with Multi-Fractured Horizontal
Wells**

A Dissertation

Presented to

the Faculty of the Department of Chemical and Biomolecular Engineering

University of Houston

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

in Chemical Engineering

by

Aaditya Khanal

December 2016

**Numerical Study and Production Forecasting of Unconventional
Liquid-Rich Shale Reservoirs with Multi-Fractured Horizontal
Wells**

Aaditya Khanal

Approved:

Committee Members

Chair of the Committee,
Guan Qin, Associate Professor,
Chemical and Biomolecular Engineering

William John Lee, Professor,
Texas A&M

Michael Nikolaou, Professor,
Chemical and Biomolecular Engineering

Reinaldo Gonzalez, Adjunct Faculty,
Petroleum Engineering

Javad Tavakoli, Professor Emeritus,
Chemical and Biomolecular Engineering,
Lafayette College

Suresh K. Khator, Associate Dean,
Cullen College of Engineering

Michael P. Harold, Professor and Chair,
Chemical and Biomolecular Engineering

Acknowledgments

This work would not have been possible without the help and support of many people around me. First of all, I would like to express my sincere gratitude to my advisor, Dr. John Lee, for his continuous guidance, motivation, immense knowledge and support during my time here at UH. His ideas and encouragement helped me learn new things and allowed me to complete my dissertation.

I am grateful to Dr. Guan Qin for agreeing to be the chair of my dissertation committee during the difficult transitional period. I would also like to thank Dr. Michael Nikolaou for guiding me towards a new topic for my research and would like to thank Dr. Reinaldo Gonzalez for agreeing to serve in my committee on such a short notice. I would like to give special thanks to Dr. Javad Tavakoli, for constantly encouraging me, helping me with my edits and making a long trip just for my dissertation defense. I will be always grateful to him for his kindness and generosity. I'd like to thank Dr. Thomas Holley, Ms. Anne Sturm for helping with my research in any ways they could.

I would like to mention my deep appreciation to my wife Prativa, for her help and support during the research. A special thanks to my family for all their prayers, and emotional support for me. Finally, I would like to thank the Chemical and Petroleum Engineering Department, my fellow group mates, Mohammad Khoshghadam, Chen Yu, and others for their support during my time as a graduate student.

**Numerical Study and Production Forecasting of Unconventional
Liquid-Rich Shale Reservoirs with Multi-Fractured Horizontal
Wells**

An Abstract
of a
Dissertation

Presented to

the Faculty of the Department of Chemical, and Biomolecular Engineering
University of Houston

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy
in Chemical Engineering

by

Aaditya Khanal

December 2016

Abstract

Accurate production performance evaluation and forecasting in shales during the early stages of development can play an important role in minimizing uncertainties associated with unconventional reservoirs. The liquid-rich shale reservoirs, due to their complex rock and fluid behaviors, require different analysis than the conventional reservoirs.

In this work, we considered both reservoir simulation and analytical models to forecast the production from liquid rich shales. We modified the tri-linear flow model derived for single phase flow to use it for multiphase flow with some simplifying assumptions. We were able to validate the results obtained from the analytical model with errors of less than 10% when used to forecast liquid rich shale volatile oil reservoirs.

Additionally, reservoir simulation was used to identify the effect of several parameters on the estimated ultimate recovery (EUR) of gas condensate reservoirs. Fracture half-length, permeability, and fracture spacing was identified to be the most important parameters for maximizing the cumulative gas production. It was also seen that the interaction of different parameters with each or their combined effect was important in optimizing the final EUR for oil and gas. We also identified the effect of fluid composition on well-spacing in the Eagle Ford Shale. For critical fluids, liquid dropout and condensate banking had a huge impact on the final production.

It was seen that gas condensate wells in shales exhibit a long transition period between the end of linear flow and the start of boundary dominated flow. Pressure normalization was found to be an effective method to identify flow regimes in a gas condensate reservoir. Results also showed that transient linear flow model with no modification for boundary-dominated flow overestimates the production in almost all cases.

Finally, compositional reservoir model has been used to create several iterations of synthetic production histories from liquid rich shales (LRS) wells based on Monte Carlo simulation with predefined probability distributions. Cumulative gas, gas rate, and condensate-to-gas ratio (CGR) for the simulated cases were decomposed into principal component (PC) scores and coefficients. The dataset was cross-validated to check its ability to predict the missing production data based on PC scores and coefficients of the limited production data. Principal component analysis was further applied to the field data from several wells from Eagle Ford shale. Two to three PCs were required to recreate the initial data with reasonable accuracy depending on the quality of the input data. During the validation step, we observed that some of the wells exhibited significant error which could be attributed to significantly different production profiles of those wells compared to the other wells. For simulated data, four PCs were enough to yield the prediction with average errors of 0.16%, 0% and 0.77% respectively for gas rate, cumulative gas and CGR respectively. For field data, three PC yielded the best prediction with average error of 1.63% and 2.98% for gas rate and oil rate respectively. This shows that multivariate statistics and data driven methods can be used as an important

approach to complement existing tools like reservoir simulation and decline curve analysis to perform production data analysis. We recognize that even more rapid approximate methods will be required for routine analysis. Understanding the limitations of different approximate methods and application of methods to overcome these limitations in given circumstances should lead to optimal use of these methods.

Table of Contents

Acknowledgments	v
Abstract	vii
Table of Contents	x
List of Figures	xiv
List of Tables	xviii
Nomenclature	xx
1. Introduction	1
1.1 Problem Statements	8
1.2 Contributions and Dissertation Outline	8
2. Analytical models for liquid rich shale reservoirs	10
2.1 Application of trilinear flow model to a multiphase flow	11
2.2 Verification of Proposed Approach	17
2.3 Conclusions	27
3. Statistical study of reservoir and completion parameters in liquid rich shale	28

3.1 Base Case	28
3.2 Base Case Results	34
3.3 Statistical Study	37
3.4 Results	40
3.5 Study of Combined Effects	42
3.6 Observations and Conclusions	46
4. Effect of well spacing on productivity of liquid rich shale (LRS) reservoirs with multiphase flow: A simulation study	48
4.1 Workflow	49
4.2 Numerical Simulation	50
4.3 Economic Analysis	53
4.4 Results for Base Case	54
4.5 Sensitivity Study	56
4.6 Conclusions.....	59
5. Optimal decline curve analysis (DCA) models for liquid rich shale (LRS) reservoirs	61
5.1 Background	61
5.2 Arps Decline Model.....	62
5.3 Duong's Production Decline.....	63

5.4 Stretched Exponential Production Decline	64
5.5 Diagnostic Plots	65
5.6 Generation of Synthetic Field Data	67
5.7 Flow Regime Identification	70
5.8 Yu Plot.....	74
5.9 Duong’s Production Decline	76
5.10 Arps Hyperbolic Model.....	77
5.11 Production decline.....	78
5.12 Determination of Cumulative Oil Production	80
5.13 Application to field history Data.....	82
5.14 Conclusions.....	85
6. Data driven study of liquid Rich Shale Reservoirs.....	87
6.1 Background and Introduction.....	87
6.2 Literature Review and Problem Statement.....	87
6.3 Theory and Background.....	89
6.4 Prediction and Forecasting for New Wells	92
6.5 Workflow	93
6.6 Simulated Case	94

6.7 Parameter Bounds and Distribution	94
6.8 Reservoir Description and Simulation.....	95
6.9 Principal Component Analysis of the Simulated Data	97
6.10 Recreation of Original Data with Limited Principal Components	100
6.11 Analysis for Normalized (or Standardized) Data	104
6.12 Prediction of New Well Data	106
6.13 Field Example	109
6.14 Data Acquisition.....	109
6.15 Principal Component Analysis for Field Case	111
6.16 Recreation of Original Data with Limited Principal Components	112
6.17 Prediction of New Well Data	115
6.18 Conclusions.....	117
References	118
Appendix.....	127

List of Figures

Figure 1. Major shale plays in the US (EIA, 2013).....	3
Figure 2. PVT diagram for different types of hydrocarbon fluids and their production trajectory (Corelabs, 2016)	5
Figure 3 Schematic for trilinear flow, where a reservoir is divided into three flow regions	13
Figure 4. PT diagram for two sample reservoir fluids	18
Figure 5. Phase saturation with pressure (Saturation pressure marked).....	19
Figure 6. Bo and Rs for Sample 1	20
Figure 7. Bg(Gas FVF) and Rv(CGR) for Sample 1	20
Figure 8. Oil and gas viscosity (cp).....	21
Figure 9. Water-Oil relative permeability curve.....	22
Figure 10. Gas-Oil relative permeability curve.....	23
Figure 11. Comparison of cumulative oil calculated from analytical and numerical methods for Sample 1	24
Figure 12. Comparison of cumulative gas calculated from analytical and numerical methods for Sample 1	24
Figure 13. Comparison of cumulative oil calculated from analytical and numerical methods for Sample 2	25
Figure 14. Comparison of cumulative gas calculated from analytical and numerical methods for Sample 2	25
Figure 15. Oil saturation with time for analytical and numerical calculation.....	26
Figure 16. Gas Saturation for analytical and numerical calculation	26
Figure 17. Schematic for base case reservoir simulation and symmetrical element	30
Figure 18. Gas history match for Eagle Ford Shale.....	31

Figure 19. Oil history match for Eagle Ford Shale	31
Figure 20. Pressure history match for Eagle Ford Shale	32
Figure 21. Pressure-Temperature (PT) diagram for sample fluids.....	33
Figure 22. Liquid dropout curve for.....	33
Figure 23. Relative permeability curves.....	34
Figure 24. Evolution of pressure with depletion	34
Figure 25. Cumulative gas production for each fluid samples	36
Figure 26. Cumulative oil production for each fluid samples	36
Figure 27. Oil saturation in fractures and matrix during production	37
Figure 28. One parameter at a time sensitivity plot for cumulative gas	40
Figure 29. One parameter at a time sensitivity plot for cumulative oil.....	42
Figure 30. Quality check for proxy model generated from RSM (cumulative oil)	43
Figure 31. Verification for cumulative gas from proxy model	44
Figure 32. Verification for cumulative oil from proxy model	44
Figure 33. Sensitivity analysis to study the interaction of parameters with each other	45
Figure 34. Sensitivity analysis to study the interaction of parameters	45
Figure 35. Evolution of pressure with time due to depletion	54
Figure 36. Cumulative gas production for each fluid sample	55
Figure 37. Cumulative oil production for each fluid sample	55
Figure 38. NPV for different fluids for base case	56
Figure 39. NPV for lean gas condensate sample	57
Figure 40. NPV for wells with fluid around critical point.....	58
Figure 41. NPV for rich fluid sample	59

Figure 42. Parameter estimation to calculate cumulative oil by using Yu method	65
Figure 43. Diagnostic plot to identify flow regimes for a typical shale reservoir..	67
Figure 44. Schematic for reservoir to generate synthetic production data.....	68
Figure 45. Normalized rate vs MBT diagnostic plot	71
Figure 46. Normalized rate vs time diagnostic plots	72
Figure 47. Normalized rate vs MBT diagnostic plot (constant BHP).....	72
Figure 48. Normalized rate vs Time diagnostic plot (constant BHP).....	73
Figure 49. Yu plot to calculate SEPD parameters for sample 1.....	75
Figure 50. Duong's Plot to calculate parameters a & m.....	76
Figure 51. Initial rate determination by using parameters a & m.....	77
Figure 52. Forecasted cumulative gas and gas rates	79
Figure 53. Yu plot for determination of condensate production	81
Figure 54. Pressure normalized vs MBT diagnostic plot for a LRS reservoir.....	83
Figure 55. Gas rate vs time diagnostic plot for LRS reservoir	83
Figure 56. Pressure normalized vs MBT diagnostic plot for a dry gas reservoir .	85
Figure 57. Gas rate vs time diagnostic plot for a dry gas reservoir.....	85
Figure 58. Symmetric element with logarithmically spaced grids for base case simulation	96
Figure 59. . Simulated production data: Left: Gas Rate Middle: Cumulative Gas Right: CGR	98
Figure 60. Eigenvalues for gas rate, cumulative gas and CGR	99
Figure 61. Principal component coefficient/loadings for gas rate.....	99
Figure 62. Principal component coefficient/loadings for cumulative gas.....	100
Figure 63. Principal component coefficient/loadings for CGR	100

Figure 64. Maximum error seen from recreation of original data with limited principal components (gas rate)	101
Figure 65. . Minimum error seen from recreation of original data with limited principal components (gas rate)	102
Figure 66. Maximum error seen from recreation of original data with limited principal components (Cumulative Gas).....	102
Figure 67. Minimum error seen from recreation of original data with limited principal components (Cumulative Gas)	102
<i>Figure 68. Maximum error seen from recreation of original data with limited principal components (CGR)</i>	<i>103</i>
Figure 69. Minimum error seen from recreation of original data with limited principal components (CGR).....	103
Figure 70. Error in recreation each number of principal components	104
Figure 71. Cross-validation error/prediction error for gas rate, cumulative gas and CGR	108
Figure 72. Wells from LaSalle County selected for analysis (shown in green) .	110
Figure 73. Field gas and oil data	110
Figure 74. Principal components for gas rate	111
Figure 75. Principal components for oil rate	111
Figure 76. Eigenvalues (blue) and cumulative eigenvalues (magenta) for gas rate and oil rate.....	112
Figure 77. Maximum error in recreation using limited number of PC (gas rate)	113
Figure 78. Minimum error in recreation using limited number of PC (gas rate).	113
Figure 79. Maximum error in recreation using limited number of PC (oil rate)..	113
Figure 80. Minimum error in recreation using limited number of PC (oil rate)...	114
Figure 81. Error in re-creation of field data using various number of PC's	115
Figure 82. Cross-validation/prediction error for field data	116

List of Tables

Table 1. Reservoir fluid properties and composition.....	18
Table 2 Reservoir properties for numerical and analytical simulation.....	21
Table 3. Reservoir properties for base case simulation.....	30
Table 4. Peng-Robinson EOS parameters for the base case fluid sample.....	32
Table 5 Parameter bounds selected for sensitivity study.....	40
Table 6. Reservoir properties for base case simulation.....	50
Table 7. Fluid Compositions for the selected liquid rich shale (LRS).....	51
Table 8. Fluid Properties for each sample	52
Table 9. Near critical fluid with varying compositions used for simulation	68
Table 10. Tuned EOS parameter for sample with C ₇ ⁺ concentration of 7.1%.....	69
Table 11. Start and end of the boundary dominated flow for each sample.....	74
Table 12. SEPD parameters based on year 2 to 3 using Yu Plot	75
Table 13. SEPD parameters based on year 3 to 4 using Yu Plot	76
Table 14. Duong's a, m and q ₁ parameters for each samples	77
Table 15. Error calculation for each sample by using different methods	80
Table 16. Cumulative oil estimation a. Simulated b. Year 1-2 c. Year 2-3 d. Year 3-4	82
Table 17. Parameters, distribution and description.....	95
Table 18. Summary of reservoir property used for the base case	97
Table 19. Average error in recreation for each case.....	100
Table 20. Maximum error in recreation for each case	101
Table 21. Average error in recreation for each case (Normalized Sample)	105

Table 22. Maximum error in recreation for each case (Normalized Sample)	105
Table 23. Results from multiple regression for simulated data	107
Table 24. Average validation (prediction) error	109
Table 25. Maximum validation (prediction) error.....	109
Table 26. Average error in recreation for each case.....	114
Table 27. Maximum error in recreation for each case	114
Table 28. Average error in recreation for each case (normalized data)	115
Table 29. Maximum error in recreation for each case (normalized data).....	115
Table 30. Average error in validation for each case (normalized data).....	117
Table 31. Maximum error in validation for each case (normalized data)	117

Nomenclature

ANN	=	Artificial Neural Network
b	=	Arps Hyperbolic Exponent
BDF	=	Boundary Dominated Flow
BFIT	=	Before Federal Income Tax
BHP	=	Bottomhole Pressure
BTU	=	British thermal unit
Capex	=	Capital expenditure
CGR	=	Condensate to gas ratio
CMG	=	Computer Modelling Group Ltd.
DCA	=	Decline Curve Analysis
DOE	=	Design of Experiments
EFS	=	Eagle Ford Shale
EIA	=	Energy Information Administration
EOS	=	Equation of State
EUR	=	Estimated ultimate recovery
GOR	=	Gas oil ratio
LRS	=	Liquid Rich Shale
LSE	=	Least Square of Errors
LS-LGR	=	Logarithmically-spaced local grid refinement
MBT	=	Material Balance Time
MCMC	=	Markov Chain Monte Carlo
NPV	=	Net Present Value

OPAAT	=	One parameter at a time
Opex	=	Operating expenditure
PC	=	Principal Component
PCA	=	Principal component Analysis
PT	=	Pressure Temperature
PVT	=	Pressure Volume Temperature
R	=	Net Cash Inflow
RSM	=	Response Surface Modelling
RTA	=	Rate Transient Analysis
SEPD	=	Stretched Exponential Production Decline
SRV	=	Stimulated reservoir volume
SVD	=	Singular Value Decomposition
TSCF	=	Trillion standard cubic feet
TVD	=	True Vertical Depth
USGS	=	United States Geological Survey
WTI	=	West Texas Intermediate
XSRV	=	External reservoir volume
TCF	=	Trillion Cubic Feet

1. Introduction

The US Energy Information Administration (EIA) predicts that the world energy consumption will increase by 56% between 2010 and 2040 from 524 quad BTU to 820 quad BTU. Almost 80% of that demand will be fulfilled by fossil fuels through 2040 (EIA, 2013). Thus far, this increase in the demand for energy has been sustained by unconventional oil and gas resources in North America.

The production of hydrocarbons from unconventional sources is not as straightforward as production from high to medium permeability conventional reservoirs. Hydrocarbons from conventional reservoirs can be produced at economic flow rates and volumes without large stimulation treatments or special recovery processes. On the other hand, unconventional reservoirs require horizontal drilling, stimulation treatments and other expensive recovery processes, which make them more challenging to exploit than conventional reservoirs. Conventional reservoir performance, depending on the fluid system, is observed for permeability of greater than about 0.5 md, whereas of unconventional shale reservoirs, with permeability in the range of 1000 nd or less, behave quite differently (Whitson, 2014). The examples of unconventional resources are shale gas and liquids, tight gas and oil, coal bed Methane (CBM), heavy oil, and gas hydrates.

Technological advances like horizontal drilling, multistage hydraulic fracturing and the use of advanced proppants have allowed us to exploit these resources with ultralow permeability at a reasonable profit. However, there are still several factors that need to be considered in unconventional reservoirs to reduce the cost of production and remain competitive in the world energy market. For

example, gas condensate reservoirs show complex subsurface pressure, volume, and temperature (PVT) behavior, which results in difficult and often inaccurate production forecasting and reserve estimation. Gas condensate, initially a single-phase gas, undergoes retrograde condensation in the reservoir below the dew point pressure. This behavior results in more complicated multiphase flow than in black oil and dry gas reservoirs. Furthermore, the retrograde condensate forms a condensate bank near the well and reduces the overall well productivity for both oil and gas. Similarly, volatile oil which initially is in a single phase oil, show multiphase flow below the bubble-point pressure.

Historically, the commercial value of the liquefiable hydrocarbons extracted from North American natural gas has been greater than the commercial value of the thermal content that would be obtained if the entire production consisted of dry gas. When converted to barrels using BTU equivalence, liquid hydrocarbons (WTI, Brent Crude) are up to 6 times more valuable than natural gas. Even with the current deflated prices, BTU's from liquid hydrocarbons are several times more valuable than BTU's from gas. Historical data, falling rig counts and reduction in capital expenditures by major oil companies all point toward a future where the demand for oil is higher than present. Liquid rich shale (LRS) reservoirs present a very attractive source to support that demand.

Figure 1 shows various shale oil and gas plays in the US as of 2011. The Eagle Ford in South Texas and the Bakken in North Dakota have changed the energy landscape of the world in last few years. The Eagle Ford Shale is one of the world's fastest growing shale plays with more than 7 billion barrels of proved

light sweet crude oil according to United States Geological Survey (USGS, 2013). The formation has three distinct regions producing different hydrocarbon fluids ranging from dry gas in the southern region to black oil in the northern region of the play. The intermediate regions produce volatile oil, retrograde gas condensate and wet gas. Bakken, located in Williston Basin is the largest continuous oil accumulation ever accessed in the US (USGS, 2013). A survey in 2013 by USGS reported that the Bakken hold 7.4 billion barrels of oil, 6.7 TSCF of natural gas and 0.53 barrels of natural gas liquids which are recoverable with the current technology. The formation's remarkable production has helped North Dakota become the second-largest oil producing state, behind Texas (USGS, 2013). In the last several years extensive literatures have been published describing the geological properties of these shale plays.

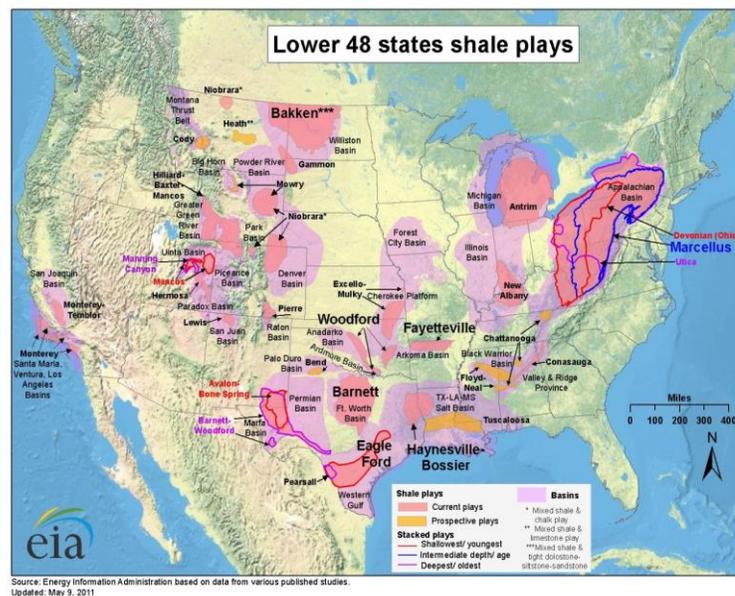


Figure 1. Major shale plays in the US (EIA, 2013)

There are five main types of hydrocarbon reservoirs, namely dry gas, wet gas, gas condensate, volatile oil and black oil. There are many differences among

them, in terms of properties and composition. The focus of this study is liquid rich shale, namely gas condensate and volatile oil reservoirs which are mainly differentiated by the initial phase. A gas condensate consists predominantly of methane and other short chain hydrocarbons, but it also consists of long-chain hydrocarbons also called the heavy ends (C^{7+}) in significant amount. The heavy ends range from 5% to 12.5 mol% and gas to oil ratio (GOR) varies from 3300 to 50000 Scf/Stb in typical gas condensates. The stock tank oil gravity ($^{\circ}$ API) for a gas condensate ranges from 45 to 60. Volatile oils have heavy ends ranging from 10 to 30 mol%, GOR from 900 to 3500 Scf/Stb and $^{\circ}$ API of around 42 to 55 (McCain Jr., 1990).

Pressure-temperature (PT) phase envelope for different kinds of reservoir fluids is shown in Figure 2. Production of gas condensate and volatile oil fluids follow a complex trajectory. Initially, an under-saturated fluid is in single phase oil or gas, as the reservoir is depleted, secondary phase appears in the reservoir. For gas condensate fluids, retrograde condensation occurs causing liquid to drop out from the gas phase. For volatile oil reservoirs, considerable vaporization occurs below the bubble point pressure. In both cases, the secondary phase leads to a multiphase flow in the reservoir and the wellbore. This causes composition of the reservoir fluid to become progressively different from the initial fluid composition.

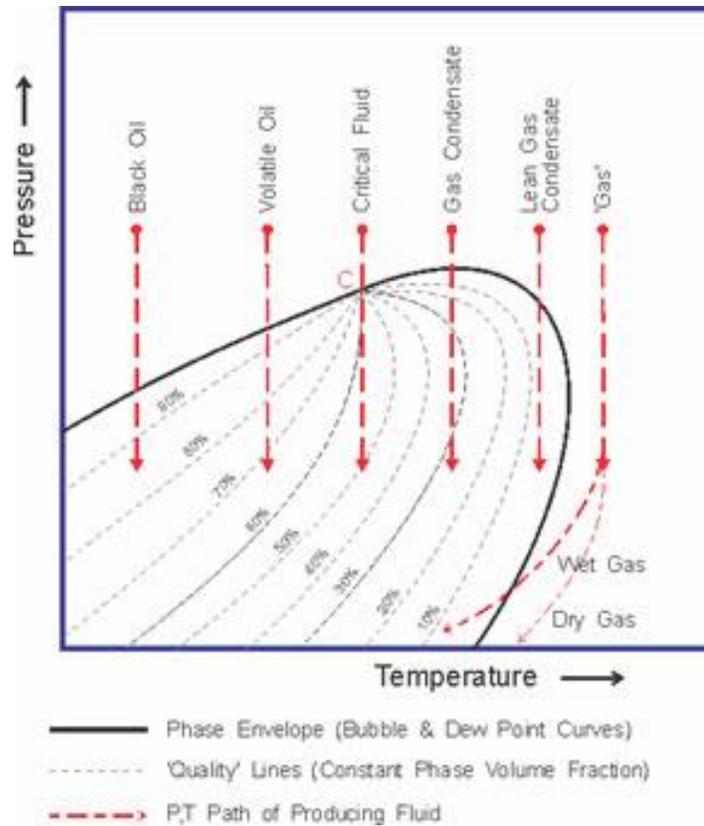


Figure 2. PVT diagram for different types of hydrocarbon fluids and their production trajectory (Corelabs, 2016)

Lee and Sidle (2011) concluded that reservoir simulation can be used as a reliable method for reserve estimation when verified with other methods like hindcasting and analogues. Compositional reservoir simulation is an effective method to simulate multiphase flow in liquid rich shale gas condensate and volatile oil reservoirs, as each individual component is tracked in this method. A compositional simulation, unlike the traditional black oil model, realistically captures changes in composition and liquid dropout below the saturation pressure. Type-curve and decline-curve analyses are other useful methods in estimating reserves.

A decline curve fitting method presented by Arps (1945) is the most widely used method for both conventional and unconventional wells. This is an empirical method based on curve fitting and it is not based on fundamentals of fluid flow in porous media. For this method to be applicable, a reservoir should be in BDF with constant drainage area and constant flowing bottomhole pressure (BHP). Unconventional reservoirs exhibit long transient linear flow periods which could last for years because of their low matrix permeability (10^{-9} md $< k < 10^{-3}$ md). In extreme cases, the wells could still be in the transient linear flow regime when they reach their economic limits. For this period, the Arps decline model exponent, b , is significantly greater than 1 (2 for linear and 4 for bilinear flow) and keeps on changing throughout the life of the well. Usually, however, transient linear flow is followed by boundary dominated flow due to intra-fracture pressure interference, followed by a transition flow regime and finally by BDF influenced by the well's drainage boundaries.

Due to the long transient-flow and limited production history, Arps model is not widely applicable to unconventional reservoirs with nano-darcy permeability. Several new decline curve models have been developed to estimate reserves in unconventional reservoirs with long transient flow. Ilk et al. (2008) introduced power-law decline curve to model the change in decline exponent with time. This model has four parameters (as opposed to three in hyperbolic decline model), which results in non-unique solution. Valko and Lee (2010) introduced the stretched exponential production decline (SEPD) model that changes from transient flow to boundary dominated flow smoothly. This model, although

applicable to transient flow, often delivers conservative estimate of the reserves in absence of limited field data. Duong (2011) developed a method based on linear flow in unconventional reservoirs. However this method usually overestimates the reserves if used without any modification as the reservoirs eventually enter boundary dominated flow. Other methods like modified stretched exponential production decline by Yu(2013) and extended exponential decline model by Zhang et al., (2015) provide alternative methods for reserves forecasting and decline curve analysis for unconventional reservoirs. There is no general consensus among industry professionals on which methods is the most suitable one for forecasting unconventional oil and gas reservoirs.

Analytical and numerical methods are the primary means of reservoir study and production forecasting in the industry. These methods are widely understood and have worked well for many cases. Tools like decline curve analysis (DCA), straight-line methods, rigorous reservoir simulation, and history matching provide valuable insights to mitigate these risks and uncertainties. Additionally, in the past few years there has been a huge interest in data driven methods where production data from the existing wells are used to forecast the performance and production from the new wells. Each of these tools and others has its own utility depending on available information and resources, and each often provide complementary answers to the problem faced. Data driven methods like Principal Component Analysis (PCA) provides a good supplement to these tools and can be used in addition to the traditional methods to obtain a quick, yet reliable, result.

1.1 Problem Statements

This work addresses the following problem states for the liquid rich gas condensate reservoirs:

1. Application of simple analytical model to forecast production from Liquid rich shale reservoirs: Liquid rich shale reservoirs with multiphase flow are described by complex non-linear partial differential equations. The objective is to use simplified trilinear flow formulation to analyze simple block reservoir and compare it with compositional reservoir simulation.
2. Statistical study of reservoir parameters: The objective is to perform statistical analysis of liquid rich shale reservoirs to understand how the completion and reservoir parameters interact with each other, and to identify the optimal number of fractures and wells in a field.
3. Application of decline curve models to forecast liquid rich shale reservoirs: The objective is to evaluate the use of decline curve analysis for liquid rich shale reservoirs and identify their effectiveness in production forecasting.
4. Application of data driven analysis for production forecasting: The objective is to develop a simple yet intuitive tool that employs the existing production data from wells, completion parameters to make decisions on planning new wells in liquid rich shale plays.

1.2 Contributions and Dissertation Outline

The first part of the dissertation (Chapter 2) explores the application of analytical methods for liquid rich shale reservoirs with multiphase flow. Diffusivity

model for multiphase flow is analyzed, and simplified analytical model is derived which matches well with compositional reservoir simulation in spite of few limitations. The advantage of using such model is illustrated by comparing the runtime with complex reservoir simulation.

The second part of the dissertation (Chapter 3 and Chapter 4) is focused on the application of statistical methods, Design of Experiments (DOE) and Response Surface Modeling (RSM), to identify the parameters that affect the production of oil and gas condensate reservoirs. This provides an insight on how each parameter affects the production performance of liquid rich shale reservoirs. Furthermore, the problem of optimal well-spacing is analyzed by using a workflow which can be used for similar problems in different plays.

In Chapter 5, several decline curve analysis methods are evaluated to generate a workflow for optimal use of decline curve analysis for production forecasting. The transitional period between the linear flow and the boundary dominated flow is identified to have a significant effect on production forecasting of liquid rich shale reservoirs.

Chapter 6 is focused on production data analysis by using data driven methods. We use principal component analysis (PCA) and multivariate regression to perform production forecasting from liquid rich shale reservoirs. Production data from new wells are forecasted by using regression when reservoir parameters are available by creating a proxy model. When reservoir parameters are not available, least sum of squares is used to estimate the well specific principal component scores.

Chapter 7 includes the conclusions and recommendations for future work.

2. Analytical models for liquid rich shale reservoirs

Diffusivity equation for a slightly compressible liquid in a homogenous reservoir is

$$\nabla^2 p = \frac{\mu \phi c_t}{k} \frac{\partial p}{\partial t}, \quad (2.1)$$

where p is pressure (psia), μ (cp) is viscosity, ϕ is porosity, c_t (psi^{-1}) is total compressibility, k (md) is permeability, and t (days) is time.

This can be extended to multiphase flow (in terms of surface volume) as follows

$$\nabla \cdot \left[\left(\frac{k_{rg}}{\mu_g B_g} + \frac{R_s k_{ro}}{\mu_o B_o} \right) \nabla p \right] = \frac{\phi}{k} \frac{\partial}{\partial t} \left(\frac{S_g}{B_g} + \frac{R_s S_o}{B_o} \right) \text{ and} \quad (2.2)$$

$$\nabla \cdot \left[\left(\frac{k_{ro}}{\mu_o B_o} + \frac{R_v k_{rg}}{\mu_g B_g} \right) \nabla p \right] = \frac{\phi}{k} \frac{\partial}{\partial t} \left(\frac{S_o}{B_o} + \frac{R_v S_g}{B_g} \right), \quad (2.3)$$

where k_{rg} and k_{ro} are gas relative permeability and oil relative permeability, R_s and R_v are solution gas oil ratio (Scf/Stb) and oil gas ratio (Stb/Scf), B_g and B_o are gas formation volume factor (RB/Scf) and oil formation volume factor (RB/Stb), S_g and S_o are gas saturation and oil saturation.

The equations above represent the non-linear partial differential equations, as the terms within the braces are function of both pressure and saturation.

The single phase diffusivity equation has been solved analytically (Brown et al., 2009), assuming a trilinear flow formulation. It is assumed that the flow in the reservoir can be divided into three regions, inner reservoir, outer reservoir, and hydraulic fracture as shown in Figure 3. This was further expanded (Stalgorova et

al., 2013) to five region model where inner and outer reservoir was further subdivided into 2 regions with respective rock and fluid properties. In both of these cases, the model generated was tested for single phase flow. In this work, we expand this model with simplified assumptions to forecast multiphase flow in liquid rich shale reservoirs with multiphase flow.

2.1 Application of trilinear flow model to a multiphase flow

A homogeneous bounded reservoir with uniformly spaced hydraulic fractures is considered as shown in the Figure 3. The notations used in chapters are based on Brown et al. The trilinear flow model considers three regions the inner reservoir, the outer reservoir and the hydraulic fractures. Only one fracture is modeled as all the fractures are assumed to be identical and uniformly spaced. The drainage area for each fracture is $2x_e$ by $2y_e$, where x_e is the well spacing (or reservoir width), y_e is the half distance between the fractures and x_f the fracture half length.

The dimensionless time is defined as

$$t_D = 0.000264 \frac{kt}{\phi \mu_{oi} c_{ti} x_f^2}. \quad (2.4)$$

For single phase flow, Brown et al. (2009) defined dimensionless pressure (p_D) for oil as

$$p_D = \frac{k_1 h}{141.2 q_F B_o \mu} (p_i - p). \quad (2.5)$$

For gas reservoirs, fluid properties are dependent on reservoir pressure and temperature, so pseudo-pressure formulation is

$$p_D = \frac{k_1 h}{141.2 q_F T} (m(p_i) - m(p)), \quad (2.6)$$

where T is reservoir temperature (°R) AND m(p) (psi²/cp) is pseudo-pressure defined as

$$m(p) = \text{gas pseudopressure} = 2 \int_0^p \frac{p dp}{\mu(p) Z(p)}. \quad (2.7)$$

Analogous to a single phase formulation, three phase dimensionless pressure is defined as

$$p_D = \lambda_t \frac{kh(p_i - p)}{141.2(q_o B_o + q_g B_g + R_w B_w)}. \quad (2.8)$$

where λ_t (md/cp) is the total mobility defined as the sum of oil, water, and gas mobility given as

$$\lambda_t = \left(\frac{k_{ro}}{\mu_o} + \frac{k_{rg}}{\mu_g} + \frac{k_{rw}}{\mu_w} \right). \quad (2.9)$$

At this point, initial water saturation is assumed to be irreducible to simplify the calculations. This reduces this three phase flow problem to two phase flow problem. This assumption is made to simplify the spreadsheet/calculation code and can be relaxed depending on the reservoir conditions in future work.

The dimensionless lengths are defined as

$$x_D = \frac{x}{x_f} \text{ and} \quad (2.10)$$

$$y_D = \frac{y}{x_f}. \quad (2.11)$$

where $h(\text{ft})$ is the thickness of the reservoir, $r_w(\text{ft})$ is the wellbore radius. The parameter α_F is

$$\alpha_F = 2 \frac{\beta_F}{C_{FD}} + \frac{s}{\eta_{FD}}, \quad (2.15)$$

where η_{FD} is the dimensionless diffusivity is

$$\eta_{FD} = \frac{\eta_F}{\eta_I}, \quad (2.16)$$

$$\eta_F = 2.637 \times 10^{-4} \frac{k_F}{(\phi c_t)_F \mu} \text{ and} \quad (2.17)$$

$$\eta_I = 2.637 \times 10^{-4} \frac{k_I}{(\phi c_t)_I \mu}, \quad (2.18)$$

where β is defined as

$$\beta_F = \sqrt{\alpha_0} \tanh\left(\sqrt{\alpha_0} \left(y_{eD} - \frac{w_D}{2}\right)\right), \quad (2.19)$$

where y_{eD} and w_D are the dimensionless inter-fracture spacing and fracture width respectively. Constant α_0 is defined as

$$\alpha_0 = \frac{\beta_0}{C_{RD} y_{eD}} + s, \quad (2.20)$$

where β_0 and C_{RD} are defined

$$\beta_0 = \sqrt{s} + \tanh(\sqrt{s}(x_{eD} - 1)) \text{ and} \quad (2.21)$$

$$C_{RD} = \frac{x_f}{y_e}. \quad (2.22)$$

These expressions assume a homogeneous reservoir with no natural fractures. The dual porosity case is given by Brown et al. (2009), requires further dual porosity parameters.

The dimensionless wellbore pressure in Laplace domain is numerically inverted by using Gaver-Stehfast algorithm. Oil production rate is calculated by using the previously defined dimensionless variable for pressure as

$$q_o = n_f \left(\frac{k_{ro}}{\mu_o} + \frac{k_{rg}}{\mu_g} \right) \frac{kh(p_i - p_{wf})}{141.2(B_o + R_g B_g)}. \quad (2.23)$$

The constants n_f and p_{wf} represents the number of fractures and wellbore pressure respectively. The gas rate is estimated by using

$$q_g = R_g q_o, \quad (2.24)$$

where R_g is producing gas-oil ratio defined as

$$R_g = R_{so} + \frac{\lambda_g B_o}{\lambda_o B_g}. \quad (2.25)$$

where R_{so} is the solution gas-oil ratio(Mscf/Stb).

The equations above can be used to calculate the oil, water and gas rate from a liquid rich shale reservoir in presence of reservoir and fluid properties. These properties are functions of reservoir pressure and temperature. For simplicity, in this work it is assumed that the reservoir permeability and porosity are constant (pressure independent). Moreover, it is also assumed that the initial reservoir water saturation is irreducible. PVT properties including formation volume factors, fluid viscosities, which are pressure dependent, are calculated in laboratory. For this work, a PVT software, CMG-Winprop is used to calculate these “modified-black oil” properties by using a fluid compositions of Liquid Rich Shale samples. Two phase relative-permeability curves are calculated by using Corey’s exponents. Pressure saturation relationship is required to calculate relative permeability curves and PVT properties as the reservoir is depleted with time. For this reason, a simple tank model is used to model depletion from the reservoir. Initially only oil is assumed to be present in the reservoir and the reservoir pressure is assumed to be above the saturation pressure. Rock and liquid expansion drive is assumed above the bubble point pressure and depletion drive mechanism is

assumed below the bubble point pressure. Above the bubble point, single phase flow is seen, below the bubble point both oil and gas flow to the wellbore.

The compressibility coefficient c which explains the change in volume due to changing pressure is given as

$$c = -\frac{1}{V} \frac{dV}{dp}, \quad (2.26)$$

where c is compressibility, V is rock volume dependent on pressure.

Rock pore volume can be calculated by modifying Equation 2.26 as

$$V = V_i \exp(-c_r(p_i - p_a)). \quad (2.27)$$

The incremental oil produced from the reservoir with dp drop in pressure is given as

$$dN_p = \frac{V c_t dp}{B_o}, \quad (2.28)$$

where c_t is the total compressibility given by the sum of rock and water compressibility.

For liquid rich shale, the oil saturation keeps on changing as the reservoir is depleted. Hence, ideally distance of investigation should be used to identify the region of influence and calculate the saturation in the discretized reservoir. We assume that the oil saturation remains constant till the average reservoir pressure falls below the saturation pressure. This is only valid till reservoir reaches the saturation pressure, after which we see the evolution of the second phase (gas phase). Below the bubble point the oil saturation pressure relationship is assumed to follow the solution-gas drive model (Muscat, 1981, Shojaei et al., 2013) as

$$\frac{dS_o}{dp} = \frac{\frac{S_o B_g dR_{so}}{B_o dp} + S_o \frac{k_{rg} \mu_o}{k_{ro} \mu_g B_o} \frac{dB_o}{dp} - S_g \frac{1}{B_g} \frac{dB_g}{dp}}{1 + \frac{k_{rg} \mu_o}{k_{ro} \mu_g}}. \quad (2.29)$$

The incremental oil production below the bubble point pressure is calculated by

$$dN_p = \frac{\frac{dS_o}{dp} dp V}{B_o}. \quad (2.30)$$

The incremental gas production is calculated by

$$dG_p = R_g dN_p. \quad (2.31)$$

If we assume the rate obtained from the analytical model (Equation 2.23) is identical to the one obtained from material balance equations (Equation 2.29 and 2.30) , oil rate is defined as

$$q_o = \frac{dN_p}{dt}. \quad (2.32)$$

Based on Equations 2.1 through 2.32, production of liquid rich volatile oil reservoirs for both oil and gas was calculated which is explained below.

2.2 Verification of Proposed Approach

The proposed analytical model was compared to the compositional numerical simulation by using the commercial reservoir simulator CMG-GEM. Composition and other properties of the samples are listed in Table 1 below (Whitson et al., 2012).

Table 1. Reservoir fluid properties and composition

	Sample 1	Sample 2	Composition	Sample 1	Sample
Temperature (°F)	250	250	CO2	1.95	1.51
Saturation psia	3800	2560	N2	0.13	0.1
OGR, STB/MMScf	1000	2000	C1	48.85	37.89
GOR, Scf/STB	1000	500	C2	6.25	4.85
° API	40	37.7	C3	3.5	2.71
C7+ Mol Wt	195	216	I-C4	0.81	0.62
C7+ Mol%	34.88	49.49	N-C4	1.37	1.07
SG C7+	0.81	0.82	I-C5	0.63	0.49
			N-C5	0.68	0.52
			C6	0.96	0.74
			C7+	34.88	49.49

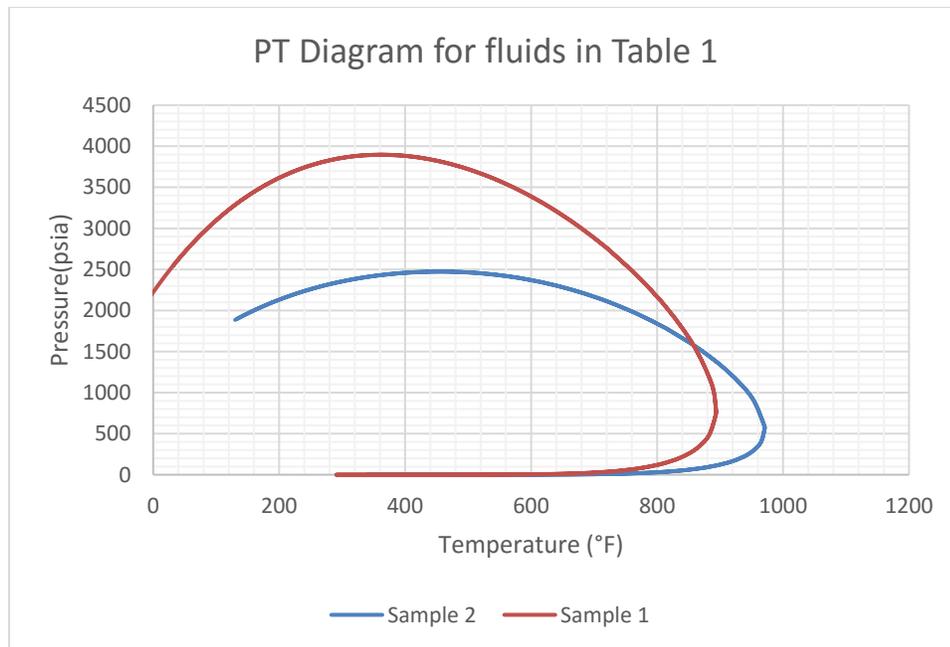


Figure 4. PT diagram for two sample reservoir fluids

Compositions of sample 1 and 2 are shown in Table 1. Each fluid consists of significant amount of light hydrocarbons including methane which leads to a high degree of shrinkage. The initial solution GOR is calculated from a separator test at the pressure of 400 psia and temperature of 100 °F. The PT diagram for samples

1 and 2 are shown in Figure 4. For the reservoir temperature of 250 °F, each sample is initially assumed to be single phase liquid. Figure 5 shows that when the reservoir pressure drops below the saturation pressure, secondary gas phase is seen in the reservoir.

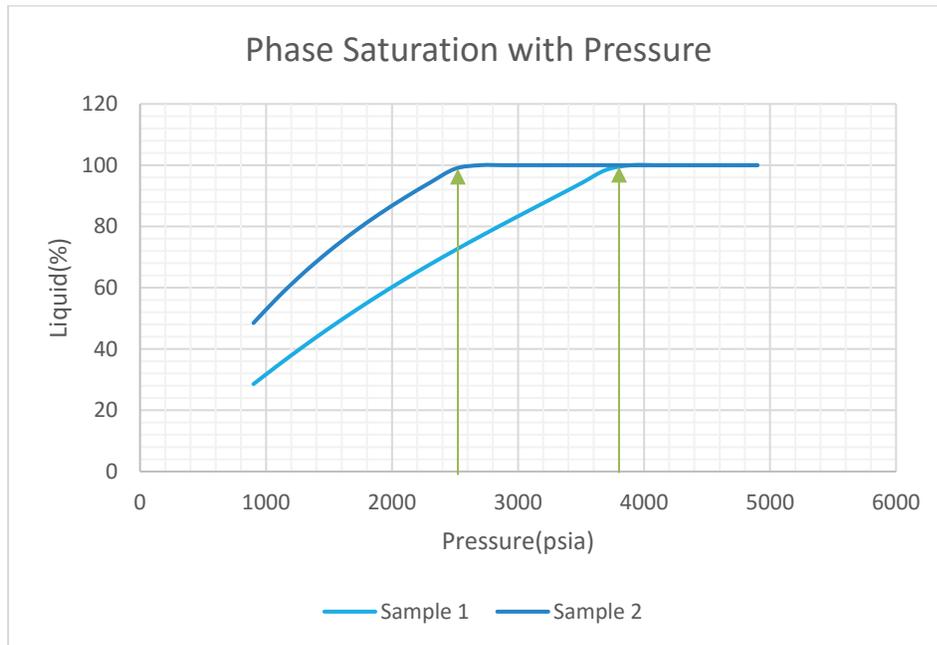


Figure 5. Phase saturation with pressure (Saturation pressure marked)

To use the analytical equations derived above, pressure dependent black oil PVT properties are required for each fluid. Gas formation volume factor (B_g), oil formation volume factor (B_o), gas viscosity (μ_g), oil viscosity (μ_o) and solution gas-oil ratio (R_s) was calculated by using the flash calculation (Whitson and Torp 1983). Properties above original saturation pressure were extrapolated based on linear relationship between R_s and B_o . The black oil properties for Sample 1 is given in Figures 6 through 8.

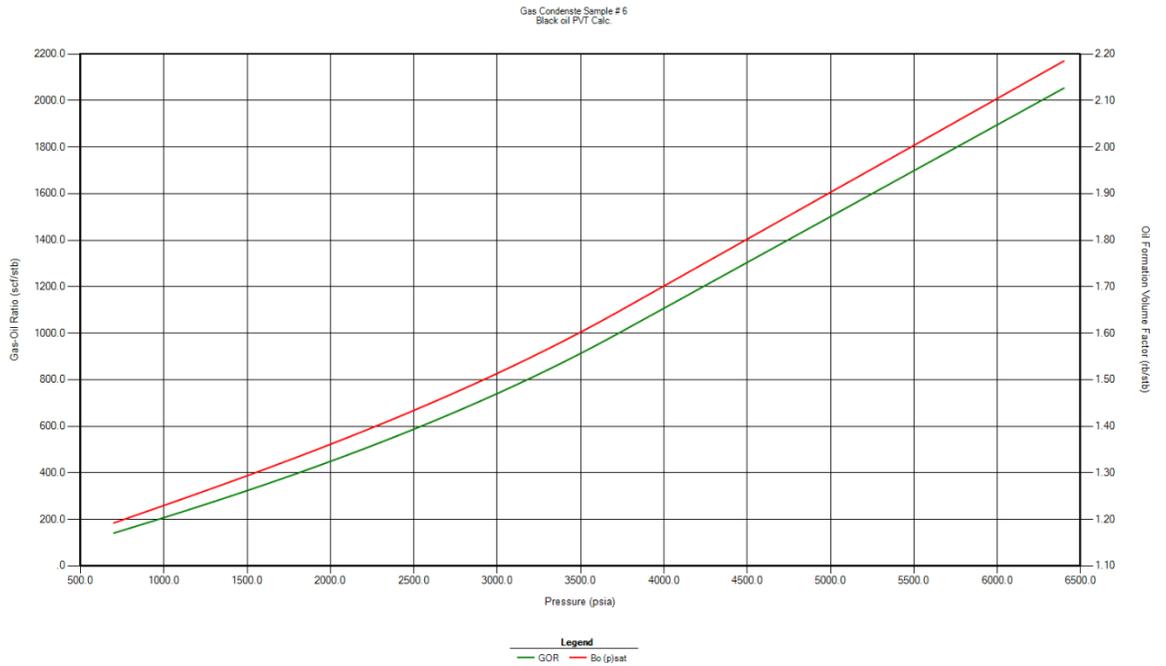


Figure 6. Bo and Rs for Sample 1

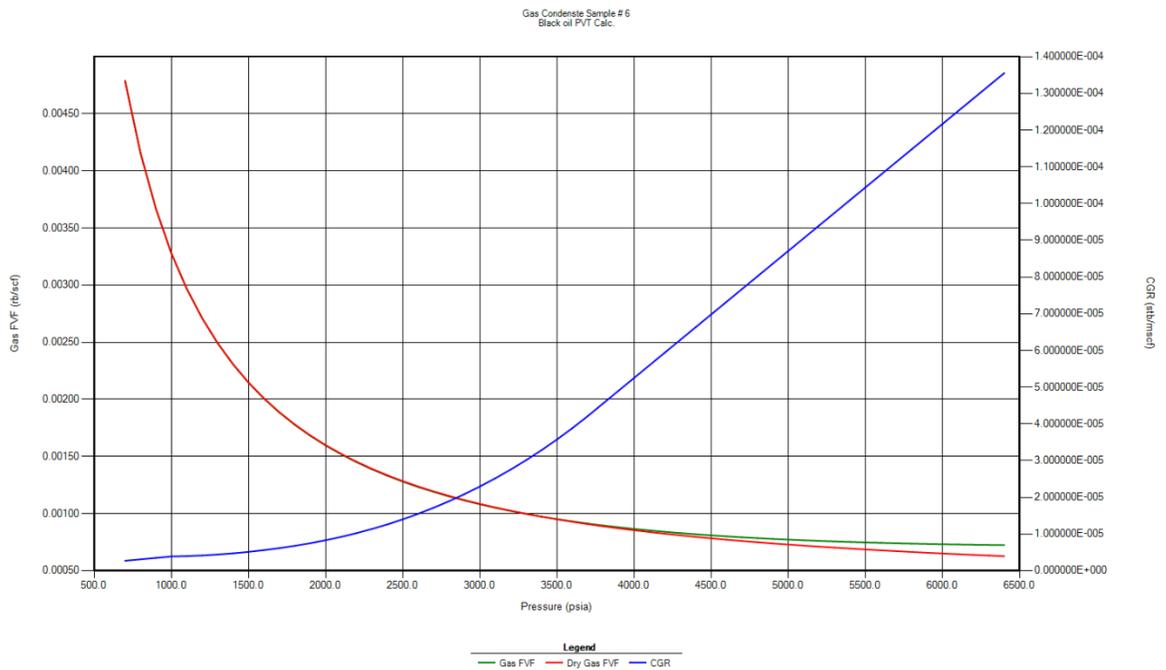


Figure 7. Bg(Gas FVF) and Rv(CGR) for Sample 1

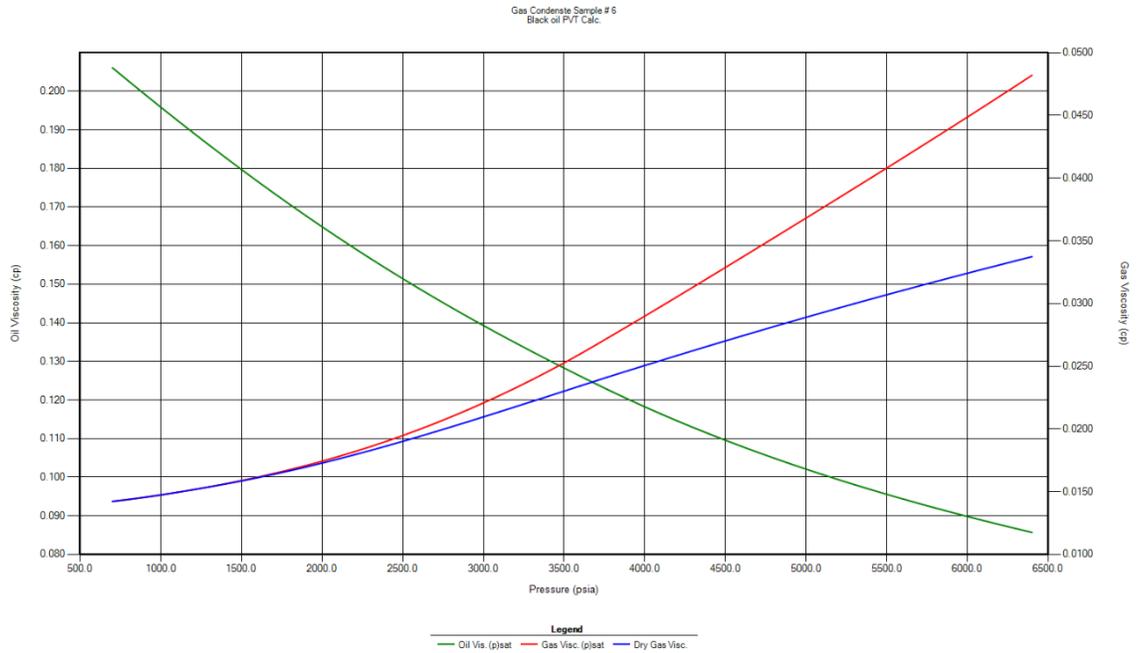


Figure 8. Oil and gas viscosity (cp)

The reservoir, well and fracture properties for analytical and numerical model are presented in Table 2:

Table 2 Reservoir properties for numerical and analytical simulation

Reservoir Properties	Value	Units
Matrix Permeability	500	nd
Porosity	7	%
Initial Pressure	5800	psia
Minimum BHP	1000	psia
Temperature	250	°F
Initial Water Saturation	0.22	
Initial Oil Saturation	0.78	
Well Radius	0.365	ft
Well Spacing (x_e)	1320	ft
Thickness(h)	45	ft
Fracture Spacing (y_e)	264	ft
Fracture Half Length (x_f)	150	ft
Fracture Conductivity	200	md.ft

For both numerical and analytical model, relative permeability curves are required to account for multiphase flow below the saturation pressure. Since water saturation is assumed to be constant (irreducible water), matrix permeability is used when the reservoir is above the saturation pressure. Below the saturation pressure, gas-oil relative permeability curves are used to account for multiphase flow of oil and gas. The water-oil relative permeability curve shown in Figure 9 should be used for cases with initial saturation greater than critical water saturation. The two phase relative permeability curves were calculated by using Corey's exponents from Shojaei et al. (2013) as shown in Figure 9 and Figure 10. This work only accounts for two-phase flow, however three phase relative involving additional water phase can be calculated by three-phase relative permeability curve generated from Stone's model(Stone 1970 ; Stone 1973).

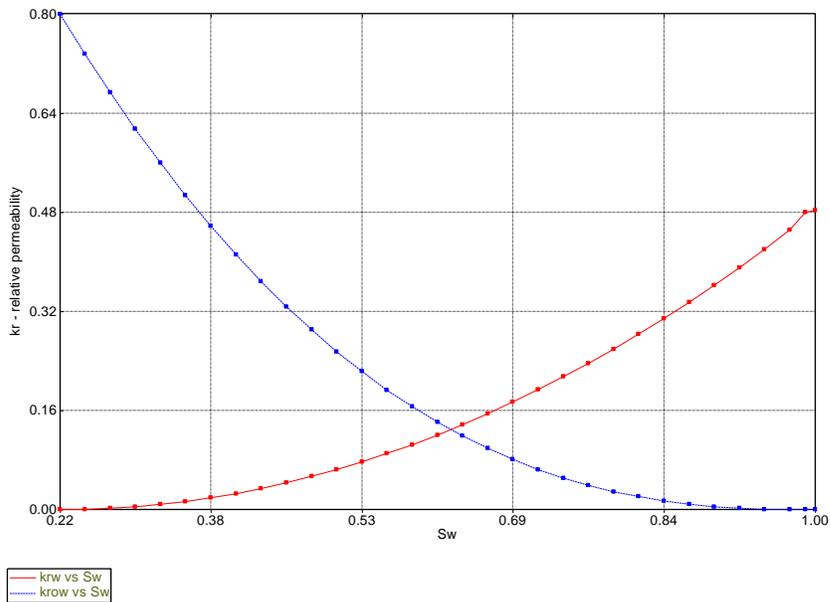


Figure 9. Water-Oil relative permeability curve

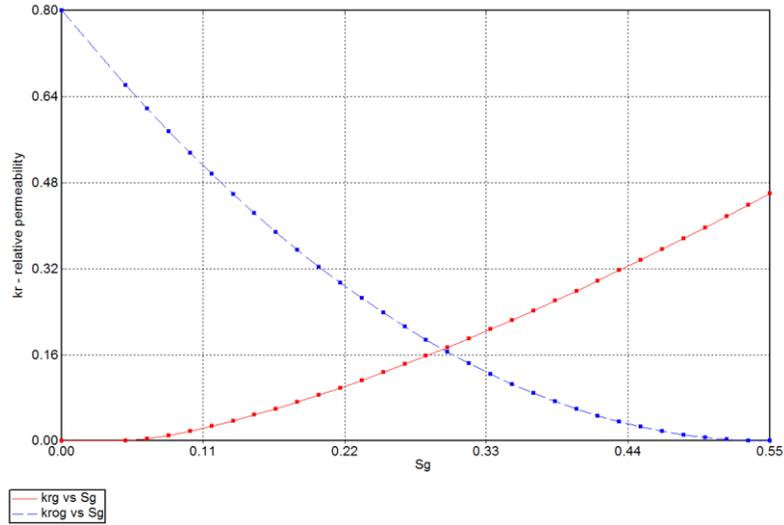


Figure 10. Gas-Oil relative permeability curve

The analytical model and the numerical model were prepared in Matlab and CMG-GEM respectively. For numerical model, Peng-Robinson equation of state (Peng and Robinson, 1976) was used to generate the equation of state (EOS) parameters. The EOS parameters were tuned to match the reported saturation pressure, GOR and API. The reservoir simulation grid-blocks were logarithmically refined around the fractures to capture the rapid change in saturation and pressure around the region. Results of both analytical and numerical simulation are shown in Figure 11 through Figure 14.

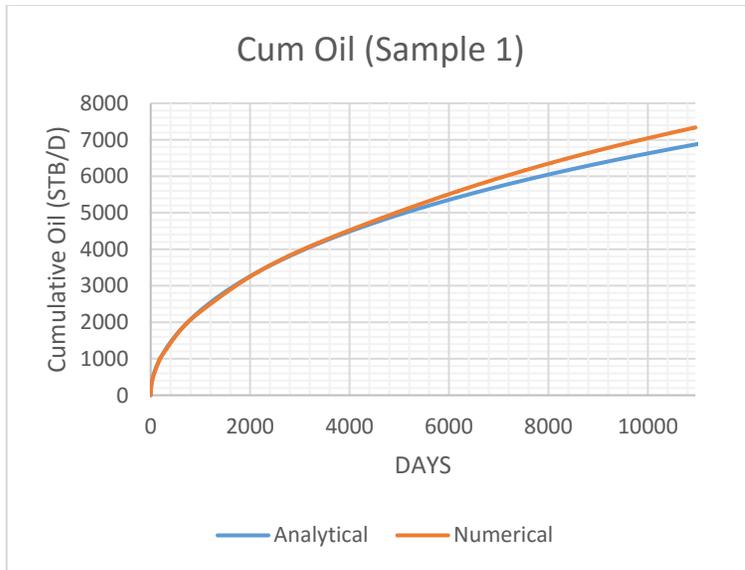


Figure 11. Comparison of cumulative oil calculated from analytical and numerical methods for Sample 1

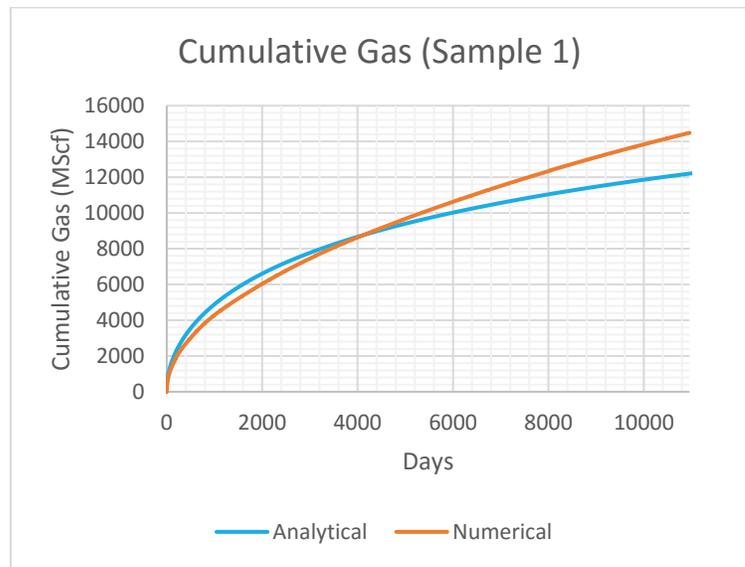


Figure 12. Comparison of cumulative gas calculated from analytical and numerical methods for Sample 1

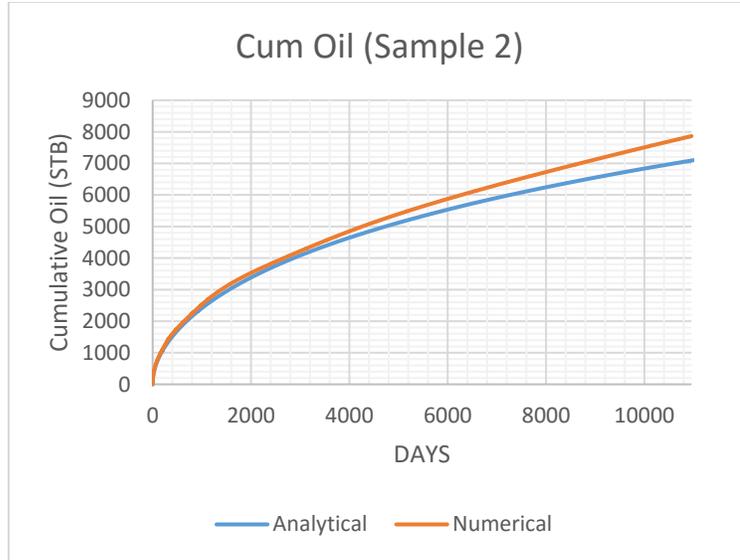


Figure 13. Comparison of cumulative oil calculated from analytical and numerical methods for Sample 2

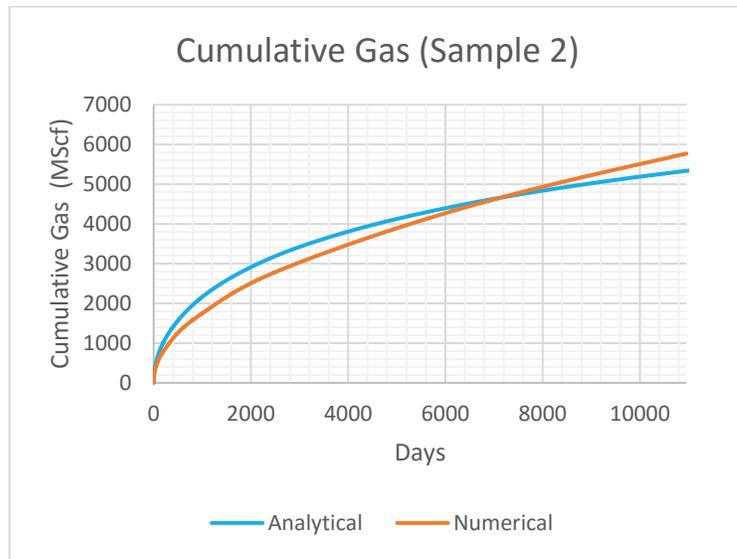


Figure 14. Comparison of cumulative gas calculated from analytical and numerical methods for Sample 2

The final cumulative production calculated from analytical equations show a good agreement with the numerical solutions. For cumulative oil, the difference in estimated ultimate recovery (EUR) was 6.4% and 3.2% for sample 1 and sample 2 respectively. Similarly, for cumulative gas, the difference in EUR was 15.5% and 7.4% for sample 1 and sample 2 respectively. The average oil and gas saturation

also show a good match for each samples. The average reservoir pressure at the end of the production for each sample was comparable at around 3500 psia.

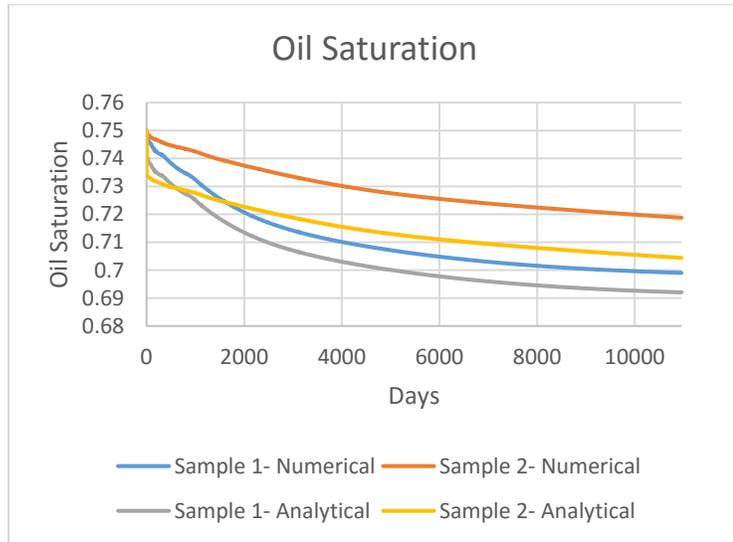


Figure 15. Oil saturation with time for analytical and numerical calculation

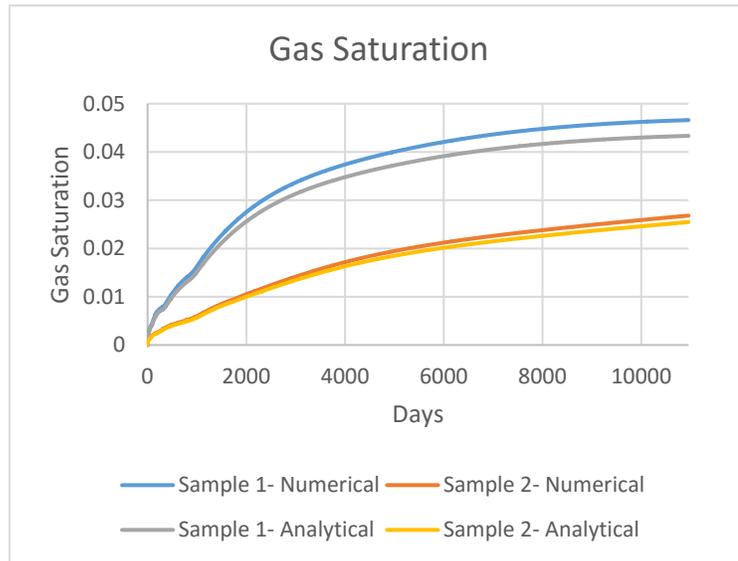


Figure 16. Gas Saturation for analytical and numerical calculation

The final cumulative production calculated from analytical equations show a good agreement with the numerical solutions. For cumulative oil, the difference in estimated ultimate recovery (EUR) was 6.4% and 3.2% for sample 1 and sample 2 respectively. Similarly, for cumulative gas, the difference in EUR was 15.5% and

7.4% for sample 1 and sample 2 respectively. The average oil and gas saturation also show a good match for each sample. Results show that this simplified modification of tri-linear flow formulation yields a comparable result to numerical simulation. Numerical simulations are significantly more complex, time consuming, and resource intensive. For this work, the analytical models yielded results in the matter of seconds. Numerical models, however required up to 15 minutes for completion. The choice of model for a certain analysis should be made based on tradeoff between accuracy and simplicity.

2.3 Conclusions

The presented simplified analytical method provides a quick yet efficient method for estimating the production from liquid rich shale reservoirs initially in liquid phase. The results highlight that the trilinear flow solution can be adapted to forecast production from liquid rich shale reservoirs.

In this work, we have modified the trilinear flow solution for single phase flow and used it to forecast production from liquid rich shale reservoirs. This method was verified to be quick and reasonably accurate compared to compositional reservoir simulation. This method needs to be modified by using pseudo-pressure and pseudo-time for reservoirs initially in gas phase to account for pressure dependence of gas viscosity.

3. Statistical study of reservoir and completion parameters in liquid rich shale

Production from unconventional liquid rich shale reservoirs is quite different compared to conventional reservoirs due to their complex PVT properties and heterogeneous rock properties. This brings uncertainties when analyzing the production performance of a liquid rich shale reservoir. The goal of this chapter is to identify various factors that affect the production of oil and gas from Liquid Rich Shale (LRS) gas condensate reservoirs, using the extensive public data base of reservoir properties, fluid properties, completions, etc., to identify the parameters that affect production performance from liquid rich shale reservoirs.

3.1 Base Case

Compositional reservoir simulation is an effective method to simulate multiphase flow in retrograde condensate reservoirs, especially in early stages of field developments. A compositional simulation, unlike the traditional black oil model, realistically captures changes in composition and liquid dropout below the saturation pressure. In this work, CMG-GEM, a compositional simulator provided to us by Computer Modelling Group, was used to model a single porosity and permeability reservoir with multiple hydraulic fractures and fluid flow in geology similar to the Eagle Ford Shale in South Texas. Hydraulic fractures are modeled by using logarithmically-spaced local grid refinement (LS-LGR), which captures the flow of fluid and change in pressure in the near-fracture region.

For the base case, a 640 acres reservoir with 4 horizontal wells and 20 stages of bi-wing hydraulic fractures was considered. Hence, the well spacing and the fracture spacing for the base case is 160 acres/well and 8 acres/fracture respectively. A symmetric element with dimensions of 1320 ft., 264 ft. and 200 ft. for length, breadth and thickness respectively was simulated to reduce the computational time. As shown in Figure 17, all fractures are assumed to be identical in dimensions and properties. TVD for each well was assumed to be 12,000 ft. Dimensionless fracture conductivity of greater than 300 was used for each case to ensure negligible pressure drop in hydraulic fractures. The fracture width was set as 2 ft for the simulation, but keeping the same fracture conductivity in the computations. Dimensionless fracture conductivity is defined as

$$F_{CD} = \frac{w_F \times k_F}{x_F \times k_{Matrix}} \quad (3.1)$$

Initial reservoir pressure was calculated by using a pressure gradient of 0.80psi/ft. Similarly, reservoir temperature was calculated by assuming a temperature gradient of 0.016 °F/ft added to the assumed surface temperature of 60°F. So, at a depth of 12,500 ft, initial reservoir pressure is 9500 psia and is significantly higher than the bubble point pressure. Reservoir permeability and fracture half-length were obtained by conducting a history match on production data from a well in the Eagle Ford. The rate history along with the tubing head pressure (THP) for the well in the Eagle Ford condensate region was available for 678 days. THP was converted to bottomhole pressure using Gray's correlation for multiphase flow (Gray, 1978). The average producing CGR for the entire production period was 156 STB/MMScf with several fluctuations during the production period. A plausible

Eagle Ford composition with CGR of 150 STB/MMscf was selected and was tuned to match the predicted liquid API (47.5°), saturation pressure(4729 psia) and CGR (150 STB/MMScf) to the field observed data by using Peng-Robinson Equation of State (EOS). Matrix porosity and permeability is homogeneous in each layer. Table 3 lists some of the parameters for the reservoir.

Table 3. Reservoir properties for base case simulation

Properties	Value	Unit
Initial Reservoir Pressure	9500	psi
Reservoir Permeability	562.5	nd
Reservoir Temperature	250	°F
Reservoir Porosity	7	%
Initial Water Saturation	20	%
Total Compressibility	1x10-6	psi-1
Horizontal Well Length(Simulated)	5,280 (210)	ft
Well Spacing(Number of Wells)	1,320 (4)	ft
Fracture Half Length	265	ft
Number of Fractures	20	-
Minimum Bottomhole Pressure	1000	psia

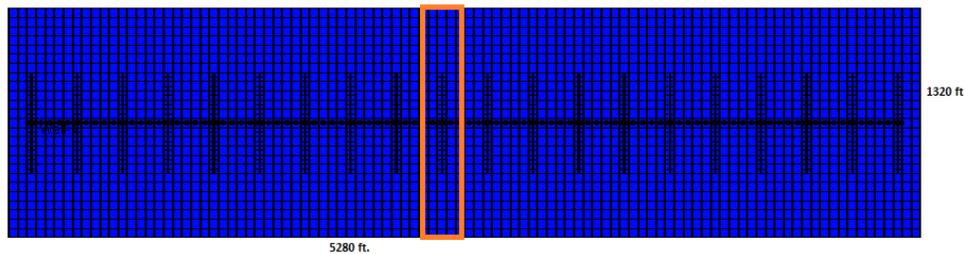


Figure 17. Schematic for base case reservoir simulation and symmetrical element

The rate and pressure history matches are shown in Figures 18 through

20:

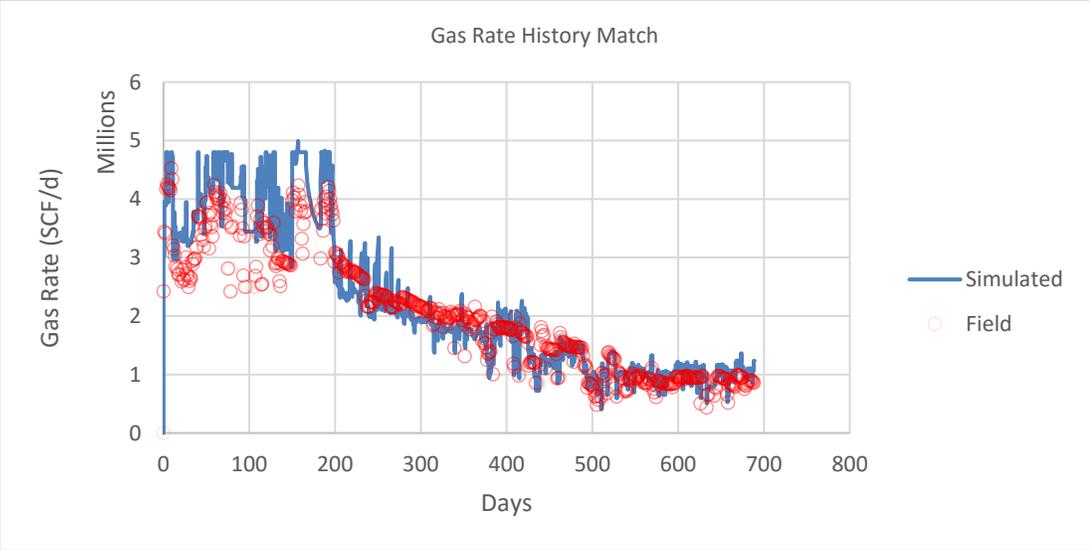


Figure 18. Gas history match for Eagle Ford Shale

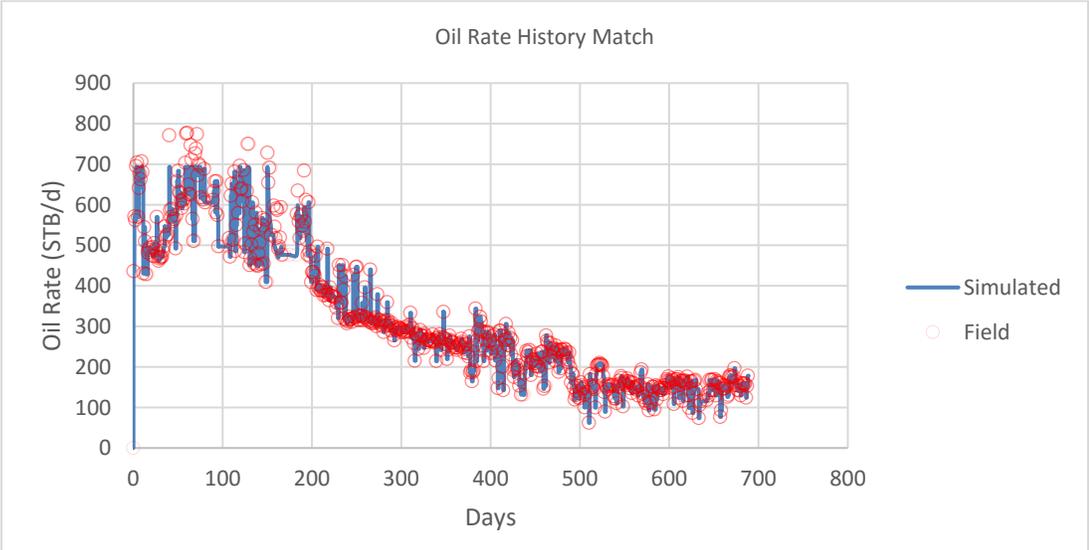


Figure 19. Oil history match for Eagle Ford Shale

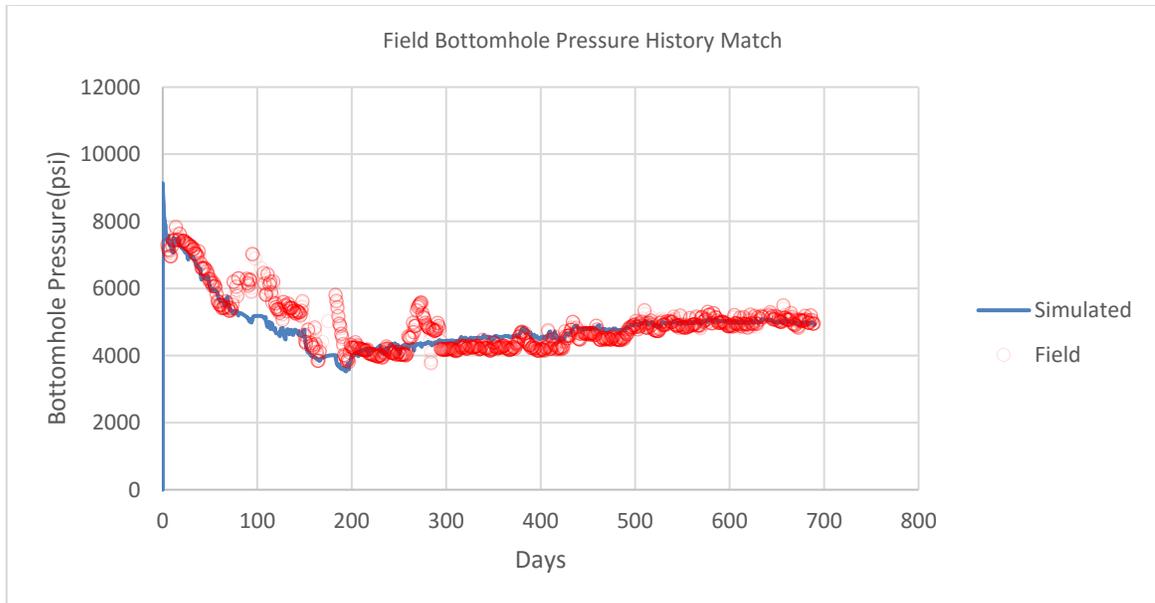


Figure 20. Pressure history match for Eagle Ford Shale

Peng Robinson EOS parameters for the base fluid sample are given in Table 4:

Table 4. Peng-Robinson EOS parameters for the base case fluid sample

Comp.	Mole Frac.	Pc (atm)	Tc (K)	Molecular Wt.	Acentric Factor
CO2	0.027	72.80	304.20	44.01	0.23
N2	0.002	33.50	126.20	28.01	0.04
CH4	0.679	45.40	190.60	16.04	0.01
C2H6	0.087	48.20	305.40	30.07	0.10
C3H8	0.049	41.90	369.80	44.10	0.15
IC4	0.011	36.00	408.10	58.12	0.18
NC4	0.019	37.50	425.20	58.12	0.19
IC5	0.009	33.40	460.40	72.15	0.23
NC5	0.009	33.30	469.60	72.15	0.25
FC6	0.013	32.46	507.50	86.00	0.28
C07-C09	0.046	28.62	567.08	108.40	0.31
C10-C12	0.024	22.67	642.89	150.48	0.43
C13-C14	0.009	19.48	693.67	186.83	0.52
C15-C17	0.008	17.36	733.17	220.61	0.60
C18+	0.008	13.91	814.23	307.29	0.79

The EOS parameters were generated for each of 5 fluid samples with different CGR. These parameters were then used to generate the pressure

temperature (PT) phase diagrams and the liquid dropout curves from two-phase flash calculations for each case as shown in Figures 21 and 22.

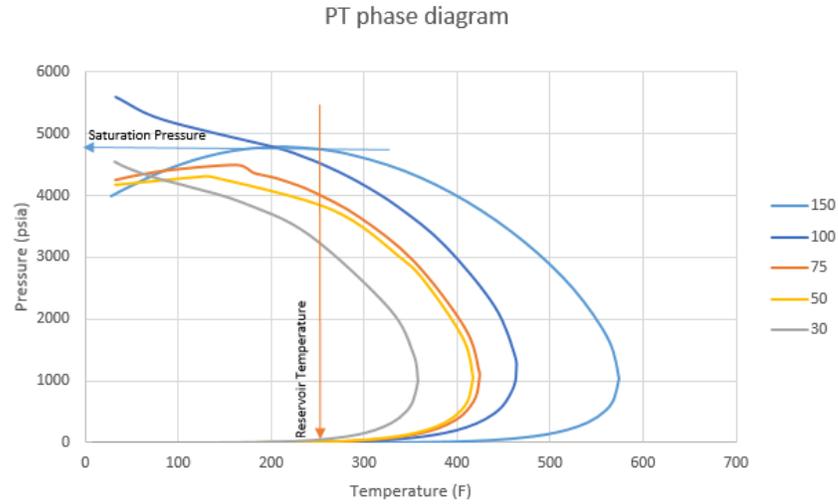


Figure 21. Pressure-Temperature (PT) diagram for sample fluids

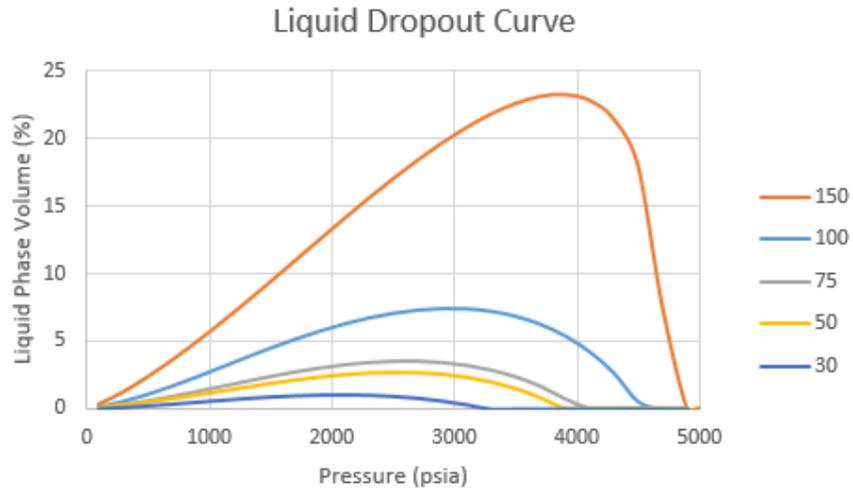


Figure 22. Liquid dropout curve for

Separate sets of relative permeability curves were used for the matrix and hydraulic fractures. Endpoints and Corey exponents for the following curves (Figure 23) were obtained from the literature (Nagarajan, 2013).

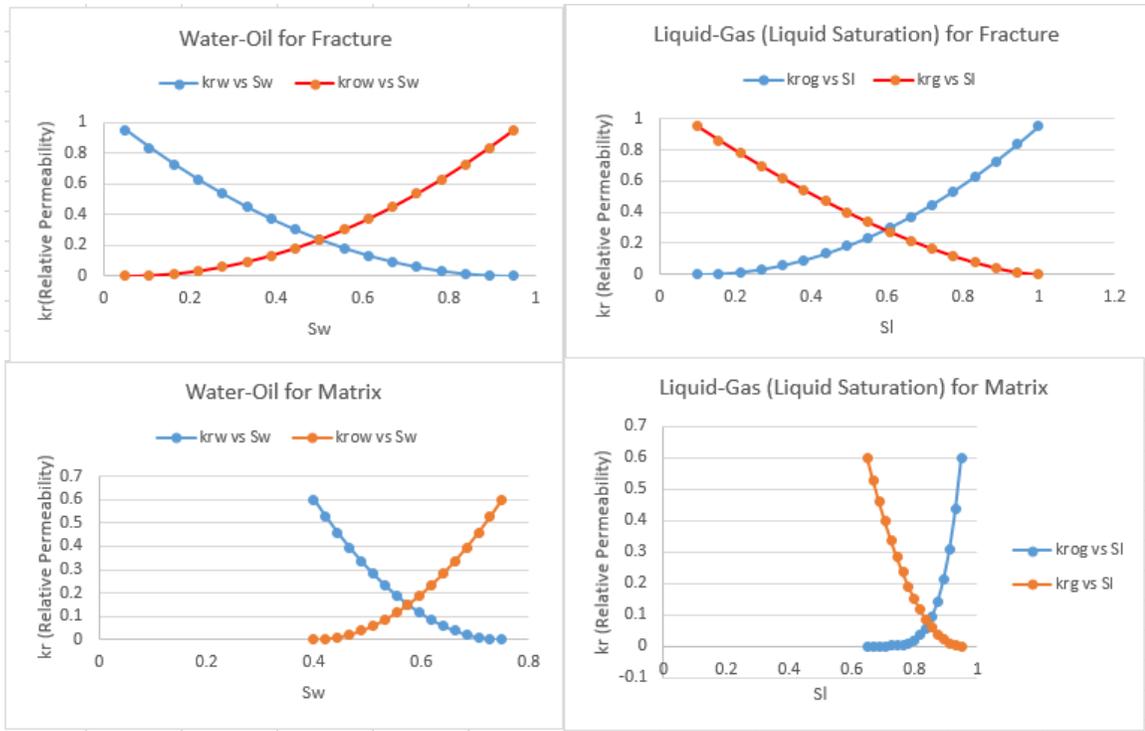


Figure 23. Relative permeability curves

3.2 Base Case Results

After history matching and parameter identification, the model was run in the prediction mode with the minimum bottomhole pressure constraint of 1000 psia for 30 years. The evolution of pressure with time for the sample with CGR of 150 STB/MMScf is shown in Figure 24:

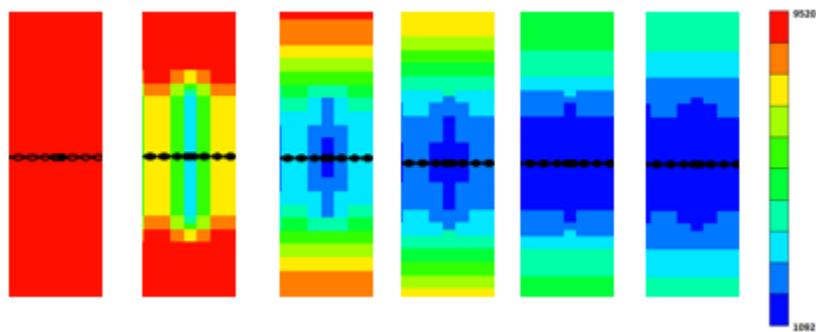


Figure 24. Evolution of pressure with depletion

As shown in Figure 24, pressure drops rapidly close to the fractures. The reservoir pressure outside of the SRV remains unaffected for up to 5 years. After 10 years of production, there is a slight drop in pressure: pressure at the outer boundary decreases to 8000 psia, compared to 1000 psia inside the SRV. After 20 years of production, the SRV seems to be completely depleted with a pressure of 1000 psia throughout. Similar trends were seen for all other fluid samples.

Figures 25 and 26 show the cumulative oil and gas production for each of the reservoir fluids. It can be seen that the oil production increases with increase in initial CGR, with 150 STB/MMScf yielding the maximum oil production. For the given relative permeability curves, the critical saturation of liquid for any flow in presence of gas is close to 75%. For the fluids with CGR 30 through 100, this critical saturation is not reached for majority of the grid blocks resulting in a condensate bank formation and trapped oil. The cumulative gas production curves show that maximum gas is produced from the leanest reservoir fluid. It is interesting to see that the fluid with intermediate CGR (100 STB/MMScf) has both oil and gas production greatly hindered due to liquid dropout and reduced permeability for gas.

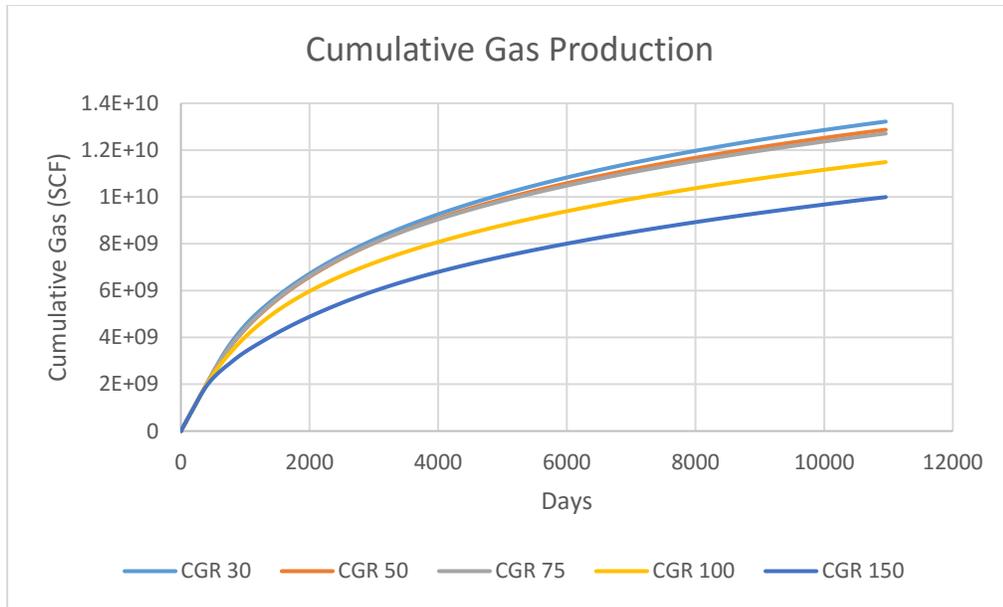


Figure 25. Cumulative gas production for each fluid samples

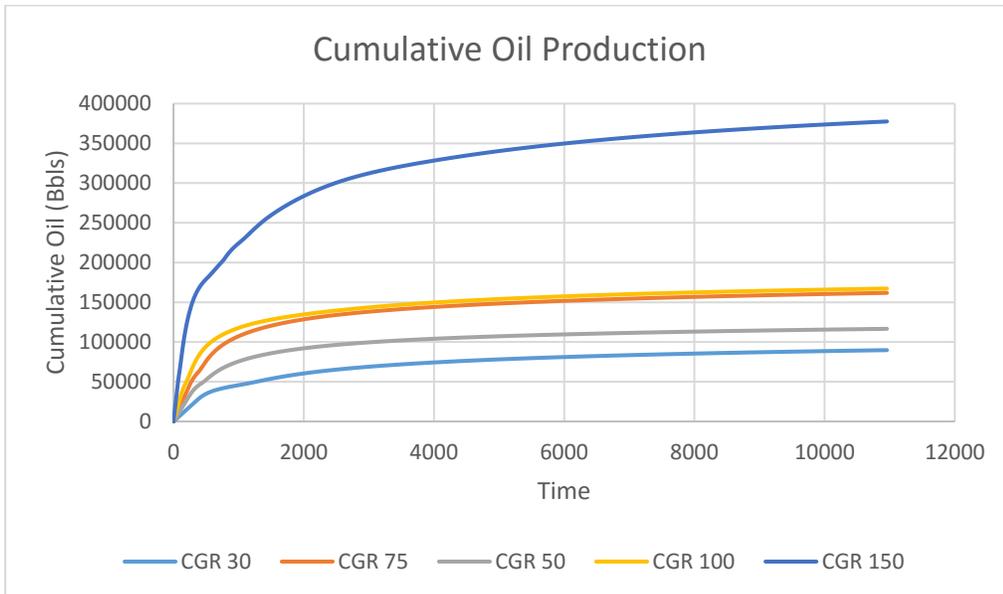


Figure 26. Cumulative oil production for each fluid samples

The oil saturation of a fracture grid and a SRV grid for three samples are shown in Figure 27. It can be seen that the oil saturation declines slowly in the fractures after a steep increase at the beginning of production. For a grid in a non-fractured region, the oil saturation is seen only after a significant period of

production. The rich condensate sample shows high oil saturation for both fracture and matrix. Oil saturation outside of SRV remains negligible for the entire production period.

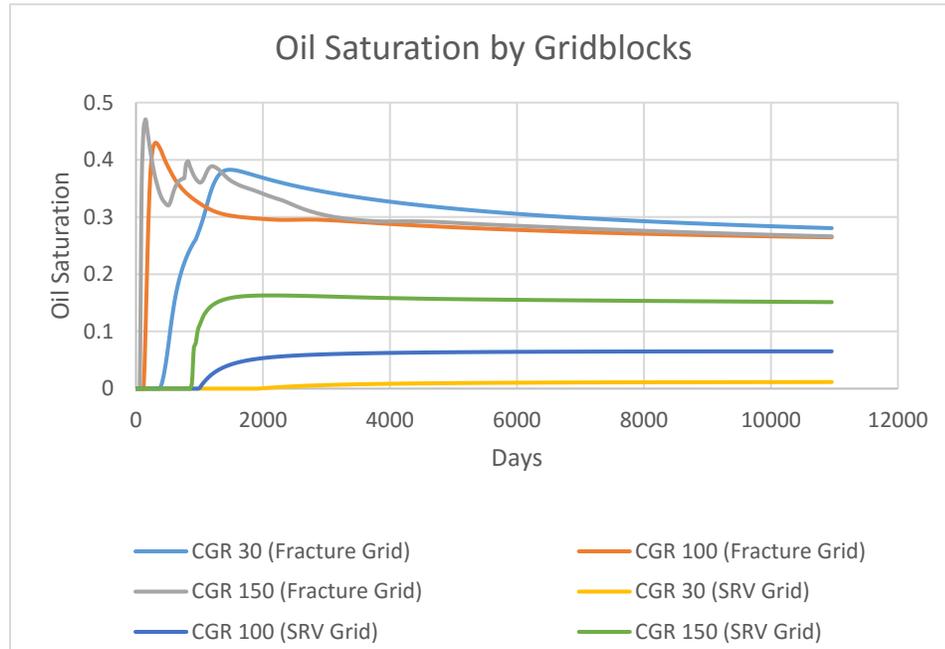


Figure 27. Oil saturation in fractures and matrix during production

3.3 Statistical Study

The production behavior and recovery of oil and gas from a low permeability gas condensate reservoir is a function of the rock, fluid and hydraulic fracture properties. Several parameters can affect the production, and most of these parameters have coupled effects in which one parameter could influence the impact of other parameters. The parameters studied in this work are explained briefly below.

Fracture Spacing

Fracture spacing, which determines the number of fractures, is a key parameter in a low permeability reservoir. The number of hydraulic fractures establishes the connected fracture surface area of the reservoir and is directly related to the production rate. Closer fracture spacing means more fracture stages and increased completion cost per well; however, it also results in greater cumulative oil and gas production.

Fracture Half-length

In shale reservoirs, fracture length and height determine the drainage volume of the reservoir. Most of the production, especially during the early life of the reservoir, comes from stimulated reservoir volume (SRV). This is contrary to conventional reservoirs where production is controlled mostly by well spacing. Since the ultimate oil and gas recovery in unconventional reservoirs is directly related to the SRV, fracture half-length is an important completion parameter.

Matrix Permeability

Matrix permeability determines the ability of hydrocarbons to flow into the fractures and thus to the wellbore. Permeability in unconventional reservoir range have high degree of uncertainty due to heterogeneities. As a result, understanding of reservoir permeability in these low permeability reservoirs is very important.

Fracture Conductivity

Fracture conductivity determines the ability of produced fluid to flow from fracture to wellbore. The problems of fracture embedment and multiphase flow can reduce the fracture conductivity and dimensionless conductivity to a value an order

of magnitude lower than the values determined in the lab. Fracture conductivity in this study was varied by changing the fracture permeability when all other parameters are kept constant.

Minimum BHP and Maximum gas rate

For gas condensate wells, the way the well is produced plays a vital role in liquid dropout behavior. A high maximum gas rate and low minimum BHP pressure yields a high cumulative EUR for gas; however, liquid dropout is increased. Higher drawdown drops the pressure in the reservoir below the dew point more quickly and the resulting liquid dropout can adversely affect the production. The Eagle Ford shale is over-pressured and most of the liquid-rich areas have initial pressures above the saturation pressure.

Well Placement

Well placement is another important parameter influencing the productivity of liquid rich shale wells. This requires a more detailed explanation, and is presented in a separate chapter.

Parameter Bounds

There is a large variability in each of the parameters depending on the regions within each play. The range of parameters was selected by choosing representative values from the literature (Orangi, 2011; Gong, 2013; Elamin, 2013; Nagarajan, 2013; Kumar, 2013). 52 one-parameter-at-a-time experiments (OPAAT) were conducted for each parameter. To confirm with the history match and to maintain simplicity, the reservoir fluid with an initial CGR of 150 STB/MMScf

was used for each case. The range of values used for the experiments are listed in Table 5.

Table 5 Parameter bounds selected for sensitivity study

Factor	Units	Minimum	Maximum
Fracture Spacing	ft	80 (66 fracs)	500 (11 fracs)
Fracture Half-Length	ft	100	450
Matrix Permeability	md	1e-3	1e-2
Fracture Conductivity	md	1,000	15,000
Minimum BHP	psia	750	2,000
Gas Rate	ft ³ /d	1,000,000	9,000,000
Matrix Porosity	fraction	0.06	0.1

3.4 Results

The cumulative oil and cumulative gas after 30 years of production for the mean case (average of minimum and maximum parameters) was 238 MSTB and 8,600 MMScf respectively. The effect on the cumulative oil and gas when one parameter at a time is changed is shown in Figure 28.

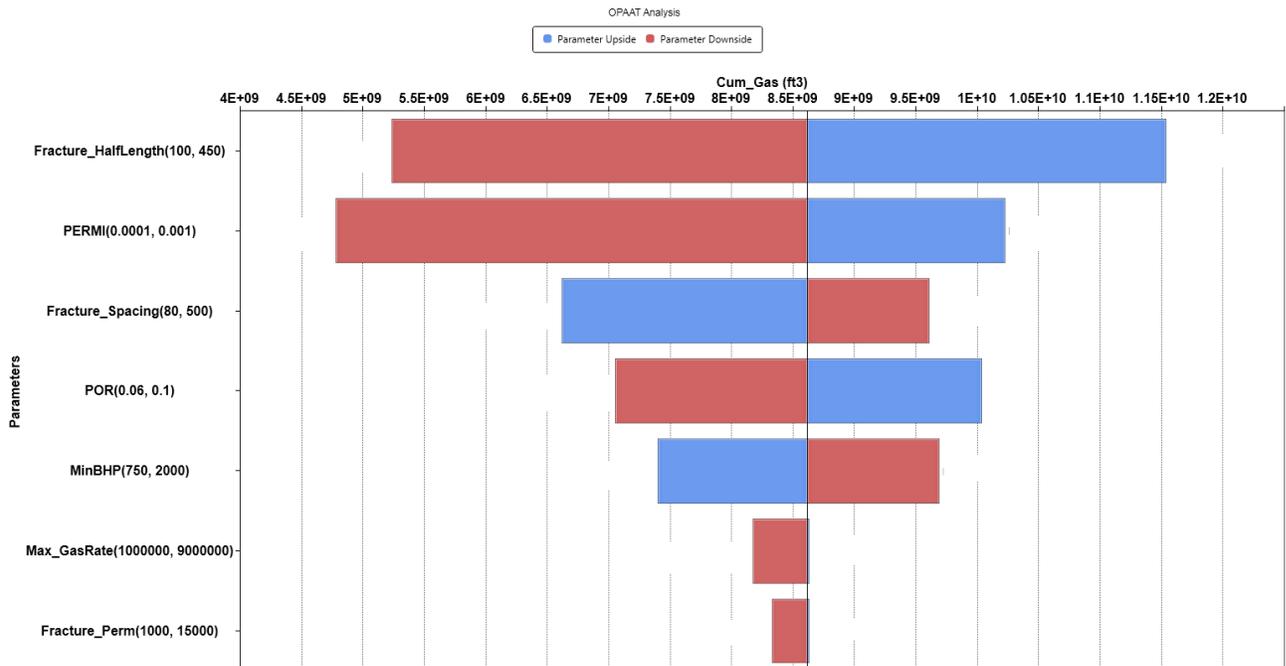


Figure 28. One parameter at a time sensitivity plot for cumulative gas

The tornado plot above shows that the fracture half-length is the most important parameter that affects the ultimate gas production, followed by matrix permeability. This result further indicates that the production of oil and gas from a liquid rich shale depends on the size of SRV (stimulated reservoir volume) and matrix permeability. Closer fracture spacing results in higher cumulative gas production due to the higher number of fractures. However, Figure 28 shows that the cumulative gas production does not increase proportionately from the mean value for the minimum fracture spacing case. This indicates that there is fracture interference which results in less than optimal return on investment. It can also be seen that maximum gas rate and fracture permeability have a relatively low impact on final cumulative gas production. This is because for the mean values of parameters, the reservoir is drained to the maximum extent possible in 30 years. Thus, these factors only affect the initial rate of production but not the final cumulative production from the reservoir.

Similarly, Figure 29 shows that fracture half-length is again the most important parameter for cumulative oil production followed by fracture spacing. However, unlike cumulative gas production, maximum gas rate has a significant impact on the final cumulative oil production from a reservoir. Figure 29 also shows that a lower rate of production results in higher cumulative oil production. This is because lower production rates result in sustained single-phase flow and lower overall liquid dropout. This can also be seen from the result of minimum bottom-hole pressure, where higher minimum bottom-hole pressure results in higher cumulative oil production contrary to cumulative gas production case. Finally,

increased fracture conductivity results in rapid pressure loss and more liquid dropout resulting in lower cumulative oil production. Porosity, which determines the overall hydrocarbon in place, shows a linear impact on both cumulative oil and gas production.



Figure 29. One parameter at a time sensitivity plot for cumulative oil

3.5 Study of Combined Effects

The simulations involving OPAAT showed that there is a correlation of cumulative production from a reservoir with the reservoir and completion properties. Design of experiment (DOE) and response surface modeling (RSM) was used in addition to reservoir simulation to find this correlation. Response surface methodology (RSM) explores the relationships between input variables (parameters) and responses (objective functions). A set of designed experiments is used to build a proxy model (approximation) of the reservoir objective function. The most common proxy models take either a linear or quadratic form (CMOST,

2013). More details on theory of DOE and RSM can be found in Box et al. (2005) and Myers and Montgomery (2002).

For analysis, a set of random experiments (49 in total) were generated by varying the parameters within the bounds specified above. A proxy model was created by using polynomial regression and significance probability alpha of 0.05. A statistically significant result is obtained when global p-value is less than the significance level. The p-value is the probability of obtaining at least as extreme results given that the null hypothesis is true whereas the significance or alpha level is the probability of rejecting the null hypothesis given that it is true (Schlotzhauer and Sandra, 2007). Figures 31 and 32 show that the proxy model created from RSM is very accurate in predicting the final cumulative oil and gas production as shown by all the data points lying close to the 45° line. The R-square value for cumulative gas and cumulative oil was 0.97 and 0.98 respectively suggesting a good match. ANOVA test showed that Probability > F is less than 0.00001 for both cases suggesting the presence of significant effects.

All of the insignificant parameters, whose coefficients are most likely zero, were removed by using the t-statistic from regression analysis.

Following this step, tornado plots were prepared for both cumulative gas and cumulative oil by removing the insignificant terms. Figures 33 and 34 show that fracture half-length is the most significant parameter for both cumulative gas production and cumulative oil production. For cumulative gas production, no interactions seem to have significant effect on the final result. However, for cumulative oil, fracture half-length, fracture spacing and maximum gas rate interact

with each other and have a combined effect on the final result. These interaction effects are more important than some of the independent parameters like matrix permeability in their impact on cumulative oil production.

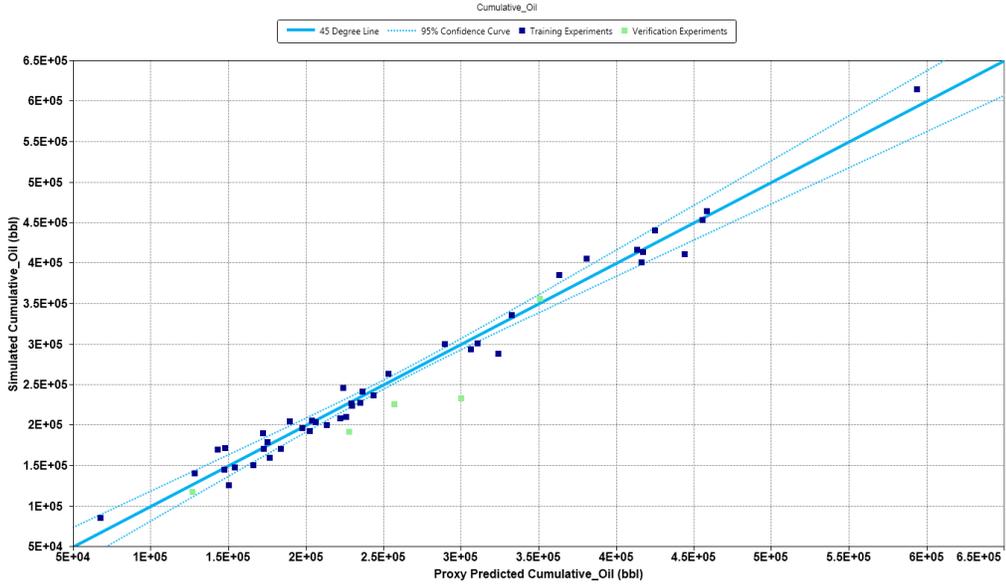


Figure 31. Verification for cumulative gas from proxy model

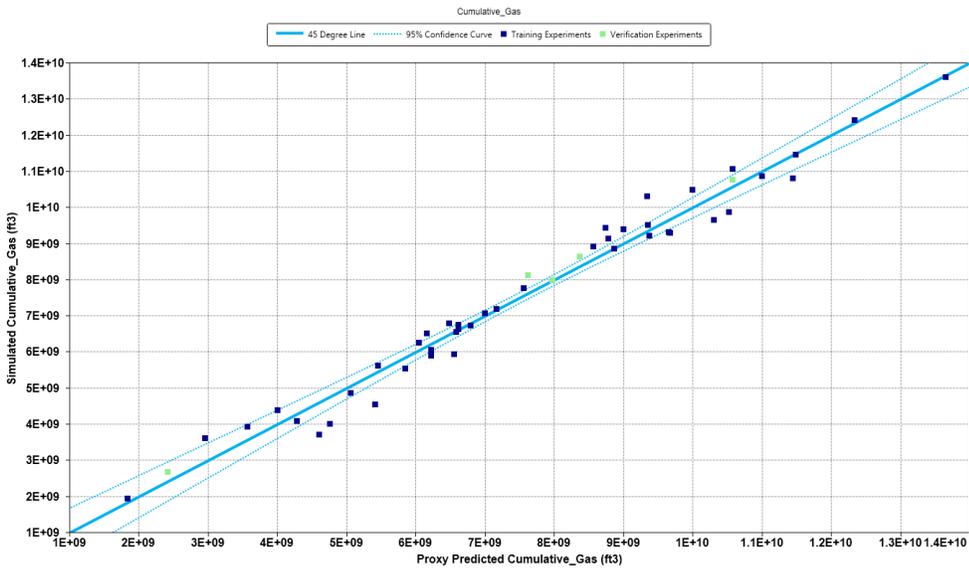


Figure 32. Verification for cumulative oil from proxy model

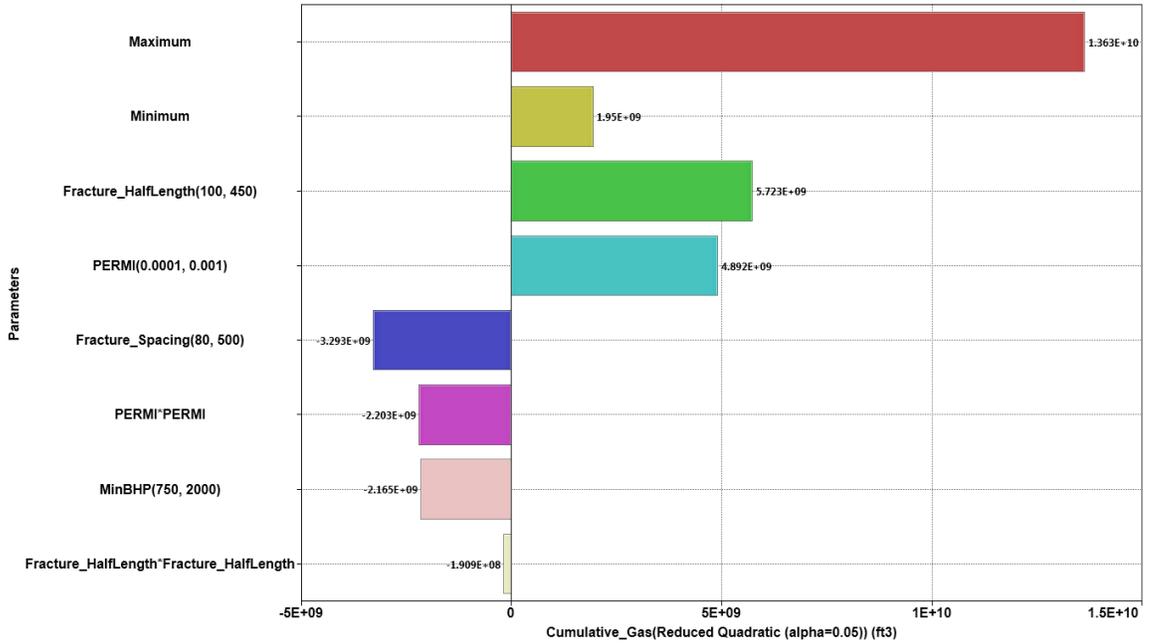


Figure 33. Sensitivity analysis to study the interaction of parameters with each other

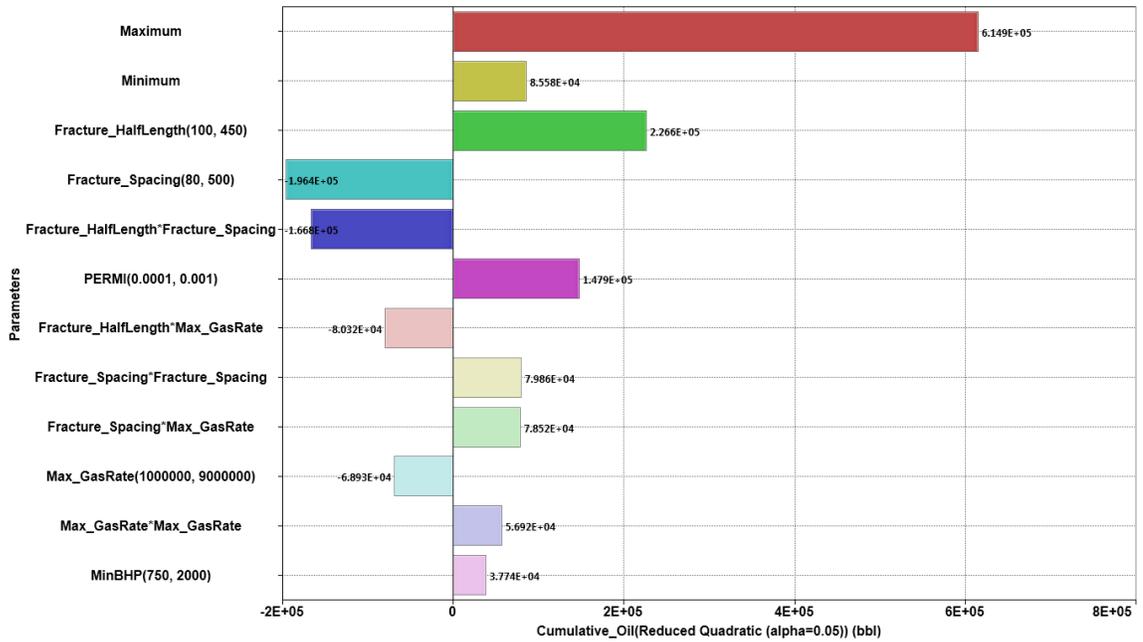


Figure 34. Sensitivity analysis to study the interaction of parameters

The equations generated from response surface modelling (RSM) and their coefficients are:

$$\begin{aligned} \text{Cumulative_Gas} = & 3.6E09 + 1.8E07 * \text{Fracture_HalfLength} - 7.8E06 * \text{Fracture_Spacing} - \\ & 1.7E06 * \text{MinBHP} + 1.1E13 * \text{PERMI} - 3117 * \text{Fracture_HalfLength} * \text{Fracture_HalfLength} - \\ & 5.4E15 * \text{PERMI}^2 \end{aligned} \quad (3.2)$$

$$\begin{aligned} \text{Cumulative_Oil} = & 40479 + 1592 * \text{Fracture_HalfLength} - 602 * \text{Fracture_Spacing} - 0.024 * \\ & \text{Max_GasRate} + 30 * \text{MinBHP} + 1.6E08 * \text{PERMI} - 2.27 * \text{Fracture_HalfLength} * \\ & \text{Fracture_Spacing} - 5.7E - 05 * \text{Fracture_HalfLength} * \text{Max_GasRate} + 0.90 * \\ & \text{Fracture_Spacing}^2 + 4.7E - 05 * \text{Fracture_Spacing} * \text{Max_GasRate} + 1.8E - 09 * \text{Max_GasRate}^2 \end{aligned} \quad (3.3)$$

3.6 Observations and Conclusions

From this study, we drew the following conclusions about production from gas-condensate liquid-rich shale reservoirs:

- Ultimate oil recovery is highly dependent on the initial condensate gas oil ratio.
- Ultimate gas recovery is most affected by fracture half-length, matrix permeability, fracture spacing and minimum bottomhole pressure.
- Ultimate oil recovery is affected by fracture half length, fracture spacing, permeability and interaction of fracture half length, minimum bottomhole pressure and fracture spacing.
- Interaction of different parameters affect the final cumulative oil and gas production from a LRS reservoir.
- Reduced gas production rates which allow a sustained period of single-phase flow lead to a significant increase in ultimate oil recovery but do not affect ultimate gas recovery.
- DOE and RSM using a proxy model can be used to determine the ultimate

recovery from LRS gas condensate reservoirs with reasonable accuracy, and are thus useful in determining the effects of parameter interactions.

4. Effect of well spacing on productivity of liquid rich shale (LRS) reservoirs with multiphase flow: A simulation study

Unconventional oil and gas reservoirs have poor permeability and require multiple wells in the same section in addition to hydraulic fracturing to obtain an economic rate. Multi-fractured horizontal wells generate a large contact by opening the paths for flow of hydrocarbon into the wellbore. However, due to the large uncertainties in reservoir and fracture properties, well spacing needs to be optimized and clearly studied to minimize the drilling and the completion cost. Most of the hydrocarbons in unconventional reservoirs is obtained from stimulated reservoir volume (SRV) with little pressure drop outside of SRV (XSRV) (Ozkan, 2009). Thus, it is essential to optimize the number of wells in a reservoir to prevent overcapitalization and unsatisfactory returns on investment. Recently several authors (Yu, 2014; Lalehrokh, 2014; Sahai, 2012) have studied economic viability and well optimization for unconventional wells. Their results showed that overall well placement and fracture design strongly affects well economics. However, most of these studies are performed by using either decline curve/rate transient analysis or black oil reservoir simulations. Retrograde gas condensate and volatile oils exhibit multiphase flow and change in composition during production. Compositional reservoir simulation is a tool that accurately captures this phenomenon and can be easily used due to computational power available today. Moreover, most of these well spacing studies perform a sensitivity analysis of different formations for different permeability and fracture half length. This work will try to expand that by performing a sensitivity analysis on initial reservoir fluid,

which can have an important impact on the number of wells considered in a play. An informed investment in multi-well development of unconventional liquid rich resources can be attained to maximize reserves and profits instead of just accelerating the production. The goal of this chapter is to utilize the compositional reservoir simulation to understand a practical problem of optimal well spacing.

4.1 Workflow

Liquid rich shale (LRS) reservoirs are intrinsically heterogeneous and there are many uncertainties in reservoir and fluid properties. The statistical study of several reservoir parameters on production performance of LRS was performed in Chapter 3. It showed that various history matching parameters like permeability, porosity, well constraints have a big impact on final oil and gas production from a reservoir. For this study we select reservoir parameters based on field data and literature survey of the Eagle Ford shale play.

Following workflow was used for this study:

- A base case was modelled after the reservoir and fluid properties similar to the LRS reservoirs from literature review and field data.
- A compositional reservoir simulation was performed for the period of 30 years to obtain cumulative oil and gas for base case.
- Additional simulation was performed for each fluid case ranging from a very lean wet gas to black oil for different well spacing.
- A simple cash flow model was used to calculate the NPV of each fluid case for different well and compared.

4.2 Numerical Simulation

Compositional reservoir simulation is an effective method to simulate multiphase flow in volatile oil and retrograde condensate reservoirs, especially in early stage of field developments. In-situ fluid data can be obtained from separator tests accurately when the reservoir is still over the saturation pressure. In this work, CMG-GEM, a compositional simulator by Computer Modelling Group was used to model the reservoir with multiple hydraulic fractures and fluid flow in geology similar to liquid rich shale reservoirs. Hydraulic fractures are modeled by using logarithmically-spaced local grid refinement (LS-LGR), which captures the flow of fluid and change in pressure in the near-fracture region.

For the base case, a 640 acres reservoir with 4 horizontal wells and 25 stages of bi-wing hydraulic fractures was considered. Hence, the well spacing and the fracture spacing for the base case is 1320 ft. and 210 ft. respectively. A symmetric element with dimensions of 1320 ft., 210 ft. and 100 ft. for length breadth and thickness respectively was simulated to reduce the computational time. TVD for each well was assumed to be 10,000 ft. Dimensionless fracture conductivity of greater than 300 was used for each case to ensure no pressure loss in fractures. The reservoir properties used for the base case are presented in Table 6.

Table 6. Reservoir properties for base case simulation

Properties	Value	Unit
Initial Reservoir Pressure	7,750	psi
Reservoir Permeability	500	nd
Reservoir Temperature	250	°F
Reservoir Porosity	7	%
Initial Water Saturation	20	%
Total Compressibility	1x10 ⁻⁶	psi ⁻¹

Table 6. Continued...

Horizontal Well Length(Simulated)	5,280 (210)	ft.
Well Spacing(Number of Wells)	1,320 (4)	ft.
Fracture Spacing(Number of Fracs)	210 (25)	ft.

Table 7. Fluid Compositions for the selected liquid rich shale (LRS)

Comp.	Gas Condensates				Volatile Oils					
	CO2	2.89	2.86	2.87	2.78	2.71	2.58	2.47	2.32	1.95
N2	0.19	0.19	0.19	0.18	0.18	0.17	0.16	0.15	0.13	0.1
C1	72.4	71.6	71.8	69.7	67.9	64.7	61.8	58.1	48.9	37.9
C2	9.26	9.16	9.19	8.92	8.69	8.27	7.91	7.43	6.25	4.85
C3	5.18	5.13	5.14	4.99	4.86	4.63	4.42	4.16	3.5	2.71
I-C4	1.19	1.18	1.18	1.15	1.12	1.07	1.02	0.96	0.81	0.62
N-C4	2.04	2.01	2.02	1.96	1.91	1.82	1.74	1.63	1.37	1.07
I-C5	0.93	0.92	0.93	0.9	0.88	0.83	0.8	0.75	0.63	0.49
N-C5	1.00	0.99	0.99	0.96	0.94	0.89	0.86	0.8	0.68	0.52
C6	1.42	1.41	1.41	1.37	1.33	1.27	1.21	1.14	0.96	0.74
C7+	3.48	4.51	4.23	7.08	9.48	13.8	17.6	22.6	34.9	49.5

Several representative reservoir fluids ranging from volatile oil to gas condensate fluid were obtained from a literature survey (Whitson, 2012). The Peng-Robinson equation of state (EOS) was used to generate PVT parameters, which was tuned to match the available properties like saturation pressure, gas oil ratio (GOR) and API. This tuning was performed by using CMG's phase modeling software Winprop. For the base case, the leanest sample with OGR of 30 STB/MMScf was selected. The detailed fluid properties and PVT parameters for the base are shown in Tables 7 and 8.

Table 8. Fluid Properties for each sample

Properties	Gas Condensates					Volatile Oils				
Saturation Pressure(psi)	3119	3759	4022	4490	4729	4754	4572	4217	3265	2260
OGR, STB/MMScf	30	50	75	100	150	250	350	500	1000	2000
GOR, scf/STB	33333	20000	13333	10000	6667	4000	2857	2000	1000	500
° API	51.4	49.8	48.4	47.5	46.2	44.5	43.4	42.3	40	37.7
C7+ Mol Wt	123	132	138	145	153	164	171	178	195	216
SG C7+	0.76	0.77	0.78	0.78	0.79	0.79	0.80	0.80	0.81	0.82

The samples ranged from lean gas condensate with OGR (oil gas ratio) of 30 STB/MMScf to volatile oils with OGR of 2,000 STB/MMScf. The pressure temperature phase diagram (PT-diagram) showed that some of these fluids with OGR of 100 STB/MMScf to 350 STB/MMScf lie extremely close to the critical temperature and may act like a near-critical fluid. The OGR for each case was matched to the separator test at 400 psia and 100°F. Each reservoir is under-saturated for an initial reservoir pressure of 7750 psi. The C₇⁺ fraction is split into 3 to 5 pseudo-components before generating the Peng-Robinson EOS parameters.

Separate sets of relative permeability curves were used for the matrix and hydraulic fractures. Separate relative permeability curves were used depending on the initial phase of fluid in the reservoir. Endpoints and Corey exponents for the following curves were obtained from a literature survey (Nagarajan, 2013). The relative permeability curves are shown in Chapter 3 (Figure 23).

4.3 Economic Analysis

Well drilling and completions are among the costliest factors in a project and require proper optimization for higher returns on investment. Net present value (NPV) is the objective function used for optimization of well spacing in this work and is calculated by

$$NPV = \sum \left(\frac{R}{(1+i)^t} \right) - CAPEX_0, \quad (4.1)$$

where R is the net cash inflow discounted to the current price by using a discount rate i, t is the number of years, CAPEX₀ is the capital expenditure like drilling and completions, land acquisition, seismic acquisition, roads, bridges at time zero (not discounted). For simplicity, it is assumed that all the wells are drilled at the same time. The net cash inflow, R, is calculated by subtracting the operating expenditures(OPEX) like labor, taxes, repairs and maintenance etc., taxes (severance, ad valorem) from oil and gas revenues. Federal income tax is one of the big cash outflow that needs to be accounted for the most accurate picture of NPV. However, for simplicity all the calculations in this paper have been performed before federal income tax (BFIT).Pioneer Natural Resources reported that the cost of drilling and completing a well in the Eagle Ford Shale was about \$8 million in 2013. In this work, \$7 million is assumed, a decrease in cost due to maturity of the techniques. The gas price of \$3.5/MSCF, the oil price of \$60 per barrel, tax expenditure (severance, ad valorem etc.) of 10%, OPEX of 10% of the revenue was used to calculate the net cash inflow, R, which is discounted by a discount factor of 10%.

4.4 Results for Base Case

The base case reservoir was simulated for each fluid sample for 30 years using the well constraints of 1000 Mscf/day followed by a minimum bottomhole pressure of 1000 psia. The change in pressure with time is shown Figure 35 below for the fluid with OGR of 30 STB/MMScf.

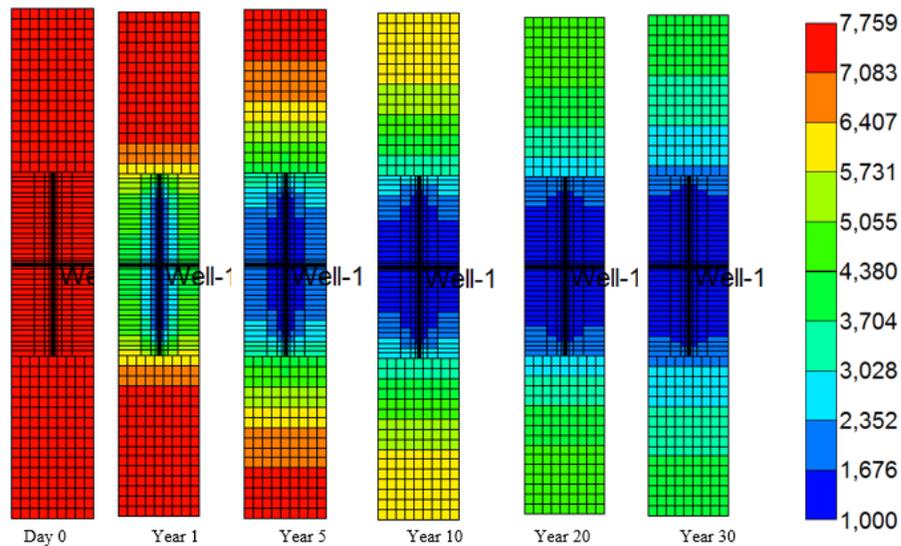


Figure 35. Evolution of pressure with time due to depletion

Figure 35 shows that pressure transient moves very rapidly in the SRV. Within the first 5 years, grid-blocks around the fractures deplete to the minimum bottomhole pressure. Figure 35 also shows that the outer reservoir boundary remains at the initial pressure even after 5 years of production. Finally, after 30 years of production, the outer boundary of the XSRV drops to around 5000 psia, which shows that a tighter well spacing would be beneficial in this case.

Figure 36-37 shows the base case cumulative oil and gas production for each sample. It can be seen from the first graph that cumulative gas for lean samples would benefit from added wells but not so much for rich samples.

However, from the second graph it can be seen that cumulative oil for rich gas has a big potential to increase with added wells. The impact of different initial fluid type on the production performance of oil and gas is shown in Figure 36-37.

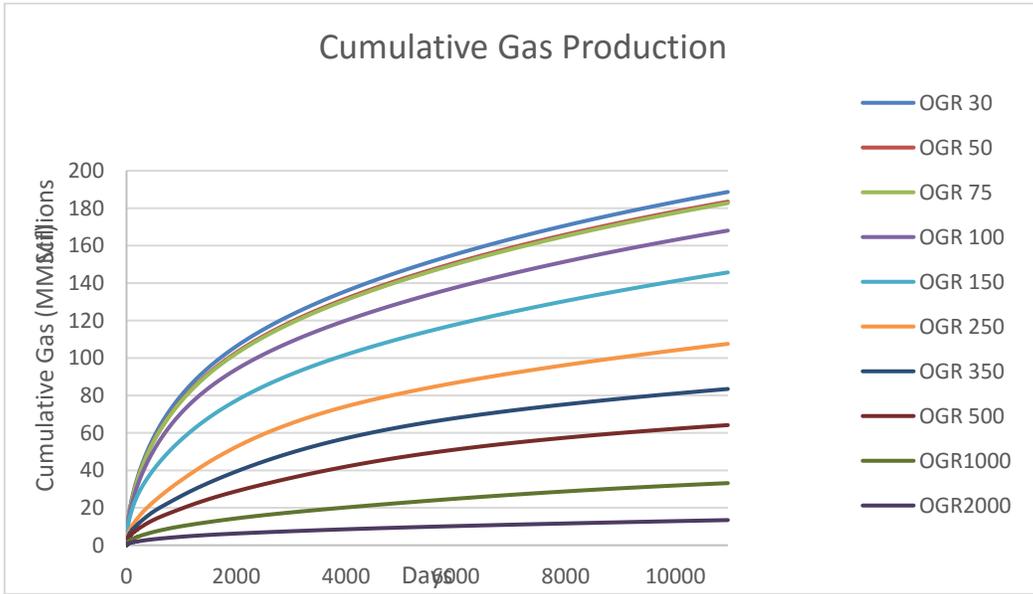


Figure 36. Cumulative gas production for each fluid sample

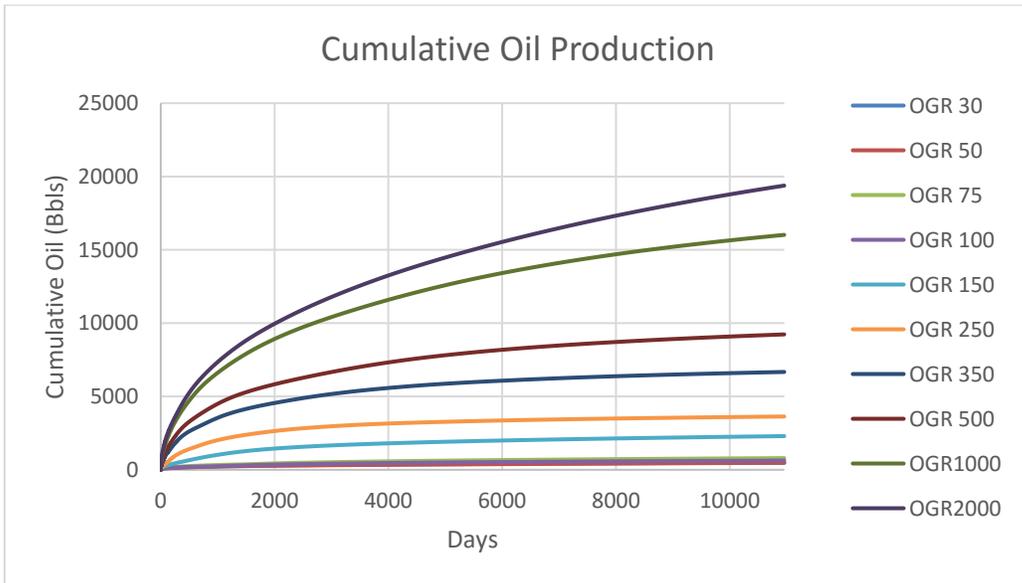


Figure 37. Cumulative oil production for each fluid sample

Based on the cumulative oil and gas produced for different fluid types in the base case, NPV is calculated using Equation 1 for 5 years, 10 years and 30 years. Figure 38 shows the NPV for different fluid types and it can be seen that the maximum NPV is obtained for samples with high OGRs. The NPV increases significantly beyond the OGR of 500 STB/MMScf. A peculiar trend is seen for the samples with OGR beyond 75 STB/MMScf where NPV remains negative even after 30 years of production. Moreover, the least NPV is seen for the case with OGR of 250 STB/MMScf. Although richer than all the samples up to 150 STB/MMScf, heavy liquid dropout and condensate banking seems to hinder the production in the samples close to critical point.

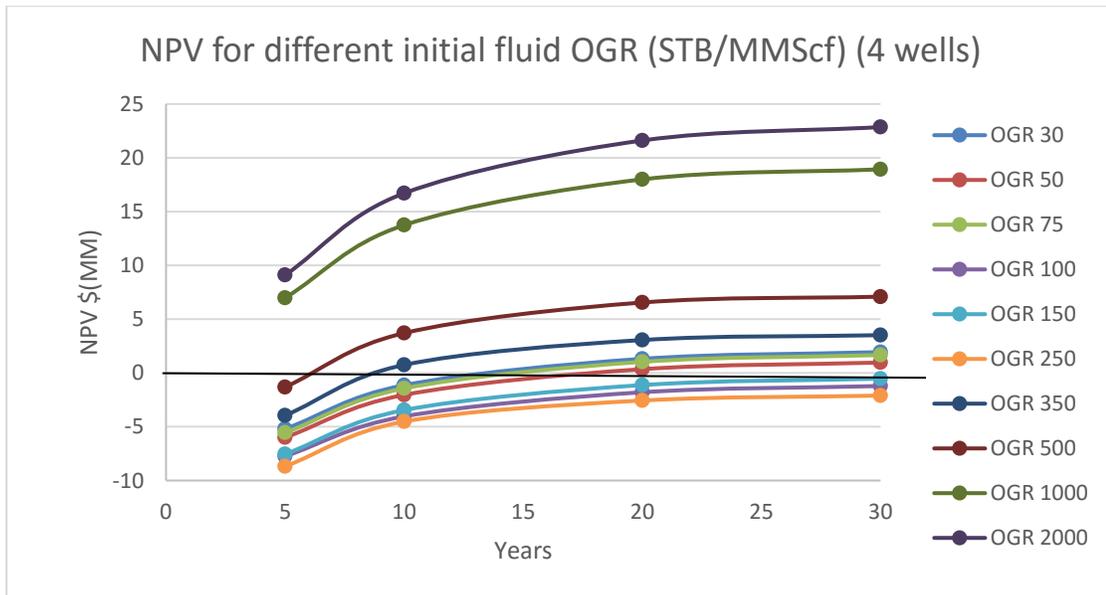


Figure 38. NPV for different fluids for base case

4.5 Sensitivity Study

Based on the information above from the base case, a sensitivity study was performed to identify the optimal well spacing for each sample. The purpose

is to identify the optimal number of horizontal wells per section based on the fluid type. In this work, we simulated 4 to 10 wells in 640 acres. This translates to well spacing of 1320 ft. to 528 ft. respectively. In total, 70 simulation runs were carried out including the base case, for each of the 10 fluid and 7 well spacing.

The NPV for each of the 70 cases was calculated and analyzed based on the assumptions given above. For the very lean samples, with OGR of 30 STB/MMScf to 75 STB/MMScf, the 5 year NPV remains negative for all well spacing. Positive return on investment is only seen after 10 years of production. It was also seen that after 10 years, 5 wells or a well spacing of 1,056 ft. provide the maximum NPV. Compared to the NPV at 20 years, NPV at 30 years grows by 17% of NPV at 20 years. Figure 39 shows the NPV by year for the sample with OGR of 30 STB/MMScf.

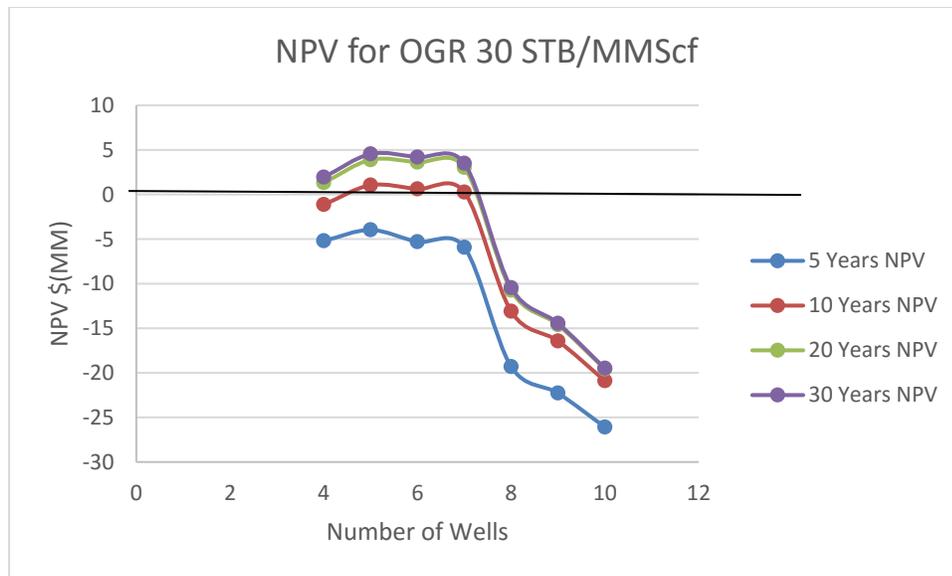


Figure 39. NPV for lean gas condensate sample

For samples in and around the critical region from 100 STB/MMScf to 250 STB/MMScf, both cumulative oil and gas production were dramatically small. This

led to NPV being negative for the entire production period. This could be due to liquid dropout and condensate banking, which reduces both oil and gas production from a reservoir mainly due to relative permeability effects. For this fluid type, natural depletion might not be appropriate and efforts should be taken to re-vaporize the liquid dropout. This could increase the mobility of the fluid and lead to higher production and consequently higher NPV. Figure 40 shows the NPV by years for the sample with OGR of 100 STB/MMScf.

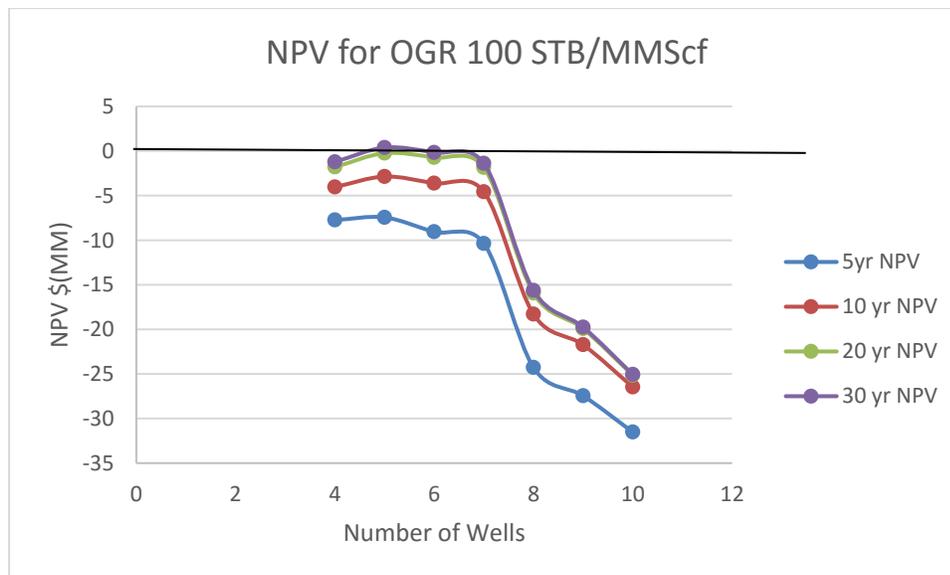


Figure 40. NPV for wells with fluid around critical point

For samples with rich oil content, from OGR of 350 STB/MMScf to 2000 STB/MMScf, positive NPV is seen as early as 5 years (OGR 1000 and 2000 STB/MMScf) even for 4 wells per section. All the samples with and beyond the OGR of 350 SCF/MMStb show positive NPV after 10 years of production. As seen from the base case, the maximum NPV is seen for the case with high OGR. Unlike the lean samples, for the rich samples the optimal well spacing was 7 wells per

section as it generated the maximum NPV. Figure 41 shows the NPV for the richest sample with OGR of 2000 STB/MMScf.

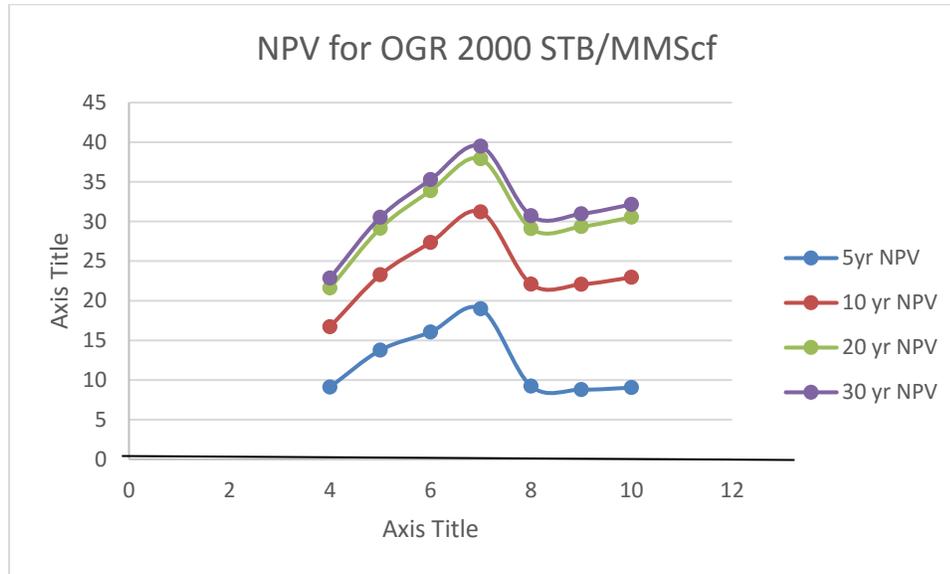


Figure 41. NPV for rich fluid sample

4.6 Conclusions

Several studies have been performed in the past to understand the effect of permeability and fracture half-length on optimal well spacing. These studies show that, for a given hydraulic fracture geometry, more wells are required to develop a shale reservoir with low matrix permeability than high matrix permeability. However, the impact of initial fluid types are usually ignored while performing the analysis. This study shows the impact of initial fluid composition on well spacing in liquid rich oil and gas shale reservoirs. The following conclusions can be drawn from this work:

- Wells can be placed closer to each other in liquid rich areas. The optimal number of wells for samples with OGR of greater than 350 STB/MMScf was found to be 7.

- For lean samples, the optimal number of wells was found to be 5. This shows that in the areas with little liquid saturation, it is advisable to drill fewer wells.
- For critical fluids, liquid dropout and condensate banking has a huge impact on the final production. Secondary methods should be employed to make this fluid region profitable at the current commodity prices.
- Most of the production in liquid rich shale is obtained in the first 20 years. Final 10 years for production has a negligible impact on the overall NPV of a project.

5. Optimal decline curve analysis (DCA) models for liquid rich shale (LRS) reservoirs

A decline curve fitting method presented by Arps (1945) is the most widely used method for both conventional and unconventional wells. This is an empirical method based on curve fitting and it is not based on fundamentals of fluid flow in porous media. For this method to be applicable, a reservoir should be in boundary dominated flow (BDF) with constant drainage area and constant flowing bottomhole pressure. Unconventional reservoirs exhibit long transient linear flow periods which could last for years because of their low matrix permeability ($10E-9$ md $< k < 10E-3$ md). In extreme cases, the wells could still be in the transient linear flow regime when they reach their economic limits. Usually, however, transient linear flow is followed by boundary dominated flow due to intra-fracture pressure interference, followed by a transition flow regime and finally by BDF influenced by the well's drainage boundaries. This paper evaluates a multi-segment or hybrid approach in which the transient (usually linear) flow regime is modeled by either the Duong or Stretched Exponential Production Decline (SEPD) models followed by Arps' hyperbolic model for BDF. Condensate production is forecasted by using a method developed by Yu (2014).

5.1 Background

Decline curve analysis has been used successfully for many traditional black oil and dry gas reservoirs in the past to forecast future production. However, these techniques might be problematic for gas condensate reservoirs due to multi-

phase flow in the reservoir below the dew point pressure. A brief review of the production data analysis used in this report is given below.

5.2 Arps Decline Model

The production decline method presented by Arps (1945) is the most widely used procedure in the industry to forecast the EUR for both oil and gas reservoirs. This method is applicable only after the onset of BDF. The most common Arps' model, the hyperbolic decline model is

$$q = \frac{q_i}{(1+bDt)^{\frac{1}{b}}}. \quad (5.1)$$

where q is the production rate at time t , q_i is the initial production rate, b is the hyperbolic decline parameter ($0 < b < 1$) and D is the decline rate defined as follows:

$$D = -\frac{dq}{dt} \times \frac{1}{q}. \quad (5.2)$$

For exponential decline ($b=0$):

$$q = q_i \exp[-D_i t]. \quad (5.3)$$

For harmonic decline ($b=1$):

$$q = q_i \frac{1}{(1+D_i t)}. \quad (5.4)$$

Unconventional reservoirs, which often have permeability in the range of hundreds of nano-darcy, exhibit long-duration transient flow. Thus during the early transient flow regime, the Arps hyperbolic decline model is likely to be unsuitable. An Arps' decline constant, b , greater than one causes forecasted cumulative production to increase without limit as time increases; thus, the forecast is

unbounded (Lee and Sidle, 2010). However, decline constants greater than 1 can be good matches with different transient flow regimes (e.g., $b = 2$ for transient linear flow, or 4 for transient bilinear flow) (Spivey, 2001). When using a b value greater than 1 for transient data, a common approach is to switch to a minimum terminal decline rate (D_{\min}) based on intuition or analogy to more accurately model BDF, and thus force an exponential decline with a constant value of D_{\min} for the remaining life of the well (Lee and Sidle, 2010).

5.3 Duong's Production Decline

Duong (2011) proposed a new production decline technique for long-duration transient linear (or near linear) or bilinear flow. A log-log plot of the ratio of rate and cumulative production with time shows a straight line for the wells in the transient flow regime. Duong's equations are

$$q = q_1 t^{-m}, \quad (5.5)$$

$$\frac{Q}{q} = \left(\frac{1}{a}\right) t^m, \quad (5.6)$$

$$t_m = t^{-m} \exp\left(\frac{a}{1-m} (t^{1-m} - 1)\right) \text{ and} \quad (5.7)$$

$$q = q_i t_m + q_{\text{inf}}, \quad (5.8)$$

where q is the production rate at time t , q_1 is the rate at $t=1$, a and m are empirical constants, t_m is "modified" time, Q is cumulative production and q_{inf} is the intercept of a plot of q vs. t_m .

The parameters in Duong's equations can be determined in two steps: the first step is plotting ratio of production rate, q , and cumulative production, Q , vs. t on a log-log coordinates. The parameters a and m can be obtained from intercept

and slope respectively. The second step is to plot modified time vs rate, to solve for q_1 . The intercept of this plot (q_{int}) should be forced to zero; otherwise, it is inconsistent with Equation 5.5.

5.4 Stretched Exponential Production Decline

The SEPD is a production decline model proposed for transient flow by Valko and Lee (2010). The equation for the SEPD model is

$$q = q_i \exp\left(\frac{-t}{\tau}\right)^n, \quad (5.9)$$

where τ is a characteristic time constant, n is an empirical parameter and q_i the initial production rate. EUR obtained from the SEPD model is bounded unlike Arps' hyperbolic model for $b > 1$.

To calculate the parameters τ and n , Yu and Miocevic (2013) proposed a method in which $\ln\left(\frac{q_i}{q}\right)$ vs t is plotted on a log-log scale. Yu proposed that q_i be approximated as the peak rate observed in a well. A straight line is usually obtained after several months of history, and the parameter n is determined from the slope. The characteristic time constant τ is calculated

$$\tau = \exp\left(\frac{-\ln(\text{Intercept})}{n}\right). \quad (5.9)$$

After the end of the transient flow period, Arps hyperbolic decline model can be used to forecast production. The Arps' hyperbolic parameters, initial decline rate, D_i , and b are calculated from the parameters in the SEPD model from

$$D = n\tau^{-n}t \text{ and} \quad (5.10)$$

$$b = \frac{\tau^n(1-n)}{n} t^{-n}. \quad (5.11)$$

Yu (2014) also proposed a method to forecast the condensate production using a simple empirical method which involves plotting cumulative CGR (Condensate Gas Ratio) vs cumulative gas on a semi-log graph to obtain a straight line given by

$$(CGR)_{cum} = A_1 + B_1 \ln(G_p) . \quad (5.12)$$

This equation is used to calculate the $(CGR)_{cum}$ from G_p , which is calculated using the SEPD method. Finally, condensate volume is calculated by using the $(CGR)_{cum}$ from the SEPD (cumulative) equation.

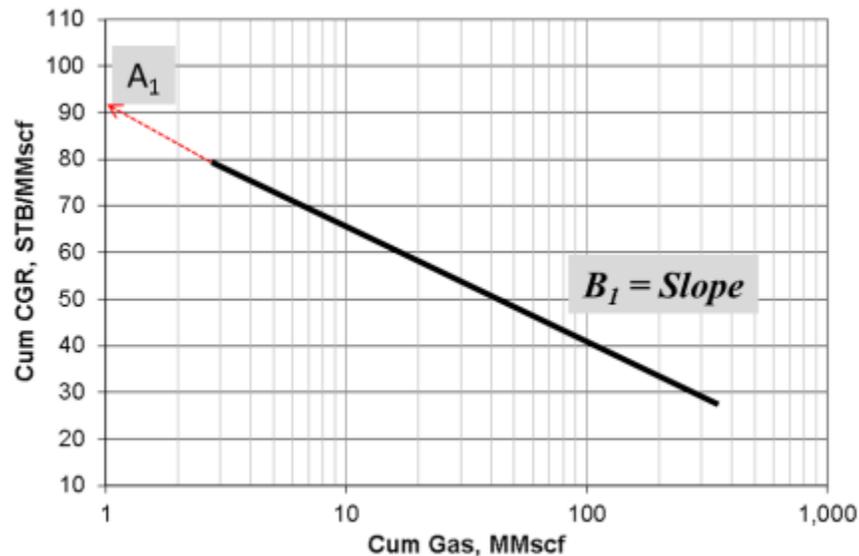


Figure 42. Parameter estimation to calculate cumulative oil by using Yu method

5.5 Diagnostic Plots

The first step in production forecasting is to prepare diagnostic plots to identify flow regimes. Flow regime identification is important because each decline model is valid only for the flow regime for which it was derived. Application of decline models without identifying flow regimes usually leads to erroneous production forecasts. The most common diagnostic plot used to identify the flow

regimes is a log-log rate-time plot. This diagnostic plot ignores possible varying bottom-hole pressure (BHP) and will lead to misleading results whenever BHP is changing with time. In particular, for a rate restricted well (constant rate production), this model does not show any decline in rate with time and cannot be used to identify the flow regime. Thus, we recommend that, where possible, analysts use pressure normalized diagnostic plots for flow regime identification purposes. The normalized rate is generated by dividing the flow rate by flowing pressure drawdown (or pseudo-pressure drawdown for gas wells) as

$$q_{o\text{Normalized}} = \frac{q_o}{p_i - p_{wf}}, \quad (5.13)$$

$$q_{g\text{Normalized}} = \frac{q_g}{m(p)_i - m(p)_{wf}}, \quad (5.14)$$

$$m(p) = \text{gas pseudopressure} = 2 \int_0^p \frac{p dp}{\mu z}. \quad (5.15)$$

The most common flow regimes observed in unconventional shale reservoirs are transient linear flow and boundary dominated flow. On a log-log plot of normalized rate vs. time, transient linear flow can be identified by a straight line with a slope of -1/2. To identify boundary dominated flow with confidence, a modified log-log diagnostic plot of pressure-normalized rate vs material balance time is preferred (when possible). Blasingame's material balance time, defined as the ratio of cumulative production to rate, provides a convenient way to convert variable-rate production into equivalent constant rate production (Mattar, 2003). On a material balance time plot, boundary dominated flow is identified with a straight line of slope -1. Figure 43 shows the log-log diagnostic plots with pressure

normalized gas rate and time or material balance time for a typical shale gas reservoir.

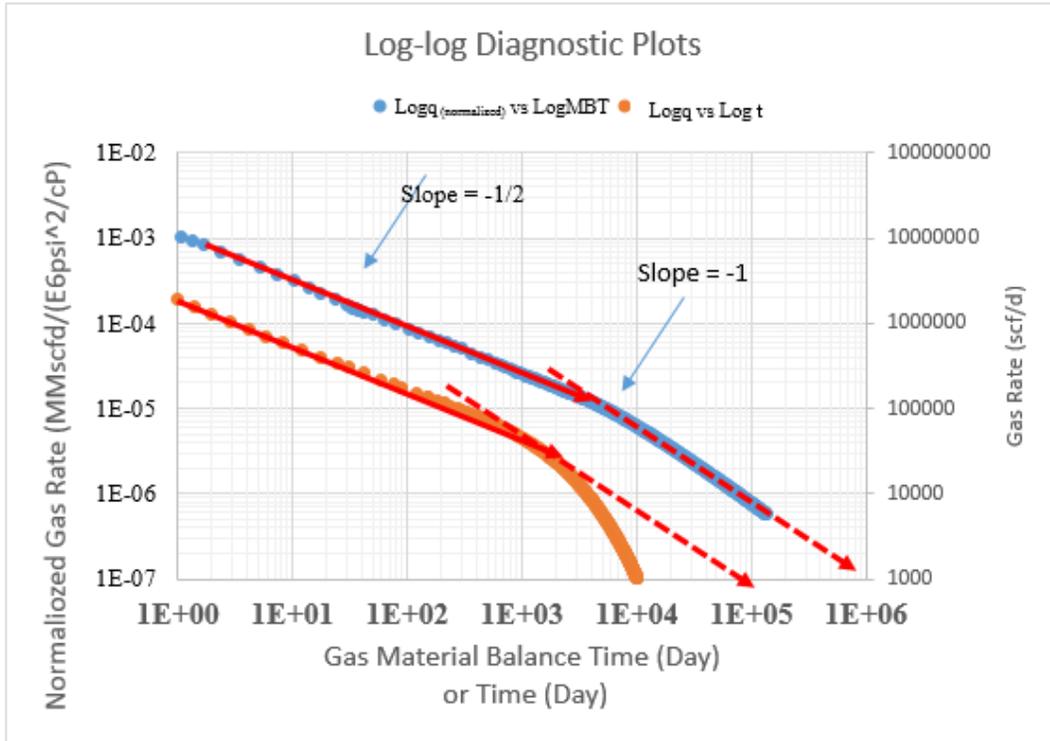


Figure 43. Diagnostic plot to identify flow regimes for a typical shale reservoir

5.6 Generation of Synthetic Field Data

In compositional simulation, material balances for $2n+1$ phases (liquid and vapor for each of n components and water) are performed by a simulator. Moreover, in compositional simulation phase behavior depends on an Equation of State (EOS) with three or more parameters. In our compositional simulation, we used a drainage area of about 3.03 acres (200 ft x 660 ft) with a single fracture as a symmetry element of a well with 25 fractures to reduce the computational time. The schematic of the reservoir is shown in Figure 44.

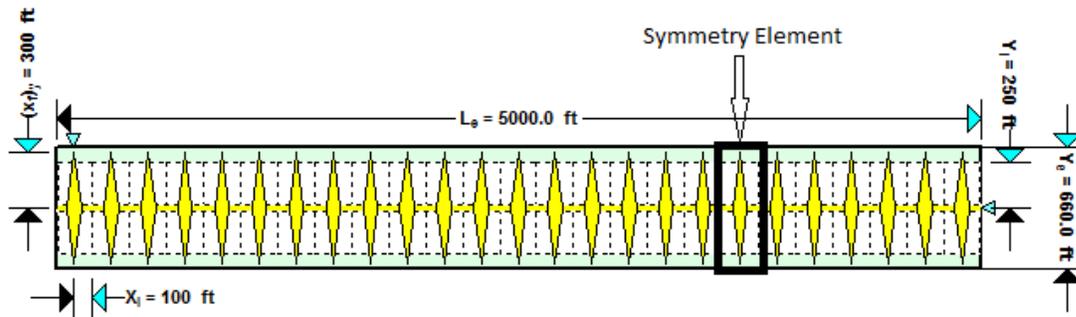


Figure 44. Schematic for reservoir to generate synthetic production data

Table 9. Near critical fluid with varying compositions used for simulation

	Sample 1	Sample 2	Sample 3	Sample 4	Sample 5
CO2	2.66	4.48	0.01	0.14	0.18
N2	0.17	0.7	0.11	1.62	0.13
CH4	60.0	66.2	68.9	66.8	61.7
C2H6	7.72	7.21	8.63	11.35	14.1
C3H8	6.5	4	5.34	6.01	8.37
IC4	1.93	0.84	1.15	1.37	0.98
NC4	3	1.76	2.33	1.94	3.45
IC5	1.64	0.74	0.93	0.84	0.91
NC5	1.35	0.87	0.85	0.97	1.52
FC6	2.38	0.96	1.73	1.02	1.79
C7+	12.69	12.2	9.99	7.98	6.85
C7+ Mol Wt	179.2	170	169	158	143
C7+ Sp Gr	0.8	0.827	0.813	0.827	0.795
Reservoir Temp	312	286	238	256	186
Liquid API	45	44.82	45.43	48.32	48.22
Sat Pressure	5210	5410	4453.81	5805	4710.41

The well was simulated for 10,957 days (30 years). For this reservoir, the fracture half was assumed to be 300 ft and fracture conductivity of 2000 md.ft. The reservoir thickness is assumed to be around 100 ft with porosity and permeability of 7% and 200 nd respectively. The reservoir is assumed to have initial gas and

water saturation of 60% and 40% respectively. Relative permeability curves used in Chapter 3 (Figure 23) is used to account for multiphase flow effects.

In addition to the condensate samples in Chapter 4, five more near-critical samples (fluids close to the critical point) or rich condensate samples were selected for simulation (Ahmed, 2005). The composition and heavier compound (C_7^+) for the near critical fluid samples are given in Table 9. The initial Condensate to Gas Ratio (CGR) was determined by simulating separator experiments with first stage separator conditions of 400 psia and 100°F.

Table 10. Tuned EOS parameter for sample with C_7^+ concentration of 7.1%

Components	P_c (atm)	T_c (K)	V_c (l/mol)	Ascentric Factor	Molecular Wt.
CO2	72.8	304.2	0.1	0.2	44.0
N2	33.5	126.2	0.1	0.0	28.0
CH4	45.4	190.6	0.1	0.0	16.0
C2H6	48.2	305.4	0.1	0.1	30.1
C3H8	41.9	369.8	0.2	0.2	44.1
IC4	36.0	408.1	0.3	0.2	58.1
NC4	37.5	425.2	0.3	0.2	58.1
IC5	33.4	460.4	0.3	0.2	72.2
NC5	33.3	469.6	0.3	0.3	72.2
FC6	32.4	507.0	0.3	0.3	86.0
C07-C09	26.2	568.0	0.4	0.3	108.1
C10-C11	23.4	635.1	0.6	0.4	144.6
C12-C13	20.7	676.6	0.7	0.5	172.7
C14-C15	18.7	712.3	0.8	0.6	200.7
C16+	18.4	674.6	1.0	0.7	271.2

A three-parameter Peng Robinson (Peng and Robinson, 1976) EOS was used to model PVT phase behavior. The lumped C_7^+ fraction was split into 3 to 5 pseudo-components before attempting to match saturation pressure, separator API gravity and CGR. Finally, the EOS was tuned by using linear regression to

match reported experimental values. The PT diagram for each sample was created and for the reservoir temperature listed below, the first five samples are very close to the critical point.

The tuned EOS parameters for sample with C_{7+} concentration of 7.1% (shown in Chapter 4, Table 7) is shown in Table 10.

5.7 Flow Regime Identification

Hydraulically fractured shale reservoirs with infinite fracture conductivity show linear flow during early production life. Fractures with lower values of dimensionless fracture conductivity (finite conductivity fractures) exhibit bilinear flow. When the pressure transients reach the no-flow boundaries of the reservoir (fracture interference, well interference, or outer boundaries), boundary dominated flow appears. These transitions are straightforward and easy to identify in conventional reservoirs. However, these flow regimes are not so straightforward to identify in unconventional reservoirs with multiphase flow. Often, the transition from linear flow to boundary dominated flow is not abrupt but can take more than a log cycle of time depending on reservoir characteristics.

Production during the transient linear flow period can be forecasted by Duong's transient model (Duong, 2011), and the stretched exponential production decline (SEPD) model. For boundary dominated flow, the Arps hyperbolic decline model is appropriate for forecasting. For dry gas wells, a production decline exponent value "b" of 0.4- 0.5, has been observed to be appropriate (Fetkovich et al., 1996). For the transition period between transient flow and boundary dominated flow, there is no consensus method for forecasting. In this work,

transient models (Duong, SEPD) are used for modeling during the transition flow period. This generalization, although not rigorously correct, can be used as both Duong's transient model and the SEPD model work well for near linear flow regimes.

Flow regime identification for simulated production data is performed by diagnostic plots, where pressure normalized flow rate is plotted against material balance time (MBT) on log-log coordinates. Linear flow is represented by a line of slope $-1/2$ and the boundary dominated flow is represented by a straight line with negative unit slope. For this work, we use two diagnostic plots namely, pressure normalized rate versus time and pressure normalized rate versus material balance time to identify the flow regimes. Each of these diagnostic plots, in conjunction with one another, can be used to identify the start and end of linear and boundary dominated flow regime. Figure 45 and 46 show the diagnostic plots for Sample 1 produced at a constant bottomhole pressure of 1000 psia.

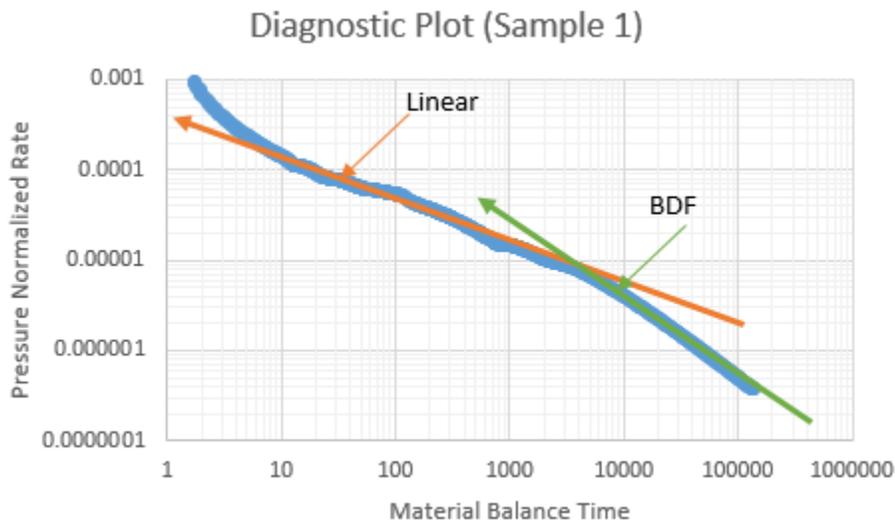


Figure 45. Normalized rate vs MBT diagnostic plot

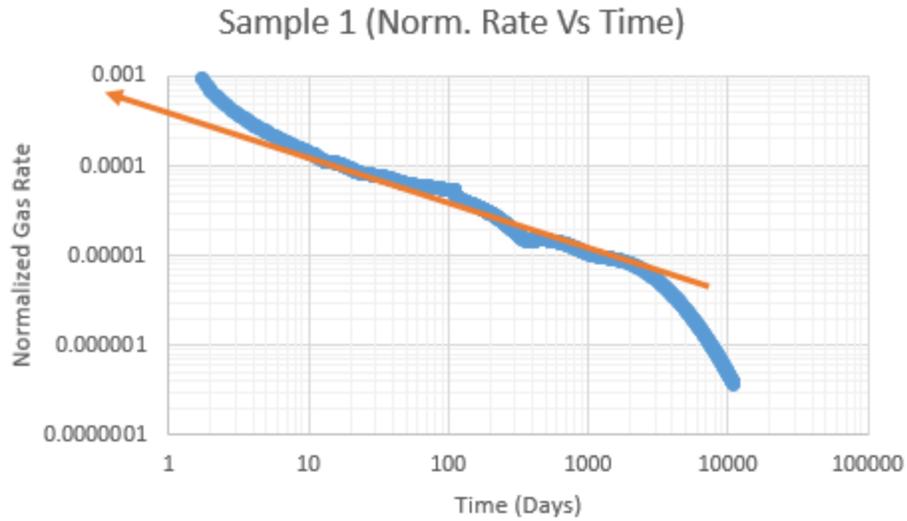


Figure 46. Normalized rate vs time diagnostic plots

Kinks in the diagnostic plots appear at the time when rate changed from constant rate (choked flow) to production at constant bottom-hole pressure. For cases with constant bottomhole production the diagnostic plots are smooth for the entire production period as seen in the Figure 47 and Figure 48.

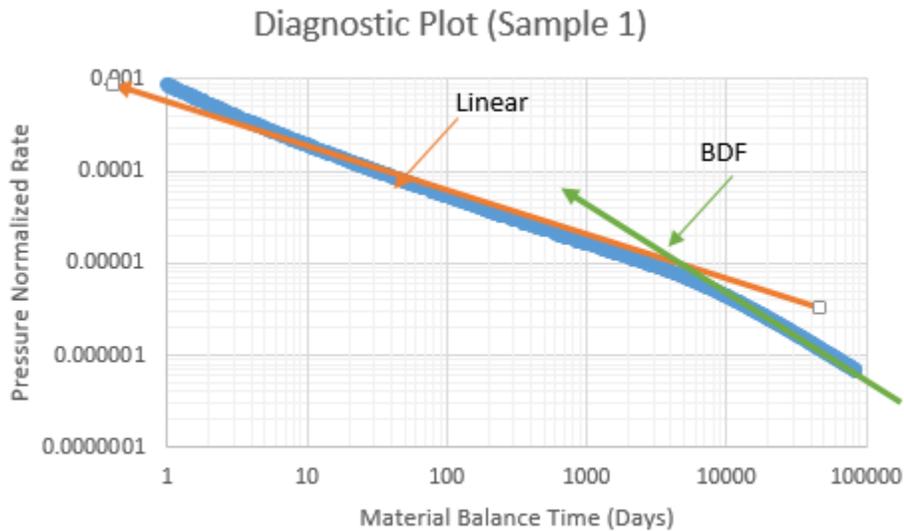


Figure 47. Normalized rate vs MBT diagnostic plot (constant BHP)

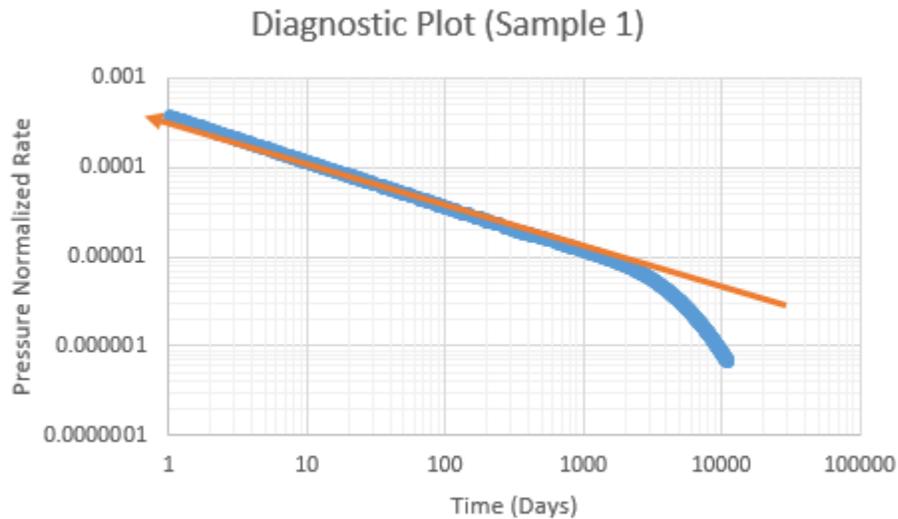


Figure 48. Normalized rate vs Time diagnostic plot (constant BHP)

Comparison of Figure 45 with Figure 47 shows that diagnostic plots for cases with constant rate (varying bottomhole pressure constrained by a minimum value) fluctuates significantly compared to the constant bottomhole case. Moreover, pressure normalization helps in identifying the flow regime even during the period of constant flow rate. Diagnostic plots for each of the samples studied follow the same trend, linear flow followed by the boundary dominated flow, as all the reservoir characteristics are same except for the initial fluid composition.

For this study, the maximum initial rate allowed for each well was 100 Mscf/D until the bottomhole pressure reaches 1000 psia. By visual inspection of the log-log diagnostic plots, time for the end of linear flow (t_{elf}) and start of boundary dominated flow (t_{sbd}) is identified and is shown in Table 11. It can be seen that for all the samples studied, the intermediate period between the transient flow and the boundary dominated flow is extremely long. For this reason, the pre-boundary-dominated flow periods are forecasted by using the SEPD and Duong flow models.

After t_{sbdf} , the Arps hyperbolic model is used with a b value of 0.5. It can also be seen that the linear flow period generally lasts longer for samples with low C^{7+} content.

Table 11. Start and end of the boundary dominated flow for each sample

Sample	1	2	3	4	5	6	7	8	9
telf(MBT days)	5000	5000	5000	5000	3000	2000	2000	1000	2000
tsbdf(MBT days)	20000	20000	20000	15000	10000	30000	20000	20000	20000
telf(days)	2069	2222	2372	2099	1339	1034	1049	615	1049
Tsbdf (days)	4473	4838	5447	3711	3195	5417	4383	4473	4473

5.8 Yu Plot

Yu and Miocevic (2013) modified the existing SEPD method so that the parameters η and τ can be calculated easily and more accurately. The method involves plotting $Ln \left[\frac{q_i}{q_t} \right]$ versus time on a log-log scale, referred to as the Yu Plot. All field data prior to appearance of a straight line on the plot is discarded and the straight line is manually matched to calculate the parameters. For this work, Yu's method indicated that the model became applicable after about a year. Yu and Miocevic also proposed that this modified method works for the entire life of the reservoir regardless of the flow regime. Yu plots for the 2nd to 3rd year production data is given in Figure 49.

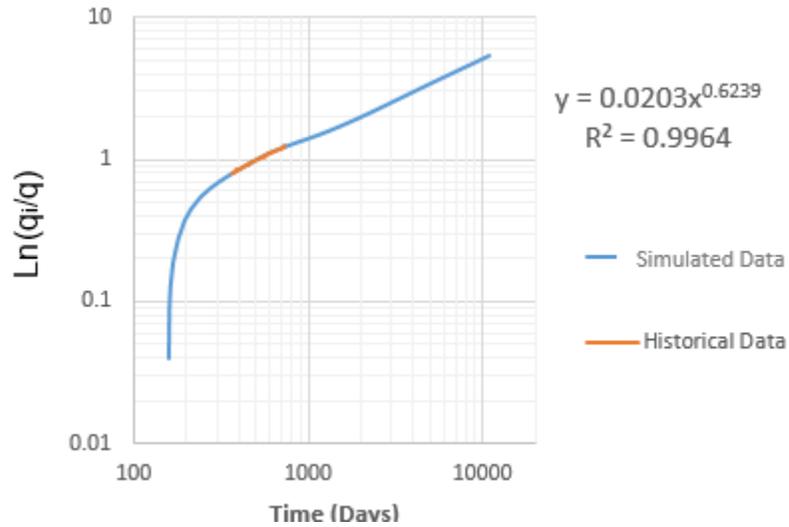


Figure 49. Yu plot to calculate SEPD parameters for sample 1

Similarly, data from year 3 to 4 (1095 days to 1460 days) was used to generate different sets of slope and intercept for the period. The slope of the Yu plot gives n and τ , which is calculated using equation 5.9. From the Yu-Plots, n and τ were calculated; they are summarized in Table 12 and Table 13.

Table 12. SEPD parameters based on year 2 to 3 using Yu Plot

Sample	Intercept	Slope	n	Tau, days
1	0.020	0.624	0.624	516
2	0.016	0.652	0.652	553
3	0.049	0.506	0.506	382
4	0.026	0.017	0.622	350
5	0.002	0.948	0.948	741
6	0.007	0.733	0.733	857
7	0.003	0.829	0.829	947
8	0.004	0.812	0.812	957
9	0.002	0.873	0.873	1006

Table 13. SEPD parameters based on year 3to 4 using Yu Plot

Sample	Intercept	Slope	n	Tau, days
1	0.048	0.485	0.485	512
2	0.074	0.421	0.421	488
3	0.235	0.272	0.272	206
4	0.019	0.607	0.622	613
5	0.025	0.556	0.556	761
6	0.014	0.632	0.632	837
7	0.007	0.722	0.722	946
8	0.007	0.722	0.722	949
9	0.006	0.746	0.746	972

5.9 Duong's Production Decline

Duong's equation can be solved in two simple steps; the first step is plotting ratio of cumulative production to production rate versus time on a log-log scale. The parameters a and m can be obtained from intercept and slope respectively. The next step is to plot rate versus modified time, t_m , to obtain q_1 . The intercept of the line, q_{inf} , is forced to zero. Figure 50 and Figure 53 shows the plots used to calculate the Duong model parameters.

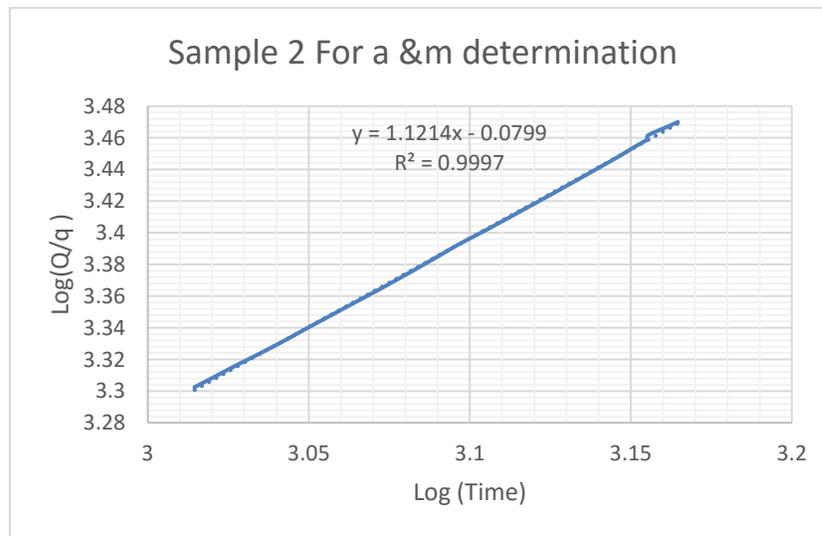


Figure 50. Duong's Plot to calculate parameters a & m

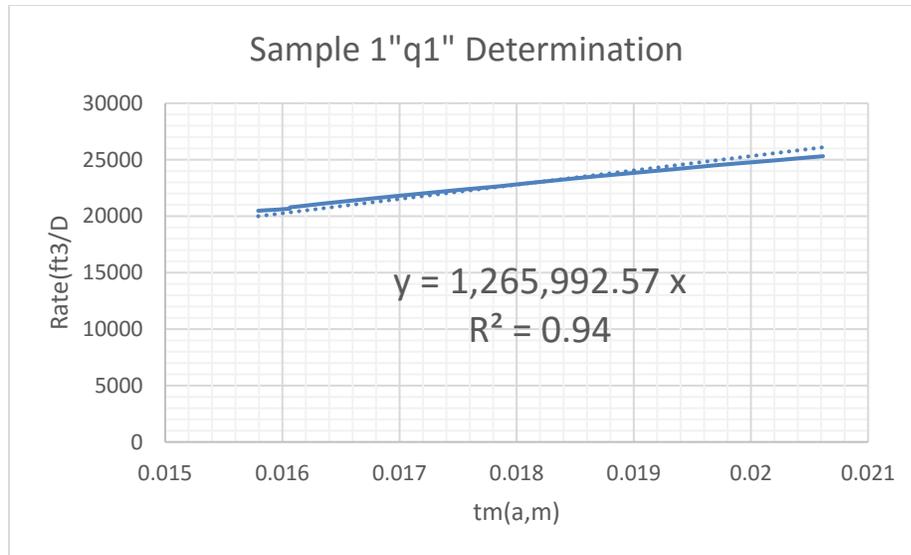


Figure 51. Initial rate determination by using parameters a & m

The same procedure was applied as before and Duong's parameters were calculated for each fluid sample as shown in Table 14.

Table 14. Duong's a, m and q_1 parameters for each samples

Sample	a	m	q_1 (SCF/day)
1	0.39	1.23	3.33E+07
2	0.83	1.12	1.27E+06
3	0.66	1.15	2.72E+06
4	0.74	1.1	1.43E+06
5	0.41	1.22	4.33E+07
6	0.24	1.3	1.48E+08
7	0.15	1.37	3.53E+08
8	0.15	1.37	3.58E+08
9	0.14	1.38	3.58E+08

5.10 Arps Hyperbolic Model

The time for start of boundary dominated flow ($t_{s b d f}$) is determined by visual inspection of the diagnostic plots. However, from the diagnostic plot it can be seen that the transition period between the end of linear flow and the start of boundary-

dominated flow lasts for a long time and, for most of the samples, boundary dominated flow starts at about 20,000 days MBT. In this chapter, the later time was used as the start of boundary dominated flow and the beginning of the transition region was 1461 days. Production during the transition region was forecasted by using one of the methods used for the transient flow period and Arps' equation with b value of 2. During BDF, the initial decline rate D_i was calculated with equation 5.2 and b of 0.5 was used for all cases. For each case, D_i about 5% was used per year regardless of the initial CGR.

5.11 Production decline

Gas production was forecasted based on decline model parameters determined from analysis of historical data available to 1460 days (4 years). The forecasted production rates and cumulative production volumes for the reservoir fluid with CGR 120 is given in Figure 52.

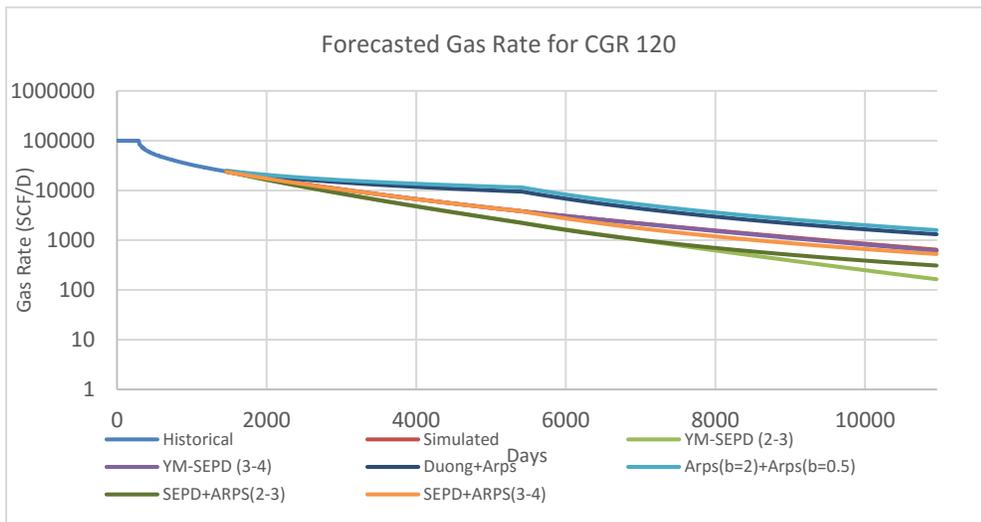
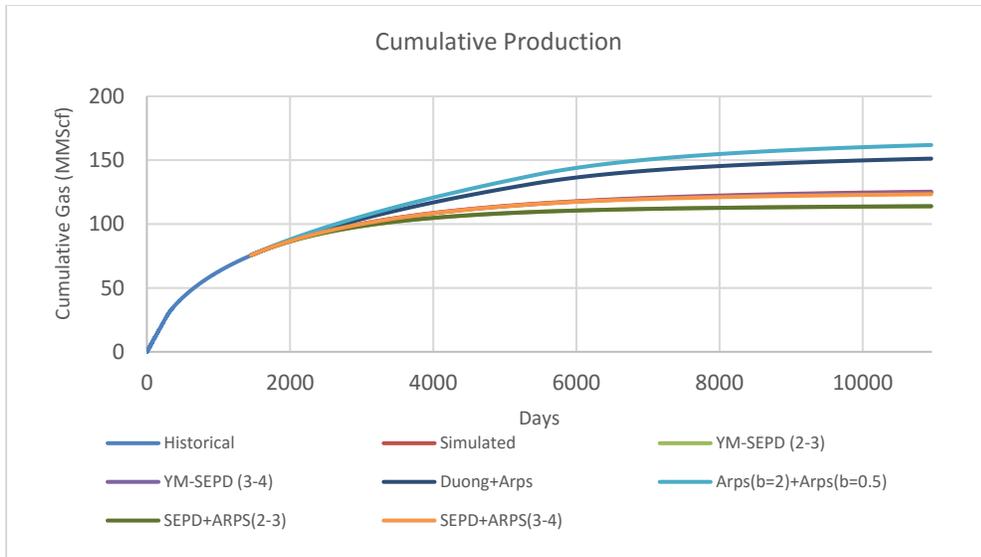


Figure 52. Forecasted cumulative gas and gas rates

Figure 52 shows that YM-SEPD with up to 4 years of data produces a good forecast that matches the historical data almost perfectly. The EUR calculated from each case was analyzed and the error in EUR was calculated by considering the EUR from simulation to be correct.

Table 15. Error calculation for each sample by using different methods

Sample	YM-SEPD (2-3)	YM-SEPD (3-4)	Arps(b= 2) +Arps	Duong + Arps	SEPD+ ARPS(2-3)	SEPD+ ARPS(3- 4)
1	-10.3	10.5	32.3	7.7	-9.4	5.6
2	-16.4	16.9	22.4	13.2	-15.1	9.9
3	-13.6	32.3	21.4	5.3	-12.8	9.8
4	-15.0	-4.6	22.1	13.4	-13.1	-4.5
5	-25.5	5.8	23.3	4.8	-21.8	1.8
6	-9.3	-0.4	17.2	14.4	-9.0	-1.6
7	-13.1	-5.8	21.5	8.6	-11.7	-5.4
8	-12.6	-6.5	21.2	7.0	-11.4	-6.3
9	-14.6	-7.6	19.3	8.1	-13.4	-7.3

In Table 15, it can be seen that YM-SEPD with 4 years of history followed by Arps provides the most accurate result for final cumulative gas production for most cases investigated in this study. Using YM-SEPD alone doesn't produce good results for samples close to the critical point. However, this method gives a reasonable result for lean samples. Duong, followed by Arps, gives reasonable results for rich samples but not for lean samples. Arps with b value of 2 followed by Arps hyperbolic with b value of 0.5 overestimates the production in each case.

5.12 Determination of Cumulative Oil Production

Cumulative oil production from a gas condensate reservoir depends on several factors including the initial production rate, minimum bottomhole pressure, fracture conductivity and their interactions. This was discussed by Khanal et al. (2015). for this paper, a method proposed by Yu (2014) was used in which cumulative CGR is regressed against cumulative gas production. The slope and intercept from the regression is used to estimate the final cumulative CGR at the

end of the production period. Finally, the cumulative oil production is calculated by multiplying the estimated cumulative CGR by the final cumulative gas volume.

This method was applied to all the samples first by regressing on the historical data from year 2 to 3 and then year 3 to 4. Figure 53 is a semi-log plot of cumulative CGR and cumulative gas production for a sample with CGR of 64 STB/MMScf.

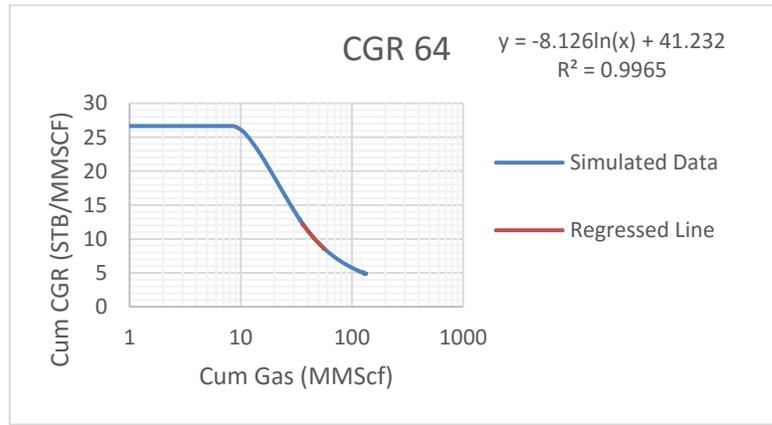


Figure 53. Yu plot for determination of condensate production

Each column in Table 16 (a, b, c and d) labeled Simulated Year 1-2, 2-3 and 3-4 represents calculation of cumulative oil production (N_p) and error using the historical data from the same time period. In most cases the ultimate production was underestimated by this method with the best results obtained by using 4 years of historical data.

Table 16. Cumulative oil estimation a. Simulated b. Year 1-2 c. Year 2-3 d. Year 3-4

Sample	a.		b.		c.		d.	
	Simulated Gp (Mscf)	Simulated Np (bbl)	Calculated Np (bbl)	Error, %	Calculated Np(bbl)	Error, %	Calculated Np (bbl)	Error, %
1	96	601	125	-79	402	-33	499	-17
2	108	2479	1491	-40	1994	-20	2186	-12
3	104	1250	148	-88	1369	10	1369	10
4	111	1727	1069	-38	1467	-15	1596	-8
6	125	789	180	-77	551	-30	669	-15
7	133	910	308	-66	686	-25	798	-12
8	134	650	191	-71	477	-27	564	-13
9	137	575	260	-55	459	-20	517	-10

5.13 Application to field history Data

To verify the importance of pressure normalization and identification of flow regimes prior to decline curve analysis, production history from a gas condensate well in the Eagle Ford shale (EFS) was analyzed. The well is located in the gas condensate window of the EFS and produced at an initial CGR of 156 STB/MMScf.

Rate and pressure history is available for a period of 688 days (~2 years). The diagnostic plots for the reservoir are shown in Figures 54 and 55.

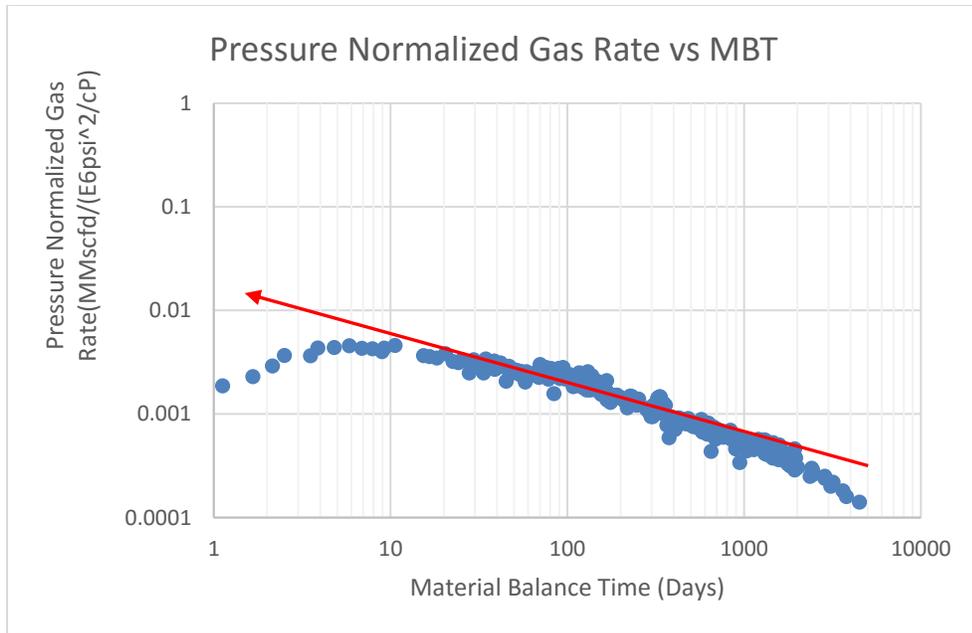


Figure 54. Pressure normalized vs MBT diagnostic plot for a LRS reservoir

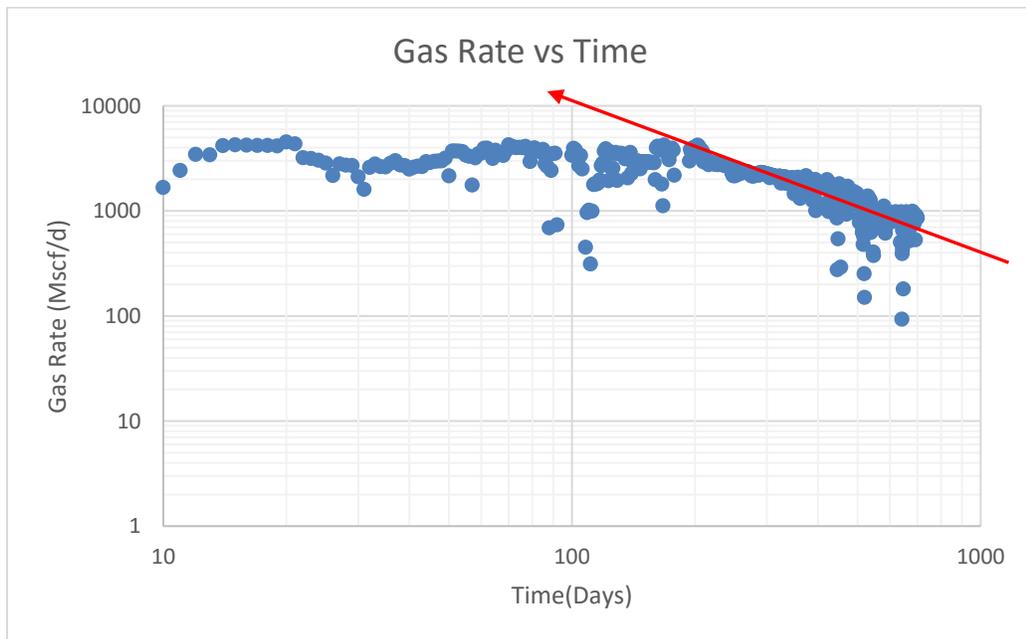


Figure 55. Gas rate vs time diagnostic plot for LRS reservoir

Figure 54 and 55 above shows that flow regimes can be identified by using pressure normalized gas rate for less than 2 years of production history more efficiently and with greater confidence than with the more basic log-log rate vs time plot. A log-log pressure normalized rate versus time plot shows that the well

displayed linear flow almost for the entire production period. BDF may be identified by a few isolated data points toward the end of history (but this could also be liquid loading). In contrast, on diagnostic plot of gas rate vs. time, linear flow is completely masked. Moreover, the log-log diagnostic plot of gas rate and time suggests that boundary influenced flow starts after 150 days. The reservoir could well be in boundary dominated flow after 700 days. However, it is unlikely that a reservoir with nano-darcy permeability will show BDF after 150 days of production and no linear flow. If a forecast is made based on this analysis the EUR from this well could be seriously underestimated. In this case, the pressure normalized diagnostic plot correctly identifies flow regimes which are not seen in rate-time diagnostic plots.

A similar analysis was performed on a dry gas reservoir initially at 5,700 psia in the Woodford shale. As seen in Figures 56 and 57, both diagnostic plots show linear flow throughout the production history. However, the normalized diagnostic plot shows that the linear flow starts before shut-in period at 80 days. From these figures, it can be inferred that if the reservoir is not over-pressured and is in single phase flow, and a simple rate vs time diagnostic plot might be sufficient to identify the flow regimes. After the flow regimes are identified, the forecasting should be performed by using linear model and Arps' minimum decline model based on field analogs. From the simulation study, boundary dominated flow starts at about 5% of the initial rate regardless of the initial CGR.

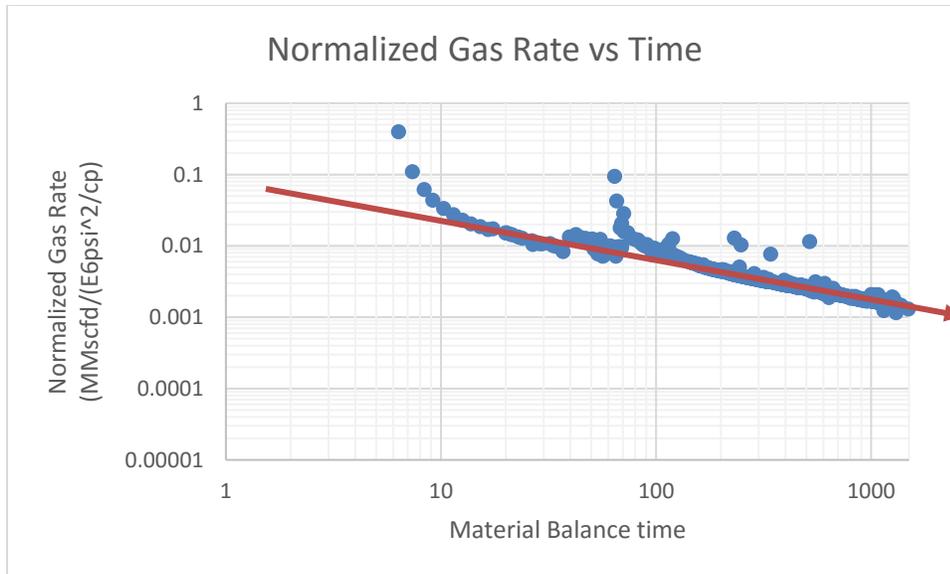


Figure 56. Pressure normalized vs MBT diagnostic plot for a dry gas reservoir

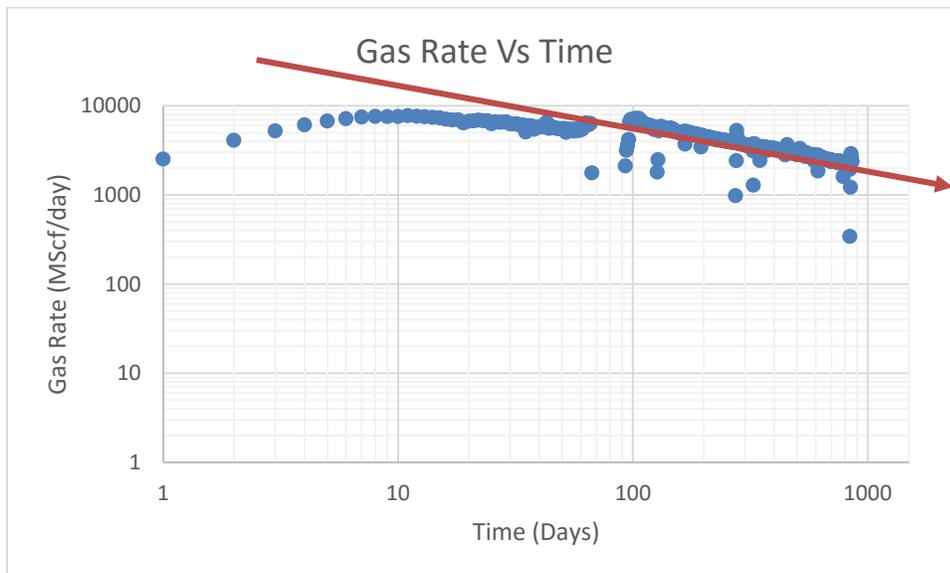


Figure 57. Gas rate vs time diagnostic plot for a dry gas reservoir

5.14 Conclusions

The following conclusions can be drawn from this work:

- Gas condensate wells in shales exhibit a long transition period between the end of linear flow and the start of boundary dominated flow.

- Pressure normalization is an effective method to identify flow regimes in a gas condensate reservoir. Using pressure normalized diagnostic plots may help in identification of flow regimes which could otherwise be difficult to identify.
- YM-SEPD provides a good forecast for lean gas condensate samples.
- YM-SEPD combined with Arps with b value of 0.5 provides a result closer to the simulated value than other methods considered provide for all reservoir fluids
- The transient linear flow model with no modification for boundary-dominated flow overestimates the production in almost all cases.
- Estimation of ultimate condensate production using Yu's method provides a reasonable result when four years of historical production data are available and used.

6. Data driven study of liquid Rich Shale Reservoirs

6.1 Background and Introduction

Principal component analysis (PCA) is regarded as one of the most valuable tools from applied linear algebra (Shlens, 2003). PCA is used in different areas of studies, from neuroscience to signal processing, because it is simple, non-parametric method of extracting relevant information from large and confusing data sets. PCA is a powerful tool which allows researchers to reduce a complex data set to a lower dimension and to reveal the hidden underlying dynamics with relatively minimal additional effort. In this chapter, we use this technique to forecast the performance of oil and gas reservoirs using simulated and field data.

2. Principal Component Analysis

6.2 Literature Review and Problem Statement

Production data analysis is an evolving field where several new techniques are applied together to develop, analyze, forecast, and evaluate the production from oil and gas reservoirs. These techniques have been applied to various types of reservoirs with varying degrees of success in the past. Several authors (Denney 1999; Srinivasan and Ertekin 2008; Mohaghegh 2009; Fulford et al., 2016) have recently discussed the use of artificial intelligence techniques like neural networks and fuzzy logic to forecast production for oil and gas reservoirs.

One such method which can be used for production data analysis is PCA, which has been used extensively in geosciences to reduce redundant data and retain useful data obtained from seismic data acquisition (Saleh and Bruin 2000, Tingdahl and Hemstra, 2003; Guo et al., 2006, Chopra and Marfurt, 2014). It has also been used for other reservoir engineering applications such as history matching (Honorio et al., 2015; Chen et al., 2016), reservoir property estimation (Scheeval and Payrazyan 2001; Dadashpour et al., 2011), gas flow in nano-pores and the analysis of production history for unconventional gas reservoirs (Bhattacharya and Nikolaou 2011, 2013). Researchers have also used PCA to perform fluid composition characterization of crude oils from different depths and wells to resolve compositional changes related to the source of the oil generating sediments and its exposure to biological and/or physical weathering processes (Ventura et al., 2011). More recently, PCA has been used to reduce uncertainty in history matching and to identify reservoir measurements that best represent the overall reservoir behavior (Bertolini and Schiozer, 2016).

To our knowledge, this method has not been applied to forecast production and condensate-to-gas ratio (CGR) in liquid-rich shale (LRS) gas condensate reservoirs. Due to the complex nature of gas condensate reservoirs, several complex reservoir simulations need to be performed, which generally requires several days and weeks to complete even with the most powerful computing resources available today. Statistical and data-driven approaches such as PCA reduce the number of such simulations required by providing a relatively quick recognition of important patterns, thus allowing a better framework for

planning and simulation of new wells. This also reduces the several steps required for proper application decline curve methods and simplifies the workflow.

6.3 Theory and Background

Principal component analysis is regarded as one of the most valuable tools from applied linear algebra (Shlens, 2003). Principal component analysis reduces the dimensionality of multivariate datasets by reducing the variables in a data-set into smaller number of derived variables, principal components (PC), which are linear combinations of original variables (Jolliffe, 2014). Principal Components are arranged in descending order, and the axis corresponding to the first principal component (or eigenvector) is the one along which the variance of the data is maximized, followed by the second principal component and so on (Jolliffe, 2014). A matrix X of rank m (and variables) can be represented by a matrix X' of lower rank p as

$$[X]_{n \times m} \approx [X']_{n \times p}, \tag{6.1}$$

$$= (PC_1)v_1^T + \dots + (PC_m)v_m^T \text{ and} \tag{6.2}$$

$$\approx (PC_1)v_1^T + \dots + (PC_p)v_p^T + E^p. \tag{6.3}$$

In the equations above, the PC_i are the principal components scores, v_i are the loadings or principal component coefficients which describe the systematic part of the data, and E^p is the residual matrix which describes the model and measurement errors (Shlens, 2003). The number of principal components required to accurately represent the original data depends on two factors: noise and redundancy and usually done empirically (Jolliffe, 2014).

Principal Component Analysis is performed by creating the covariance matrix, calculating the respective eigenvectors and eigenvalues of the matrix and finally ranking them based on their respective eigenvalues (Jolliffe, 2014). The eigenvectors with the greatest eigenvalues are the Principal Components of the data matrix. Principal components can also be calculated by using singular value decomposition (SVD) where a matrix M is decomposed into matrices U , S and V as

$$M = USV^T, \quad (6.4)$$

where U and V are orthonormal vectors (vectors with unit norm and zero inner product) and S is a diagonal matrix with eigenvalues in the main diagonal (Jolliffe, 2014). The product of matrix U and S yields the PC $_i$ represented in Equation 1 and V represents the principal component coefficients v_i .

The field of petroleum engineering is inundated with data from various sources, such as geological, production, and experimental data. In the United States, most of these data are proprietary except for a few publicly reported data sets such as rate-time data or occasional well testing data. This collection of data can be viewed as a large matrix, which, when analyzed using multi-variate statistical methods like PCA, can be reduced to a smaller matrix that retains the important pattern from the original matrix. These smaller matrices, with fewer rows and columns, can be used more efficiently to predict the performance of the original wells and the wells with similar characteristics.

If we assume $q_i(t)$ to be the gas rate and $c_i(t)$ to be the corresponding producing condensate-to-gas ratio (CGR) for well i , the production data can be represented in matrix as

$$X = \begin{bmatrix} q_1(t_1) & \cdots & q_1(t_n) \\ \vdots & \ddots & \vdots \\ q_i(t_1) & \cdots & q_i(t_n) \end{bmatrix} = \begin{bmatrix} Q_1^T \\ \vdots \\ Q_i^T \end{bmatrix} \text{ and} \quad (6.5)$$

$$Y = \begin{bmatrix} c_1(t_1) & \cdots & c_1(t_n) \\ \vdots & \ddots & \vdots \\ c_i(t_1) & \cdots & c_i(t_n) \end{bmatrix} = \begin{bmatrix} C_1^T \\ \vdots \\ C_i^T \end{bmatrix}. \quad (6.6)$$

The rows of matrices X and Y represent gas rate and CGR for 'n' time steps for 'i' wells respectively. Q_1^T and C_1^T represent the transpose of the respective data in column form.

The gas rate and CGR for well X_i can be represented as

$$X'_i = (PC_{i_1})v_1^T + (PC_{i_2})v_2^T + \dots \text{ and} \quad (6.7)$$

$$Y'_i = (PC_{i_1})v_1^T + (PC_{i_2})v_2^T + \dots \quad (6.8)$$

As seen in Equation 6.10, E^p represents the residual error, which is discarded in the approximated rate (X'_i) data and can be represented as follows:

$$E^p = (PC_{i_{p+1}})v_{p+1}^T + \dots + (PC_{i_m})v_m^T \text{ and} \quad (6.9)$$

$$= \sum_{p+1}^m PC_i v^T. \quad (6.10)$$

The normalized percent square residual error can be calculated by square root mean squared error with respect to the original data point or the square of Frobenius norm of the error with respect to original data as

$$X_{error} = \sum_1^m \frac{\|(X-X')\|^2}{\|X\|^2}. \quad (6.11)$$

6.4 Prediction and Forecasting for New Wells

Principal component analysis can be used to predict the rates (gas/oil rate and cumulative, CGR) for new wells by using the PC coefficients of existing wells. Equations. 6.1 – 6.10 show that rates for a well can be represented as a weighted linear combination of PC coefficients. The individual weights or PC scores are functions of several reservoir, fluid and completion properties (Bhattacharya and Nikolaou 2013) as

$$PC_i = f(\textit{reservoir} + \textit{fluid} + \textit{completion}). \quad (6.12)$$

In presence of these properties, the PC weights for each new well can be calculated, which can then be used to predict the production behavior of the new wells. Dependence of the PC weights on these several properties can be determined by using a statistical tool like multivariate linear regression. Analysis of Variance (ANOVA) from the regression analysis can be used to calculate the confidence interval for the prediction.

Cross-validation is a technique used to analyze the effectiveness of results of regression or some other statistical analysis to generalize an independent data set (Geisser, 1993). Cross-validation tests the effectiveness of a model generated from a training data set to predict a new set of data (testing data set). Cross validation can be performed by dividing the original data sets into two groups: training set and validation set. However, this method could lead to high variance due to the fact that the model depends heavily on the data points selected for the training set (Schneider, 2016). For this reason, leave-one-out-cross-

validation method is used where one of the m observations is used as the validation (or test) set and $m-1$ observations are used as the training set (Schneider, 2016). In this case, one set of well production data is excluded while generating the PC scores and coefficients. This excluded rate data is then predicted by using the generated PC parameters. This process is repeated for each well to conduct a comprehensive cross-validation.

6.5 Workflow

To illustrate the use of PCA as a forecasting tool, we used both simulated data and field data. A commercial reservoir simulator was used to model an unconventional reservoir with properties similar to a liquid-rich shale reservoir (Eagle Ford) and was run for 30 years. We identified the uncertain parameters and defined the probability distribution for them from literature review. Based on the probability distribution for the uncertain parameters, we performed Markov Chain Monte Carlo (MCMC) simulation by using to obtain the production scenario for many cases and reservoir conditions. Following this step, we applied PCA to gas rate, cumulative gas, and CGR to obtain the respective PC coefficients and scores. Cross-validation was performed to check the predictive ability of the model generated. CMG-GEM(v2015.10), CMG-CMOST(v2015.10) and Matlab(vR2015a) was used to perform reservoir simulation, MCMC simulation and PCA calculation respectively. The simulated case is shown in Section 4. The truncated Matlab code, enough to replicate the procedure for simulated data is given in Appendix A.

Following this analysis, publicly available monthly field production data from the Eagle Ford Shale was obtained from DI Desktop by specifying certain constraints such as the well direction, starting production date, and primary fluid type (oil/gas). We checked the quality of the extracted data to ensure that only meaningful data was retained for analysis. The initial production time for each well was normalized so that production started from the same month. Similar to the simulated case, cross-validation was performed to analyze the predictive ability of the model. The field case is shown in Section 5.

6.6 Simulated Case

Reservoir, geological and completion data for the Eagle Ford Shale was obtained from a literature survey (Mullen, 2010; Orangi et al., 2011; Nagarajan, 2013; Kumar, 2013; Gong et al., 2013). We identified the uncertainty parameters and defined distributions for those parameters (Gong et al., 2013; Khanal et al., 2015; Khoshghadam et al., 2015). These uncertain parameters have a major impact in the overall production performance of a reservoir. The uncertain parameters studied for the Monte-carlo simulation were fracture half length, maximum fracture permeability, fracture spacing, minimum bottom-hole pressure, matrix permeability, porosity, and initial reservoir pressure.

6.7 Parameter Bounds and Distribution

A wide range of natural phenomena in biological, and physical sciences have been found to be closely approximated by normal and log-normal distributions (Ringrose and Bentley, 2015). These probability distributions are also

widely used in geostatistical reservoir modelling to account for uncertainties and heterogeneities in a reservoir (Pyrzcz and Duetsch, 2002). The geological properties like permeability and porosity in most reservoirs also appear to exhibit log normal distributions (Engler, 2010). Thus, the naturally occurring (uncertainty) parameters such as porosity, permeability and initial reservoir pressure are assumed to follow the log-normal distribution (Engler, 2010; Gong et al., 2013). The (decision) parameters selected by an operator like fracture spacing, half-length are selected based on uniform probability distribution constrained by values normally seen in the Eagle Ford Shale in the liquid rich regions(Kumar, 2013; Gong et al., 2013). The parameters and their distribution are shown in Table 17.

Table 17. Parameters, distribution and description

Parameter	Distribution	Values
Fracture Half-Length(f_{hl})	Uniform	200 ft to 350 ft
Fracture Permeability(F_k)	Lognormal	Mean = 50 md; SD= 20 md
Fracture Spacing(f_s)	Uniform	14 fracs to 24 fracs
Minimum BHP(BHP)	Uniform	500 psia to 1500 psia
Matrix Permeability(k)	Lognormal	Mean = 450nd; SD =170nd
Porosity(ϕ)	Lognormal	Mean = 7.4%; SD = 2.1%
Initial Pressure(p)	Lognormal	Mean =6200 psia; SD = 1870psia

6.8 Reservoir Description and Simulation

Before constructing the base case reservoir simulation, geological and completion data for the Eagle Ford Shale was obtained from a literature survey. The average lateral in the Eagle Ford was found to be greater than 5,000 ft. with 10-17 hydraulic fracture stages per lateral. Depth ranged from 4,000 ft. to 14,000 ft. and thickness ranged from 100 to 300 ft. This play has three distinct fluid regions, dry gas, condensate and oil from South to North. The porosity and

permeability range up to 11% and 1,200 nanodarcies respectively. The best part of the play has the potential to deliver an initial rate of 1,000 bbl/d of oil and up to 5 MMscf/d of gas. Based on these data from the literature, we modeled a 640-acre section of a reservoir with 5280 ft. long horizontal laterals and 4 laterals per section. Each well has 20 bi-wing fractures (fracture spacing of 264 ft.) which were gridded with logarithmically-spaced local grid refinement (LS-LGR) to account for rapidly changing pressure and composition around the wellbore. The reservoir temperature for Peng-Robinson Equation of State (EOS) was assumed to be around 186 °F, which was tuned by changing the binary interaction parameters to match the reported saturation pressure, and condensate to gas ratio. The initial and saturation pressure for the fluid was 9500 psia and 4729 psia respectively. Separate sets of relative permeability curves were used for matrix and hydraulic fractures using Corey exponents published in literature (Nagarajan, 2013). Additional reservoir attributes are described in Khanal et al. (2015). A single symmetric hydraulic fracture shown in Figure 58 was simulated for the period of 30 years.

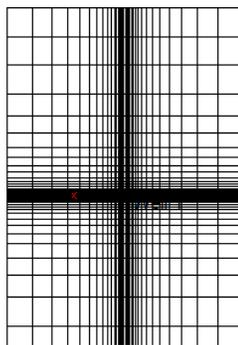


Figure 58. Symmetric element with logarithmically spaced grids for base case simulation

Some of the reservoir properties used for the base case simulation is summarized in Table 18:

Table 18. Summary of reservoir property used for the base case

Simulated Drainage Area	3 acres (200 ft. by 660 ft.)
Number of Fractures	20 (1 fracture simulated)
Fracture Half Length	264 ft.
Fracture Conductivity	2,000 md.ft.
Reservoir Thickness	100 ft.
Matrix Porosity	0.07
Matrix Permeability	200 nd.
Initial Gas Saturation	0.60
Initial Water Saturation	0.40

6.9 Principal Component Analysis of the Simulated Data

Using the pre-defined probability distributions, we performed Monte Carlo to generate 335 compositional simulations. We used the SVD algorithm to calculate the orthonormal matrices U , V and the singular values S . The product of matrices U and S yields the Principal Component Scores and the matrix V represents the Principal Component coefficients (or loadings). As mentioned earlier, only a few principal components and principal component scores are sufficient to adequately represent a large matrix. There is no certain rule on how many principal components to retain: however, a useful rule of thumb is to retain enough singular values to constitute 95% of the sum of squares of all the singular values (Shlens, 2003). Figure 59 shows the cumulative gas, gas rate, and CGR

for all the simulated cases. Depending on the input for simulation, wide ranges data are observed for each case. Although the CGR profile is erratic compared to the rate profiles, all profiles exhibit a distinct pattern applicable to a majority of the data.

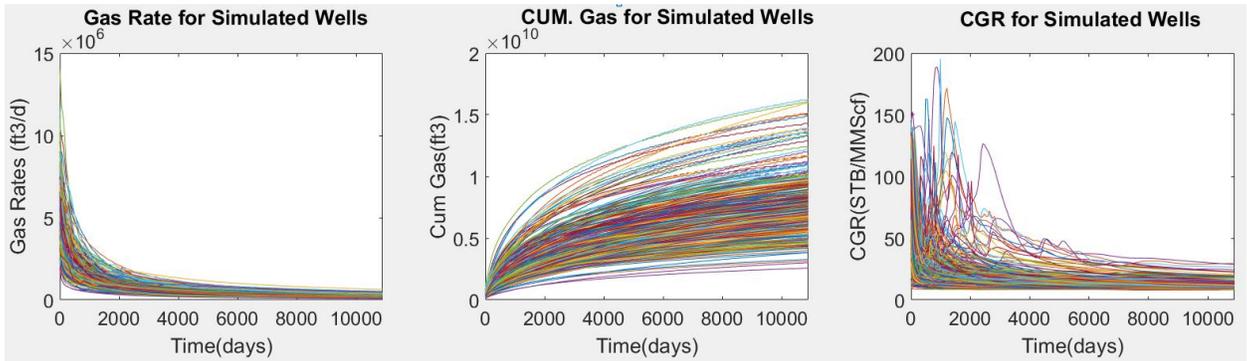


Figure 59. Simulated production data: Left: Gas Rate Middle: Cumulative Gas Right: CGR

Principal component analysis was performed on gas rate, cumulative gas, and CGR by using SVD to obtain eigenvalues in descending order as shown below. Figure 60 shows that CGR has significantly larger number of prominent eigenvalues compared to gas rate and cumulative gas. The number of eigenvalues that accounts for 90% of weights is 4, 1, and 8 for gas rate, cumulative gas and CGR respectively. The sum of squares of the singular values retained should be at least 90% of the sum of the squares of all the singular values, also referred to as energy (Rajaraman & Ullman 2014). In all cases, this value was over 95% with just two eigenvalues.

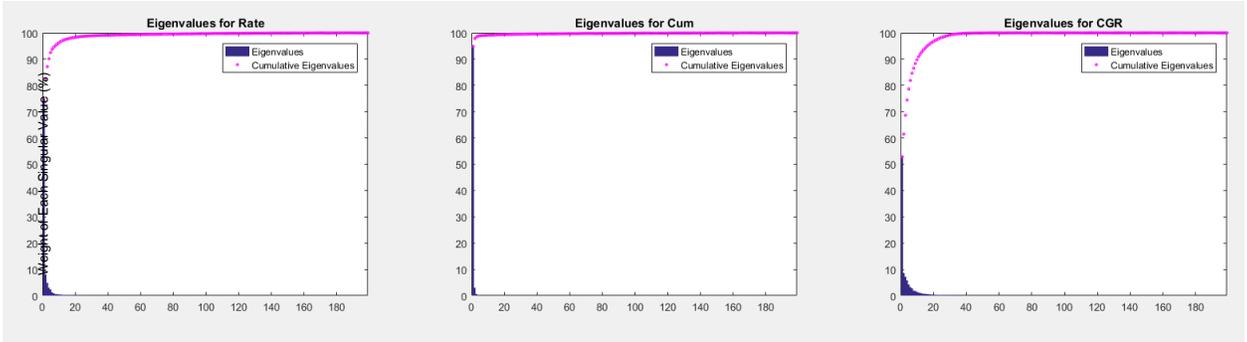


Figure 60. Eigenvalues for gas rate, cumulative gas and CGR

The PC coefficients for gas rate, cumulative gas and CGR are shown in Figure 61, which indicates that the first PC depicts the major pattern for each of the rates. The first PC for CGR also shows the slight kink observed for the simulated data where CGR falls to a local minimum before rising and then falling to an almost constant value after about 35 months. Other principal components further add to the characteristic profiles seen for each of the rates in the original data.

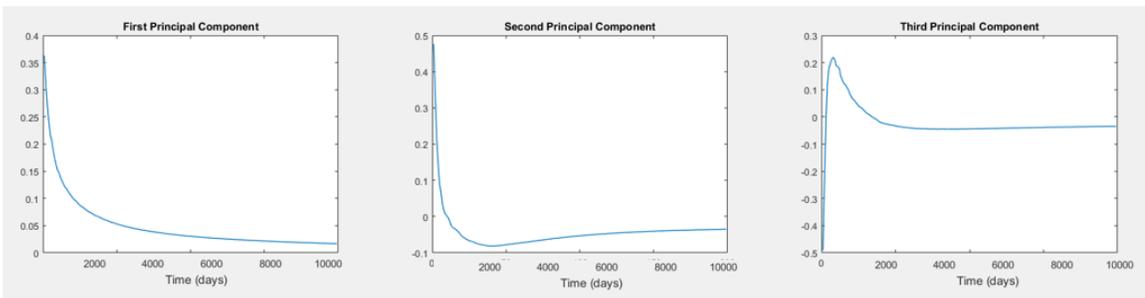


Figure 61. Principal component coefficient/loadings for gas rate

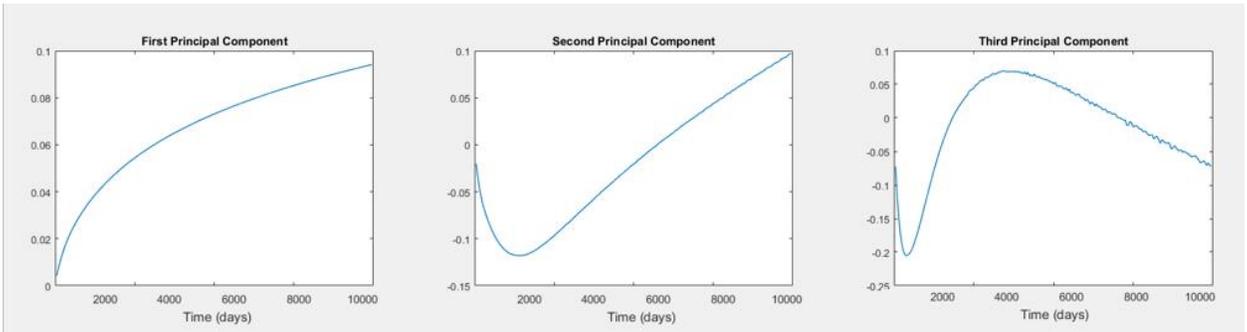


Figure 62. Principal component coefficient/loadings for cumulative gas

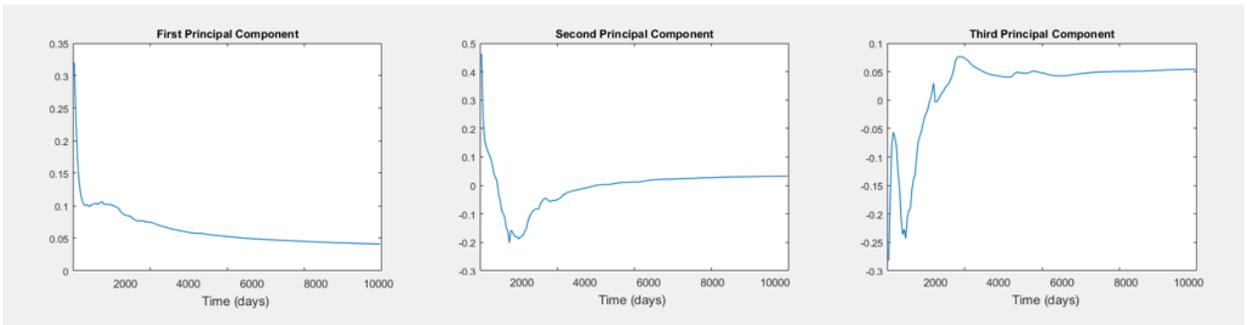


Figure 63. Principal component coefficient/loadings for CGR

6.10 Recreation of Original Data with Limited Principal Components

Original profiles were recreated for gas rate, cumulative gas and CGR by using the limited number of PC coefficients. The error for each case was calculated by using Equation 6.11 shown above. The average and the maximum error for each case are shown in Table 19 and Table 20 respectively. The minimum error for each case was observed to be less than 1%.

Table 19. Average error in recreation for each case

	1 PC	2 PC	3 PC	4 PC
Gas				
Rate	1.78	0.68	0.32	0.14
Cum.				
Gas	0.11	0.01	0.00	0.00
CGR	5.23	3.88	1.89	0.82

Table 20. Maximum error in recreation for each case

	1 PC	2 PC	3 PC	4 PC
Gas				
Rate	8.18	3.84	2.45	1.66
Cum.				
Gas	0.63	0.08	0.01	0.00
CGR	25.0	22.8	11.3	8.76

On closer examination of error for each case, it was seen that maximum and minimum error for a particular number of principal components varied among the wells. For this reason, the maximum and minimum error was identified by averaging the errors for each principal component case. The maximum average error was identified as 3.1% for well 301 shown in Figure 64. Similarly, the minimum error of 0.05% was identified for well 146 shown in Figure 65.

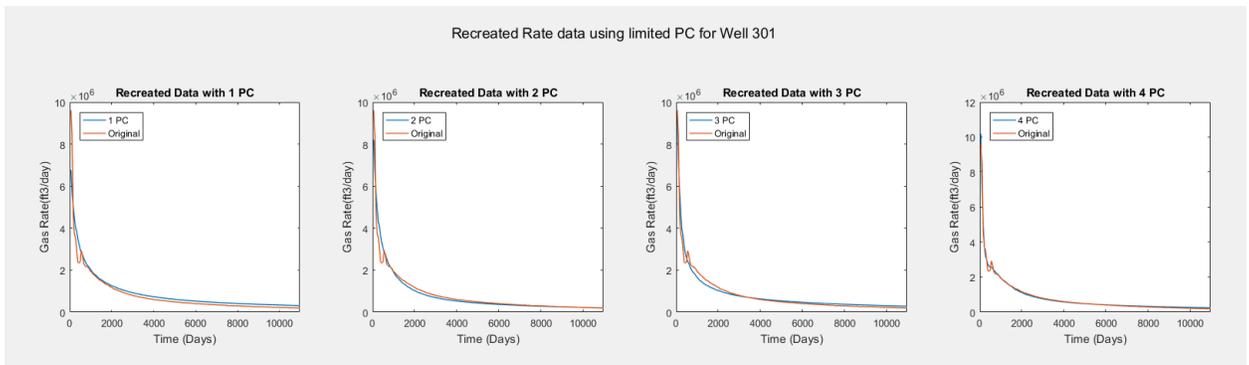


Figure 64. Maximum error seen from recreation of original data with limited principal components (gas rate)

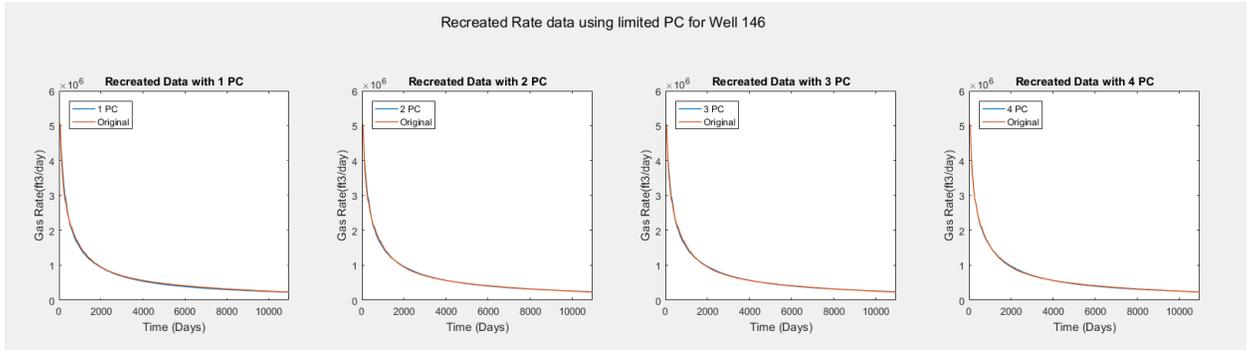


Figure 65. Minimum error seen from recreation of original data with limited principal components (gas rate)

The maximum and minimum error for cumulative gas was 0.16% for well 315 and close to 0(7E-04) for well 66 as shown in Figures 66 and 67 respectively.

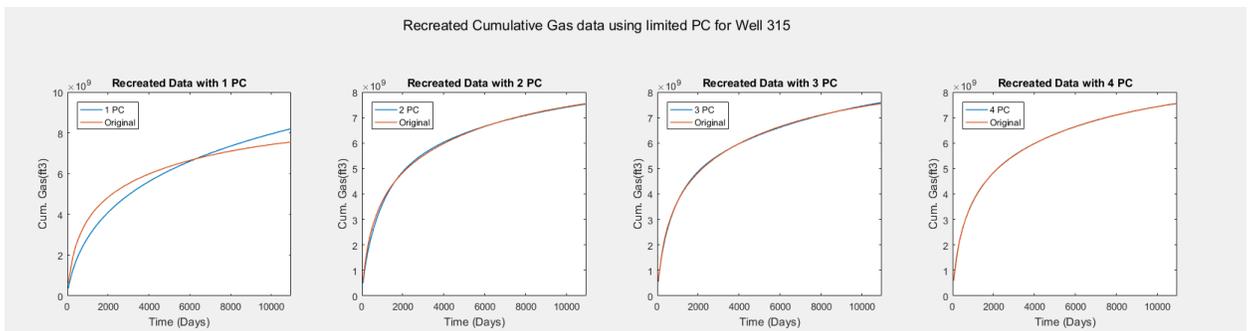


Figure 66. Maximum error seen from recreation of original data with limited principal components (Cumulative Gas)

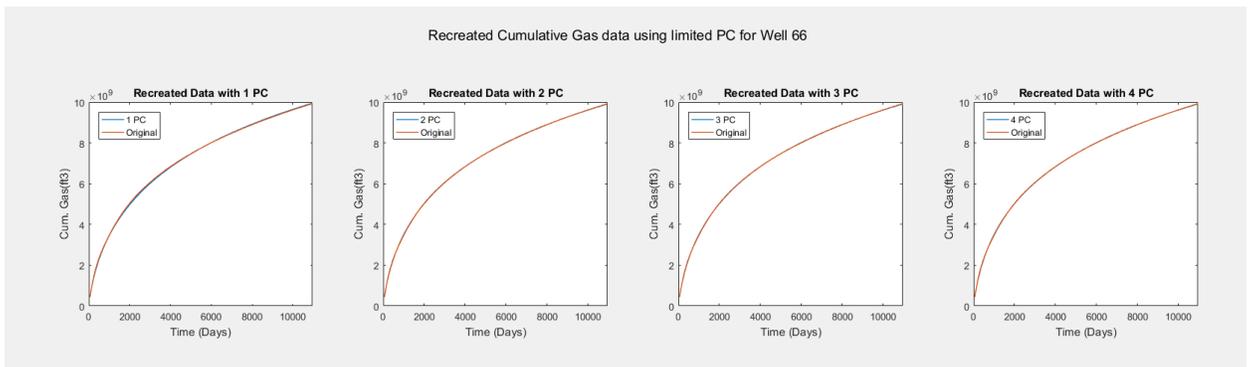


Figure 67. Minimum error seen from recreation of original data with limited principal components (Cumulative Gas)

The maximum and minimum error for CGR was 15.65% for well 288 and 0.34 for well 42 as shown in Figures 68 and 69 respectively.

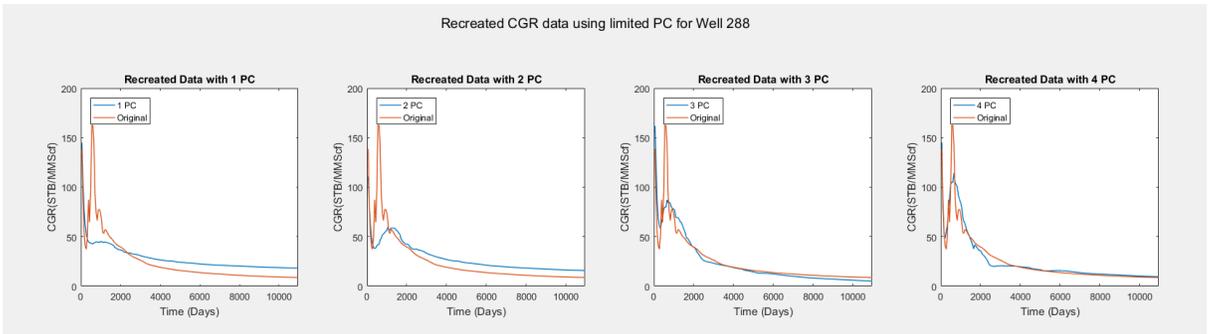


Figure 68. Maximum error seen from recreation of original data with limited principal components (CGR)

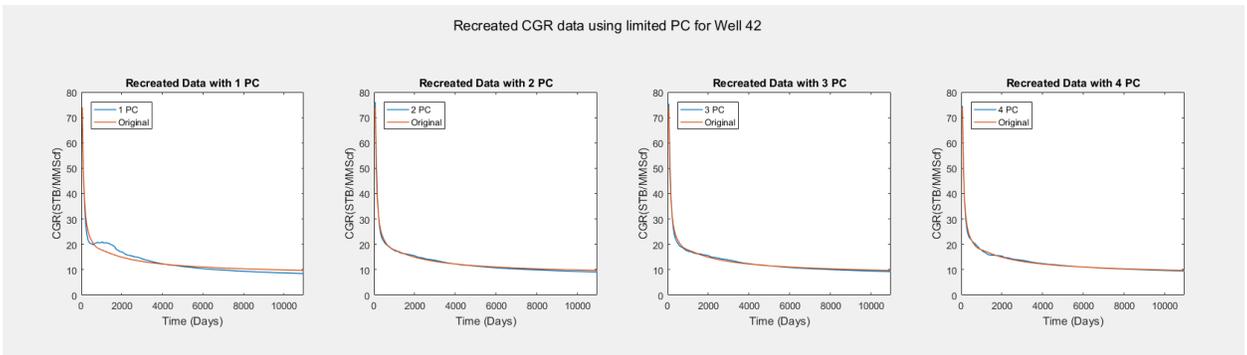


Figure 69. Minimum error seen from recreation of original data with limited principal components (CGR)

Figures 64 through 69 indicate that using only one PC yields a relatively poor result for all cases. The error in recreation is alleviated when the number of principal components are increased. Figures 64 through 69 also show that for the cases with maximum error, the simulated data are usually erratic and show sudden changes in the rates. Similarly, for the cases with minimum error, the simulated data are smooth and follow the overall trend of the first principal component.

Error for each of the simulated wells is shown in Figure 70. Error was as large as 25% for CGR, 8.5% for gas rate, and 0.7% for cumulative gas when only

one PC was used, also shown in Tables 19 and 20. Cumulative gas shows a good result just with one PC whereas gas rate needs at least two PC's to reduce the error to less than 5%. However, CGR requires at least four PC's for the error to be less than 5%. From this analysis, we conclude that, if data (field or simulated) contains fluctuations, noise or other variations, a larger number of PC's are required to re-create the original data. However, using too many PC component could lead to high variance due to overfitting. So, it is important for an engineer to identify the optimal number of principal components based on the noise and complexity of the field data.

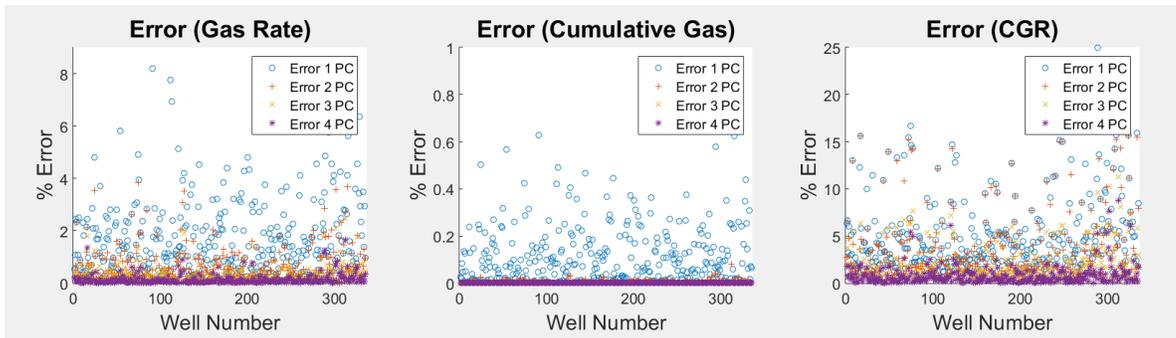


Figure 70. Error in recreation each number of principal components

6.11 Analysis for Normalized (or Standardized) Data

Normalization is performed by dividing each production value by its maximum value. This reduces the maximum value of a data point to one for all cases. If we consider gas rates 'X' and CGR 'Y' from section 2.2 again to represent gas rates and CGR, normalized rate data can be written as follows:

$$\bar{X} = \frac{q_{ij}}{q_{max}} \text{ and} \tag{6.12}$$

$$\bar{Y} = \frac{c_{ij}}{c_{max}}, \tag{6.13}$$

where q_{\max} and c_{\max} represent the maximum gas rate and CGR, respectively.

Normalization is extremely useful during cross-validation. Principal component scores for the testing set is calculated by minimizing the least squares of errors (LSE) which requires an initial guess for the unknown value (PC scores). When a normalized data set is used, the initial guess can be easily selected as it lies between 0 and 1. This allows for the least-squares algorithm to be used for all cases regardless of their magnitude with the same initial guess. If a dataset is not normalized, the initial guess needs to be properly selected as it could lead to local minima instead of global minima. The eigenvalues, PC's, and other values for normalized data exhibit trend similar to the trends in actual data. Average Error and maximum error in recreation for normalized cases are shown in Table 21 and Table 22 respectively. Comparing with the non-normalized cases above in Tables 19 and 20, normalized samples show lower error when more than 1 PC is used.

Table 21. Average error in recreation for each case (Normalized Sample)

	1 PC	2 PC	3 PC	4 PC
Gas Rate(N)	1.81	0.63	0.31	0.14
Cum. (N)	0.11	0.01	0.00	0.00
CGR (N)	7.06	2.18	1.15	0.66

Table 22. Maximum error in recreation for each case (Normalized Sample)

	1 PC	2 PC	3 PC	4 PC
CGR (N)	34.98	23.06	17.88	9.29
Cum. (N)	0.66	0.08	0.01	0.00
Gas Rate(N)	9.40	3.38	2.35	1.50

6.12 Prediction of New Well Data

Production of oil and gas from gas condensate reservoirs is affected by geological, fluid, and completion properties. Each property has a varying effect on how much and how fast oil is produced from a reservoir. Several authors (Khoshghadam et al., 2015) have studied the effects of these properties on underlying production mechanisms in liquid-rich shale reservoirs. For this study, we used multiple linear regression with a 95% confidence interval to generate the relationship of PC scores and the reservoir properties studied. Higher-order interactions are excluded as they add large number of variables ($O(n^2/2)$ for 2nd order, $O(n^3)$ for 3rd order) without adding significant insight to the result. Important higher order interactions can be selectively chosen in future studies by using techniques like design of experiments (DOE) or artificial neural networks (ANN). Several other reservoir and geological properties should be accounted for, as more data become available through experimental and field studies.

In this study, we regressed on seven variables (centered by mean and normalized by standard deviation) so the output of the regression as

$$PC_i = \beta_0 + \beta_1 \widehat{f_{hl}} + \beta_2 \widehat{F_k} + \beta_3 \widehat{f_s} + \beta_4 \widehat{bhp} + \beta_5 \widehat{k} + \beta_6 \widehat{\phi} + \beta_7 \widehat{p}, \quad (6.14)$$

where β_i represents the standardized regression coefficients for the standardized variables.

Table 23 summarizes the output from multiple linear regression. It shows that all the variables are significant to explain PC_i . The R^2 value of 92% indicates a good fit of the variables

Table 23. Results from multiple regression for simulated data

<i>Regression Statistics</i>		ANOVA					
Multiple R	0.96						
R Square	0.92	Regression	7	3.76E+15	5.37E+14	393	1.02E-122
Adjusted R Square	0.92	Residual	226	3.09E+14	1.37E+12		
Standard Error	1169098	Total	233	4.07E+15			
Observations	234						

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>
Intercept	630075	1009259	0.62	0.53	-1358686	2618836
Frac_HL	-36008	1844	-19.53	0.00	-39641	-32375
Frac_Perm	-180	50	-3.59	0.00	-279	-81
Min_BHP	2658	269	9.89	0.00	2128	3187
PERMI	-10077228728	396713763	-25.40	0.00	-10858959642	-9295497814
POR	-83677799	3698980	-22.62	0.00	-90966700	-76388898
Frac_Spcing	37977	1807	21.01	0.00	34415	41539
Init_Pressure	-919	44	-20.86	0.00	-1005	-832

Multiple linear regression was performed for each of the PC scores deemed to be important to re-create the original dataset. The results from ANOVA was used to determine the upper and lower 95% confidence interval for each factor. In the presence of geological and completion properties, we can forecast production from new wells by using the relevant PC coefficients from the existing data and scores calculated from regression. In absence of these field data, the PC scores could be estimated by minimizing the LSE or the available historical data. Cross-validation was performed by using the leave-one-out method to check the predictive ability of the model. Error for each cross-validated case is shown in Figure 71, which indicates that the validation error for CGR is as large as 35% with only one PC and is less than 10% after using 3 PC's. For gas rate, two PC's was sufficient to reduce the error to less than 5%.

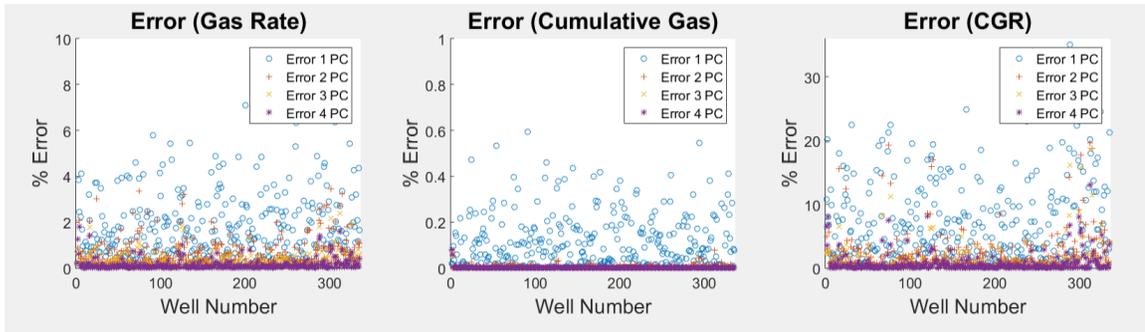


Figure 71. Cross-validation error/prediction error for gas rate, cumulative gas and CGR

Table 24 and Table 25 show the average and maximum validation errors for each normalized simulated case. Average validation error for each of the cases are below 2.21% when 2 PC are used which shows an excellent predictive capability. The maximum error for normalized gas rate and cumulative gas using 2 PC is less than 4%. However, the maximum error for CGR remains as high as 13% even when 4 PC is used. The reason why CGR shows higher error compared to gas rate could be that producing CGR depends on both oil rate and gas rate, so errors are compounded when calculating the CGR. The cause could also be that oil rate (consequently CGR) depends on interaction of several parameters including maximum production rate, minimum bottom-hole pressure, and permeability not accounted for in multiple linear regression model used in this study. This could cause some problems when used to forecast noisy or highly fluctuating data. This can potentially be alleviated by gathering more samples for analysis.

Table 24. Average validation (prediction) error

	1 PC	2 PC	3 PC	4 PC
Gas				
Rate(N)	1.82	0.64	0.32	0.16
Cum.(N)	0.11	0.01	0.00	0.00
CGR(N)	7.09	2.21	1.19	0.77

Table 25. Maximum validation (prediction) error

	1 PC	2 PC	3 PC	4 PC
Gas				
Rate(N)	9.42	3.44	2.39	1.79
Cum.(N)	0.66	0.08	0.08	0.08
CGR(N)	35.1	23.5	18.59	13.1

6.13 Field Example

Simulated data showed that two PC's can predict trends of several dynamic production metrics of a well with reasonable accuracy. Although in some cases the error could be as high as 23%, the average error is less than 5%. Reservoir and geological properties were not reported for any of the cases so LSE was used to cross-validate and calculate the PC scores for test well data.

6.14 Data Acquisition

We used DI Desktop (DrillingInfo, Version 6.2.1) to obtain production data from La Salle County (Eagle Ford shale). Horizontal "gas" wells classified with first production data from 1/1/2008 with at least 60 months of production history were selected as DI Desktop does not further classify gas wells as condensate wells nor oil wells as volatile oil wells (Figure 72).

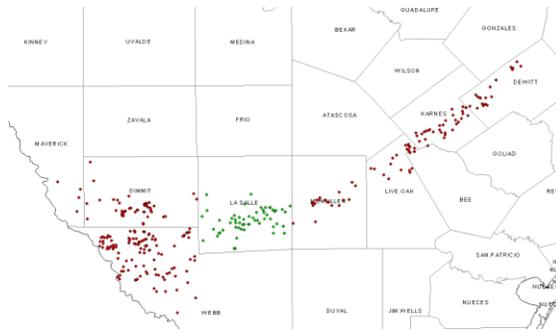


Figure 72. Wells from LaSalle County selected for analysis (shown in green)

The maximum CGR of the wells selected was 189 STB/MMscf (Gas to oil ratio (GOR) = 5,282 Scf/STB) and the minimum was zero (dry gas wells). The initial production date for each well was normalized so that production starts from the same time zero. Additionally, data was manually preprocessed to remove sudden increases or decreases in rate. Months with production rate of zero were also removed from the data. After these preprocessing steps, 46 wells were identified with 52 months of monthly production data for gas and oil. Producing GOR (or CGR) can be calculated by using the reported gas and oil rate data. The monthly gas rate for each well is shown in Figure 73 Oil rates were also analyzed. Of 52 cases, 24 cases showed little to no oil production. In addition, 6 wells showed erratic production behavior and were excluded from the study.

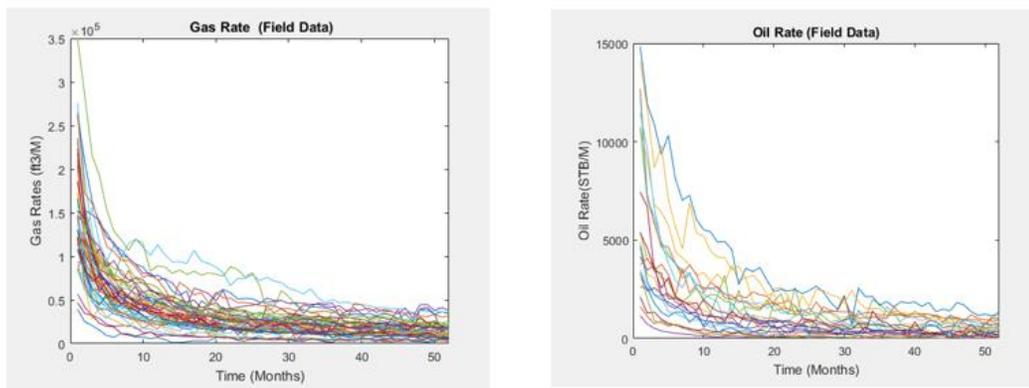


Figure 73. Field gas and oil data

6.15 Principal Component Analysis for Field Case

Pre-processed gas and oil rate data was subjected to PCA by creating 46 by 52 and 21 by 52 matrices for gas and oil respectively (rows represent transposed rate data). Principal components for each gas and oil data set are shown in Figures 74 and 75, which indicate the first PC for both gas rate and oil rate accurately represents the sharp decline seen in unconventional reservoirs. Second and third PC's add characteristic features seen in the final data.

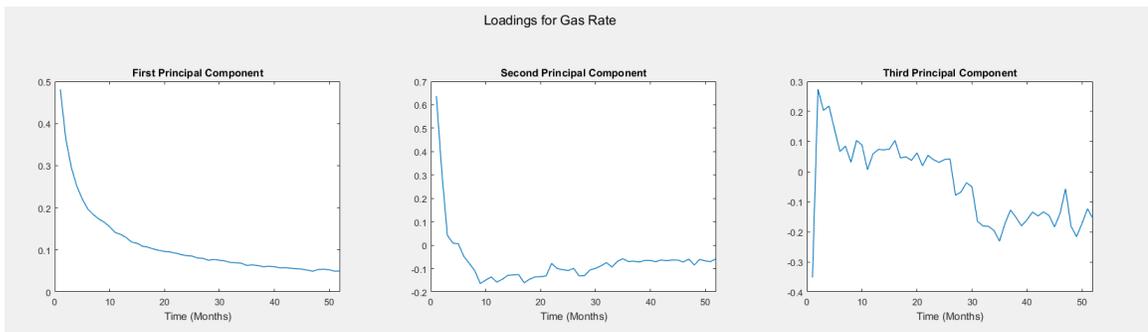


Figure 74. Principal components for gas rate

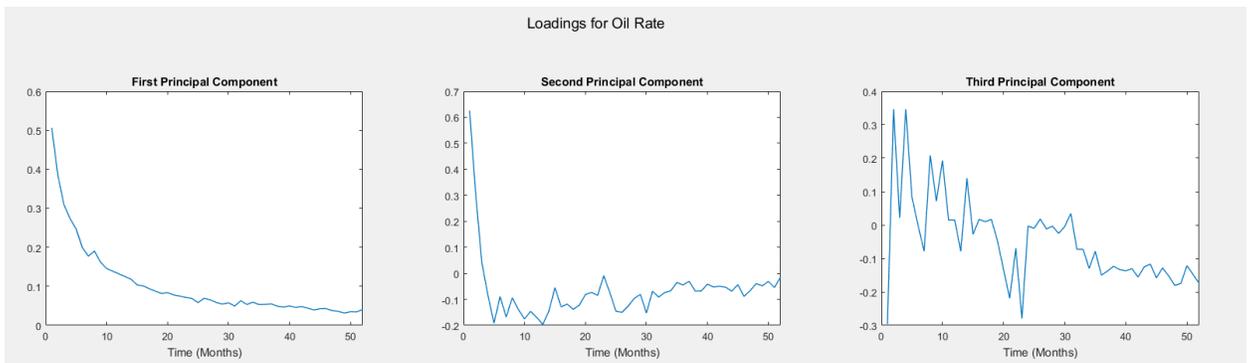


Figure 75. Principal components for oil rate

Eigenvalues for both gas rate and oil rate are shown in Figure 76 Field data, unlike simulated data, is noisy, which results in weights of subsequent

eigenvalues having higher values than for simulated cases. Still, first and second eigenvalues accounted for over 95% of energy as defined earlier.

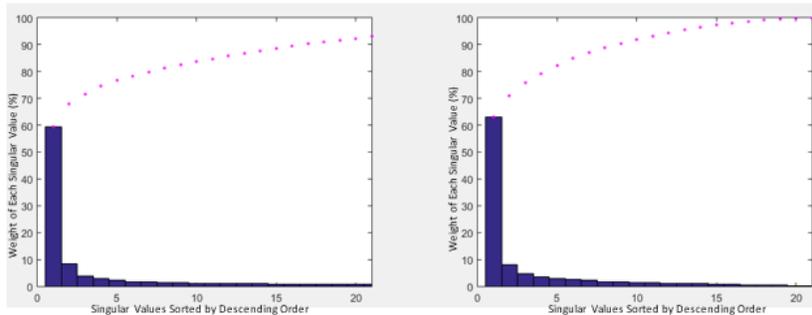


Figure 76. Eigenvalues (blue) and cumulative eigenvalues (magenta) for gas rate and oil rate

6.16 Recreation of Original Data with Limited Principal Components

Oil and gas rate data for each well was re-created using a limited number of principal components. Figure 77 and 78 show the maximum and minimum error of 5.1% and 0.29% respectively in recreation for gas rate data. Similarly, Figure 79 and 80 show the maximum and minimum error of 7.01% and 0.78% respectively in recreation of oil rate data. Figure 77 through 80 show that quality of re-creation of the original data increases significantly with increases in the number of PC's. It is also evident that unexplained fluctuations cannot be recreated by using a limited number PC's, especially for oil rate.

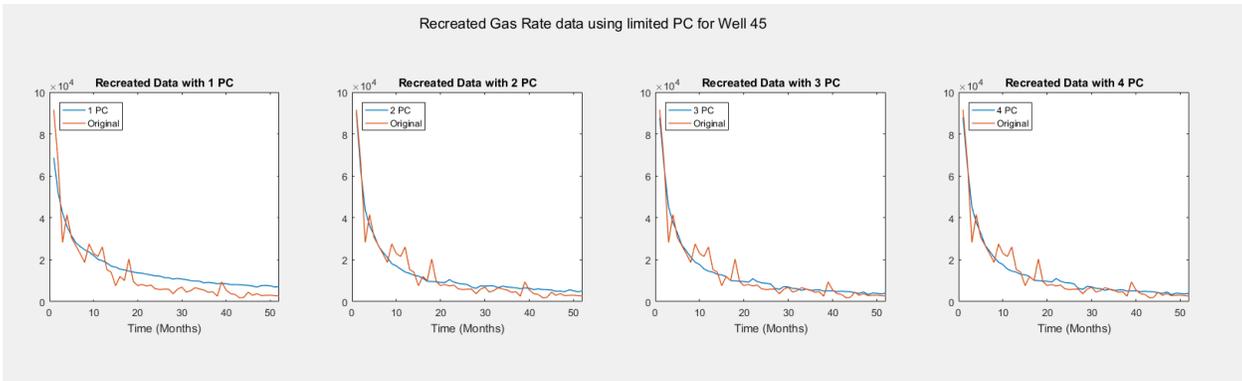


Figure 77 . Maximum error in recreation using limited number of PC (gas rate)

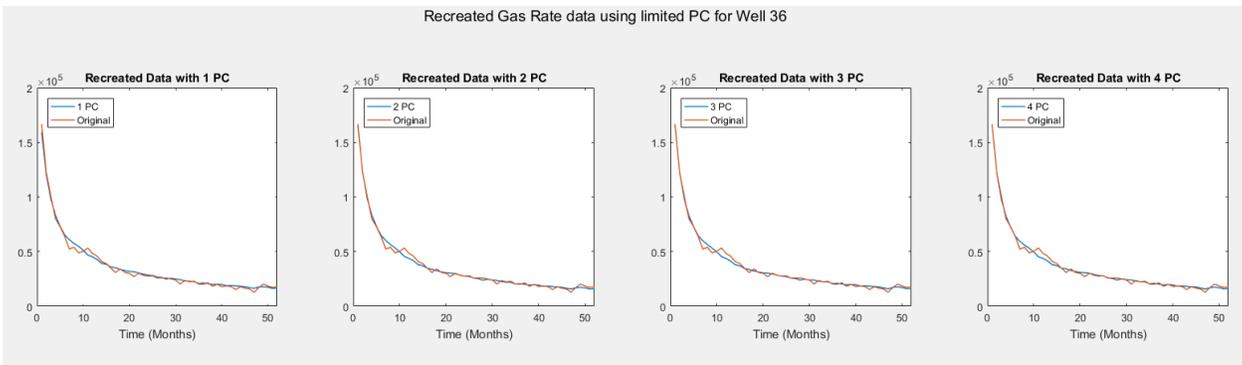


Figure 78. Minimum error in recreation using limited number of PC (gas rate)

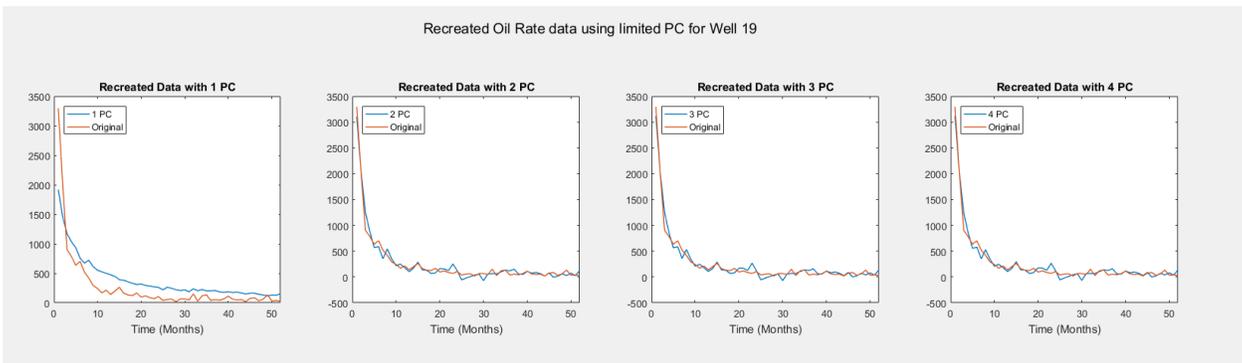


Figure 79. Maximum error in recreation using limited number of PC (oil rate)

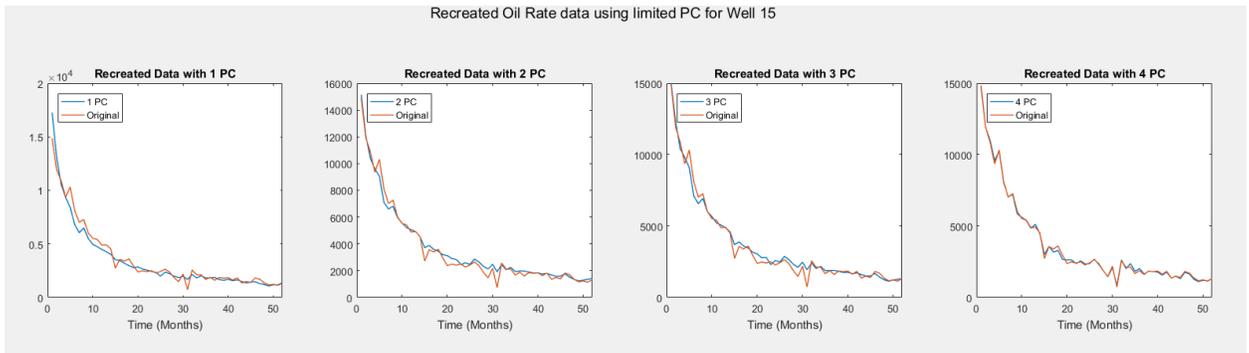


Figure 80. Minimum error in recreation using limited number of PC (oil rate)

The average and maximum error in recreation for field data is shown in Table 26 and Table 27 respectively. It can be seen that the error for gas rate is significantly lower compared to oil rate. This is due to the field oil rate data for condensate reservoirs being erratic and noisy as seen in Figure 73.

Table 26. Average error in recreation for each case

	1 PC	2 PC	3 PC	4 PC
Gas				
Rate	3.36	1.57	1.19	1.01
Oil Rate	5.52	2.58	2.06	1.82

Table 27. Maximum error in recreation for each case

	1 PC	2 PC	3 PC	4 PC
Gas				
Rate	14.5	4.2	3.8	3.8
Oil Rate	21.7	7.7	5.5	5.0

Error in re-creation of the field data for the entire set of simulated data is shown in Figure 81. Oil rates have larger errors than gas rates due mainly to large fluctuations in production rates. Excepting a few outliers, gas rate data for each well is re-created sufficiently accurately using at least two PC's.

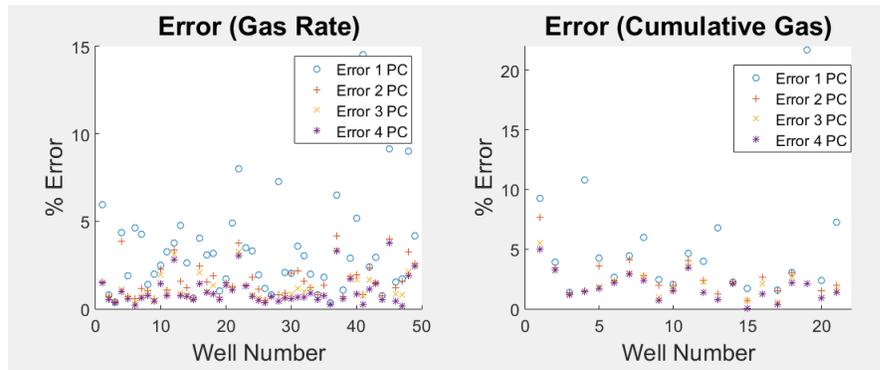


Figure 81. Error in re-creation of field data using various number of PC's
Gas and oil rates were normalized using Equation 6.12 and 6.13. This

step is especially important for field data as PC scores cannot be calculated by multiple linear regression, since values of field parameters are not known. The average and maximum error in recreation for normalized data is shown in Table 28 and 29. The error for normalized data is comparable to un-normalized data as seen in Table 26 and Table 27.

Table 28. Average error in recreation for each case (normalized data)

	1 PC	2 PC	3 PC	4 PC
Gas				
Rate	3.52	1.56	1.13	0.95
Oil Rate	5.63	2.28	1.79	1.49

Table 29. Maximum error in recreation for each case (normalized data)

	1 PC	2 PC	3 PC	4 PC
Gas				
Rate	13.4	4.2	3.8	3.1
Oil Rate	23.7	4.6	3.4	3.2

6.17 Prediction of New Well Data

Principal component scores were calculated by using LSE for validation and prediction. Cross-validation was performed with the leave-one-out method to check the predictive ability of the model. Errors for each cross-validated case are

shown in Figure 82. Validation error for CGR was as high as 25% with only one PC and was below 10% only after using 3 PC's. For gas rate, two PC's were sufficient to reduce the error to less than 5%. A few cases showed larger errors even when three or four PC's were used. This could be due to a drastically different production profile for these wells compared to the whole set.

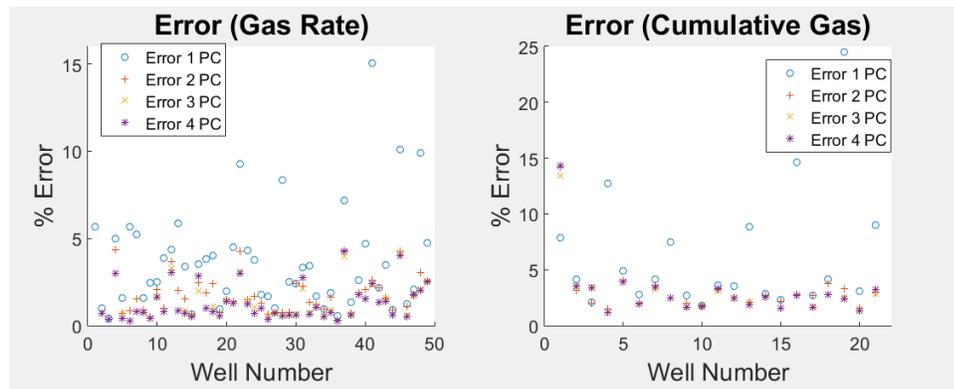


Figure 82. Cross-validation/prediction error for field data

The average and maximum error in recreation is shown in Table 30 and Table 31 respectively. The average validation error for cases with just two PC's was less than 4% with the maximum being 14 % for oil rate. The average and maximum validation error for gas rates with two PC's was 2% and 17% respectively. For each gas rate and oil rate, 3 PC yielded the best prediction with the least error. It can be seen that 4 PC, actually yields a slightly worse result for both gas rate and oil rate. This could be due to overfitting, which happens when a model performs well for the test set (recreation) but doesn't perform well for prediction (Shlens, 2003). Thus, careful analysis much be performed to select the optimum number of principal components while using this method to predict the field data.

Table 30. Average error in validation for each case (normalized data)

	1 PC	2 PC	3 PC	4 PC
Gas				
Rate	3.68	2.02	1.63	1.62
Oil Rate	6.19	3.16	2.98	3.02

Table 31. Maximum error in validation for each case (normalized data)

	1 PC	2 PC	3 PC	4 PC
Gas				
Rate	15.0	17.2	17.1	17.5
Oil Rate	24.5	14.2	13.4	14.4

These results show that field data can be used to identify and forecast production for similar wells in the same play. This process could be used iteratively to obtain more refined results as more field data becomes available. This process can also provide a quick and easy alternative to reservoir simulation and add to the results from decline curve analysis (DCA) and rate transient analysis (RTA). As mentioned earlier, the PCA method should be used in addition to, not instead of other methods to quickly screen and forecast wells with similar characteristics.

6.18 Conclusions

From this study we can conclude that principal component analysis can be used successfully to identify and forecast production for wells with similar historical production profiles. It can also be used to identify wells with unique production profiles compared to the majority of the wells in a play. When extensive reservoir parameter and completion data are available, multiple linear regression can be used to calculate PC scores to forecast and re-create production data for new wells. In this work for simulated data, four PC was enough to yield the

prediction with average error of 0.16%, 0% and 0.77% respectively for gas rate, cumulative gas and CGR respectively.

However, for field data, three PC yielded the best prediction with average error of 1.63% and 2.98% for gas rate and oil rate respectively. We can also conclude that noisy and highly fluctuating production data requires a larger number of PC's for forecasting as evidenced by error for CGR. However, up to two PC's can reasonably recreate the expected production profile for a smoothly varying data. DCA, RTA and PCA are able to predict the performance of gas condensate reservoirs with varying degree of accuracy but they do not reproduce the forecasts from robust reservoir simulation.

References

Ahmed, K. (2005). Phd Dissertation. "Volatile Oil and Gas Condensate Reservoir Fluid Behavior." Cairo: Cairo University.

Arps, J. J. (1945, December 1). *Analysis of Decline Curves*. Society of Petroleum Engineers. doi:10.2118/945228-G

Brown, M.L. (2009). "Analytical Trilinear Pressure Transient Model for Multiply Fractured Horizontal Wells in Tight Shale Reservoirs" M.Sc. Thesis, Colorado School of Mines, Golden.

Bhattacharya, S., & Nikolaou, M. (2011, January 1). *Using Data from Existing Wells To Plan New Wells in Unconventional Gas Field Development*. Society of Petroleum Engineers. doi:10.2118/147658-MS

Bhattacharya, S., & Nikolaou, M. (2013, June 6). *Analysis of Production History for Unconventional Gas Reservoirs With Statistical Methods*. Society of Petroleum Engineers. doi:10.2118/147658-PA

Box G.E., Hunter J.S. and Hunter W. G. 2005. Statistics for Experimenters, John Wiley & Sons.

Chen, C., Gao, G., Ramirez, B. A., Vink, J. C., & Girardi, A. M. (2016, April 1). *Assisted History Matching of Channelized Models by Use of Pluri-Principal-Component Analysis*. Society of Petroleum Engineers. doi:10.2118/173192-PA

Chopra, S., & Marfurt, K. J. (2014, October 29). *Churning Seismic Attributes With Principal Component Analysis*. Society of Exploration Geophysicists.

CMG Release 2015.10, Computer Modelling Group Ltd., Calgary, Alberta, Canada

CMG, 2013. CMOST User's Guide. Computer Modelling Group Ltd.

Dadashpour, M., Rwechungura, R. W., & Kleppe, J. (2011, January 1). *Fast Reservoir Parameter Estimation By Using Effect Of Principal Components Sensitivities And Discrete Cosine Transform*. Society of Petroleum Engineers. doi:10.2118/141913-MS

Dake, L. P. (1978). Fundamentals of reservoir engineering. Amsterdam: Elsevier Scientific Pub.

Denney, D. (1999, October 1). *Management of Well Production With Real-Time Expert Systems*. Society of Petroleum Engineers. doi:10.2118/1099-0034-JPT

Deutsch, C. V. (2002). Geostatistical reservoir modeling. Oxford: Oxford University Press.

Duong, A. N. (2011, June 1). *Rate-Decline Analysis for Fracture-Dominated Shale Reservoirs*. Society of Petroleum Engineers. doi:10.2118/137748-PA

EIA. (November 2015), "U.S. Crude Oil and Natural Gas Proved Reserves, 2014"

Fan, L., Harris, B. W., Jamaluddin, A., Kamath, J., Mott, R., Pope, G. A., Whitson, C. H. (2005). "Understanding Gas-Condensate Reservoirs. *Oilfield Review*", 14-27.

Fathi, E., Elamin, A., & Ameri, S. (2013, April 10). *Simulation of Multicomponent Gas Flow and Condensation in Marcellus Shale Reservoir*. Society of Petroleum Engineers. doi:10.2118/164538-MS

Fetkovich, M. J., Fetkovich, E. J., & Fetkovich, M. D. (1996, February 1). *Useful Concepts for Decline Curve Forecasting, Reserve Estimation, and Analysis*. Society of Petroleum Engineers. doi:10.2118/28628-PA

Freeborn, R., & Russell, B. (2012, January 1). *How To Apply Stretched Exponential Equations to Reserve Evaluation*. Society of Petroleum Engineers. doi:10.2118/162631-MS

Fulford, D. S., Bowie, B., Berry, M. E., Bowen, B., & Turk, D. W. (2016, January 1). *Machine Learning as a Reliable Technology for Evaluating Time/Rate Performance of Unconventional Wells*. Society of Petroleum Engineers. doi:10.2118/174784-PA

Geisser, S. (1993). Predictive inference: An introduction. New York: Chapman & Hall.

Gong, X., Tian, Y., McVay, D. A., Ayers, W. B., & Lee, J. (2013, November 5). *Assessment of Eagle Ford Shale Oil and Gas Resources*. Society of Petroleum Engineers. doi:10.2118/167241-MS

Gray, H.E., 1978. Vertical flow correlation-gas wells. In User Manual for API 14b Subsurface Controlled Safety Valve Sizing Computer Program, Second Edition, Appendix B. API, Washington DC.

Guo, H., Marfurt, K. J., Liu, J., & Dou, Q. (2006, January 1). *Principal Components Analysis of Spectral Components*. Society of Exploration Geophysicists.

Holditch, S. A. (2006, June 1). *Tight Gas Sands*. Society of Petroleum Engineers. doi:10.2118/103356-JPT

Honorio, J., Chen, C., Gao, G., Du, K., & Jaakkola, T. (2015, September 28). *Integration of PCA with a Novel Machine Learning Method for Reparameterization and Assisted History Matching Geologically Complex Reservoirs*. Society of Petroleum Engineers. doi:10.2118/175038-MS

Jeff Schneider. Retrieved April 10, 2016, from "<https://www.cs.cmu.edu/~schneide/>"

Jolliffe, I. 2014. Principal Component Analysis. Wiley StatsRef: Statistics Reference Online.

Khanal, A., Khoshghadam, M., & Lee, W. J. (2015, September 2). *Optimal Decline Curve Analysis (DCA) Models for Liquid Rich Shale (LRS) Gas Condensate Reservoirs*. Society of Petroleum Engineers. doi:10.2118/175530-MS

Khanal, A., Khoshghadam, M., Lee, W., J., (2015, September 2). *Effect of Well Spacing on Productivity of Liquid Rich Shale (LRS) Reservoirs with Multiphase Flow: A simulation Study*. Manuscript submitted for publication

Khanal, A., Khoshghadam, M., & Lee, W. J. (2015, September 2). *Stochastic Study of Reservoir and Completion Parameters in Liquid Rich Shale*. Society of Petroleum Engineers. doi:10.2118/175529-MS

Khanal, A., Khoshghadam, M., Makinde, I., & Lee, W. J. (2015, October 20). *Modeling Production Decline in Liquid Rich Shale (LRS) Gas Condensate Reservoirs*. Society of Petroleum Engineers. doi:10.2118/175913-MS

Khoshghadam, M., Khanal, A., & Lee, W. J. (2015, September 28). *Numerical Study of Production Mechanisms and Gas-Oil Ratio Behavior of Liquid-Rich Shale Oil Reservoirs*. Society of Petroleum Engineers. doi:10.2118/175137-MS

Kumar, A., Dusterhoft, R. G., & Siddiqui, S. (2013, September 30). *Completion and Production Strategies for Liquids-Rich Wells in Ultra-Low-Permeability Reservoirs*. Society of Petroleum Engineers. doi:10.2118/166177-MS

Lalehrokh, F., & Bouma, J. (2014, September 30). *Well Spacing Optimization in Eagle Ford*. Society of Petroleum Engineers. doi:10.2118/171640-MS

Lee, W. J., & Sidle, R. E. (2010, January 1). *Gas Reserves Estimation in Resource Plays*. Society of Petroleum Engineers. doi:10.2118/130102-MS

Lin, Y., Wang, H., He, S., Nikolaou, M., 2015. *Increasing Shale Gas Recovery through Thermal Stimulation: Analysis and an Experimental Study*. SPE-175070-MS.

“MATLAB Release 2015a”, The MathWorks, Inc., Natick, Massachusetts, United States.

Mattar, L., & Anderson, D. (2005, January 1). *Dynamic Material Balance(Oil or Gas-In-Place Without Shut-Ins)*. Petroleum Society of Canada. doi:10.2118/2005-

113

Mohaghegh, S. D. (2009, October 1). *Artificial Intelligence and Data Mining: Enabling Technology for Smart Fields*. Society of Petroleum Engineers. doi:10.2118/0309-014-TWA

Myers R.H and Montgomery D.C. 2002. Response Surface Methodology: Process and Product.

Muskat, M. 1981. Physical Principals of Oil Production. IHRDC, Boston, MA.

Nagarajan, N. R., Honarpour, M. M., & Arasteh, F. (2013, August 12). *Critical Role of Rock and Fluid - Impact on Reservoir Performance on Unconventional Shale Reservoirs*. Society of Petroleum Engineers. doi:10.1190/URTEC2013-194

Optimization Using Design of Experiments, John Wiley & Sons. ISBN: 0-470-17446-3.

Orangi, A., Nagarajan, N. R., Honarpour, M. M., & Rosenzweig, J. J. (2011, January 1). *Unconventional Shale Oil and Gas-Condensate Reservoir Production, Impact of Rock, Fluid, and Hydraulic Fractures*. Society of Petroleum Engineers. doi:10.2118/140536-MS

Ozkan, E., Brown, M. L., Raghavan, R. S., & Kazemi, H. (2009, January 1). *Comparison of Fractured Horizontal-Well Performance in Conventional and Unconventional Reservoirs*. Society of Petroleum Engineers. doi:10.2118/121290-MS

Peng, D. Y., and Robinson, D. B. (1976). *A New Two-Constant Equation of State*. *Industrial and Engineering Chemistry: Fundamentals* 15: 59–64.

Rajaraman, Anand, and Jeffrey D. Ullman. Mining of Massive Datasets. New York, NY: Cambridge UP, 2012.

Ringrose, P., & Bentley, M. (2015). Reservoir model design: A practitioner's guide. Dordrecht: Springer.

Saleh, S. J., & de Bruin, J. A. (2000, January 1). *AVO Attribute Extraction Via Principal Component Analysis*. Society of Exploration Geophysicists.

Sattar, A., Iqbal, G. M., and Buchwalter, J. L. (2008). Practical Enhanced Reservoir Engineering Assisted with Simulation Software. PenWell Corp.

Scheevel, J. R., & Payrazyan, K. (2001, February 1). *Principal Component Analysis Applied to 3D Seismic Data for Reservoir Property Estimation*. Society of Petroleum Engineers. doi:10.2118/69739-PA

Schlotzhauer, Sandra (2007). Elementary Statistics Using JMP (SAS Press) (PAP/CDR ed.). Cary, NC: SAS Institute. pp. 166–169. ISBN 1-599-94375-1.

Shlens, J. (2003, March 25). “A Tutorial on Principal Component Analysis” Web

Shojaei, H., & Tajer, E. S. (2013, August 20). *Analytical Solution of Transient Multiphase Flow to a Horizontal Well With Multiple Hydraulic Fractures*. Society of Petroleum Engineers. doi:10.2118/165703-MS

Spivey, J. P., Frantz, J. H., Williamson, J. R., & Sawyer, W. K. (2001, January 1). *Applications of the Transient Hyperbolic Exponent*. Society of Petroleum Engineers. doi:10.2118/71038-MS

Srinivasan, K., & Ertekin, T. (2008, January 1). *Development and Testing of an Expert System for Coalbed Methane Reservoirs Using Artificial Neural Networks*. Society of Petroleum Engineers. doi:10.2118/119935-MS

Stalgorova, K., & Mattar, L. (2013, July 31). *Analytical Model for Unconventional Multifractured Composite Systems*. Society of Petroleum Engineers. doi:10.2118/162516-PA

Tingdahl, K., & Hemstra, N. (2003, January 1). *Estimating Fault-attribute Orientation With Gradient Analysis, Principal Component Analysis And the Localized Hough-transform*. Society of Exploration Geophysicists.

“United States Geological Survey (USGS), 2013, USGS Releases New Oil and Gas Assessment for Bakken and Three Forks Formation. USGS Webpost,” http://www.usgs.gov/blogs/features/usgs_top_story/usgs-releases-new-oil-and-gas-assessment-for-bakken-and-three-forks-formations/”, May 2

Valko, P. P., & Lee, W. J. (2010, January 1). *A Better Way To Forecast Production From Unconventional Gas Wells*. Society of Petroleum Engineers. doi:10.2118/134231-MS

Whitson, C. H., & Sunjerga, S. (2012, January 1). *PVT in Liquid-Rich Shale Reservoirs*. Society of Petroleum Engineers. doi:10.2118/155499-MS

Whitson, C. H., & Sunjerga, S. (2012, January 1). *PVT in Liquid-Rich Shale Reservoirs*. Society of Petroleum Engineers. doi:10.2118/155499-MS

Yu, S. (2014, April 1). *A New Methodology to Predict Condensate Production in Tight/Shale Retrograde Gas Reservoirs*. Society of Petroleum Engineers. doi:10.2118/168964-MS

Yu, S., & Miocevic, D. J. (2013, August 12). *An Improved Method to Obtain Reliable Production and EUR Prediction for Wells with Short Production History in Tight/Shale Reservoirs*. Society of Petroleum Engineers. doi:10.1190/URTEC2013-003

Yu, W., & Sepehrnoori, K. (2014, August 28). *Optimization of Well Spacing for Bakken Tight Oil Reservoirs*. Society of Petroleum Engineers. doi:10.15530/urtec-2014-1922108

Appendix

Matlab Code:

Main File: This file takes the gas rate, cumulative and cgr data for simulated case, performs PCA, recreates the data. This step helps identify the number of principal components required for forecasting the gas rate, cumulative gas and cgr data. Following this step, this file performs multiple linear regression, which calculates

the principal components for new wells based on the reservoir properties. Using this, the code automatically calculates the final rates for new well with the given reservoir properties.

This code is automated and reusable, as the only thing that needs to be changed are the input rate data and reservoir parameters. The other functions are given as follows.

```
%% Initialization
clear; close all; clc

%% Load Data

gas_rate_data = load('data_gas_rate.txt');
gas_cum_data = load('data_gas_cum.txt');
cgr_data = load('data_cgr.txt');
res_pro = load('multiple_regression.txt');
res_pro = [ res_pro(1:332, 1:7); res_pro(334:335,1:7)];

gas_rate_data = [gas_rate_data(1:199,1:333)
gas_rate_data(1:199,335:336)];
gas_cum_data = [gas_cum_data(1:199,1:333) gas_cum_data(1:199,335:336)];
cgr_data = [cgr_data(1:199,1:333) cgr_data(1:199,335:336)];

Time = gas_rate_data(:,1);
G_Rates= gas_rate_data(:,2:size(gas_rate_data,2));
G_cum= gas_cum_data(:,2:size(gas_cum_data,2));
Cgr = cgr_data(:,2:size(cgr_data,2));

fprintf(['Plotting the relevant rates ' ...
        '\n'...
        ' Gas Rates, Cumulatives and CGR (Press enter to
continue)\n']);
%plot gas rate
Figure1 = figure;hold on;
subplot (2,3,1)
plot(Time,G_Rates);
ylabel('Gas Rates (ft3/d)'); xlabel('Time(days)');
head = title(' Gas Rate for Simulated Wells');
SetPos(head);
xlim([0 10900]);

%plot cumulative gas
```

```

subplot (2,3,2)
plot(Time,G_cum);
ylabel('Cum Gas (ft3)'); xlabel('Time(days)');
head = title('CUM. Gas for Simulated Wells');
SetPos(head);
xlim([0 10900]);

%plot CGR
subplot (2,3,3)
plot(Time,Cgr);
ylabel('CGR(STB/MMScf)'); xlabel('Time(days)');
head = title('CGR for Simulated Wells');
SetPos(head);
set(gca,'fontsize', 18);
xlim([0 10900]);

hold off;
% divide data by 70/15/15

[train_rate,val_rate,test_rate]= splitdata(G_Rates);
[train_cum,val_cum,test_cum]= splitdata(G_cum);
[train_cgr,val_cgr,test_cgr]= splitdata(Cgr);

fprintf(['Normalize all the rates and values' ...
        '(Press enter to continue)\n']);

% use normalize function to normalize each element in the original
matrix
% by using the max values

normalized_gasrate = (normalize(G_Rates));
normalized_cum = (normalize(G_cum));
normalized_cgr = (normalize(Cgr));

subplot (2,3,4)
plot(Time,normalized_gasrate);
ylabel('Normalized Gas Rates '); xlabel('Time(days)');
head = title(' Gas Rate for Simulated Wells');
xlim([0 10900]);
SetPos(head);

%plot cumulative gas
subplot (2,3,5)
plot(Time,normalized_cum);
ylabel('Normalized Cum Gas'); xlabel('Time(days)');
head = title('CUM. Gas for Simulated Wells');
xlim([0 10900]);
SetPos(head);

```

```

%plot CGR
    subplot (2,3,6)
    plot(Time,normalized_cgr);
    ylabel('Normalized CGR'); xlabel('Time(days)');
    head = title('CGR for Simulated Wells');
    set(gca,'fontsize', 18);
    xlim([0 10900]);
    SetPos(head);

[train_rate_norm,val_rate_norm,test_rate_norm]=
splitdata(normalized_gasrate);
[train_cum_norm,val_cum_norm,test_cum_norm]= splitdata(normalized_cum);
[train_cgr_norm,val_cgr_norm,test_cgr_norm]= splitdata(normalized_cgr);

fprintf(['Perform SVD to Calculate the Principal Component Scores '
...
' and Principal Component coefficeints(Press enter to
continue)\n']);

[U_nrate, S_nrate, V_nrate] = svd (train_rate_norm');
[U_ncum, S_ncum, V_ncum] = svd (train_cum_norm');
[U_ncgr, S_ncgr, V_ncgr] = svd (train_cgr_norm');

[U_rate, S_rate, V_rate] = svd (train_rate');
[U_cum, S_cum, V_cum] = svd (train_cum');
[U_cgr, S_cgr, V_cgr] = svd (train_cgr');

fprintf(['Eigenvalues weight and Error for normalized data ' ...
' (Press enter to continue)\n']);

cumnMaterr = weighteigen(S_nrate,S_ncum,S_ncgr,'Eigenvalues for Norm.
Rate','Eigenvalues for Norm. Cum','Eigenvalues for Norm. CGR');
cumMat = weighteigen(S_rate,S_cum,S_cgr,'Eigenvalues for
Rate','Eigenvalues for Cum','Eigenvalues for CGR');

plotPCA(-V_nrate, 'Loadings for Norm. Rate');
plotPCA(-V_ncum, 'Loadings for Norm. Cum');
plotPCA(-V_ncgr, 'Loadings for Norm. CGR');

plotPCA(-V_rate, 'Loadings for Rate');
plotPCA(-V_cum, 'Loadings for Cum');
plotPCA(-V_cgr, 'Loadings for CGR');

%Rates

Error_Rate = cell2mat(recreate(train_rate','Rate'));

```

```

Error_CUM= cell2mat(recreate(train_cum','CUM'));
Error_CGR = cell2mat(recreate(train_cgr','CGR'));

Error_NRate = cell2mat(recreate(train_rate_norm','CGR'));
Error_NCUM= cell2mat(recreate(train_cum_norm','CUM'));
Error_NCGR = cell2mat(recreate(train_cgr_norm','Rate'));

%Multiple Regression Validation Section
train_ratePC = U_rate *S_rate;
train_cumPC = U_cum *S_cum;
train_cgrPC = U_cgr *S_cgr;

%for this section splitting is done by 70 15 15 check splitdata for more
%details

%this does multiple regression for each principal component
PC1_PRED_Gas = mult_reg_PC (train_ratePC,res_pro,1);
PC1_PRED_Cum = mult_reg_PC (train_cumPC,res_pro,1);
PC1_PRED_Cgr = mult_reg_PC (train_cgrPC,res_pro,1);

PC2_PRED_Gas = mult_reg_PC (train_ratePC,res_pro,2);
PC2_PRED_Cum = mult_reg_PC (train_cumPC,res_pro,2);
PC2_PRED_Cgr = mult_reg_PC (train_cgrPC,res_pro,2);

PC3_PRED_Gas = mult_reg_PC (train_ratePC,res_pro,3);
PC3_PRED_Cum = mult_reg_PC (train_cumPC,res_pro,3);
PC3_PRED_Cgr = mult_reg_PC (train_cgrPC,res_pro,3);

PC4_PRED_Gas = mult_reg_PC (train_ratePC,res_pro,4);
PC4_PRED_Cum = mult_reg_PC (train_cumPC,res_pro,4);
PC4_PRED_Cgr = mult_reg_PC (train_cgrPC,res_pro,4);

val_rate =[val_rate test_rate];
val_cum =[val_cum test_cum];
val_cgr =[val_cgr test_cgr];

PC_Gas={PC1_PRED_Gas PC2_PRED_Gas PC3_PRED_Gas PC4_PRED_Gas};
PC_Cum={PC1_PRED_Cum PC2_PRED_Cum PC3_PRED_Cum PC4_PRED_Cum};
PC_Cgr={PC1_PRED_Cgr PC2_PRED_Cgr PC3_PRED_Cgr PC4_PRED_Cgr};

[mr_gas_pc1, mr_gas_pc2, mr_gas_pc3, mr_gas_pc4]= mult_reg_Rates(
PC_Gas,V_rate,2 );
[mr_cum_pc1 ,mr_cum_pc2, mr_cum_pc3 ,mr_cum_pc4] = mult_reg_Rates(
PC_Cum,V_cum,2 );
[mr_cgr_pc1 ,mr_cgr_pc2, mr_cgr_pc3, mr_cgr_pc4] = mult_reg_Rates(
PC_Cgr,V_cgr,2 );

```

```

ErrorMatRate = validation_Err( val_rate, mr_gas_pc1, mr_gas_pc2,
mr_gas_pc3, mr_gas_pc4, 1 );
ErrorMatCum = validation_Err( val_cum, mr_cum_pc1, mr_cum_pc2,
mr_cum_pc3, mr_cum_pc4, 1 );
ErrorMatCgr = validation_Err( val_cgr, mr_cgr_pc1, mr_cgr_pc2,
mr_cgr_pc3, mr_cgr_pc4, 1 );

err_sctter_fcn( ErrorMatRate,ErrorMatCum,ErrorMatCgr, ' (Lower 95% CI)'
);

ll_Rate = [min(ErrorMatRate) mean(ErrorMatRate) max(ErrorMatRate)];
ll_Cum = [min(ErrorMatCum) mean(ErrorMatCum) max(ErrorMatCum)];
ll_Cgr = [min(ErrorMatCgr) mean(ErrorMatCgr) max(ErrorMatCgr)];

ErrorMatRate = validation_Err( val_rate, mr_gas_pc1, mr_gas_pc2,
mr_gas_pc3, mr_gas_pc4, 2 );
ErrorMatCum = validation_Err( val_cum, mr_cum_pc1, mr_cum_pc2,
mr_cum_pc3, mr_cum_pc4, 2 );
ErrorMatCgr = validation_Err( val_cgr, mr_cgr_pc1, mr_cgr_pc2,
mr_cgr_pc3, mr_cgr_pc4, 2 );

err_sctter_fcn( ErrorMatRate,ErrorMatCum,ErrorMatCgr, ' (Predicted)' );

pv_Rate = [min(ErrorMatRate) mean(ErrorMatRate) max(ErrorMatRate)];
pv_Cum = [min(ErrorMatCum) mean(ErrorMatCum) max(ErrorMatCum)];
pv_Cgr = [min(ErrorMatCgr) mean(ErrorMatCgr) max(ErrorMatCgr)];

ErrorMatRate = validation_Err( val_rate, mr_gas_pc1, mr_gas_pc2,
mr_gas_pc3, mr_gas_pc4, 3 );
ErrorMatCum = validation_Err( val_cum, mr_cum_pc1, mr_cum_pc2,
mr_cum_pc3, mr_cum_pc4, 3 );
ErrorMatCgr = validation_Err( val_cgr, mr_cgr_pc1, mr_cgr_pc2,
mr_cgr_pc3, mr_cgr_pc4, 3 );

ul_Rate = [min(ErrorMatRate) mean(ErrorMatRate) max(ErrorMatRate)];
ul_Cum = [min(ErrorMatCum) mean(ErrorMatCum) max(ErrorMatCum)];
ul_Cgr = [min(ErrorMatCgr) mean(ErrorMatCgr) max(ErrorMatCgr)];

err_sctter_fcn( ErrorMatRate,ErrorMatCum,ErrorMatCgr, ' (Upper 95% CI)'
);

plotVal( val_rate,mr_gas_pc1, mr_gas_pc2, mr_gas_pc3)
plotVal( val_cum,mr_cum_pc1, mr_cum_pc2, mr_cum_pc3);

```

```
plotVal( val_cgr,mr_cgr_pc1, mr_cgr_pc2, mr_cgr_pc3);
```

Function Split data:

This function is used in the main file to split the input data into training, validation and test set

```
function [ train, val, test ] = splitdata( G_Rates )
%SPLITDATA Summary of this function goes here
% Detailed explanation goes here

num_points = size(G_Rates,2);
split_test = round(num_points*0.7);
split_val = round(num_points*0.85);
train = G_Rates(1:size(G_Rates,1), 1:split_test);
val = G_Rates(1:size(G_Rates,1), (split_test+1):split_val );
test = G_Rates(1:size(G_Rates,1), (split_val+1):num_points );

end
```

Function Normalize: This function is used to normalize the input data

```
function [ output_mat ] = normalize( input_mat )
%NORMALIZE Summary of this function goes here
% Detailed explanation goes here

maxVal = max(max(input_mat));
for j = 1:size(input_mat,2)
    for i = 1:size(input_mat,1)
        input_mat(i,j) = input_mat(i,j)/maxVal;
    end
end

output_mat = input_mat;

end
```

Function weighteigen:

This function calculates and plots the weighted eigenvalues for the simulated data.

```
function [cumMat] = weighteigen( S1,S2,S3, tt11,tt12,tt13 )
%WEIGHTEIGEN Summary of this function goes here
% Detailed explanation goes here

%figure; hold on;

%find the diagonal matrix
```

```

d1 = diag(S1);
d2 = diag(S2);
d3 = diag(S3);

pc1 = 100 * d1 / sum(d1);
pc2 = 100 * d2 / sum(d2);
pc3 = 100 * d3 / sum(d3);

x =1:1:199;

cumMat1 = pc1;
cumMat2 = pc2;
cumMat3 = pc3;

for i = 2:size(pc1,1)
    cumMat1(i,1) = pc1(i,1)+cumMat1((i-1),1);
    cumMat2(i,1) = pc2(i,1)+cumMat2((i-1),1);
    cumMat3(i,1) = pc3(i,1)+cumMat3((i-1),1);
end

cumMat = cumMat1;
figure;
subplot (2,3,1)
bar(x, pc1, 'barwidth', 1);
hold on;
plot(x,cumMat1,'mo','LineWidth',1,'MarkerSize',2);
title(ttl1);
legend('Eigenvalues', 'Cumulative Eigenvalues');
axis([0,199,0,100])

hold off;

subplot (2,3,2)
bar(x, pc2, 'barwidth', 1);
hold on;
plot(x,cumMat2,'mo','LineWidth',1,'MarkerSize',2);
legend('Eigenvalues', 'Cumulative Eigenvalues');
title(ttl2);
axis([0,199,0,100])
hold off;

subplot (2,3,3)
bar(x, pc3, 'barwidth', 1);
hold on;
plot(x,cumMat3,'mo','LineWidth',1,'MarkerSize',2);
legend('Eigenvalues', 'Cumulative Eigenvalues');
title(ttl3);
axis([0,199,0,100])
hold off;

```

```
end
```

Function recreate:

This function recreates the original data and checks the error in recreation.

```
function [errorMat] = recreate( Cgr,heading )
%RECREATE the heading goes to create the suptitle for all the graphs
%combined
k = randi([1, size(Cgr,1)]);

[U,S,V] = svd(Cgr);

loadings =zeros(size(V,1),size(V,2));

product = U*S;

%1 PRINCIPAL COMPONENT
for i = 1:size(V,1)
    loadings(i,1) = V(i,1);
end
firstPC = (product* loadings)';

figure ; hold on;
subplot(2,4,1);
plot(1:1:size(Cgr,2),firstPC(:,k),'-','LineWidth',1);
hold on;
plot(1:1:size(Cgr,2),(Cgr(k,:))','-','LineWidth',1);
legend('1 PC','Original','Location','northwest');
title('Recreated Data with 1 PC');
xlabel('Time Steps');
%Calculate Error
errPC1 = errorCalc(Cgr,firstPC)';

%TWO PRINCIPAL COMPONENTS
for i = 1:size(V,1)
    loadings(i,2) = V(i,2);
end
firstPC = (product* loadings)';

subplot(2,4,2);
plot(1:1:size(Cgr,2),firstPC(:,k),'-','LineWidth',1);
hold on;
plot(1:1:size(Cgr,2),(Cgr(k,:))','-','LineWidth',1);
legend('2 PC','Original','Location','northwest');
title('Recreated Data with 2 PC');
```

```

xlabel('Time Steps');
hold off;
errPC2 = errorCalc(Cgr,firstPC');

%3 pc
for i = 1:size(V,1)
    loadings(i,3) = V(i,3);
end
firstPC = (product* loadings)';

subplot(2,4,3);
plot(1:1:size(Cgr,2),firstPC(:,k),'-','LineWidth',1);
hold on;
plot(1:1:size(Cgr,2),(Cgr(k,:))','-','LineWidth',1);
legend('3 PC','Original','Location','northwest');
title('Recreated Data with 3 PC');
xlabel('Time Steps');
hold off;
errPC3 = errorCalc(Cgr,firstPC');

%for 4 principal components
for i = 1:size(V,1)
    loadings(i,4) = V(i,4);
end
firstPC = (product* loadings)';

subplot(2,4,4);
plot(1:1:size(Cgr,2),firstPC(:,k),'-','LineWidth',1);
hold on;
plot(1:1:size(Cgr,2),(Cgr(k,:))','-','LineWidth',1);
legend('4 PC','Original','Location','northwest');
title('Recreated Data with 4 PC');
xlabel('Time Steps');
hold off;
errPC4 = errorCalc(Cgr,firstPC');

Well_Number = num2str(k);
Graph_heading = strcat({'Recreated '},{heading},{' data using limited
PC for Well '}, Well_Number) ;
suptitle (Graph_heading);

errorMat = {errPC1 errPC2 errPC3 errPC4};

end

```

Function mult_reg_PC:

This function is invoked in the validation stage/prediction stage to predict the production from new wells. This function calculates the principal component for new well using multiple linear regression. In this study, we use cross validation to check the effectiveness of the model generated from PCA.

```
function [ PC1_PRED_Gas ] = mult_reg_PC( train_ratePC, res_pro, i )

train_respro = res_pro (1:round(size(res_pro,1)*0.7), 1:7);
val_respro = res_pro ((round(size(res_pro,1)*0.7)+1):size(res_pro,1),
1:7);

mult_reg_rate =fitlm (train_respro,
train_ratePC(1:size(train_ratePC,1), i));
[PC1_PRED_Gas, PC1_PRED_Gas_CI] = predict(mult_reg_rate, val_respro);
PC1_PRED_Gas = [PC1_PRED_Gas_CI(1:size(val_respro,1), 1) PC1_PRED_Gas
PC1_PRED_Gas_CI(1:size(val_respro,1), 2)];

end
```

Function mult_reg_Rates:

This function calculates the rates of new using multiple regression.

```
function [ A B C D ] = mult_reg_Rates( PC_Gas,V_rate,i )
%MULT_REG_RATES Summary of this function goes here
% Detailed explanation goes here

PC1 = cell2mat(PC_Gas(1));
%l1 = lower 95% , pv = predicted value ul= upper limit
l1 = [PC1(1:size(PC1,1),1) zeros(size(PC1,1), (size(V_rate,1)-1))];
pv =[PC1(1:size(PC1,1),2) zeros(size(PC1,1), (size(V_rate,1)-1))];
ul =[PC1(1:size(PC1,1),3) zeros(size(PC1,1), (size(V_rate,1)-1))];

rate_l1_pc1 = (l1*V_rate)';
rate_pv_pc1 = (pv*V_rate)';
rate_ul_pc1 = (ul*V_rate)';

A = { rate_l1_pc1 rate_pv_pc1 rate_ul_pc1};

PC2 = cell2mat(PC_Gas(2));

l1 = [PC1(1:size(PC1,1),1) PC2(1:size(PC1,1),1) zeros(size(PC1,1),
(size(V_rate,1)-2))];
pv =[PC1(1:size(PC1,1),2) PC2(1:size(PC1,1),2) zeros(size(PC1,1),
(size(V_rate,1)-2))];
```

```

ul =[PC1(1:size(PC1,1),3) PC2(1:size(PC1,1),3) zeros(size(PC1,1),
(size(V_rate,1)-2))];

rate_ll_pc2 = (ll*V_rate)';
rate_pv_pc2 = (pv*V_rate)';
rate_ul_pc2 = (ul*V_rate)';

B = { rate_ll_pc2 rate_pv_pc2 rate_ul_pc2};

PC3 = cell2mat(PC_Gas(3));

ll = [PC1(1:size(PC1,1),1) PC2(1:size(PC1,1),1) PC3(1:size(PC1,1),1)
zeros(size(PC1,1), (size(V_rate,1)-3))];
pv = [PC1(1:size(PC1,1),2) PC2(1:size(PC1,1),2) PC3(1:size(PC1,1),2)
zeros(size(PC1,1), (size(V_rate,1)-3))];
ul =[PC1(1:size(PC1,1),3) PC2(1:size(PC1,1),3) PC3(1:size(PC1,1),3)
zeros(size(PC1,1), (size(V_rate,1)-3))];

rate_ll_pc3 = (ll*V_rate)';
rate_pv_pc3 = (pv*V_rate)';
rate_ul_pc3 = (ul*V_rate)';

C = { rate_ll_pc3 rate_pv_pc3 rate_ul_pc3};

PC4 = cell2mat(PC_Gas(4));
ll = [PC1(1:size(PC1,1),1) PC2(1:size(PC1,1),1) PC3(1:size(PC1,1),1)
PC4(1:size(PC1,1),1) zeros(size(PC1,1), (size(V_rate,1)-4))];
pv = [PC1(1:size(PC1,1),2) PC2(1:size(PC1,1),2) PC3(1:size(PC1,1),2)
PC4(1:size(PC1,1),2) zeros(size(PC1,1), (size(V_rate,1)-4))];
ul =[PC1(1:size(PC1,1),3) PC2(1:size(PC1,1),3) PC3(1:size(PC1,1),3)
PC4(1:size(PC1,1),3) zeros(size(PC1,1), (size(V_rate,1)-4))];

rate_ll_pc4 = (ll*V_rate)';
rate_pv_pc4 = (pv*V_rate)';
rate_ul_pc4 = (ul*V_rate)';

D = { rate_ll_pc4 rate_pv_pc4 rate_ul_pc4};

```

end

Function validation_Err:

This function calculates the validation error for the validation set

```

function ErrorMatRate = validation_Err( val_rate, mr_gas_pc1,
mr_gas_pc2, mr_gas_pc3, mr_gas_pc4, i )

```

```

error_gas1 = errorCalc(val_rate', (cell2mat(mr_gas_pc1(i)))');
error_gas2 = errorCalc(val_rate', (cell2mat(mr_gas_pc2(i)))');
error_gas3 = errorCalc(val_rate', (cell2mat(mr_gas_pc3(i)))');
error_gas4 = errorCalc(val_rate', (cell2mat(mr_gas_pc4(i)))');

```

```
ErrorMatRate = [error_gas1 error_gas2 error_gas3 error_gas4];
```

```
end
```

Function `err_sctter_fcn`:

This function calculates scatter plot for the error for both validation and recreation.

```
function err_sctter_fcn( ErrorMatRate,ErrorMatCum,ErrorMatCgr,head )
%ERR_SCTTER_FCN Summary of this function goes here
% Detailed explanation goes here

figure;
for i=1:4
hold on
subplot(2,3,1);
ret = scatter_notation(i);
scatter(1:size(ErrorMatRate,1),ErrorMatRate(:,i),ret);
end
legend ('Error 1 PC', 'Error 2 PC', 'Error 3 PC', 'Error 4 PC')
xlabel('Well Number', 'FontSize', 22)
ylabel('% Error', 'FontSize', 22);
str = strcat('Error (Gas Rate) ', head);
title(str, 'FontSize', 25)
xlim([0 (size(ErrorMatRate,1))])
hold off;

for i=1:4
hold on
subplot(2,3,2);
ret = scatter_notation(i);
scatter(1:size(ErrorMatRate,1),ErrorMatCum(:,i),ret);
end
legend ('Error 1 PC', 'Error 2 PC', 'Error 3 PC', 'Error 4 PC')
xlabel('Well Number', 'FontSize', 22)
ylabel('% Error', 'FontSize', 22);
title('Error (Cumulative Gas)', 'FontSize', 25)
xlim([0 (size(ErrorMatRate,1))])
hold off;

for i=1:4
hold on
subplot(2,3,3);
ret = scatter_notation(i);
scatter(1:size(ErrorMatRate,1),ErrorMatCgr(:,i),ret);
end
legend ('Error 1 PC', 'Error 2 PC', 'Error 3 PC', 'Error 4 PC')
xlabel('Well Number', 'FontSize', 22)
ylabel('% Error', 'FontSize', 22);
title('Error (CGR)', 'FontSize', 25)
```

```
xlim([0 (size(ErrorMatRate,1))])
hold off;

end
```

Function plotVal:

This plots the validation and confidence interval for the validation set

```
function plotVal( val_rate,mr_gas_pc1, mr_gas_pc2, mr_gas_pc3)
%PLOTVAL Summary of this function goes here
% Detailed explanation goes here

k = randi([1, size(val_rate,2)]);
%k=31;

figure ; hold on;
subplot(2,3,1);

l1 =cell2mat(mr_gas_pc1(1));
pval =cell2mat(mr_gas_pc1(2));
ul = cell2mat(mr_gas_pc1(3));

Time =((10958/199)*(1:1:size(val_rate,1)))';

plot(Time,l1(:,k),'--m','LineWidth',1);
hold on
plot(Time,pval(:,k),'-k','LineWidth',1);
hold on
plot(Time,ul(:,k),'--m','LineWidth',1);
hold on
plot(Time,val_rate(:,k),'--g','LineWidth',1);
legend('Lower 95%','Predicted','Upper 95%','Simulated');
title('1 PC');
xlabel('Time (days)');
xlim([0 10958]);

subplot(2,3,2);

l1 =cell2mat(mr_gas_pc2(1));
pval =cell2mat(mr_gas_pc2(2));
ul = cell2mat(mr_gas_pc2(3));

plot(Time,l1(:,k),'--m','LineWidth',1);
hold on
plot(Time,pval(:,k),'-k','LineWidth',1);
hold on
plot(Time,ul(:,k),'--m','LineWidth',1);
hold on
plot(Time,val_rate(:,k),'--g','LineWidth',1);
```

```

legend('Lower 95%', 'Predicted', 'Upper 95%', 'Simulated');

title('2 PC');
xlabel('Time (days)');
xlim([0 10958]);

subplot(2,3,3);

l1 =cell2mat(mr_gas_pc3(1));
pval =cell2mat(mr_gas_pc3(2));
ul = cell2mat(mr_gas_pc3(3));

plot(Time,l1(:,k), '--m', 'LineWidth',1);
hold on
plot(Time,pval(:,k), '-k', 'LineWidth',1);
hold on
plot(Time,ul(:,k), '--m', 'LineWidth',1);
hold on
plot(Time,val_rate(:,k), '--g', 'LineWidth',1);
legend('Lower 95%', 'Predicted', 'Upper 95%', 'Simulated');

title('3 PC');
xlabel('Time (days)');
xlim([0 10958]);

Well_Number = num2str(k);
Graph_heading = strcat({'Cross Validation '},{' for Well '},
Well_Number) ;
suptitle (Graph_heading);

end

```

Function ValScoCal:

For field case ValScoCal function is used to calculate the principal components of by using least sum of squares fit.

```
function [ outputMat ] = ValScoCal( Cgr )
```

```
%This function calculates the principal component scores by using
```

```

%leave-one-out cross validation method. 'for loops' remove each row
(the
%cgr_tru), and calculates the PC. That PC is in turn used to evaluate
the
%truncated row of value by using least squares fit

x_PC4 = zeros(size(Cgr,1),size(Cgr,1));

for i = 1 :size(Cgr,1)
    cgr_tru = [Cgr(1:(i-1),:); Cgr((i+1):size(Cgr,1),:)];

    [U S V] = svd(Cgr);
    PC1 = V(:,1);PC2 =V(:,2);PC3 = V(:,3);PC4 = V(:,4);
    predict = (Cgr(i,:))' ;
    fun = @(x,PC1) (x(1)*PC1+x(2)*PC2+x(3)*PC3+x(4)*PC4);
    x0 = [-20, -5 -4 -1];
    x_INT = lsqcurvefit(fun,x0,PC1,predict);
    x_PC4(i,1) =x_INT(1,1);
    x_PC4(i,2) =x_INT(1,2);
    x_PC4(i,3) =x_INT(1,3);
    x_PC4(i,4) =x_INT(1,4);

end

%this yields the cell matrix with pricipal component scores when 1, 2,
3
%and 4 PC coefficeints are used. Note that since they are individually
%calculated X_PC1 has different pc score compared to others and so on

    outputMat = x_PC4;

end

```